Proceedings of the 16th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty *(CJS-2013)* Mariánské Lázně, Czech Republic

Václav Kratochvíl, Jiřina Vejnarová
(editors)

September 19-22, 2013

Published by:
Faculty of Management
University of Economics
Jindřichův Hradec
Czech Republic
http://cjs.fm.vse.cz

Organized and sposored by:
Faculty of Management, University of Economics, Jindřichův Hradec
Centre of Excellence IT4Innovations, Division Univ. of Ostrava, IRAFM
Inst. of Information Theory and Automation, Acad. of Sciences of ČR
Czech Society for Cybernetics and Informatics

Credits:
Cover design: Jiří Přibil
Logo photo: provided by Mariánské Lázně City Hall Council
Editors: Václav Kratochvíl, Jiřina Vejnarová
LaTeX editor: Václav Kratochvíl

Printed by Typos, tiskařské závody, s. r. o., Plzeň — September 2013

ISBN 978-80-245-1950-0

Steering Commitee Chairs:
Hiroaki Ishii, Kwansei Gakuin University, Emeritus Professor of Osaka University
Milan Vlach, Charles University in Prague

Seminar Chair:
Radim Jiroušek, University of Economics, Faculty of Management, Jindřichův Hradec

Programme Committee:
Roman Barták, Charles University in Prague
Ondřej Čepek, Charles University in Prague
Milan Daniel, Academy of Sciences of the Czech Republic, Prague
Martin Gavalec, University of Hradec Králové
Takeshi Itoh, Tohoku University
Masahiro Inuiguchi, (Co-Chair) Osaka University
Vilém Novák, University of Ostrava
Jaroslav Ramík, (Co-Chair) Silesian University
Shogo Shiode, Kobe Gakuin University
Shao Chin Sung, Aoyama Gakuin University
Tetsuzo Tanino, Osaka University
Junzo Watada, Waseda University

Organizing Committee:
Vladislav Bína, (Chair) Univ. of Economics, Faculty of Management, Jindrichuv Hradec
Václav Kratochvíl, Academy of Sciences of the Czech Republic, Prague
Iva Krejčová, University of Economics, Faculty of Management, Jindrichuv Hradec
Lucie Váchová, University of Economics, Faculty of Management, Jindrichuv Hradec
Jiřina Vejnarová, Academy of Sciences of the Czech Republic, Prague

This proceedings contains 28 contributions accepted for presentation at the 16th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty that was held in Mariánské Lázně, Czech Republic, September 19-23, 2013. This conference is organized at the occasion of 60th anniversary of the University of Economics, Prague, and the organizers have a pleasure to thank all the organizing institutions and sponsors, especially the Czech Society for Cybernetics and Informatics http://www.cski.cz, and Centre of Excellence IT4Innovations http://www.it4i.cz/en/.

Conference Program

Friday, September 20, 2013

Session 1 *(8:40–10:20)*

A Promotion Sale Problem for a Perihsable Product
Hiroaki Ishii 1

On Modeling Planning Problems: Experience From The Petrobras Challenge
Roman Barták, Neng-Fa Zhou 7

A Generalization of the Noisy-Or Model to Multivalued Parent Variables
Jiří Vomlel 19

Fuzzy Relational Compositions Based on Generalized Intermediate Quantifiers
Martin Štěpnička, Michal Holčapek 29

Session 2 *(10:50–12:30)*

Mining Information from Time Series using Fuzzy Natural Logic
Vilém Novák 43

A Karhunen-Loeve Transform Approach to Robust Color Image Segmentation
Junzo Watada, Chenxue Wang 51

perator of composition for continuous copulas
Vladislav Bína 65

On Causal Compositional Models: preliminaries
Radim Jiroušek 73

Session 3 *(14:00–15:40)*

Portfolio Selection with Interactive Fuzzy Coefficients
Masahiro Inuiguchi 83

Decision Making in MAX-PROD Algebra
Hana Tomášková, Martin Gavalec 91

Decision Making Process in Agent-based Roundabout Simulation
Richard Cimler 99

Location, Allocation and Capacity Design for Emergency Medical Centers
Young Dae Ko, Byung Duk Song, Hark Hwang 107

Session 4 *(16:10–17:50)*

Credal Networks and Compositional Models: Preliminary Considerations
Jiřina Vejnarová 119

Hierarchical Clustering Using Proximity Measures Based on Discernibility of Clusters
Yoshifumi Kusunoki, Tetsuzo Tanino 129

Rule Based Systems for Monitoring and Controller Re-adaptation
Jana Nowaková, Miroslav Pokorný 137

Triads of Implicational, Double-Implicational and Equivalency Data-Mining Quantifiers
Jiří Ivánek 145

Saturday, September 21, 2013

Session 5 *(8:40–10:20)*

Three-Person Bargaining Problems
Milan Vlach 153

Role of Pavlov-like Strategy on Emergence of Cooperation in Demographic Donor-Recipient Game
Tsuneyuki Namekata, Yoko Namekata 163

Additively and Multiplicatively Transitive Fuzzy Relations in Ranking of Alternatives
Jaroslav Ramík, Petr Korviny 171

An Interpretation of Conflicting Parts of Belief Functions on Two-Element Frame of Discernment
Milan Daniel 187

Session 6 *(10:50–12:30)*

The Linear Space of Transferable Utility Games
Tetsuzo Tanino, Yoshifumi Kusunoki 197

F-Transform for Image Reduction
Irina Perfilieva, Petr Hurtík 205

Discrete Representation of Stationary Random Processes Using Fuzzy Transform
Michal Holčapek, Vilém Novák, Irina Perfilieva 215

UAV Environment Recognition for Better Horizon Detection with Usage of FUZZY IS
David Novák, Petr Čermák 223

Session 7 *(14:00–15:40)*

On Stable Solutions of Atomic Low-Carbon Growth Partnershon
Mayuko Miyata, Shao-Chin Sung 231

Notes on Solving the Pipelined Set Cover Problem
Václav Lín 239

Fuzzy Rule Based Behavioral Models in Boarding a Train
Petr Čermák, Martin Morawiec 245

On the Practical Issues of Document Context Extraction And Similarity Measuring
Jiří Přibil 253

Index of Authors 263

A Promotion Sale Problem for a Perihsable Product

Hiroaki Ishii
Department od Mathematical Sciences, School of Science and Technology
Kwansei Gakuin University
e-mail:ishiihiroaki@kwansei.ac.jp

Abstract

In the field of inventory problems, several researchers have been interested in inventory control for a perishable product such as blood, fresh fruit, milk, film etc. Here we introduce a promotion sale in order to reduce the outdating quantity. We consider the following model. (1) Single perishable product with two life time period is considered. That is, for a period, an amount of old one (remaining life time period is 1) in the stock is given as x_2 and under this condition we should determine an ordering quantity x_1 of the fresh one(remaining life time period is 2). (2) Ordering takes a place at the start of the period. The unit purchasing cost of the product is c. (3)Issuing policy is LIFO, that is, customer buys fresh one first and if fresh one is sold out, only some percent of customers overflowed from purchase of the fresh one buy the old one. We assume that this percentage is at most $100q$. Unit price of the fresh one is r_1 and that of old one is r_2. (4) Prominent feature of our model is a promotion sale, that is, we sell a set of products consisting of the fresh and old one and it is sold at the discount unit price r_B less than the sum,$r_1 + r_2$. We assume at most $100p$ percent of the customers puchsing the fresh one accept the promotion set, that is, buy the old one with the fresh one at the same time where we assume that $p > q$. (5) The old one that is not purchased by the customer and remained outdates and is discarded at the unit cost θ. While the fresh one not purchsed by the customer is stocked at the unit cost h. (6) The demand D of the customer is a nonnegative random variable. Its cummulative distribution function and density function are $F(D)$ and $f(D)$ respectively where $F(0) = f(0) = 0$.

Under the above setting, we calculate an expected profit function $E(x_1)$. $E(x_1)$ is divided two parts, that is, $0 \leq x_1 \leq \frac{x_2}{p}$ and $x \geq \frac{x_2}{p}$. We show $E(x_1)$is concave in either part. Then we investigate an optimal ordering quantity depending on the conditions ofprices, costs and parameters p, q.

Finally we discuss many further research problems including a more suitable promotion sale and customer preference between price and freshness.

1 Introduction

In the field of inventory problems, several researchers have been interested in inventory control for a perishable product such as blood, fresh fruit, milk, film etc. Though there are huge number of research papers on perishable inventory, we only cite related papers ([1],[2],[3]). This paper introduce a promotion sale to the perishable cpntrol problem with life time two and LIFO issuing policy. Section 2 formulates the problem and calculates total expected profit function. Section 3 investigates an optimal ordering quantity. Finally section 4 summarizes results of this paper and discusses further research problems.

2 Problem Formulation

We consider the following model.

1. Single perishable product with two life time period is considered. That is, for a period, an amount of old one (remaining life time period is 1) in the stock is given as x_2 and under this condition we should determine an ordering quantity x_1 of the fresh one(remaining life time period is 2).

2. Ordering takes a place at the start of the period. The unit purchasing cost of the product is c.

3. Issuing policy is LIFO, that is, customer buys fresh one first and if fresh one is sold out, only some percent of customers overflowed from purchase of the fresh one buy the old one. We assume that this percentage is at most $100q$. Unit price of the fresh one is r_1 and that of old one is r_2. We assume that $r_1 \geq r_2 > c$ without any loss of generality.

4. Prominent feature of our model is a prpmotion sale, that is, we sell a set of products consisting of the fresh and old ones and it is sold at the discount unit price r_B less than the sum, $r_1 + r_2$ but not less than $r_2 + c$. We assume at most $100p$ percent of the customers purchasing the fresh one accept the promotion set, that is, buy the old one with the fresh one at the same time where we assume that $p \geq q$.

5. The old one that is not purchased by the customer and remained outdates and is discarded at the unit cost θ. While the fresh one not purchsed by the customer is stocked at the unit cost h.

6. The demand D of the customer is a nonnegative random variable. Its cummulative distribution function and density function are $F(D)$ and $f(D)$ respectively where $F(0) = f(0) = 0$.

Under the above setting, we calculate total expected profit function $E(x_1)$

Total expected profit function $E(x_1)$

Condition that $x_1 \geq \frac{x_2}{p}$

(i) Case $D < x_1$
Amount $(x_1 - D)$ (remaining quantity after issued) is stocked.

(a) If $pD \geq x_2$, then purchased amount as promotion set$= x_2$. Therefore old ones are sold out. Total profit is

$$r_1(D - x_2) + r_B x_2 - h(x_1 - D) - cx_1.$$

(b) If $pD \leq x_2$, then purchased amount as the fresh one is $(1-p)D$, purchased amount as a promotion set pD and outdating quantity is $(x_2 - pD)$. Therefore total profit is $(1-p)D + r_B pD - \theta(x_2 - pD) - h(x_1 - D) - cx_1$.

(ii) Case $D \geq x_1$
Note that fresh one and old one are sold out, Thatis, amount $(x_1 - x_2)$ is purchased as a fresh one and amount x_2 is purchased as promotion set. Therefore total profit is $r_1(x_1 - x_2) + r_B x_2 - cx_1$.

Condition that $x_1 \leq \frac{x_2}{p}$

(iii) Case $D < x_1$
Purchased amount as a fresh one is $(1-p)D$ and that as promotion set is pD. Further amount $x_1 - D$ (remaining quantity after issued) is stocked and amount $(x_2 - pD)$ is outdated. Therefore total profit is $r_1(1-p)D + r_B pD - h(x_1 - D) - \theta(x_2 - pD)$

(iv) Case $D \geq x_1$
Purchased amount as fresh one is $(1-p)x_1$ and purchased amount as a promotion set is px_1.

(a) Subcase $q(D - x_1) \geq x_2 - px_1$
Purchased amount as the old one is $x_2 - px_1$. Note that old one is sold out and so outdated amount is 0. Therefore total profit is $r_1(1-p)x_1 + r_B px_1 + r_2(x_2 - px_1) - cx_1$

(b) Subcase $q(D-x_1) \leq x_2 - px_1$ Purchased amount as old one is $q(D-x_1)$ and outdated amount is $x_2 - px_1 - q(D - x_1)$. Therefore total profit is $r_1(1-p)x_1 + r_B px_1 + r_2 q(D - x_1) - \theta(x_2 - px_1 - q(D - x_1)) - cx_1$.

Then expected total profit function $E(x_1)$:
For $x_1 \geq \frac{x_2}{p}$

$$E(x_1) = \int_{\frac{x_2}{p}}^{x_1} f(D)\{r_1(D - x_2) + r_B x_2 - h(x_1 - D)\}dD$$

$$+ \int_0^{\frac{x_2}{p}} f(D)\{r_1(1-p)D + r_B pD - \theta(x_2 - pD) - h(x_1 - D)\}dD$$

$$+\int_{x_1}^{\infty} f(D)\{(x_1-x_2)r_1+r_Bx_2\}dD-cx_1$$

For $x_1 \le \frac{x_2}{p}$

$$E(x_1)=\int_0^{x_1} f(D)\{r_1(1-p)D+r_BpD-h(x_1-D)-\theta(x_2-pD)\}dD$$
$$+\int_{x_1+\frac{x_2-px_1}{q}}^{\infty} f(D)\{r_1(1-p)x_1+r_Bpx_1+r_2(x_2-px_1)\}dD$$
$$+\int_{x_1}^{x_1+\frac{x_2-px_1}{q}} f(D)\{r_1(1-p)x_1+r_Bpx_1-\theta(x_2-px_1-q(D-x_1))\}dD-cx_1$$

3 Optimal ordering quantity

For $x_1 \ge \frac{x_2}{p}$

$$\frac{dE(x_1)}{dx_1}=r_1-(r_1+h)F(x_1)-c$$

and

$$\frac{d^2E(x_1)}{dx_1^2}=-f(x_1)(r_1+h)\le 0$$

For $x_1 \le \frac{x_2}{p}$

$$\frac{dE(x_1)}{dx_1}=\{r_1(1-p)+r_Bp-r_2p\}+F(x_1+\frac{x_2-px_1}{q})\{r_2+(p-q)\theta\}$$
$$-\{h+r_1(1-p)+r_Bp+\theta(p-q)\}F(x_1)-c.$$
$$\frac{d^2E(x_1)}{dx_1^2}=f(x_1+\frac{x_2-px_1}{q})(1-\frac{p}{q})\{r_2+(p-q)\theta\}$$
$$-\{h+r_1(1-p)+r_Bp+\theta(p-q)\}f(x_1)\le 0$$

since $p \ge q$. This shows $E(x_1)$ is concave for both parts, that is, for $x_1 \ge \frac{x_2}{p}$ and for $x_1 \le \frac{x_2}{p}$.

$$\lim_{x_1\to\frac{x_2}{p}-0}\frac{dE(x_1)}{dx_1}=r_1(1-p)+r_Bp-r_2p-c-\{r_1(1-p)+r_Bp-r_2p+h\}F(\frac{x_2}{p})$$

and

$$\lim_{x_1\to\frac{x_2}{p}+0}\frac{dE(x_1)}{dx_1}=r_1-c-(r_1+h)F(\frac{x_2}{p}).$$

Further

$$\lim_{x_1\to\frac{x_2}{p}+0}\frac{dE(x_1)}{dx_1}-\lim_{x_1\to\frac{x_2}{p}-0}\frac{dE(x_1)}{dx_1}=p(r_1+r_2-r_B)F(\frac{x_2}{p})\ge 0(1)$$

and stationary point x_1^a that $r_1 - c - (r_1 + h)F(x_1) = 0$ is

$$F^{-1}(\frac{r_1 - c}{r_1 + h}).$$

Therefore depending on whether $\frac{r_1-c}{r_1+h} \leq F(\frac{x_2}{p})$ or not, optimal ordering quantity is determined as follows:

Case I

$$\frac{r_1 - c}{r_1 + h} \leq F(\frac{x_2}{p})$$

In this case, note that the stationary point $x_1^a \leq \frac{x_2}{p}$. Therefore replacing $F(\frac{x_2}{p})$ by $\frac{r_1-c}{r_1+h}$, we have

$$\lim_{x_1 \to \frac{x_2}{p} - 0} \frac{dE(x_1)}{dx_1} \leq \frac{hp + c}{r_1 + h}(r_B - r_1 - r_2) \leq 0.$$

Therefore an optimal ordering quantity x_1^o is the stationary point x_1^b that

$$\begin{aligned} \{r_1(1-p) + r_B p - r_2 p\} \quad &+ \quad F(x_1 + \frac{x_2 - p x_1}{q})\{r_2 + (p-q)\theta\} - \\ &- \quad \{h + r_1(1-p) + r_B p + \theta(p-q)\}F(x_1) - \\ &- \quad c = 0 \end{aligned}$$

Case II

$$\frac{r_1 - c}{r_1 + h} \geq F(\frac{x_2}{p})$$

In this case, the stationary point $x_1^a \geq \frac{x_2}{p}$ and so

$$\lim_{x_1 \to \frac{x_2}{p} + 0} \frac{dE(x_1)}{dx_1} \geq 0$$

due to the concavity.

Subcase IIa

$$\frac{r_1 - c}{r_1 + h} \geq F(\frac{x_2}{p}) \geq \frac{\{r_1(1-p) + r_B p - r_2 p - c\}}{\{r_1(1-p) + r_B p - r_2 p + h\}}.$$

In this subcase, stationary point $x_1^b \leq \frac{x_2}{p}$. and so we compare the value $E(x_1)|_{x_1 = x_1^a}$ for $x_1 \geq \frac{x_2}{p}$ and the value $E(x_1)|_{x_1 = x_1^b}$ for $x_1 \leq \frac{x_2}{p}$. Corresponding ordering quantity to the greater one is an optimal ordering quantity x_1^o.

Subcase IIb

$$F(\frac{x_2}{p}) \leq \frac{\{r_1(1-p) + r_B p - r_2 p - c\}}{\{r_1(1-p) + r_B p - r_2 p + h\}}$$

In this subcase, there exists no stationary point that

$$\begin{aligned} \{r_1(1-p) + r_B p - r_2 p\} \quad &+ \quad F(x_1 + \frac{x_2 - p x_1}{q})\{r_2 + (p-q)\theta\} - \\ &- \quad \{h + r_1(1-p) + r_B p + \theta(p-q)\}F(x_1) - \\ &- \quad c = 0 \end{aligned}$$

for the part $x_1 \leq \frac{x_2}{p}$ and so optimal ordering quantity $x_1^o = x_1^a$.

4 Conclusion

We have discussed a promotion sale problem for the perishable product. But in our model, we have not considered the shortage cost. Shortage cost is usually hard to be estimated. Therefore L fuzzy number may be considered and if it is introduced in our model, the expected total profit function becomes a L fuzzy number. Using some fuzzy order, we need to seek some non-dominated ordering quantities since the fuzzy order is not linear order. Further we should investigate other suitable promotion sales in order to cope with other dealers and sensitivity about stock amount, prices etc. But these are furture research problems.

References

[1] Ishii H. (1993), Perishable inventory problem with two types of customers and different selling prices, *Journal of the Operations Reserch Society of Japan*,30(4)199-205

[2] Nahmias S. (1963), Perishable inventory theory: a Review, *Operations Research*, 30, 680-708.

[3] Katagiri H. and Ishii H. (2003), An inventory problem with a perishable and non-perishable one, *Asia Pacific Management Review*, 8(4)477-485

On Modeling Planning Problems: Experience From The Petrobras Challenge

Roman Barták
Faculty of Mathematics and Physics
Charles University in Prague
bartak@ktiml.mff.cuni.cz

Neng-Fa Zhou
Brooklyn College
The City University of New York
nzhou@acm.org

Abstract

The International Planning Competitions have led to the development of the standard modeling framework for describing planning domains and problems Planning Domain Description Language (PDDL). The majority of planning research is done around problems modeled in PDDL though there are only a few applications adopting PDDL. The planning model of independent actions connected only via causal relations is very flexible, but it also makes plans less predictable (plans look different than expected by the users) and it is probably also one of the reasons of bad practical efficiency of current planners (visibly wrong plans are blindly explored by the planners). In this paper we argue that grouping actions into flexible sub-plans is a way to overcome the efficiency problems. The idea is that instead of seeing actions as independent entities that are causally connected via action preconditions and effects, we suggest using a form of finite state automaton (FSA) to describe the expected sequences of actions. The arcs in FSA are annotated by conditions guiding the planner to explore only proper paths in the automaton. The second idea is composing primitive actions into meta-actions, which decreases the size of FSA and makes planning much faster. The main motivation is to give users more control over the action sequencing with two primary goals: obtaining more predictable plans and improving efficiency of planning. The presented ideas originate from solving the Petrobras logistic problem, where this technique outperformed classical planning models.

1 Introduction

Recent research in the area of planning is centered on the representation of problems in the Planning Domain Description Language introduced for International Planning Competitions [2]. Having a standard modeling language accelerated research in the area of planning and led to development of many benchmark problems that are used to experimentally evaluate and compare various solving

approaches. On the other hand, PDDL is based on the original STRIPS idea of having "independent" actions that can be causally connected via their preconditions and effects into a sequence – a *plan*. This makes planning flexible but it also introduces some undesirable behaviors. For example, an action for unloading an item can be planned immediately after an action that loaded the item. This is causally correct (unloading only requires the item to be loaded which is achieved by the previous loading action) though from a human perspective such action sequences are not desirable (the state after unloading will be identical to the state before loading). It is possible to forbid such situations by changing the model in such a way that, for example, a transportation action must be planned between loading and unloading[1]. However, such enhanced models are less natural and less readable by humans, and flaws can be easily introduced in the models if more such modifications are required. It is more natural, if a human modeler prescribes possible (reasonable) action sequences. There exist two modeling approaches based on this idea, *hierarchical task networks* (HTNs) [9] and *timelines* [10]. While an HTN uses the notion of task that decomposes into sub-tasks until primitive actions are obtained, timelines focus on modeling allowed time evolutions of state variables and synchronizations between them.

In this paper we study a modeling framework positioned half way between timelines and HTNs. Similarly to [4] we propose to use a finite state automaton (FSA) describing allowed sequences of actions. An FSA plays the role of effects and conditions from classical planning as it says which actions may follow a given action. This FSA can be accompanied by additional constraints restricting when some transitions between the actions may occur based on the current state of the world. These conditions are different from classical action preconditions as they can involve information about the goal (for example, pickup cargo only when there is some cargo to deliver). This is much closer to control rules [7], but rather than specifying control rules separately from the description of the planning domain, we suggest integrating them in the FSA (and also into meta-actions, see below). This is an original idea of this paper.

We have found that the above modeling approach is not sufficient enough when solving larger real-life problems and we suggest additional extensions motivated by the Petrobras challenge. The Petrobras challenge [12] is a logistic problem of transporting cargo items between ports and oil platforms using vessels with limited capacity. The paper [11] studied three approaches to solve this problem. The winning technique was based on Monte Carlo Tree Search (MCTS) where the search was done over sequences of abstract actions. Each abstract action was then decomposed to primitive actions based on the state of the world. In this paper we use a very similar idea where we take reasonable sub-plans (sequences of actions) and encode them as a single meta-action. The FSA is then defined over these meta-actions. During planning the selected meta-action decomposes into a sequence of primitive actions depending on the current situation. There already exists a concept of *macro-actions* in planning

[1]The action *unload* may use some new proposition – a semaphore – as its precondition, and this proposition is removed by the *load* action while added by the *transport* actions.

[5]. However, while a macro-action decomposes to a fixed sequence of primitive actions, a meta-action may decompose to different sequences of primitive actions based on the situation. The concept of meta-actions is closer to HTNs, though we use only one level of decompositions – from a meta-action to a sequence of primitive actions. Also, planning in our framework is based on classical action sequencing rather than on task decomposition.

This paper shows that action grouping and prescribed action sequencing are very important for the Petrobras challenge. Instead of sophisticated MCTS method, we use backtracking accompanied by tabling [13] to solve the problem. As we shall experimentally show the resulting method achieves very similar performance to the MCTS algorithm. Hence we believe the presented modeling concepts are important for solving real-life planning problems and may bring significant efficiency boost.

The paper is organized as follows. We will first briefly introduce the Petrobras challenge, highlight some of its important components, and describe the three techniques already applied to this problem. Then we will present the proposed modeling framework based on finite state automata, explain the solving technique, and show how actions can be grouped to meta-actions. After that we will experimentally compare our method with the current best methods using the Petrobras benchmarks from [11]. The description of possible future directions of research will conclude the paper.

2 The Petrobras Challenge

International Competition on Knowledge Engineering for Planning and Scheduling (ICKEPS 2012) brought several real-life motivated modeling and solving challenges including the Petrobras problem. The Petrobras problem [12] deals with planning deliveries of cargo items between ports and platforms while respecting the capacity limits of vessels, ports, and platforms. The ports and platforms can serve a limited number of vessels at the same time; the vessels can transport limited weight of cargo items, vessel's load influences vessel's speed and fuel consumption, and the limited capacity of fuel tanks must also be assumed when planning transport. Given a set of cargo items to deliver (including cargo weights and initial ports), the problem is to find a feasible plan that guarantees the delivery of all cargo items to given platforms and respects the constraints on the vessel, port, and platform capacities. Vessels should leave a waiting area, perform their mission and go back to one of the waiting areas. Loading and unloading of cargo items are done at ports and platforms and require some time (vessels must be docked before any operation and undocked before moving elsewhere). Vessels can be refueled at ports and certain platforms and each vessels must always have enough fuel to go to the nearest refueling station. The following primitive actions with self-explaining names can be used in the plan: *navigate-empty-vessel, navigate-nonempty-vessel, load-cargo, unload-cargo, refuel-vessel-platform, refuel-vessel-port, dock-vessel, undock-vessel*.

So far there was only one attempt to solve the full Petrobras challenge.

Toropila et al. [11] applied classical planning, temporal planning, and Monte Carlo Tree Search (MCTS) to solve the Petrobras challenge. The classical planning approach used SGPlan 6.0 to generate the initial plan and actions were scheduled to particular times in the post-processing stage. The temporal planning approach used PDDL 3.1 with fluents and durative actions and the plan was generated by the Filuta planner [8]. The last approach exploited Monte Carlo Tree Search (MCTS) techniques that required a different action model with four abstract actions: *Load, Unload, Refuel, GoToWaitingArea*. These actions describe intentions and they are decomposed to primitive actions based on the situation. For example, the action *Unload* assumes that cargo is loaded and vessel is either in a port or a platform. This action is decomposed in the following way. If the current location of the vessel is the same as the target location of the cargo then only a single underlying action *unload-cargo* is performed, otherwise the abstract action is decomposed to the sequence of the original actions: *undock-vessel, navigate-nonempty-vessel, dock-vessel, unload-cargo*.

The experiments with the challenge data and randomly generated data, where the number of vessels and cargo items varied (3-10 vessels, 1-15 cargo items), showed that the classical planning approach is not viable as it cannot solve problems with more (7+) cargo items. The MCTS planner was the clear winner, followed by Filuta that can solve all problems, but the quality of plans was lower (30% for makespan, 130% for fuel consumption).

3 The Novel Modeling Approach

The Petrobras challenge inspired us to explore the reasons of success of the MCTS technique. In particular we focused on the "predefined" action sequences that are hidden in the abstract actions of the MCTS approach and partly also in the special resource solvers in the Filuta planner. Another motivation for our research went from the efficient model of the Sokoban game implemented in B-Prolog [14]. This model was also based on grouping actions into specific sequences. Our hypothesis is that even a simple search algorithm, for example depth-first search with tabling, can solve complex planning problems provided that the model itself guides the solver by describing expected action sequences rather than giving only independent actions connected via causal relations.

3.1 Model Based on Finite State Automata

Let us first describe the possible plans of each vessel in the Petrobras challenge as a finite state automaton. Finite state automata (FSA) were shown to significantly improve efficiency of constraint-based planners [4] and they are also used by the Filuta planner [8]. We use the concept where states of the automaton (nodes) correspond to actions and transitions (arcs) describe allowed action sequences. Figure 1 shows the FSA that models all possible actions and transitions between the actions for a single vessel in the Petrobras domain. We already use some abstract actions there, for example the action *navigate* means *navigate-*

empty-vessel, if the vessel is empty, or *navigate-nonempty-vessel*, if some cargo is loaded to the vessel. Similarly the action *refuel* means either *refuel-vessel-platform* or *refuel-vessel-port* depending on whether the vessel is docked in a platform or in a port. Notice also that the presented FSA restricts some action sequences. In particular, it is not possible to dock immediately after undocking there is no practical reason for such a sequence of actions though the classical PDDL model allows it. Similarly, refueling is done only after loading/unloading – this removes symmetrical sequences of actions with an identical overall effect (the vessel is loaded/unloaded and refueled). Finally, during one stop at the port or platform, the FSA allows either loading of cargo or unloading of cargo, but not both operations together. This is motivated by the particular problem to be solved. We need to move cargo from ports to platforms and there is no need to unload a cargo (at some platform) and load another cargo at the same location. In principle, it might be possible to move cargo to some intermediate location where it will be picked up by another vessel. However such flexible plans were not found necessary in the Petrobras challenge. Note finally that all these sequencing restrictions are naturally modeled using the transitions in the FSA. If more flexible plans are desirable then the corresponding transitions can be added to the FSA. We also use another mechanism to restrict sequencing by putting constraints on the transitions. These constraints describe situations when the transition is allowed. We will describe these constraints in more detail later in the text. The classical planning model with action preconditions and effects makes expressing such allowed action sequences much more complicated and not very natural (see the footnote 1 in the Introduction).

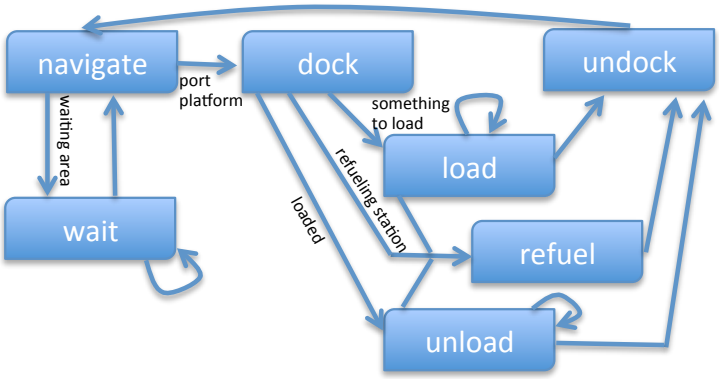

Figure 1: Finite state automaton describing actions and allowed transitions between the actions in the Petrobras challenge.

Actions in the model have specific durations that are given by action parameters (such as locations and current load for the navigation actions). Hence, in the planning terminology we should rather use the notion of a planning operator and actions are obtained by setting values of the parameters of the operator.

The capacity of the vessel is modeled within the action, for example, the transition to a loading action is allowed only if there is enough capacity in the vessel.

So far we discussed plans for a single vessel, but if there are more vessels in the problem, their plans interact. For example at most two vessels can be docked at the port at the same time. We check these synchronization constraints when adding a new action to the plan as described in the next section.

3.2 Solving Approach

The solving algorithm uses the round-robin approach, where an action is selected for each vessel provided that the last action for that vessel finished before the rolling horizon. Figure 2 demonstrates the left-to-right round-robing solving approach that combines planning (action selection) with scheduling (allocation to time and resources). At the beginning all vessels are waiting so in the first step, we select an action for each vessel. There is only one exception of this process – if a vessel is waiting and a new waiting action is selected, we only prolong the existing waiting action for that vessel. The waiting action is the only action with arbitrary duration so it is possible to set its duration to any time. Action selection represents the choice point of the search algorithm. The only heuristic for action selection is the fixed order of actions in the model specification, where for example unloading is before loading which is before refueling for a docked vessel. There are also control rules encoded in the action descriptions – for example, the navigation action for an empty vessel goes only to a port with some remaining cargo or to a waiting area.

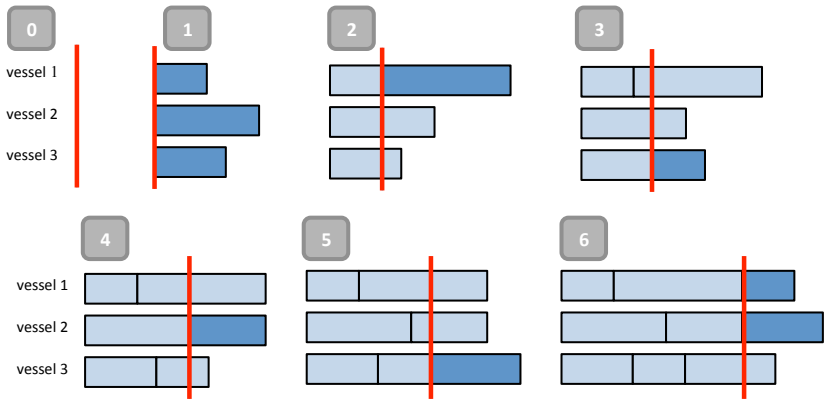

Figure 2: Illustration of the left-to-right integrated planning and scheduling approach with a rolling horizon (vertical line). Newly added actions are displayed as dark rectangles; x-axis represents time.

We have implemented the above solving approach in B-Prolog using tabling [1]. It means that we use depth-first search with remembering visited states – the state is represented by the current states of vessels and the list of cargo items

to be still delivered. In each step, we select an action for each vessel and move the time horizon. The core idea of this planning algorithm can be described using the following abstract code (this is actually an executable code in Picat [3] the follower of B-Prolog):

```
table (+,-,min)
plan(S,Plan,Cost),final(S) =>
        Plan=[],Cost=0.
plan(S,Plan,Cost) =>
        action(Action,S,S1,ActionCost),
        plan(S1,Plan1,Cost1),
        Plan = [Action|Plan1],
        Cost = Cost1+ActionCost.
```

As the reader can see, the search procedure is very simple. The real power of the solving approach is hidden in the action model and in the tabling mechanism. Tabling is important to save visited states so they are not re-explored. In the above code, for each state S the tabling mechanism stores the found *Plan* while minimizing the *Cost* of the plan (the cost is measured by makespan in the Petrobras problem). It is a form of a branch-and-bound procedure.

3.3 Meta-actions

Though the presented action model already included some sequencing restrictions, we have found experimentally that the model did not scale up well. In fact, it worked only for a single vessel with a few cargo items to deliver. By exploring the generated plans we noticed two types of erratic behavior. If more cargo items were available for delivery, all free vessels headed to the port, where cargo was located. This behavior was caused by preferring the navigation action to other actions if some cargo should be delivered. As the cargo was available before the first vessel loaded it, the other vessels "believed" that there is still some cargo to deliver and so transport to the port was planned for them. The second problematic behavior was that vessels left the waiting area just to refuel and then returned back to the waiting area.

Though the naïve model was not competitive to solve the problem, it showed the core ideas of our proposed modeling approach. The reasonable sequences of actions are modeled using a finite state automaton. To follow some transition a specific condition must be satisfied. We can make this model more efficient by grouping sequences of actions into a meta-action similarly to the MCTS approach [11]. Each meta-action represents a sequence of primitive actions with possible alternatives. We propose a model using four meta-actions with more specific conditions to apply the actions. Figure 3 shows the resulting finite state automaton including the transition conditions. Each meta-action decomposes into primitive actions while applying additional conditions on the actions. For example, the *Deliver* action starts with the *navigate* action where the destination is selected only from the destinations of loaded cargo items (a choice point). The next action in the sequence is *docking* followed by *unloading* and, if pos-

sible *refueling* done in parallel with *unloading*. We always unload all cargo items for a given destination and we always refuel the full tank (deterministic choice). The last action in the sequence is *undocking*. The *Deliver* action can only be used if some cargo is loaded to the vessel. Similarly, the *Pickup* action is applicable only if the vessel is empty and there is some cargo to deliver. We select the port where cargo is available (choice point) and pre-allocate some cargo items to the vessel (choice point). These conditions ensure that vessels are moving only when necessary. The *Pickup* action then decomposes to *navigate*, *docking*, *loading* and *refueling*, and *undocking* actions. Note that there could be more loading actions if more cargo items are loaded. As the order of loaded items is not important, only a single sequence of loading actions is explored during the decomposition (based on the fixed order of cargo items). This further reduces the search space – equivalent permutations of loading actions are not explored (similarly for unloading actions). The last two actions are *Waiting* and *Go2Wait* that are applicable if the vessel is empty and it is in the waiting area (then *Waiting*) or elsewhere (*Go2Wait*). The *Go2Wait* action decomposes to the navigate action, but if there would not be enough fuel then the vessel navigates to a refueling station before navigating to the waiting area. The three actions that include transport – *Pickup*, *Delivery*, *Go2Wait* – force the vessels to do only"reasonable" moves. If the vessel is empty, it can either go to a waiting area or to some port to pickup cargo, if any cargo is available. No other movement is allowed. Similarly, if the vessel is loaded, it can go only to some platform where some loaded cargo should be delivered.

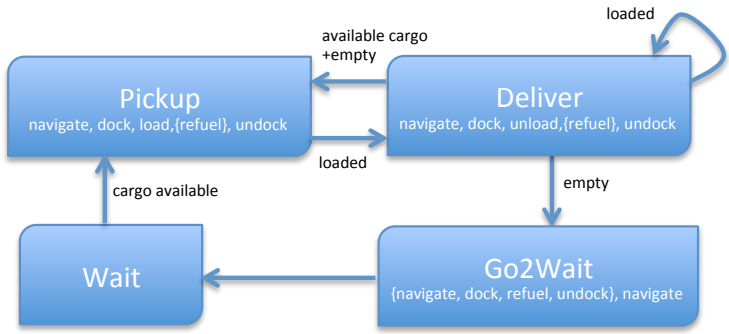

Figure 3: Finite state automaton describing actions and allowed transitions between the actions in the Petrobras challenge.

The meta-actions allow users to specify expected sub-plans with conditions when to apply the sub-plans and non-determinism to be resolved by the planner (what cargo by which vessel). The user has better control about how the plans look like while leaving some decisions to the solver. The major caveat is some loss of flexibility. By specifying the sub-plans we may omit possible plans that were not assumed by the user. For example, in our model, we do not allow to

pickup new cargo while some cargo is still loaded to the vessel. Also the cargo is only unloaded at its final destination. In particular, it is not possible to deliver the cargo half-way and using another vessel to deliver it to the final destination. These restrictions were intentional to reduce exploration of unwanted plans.

4 Experimental Results

To evaluate the proposed modeling and solving techniques we compare them with the best approaches from [11], namely the Filuta planner and the MCTS approach. We do not include SGPlan as its results were poor compared to other approaches. The Filuta planner minimized makespan identically to our method, while the MCTS planner used a manually tuned ad-hoc objective function that combined makespan, the number of actions, and fuel consumption: $usedFuel + 10 * numActions + 5 * makespan$. We re-use here the results reported in [11] where the experiments were run on the Ubuntu Linux machine equipped with Intel Core i7-2600 CPU @ 3.40GHz and 4GB of memory and the planners were allowed to use approximately 10 minutes of runtime. Our method was implemented in B-Prolog, we run the experiments on the MacOS X 10.7.5 (Lion) machine with Intel Core i7 CPU @ 1.8 GHz and 4GB of memory, and we report the best results found within one minute of runtime (we did not observe any further improvement when running B-Prolog longer).

We will first present the results for the original Petrobras problem described in [12], which consisted of 10 vessels and 15 cargo items to deliver, with the fuel tank capacity for all vessels set to 600 liters. Table 1 shows the comparison of makespan and fuel consumption that were two major objectives in the Petrobras challenge. We can see that our approach is significantly better than the Filuta planner in both objectives and it also beats the best so-far approach based on MCTS in the makespan though the fuel consumption is worse.

Table 1: Results for the Petrobras challenge from [12].

Planner	**Makespan**	**Fuel**
Filuta	263 (1.62x)	1989 (2.24x)
MCTS	204 (1.26x)	**887** (1.00x)
B-Prolog	**162** (1.00x)	1263 (1.42x)

To compare the approaches in more detail, the paper [11] proposed several random benchmarks based on the Petrobras domain, where the number of cargo items to deliver and the number of available vessels varied. They also varied the fuel tank capacity, but according to the experiments this had a limited impact of performance so we kept the fuel tank capacity at 600 liters. We used two scenarios from [11] with 3 vessels and 10 vessels (called Group A and Group B in [11]), and we varied the number of cargo items from 1 to 15. Figure 4 shows the comparison of makespan for all three planners. We can observe that

our approach is better for problems with the smaller number of cargo items, while the MCTS method takes lead when the number of cargo items increases. In fact, for the first six problems in each group, our system found (and proved) makespan-optimal plans, while the other two planners are sub-optimal only. We were also consistently better than the Filuta planner.

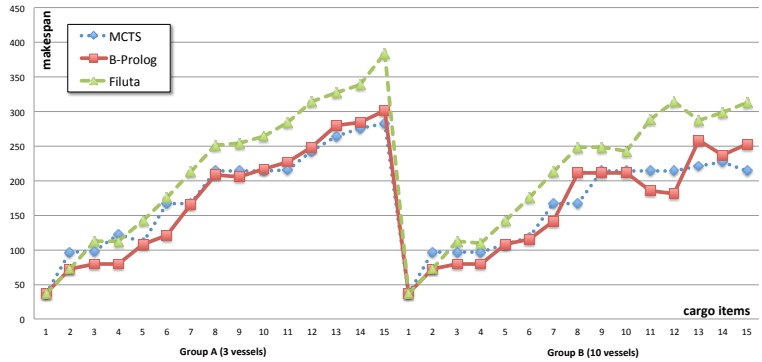

Figure 4: Comparison of makespan for problems with different numbers of cargo items.

Regarding the fuel consumption (Figure 5), our planner is closer to the Filuta planner. We are mostly better but there were problems on which Filuta found plans with less fuel consumption. The MCTS method was consistently the best planner regarding fuel consumption, though this is not surprising as the other two planners optimized makespan only.

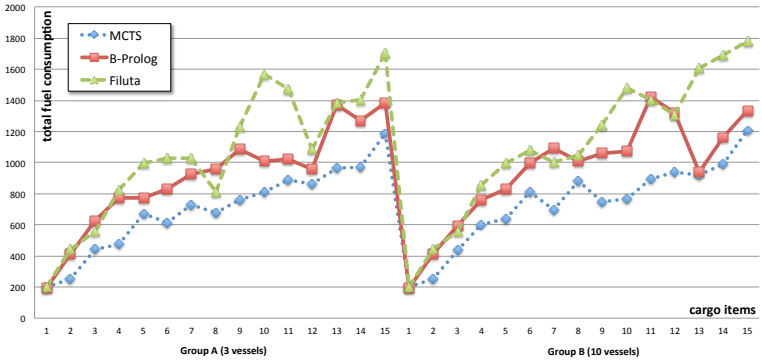

Figure 5: Comparison of fuel consumption for problems with different numbers of cargo items.

5 Conclusions

We proposed a novel framework for describing planning problems that is based on meta-actions and transitions between them accompanied by conditions when the transitions can be used. Each meta-action decomposes into primitive actions; again specific conditions can be imposed on the parameters of these actions and also on their composition. The main motivation was to give users more control over action sequencing with two primary goals: obtaining more predictable plans and improving efficiency of planning. We demonstrated the modeling principles using a single domain – the Petrobras challenge – and using a single solving approach – B-Prolog with tabling. Though the solving technique is very simple, the proposed approach was shown to be competitive with the leading technique for the Petrobras problem in terms of plan quality (our technique was also faster).

The next step is to decouple the modeling approach from the solving mechanism and applying it to other planning problems. We sketched the modeling principles informally; a formal description in the form of a modeling language is necessary for more general applicability. The presented approach can also drive further research in the tabling methods applied to optimization problems. Other solving techniques may also be applied, we are not aware of any current planner supporting all three presented features: meta-actions, predefined action sequencing, and control rules. It would be interesting to study how these three modeling techniques contribute to plan efficiency. Our preliminary experiments showed that their combination helped to successfully solve the Petrobras challenge.

6 Acknowledgements

Roman Barták is supported by the Czech Science Foundation under the project KnowSched (P202-10-1188).

References

[1] B-Prolog, http://www.probp.com

[2] International Planning Competitions, http://ipc.icaps-conference.org

[3] Picat, http://www.picat-lang.org

[4] Barták, R. (2011), On Constraint Models for Parallel Planning: The Novel Transition Scheme. *Proceedings of SCAI 2011*, pp. 50–59, *Frontiers of Artificial Intelligence*, Vol. 227, IOS Press.

[5] Botea, A.; Enzenberger, M.; Muller, M.; Schaeffer, J. (2005), Macro-FF: Improving AI planning with automatically learned macro-operators. *Journal of Artificial Intelligence Research*, 24, pp. 581–621.

[6] Chaslot, G.; Bakkes, .; Szita, I.; Spronck, P. (2008), Monte-Carlo tree search: A new framework for game AI. *Proceedings of the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference*, The AAAI Press.

[7] Doherty, P.; Kvarnstršm, J. (1999), TALplanner: An empirical investigation of a temporal logic-based forward chaining planner. *Proceedings of the Sixth International Workshop on Temporal Representation and Reasoning (TIME99)*, pp. 47–54, Orlando, Florida, USA. IEEE Computer Society Press.

[8] Dvořák F. and Barták, R. (2010), Integrating time and resources into planning. *Proceedings of ICTAI 2010*, Volume 2, pp. 71–78, IEEE Computer Society.

[9] Erol, K.; Hendler, J.; Nau, D. (1994), HTN Planning: Complexity and Expressivity. *Proceedings of AAAI-94*, pp. 1123–1128, AAAI Press.

[10] Muscettola, N. (1994), HSTS: Integrating Planning and Scheduling. *Intelligent Scheduling*. Morgan Kauffmann.

[11] Toropila, D.; Dvořák, F.; Trunda, O.; Hanes, M.; Barták, R. (2012), Three Approaches to Solve the Petrobras Challenge: Exploiting Planning Techniques for Solving Real-Life Logistics Problems. *Proceedings of ICTAI 2012*, pp. 191–198, IEEE Conference Publishing Services.

[12] Vaquero, T. S.; Costa, G.; Tonidandel, F.; Igreja, H.; Silva J. R. and Beck, C. (2012), Planning and scheduling ship operations on petroleum ports and platform. *Proceedings of the Scheduling and Planning Applications Workshop*, pp. 8–16.

[13] Warren. D.S. (1992), Memoing for Logic Programs, *CACM*, Vol. 35, No. 3, pp. 93–111.

[14] Zhou, N.-F., Dovier, A. (2011), A Tabled Prolog Program for Solving Sokoban. *Proceedings of the 26th Italian Conference on Computational Logic (CILC 2011)*, pp. 215–228.

A generalization of the noisy-or model to multivalued parent variables*

Jiří Vomlel
Institute of Information Theory and Automation of the AS CR
Academy of Sciences of the Czech Republic
http://www.utia.cas.cz/vomlel

Abstract

In this paper we propose a generalization of the noisy-or model to multivalued parent variables. Albeit the proposed generalization is more restrictive than previous proposals, it has several nice properties. In this paper we suggest a method for learning this model and report results of experiments on the Reuters text classification data.

1 Introduction

The conditional probability tables (CPTs) that are the basic building blocks of Bayesian networks [9, 6] have, generally, an exponential size with respect to the number of variables of the CPT. This has two unpleasant consequences. First, during the elicitation of model parameters one needs to estimate an exponential number of parameters. Second, in case of a high number of parent variables the exact probabilistic inference may become intractable.

On the other hand real implementations of Bayesian networks (see e.g. [8]) often have a simple local structure of the CPTs. The noisy-or model [9] is a popular model for describing relations between variables in one CPT of a Bayesian network. Noisy-or is member of a family of models called models of independence of causal influence [4] or canonical models [2]. The advantage of these models is that the number of parameters required for their specification is linear with respect to the number of variables in CPTs and that they allow applications of efficient inference methods, see for example [3, 11].

In this paper we propose a generalization of the noisy-or model to multivalued parent variables. Our proposal differs from the noisy-max model [5] since we keep the child variable binary no matter what the number of states of the parent variables are. Also we have only one parameter for each parent no matter

*This work was supported by the Czech Science Foundation through the project 13-20012S.

the number of states of the parent variables. Our generalization is also different than the generalization of the noisy-or model proposed by Srinivas [12] since in his model the inhibitor probabilities cannot depend on the state of the parent variables if the state differs from the state of the child. We find this to be a quite restrictive requirement for some applications.

We will show that our proposal is closely connected with the Poisson Regression of Generalized Linear Models [7]. Due to this connection we can use methods from Poisson Regression for learning parameters of the generalized noisy-or model from data. In the paper we present results of numerical experiments on the well-known Reuters text classification data. We use this dataset to compare the performance of our multinomial generalization of noisy-or with the standard noisy-or.

2 Multinomial noisy-or

In this section we propose a generalization of noisy-or for multivalued parent variables. Let Y be a binary variable taking states $y \in \{0, 1\}$ and $X_i, i = 1, \ldots, n$ be multivalued discrete variables taking states $x_i \in \{0, 1, \ldots, m_i\}$, $m_i \in \mathbb{N}^+$. The local structure of both the standard (see, e.g., [2]) and the multinomial generalization of the noisy-or can be made explicit as it is shown in Figure 1.

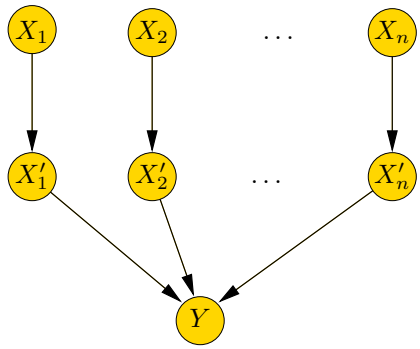

Figure 1: Noisy-or model with the explicit deterministic (OR) part.

The conditional probability table $P(Y|X_1, \ldots, X_n)$ is defined using CPTs $P(X'_i|X_i)$ as

$$P(X'_i = 0|X_i = x_i) = (p_i)^{x_i} \quad (1)$$
$$P(X'_i = 1|X_i = x_i) = 1 - (p_i)^{x_i} \ , \quad (2)$$

where $p_i \in [0, 1]$ is the parameter that defines the probability that the positive value x_i of variable X_i is inhibited. In the formula, we use parenthesis to emphasize that x_i is an exponent, not an upper index of p_i. The CPT $P(Y|X'_1, \ldots, X'_n)$ is deterministic and represents the logical OR function.

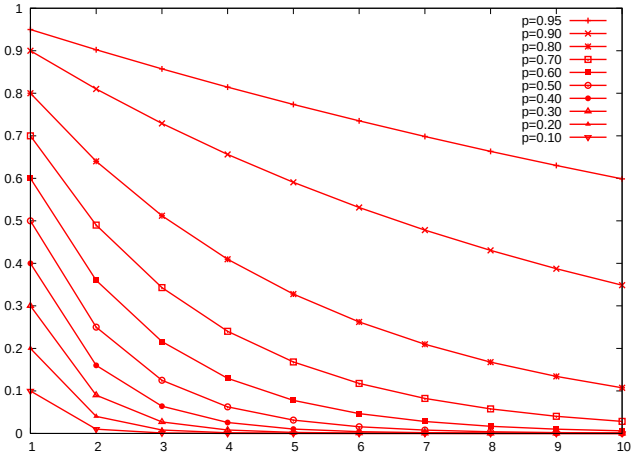

Figure 2: The dependence of $P(X'=0|X=x)$ on p and x.

Remark. Note that the higher is the value x_i of X_i the lower the probability of $X_i'=0$, which is a desirable property in many applications.

The conditional probability table $P(Y|X_1,\ldots,X_n)$ is then defined as

$$\begin{aligned} P(Y=0|X_1=x_1,\ldots,X_n=x_n) &= \prod_{i=1}^{n} P(X_i'=0|X_i=x_i) \\ &= \prod_{i=1}^{n} (p_i)^{x_i} \qquad (3) \\ P(Y=1|X_1=x_1,\ldots,X_n=x_n) &= 1-\prod_{i=1}^{n} (p_i)^{x_i} \ . \qquad (4) \end{aligned}$$

Remark. Note that if $m_i=1$, i.e. the values x_i of X_i are either 0 or 1, then we get the classical noisy-or model.

In Figure 2 dependence of the inhibitory probability $P(X'=0|X=x)$ on the value x of a variable X is depicted for different values of the parameter p.

It is important to note that contrary to the definition of causal noisy-max [2, Section 4.1.6] we have only one parameter p_i for each parent X_i of Y no matter what is the number of states of X_i. This implies that our model is more restricted. But, on the other hand, the suggested simple parametrization guarantees ordinality, which is in many application a desirable property (as it is also discussed in [2]). Also, since we estimate or elicite (from domain experts) fewer parameters, the estimates are more reliable.

3 Correspondence to Poisson Regression

Next we will show the correspondence of the multionomial noisy-or to the Poisson Regression of Generalized Linear Models [7].

By taking the logarithm of both sides of equation (3) we get

$$\log P(Y = 0|X_1 = x_1, \ldots, X_n = x_n) \quad = \quad \sum_{i=1}^{n} x_i \cdot \log p_i \ .$$

Define a new parameter $r_i = \log p_i$. Note that $r_i \in (-\infty, 0]$. Then we get

$$\log P(Y = 0|X_1 = x_1, \ldots, X_n = x_n) \quad = \quad \sum_{i=1}^{n} x_i \cdot r_i \ .$$

which is the formula of the Poisson regression of the binary variable $1 - Y$. Please, note that the expected value $E(1 - Y|x_1, \ldots, x_n) = P(Y = 0|X_1 = x_1, \ldots, X_n = x_n)$. Therefore

$$\log E((1 - Y)|x_1, \ldots, x_n) \quad = \quad \sum_{i=1}^{n} x_i \cdot r_i \ .$$

This correspondence allows us to apply standard maximum likelihood estimation methods for Poisson regression models to learning multinomial noisy-or. A method typically used to learn the generalized linear models is the iteratively reweighted least squares method [7].

When using a real data that might be modified by a noise or might be generated from a different model it can happen that for some of the $r_i, i = 1, \ldots, n$ parameters we learn positive values. This has a quite natural interpretation in the multinomial noisy-or model. It means that higher values of X_i imply higher inhibitory probability. Therefore we decided to treat positive values of r_i parameters by relabeling the values of X_i from $x_i = 0, 1, \ldots, m_i$ to $m_i - x_i$ in the multinomial noisy-or model. In this way the generalized noisy-or is now capable to treat not only positive (presence of X_i increases probability of $Y = 1$) but also negative influences (presence of X_i decreases probability of $Y = 1$).

4 Experiments

In this section we describe experiments we performed with the well known Reuters-21578 collection (Distribution 1.0) of text documents. The text documents from this dataset appeared on the Reuters newswire in 1987 and were manually classified by personnel from Reuters Ltd. and Carnegie Group, Inc. to several classes according to their topic. In the test we used the split of documents to training and testing sets according to Apté et al. [1]. We performed experiments with preprocessed data for eight largest classes[1].

[1]The preprocessed dataset is available at `http://web.ist.utl.pt/acardoso/datasets/`.

In the experiments we compare the standard noisy-or classifier [13] and our generalized multinomial noisy-or. Both models were learned using the iteratively reweighted least squares method [7] implemented in R – a language and environment for statistical computing [10]. We performed experiments with two versions of both classifiers:

(a) features X_i with both a positive (+) or a negative influence (-) on probability of $Y = 1$ were allowed and treated as it was described in previous section,

(b) features X_i with a negative influence (-) on probability of $Y = 1$ were omitted.

Table 1: Comparisons of the accuracy of the noisy-or and its multinomial generalization. The best achieved accuracy is printed boldface and framed.

Class	# test documents	binomial (+ and -)	binomial (only +)	multinomial (+ and -)	multinomial (only +)
earn	1083	**95.61**	95.02	94.29	94.66
acq	696	**94.20**	91.78	92.01	91.87
crude	121	**97.58**	**97.58**	96.12	96.12
money-fx	87	**96.67**	**96.67**	96.30	96.44
interest	81	96.67	96.67	**97.03**	**97.03**
trade	75	97.44	97.44	**98.13**	**98.13**
ship	36	98.77	98.77	**99.13**	**99.13**
grain	10	**99.91**	**99.91**	99.86	99.86
total	2189				

The results of experiments are summarized in Table 1. The accuracy is reported using the percentage scale, it is the relative proportion of correctly classified documents either as belonging to the given class or not. From Table 1 we can see that standard noisy-or performs better for larger models, while multinomial noisy-or is better at smaller models. The model for the class *grain* is very small, it has one feature only and also the difference between the models' accuracy is very small – it is 0.046, which corresponds to one test case only. In Table 2 we provide the number of selected features for models from Table 1.

We decided to include into the models all features that were not rejected as irrelevant at the significance level 0.1. In the experiments, we observed that the classification accuracy could be slightly improved if the significance was increased to 0.3, this would also slightly improve the AIC criteria[2] However, since the gain was not large we decide to prefer simpler models. Also, it has

[2]The AIC criteria takes into account both the log-likelihood and the number of parameters of the learned model. The lower the AIC the better the model.

Table 2: Comparisons of the number of selected features for the noisy-or and its multinomial generalization.

Class	# test documents	binomial (+ and -)	binomial (only +)	multinomial (+ and -)	multinomial (only +)
earn	1083	17	14	13	12
acq	696	28	20	23	20
crude	121	4	4	3	3
money-fx	87	4	4	4	3
interest	81	3	3	2	2
trade	75	5	6	4	4
ship	36	2	2	3	3
grain	10	1	1	1	1
total	2189				

a very limited influence on the two tested models' preference, which is of our major interest in this paper. However, it might be topic for a future research to apply exhaustive feature selection methods that would find optimal models for the families of our interest.

5 An example

In this section we use the class *ship* to illustrate the benefits of treating the features as multinomial. In the first example we present the standard noisy-or model and in the second the multinomial noisy-or model. Both models were learned by the iteratively reweighted least squares method [7] and contain only significant features for the significance level 0.1. The accuracy of the noisy-or model was 98.77%, while the multinomial noisy-or model achieved accuracy 99.13%. Even if more features are included in the standard noisy-or model the accuracy remains lower than the accuracy of the multinomial noisy-or model.

Example 1 (The noisy-or model for the ship class)**.** In Figure 3 the structure of the noisy-or model for the ship class is presented (in the examples we do not make the deterministic part explicit). The variables are all binary, taking values 0 or 1. The leaky cause has a fixed value 1. The conditional probability $P(class.ship = 0|chip = s, vessel = v)$ is defined as

$$P(class.ship = 0|ship = s, vessel = v) \quad = \quad (p_1)^s \cdot (p_2)^v \cdot p_0 \ ,$$

where $s \in \{0, 1\}$ is the state of feature *ship* and $v \in \{0, 1\}$ is the state of feature *vessel*. The values of parameters p_1, p_2 were estimated to be

$$\begin{aligned} p_1 &= \exp(r_1) = \exp(-0.773407) \doteq 0.461438 \\ p_2 &= \exp(r_2) = \exp(-1.980023) \doteq 0.138066 \end{aligned}$$

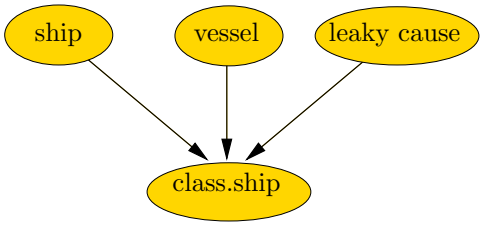

Figure 3: Noisy-or model for the ship class.

and the leaky parameter $p_0 = \exp(r_0)$ was estimated to be

$$p_0 \quad = \quad \exp(r_0) \;=\; \exp(-0.005252) \doteq\; 0.994762 \;\; .$$

This model has accuracy 98.77%.

Example 2 (The multinomial noisy-or model for the ship class)**.** In Figure 4 the structure of the multinomial noisy-or model for the ship class is presented. The variable *ship* takes values from the set $\{0, 1, \ldots, 9\}$, variables *vessel* and

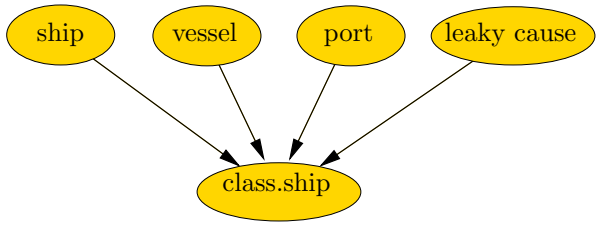

Figure 4: Noisy-or model for the ship class.

port take values from the set $\{0, 1, \ldots, 5\}$. The leaky cause has fixed state 1. The conditional probability $P(Class.ship = 0|Ship = s, Vessel = v, Port)$ is defined as

$$P(Class.ship = 0|Ship = s, Vessel = v) \quad = \quad (p_1)^s \cdot (p_2)^v \cdot (p_3)^p \cdot p_0 \;\; ,$$

where $s \in \{0, 1, \ldots, 9\}$ is the state of feature *ship*, $v \in \{0, 1, \ldots, 5\}$ is the state of feature *vessel*, and $p \in \{0, 1, \ldots, 5\}$ is the state of feature *port*. The values of parameters p_1, p_2, p_3 were estimated to be

$$\begin{aligned} p_1 \quad &= \quad \exp(r_1) \;=\; \exp(-0.467276) \;\doteq\; 0.626707 \\ p_2 \quad &= \quad \exp(r_2) \;=\; \exp(-1.361929) \;\doteq\; 0.256166 \\ p_3 \quad &= \quad \exp(r_3) \;=\; \exp(-0.500009) \;\doteq\; 0.606525 \end{aligned}$$

and the leaky parameter $p_0 = \exp(r_0)$ was estimated to be

$$p_0 \quad = \quad \exp(r_0) \;=\; \exp(-0.001273) \doteq\; 0.998728 \;\; .$$

This model has accuracy 99.13%, which is higher than the accuracy of noisy-or from Example 1.

6 Conclusions

In this paper we proposed a generalization of the popular noisy-or model to multivalued explanatory variables. We showed the correspondence of this model to the Poisson family of generalized linear models and applied iteratively reweighted least squares method to learning of these models. In the experiments with the Reuters text collection the standard noisy-or performed better for larger models, while the multinomial noisy-or was better for smaller models.

Acknowledgments

I am grateful to Remco Bouckaert from The University of Auckland, New Zealand for his suggestion to consider generalizations of noisy-or classifier [13] to multinomial variables.

References

[1] Ch. Apté, F. Damerau, and S. M. Weiss. Automated learning of decision rules for text categorization. *ACM Transactions on Information Systems*, 12(3):233–251, 1994.

[2] F. J. Díez and M. J. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical Report CISIAD-06-01, UNED, Madrid, Spain, 2006.

[3] F. J. Díez and S. F. Galán. An efficient factorization for the noisy MAX. *International Journal of Intelligent Systems*, 18:165–177, 2003.

[4] D. Heckerman and J. Breese. A new look at causal independence. In *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA*, pages 286–292. Morgan Kaufmann, 1994.

[5] Max Henrion. Practical issues in constructing a Bayes' Belief Network. In *Proceedings of the Third Conference Annual Conference on Uncertainty in Artificial Intelligence*, pages 132–139. AUAI Press, 1987.

[6] F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs, 2nd ed.* Springer, 2007.

[7] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1989.

[8] R. A. Miller, F. E. Fasarie, and J. D. Myers. Quick medical reference (QMR) for diagnostic assistance. *Medical Computing*, 3:34–48, 1986.

[9] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, CA, 1988.

[10] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[11] P. Savicky and J. Vomlel. Exploiting tensor rank-one decomposition in probabilistic inference. *Kybernetika*, 43(5):747–764, 2007.

[12] Sampath Srinivas. A generalization of the noisy-or model. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 208–215. Morgan Kaufmann, 1993.

[13] J. Vomlel. Noisy-or classifier. *International Journal of Intelligent Systems*, 21:381–398, 2006.

Fuzzy Relational Compositions Based on Generalized Intermediate Quantifiers

Matin Štěpnička
Division of UO – IRAFM
Centre of Excellence IT4Innovations
Martin.Stepnicka@osu.cz

Michal Holčapek
Division of UO – IRAFM
Centre of Excellence IT4Innovations
Michal.Holcapek@osu.cz

Abstract

Fuzzy relational composition has been extensively studied by many authors. Especially, we would like to highlight initial studies of the fuzzy relational compositions motivated by their applications to medical diagnosis by W. Bandler and L.J. Kohout. In this investigation, we provide a brief survey of the main pilots of the fuzzy relational compositions and repeat the medical diagnosis motivation. We revisit these types of compositions and demonstrate a big gap between sup-T compositions and Bandler-Kohout inf-R compositions (also called subproduct, superproduct and square product). We show, that the gap is caused by the use of the classical (existential and universal) quantifiers. Therefore, we suggest an implementation of generalized intermediate quantifiers and demonstrate their influence on filling the gap with a brief prospect on the potential in applications. This paper is an introductory study that should demonstrate the needs and motivate further research in this area.

1 Introduction

Fuzzy relational composition are widely used in many areas of fuzzy mathematics, including the formal constructions of fuzzy inference systems [1, 2, 3, 4] , medical diagnosis [5] or architectures for information processing and protection of IT systems [6]. Since late 70's and early 80's when W. Bandler and L.J. Kohout studied classical relational compositions and extended the concept in order to define and deal with fuzzy relational compositions, these area became deeply elaborated by numerous researchers. Let us recall mainly R. Bělohlávek's book [7], an article by B. De Baets and E.E. Kerre [8] and finally an exhaustive investigation in the so called Fuzzy Class Theory [9] by L. Běhounek and M. Daňková [10].

As one may find on the very early articles [5, 11], the fuzzy relational compositions came to live as very natural generalization of well motivated compositions

of classical relations. For example, the basic sup-T compositions is nothing else but a generalization of the classical composition of two classical binary relations. It is sufficient to consider binary fuzzy relations and to deal with the operations that serve as interpretations of fuzzy connectives, which involves the t-norm T in the formula that gave rise to the notion "sup-T composition". However, what remained untouched and up to the best knowledge of the authors, never generalized so far, is the nature of the quantifiers that are used in the definitions of the compositions. Particularly, sup-T compositions that use the operation of supremum, implicitly employ the existential quantifier while inf-R compositions that use the operation of infimum, implicitly employ the universal quantifier. The fact that there is nothing in between the option of the existential quantifier where just one element is enough to result the truth and the other option of the universal quantifier where all elements have to fulfill a given formula in order to result the truth, may be very limiting in distinct applications. Therefore, the introduction of fuzzy relational compositions based on generalized quantifiers, such as '**Most**' or '**Many**', is a well motivated natural step that is, up to our best knowledge, firstly elaborated in this paper.

2 Relational compositions and fuzzy relational compositions

2.1 Circlet composition

Let us consider three non-empty finite universes X, Y, Z of elements. Following the work of W. Bandler and L.J. Kohout [5], for the sake of illustrative nature, we can assume that X is a finite set of patients, Y is a finite set of symptoms and Z is a finite set of diseases.

Let us be given two binary relations $R \subseteq X \times Y$ and $S \subseteq Y \times Z$, i.e., if a pair $(x, y) \in X \times Y$ belongs to relation R then it means that patient x has symptom y and similarly, if a pair $(y, z) \in Y \times Z$ belongs to relation S then it means that symptom y belongs to disease z. Both relations are usually at disposal since R can be easily obtained by asking patients or by measuring symptoms (body temperature, cholesterol, blood pressure etc.) and S constitutes an expert medical knowledge that is at disposal e.g. in literature. The usual diagnosis task of a physician is from the mathematical point of view nothing else but a composition of these two relation in order to obtain a relation between patients and diseases. In order words, to state what are the potential diseases of a given patient. Obviously, formally, a similar job may be done by a relational composition using e.g. the standard (also "circlet" [6]) one which gives a binary relation $R \circ S$ on $X \times Z$:

$$\begin{array}{lllllll} R & \subseteq & X & \times & Y & & \\ S & \subseteq & & & Y & \times & Z \\ \hline R \circ S & \subseteq & X & & & \times & Z. \end{array}$$

Relation $R \circ S \subseteq X \times Z$ is given as follows

$$R \circ S = \{(x,z) \in X \times Z \mid \exists\, y \in Y : (x,y) \in R \,\&\, (y,z) \in S\}. \tag{1}$$

Using the fact that the existential quantifier may be interpreted by the operation of supremum, formula (1) may be rewritten into the following functional form:

$$\chi_{(R\circ S)}(x,z) = \bigvee_{y\in Y} (\chi_R(x,y) \wedge \chi_S(y,z)) \tag{2}$$

where χ_R, χ_S and $\chi_{(R\circ S)}$ denote characteristic functions of relations R, S and $R \circ S$ respectively, and symbol $\wedge$ denotes the minimum.

The relation $R \circ S$ then expresses a sort of suspicion of a disease for a particular patient. It is a basic relation – for each patient it is sufficient to have only a single symptom related to a particular disease in order to detect the suspicion. Thus, a patient having a very general symptom related to many diseases is immediately suspicious of having all these diseases.

2.2 Triangle and square compositions

Obviously, some sort of more accurate specification or strengthening the initial suspicion is desirable. This may be done by further composition suggested in [5]. Particularly, we talk about two triangle and one square compositions that are denoted by symbols $\triangleleft$, $\triangleright$ and $\square$, respectively. All of the mentioned compositions are defined on the same universes, i.e. similarly to the circlet compositions:

R	$\subseteq$	X	$\times$	Y		
S	$\subseteq$			Y	$\times$	Z
$R@S$	$\subseteq$	X			$\times$	Z

where $@ \in \{\triangleleft, \triangleright, \square\}$.

The first triangle compositions, also called *Bandler-Kohout subproduct (abb. BK-subproduct)*, defines the relation $R \triangleleft S \subseteq X \times Z$ as follows

$$R \triangleleft S = \{(x,z) \in X \times Z \mid \forall\, y \in Y : (x,y) \in R \Rightarrow (y,z) \in S\}. \tag{3}$$

In other words, the BK subproduct is defined as a relation of patients and diseases such that, for all symptoms that a given patient has it holds, that they belong to the given diseases that is in a relation with the patient. Note, that the patient does not have to have all of the symptoms and that he or she may have symptoms belonging also to other diseases however, only in case the symptoms simultaneously belong to the given disease.

Using the fact that the universal quantifier may be interpreted by the operation of infimum, formula (3) may be rewritten into the following functional form:

$$\chi_{(R\triangleleft S)}(x,z) = \bigwedge_{y\in Y} (\chi_R(x,y) \Rightarrow \chi_S(y,z)) \tag{4}$$

where symbol $\Rightarrow$ expresses the binary operation of the classical implication.

A sort of inverse strengthening of the initial suspicion to the BK subproduct is provided by the second triangle composition called *Bandler-Kohout superproduct (abb. BK-superproduct)* which defines the relation $R \rhd S \subseteq X \times Z$ as follows

$$R \rhd S = \{(x,z) \in X \times Z \mid \forall\, y \in Y : (x,y) \in R \Leftarrow (y,z) \in S\}. \tag{5}$$

In other words, the BK superproduct is defined as a relation of patients and diseases such that, for all symptoms that belong to a given disease it holds, that the given patient that is in a relation with the disease necessarily must have them. Note, that the patient may have symptoms belonging also to other diseases however, he or she cannot miss any of the symptoms belonging to the given disease.

Similarly to the previous case, we may rewrite formula (5) into the following functional form:

$$\chi_{(R \rhd S)}(x,z) = \bigwedge_{y \in Y} \left(\chi_R(x,y) \Leftarrow \chi_S(y,z)\right). \tag{6}$$

Finally, we may recall the square compositions, also called *Bandler-Kohout square product (abb. BK square product)*, which defines the relation $R \,\square\, S \subseteq X \times Z$ as follows

$$R \,\square\, S = \{(x,z) \in X \times Z \mid \forall\, y \in Y : (x,y) \in R \Leftrightarrow (y,z) \in S\} \tag{7}$$

and may be easily rewritten to the following functional form:

$$\chi_{(R \,\square\, S)}(x,z) = \bigwedge_{y \in Y} \left(\chi_R(x,y) \Leftrightarrow \chi_S(y,z)\right). \tag{8}$$

One may see, that the square compositions models an ideal example when a given patient has all the symptoms of a given disease and all the symptoms of the patient belong to the given disease, i.e., he or she has no symptoms that we could not connect to the disease.

2.3 Compositions of fuzzy relations

Since usual symptoms such as *high temperature*, *increased cholesterol* or *very high blood pressure* are basically vaguely specified and imprecisely measured (all these values oscillate during a day) and very often some symptoms do not clearly or necessarily belong to a given diseases however, they might belong to it under some assumptions or conditions, the extension of the compositions for fuzzy relations $R \underset{\sim}{\subseteq} X \times Y$ and $S \underset{\sim}{\subseteq} Y \times Z$ was highly desirable. Obviously, since such a extension causes that we deal with fuzzy relations which contain pairs of elements up to some degrees from the interval $[0,1]$, we have to take into account appropriate operations. Basically, it is appropriate to deal with a residuated lattice as the underlying algebraic structure [12] and the used operations will be left-continuous t-norms [13] and their residual (bi)implications [14].

In this article, we only briefly recall the basic definitions of the fuzzy relational compositions as introduced by W. Bandler and L.J. Kohout.

Definition 1 Let X, Y, Z be non-empty universes, let $R \underset{\sim}{\subseteq} X \times Y$, $S \underset{\sim}{\subseteq} Y \times Z$ and let $*$ be a left-continuous t-norm. Then the $\circ$ *(sup-$*$) composition* of fuzzy relations R and S is a fuzzy relation on $X \times Z$ defined as follows:

$$(R \circ S)(x, z) = \bigvee_{y \in Y} (R(x, y) * S(y, z))$$

for all $x \in X$ and $z \in Z$.

Analogously, we can define generalized Bandler-Kohout products.

Definition 2 Let X, Y, Z be non-empty universes, let $R \underset{\sim}{\subseteq} X \times Y$, $S \underset{\sim}{\subseteq} Y \times Z$ and let $\to$ be a residual implication. Then the $\triangleleft, \triangleright, \square$ *(inf-$\to$, inf-$\leftarrow$ and inf-$\leftrightarrow$) compositions* of fuzzy relations R and S are fuzzy relations on $X \times Z$ defined as follows:

$$(R \triangleleft S)(x, z) = \bigwedge_{y \in Y} (R(x, y) \to S(y, z)),$$
$$(R \triangleright S)(x, z) = \bigwedge_{y \in Y} (R(x, y) \leftarrow S(y, z)),$$
$$(R \square S)(x, z) = \bigwedge_{y \in Y} (R(x, y) \leftrightarrow S(y, z)),$$

for all $x \in X$ and $z \in Z$.

Since $*$ is a t-norm, often denoted by the capital T, the sup-$*$ composition is also called the sup-T composition. Similarly, the Bandler-Kohout products since being constructed with help of the infimum and the residual operation, are called inf-R compositions.

Remark 1 *Note, that for $x \in X$ such that $R(x, y) = 0$ for all $y \in Y$, the composed relation $(R \triangleleft S)(x, z) = 1$ for any $z \in Z$. More illustratively, if there is a patient with no symptoms, it is trivially true that for any given disease, all his symptoms are related to the given disease. Similarly, for $z \in Z$ such that $S(y, z) = 0$ for all $y \in Y$, the composed relation $(R \triangleright S)(x, z) = 1$ for any $x \in X$. On the other hand, in such situations $(R \circ S)(x, z) = 0$ so, the inf-R compositions may hardly be viewed as strengthening of a suspicion determined by the sup-T composition if no suspicion was determined. De Baets and Kerre in [8] approached this problem by a redefinition of the original inf-R compositions where an existence of joining element $y \in Y$ (symptom) is assumed. Although we are aware of this solution that is both, mathematically elegant and practically useful in distinct applications, in this preliminary investigation we stay stuck to the original definitions and we leave the investigation of the later modification for further studies.*

As we may see from Definitions 1-2, the generalizations focus on the relations (newly fuzzy relations) and the internal operations only. However, concerning

the outer operations, they remain unchanged. Particularly, the definitions still deal with the supremum and the infimum and thus, inherently use the existential and universal quantifiers, respectively. The subsequent example shows the gap above mentioned between such compositions and its potential drawback for applications.

Example 1 *Let us consider just a simple artificial example for demonstrative and motivation purposes. Let us consider $*$ and $\rightarrow$ to be the Łukasiewicz operations and the following fuzzy relations R and S:*

R	y_1	y_2	y_3	y_4
x_1	0.9	1	0.8	0
x_2	0	0.9	0.8	0.1
x_3	0	0.8	0.9	0
x_4	0	0	1	0.9

S	z_1	z_2	z_3	z_4	z_5
y_1	1	1	0.1	0.9	0
y_2	0.9	0.2	0.9	0	1
y_3	0	1	0	1	1
y_4	1	0	0.7	0.1	0.9

If we compute the standard fuzzy relational compositions, e.g., the sup-T composition and the Bandler-Kohout square products:

$R \circ S$	z_1	z_2	z_3	z_4	z_5
x_1	0.9	0.9	0.9	0.8	1
x_2	0.8	0.8	0.8	0.8	0.9
x_3	0.7	0.9	0.7	0.9	0.9
x_4	0.9	1	0.6	1	1

$R \square S$	z_1	z_2	z_3	z_4	z_5
x_1	0	0.2	0.2	0	0.1
x_2	0	0	0.2	0.1	0.2
x_3	0	0	0.1	0.1	0.1
x_4	0	0	0	0.1	0

we will see, that while all patients are suspicious of having all diseases in a high degree when using $\circ$, if we want to strengthen the suspicion with help of $\square$, no patients is suspicious of having any disease in a high degree anymore.

3 Generalized quantifiers

3.1 Generalized quantifiers based on fuzzy measures

In the above sections, we have recalled relational compositions and fuzzy relational compositions. We also showed their drawback demonstrated why a use of *generalized intermediate quantifiers* [15] is highly desirable. Particularly, for our goal, the so called *monadic quantifiers of the type $\langle 1 \rangle$ determined by fuzzy measures* [16] will be applied. First of all, let us recall some basic definitions.

Definition 3 Let $U = \{u_1, \ldots, u_n\}$ be a finite universe, $\mathcal{P}(U)$ denote the power set of U and $\mu : \mathcal{P}(U) \rightarrow [0, 1]$ be a normalized fuzzy measure, i.e., a monotone mapping with $\mu(\emptyset) = 0$ and $\mu(U) = 1$. We say that the fuzzy measure μ is invariant with respect to cardinality, if the following condition holds:

$$\forall A, B \in \mathcal{P}(U) : |A| = |B| \Rightarrow \mu(A) = \mu(B)$$

where $|\cdot|$ denotes the cardinality of a set.

Example 2 *The measure called* relative cardinality *and given by*

$$\mu_{rc}(A) = \frac{|A|}{|U|} \tag{9}$$

is invariant w.r.t. cardinality. If $f : [0,1] \to [0,1]$ is a non-decreasing mapping with $f(0) = 0$ and $f(1) = 1$ then μ defined as $\mu(A) = f(\mu_{rc}(A))$ is again a fuzzy measure that is invariant w.r.t. cardinality.

Note, that all the intensions of linguistic evaluative expressions [17] of the type `Big` *and modified by arbitrary linguistic hedge (e.g.* `More or less, Very, Roughly, Extremely` *etc.) are fuzzy sets on $[0,1]$ that fulfill the boundary conditions and thus, may be used in order to modify the original relative cardinality.*

In the sequel, we will deal only with such fuzzy measures that are created by a modification of the relative cardinality by an appropriate fuzzy set (cf. Definition 3.7 in [16])

Definition 4 Let U be non-empty finite universe and μ be a fuzzy measure on U that is invariant w.r.t. cardinality. A mapping $Q : \mathcal{F}(U) \to [0,1]$ defined by

$$Q(C) = \bigvee_{D \in \mathcal{P}(U) \setminus \{\emptyset\}} \left(\left(\bigwedge_{u \in D} C(u) \right) * \mu(D) \right), \quad C \in \mathcal{F}(U) \tag{10}$$

where $*$ is a left-continuous t-norm, is called *fuzzy quantifier determined by fuzzy measure μ.*

Example 3 *Let us assume that the fuzzy measures μ defined as follows*

$$\mu^{\forall}(B) = \begin{cases} 1 & B \equiv U \\ 0 & \textit{otherwise}, \end{cases} \qquad \mu^{\exists}(B) = \begin{cases} 0 & B \equiv \emptyset \\ 1 & \textit{otherwise}. \end{cases} \tag{11}$$

Then the derived quantifiers $Q^{\forall}$ and $Q^{\exists}$ are exactly the classical universal and existential quantifiers, respectively. Note that similarly one may define fuzzy quantifiers also for other algebras of sets than power sets considered here.

One can immediately see, that formula (10) is not very appropriate from the computational point of view since it requires calculation over all sets from $\mathcal{P}(U) \setminus \{\emptyset\}$. However, we may use the property of fuzzy measure being invariant w.r.t. cardinality and simply show that the fuzzy quantifier may be very efficiently computed.

Theorem 1 *Let Q be a fuzzy quantifier on U determined by a fuzzy measure μ that is invariant w.r.t. cardinality. Then*

$$Q(C) = \bigvee_{i=1}^{n} C(u_{\pi(i)}) * \mu(\{u_1, \ldots, u_i\}), \quad C \in \mathcal{F}(U) \tag{12}$$

where π is a permutation on U such that $C(u_{\pi(1)}) \geq C(u_{\pi(2)}) \geq \cdots \geq C(u_{\pi(n)})$.

Proof: Let C be an arbitrary fuzzy set. It is easy to see that, for any $i = 1, \ldots, n$ and $D \in \mathcal{P}(U)$ with $|D| = i$, it holds

$$C(u_{\pi(i)}) = \bigwedge_{u \in \{u_{\pi(1)}, \ldots, u_{\pi(i)}\}} C(u) \geq \bigwedge_{u \in D} C(u).$$

The statement immediately follows from the invariance of μ w.r.t. cardinality. □

In other words, if we again apply the fuzzy measure that is constructed from the relative cardinality by some modifying fuzzy set f, formula (12) turns into the following equality:

$$Q^f(C) = \bigvee_{i=1}^{n} C(u_{\pi(i)}) * f(i/n) \tag{13}$$

which is easy to calculate.

3.2 Fuzzy relational compositions based on generalized quantifiers

In this part of the text, we directly apply the above introduced theory of generalized quantifiers to our problem of fuzzy relational compositions.

Let us recall, e.g., the definition of the Bandler-Kohout subproduct of two classical relations R and S given by formula (3)

$$R \lhd S = \{(x, z) \in X \times Z \mid \forall\, y \in Y : (x, y) \in R \Rightarrow (y, z) \in S\}.$$

Using the generalized quantifier Q defined on Y, formula (3) modifies into

$$R \lhd^Q S = \{(x, z) \in X \times Z \mid Q\, y \in Y : (x, y) \in R \Rightarrow (y, z) \in S\}. \tag{14}$$

For example, if Q represents the quantifier '**Most**', the formula can be easily read as a set of pairs of patients and diseases such that for most of the symptoms that a given patient has it holds, that they belong to the given disease, which is very natural.

Transformation of the procedure leading from (3) to (14) in the case of fuzzy relational compositions is then straightforward and realized in the following definition.

Definition 5 Let X, Y, Z be non-empty finite universes, let $R \subsetneq X \times Y$, $S \subsetneq Y \times Z$, let $*$ be a left-continuous t-norm and $\to$ be its residual implication. Let μ be a fuzzy measure on Y that is invariant w.r.t. cardinality and let Q be a fuzzy quantifier on Y determined by the fuzzy measure μ. Then the $\circ^Q, \lhd^Q, \rhd^Q, \Box^Q$ *compositions* of fuzzy relations R and S are fuzzy relations on $X \times Z$ defined

as follows:

$$(R \circ^Q S)(x,z) = \bigvee_{D \in \mathcal{P}(Y) \setminus \{\emptyset\}} \left(\left(\bigwedge_{y \in D} R(x,y) * S(y,z) \right) * \mu(D) \right),$$

$$(R \triangleleft^Q S)(x,z) = \bigvee_{D \in \mathcal{P}(Y) \setminus \{\emptyset\}} \left(\left(\bigwedge_{y \in D} R(x,y) \to S(y,z) \right) * \mu(D) \right),$$

$$(R \triangleright^Q S)(x,z) = \bigvee_{D \in \mathcal{P}(Y) \setminus \{\emptyset\}} \left(\left(\bigwedge_{y \in D} R(x,y) \leftarrow S(y,z) \right) * \mu(D) \right),$$

$$(R \,\square^Q S)(x,z) = \bigvee_{D \in \mathcal{P}(Y) \setminus \{\emptyset\}} \left(\left(\bigwedge_{y \in D} R(x,y) \leftrightarrow S(y,z) \right) * \mu(D) \right),$$

for all $x \in X$ and $z \in Z$.

Though the definition is general and enables to use any quantifier Q, obviously, for application purposes, when dealing with the inf-R compositions $\triangleleft^Q, \triangleright^Q, \square^Q$ quantifiers weakening the universal quantifier such as '**Most**' or '**Many**' are expected to be applied. These quantifiers may be applied for example using the fuzzy sets modeling the meaning of evaluative linguistic expression `Very Big` and `not Small`, see [15]. Similarly, in the case of sup-T composition $\circ^Q$, we should apply quantifiers that slightly strengthen the expectations from the existential quantifiers, such as '**A Few**'[1] that may be modeled by a fuzzy sets representing the meaning of the expression `not Very Small`.

Corollary 1 *Let μ be a fuzzy measure that is constructed from the relative cardinality by the modification using function f. Then for all $x \in X$ and $z \in Z$:*

$$(R \circ^Q S)(x,z) = \bigvee_{i=1}^{n} \big(\big(R(x, y_{\pi(i)}) * S(y_{\pi(i)}, z) \big) * f(i/n) \big),$$

$$(R \triangleleft^Q S)(x,z) = \bigvee_{i=1}^{n} \big(\big(R(x, y_{\pi(i)}) \to S(y_{\pi(i)}, z) \big) * f(i/n) \big),$$

$$(R \triangleright^Q S)(x,z) = \bigvee_{i=1}^{n} \big(\big(R(x, y_{\pi(i)}) \leftarrow S(y_{\pi(i)}, z) \big) * f(i/n) \big),$$

$$(R \,\square^Q S)(x,z) = \bigvee_{i=1}^{n} \big(\big(R(x, y_{\pi(i)}) \leftrightarrow S(y_{\pi(i)}, z) \big) * f(i/n) \big),$$

where π is a permutation such that (putting $\circledast \in \{*, \to, \leftarrow, \leftrightarrow\}$)

$$R(x, y_{\pi(i)}) \circledast S(y_{\pi(i)}, z) \geq R(x, y_{\pi(i+1)}) \circledast S(y_{\pi(i+1)}, z), \quad i = 1, \ldots, n-1.$$

[1] Note the difference between '**A Few**' and '**Few**'. Quantifier '**Few**' implicitly means something like 'not many' and thus, the application of a non-decreasing function f is impossible and the quantifier requires a different construction, see [15].

Indeed, the original compositions are special cases of the newly defined ones. Using the fuzzy measure $\mu^{\forall}$ given by (11), one may easily check that $R \lhd S \equiv R \lhd^{\forall} S$ and similarly that $R \rhd S \equiv R \rhd^{\forall} S$, $R \,\square\, S \equiv R \,\square^{\forall}\, S$. Indeed, since $f(i/n) = 0$ for all $i < n$ and $f(1) = 1$ then

$$(R \lhd^{\forall} S)(x, z) = \big(R(x, y_{\pi(n)}) \to S(y_{\pi(n)}, z)\big) * f(n/n)$$

which due to the fact that

$$R(x, y_{\pi(n)}) \to S(y_{\pi(n)}, z) = \bigwedge_{i=1}^{n} (R(x, y_i) \to S(y_i, z))$$

proves $R \lhd S \equiv R \lhd^{\forall} S$. The other equalities may be proved analogously.

Example 4 *Let us consider fuzzy relations from Example 1 and assume again the Łukasiewicz operations. Furthermore, let us consider the fuzzy set modeling the meaning of the linguistic expression* `Roughly Big` *which enables us to construct a generalized quantifier* **'Majority'**. *In a standard context, this fuzzy set takes values* $RoBi(1/4) = 0, RoBi(2/4) = 0, RoBi(3/4) = 0.95$ *and* $RoBi(1) = 1$. *Then the newly suggested fuzzy relational composition* $\square^{Q}$, *that is built using the above mentioned fuzzy set, gives the following results:*

$R \,\square^{Q}\, S$	z_1	z_2	z_3	z_4	z_5
x_1	*0.15*	*0.75*	*0.2*	*0.75*	*0.1*
x_2	*0.05*	*0.25*	*0.35*	*0.1*	*0.75*
x_3	*0*	*0.35*	*0.25*	*0.15*	*0.75*
x_4	*0*	*0.05*	*0.05*	*0.15*	*0.95*

As we may see, composition $\square^{Q}$ *strengthened the original suspicion given by* $\circ$ *but no so strictly as by* $\square$ *which would require to have* **'All'** *symptoms of a given disease and to have* **'All'** *symptoms related a given disease. Composition* $\square^{Q}$ *requires only* **'Majority'** *of the symptoms to have such properties (connections) and thus, better captures the natural vagueness and fuzziness of the real-world situation which leads to a strengthen suspicion of diseases* z_5 *(in the case of patients* x_2, x_3, x_4*) and of diseases* z_2, z_4 *(in the case of patient* x_1*).*

4 Properties

In the sections above, we have recalled classical and fuzzy relational composition, we have demonstrated that there is a big gap between sup-T and inf-R compositions that motivated us to involve generalized quantifiers, we have recalled brief facts about monadic quantifiers of type $\langle 1 \rangle$ and used them in a construction of fuzzy relational compositions based on generalized quantifiers. Many appropriate properties were proved for the original classical as well as fuzzy relational compositions. In this section, we face the question whether the same or similar properties may be valid also for the compositions based on generalized quantifiers. As we will show, the answer is positive.

Theorem 2 *Let X, Y, Z, U are finite universes and let $R_1, R_2 \subsetneq X \times Y$, $S_1, S_2 \subsetneq Y \times Z$ and $T \subsetneq Z \times U$. Furthermore, let $\cup, \cap$ denote the Gödel union and intersection, respectively. Then*

1. $R \circ^Q (S \circ^Q T) = (R \circ^Q S) \circ^Q T$
2. $R \,\square^Q\, S \leq (R \triangleleft^Q S) \cap (R \triangleright^Q S)$
3. $R_1 \leq R_2 \;\Rightarrow\; (R_1 \circ^Q S) \subseteq (R_2 \circ^Q S)$ *and* $S_1 \leq S_2 \;\Rightarrow\; (R \circ^Q S_1) \subseteq (R \circ^Q S_2)$
4. $R_1 \leq R_2 \;\Rightarrow\; (R_1 \triangleleft^Q S) \supseteq (R_2 \triangleleft^Q S)$ *and* $(R_1 \triangleright^Q S) \subseteq (R_2 \triangleright^Q S)$
5. $(R_1 \cup R_2) \circ^Q S = (R_1 \circ^Q S) \cup (R_2 \circ^Q S)$
6. $(R_1 \cap R_2) \triangleleft^Q S = (R_1 \triangleleft^Q S) \cup (R_2 \triangleleft^Q S)$
7. $(R_1 \cup R_2) \triangleright^Q S = (R_1 \triangleright^Q S) \cup (R_2 \triangleright^Q S)$
8. $(R_1 \cap R_2) \circ^Q S \leq (R_1 \circ^Q S) \cap (R_2 \circ^Q S)$
9. $(R_1 \cup R_2) \triangleleft^Q S \leq (R_1 \triangleleft^Q S) \cap (R_2 \triangleleft^Q S)$
10. $(R_1 \cap R_2) \triangleright^Q S \leq (R_1 \triangleright^Q S) \cap (R_2 \triangleright^Q S)$

Sketch of the proof: All the properties are proved based on the properties of left-continuous t-norms and their residual implications on a linearly order set $[0, 1]$, i.e., using

$$
\begin{aligned}
(a \wedge b) * c &= (a * c) \wedge (b * c), & (a \vee b) * c &= (a * c) \vee (b * c),\\
(a \wedge b) \to c &= (a \to c) \vee (b \to c), & (a \vee b) \to c &= (a \to c) \wedge (b \to c),\\
a \to (b \wedge c) &= (a \to b) \wedge (a \to c), & a \to (b \vee c) &= (a \to b) \vee (a \to c),\\
(a \leftrightarrow b) &= (a \to b) \wedge (a \leftarrow b), & \bigvee_i ((a_i * b) \wedge (a_i * c)) &\leq \bigvee_i (a_i * b) \wedge \bigvee_i (a_i * c)
\end{aligned}
$$

and the antitonicity and the isotonicity of $\to$ in its first and second argument, respectively. Furthermore, the monotonicity properties *3.-4.* are extensively used in proving the latter properties. □

Remark 2 *Obviously, items* 5.-10. *may be also read as follows:*

12. $R \circ^Q (S_1 \cup S_2) = (R \circ^Q S_1) \cup (R \circ^Q S_2)$
13. $R \triangleright^Q (S_1 \cap S_2) = (R \triangleright^Q S_1) \cup (R \triangleright^Q S_2)$
14. $R \triangleleft^Q (S_1 \cup S_2) = (R \triangleleft^Q S_1) \cup (R \triangleleft^Q S_2)$
15. $R \circ^Q (S_1 \cap S_2) \leq (R \circ^Q S_1) \cap (R \circ^Q S_2)$
16. $R \triangleright^Q (S_1 \cup S_2) \leq (R \triangleright^Q S_1) \cap (R \triangleright^Q S_2)$
17. $R \triangleleft^Q (S_1 \cap S_2) \leq (R \triangleleft^Q S_1) \cap (R \triangleleft^Q S_2)$

5 Conclusions

We have shown, how crucial the gap between the sup-T and inf-R compositions might be. Luckily, there is a deep theory of generalized quantifiers that might be very helpful for such situations. Particularly, they allow us to define fuzzy relational compositions with help of linguistically very natural quantifiers such as **'A Few'**, **'Many'**, **'Majority'** or **'Most'**. These quantifiers provide us with a wider choice for such fuzzy relational compositions that may better fit for each particular practical problem. Besides the fundamental definitions and motivating examples, we have shown that many of the appreciated properties of the standard fuzzy relational compositions are preserved. This gives a huge potential to employ the fuzzy relational compositions based on generalized quantifiers in many other areas of application, such as inference systems, where fuzzy relational compositions play a crucial role.

Acknowledgments This investigation was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] W. Pedrycz (1985), Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data, *Fuzzy Sets and Systems*, 16, 163–175.

[2] M. Štěpnička, B. De Baets, and L. Nosková (2007), On additive and multiplicative fuzzy models, in *New Dimensions in Fuzzy Logic and Related Technologies. Proc. 5th EUSFLAT Conference*, M. Štěpnička, V. Novák, and U. Bodenhofer, Eds., vol. 2, Ostrava, 95–102.

[3] M. Štěpnička, B. De Baets, and L. Nosková (2010), Arithmetic fuzzy models, *IEEE Transactions on Fuzzy Systems*, 18, 1058–1069.

[4] M. Štěpnička and B. D. Baets (2013), Interpolativity of at-least and at-most models of monotone single-input single-output fuzzy rule bases, *Information Sciences*, 234, 16–28.

[5] W. Bandler and L. J. Kohout (1978), Fuzzy relational products and fuzzy implication operators, in *Proc. Int. Workshop on Fuzzy Reasoning Theory and Applications*. London: Queen Mary College.

[6] W. Bandler and L. J. Kohout (1985), Relational-product architectures for information processing, *Information Sciences*, 37, 25–37.

[7] R. Bělohlávek (2002), *Fuzzy relational systems: Foundations and principles*. Dordrecht, New York: Kluwer Academic, Plenum Press.

[8] B. De Baets and E. Kerre (1993), Fuzzy relational compositions, *Fuzzy Sets and Systems*, 60, 109–120.

[9] L. Běhounek and P. Cintula (2005), Fuzzy class theory, *Fuzzy Sets and Systems*, 154, 34–55.

[10] L. Běhounek and M. Daňková (2009), Relational compositions in fuzzy class theory, *Fuzzy Sets and Systems*, 160, no. 8, 1005–1036.

[11] W. Bandler and L. J. Kohout (1980), Fuzzy relational products as a tool for analysis and synthesis of the behaviour of complex natural and artificial systems, in *Fuzzy Sets: Theory and Application to Policy Analysis and Information Systems*, S. K. Wang and P. P. Chang, Eds. New York: Plenum Press, 341–367.

[12] V. Novák, I. Perfilieva, and J. Močkoř (1999), *Mathematical Principles of Fuzzy Logic*. Boston: Kluwer Academic Publishers.

[13] E. P. Klement, R. Mesiar, and E. Pap (2000), *Triangular Norms*, ser. Trends in Logic. Dordrecht: Kluwer Academic Publishers.

[14] M. Baczyński and B. Jayaram (2008), *Fuzzy Implications*. Heidelberg: Springer-Verlag.

[15] P. Murinová and V. Novák (2012), A formal theory of generalized intermediate syllogisms, *Fuzzy Sets and Systems*, 186, 47–80.

[16] A. Dvořák and M. Holčapek (2009), L-fuzzy quantifiers of type $\langle 1 \rangle$ determined by fuzzy measures, *Fuzzy Sets and Systems*, 160, 3425–3452.

[17] V. Novák (2008), A comprehensive theory of trichotomous evaluative linguistic expressions, *Fuzzy Sets and Systems*, 159, 2939–2969.

Mining Information from Time Series using Fuzzy Natural Logic

Vilém Novák
IT4Innovations National Supercomputing Center
division of the University of Ostrava
Institute for Research and Applications of Fuzzy Modelling,
30. dubna 22, 701 03 Ostrava 1, Czech Republic
Vilem.Novak@osu.cz

Abstract

In this paper, we present methods using which information expressed in natural language is automatically generated from a given set of time series. The information consists in description of trend of the time series in various time slots that is estimated using F^1-transform. This description is based on the theory of evaluative linguistic expressions.

1 Introduction

Fuzzy natural logic (FNL) is a group of mathematical theories that extend mathematical fuzzy logic in narrow sense. Its goal is to develop a mathematical model of the semantics of some parts of natural language and a model of special human reasoning schemes that employ natural language but are independent on a concrete one. The main constituents of FNL are:

- Theory of evaluative linguistic expressions. Example of such expressions are *small, very small, medium, large*, etc.
- Theory of fuzzy/linguistic IF-THEN rules and logical inference from them.
- Theory of intermediate quantifiers that are linguistic expressions such as *most, almost all, many, a lot of*, etc.

In this paper, we focus on the application enabling us to generate automatically natural language comments to general tendency of time series in a given time slot. The application is based on the theory of evaluative linguistic expressions in combination with the technique of F-transform (see [1, 2]).

By a time series, we understand a discrete stochastic process (see [3, 4])

$$X : Q \times \Omega \longrightarrow \mathbb{R} \tag{1}$$

where Q is a finite set $Q \subset \mathbb{N}$ whose elements are interpreted as time moments.

It follows from (1) that for each $t \in Q$ the mapping $X(t,\omega)$, $\omega \in \Omega$ is a random variable. If we fix $\omega \in \Omega$ then we obtain one realization of (1) and in this case, we will write simply $X(t)$.

Our basic assumption is that $X(t,\omega)$ can be decomposed into three components, namely

$$X(t,\omega) = TC(t) + S(t) + R(t,\omega), \qquad t \in [a,b], \omega \in \Omega, \tag{2}$$

where TC is a *trend-cycle*, S is a seasonal component and R is a random noise. Both TC and S are usual (i.e. non-random) real or complex-valued functions of a real variable.

2 Linguistic evaluation of the behavior of time series

2.1 Estimation of trend using F^1-transform

Important feature of the trend-cycle TC is its *trend* (tendency). We may distinguish it on the whole time series as well as in local time slots (for example, quarter of year, production period, etc.). The course of the time series, however, can be largely volatile and so, it may be quite difficult to recognize its trend. An objective tool using which the trend can be clearly recognized even on the volatile time series can be thus valuable. Such a tool is the F^1-transform because of its ability to estimate the tangent.

Definition 1
Let $f : [a,b] \longrightarrow \mathbb{R}$ be a continuous function and $\mathcal{A} = \{A_0, \ldots, A_n\}$, $n \geq 2$ be a fuzzy partition of $[a,b]$. The vector of linear functions

$$\mathbf{F}^1[f] = (\beta_1^0 + \beta_1^1(x - c_1), \ldots, \beta_{n-1}^0 + \beta_{n-1}^1(x - c_{n-1})) \tag{3}$$

is called the F^1-transform of f with respect to the fuzzy partition $\mathcal{A}$, where

$$\beta_k^0 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)A_k(x)dx}{h}, \tag{4}$$

$$\beta_k^1 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)(x - x_k)A_k(x)dx}{\int_{c_{k-1}}^{c_{k+1}} (x - c_k)^2 A_k(x)dx} \tag{5}$$

for every $k = 1, \ldots, n-1$.

The following theorem plays an important role in our application to time series trend evaluation.

Theorem 1
Let $\mathcal{A} = \{A_1, \ldots, A_{n-1}\}$ be an h-uniform partition of $[a, b]$, let functions f and $A_k \in \mathcal{A}$, $k = 1, \ldots, n-1$, be four times continuously differentiable on $[a, b]$. Finally, let $\mathbf{F}^1[f]$ be the F^1-transform (3) of f. Then

$$\beta_k^1 = f'(c_k) + O(h), \qquad k = 1, \ldots, n-1. \tag{6}$$

On the basis of the previous theory, may define trend $T(A_k)$ of the time series X in the area characterized by the fuzzy set $A_k \in \mathcal{A}$ by

$$T(A_k) = \beta_k^1 \tag{7}$$

where β_k^1 is the coefficient (5). Hence, $T(A_k)$ is a weighted average tangent of the function $X(t)$ over the area determined by the fuzzy set (basic function) $A_k \in \mathcal{A}$.

2.2 Evaluative linguistic expressions

The formal theory of the semantics of *evaluative linguistic expressions* was in detail described in [5]. Essential concept in this theory is that of (linguistic) *context*. For evaluative expressions, it is determined by a triple of real numbers $\langle v_L, v_S, v_R \rangle$ where $v_L < v_S < v_R$ ($\in \mathbb{R}$). These numbers represent the *smallest, typically medium*, and the *largest* thinkable values, respectively. The context is thus a set

$$w = \{x \mid v_L \leq x \leq v_R\} \tag{8}$$

together with three distinguished points $DP(w) = \langle v_L, v_S, v_R \rangle$. By W we denote the set of all contexts (8) and by *EvExpr* the set of all considered evaluative expressions. Each evaluative expression $Ev \in EvExpr$ is assigned the meaning which is a function

$$\mathrm{Int}(Ev) : W \longrightarrow \mathcal{F}(\mathbb{R}).$$

We will call this function *intension* of the evaluative expression Ev. It assigns to each context $w \in W$ a fuzzy set $\mathrm{Ext}_w(Ev) \subsetneq w$ called *extension* of the expression Ev in the context $w \in W$.

We will distinguish abstract evaluative expressions, i.e. expressions such as *small, weak, very strong*, etc., that alone do not talk about any specific objects and *evaluative linguistic predications* such as "temperature is high, expenses are extremely low, the building is quite ugly", etc. In general, the latter have the surface form

$$\langle \text{noun} \rangle \text{ is } \langle \text{simple evaluative expression} \rangle \tag{9}$$

where

$$\langle \text{simple evaluative expression} \rangle := \langle \text{hedge} \rangle \langle \text{TE-adjective} \rangle,$$

$\langle$hedge$\rangle$ is a linguistic hedge (for example *very, rather, extremely, more or less, roughly*, etc.) and $\langle$TE-adjective$\rangle$ is a trichotomous evaluative adjective (for example *small, medium, big, large, weak, good*, etc.). The "is" takes here the role of a copula assigning property to objects and is not treated as a genuine verb.

2.3 Linguistic evaluation of the trend of time series

The theory of evaluative expressions provides a special function of *local perception*

$$LPerc : w \times W \longrightarrow EvExpr\,. \tag{10}$$

This function assigns to each value $x \in w$ in each context $w \in W$ an evaluative expression of the form (9). The function is constructed in such a way that given a linguistic context $w \in W$ and a value $x \in w$, the result of (10) is the *most plausible* evaluative expression that characterizes x in the given context w. Using (10), we will generate linguistic evaluation of the trend of time series. First we must specify the context. We start with specification of what does it mean "extreme increase (decrease)". In practice, it can be determined as the largest acceptable difference of time series values with respect to a given (basic) time interval. Hence, mathematically we speak about the tangent. The usual basic time interval is 12 months, 31 days, etc. depending on the kind of the time series. Thus, the context is determined by the three distinguished values v_L, v_S, v_R of the tangent. The largest tangent v_R is determined in the way mentioned above while the smallest one is usually $v_L = 0$. The typical medium value v_S is determined analogously as v_R. The result is the context $w_{tg} = \langle v_L, v_S, v_R \rangle$.

Now, we can linguistically characterize the trend $T(A_k)$ in (7) with respect to the context w_{tg}, i.e. we will automatically generate evaluative linguistic expressions using the function $LPerc(T(A_k), w_{tg})$ of local perception (10): This is justified by the fact that $T(A_k)$ is an average tangent over an area covered by the basic function $A_k \in \mathcal{A}$.

Predications using which we linguistically evaluate time series trend have specific form. The basic characteristic is *sign* of the trend. This is characterized by a special word, namely "+" is expressed by the word *increasing* (or *increase*) and "−" by *decreasing* (or *decrease*). This can further be completed by special expressions characterizing its *gradient*. Moreover, the obtained expressions are apparently subject to ordering that is similar to the natural ordering of the "standard" evaluative expressions. We conclude that the general syntactic form of expressions characterizing trend is either (a) or (b) specified below:

(a)

$$Trend \text{ is } \langle \text{gradient} \rangle \tag{11}$$

where

$$\langle \text{gradient} \rangle := \mathsf{stagnating} | \langle \text{hedge} \rangle \langle \text{sign} \rangle, \tag{12}$$

$$\langle \text{sign} \rangle := \mathsf{increasing} | \mathsf{decreasing} \tag{13}$$

and

$$\langle \text{hedge} \rangle := \mathsf{negligibly} | \mathsf{slightly} | \mathsf{somewhat} | \mathsf{clearly} | \mathsf{roughly} | \mathsf{sharply} | \mathsf{significantly}.$$

(b) In some cases, however, only the feature *increase (decrease) of trend* is evaluated:

$$\langle \text{sign of trend} \rangle \text{ is } \langle \text{special hedge} \rangle \tag{14}$$

where $\langle \text{sign of trend} \rangle :=$ increase|decrease and

$$\langle \text{special hedge} \rangle := \text{negligible|slight|small|clear|rough|large|fairly large|quite large|significant|huge.}$$

Note that, in fact, the *increase (decrease)* (of trend) is evaluated both in (11) as well as in (14). The difference in their use depends on the syntactic specificities but *not on their semantics*! Some cases, for example "trend is slightly increasing" and "increase of trend is slight" are even synonymous.

This suggest the idea that the above special evaluative predications (11) and (14) are semantically tantamount to the standard form

$$\langle \text{sign of trend} \rangle \text{ is } \mathcal{B} \tag{15}$$

where $\mathcal{B}$ is an evaluative expression generated by the function (10). Therefore, we can first generate the predications (15) linguistically characterizing the given tangent and then replace them by (11) or (14) according to the following tables Table 1 and Table 2:

Tantamount linguistic expressions	
⟨gradient⟩	$\mathcal{B}$
stagnating	Ze, ± extremely small
negligibly ⟨sign⟩	significantly small
slightly ⟨sign⟩	very small
somewhat ⟨sign⟩	rather small
clearly ⟨sign⟩	medium, quite roughly small, very roughly small
roughly ⟨sign⟩	quite roughly big, very roughly big
sharply ⟨sign⟩	very big
significantly ⟨sign⟩	significantly big

Table 1: Case (a)

Demonstration of generated comments to the time series is depicted in Figure 1. One can see that its trend is by no means clear. Slot 3 (time 92-115) of this time series is validation part, on which the quality of the forecast is tested and the best one is chosen. Slot 4 (time 116-127) is testing part that is not used for computation of the forecast but only for comparison of the forecast with the real data. In Fig. 2 the same time series is analyzed and its forecast is computed.

The linguistic context for the trend evaluation was set to $w_{tg} = \langle v_L = 0, v_S = 1200/12, v_R = 3000/12 \rangle$ since the time series demonstrates clear periodicity of $T = 12$ (this was obtained using periodogram — cf., e.g., [3]). The generated

Tantamount linguistic expressions	
⟨special hedge⟩	$\mathcal{B}$
negligible	significantly small
slight	very small
small	small
clear	medium, quite roughly small, very roughly small
rough	quite roughly big, very roughly big
fairly large	roughly big, more or less big
quite large	rather big
large	big
sharp	very big
significant	significantly big
huge	extremely big

Table 2: Case(b)

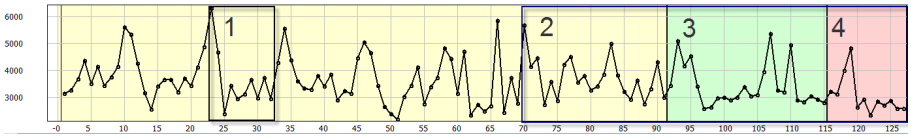

Figure 1: Demonstration of evaluation of trend of various parts of a real time series. Trend of the whole series is *stagnating*. Slot 1 (time 23-32): *clear decrease*, Slot 2 (time 70-127): *negligible decrease*, Slot 3 (time 92-115): *small increase*, Slot 4 (time 116-127): *fairly large decrease*.

evaluation of trend of the predicted values in the testing part is *rough decrease* while evaluation of trend of the real data is *fairly large decrease*. Thus, instead of presenting concrete predicted numbers, the manager might be satisfied with the information that "*rough decrease is expected*".

3 Discussion

In this paper, we presented a method based on combination of the F^1-transform and fuzzy natural logic using which we can generate linguistic comments to the trend of the time series in arbitrary time slots. We believe that such comments can be useful, for example, in situations when it is difficult to see the global trend because the time series is too volatile.

In further research, we can apply other theories of fuzzy natural logic, for example formal theory of *intermediate quantifiers* (see [6, 7]). Using this theory, we can model the meaning of sentences, such as

> "Most (many, few) analyzed time series stagnated recently but their future trend is slightly increasing."

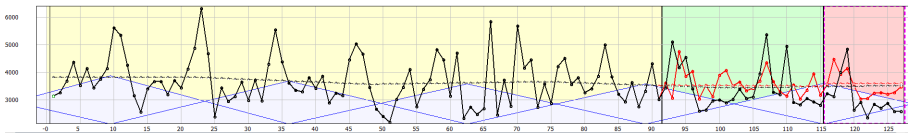

Figure 2: Analysis and forecast of the time series from Fig. 1. Slot 3 (time 92-115 — validation part) and Slot 4 (time 116-127 — testing part) contain computed and predicted trend cycle, and also real and predicted values. The generated comments in Slot 4 are: *rough decrease* for the predicted data and *fairly large decrease* for the real data.

> "Huge decrease of trend of almost all time series in the recent quarter of the year."

Moreover, we can also apply syllogistic reasoning with such expressions, for example

$$\frac{\begin{array}{l}\textit{Few analyzed time series are not from IT industry}\\ \textit{Many analyzed time series are clearly raising}\end{array}}{\textit{Few clearly raising time series are not from IT industry}}$$

It is important to realize that the latter is example of valid generalized Aristotle's syllogism which means that it is true in all situations (models).

Another possibility is to mine interesting information from the given set of time series, summarize their properties and summarize also their possible future development. Namely, we start with analysis and forecasting of all the time series. Then we generate comments to interesting time slots, or we can also determine time slots in which behavior of the time series is interesting for us, for example, "in which period was the time series sharply increasing", "how long was the time series stagnating or decreasing before sharp increase", etc. Finally, we can summarize the results using intermediate quantifiers and derive further properties on the basis of valid syllogisms.

Acknowledgment

This paper was supported by the program MŠMT-KONTAKT II, project LH 12229 "Research and development of methods and means of intelligent analysis of time series for the strategic planing problems". Additional support was given also by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, and L. Vavříčková. Analysis of seasonal time series using fuzzy approach. *Int. Journal of General Systems*, 39:305–328, 2010.

[2] I. Perfilieva, M. Daňková, and B. Bede. Towards a higher degree F-transform. *Fuzzy Sets and Systems*, 180:3–19, 2011.

[3] J. Anděl. *Statistical Analysis of Time Series.* SNTL, Praha, 1976 (in Czech).

[4] J.D. Hamilton. *Time Series Analysis.* Princeton, Princeton University Press, 1994.

[5] V. Novák. A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems*, 159(22):2939–2969, 2008.

[6] V. Novák. A formal theory of intermediate quantifiers. *Fuzzy Sets and Systems*, 159(10):1229–1246, 2008.

[7] P. Murinová and V. Novák. A formal theory of generalized intermediate syllogisms. *Fuzzy Sets and Systems*, 186:47–80, 2012.

A Karhunen-Loeve Transform Approach to Robust Color Image Segmentation

Junzo Watada and Chenxue Wang
The Graduate School of Information, Production and Systems
Waseda University
e-mail: junzow@osb.att.ne.jp

Abstract

In this paper, a novel fast approach is proposed to achieve image segmentation in color image. This method helps to refine the foreground regions and achieves the goal of robust color image segmentation throw the following four steps. First, modified Karhunen-Loeve transform is performed to reduce the redundant component, thus selecting the most important part of the color images. Second, a multi-threshold Otsu method is carried out to select the best thresholds from image histogram. Thereby, the conventional Otsu method has been extended from gray level to color level. Third, improved Sobel edge detection is added to enhance the weight of edge detail of the foreground image. Finally, a K-Means Clustering is used to merge the over-segmented regions. Experimental results prove that this method has a good performance even when the color image has a complicated structure in the background.

Keywords: Image segmentation, Karhunen-Loeve transform, background subtraction, Otsu method, K-means clustering

1 Introduction

Image segmentation is an important part in pattern recognition, computer vision and computer graphics. The main target of image segmentation is to separate an image into two or more parts, each part belongs to the same object or the pixels of this part has similar characteristic. Up till now, many researchers have been focusing on gray-level image segmentation. On the other hand, since color pictures are much more complicated, studies on color image segmentation are much less. Color image contains much more information and is a closer description for the real world, so a robust segmentation with less computational load is required for color image processing to meet the computer vision's need.

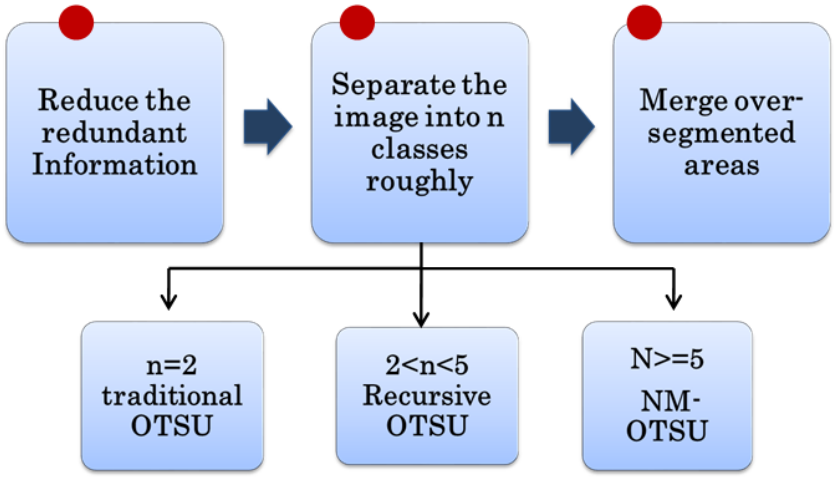

Figure 1: Processing flow of our image segmentation system

The purpose of this research is to develop a robust image segmentation system in order to separate an image into proper number of parts and consume less time. The system consists of three main modules.

First we use K-L theorem as a preprocessing in order to reduce the three dimensional data (RGB) into only one dimension and keep the important information of original color image as much as possible.

Second, we have built a relative robust Otsu-based segmentation system as a rough segmentation, for example, to separate an image into about 50 parts. There are three conditions in multi-threshold Otsu system. When we need to separate an image into simply two parts, a traditional Otsu method is implemented. When three or more classes is necessary for a proper segmentation, we should consider the time consuming problem. Here, for the segmentation more than 2 classes but less than 5 classes, we used recursive Otsu method to reduce the iteration time. If more classes(for example , 50) is needed to get a corrective segmentation, recursive Otsu can still not satisfy our needs for the purpose of less time consuming. In order to make a balance between the performance and time consuming, a heuristic algorithm called Nelder-Mead is combined with Otsu to reduce time consuming further more.

Third, we use Just-noticeable difference (JND) histogram to merge the over segmented areas. As we know that Otsu method just consider the pixel values but not the spatial conditions of these pixels, by applying JND method, a pixel can be merged to the background nearby it if the difference between its pixel values and the background can't be differentiate by human eyes. The whole process of our proposed method is shown in Fig. 1.

The remaining portion of this paper is organized as follows. Related methods are surveyed the proposed in section 2. Section 3 describes segmentation

method. In Section 4, experimental results are presented together with discussion. Finally, Section 5 concludes this research.

2 Related Method Survey

There are primarily four types of segmentation techniques: thresholding, boundary based, region based and hybrid techniques.

Among the thresholding techniques, more than ten approaches of determining the threshold have already been proposed, most representative ones are P-tile method, bimodal method, Otsu method, minimum error method, maximum entropy method and iterative method [22]. These studies are focused on the selection of the best single threshold from gray-level images. Among these methods, the Otsu method [18] is considered as the best algorithm with high robustness. In the luminance of one-dimensional gray level histogram of an image, Otsu method takes the variance between classes as the criterion to choose the optimal segmentation threshold. Maximum variance between clusters can automatically obtain the optimal threshold in the statistical sense and has a better effect on differing the classes such as foreground and background.

For the boundary based methods, these methods search for pixels that lie on a region boundary (or at the boundary between two regions). These pixels are called edges [14]. An edge is characterized by a significant local change in image intensities. Edges are detected by looking at neighboring pixels. The basic assumption is that the change in pixels values between neighboring pixels inside a region is not as significant as the change in pixels values on the regions boundary. When the difference between two regions grows, the change becomes bigger and the edge becomes stronger. Sometimes weak edges should be detected as strong edges and in other times they should not. Consequently, not all the detected edges create closed curves, which are necessary to separate between regions. Therefore, some types of post/pre-processing techniques, such as [19] [26] [16] [20] are required for grouping the detected edges into connected surfaces to represent the region. In [15], an Edge Flow method was presented that is based on the edge directions rather than the edge energy. They detected the regions boundaries by identifying a flow direction at each pixel location that point to the closest boundary. Then, it follows by detection of the locations that encounter two opposite directions of edge flow. However, the main drawback of many boundary-based methods is the over-segmentation result, which does not always correctly reflect the image nature.

The region based methods gather similar pixels according to some homogeneity criteria [3]. They are based on the assumption that pixels, which belong to the same homogeneous region, are more alike than pixels from different homogeneous regions. The split-and-merge or the region growing techniques are examples for such method [2]. The region-growing algorithm initially defines each pixel as a region. Then, it scans the image from left to right and from top to bottom and compares the current pixel with its neighboring regions that were already scanned. If the pixel is sufficiently similar to one of its adjacent regions

it is added to that region. If it is not close enough to any of them, then it is still defined as a different region. On the contrary, the split-and-merge techniques [9] initially assume that the image is composed of one region. It splits inhomogeneous segment into four rectangular segments and merge four adjacent regions if they are found to be similar. When no region can split and no four adjacent regions can merge the algorithm is terminated. Two main drawbacks characterize these techniques. They are both strongly dependent on global pre-defined homogeneous criteria thresholds while the region-growing technique depends also on initial segments, which is the first pixel/segment, that is first to be scanned and on the order of the process.

The hybrid technique improves the segmentation result by combining the above methods for segmentation. Many of the hybrid techniques combine the region-based method with the boundary-based method. Some used the combination of the histogram-based with the region-based methods. The hybrid technique for segmentation is very common since it relies on wide information as global (histogram) and local (regions and boundaries). An example of a hybrid technique was presented in [21], which integrated between regions-based and boundary-based methods. First a split-and-merge algorithm is performed in order to initially segment the image. Then, the contours of the obtained regions are refined using the edge information. Later, the watershed algorithm [25] was presented. It begins with a boundary-based method to get gradient magnitude. Then, regions are produced by a region-growing technique. In [8], Haris et al. presented a segmentation algorithm using the watershed algorithm and regions merging. They applied the watershed transform to initial partitioning of the image into primitive regions. The output of the watershed used as an input for hierarchical (bottom-up) region merging process, which produced the final segmentation.

In this paper, an effective multilevel threshold selection method of color image segmentation is proposed based on Otsu method.

In recent years, the traditional Otsu method has been improved and promoted by many researchers. By using the joint histogram of original image and its neighbor smoothing image, Liu [13] have extended the Otsu method to 2-D and switched the threshold into a vector, this concept greatly improved the segmentation result. How-ever, searching optimal threshold in two dimensional spaces increased the computation cost significantly, which limited the application of the algorithm. In paper [10], a Two-Stage Multi-threshold Otsu method has been put forwarded on multi-level thresholding on Otsu method, but still focus on gray-level image. A lot of optimizations have already been proposed in order to improve the performance of 2-D Otsu, such as by using integral image, D.Han [7] concerns the problem of tree segmentation in color image based on 2-D Otsu in HSI color space while the background of the image should be quite simple and the H component of the tree in HIS color space need to be calculated in advance, so the method meets a huge restriction.

Compared with the gray image, the most salient feature of color images is the increase of the carrying data and information. Generally, color spaces always contain three components, for instance, the most commonly used RGB

channels. Each channel of the color space has a high correlation to the others. In this way, to reduce the redundant information processing is quite important.

3 Proposed Algorithm

In this paper, a novel method of Karhunen-Loeve transform(K-L transform) based Multi- level Otsu together with Sobel edge detection and K-Means clustering method has been proposed to extend the traditional gray level Otsu method to achieve the color image segmentation.

3.1 Step One

The transformation of the color space is a vital part in image processing. Considering a color image in RGB color space as an input, all colors can be seen as different combinations of three basic colors: red (R), green (G) and blue (B). Since these three components have great correlation with each other, a robust and effective transform should be chosen to reduce the redundant relationships in the color space. Here we use the K-L transform, which is one representation of a stochastic process. The process is to map the multi-dimensional data with correlation into a new coordinate in the region of the data distribution in order to compress the data information while the orientation of the new coordinate should keep the maximum amount of the information. This method is helpful to achieve dimensionality reduction of high dimensional data.

Given a color image of size(x, y), we can consider it as a three dimensional matrix $P(x, y, 3)$ in which each dimension contains the information of each component. In order to keep the color information, we just change the three dimensional matrix P into two dimensional matrix $P'(x * y, 3)$ which keeps each component that in one column. We get three column vectors represent the three components:

$$\mathbf{R} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_{x*y} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{x*y} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_{x*y} \end{bmatrix}, \tag{1}$$

Then $\sum_{ij}$, the covariance matrix is calculated by the following formula:

$$\begin{aligned} \sum_{ij} &= cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= E(X_i \dot{X}_j) - \mu_i \mu_j. \end{aligned} \tag{2}$$

where $\mu_i = E(X_i)$ (E is the statistical expectation) and $i, j = R, G, B$

Since matrix $\mathbf{P'}$ only has three columns, the covariance matrix is calculated as equation (3):

$$\mathbf{\Sigma} = \begin{bmatrix} cov(R,R) & cov(R,G) & cov(R,B) \\ cov(R,R) & cov(R,G) & cov(R,B) \\ cov(R,R) & cov(R,G) & cov(R,B) \end{bmatrix} \tag{3}$$

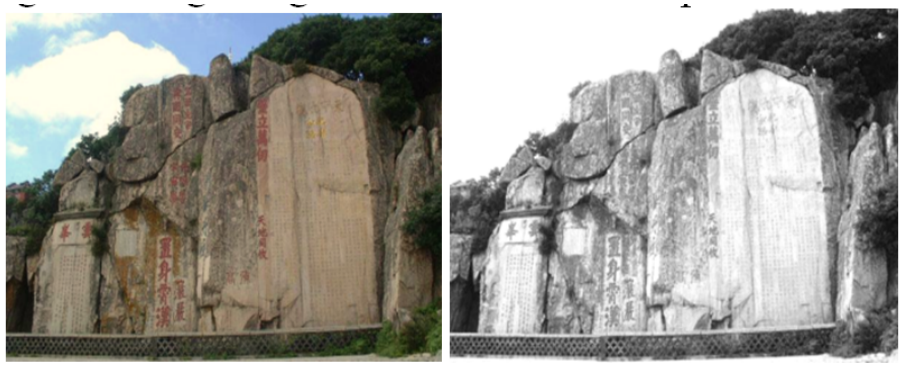

Figure 2: The right image is the component with most of the information of the left one after K-L transform.

Let ϕ_k be the eigenvector corresponding to the k-th eigenvalue λ_k of covariance matrix $\mathbf{\Sigma}$ that $\mathbf{\Sigma}\phi_k = \lambda_k\phi_k (k = R, G, B)$. As the covariance matrix is symmetric, its eigenvectors ϕ_k are mutually orthogonal and we can construct a 3×3 orthogonal matrix ϕ by formula (4):

$$\phi \triangleq [\phi_R, \phi_G, \phi_B] \quad \text{satisfying} \quad \phi^T\phi = \mathbf{I} \tag{4}$$

N eigen equations above can be combined to be expressed as $\mathbf{\Sigma}\phi = \phi\Lambda$, here $\Lambda = diag(\lambda_R, \lambda_G, \lambda_B)$ is a diagonal matrix. In this way, we choose the maximum value among λ_R, λ_G and λ_B and make P' multiply the corresponding ϕ. Finally, the new image is only composed of the component that keeps most of the information from the original image. Fig. 2 shows some examples:

3.2 Step Two

According to [18], in the case of single threshold segmentation (two partitions), the constraint equation is $\sigma_B^2 + \sigma_W^2 = \sigma^2$. In the equation, $\sigma_W^2$is within class variance and σ_B^2 is the between class variance which equals to $\omega_0\omega_1(\mu_0 - \mu_1)^2$ and σ^2 is the total variance of the image.

In accordance with the constraints above, because the total variance is constant for a given image, it is equivalent to select the optimal threshold by maximizing between class difference or minimizing within class difference. This principle reflects that the optimal threshold value by Otsu method could make the best segmentation between the target and background.

But with increasing complexity of images, a single threshold is certainly insufficient. So based on the classical Otsu method, we could extent the function to n-class segmentation:

$$\sigma_B^2 = \sigma^2 - \sigma_W^2 = \sum_{j=0}^{n-1}\sum_{k=j}^{n-1}\omega_j\omega_k(\mu_j - \mu_k)^2 \tag{5}$$

If we just use the function above to realize the multi-threshold segmentation, the computation load increases exponentially as the number of the threshold grows.

So in this paper, after get the image handled by the $K-L$ transform, a proposed multi-threshold segmentation by Otsu method is then carried out on the low dimensional image which contains most of the information.

For the segmentation of classes no more than four, we use the H-table method which was proposed by Liao and chung [12]. First, the between-class variance σ_B^2 of class n is modified as

$$\sigma_B^2 = \sum_{k=1}^{n} \omega_k \mu_k^2,$$

where

$$\omega_k = \sum_{i \in C_k} p_i, \quad \mu_k = \frac{\sum_{i \in C_k} i p_i}{\omega_k} = \frac{\mu(k)}{\omega_k}.$$

There is no μ_j^2 because it is independent of the choice of the optimal thresholds. Second, the $u-v$ interval zeroth- and first-order moment, $P(u,v)$ and $S(u,v)$ of a class with gray levels from u to v are defined in (6):

$$P(u,v) = \sum_{i=u}^{v} p_i, S(u,v) = \sum_{i=u}^{v} i p_i \tag{6}$$

Hence, all the values of $P(u,v)$ and $S(u,v)$ can be calculated recursively, so that the values of ω_k and $\mu(k)$ can be obtained by indexing the look-up table in equation (7):

$$\begin{aligned} \omega_k &= P(1,t_k) - P(1,t_{k-1}) = P(t_{k-1}+1,t_k) \\ \mu(k) &= S(1,t_k) - S(1,t_{k-1}) = S(t_{k-1}+1,t_k) \end{aligned} \tag{7}$$

Therefore, the σ_B^2 of class i can be calculated and stored in the so-called H-table:

$$H(t_(i-1)+1,t_i) = \frac{S(t_(i-1)+1,t_i)^2}{P(t_(i-1)+1,t_i)} \tag{8}$$

For the classes more than four, the H-table method still costs too much time. We use several methods to simplify the segmentation.

a) For n-class separation, we segment the histogram into $n-1$ parts.

b) From formula (5), we maxmize σ_B^2 to get the optimal segmentation. Equivalently, we could minmize the normalized class variance expressed as follows:

$$T = 1 - \frac{\sigma_B^2}{\sigma^2}.$$

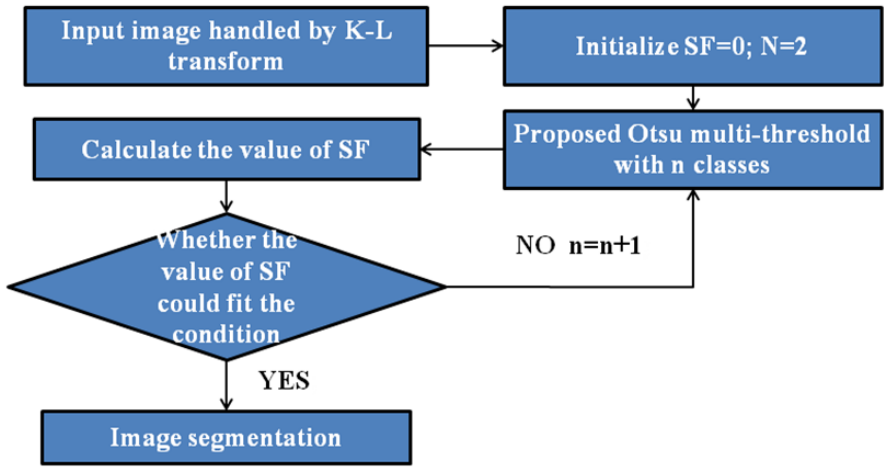

Figure 3: Fig. 2.Workflow of proposed Multi-thresholding

c) For each part, Nelder-Mead simplex algorithm [11] is utilized to calculate the local minimum values of function T.

d) Segment the image into n classes according to the $n-1$ local minimum values.

But another important factor is the method how we defined the optimal classes n to get the most suitable segmentation. Here, a separation factor is used as a condition to determine whether the segmentation is optimal or not. The SF (separation factor) is defined as $\frac{\sigma_B^2}{\sigma^2}$ and range from 0 to 1. When the value of SF tends to 1, it means that the image has been separated absolutely. The work flow of this part is showed in Fig. 3:

3.3 Step Three

Considering that the foreground objects always have relative strong edge, after obtaining the segmented image through multi-threshold Otsu, we apply Sobel operation on the original image and add the edge information to the segmented one so that we can keep more edge details.

The traditional Sobel operator uses two 3×3 kernels (one for horizontal, and another for vertical) which are convolved with the original image to calculate approximations of the brightness difference. But in many cases, only two directions are far from satisfactory. So in this paper, we use the eight-orientation

Sobel operator to keep more information. The models are as follows:

$$\underset{0\text{ degree}}{\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}} \quad \underset{45\text{ degree}}{\begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}} \quad \underset{90\text{ degree}}{\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}} \quad \underset{135\text{ degree}}{\begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}}$$

$$\underset{180\text{ degree}}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}} \quad \underset{225\text{ degree}}{\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix}} \quad \underset{270\text{ degree}}{\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}} \quad \underset{315\text{degree}}{\begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix}} \tag{9}$$

According to the n-class separation of the previous step, we also divide the edge map into n classes, and then the edge information is added to the segmented colormap to enhance the weight of the foreground objects. The enhance of the edge information makes the segmentation more clearly.

3.4 Step Four

In the final step, a K-Means method is used to merge the segmented image. The K-means clustering is a partitioning method for grouping objects so that the within-group variance is minimized. By minimizing the local dissimilarity of each subset, the algorithm will assign the optimal distribution of all subsets. From our previous result, we observed that there are many regions with similar intensities in a colormap of the segmented image, which result in many local minima that increase over segmentation. In this way, K-means to merge the over segmented regions and separate the image into background and foreground.

The method is applied by the following steps:

a) Initialize two class centers, for example, value 0 and 1. These centers represent initial group centroids.

b) Calculate the histogram bin value distance between each image pixel and class centers, assign each pixel to its closest class centroid.

c) Recalculate the positions of the centroids by calculate the mean histogram bin value of the same group.

d) Repeat Steps b) and c) until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

4 Experimental Results and Discussion

In this section, two experiments have been done to evaluate the performance of the proposed algorithm. In the first one, we compare the results of our proposed method with some others. Fig. 4 illustrates the results.

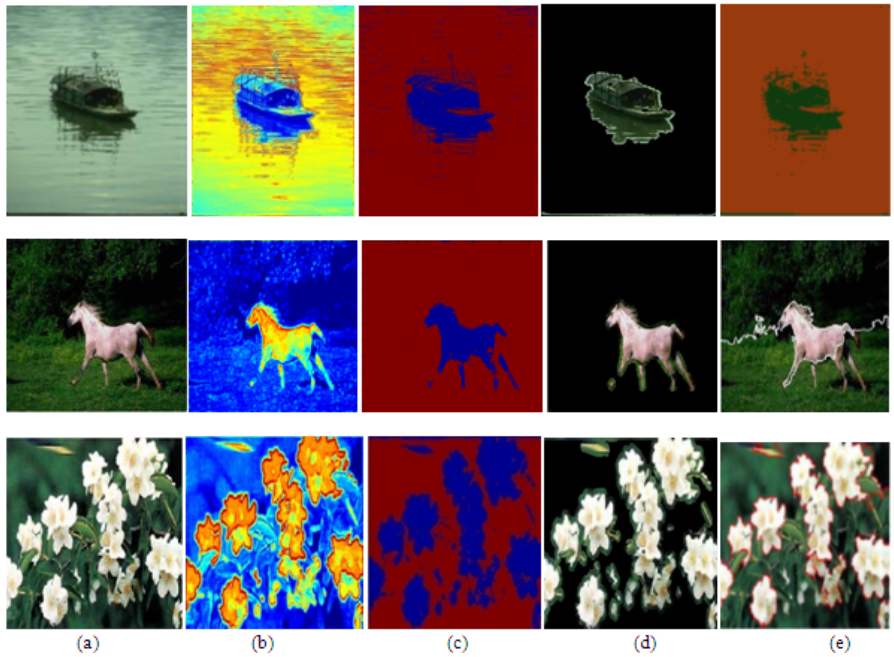

Figure 4: (a) The source images. (b) Results after step three by the proposed algorithm. (c) Results after step four by the proposed algorithm. (d) Results after background subtraction and morphology process. (e) Results from other methods.

The first row shows a comparison of the proposed algorithm and MCVT in [17] applied to an image collected from Berkeley segmentation dataset [27]. The image after Otsu multi-thresholding (b) and result after region merging (c) shows that our results keep more information than (e).

The results of the second row compare the performance with the proposed algorithm and Mean Shift and Normalized Cuts (MSNC) in [24]. From the result after background subtraction (d), we could see that our approach segments the image better than the result (e) in [24].

In the third row, the results of our method keep more details than that of automatic seeded region growing (ASRG) method in [23], the result (d) contains the entire flower pixels while (e) lose some parts.

In the second experiment, we calculate the time consumption of our approach. The experiment environment is as follows: Intel Core2 CPU, E8500 - 3.16GHz, 3.25 GB RAM. The develop tool is Matlab R2007b. The calculation time for each stage can be seen in Table.1. Compare to the time consumption in [24] [23], our method costs less time work load.

According to the experimental results, we can come to the conclusion that our method extends the traditional Otsu method from gray level to color im-

Table 1: Calculation Time For Each Step

Size	320×240			1152×872		
Class	3	16	64	3	16	64
K-L	0.010			0.053		
Multi-th	0.084	0.088	0.140	0.596	0.732	1.084
Sobel	0.085	0.138	0.221	0.323	0.453	0.803
K-means	0.220			1.126		
Total	0.399	0.456	0.591	2.098	2.364	3.066

ages. Moreover, it can achieve better performance and lower computational complexity than similar works. It is robust to the user inputs.

5 Conclusion

In this paper, we develop a robust image segmentation system in order to separate an image into proper number of parts and consume less time.

The system consists of three main modules. Firstly we use K-L theorem as a preprocessing in order to reduce the three dimensional data (RGB) into only one dimension and keep the important information of original color image as much as possible. Second, we have built a relative robust Otsu-based segmentation system as a rough segmentation, for example, to separate an image into about 50 parts. There are three conditions in multi-threshold Otsu system. When we need to separate an image into simply two parts, a traditional Otsu method is implemented. When three or more classes are necessary for a proper segmentation, we should consider the time consuming problem. Here, for the segmentation more than 2 classes but less than 5 classes, we used recursive Otsu method to reduce the iteration time. If more classes(for example , 50) is needed to get a corrective segmentation, recursive Otsu can still not satisfy our needs for the purpose of less time consuming. In order to make a balance between the performance and time consuming, a heuristic algorithm called Nelder-Mead is combined with Otsu to reduce time consuming further more. Thirdly, we use Just-noticeable difference (JND) histogram to merge the over segmented areas.

As we know that Otsu method just consider the pixel values but not the spatial conditions of these pixels, by applying JND method, a pixel can be merged to the background nearby it if the difference between its pixel values and the background can't be differentiate by human eyes.

References

[1] Edward Y. Chang, Beitao Li, Chen Li, "Toward Perception-Based Image Retrival" IEEE Trans., 2000

[2] Y.-L. Chang, X. Li, Adaptive image region-growing, IEEE Transactions on Image Processing 3 (1994) 868-872.

[3] S. Chen, W. Lin, C. Chen, Split-and-merge image segmentation based on localized feature analysis and statistical tests, CVGIP: Graph, Models Image Processing 53 (1991) 457-475.

[4] C.H. Chou, Y.C. Li, A perceptually tuned subband image coder based on the measure of Just-Noticeable-Distortion profile, IEEE Trans. Circuits Systems Vid-eo Technol. 5 (6) (December 1995) 467-476.

[5] G. Dai, "Modal wave-front reconstruction with Zernike polynomials and Karhunen-Loeve functions". JOSA A vol. 13, no. 6, pp. 1218, 1996.

[6] Chris Ding and Xiaofeng He. "K-means clustering via Principal Component Analysis". Proc. of Int'l Conf. Machine Learning (ICML 2004), pp 225-232. July 2004.

[7] D.Han and X.Huang, "A Tree Image Segmentation Method Based on 2-D OTSU in HSI Color Space", in *WiCOM*, pp.1-4, 2010. doi:10.1109/WICOM.2010.5600669

[8] K. Haris, S.N. Efstratiadis, N. Maglaveras, A.K. Katsaggelos, Hybrid image seg-mentation using watershed and fast region merging, IEEE Transactions on Image Processing 7 (1998) 1684-1699.

[9] S. Horowitz, T. Pavlidis, Picture segmentation by a directed split and merge pro-cedure, in: Proceedings of Second International Joint Conference on Pattern Recognition, 1974, pp. 424-433.

[10] D.Y.Huang and C.H.Wang, "Optimal multi-level thresholding using a two-stage Otsu optimization approach", *Pattern Recognition Letters*, vol.30, no. 3,pp. 275-284, Feb. 2009.

[11] Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright. "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions." *SIAM Journal of Optimization*, vol. 9, no 1, pp. 112-147, 1998.

[12] P.Liao and P.Chung, "A fast algorithm for multilevel thresholding", *Journal of Information Science and Engineering*, vol.17, no.5, pp. 713-727, Sept. 2001.

[13] J. Z. Liu, and W. Q. Li, "The automatic thresholding of gray-level picture via 2D Otsu method", *Acta Automatica Sinica*, vol. 19, no.1, pp. 101-105, Jan. 1993.

[14] D. Marr, E. Hildreth, Theory of edge detection, in: Proceedings of Royal Society of London B, no. 207, 1980, pp. 187-217.

[15] W.-Y. Ma, B.S. Manjunath, EdgeFlow: a technique for boundary detection and image segmentation, IEEE Transaction on Image Processing 9 (8) (2000).

[16] R. Murrieta-Cid, M. Briot, N. Vandapel, Landmark identification and tracking in natural environment, IEEE International Conference on Intelligent Robots and Systems, Victoria, BC, Canada, 1998, pp. 738- 740

[17] Nita M. Nimbarte and Milind M. Mushrif, "Multi-level Thresholding Algorithm for Color Image Segmentation", in *10 Proceedings, International Conference on Computer Engineering and Applications*, vol.2, pp.231-233, 2010. doi:10.1109/ICCEA.2010.198.

[18] N.Otsu, "A threshold selection from gray level histograms", *IEEE Trans. Systems, Man and Cybernetics*, vol.9, pp. 62-66. Mar. 1979

[19] E. Ozyildiz, N. Krahnst-over, R. Sharm, Adaptive texture and color segmentation for tracking moving objects, Pattern Recognition 35 (2002) 2013-2029.

[20] N. Pal, S. Pal, A review on image segmentation techniques, Pattern Recognition 26 (1993) 1277-1294.

[21] T. Pavlidis, Y. Liow, Integrating region growing and edge detection, IEEE Trans-actions on Pattern Analysis and Machine Intelligence 12 (1990) 225-233.

[22] M.Sezgin and B.Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", *Journal of Electronic Imaging*, vol.13, no.1, pp. 146-165, Jan. 2004.

[23] F.Y.Shih and S.Chen, "Automatic seeded region growing for color image segmentation", *Image and Vision Computing*, vol.23,no 10, pp.877-886, 2005.

[24] W.Tao, H.Jin, and Y.Zhang, "Color Image Segmentation Based on Mean Shift and Normalized Cuts", *IEEE Trans. Systems, Man and Cybernetics*, vol.37, no.5, Oct. 2007.

[25] L. Vincent, P. Soille, Watershed in digital space: an efficient algorithm based on immersion simulation, IEEE Transactions on Pattern Analysis and Machine Intel-ligence 13 (1991) 583-598.

[26] D.S. Yang, H.I. Choi, Moving object tracking by optimizing models, Proceedings of the International Conference of Pattern Recognition, Brisbane, Australia, 1998, pp. 738-740.

[27] http://www.cs.berkeley.edu/projects/version/grouping/segbench/

Operator of composition for continuous copulas

Vladislav Bína
Faculty of management in Jindřichův Hradec
University of Economics in Prague
bina@fm.vse.cz

Abstract

The contribution concerns an approach to the construction of multidimensional copula models from the low-dimensional copulas. Author define an operator of composition for continuous copulas in an analogous manner to the already developed approach for discrete probability distributions.

The operator of composition is first defined for continuous probability densities. Then the formula for special case of composition in general copulas is inferred. And finally, for the Frank copulas subclass, the definition is simplified using the properties of the corresponding generating functions.

1 Introduction

The authors present a definition of operator of composition and analysis of its properties within the theory of copulas in an analogous manner to the definitions in field of discrete probabilistic distributions (see, e.g., Jiroušek [2]), theory of possibility (see Vejnarová [7]) and generalization to the Dempster-Shafer theory of evidence (see Jiroušek and Vejnarová [3]). The employment of operator of composition provides a possibility to construct multidimensional distribution with given dependency structure using the apparatus of copula theory.

2 Preliminaries and basic notions

Throughout the paper we shall consider vectors of continuous random variables $(X_1, \ldots, X_d)$ with continuous one-dimensional marginal CDFs. The random vector obtained by the probability integral transform

$$(U_1, U_2, \ldots, U_d) = (F_1(x_1), F_2(x_2), \ldots, F_d(x_d))$$

has uniform components.

Now the *copula* of vector $(X_1, \ldots, X_d)$ is defined as a joint CDF of vector $(U_1, U_2, \ldots, U_d)$ in the following way

$$C(u_1, u_2, \ldots, u_d) = P\left[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d\right].$$

The *d-dimensional copula* $C : [0,1]^d \to [0,1]$ contains information about dependencies among particular variables $X_1, \ldots, X_d$ and the marginal CDFs F_i hold information concerning one-dimensional marginal distributions.

A theoretical background of copula theory is provided by the famous Sklar's theorem (see [6]) which states that a multivariate CDF of a random vector $(X_1, \ldots, X_d)$ can be written in the form

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right),$$

where C denotes a copula. In case of continuous F_i the copula is unique.

As we already mentioned, a copula is a CDF. But for statistical applications it is reasonable to express the multidimensional distribution in terms of densities. If the copula function is sufficiently differentiable we can express the copula density in the following manner

$$c(u_1, u_2, \ldots, u_d) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} C(u_1, u_2, \ldots, u_d) \tag{1}$$

and the multivariate density can be expressed in the form

$$\begin{aligned} f(x_1, x_2, \ldots, x_d) &= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F(x_1, x_2, \ldots, x_d) = \\ &= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) = \\ &= c(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \cdot \prod_{j=1}^{d} f_j(x_j), \end{aligned}$$

where $f_i(x_i) = \frac{dF_i(x_i)}{dx_i}$ is a one-dimensional marginal density.

The shortened notation $f(x_K)$ denotes a multidimensional density function of variables having indices from K. Having two sets of variable indices K and L, the symbol $f(x^{\downarrow K \cap L})$ denotes corresponding marginal density summed up from the multidimensional density $f(x_K)$.

In order to shorten the expressions, the symbol G_L with $L = \{\ell_1, \ldots, \ell_j\}$ stands for a vector $(G_{\ell_1}(x_{\ell_1}), \ldots, G_{\ell_j}(x_{\ell_j}))$.

3 Archimedean copulas

One of the popular classes are the Archimedean copulas. Different represents of this class allows to model different types of dependence and popularity of this class probably stems from the ability to model different strength of dependence by the choice of single parameter.

A copula C is *Archimedean* if it can be represented using generator function ψ in the following way

$$C(u_1, \ldots, u_d, \vartheta) = \psi\left(\psi^{[-1]}(u_1, \vartheta) + \cdots + \psi^{[-1]}(u_d, \vartheta), \theta\right).$$

The generator function $\psi : [0, \infty) \to [0, 1]$ is strictly decreasing convex continuous function. The function $\psi^{[-1]}$ is a pseudo-inverse of the generator function (for the properties see, e.g., Nelsen [5]).

Among the classes of the Archimedean copulas belong the well known Frank, Gumbel and Clayton copulas (for detail see, e.g., Nelsen [5], graphs in Bína and Jiroušek [1]).

By careful derivation from Formula (1) we can infer that for the copula density holds

$$c(u_1, \ldots, u_d) = \psi^{(d)}\left(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\right) \cdot \prod_{j=1}^{d} \left(\psi^{-1}\right)'(u_j). \qquad (2)$$

One of the important classes of Archimedean copulas is the family of Frank copulas. It is defined by a generator

$$\psi_\theta(t) = -\frac{1}{\theta} \log(1 + \exp(-t)(\exp(-\theta) - 1))$$

with an inverse

$$\psi_\theta^{-1}(t) = -\log\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right).$$

Notice that $\theta \in R \setminus \{0\}$.

4 Operator of composition

The operator of composition can be defined in an analogous manner to the definition in case of discrete probabilistic distributions (see Jiroušek [2]).

Definition 1. *Consider two sets of continuous variables X_K and X_L, a probability density $f(x_K)$ and a probability density $g(x_L)$ with supports fulfilling condition* Supp $f(x^{\downarrow K \cap L}) \subseteq$ Supp $g(x^{\downarrow K \cap L})$. *The* right composition *is given by*

$$f(x_K) \triangleright g(x_L) = \frac{f(x_K) g(x_L)}{g(x^{\downarrow K \cap L})}.$$

Analogously to the discrete case (see again Jiroušek [2]) we can iterate the operation of composition in order to build up a multidimensional compositional model.

Under certain conditions we can take advantage of the copula properties and reformulate the previous definition.

4.1 Composition with one-dimensional intersection

First of all, we will focus at the situation where the second operand contains only one variable appearing in the first operand.

Theorem 1. *Let X_K and X_L be two sets of variables, $L = \{\ell_1, \ldots, \ell_j\}$ and there exist exactly one $m \in \{1, \ldots, j\}$ such that $\ell_m = K \cap L$. For the supports we require* $\operatorname{Supp} f(x_{\ell_m}) \subseteq \operatorname{Supp} g(x_{\ell_m})$. *Then the right composition can be expressed using the copula density c_g corresponding to density g in the following manner*

$$f(x_K) \triangleright g(x_L) = f(x_K) c_g\left(G_L(x_L)\right) \prod_{i \in L \setminus \{\ell_m\}} g_i(x_i).$$

Proof.
The proof is straightforward. Considering that

$$g(x_L) = c_g\left(G_L(x_L)\right) \prod_{i \in L} g_i(x_i)$$

we can express and simplify the form from the assertion

$$\begin{aligned} f(x_K) \triangleright g(x_L) &= \frac{f(x_K) g(x_L)}{g(x^{\downarrow K \cap L})} \\ &= \frac{f(x_K) c_g\left(G_L(x_L)\right)) \prod_{i \in L} g_i(x_i)}{g_{\ell_m}(x_{\ell_m})} \\ &= f(x_K) c_g\left(G_L(x_L)\right) \prod_{i \in L \setminus \{\ell_m\}} g_i(x_i). \end{aligned}$$

□

Example 1. A simple application of copulas can be based on subsequent addition of two-variable copulas to the constructed model. Let us formulate the composition of two two-variable copulas with one common variable.

Consider two probability densities given by copula

$$f(x_1, x_2) = c_f(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2)$$

and

$$g(x_2, x_3) = c_g(G_2(x_2), G_3(x_3)) g_2(x_2) g_3(x_3)$$

such that $\operatorname{Supp} f(x_2) \subseteq \operatorname{Supp} g(x_2)$ their right composition is given by

$$\begin{aligned} f(x_1, x_2) \triangleright g(x_2, x_3) &= \frac{f(x_1, x_2) g(x_2, x_3)}{g(x_2)} = \\ &= \frac{c_f(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2) c_g(G_2(x_2), G_3(x_3)) g_2(x_2) g_3(x_3)}{g_2(x_2)} = \\ &= c_f(F_1(x_1), F_2(x_2)) c_g(G_2(x_2), G_3(x_3)) f_1(x_1) f_2(x_2) g_3(x_3). \end{aligned}$$

4.2 Composition in Frank copulas

Now we will focus at the composition where the right operand is a Frank copula. Its properties allows us to simplify the expression even in general case of intersection of K and L containing more variables.

But first of all we need to show that the marginalization of densities given by Frank copula retain its type and parameter.

Lemma 1. *Marginalization (over a single variable) preserves the type and parameter of Frank copula c in the following sense:*

$$c\left(G_{L\setminus\{m\}}\left(x_{L\setminus\{m\}}\right)\right) = \int_{-\infty}^{\infty} c(G_K(x_K))g_m(x_m)dx_m$$

where $m \in L = \{\ell_1, \ldots, \ell_j\}$

Proof.
Let us start from the marginalization of density of continuous variables. Above in the Section 2 we inferred that

$$g\left(x_{L\setminus\{m\}}\right) = c\left(G_{L\setminus\{m\}}\right) \prod_{i\in L\setminus\{m\}} g_i(x_i). \tag{3}$$

On the other hand, if we straightforwardly rewrite the formula of marginalization of densities into the terms of copula densities we obtain

$$\begin{aligned} g\left(x_{L\setminus\{m\}}\right) &= \int_{-\infty}^{\infty} g(x_L)dx_m = \\ &= \int_{-\infty}^{\infty} c(G_L)\prod_{i\in L} g_i(x_i)dx_m = \\ &= \int_{-\infty}^{\infty} c(G_{\ell_1}(x_{\ell_1}), \ldots, G_{\ell_j}(x_{\ell_j}))\prod_{i\in L} g_i(x_i)dx_m. \end{aligned}$$

Substituting $u_i = G_i(x_i)$ for all $i \in L$ we arrive at

$$g\left(x_{L\setminus\{m\}}\right) = \int_0^1 c(u_1, \ldots, u_j)du_m \prod_{i\in L\setminus\{m\}} g_i(x_i).$$

Using the Formula (2) for the density of Archimedean copula we get

$$g\left(x_{L\setminus\{m\}}\right) = \int_0^1 \frac{\psi^{(j)}\left(\sum_{i\in L}\psi^{-1}(u_i)\right)}{\prod_{i\in L}\psi'\left(\psi^{-1}(u_i)\right)}du_m \prod_{i\in L\setminus\{m\}} g_i(x_i).$$

This formula can be integrated over u_m with the result

$$g\left(x_{L\setminus\{m\}}\right) = \left[\frac{\psi^{(j-1)}\left(\sum_{i\in L}\psi^{-1}(u_i)\right)}{\prod_{i\in L\setminus\{m\}}\psi'\left(\psi^{-1}(u_i)\right)}\right]_{u_m=0}^{1} \prod_{i\in L\setminus\{m\}} g_i(x_i).$$

From the definition of the generating function in case of Frank copulas we can see that setting of $u_m = 0$ results to

$$\psi^{-1}(0) \to +\infty$$

and setting of $u_m = 1$ implies

$$\psi^{-1}(1) = 0.$$

We can see that $u_m = 1$ results into the cancelation of the corresponding term in the sum and therefore we obtain the desired Formula (3).

The only step remains. It is necessary to show that the term obtained by setting $u_m = 0$ is equal to zero. We can see in McNeil and Nešlehová [4] that the derivative of Frank copula generator is

$$\psi_\theta^{(j)}(t) = (-1)^j \frac{1}{\theta} \mathrm{Li}_{-(j-1)}((1 - e^{-\theta} e^{-t}))$$

for $t \in (0, +\infty)$. The symbol Li denotes a polylogarithm function defined by an infinite sum in the following way

$$\mathrm{Li}_s(z) = z + \frac{z^2}{2^s} + \frac{z^3}{3^s} + \dots.$$

Calculation of the limit of any derivative $\psi_\theta^{(j)}(t)$ for t going to infinity we arrive at the value of polylogarithm special function in point zero. This is obviously equal to zero and the term for $u_m = 0$ cancels which finishes the proof. □

Let us denote that the preceding lemma concerns the marginalization over a single variable but its use can be easily iterated. Moreover, it seems that it can be easily generalized on other classes of Archimedean copulas using the results of McNeil and Nešlehová [4] concerning the derivatives of Archimedean copula densities.

Theorem 2. *The density $g(x_L)$ in the second operand of right composition is given by a Frank copula c*

$$g(x_L) = c(G_L) \prod_{i \in L} g(x_i).$$

Then the right composition can be expressed as follows

$$f(x_K) \triangleright g(x_L) = f(x_K) \frac{c(G_L)}{c(G_{\downarrow K \cap L})} \prod_{i \in L \setminus K} g(x_i).$$

Proof.
Starting from the definition of operator of composition we have

$$f(x_K) \triangleright g(x_L) = \frac{f(x_K) g(x_L)}{g(x^{\downarrow K \cap L})}$$

and using the Lemma 1 we know that using the marginalization (iteratively) the copula type and parameter remain unchanged

$$f(x_K) \triangleright g(x_L) = f(x_K) \frac{c(G_L) \prod_{i \in L} g_i(x_i)}{c(G_{\downarrow K \cap L}) \prod_{i \in K \cap L} g(x_i)}.$$

Now it is apparent that we arrive at the assertion of the theorem. □

5 Conclusion

The paper presents a definition of operator of composition for continuous densities together with the elicitation of a simplified formula employing the copula densities in case of certain type of composed copulas and for the case of Frank copula subclass.

The work can continue with a generalization of the results on other classes of Archimedean copulas such as Clayton, Gumbel and others. Also the study of properties (such as Shannon entropy) of the composition in copulas is still opened.

Acknowledgement

The research was partially supported by Grant Agency of the Czech Republic under grant no. 403/12/2175 and University of Economics in Prague under internal grant project no. F6/12/2013.

References

[1] Bína V., Jiroušek R. (2013) A short note on multivariate dependence modeling, *Kybernetika*, **49** (3), p. 420–432.

[2] Jiroušek R. (2011) Foundations of compositional model theory, *Int. J. Gen. Syst.*, **40** (6), p. 623–678.

[3] Jiroušek R., Vejnarová J. (2011) Compositional models and conditional independence in evidence theory, *Internat. J. Approx. Reason.*, **52** (3), p. 316–334.

[4] McNeil A. J., Nešlehová J. (2009) Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions, *The Annals of Statistics*, **37** (5B), p. 3059–3097.

[5] Nelsen, R. B. (2006) *An Introduction to Copulas*, Second Edition. New York, NY 10013, USA: Springer Science+Business Media Inc.

[6] Sklar A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, p. 229–231.

[7] Vejnarová J. (1998) Composition of possibility measures on finite spaces: preliminary results, In (B. Bouchon-Meunier, R.R. Yager, eds.): Proceedings of 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems IPMU'98, Editions E.D.K. Paris, p. 25–30.

On Causal Compositional Models: preliminaries

Radim Jiroušek
Faculty of management, University of Economics
Jindřichův Hradec
and
Inst. of Information Theory and Automation AV ČR
radim@utia.cas.cz

Abstract

This is the first attempt to introduce causal models in a compositional form. This "algebraic" form of representation of multidimensional probability distributions seems to be quite useful for representation of causal models because of two reasons. First, decomposition of the model into its low-dimensional parts makes some of computations feasible, and, second, it appears that within these models, both conditioning and intervention can be realized as a composition of the model with a degenerated one-dimensional distribution. Surprisingly, the syntax of these two computational processes are almost the same; they differ from each other just by one pair of brackets.

1 Introduction

Each of us is interested in the relation of causation from childhood; it first enables us to answer the WHY questions, and a couple of years later, more sophisticated WHAT IF questions. Moreover, referring to any textbook we can see that practically all the knowledge is explained using causal relation. For example, just to be able to use causality, some properties of light are explained using wave theory of light, and to explain some other properties it is useful to consider light to be a flow of particles - photons.

The importance of causation is visible also from the fact that from most of the articles in professional journals describing data mining applications one can see that the described research was performed with the (often hidden) goal to support or to uncover some new causal relations. However, this is sometimes misleading, because, as Pearl says in his book ([4], page 40): *The sharp distinction between statistical and causal concepts can be translated into a useful principle: behind every causal claim there must lie some causal assumption that*

is not discernable from the joint distribution and, hence, not testable in observational studies. Such assumptions are usually provided by humans, resting on expert judgment. Therefore, when using causal models one should keep this fact in mind. We can construct causal models only when we have a knowledge allowing us to specify causal relations, allowing us to determine what is cause and what is effect.

In this paper we present some preliminary ideas regarding application of probabilistic causal models represented in a form of compositional models. To keep the presentation as simple and informal as possible, we will introduce most of the concepts just on examples. We will consider only finite valued random variables that will be denoted by upper case Latin characters: $X, Y, Z, W, ...$ Sets of these variables will be denoted by lower case characters $(x, y, ...)$, and their probability distributions will be denoted using characters of a Greek alphabet $\kappa, \lambda, \mu, \nu, \pi$. So, $\kappa(X_1, \ldots, X_n)$ denotes an n-dimensional probability distribution. Its $n-1$-dimensional marginal distribution will be denoted by κ^{-X_i}, or, denoting $x = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$ we use also the symbol $\kappa^{\downarrow x}$. The latter symbol $\kappa^{\downarrow y}$ can be used for any $y \subseteq \{X_1, \ldots, X_n\}$.

In the above mentioned Pearl's book [4] (from which the causal model studied in this paper is taken over), one of the most important notions is the concept of an *intervention*. It means that by an external force we change the value of an intervention variable. As a simple example consider two binary variables: A - alarm bell rings or not, and B - smoke is or is not in a room. A smoke in the room makes the alarm bell ring. Therefore, if $\pi(A, B)$ describes the relationship between these two variables then the conditional probability $\pi(A = \textit{bell rings}|B = \textit{smoke is}) = 1$, and $\pi(B = \textit{smoke is}|A = \textit{bell rings}) \gg \pi(B = \textit{smoke is})$. But when considering the intervention that will be here denoted by $do(A = \textit{bell rings})$, which means that by some way or another we make the alarm bell ring (no matter whether there is a smoke in the room or not) it does not create a smoke in the room. Therefore

$$\pi(B = \textit{smoke is}|do(A = \textit{bell rings})) = \pi(B = \textit{smoke is}).$$

2 Causal networks

Let us start considering a classical example from [4] represented by a Bayesian network, the graph of which is in Figure 1(a). From this graph we can see that this Bayesian network is specified by five (un)conditional probability distributions. Let they be $\pi_1(S)$, $\pi_2(G|S)$, $\pi_3(R|S)$, $\pi_4(W|G, R)$, $\pi_5(P|W)$.

The meaning of the variables is the following:

S - season of a year;
G - garden sprinkler is on, off;
R - it is raining/snowing or not;
W - grass is wet or not;
P - pavement is slippery or not.

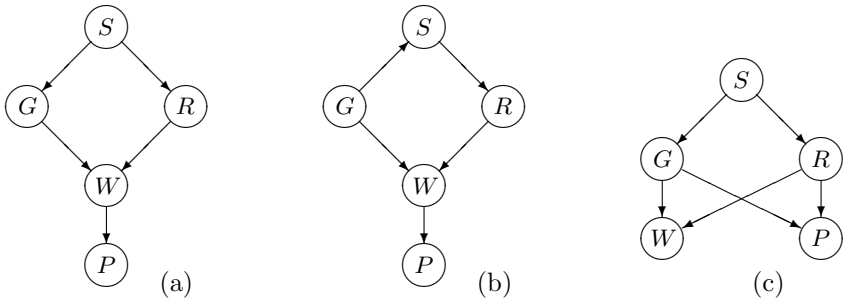

Figure 1: Wet grass example.

This Bayesian network defines a five-dimensional probability distribution

$$\kappa(S,G,R,W,P) = \pi_1(S)\pi_2(G|S)\pi_3(R|S)\pi_4(W|G,R)\pi_5(P|W) \qquad (1)$$

that represent knowledge describing the state of the pavement in front of a house. It may be slippery or not. Usually, it is slippery when it is wet, which happens, for example, during a rain/snow fall or when it is splashed by the garden sprinkler. Usage of the words "usually" and "may be" suggests that a probabilistic model is adequate (there are no deterministic relationships).

It is well-known from Bayesian network theory (see e.g. [1]) that without changing the resulting joint distribution κ we can modify the graph and the system of conditional distributions so that distribution κ is defined by another (but probabilistically equivalent) Bayesian network with a graph from Figure 1(b).

The situation is different when we start considering the model to be causal. It means that we assume that the arrows point from causes to effects. Now, if the graph from Figure 1(a) is causal then the graph from Figure 1(b) cannot be causal because of the different orientation of the arrow connecting nodes S and G. Really, one can hardly imagine that switching the sprinkler on can change winter to summer. Nevertheless, taking the second glance we can see that neither the graph from Figure 1(a) can be considered to be causal. Imagine that you take a watering can and keeping the pavement dry you carefully sprinkle grass. It means, you make an intervention on variable W - *grass is wet*. As a matter of course, it does not change the state of the pavement. Similarly, neither making the pavement slippery (for example using some oil) makes the grass wet. So, for this example, we propose to accept the causal model with the graph from Figure 1(c).

Let us, now, illustrate on this example once more the difference between conditioning and the intervention. From Bayesian network theory we know that if $\kappa(S,G,R,W,P)$ is defined by Formula (1) then $\kappa(G|S) = \pi_2(G|S)$. It is quite natural to assign much higher probability to *sprinkler=on* in summer than in winter. Therefore, since it is quite natural to assume that $\pi_1(S = summer) \doteq \pi_1(S = winter)$ and therefore also $\kappa(S = summer) \doteq \kappa(S = winter)$, we can easily deduce that

$$\kappa(S = summer|G = on) \gg \kappa(S = winter|G = on).$$

On the other hand, as said in Introduction, by the intervention we understand the situation when we change a state of a variable by an external force. It means, for example, that we approach the water tap and let the water sprinkle. Naturally, by this action we do not influence a season of the year. Therefore,

$$\kappa(S = summer|do(G = on)) = \kappa(S = summer)$$
$$\doteq \kappa(S = winter|do(G = on) = \kappa(S = winter).$$

This is why, in causal networks, the intervention is realized not only by a change of a state of the considered variable, but also by deleting all the arrows heading to the considered node [4]. As the reader can see, deleting the arrow from S to G in Figure 1(a) (or Figure 1(c)) makes the variables S and G independent, and therefore the change of a state G does not influence the probability of S.

For a nice, more sophisticated economic causal model see [5].

3 Compositional models

For a more thorough introduction to compositional model theory the reader is referred to [2]. In this paper we will introduce these models rather informally.

The basic idea is simple. Considering a three-dimensional distribution $\pi(X, Y, Z)$ and knowing that variables X and Z are conditionally independent given variable Y, one can decompose π into its two-dimensional marginals $\pi(X, Y)$ and $\pi(Y, Z)$. It means that the original three-dimensional distribution can be unambiguously reconstructed (composed) from its two-dimensional marginals using a simple formula

$$\pi(X, Y, Z) = \frac{\pi(X, Y) \cdot \pi(Y, Z)}{\pi(Y)}.$$

This formula can be rewritten using an operator of composition $\triangleright$ that is defined as follows.

Consider two (non-empty) sets of variables x and y. We do not impose any conditions regarding the mutual relation of these sets; they may be but need not be disjoint, one may be a subset of the other. Let κ and λ be distributions defined for x and y, respectively. To avoid technical problems connected with division by zero, we assume that marginal $\lambda^{\downarrow x\cap y}$ dominates $\kappa^{\downarrow x\cap y}$, i.e.,

$$\lambda^{\downarrow x\cap y}(\cdot) = 0 \implies \kappa^{\downarrow x\cap y}(\cdot).$$

In this case we can define composition of κ and λ by the formula[1]

$$\kappa \triangleright \lambda = \frac{\kappa \cdot \lambda}{\lambda^{\downarrow x\cap y}}.$$

[1]To avoid technical problems, if not specified explicitly otherwise, in this paper we will consider only positive distributions. Under this assumption the dominance assumption holds for any couple of distributions and therefore their composition is always defined.

Note that for disjoint x and y the marginal $\kappa^{\downarrow x\cap y} = \lambda^{\downarrow x\cap y} = 1$, and $\lambda \triangleright \kappa$ simplifies to a product of (independent) distributions.

It is known that the composition of distributions $\kappa(x)$ and $\lambda(y)$ is always a distribution of variables $x \cup y$. This means that an iterative application of the operator of composition to a sequence of low-dimensional distributions may yield a multidimensional distribution. Nevertheless, it is not a difficult task to show that this operator is generally neither commutative nor associative. Therefore, if the opposite is not explicitly specified by brackets, we will always apply the operator of composition from left to right. Therefore, e.g.,

$$\pi \triangleright \kappa \triangleright \lambda \triangleright \mu \triangleright \nu = (((\pi \triangleright \kappa) \triangleright \lambda) \triangleright \mu) \triangleright \nu.$$

To illustrate it let us construct a compositional model for the wet gras example from Section 2. Consider the Bayesian network that is defined by the graph from Figure 1(a), and a system of five distributions $\pi_1(S)$, $\pi_2(G|S)$, $\pi_3(R|S)$, $\pi_4(W|G,R)$, $\pi_5(P|W)$. Defining

$$\begin{aligned}
&\lambda_1(S) = \pi_1(S) = \kappa^{\downarrow\{S\}},\\
&\lambda_2(S,G) = \pi_1(S)\pi_2(G|S) = \kappa^{\downarrow\{S,G\}},\\
&\lambda_3(S,R) = \pi_1(S)\pi_3(R|S) = \kappa^{\downarrow\{S,R\}},\\
&\lambda_4(G,R,W) = (\pi_1(S)\pi_2(G|S)\pi_3(R|S))^{\downarrow\{G,R\}}\pi_4(W|G,R) = \kappa^{\downarrow\{S,R\}},\\
&\lambda_5(W,P) = (\pi_1(S)\pi_2(G|S)\pi_3(R|S)\pi_4(W|G,R))^{\downarrow\{W\}}\pi_5(P|W) = \kappa^{\downarrow\{W,P\}},
\end{aligned}$$

then the compositional model of the distribution represented by this Bayesian network is

$$\kappa(S,G,R,W,P) = \lambda_1(S) \triangleright \lambda_2(S,G) \triangleright \lambda_3(S,R) \triangleright \lambda_4(G,R,W) \triangleright \lambda_5(W,P).$$

4 Conditioning by composition

From now on, consider a general probability distribution $\kappa(X_1, X_2, \ldots, X_n)$ and define a *degenerated* one-dimensional probability distribution $\nu_{|i;\alpha}$ as a distribution of variable X_i achieving probability 1 for value $X_i = \alpha$, i.e.,

$$\nu_{|i;\alpha}(X_i) = \begin{cases} 1 & \text{if } X_i = \alpha,\\ 0 & \text{otherwise.}\end{cases}$$

Let us compute

$$\nu_{|i;a} \triangleright \kappa = \frac{\nu_{|i;a}(X_i)\cdot\kappa(X_1,\ldots,X_n)}{\kappa^{\downarrow\{X_i\}}}$$

for any combination of values of all variables $X_1, \ldots, X_n$. It is clear that if $X_i \neq \alpha$, $\nu_{|i;a} \triangleright \kappa = 0$. In opposite case, if $X_i = \alpha$, then

$$\begin{aligned}
\nu_{|i;a} \triangleright \kappa &= \frac{\kappa(X_1,\ldots,X_{i-1},X_i=\alpha,X_{i+1},\ldots,X_n)}{\kappa^{\downarrow\{X_i\}}(X_i=a)}\\
&= \kappa(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)|X_i=\alpha).
\end{aligned}$$

It means that $\nu_{|i;a} \triangleright \kappa$ is an n-dimensional distribution that equals 0 for all combinations of values for which $X_i \neq \alpha$. In case that $X_i = \alpha$, then it equals the conditional distribution $\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)|X_i = \alpha)$. Therefore

$$\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|X_i = \alpha) = (\nu_{|i;a} \triangleright \kappa)^{-\{X_i\}}.$$

Naturally, this way of expressing conditional distributions can also be used for distributions represented as compositional models. Therefore, for

$$\kappa(X_1, \ldots, X_n) = \mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m$$

we get

$$\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|X_i = \alpha) = \big(\nu_{|i;a} \triangleright (\mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m)\big)^{-\{X_i\}}.$$

As said above, in this paper we do not have space to go into theory of compositional models, nevertheless let us state that the brackets in the preceding formula are important. This is because the operator of composition is not associative. Moreover, in the next section we will show an important property: namely, if $\mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m$ is a causal model then the intervention is computed by the formula

$$\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|do(X_i = \alpha)) = \big(\nu_{|i;a} \triangleright \mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m\big)^{-\{X_i\}}.$$

5 Compositional causal models

Consider a set $\{X_1, X_2, \ldots, X_n\}$ of finite-state variables. For each variable let $\mathcal{C}(X_i)$ denote the set of the variables that are causes of X_i. Naturally, some of $\mathcal{C}(X_i)$ may be empty (in fact, to get a correct model at least one of these sets *must* be empty), and $X_i \notin \mathcal{C}(X_i)$. We say that the causal model is *correct* if there exists an ordering of variables (without loss of generality we will assume that it is the ordering $X_1, X_2, \ldots, X_n$) such that $\mathcal{C}(X_1) = \emptyset$, and for all $i = 2, 3, \ldots, n$ $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$.

For the sake of simplicity denote $x_i = \mathcal{C}(X_i) \cup \{X_i\}$. If we have probability distributions $\mu_i(x_i)$ we can construct a compositional causal model (CCM) as

$$\kappa(X_1, \ldots, X_n) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n).$$

There are several theorems in [2] saying under what conditions one can change the ordering of distributions in a compositional model without influencing the resulting joint distribution. It is important to stress that for causal models, most of such transformations are forbidden. For causal models, we can consider only those orderings that guarantee the correctness of a causal model, i.e., for which $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. And it is the result of Kratochvíl that says that all these orderings define the same joint probability distribution $\kappa(X_1, \ldots, X_n)$ (see [3]).

At the end of the last section we promised to show how to compute the result of an intervention in CCMs. Let us repeat the idea of Pearl [4], who computes it as a conditioning in a Bayesian network, in which all the arrows heading to the intervention node are deleted. To do the same in a CCM we will need a possibility to find a causal graph corresponding to a given CCM.

Consider a CCM $\mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$. If it is constructed in the way described at the beginning of this section then for all $i = 1, 2, \ldots, n$ the set $x_i \setminus (x_1 \cup \ldots \cup x_{i-1})$ is a singleton (i.e., $|x_i \setminus (x_1 \cup \ldots \cup x_{i-1})| = 1$). In a few lines below we will need a minor generalization of this condition, namely that

$$|x_i \setminus (x_1 \cup \ldots \cup x_{i-1})| \leq 1. \tag{2}$$

Let us construct a causal graph from $\mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$ meeting condition (2). The elements from $x_1 \cup \ldots \cup x_n$ are nodes of the constructed causal graph, and there is an arrow $(X_i \to X_j)$ in this graph if and only if there is a distribution $\mu_k(x_k)$ in the CCM, for which

$$\begin{aligned} &X_j \in x_k; \\ &X_j \notin x_1 \cup \ldots \cup x_{k-1}; \\ &X_i \in x_1 \cup \ldots \cup x_{k-1}. \end{aligned}$$

Now, consider two CCMs:

$$\begin{aligned} \kappa(X_1, \ldots, X_n) &= \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n), \\ \lambda(X_1, \ldots, X_n) &= \kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n), \end{aligned}$$

and construct for both of them the corresponding causal graphs. It is evident that if condition (2) holds for CCM κ, then it holds true also for CCM λ. Moreover, the reader can almost immediately see that, in the causal graph corresponding to CCM λ, there are no arrows heading to node X_i, and that all the other arrows from the causal graph corresponding to CCM κ are preserved in the causal graph corresponding to CCM λ. It means that an intervention in CCM κ can be done through conditioning in CCM λ (see [4]):

$$\begin{aligned} &\kappa(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | do(X_i = \alpha)) \\ &\qquad = \lambda(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n | X_i = \alpha) \\ &\qquad = \left(\nu_{|i;a} \triangleright (\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n))\right)^{-\{X_i\}} \\ &\qquad = \left(\nu_{|i;a} \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\right)^{-\{X_i\}}. \end{aligned}$$

(Validity of the last equality is not trivial and therefore it is proved in Appendix because it needs some more knowledge from compositional model theory.)

6 Conclusions

As promised in Introduction, we have presented some preliminary ideas regarding causal compositional models. Namely, we have shown how to compute intervention in these models. For the sake of simplicity we have assumed that the

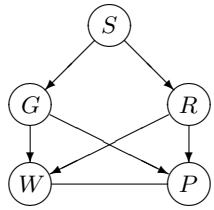

Figure 2: Generalized causal graph.

considered probability distributions are positive. Let us stress that this assumption can easily be replaced by the assumption of dominance of the respective distributions. So, the approach can be used also in situations when some of the considered dependencies are deterministic.

Another straightforward generalization concerns situations that appear in practical problems. When constructing a causal model, it happens quite often that we know about the dependence of two (or more) variables but we do not know which one is the cause of the other. It may be that both are influenced by a third (unknown) variable. For example, consider the wet grass example with the graph in Figure 1(c). We know that both the variables P and W are influenced by G and R but we can hardly assume that they are conditionally independent given G and R. For example, when the grass and the pavement are covered with hoarfrost, the grass is wet and pavement is slippery though the grass sprinkler may be off and it may not rain/snow. Therefore it may be useful to have a possibility to introduce a dependence between (among) variables that need not be causal. In the considered wet grass example it can easily be done by considering that a couple of variables P, W has a common cause $\mathcal{C}(G, R)$, and therefore the respective compositional model is

$$\bar{\kappa}(S, G, R, W, P) = \mu_1(S) \triangleright \mu_2(S, G) \triangleright \mu_3(S, R) \triangleright \mu_4(G, R, W, P).$$

Notice, that this generalization violates condition (2), and that it does not correspond to any causal graph. Here we would have to consider a generalization of such graphs allowing them to contain also indirected edges (see Figure 2).

Acknowledgement

The research was partially supported by GA ČR under the project number GA 403/12/2175.

References

[1] Jensen F. V. and Nielsen T. D. (2007) *Bayesian networks and decision graphs*, Springer.

[2] Jiroušek R. (2011), Foundations of compositional model theory, *International Journal of General Systems*, 40, 6, 623-678.

[3] Kratochvíl V. (2013), Probabilistic Compositional Models: solution of an equivalence problem , *International Journal of Approximate Reasoning* 54, 5, p. 590-601.

[4] Pearl J. (2009) *Causality: Models, Reasoning, and Inference*, Cambridge University Press, New York.

[5] Váchová L. and Bína Vl. (2013), A Note on Computation of Causal Effect in Economic Models, accepted for publication in: *Proceedings of the 31st Mathematical models in Economy.*

Appendix

To show that

$$\begin{aligned}\nu_{|i;a} \triangleright \big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\big) \\ = \nu_{|i;a} \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\end{aligned} \tag{3}$$

we need two assertions that can be found in [2]:

Assertion. If the operators of composition in the formulas below are defined then

1. $(\pi_1(y_1) \triangleright \pi_2(y_2))^{\downarrow y_1} = \pi_1(y_1)$;
2. if $y_2 \supseteq y_1 \cap y_3$ then $\pi_1(y_1) \triangleright \pi_2(y_2) \triangleright \pi_3(y_3) = \pi_1(y_1) \triangleright (\pi_2(y_2) \triangleright \pi_3(y_3))$.

When computing $\nu_{|i;a} \triangleright \big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\big)$, it is important to realize that both $\nu_{|i;a}$ and $\kappa^{\downarrow X_i}$ are distributions defined for the same variable X_i. Therefore we can apply property 2 from Assertion getting

$$\begin{aligned}\nu_{|i;a} \triangleright \Big(\Big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-1}(x_{n-1})\Big) \triangleright \mu_n(x_n)\Big) \\ = \nu_{|i;a} \triangleright \Big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-1}(x_{n-1})\Big) \triangleright \mu_n(x_n).\end{aligned}$$

The same idea can also be applied to a shorter sequence, which yields

$$\begin{aligned}\nu_{|i;a} \triangleright \Big(\Big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-2}(x_{n-2})\Big) \triangleright \mu_{n-1}(x_{n-1})\Big) \\ = \nu_{|i;a} \triangleright \Big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_{n-2}(x_{n-2})\Big) \triangleright \mu_{n-1}(x_{n-1}).\end{aligned}$$

Thus, applying property 2 from Assertion n times we get that

$$\begin{aligned}\nu_{|i;a} \triangleright \big(\kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\big) \\ = \nu_{|i;a} \triangleright \kappa^{\downarrow X_i}(X_i) \triangleright \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n),\end{aligned}$$

and to show the validity of the required equation (3) it is enough to apply property 1 to the first operator of composition.

Portfolio Selection with Interactive Fuzzy Coefficients

Masahiro Inuiguchi
Department of Systems Innovation
Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
e-mail: inuiguti@sys.es.osaka-u.ac.jp

Abstract

The conventional fuzzy programming approach to the portfolio selection problem may yield a concentrated investment solution due to the implicit assumption of non-interaction among fuzzy coefficients. Two countermeasures have been proposed: one is minimax regret model and the other is the introduction of interaction. In this paper, we demonstrate how much diversified investment solution is obtained by the combinations of those two countermeasures.

1 Introduction

In fuzzy programming problems, it has been implicitly assumed that the fuzzy coefficients are non-interactive. Because of this implicit assumption, a concentrated investment solution is frequently obtained to portfolio selection problems with fuzzy coefficients by the conventional fuzzy programming approaches. To cope with this problem, Inuiguchi et al. [1] proposed the minimax regret approach and showed that a diversified investment solution is obtained. On the other hand, we can introduce the interaction among fuzzy coefficients to the fuzzy portfolio selection problem. So far, several models treating the interaction among fuzzy coefficients have been proposed. In this paper, we introduce some models treating the interaction among fuzzy coefficients to the portfolio selection problems. By numerical examples, we demonstrate the effect of the introduction of interaction and the effect of minimax regret approach as well as the effect of the combination.

2 Portfolio Selection with Fuzzy Coefficients

We treat the following portfolio selection problem with fuzzy return rates:

$$\text{maximize } \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}, \quad \text{sub. to } \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = 1,\ \boldsymbol{x} \geq 0, \tag{1}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_n)^{\mathrm{T}}$ is a variable vector of return rates whose variation range is represented by an n-dimensional fuzzy set C.

Assuming that all return rates are non-interactive one another, we can apply the conventional fuzzy programming approaches. However, the solutions are concentrated or semi-concentrated investment solutions [1]. Then, the following minimax regret model:

$$\begin{array}{ll} \text{minimize} & z, \\ \text{sub. to} & \max\limits_{\boldsymbol{c} \in \mathrm{cl}\ (C)_{1-\alpha^0}} c_i - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \leq z,\ i = 1, 2, \ldots, n,\ \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = 1,\ \boldsymbol{x} \geq 0, \end{array} \tag{2}$$

is introduced to obtain a diversified investment solution [1]. As another way to obtain a diversified investment solution is to introduce the interaction. Then, some models to treat interaction among fuzzy coefficients without great loss of the tractability have been proposed [2, 3]. In this paper, we consider scenario decomposed fuzzy numbers and oblique fuzzy vector as models of interactive fuzzy numbers and formulate the portfolio selection problem based on the minimax regret approach.

3 Scenario Decomposed Fuzzy Numbers

The return rates of bonds are often influenced by the economic situation. Then the estimated return rates of bounds can be different by the economic situation. This kind of the estimated return rates can be represented by scenario decomposed fuzzy numbers proposed by Inuiguchi and Tanino [2]. In this approach, the possible ranges of uncertain parameters which depend on the situation are expressed by fuzzy if-then rules.

We may have a vague knowledge about the possible range of $\boldsymbol{\gamma}$ as the following k fuzzy if-then rules:

$$\text{if } s = s_k \text{ then } \boldsymbol{\gamma} \in \boldsymbol{C}^k,\ k = 1, 2, \ldots, u, \tag{3}$$

where s is a variable taking a value from $\{s_1, s_2, \ldots, s_u\}$. s is called a *scenario variable* and showing the situation. $\boldsymbol{C}^k = (C_1^k, C_2^k, \ldots, C_n^k)^{\mathrm{T}}$ is a vector of non-interactive fuzzy numbers. Namely, $\boldsymbol{C}^k$ has a membership function,

$$\mu_{\boldsymbol{C}^k}(\boldsymbol{c}) = \min\left(\mu_{C_1^k}(c_1), \mu_{C_2^k}(c_2), \ldots, \mu_{C_n^k}(c_n)\right), \tag{4}$$

and C_j^k is a fuzzy number such that $[C_j^k]_h = \{r \mid \mu_{C_j^k}(r) \geq h\}$ is a bounded closed interval, where $\mu_{C_j^k}$ is a membership function of a fuzzy number C_j^k. The body of rules (3) shows a fuzzy relation between scenario variable s and possible range of uncertain vector $\boldsymbol{\gamma}$.

When we obtain the estimated range of scenario variable s as a fuzzy set S showing a possible rage of s, the estimated fuzzy set $\boldsymbol{C}$ is obtained as

$$\mu_{\boldsymbol{C}}(\boldsymbol{c}) = \max_{k=1,2,\ldots,u} \min\left(\mu_S(s_k), \mu_{\boldsymbol{C}^k}(\boldsymbol{c}),\right) \tag{5}$$

where μ_S is a membership function of S. Inuiguchi and Tanino [2] considered a continuous scenario variable. In the continuous scenario variable case, the knowledge can be represented by a set of fuzzy if-then rules, "if s is in fuzzy set S^k then $\boldsymbol{\gamma}$ is in fuzzy set C^k". Let μ_k be the membership function of fuzzy set S^k in the antecedent part of fuzzy if-then rules. For the sake of simplicity, we assume $\sum_{k=1}^{u} \mu_k(s) = 1,\ \forall s$. Then for $s = \bar{s}$, the estimated range $C_j(\bar{s})$ of the return rate of j-th bond is defined by

$$C_j(\bar{s}) = \sum_{k=1}^{u} \mu_k(\bar{s}) C_j^k. \tag{6}$$

The extension principle is applied to calculate $C_j(\bar{s})$. Let $\boldsymbol{C}(\bar{s}) = (C_1(\bar{s}), \ldots, C_n(\bar{s}))^{\mathrm{T}}$. The estimated fuzzy set $\boldsymbol{C}$ under a fuzzy set S showing a possible realizations of s is obtained as

$$\mu_{\boldsymbol{C}}(\boldsymbol{c}) \quad = \quad \sup_s \min\left(\mu_S(s), \mu_{\boldsymbol{C}(s)}(\boldsymbol{c})\right). \tag{7}$$

The fuzzy set whose membership function is defined by (3) and (7) is called scenario decomposed fuzzy numbers.

Now let us investigate the possible range of a linear function value $\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}$ with scenario decomposed fuzzy numbers $\boldsymbol{C}$ with membership function defined by (7). Let $Y(\boldsymbol{x})$ and $Y^k(\boldsymbol{x})$ be fuzzy sets defined by membership functions,

$$\mu_{Y(\boldsymbol{x})}(y) = \sup\left\{\mu_{\boldsymbol{C}}(\boldsymbol{c}) \mid \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = y\right\}, \quad \mu_{Y^k(\boldsymbol{x})}(y) = \sup\left\{\mu_{\boldsymbol{C}^k}(\boldsymbol{c}) \mid \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = y\right\}. \tag{8}$$

Namely, $Y(\boldsymbol{x})$ shows the overall possible range of $\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}$ while $Y^k(\boldsymbol{x})$ shows the possible range of $\boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{x}$ when the possible range of $\boldsymbol{\gamma}$ is given by $\boldsymbol{C}^k$.

Because the linearity of function is preserved in the extension principle, We obtain the following relation between $Y(\boldsymbol{x})$ and $Y^k(\boldsymbol{x})$:

$$\mu_{Y(\boldsymbol{x})}(y) = \sup_s \min\left(\mu_S(s), \sup_{\boldsymbol{r}:\boldsymbol{\mu}(s)^{\mathrm{T}}\boldsymbol{r}=y} \min\left(\mu_{Y^1(\boldsymbol{x})}(r_1), \ldots, \mu_{Y^u(\boldsymbol{x})}(r_u)\right)\right), \tag{9}$$

where $\boldsymbol{r} = (r_1, r_2, \ldots, r_u)^{\mathrm{T}}$ and $\boldsymbol{\mu}(s) = (\mu_1(s), \mu_2(s), \ldots, \mu_u(s))^{\mathrm{T}}$.

From (6) and the non-negativity of $\boldsymbol{x}$, we obtain

$$\begin{aligned} & N_{Y(\boldsymbol{x})}([z, +\infty)) \geq \alpha^0 \Leftrightarrow \mathrm{cl}(Y(\boldsymbol{x}))_{1-\alpha^0} \subseteq [z, +\infty) \\ & \quad \Leftrightarrow \quad \mathrm{cl}(Y^k(\boldsymbol{x}))_{1-\alpha^0} \subseteq [z, +\infty),\ \forall s \text{ such that } \mu_S(s) > 1 - \alpha^0 \\ & \quad \Leftrightarrow \quad \sum_{j=1}^{n}\sum_{k=1}^{u} \mu_k(s) \bar{c}_{jk}^{\mathrm{L}}(1-\alpha^0) x_j \geq z,\ \forall s \text{ such that } \mu_S(s) > 1 - \alpha^0, \end{aligned} \tag{10}$$

where we define $\mathrm{cl}(C_j^k)_\alpha = [\bar{c}_{jk}^{\mathrm{L}}(\alpha), \bar{c}_{jk}^{\mathrm{R}}(\alpha)]$, $k = 1, 2, \ldots, u$.

Now we describe the minimax regret model. We assume that the range of scenario variable s is a subset of real line $\mathbf{R}$, and S is a fuzzy number.

In the same way as the calculation of a linear function value of scenario decomposed fuzzy numbers, Problem (2) is reduced to the following linear programming problem:

$$
\begin{aligned}
\text{minimize} \quad & q, \\
\text{sub. to} \quad & c_{ik}^{\mathrm{R}}(1-\alpha^0)(1-x_i) - \sum_{j=1, j\neq i}^{n} c_{jk}^{\mathrm{L}}(1-\alpha^0)x_j \leq q,\ i=1,2,\ldots,n, \\
& \qquad\qquad\qquad\qquad\qquad\qquad k \ \text{ such that } [S]_1 \cap [S]_k \neq \emptyset, \\
& \boldsymbol{\mu}(\bar{s}^{\mathrm{L}}(1-\alpha^0))^{\mathrm{T}} \bar{\boldsymbol{c}}_i^{\mathrm{R}}(1-\alpha^0)(1-x_i) \\
& \qquad - \sum_{j=1, j\neq i}^{n} \boldsymbol{\mu}(\bar{s}^{\mathrm{L}}(1-\alpha^0))^{\mathrm{T}} \bar{\boldsymbol{c}}_j^{\mathrm{L}}(1-\alpha^0)x_j \leq q \ i=1,2,\ldots,n, \\
& \boldsymbol{\mu}(\bar{s}^{\mathrm{R}}(1-\alpha^0))^{\mathrm{T}} \bar{\boldsymbol{c}}_i^{\mathrm{R}}(1-\alpha^0)(1-x_i) \\
& \qquad - \sum_{j=1, j\neq i}^{n} \boldsymbol{\mu}(\bar{s}^{\mathrm{R}}(1-\alpha^0))^{\mathrm{T}} \bar{\boldsymbol{c}}_j^{\mathrm{L}}(1-\alpha^0)x_j \leq q \ i=1,2,\ldots,n, \\
& \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = 1, \ \boldsymbol{x} \geq 0,
\end{aligned}
\tag{11}
$$

where $\bar{\boldsymbol{c}}_j^{\mathrm{L}}(\alpha) = (c_{j1}^{\mathrm{L}}(\alpha), c_{j2}^{\mathrm{L}}(\alpha), \ldots, c_{ju}^{\mathrm{L}}(\alpha))^{\mathrm{T}}$ and $\bar{\boldsymbol{c}}_j^{\mathrm{R}}(\alpha) = (c_{j1}^{\mathrm{R}}(\alpha), c_{j2}^{\mathrm{R}}(\alpha), \ldots, c_{ju}^{\mathrm{R}}(\alpha))^{\mathrm{T}}$. $\bar{s}^{\mathrm{L}}(\alpha)$ and $\bar{s}^{\mathrm{R}}(\alpha)$ are defined by $\mathrm{cl}(S)_\alpha = [\bar{s}^{\mathrm{L}}(\alpha), \bar{s}^{\mathrm{R}}(\alpha)]$.

4 Oblique Fuzzy Vector

From the historical data, we may find a vague knowledge about a linear function value of return rates of several bonds and the differences of two uncertain values, e.g., $\gamma_1 + 2\gamma_2 + \gamma_3$ is about 1.3, $\gamma_4 - \gamma_5$ is approximately 0.1, and so on. If we have only n independent pieces of vague knowledge about the linear function values of return rates of bounds, we can apply oblique fuzzy vector [3] to represent the possible range of return rate vector.

Oblique fuzzy vectors are proposed by Inuiguchi et al. [3] and each of them can express n independent pieces of vague knowledge about the linear function values of uncertain values. A non-singular matrix shows the interaction among uncertain parameters in an oblique fuzzy vector as a covariance matrix shows in a multivariate normal distribution.

An oblique fuzzy vector $\boldsymbol{C}$ is defined by the following membership function,

$$
\mu_{\boldsymbol{C}}(\boldsymbol{c}) = \min_{j=1,2,\ldots,n} \mu_{B_j}(\boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{c}), \tag{12}
$$

where μ_{B_j} is a membership function of an L-L fuzzy number $B_j = (b_j^{\mathrm{L}}, b_j^{\mathrm{R}}, \beta_j^{\mathrm{L}}, \beta_j^{\mathrm{R}})_{LL}$ and $\boldsymbol{d}_j$, $j = 1, 2, \ldots, n$ are vectors such that $D = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_n)^{\mathrm{T}}$ be a non-singular real-valued $n \times n$ matrix.

Now let us investigate a minimax regret model. From Problem (2), we consider linear function values

$$
R_i(\boldsymbol{x}) = c_i - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = \sum_{l=1, l\neq i}^{n} c_l x_l + c_i(1-x_i). \tag{13}
$$

In the same way as we did for the calculation of linear function value of an oblique fuzzy vector [3], we obtain for any $\alpha \in [0,1)$,

$$\begin{aligned} \mathrm{cl}(R_i(\boldsymbol{x}))_\alpha &= \Bigg[\sum_{j:k_j^i(\boldsymbol{x})\geq 0} \bar{b}_j^{\mathrm{L}}(\alpha) k_j^i(\boldsymbol{x}) + \sum_{j:k_j^i(\boldsymbol{x})<0} \bar{b}_j^{\mathrm{R}}(\alpha) k_j^i(\boldsymbol{x}), \\ &\qquad \sum_{j:k_j^i(\boldsymbol{x})\geq 0} \bar{b}_j^{\mathrm{R}}(\alpha) k_j^i(\boldsymbol{x}) + \sum_{j:k_j^i(\boldsymbol{x})<0} \bar{b}_j^{\mathrm{L}}(\alpha) k_j^i(\boldsymbol{x}) \Bigg], \quad (14) \end{aligned}$$

where with letting d_{ij}^* be the (i,j) component of D^{-1}, we define

$$k_j^i(\boldsymbol{x}) = \sum_{l=1, l\neq i}^{n} d_{lj}^* x_l + d_{ij}^*(1-x_i). \quad (15)$$

$\bar{b}_j^{\mathrm{L}}(h)$ and $\bar{b}_j^{\mathrm{R}}(h)$ are defined by $\bar{b}_j^{\mathrm{L}}(h) = b_j^{\mathrm{L}} - \beta_j^{\mathrm{L}} L^*(h)$ and $\bar{b}_j^{\mathrm{R}}(h) = b_j^{\mathrm{R}} - \beta_j^{\mathrm{R}} L^*(h)$, respectively. L^* is defined by $L^*(h) = \sup\{r \in \mathbf{R}_+ \mid L(r) \geq h\}$.

Introducing $\boldsymbol{y}_i^+ = (y_{i1}^+, \ldots, y_{in}^+)^{\mathrm{T}}$ and $\boldsymbol{y}_i^- = (y_{i1}^-, \ldots, y_{in}^-)^{\mathrm{T}}$ such that

$$k_j^i(\boldsymbol{x}) = y_{ij}^+ - y_{ij}^-,\ y_{ij}^+ \cdot y_{ij}^- = 0,\ y_{ij}^+ \geq 0,\ y_{ij}^- \geq 0,\ j = 1,2,\ldots,n, \quad (16)$$

we obtain the following reduced linear programming problem:

$$\begin{aligned} \text{minimize} \quad & q, \\ \text{sub. to} \quad & \sum_{j=1}^{n} \bar{b}_j^{\mathrm{R}}(1-h^0) y_{ij}^+ - \sum_{j=1}^{n} \bar{b}_j^{\mathrm{L}}(1-h^0) y_{ij}^- \leq q,\ i = 1,2,\ldots,n, \\ & x_i = 1 - \sum_{j=1}^{n} d_{ji}(y_{ij}^+ - y_{ij}^-),\ i = 1,2,\ldots,n, \\ & x_l = \sum_{j=1}^{n} d_{jl}(y_{ij}^+ - y_{ij}^-),\ l = 1,2,\ldots,n\ (l \neq i),\ i = 1,2,\ldots,n, \\ & y_{ij}^+ \geq 0,\ y_{ij}^- \geq 0,\ i = 1,2,\ldots,n,\ j = 1,2,\ldots,n,\ \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = 1,\ \boldsymbol{x} \geq 0, \end{aligned} \quad (17)$$

where d_{ij} is the (i,j)-component of D. This problem can also be solved by Bender's decomposition method.

5 Numerical Examples

Example 1. As an example of scenario decomposed fuzzy numbers, we consider a case where the possible range of return rates of five bonds in different categories of industry estimated by the following fuzzy if-then rules:

- if s is *small* then the return rate vector $\boldsymbol{\gamma}$ is in a possible range $\boldsymbol{C}_1$,
- if s is *medium* then the return rate vector $\boldsymbol{\gamma}$ is in a possible range $\boldsymbol{C}_2$,
- if s is *large* then the return rate vector $\boldsymbol{\gamma}$ is in a possible range $\boldsymbol{C}_3$, (18)

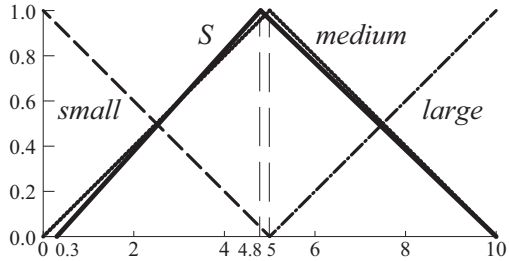

Figure 1: Fuzzy sets *small*, *medium* and *large* and fuzzy set S

Table 1: Parameters of C_{ij}

s	i	c_{i1}	w_{i1}	c_{i2}	w_{i2}	c_{i3}	w_{i3}	c_{i4}	w_{i4}	c_{i5}	w_{i5}
small	1	0.18	0.023	0.22	0.05	0.26	0.01	0.2	0.026	0.2	0.008
medium	2	0.25	0.023	0.22	0.014	0.19	0.013	0.16	0.005	0.14	0.004
large	3	0.3	0.03	0.18	0.015	0.18	0.0225	0.2	0.006	0.13	0.004

where fuzzy sets *small*, *medium* and *large* are triangular fuzzy numbers depicted in Figure 1. $\boldsymbol{C}_i$, $i = 1, 2, 3$ are non-interactive fuzzy numbers whose component C_{ij} has the following type of membership function:

$$\mu_{C_{ij}}(r) = \exp\left(-\frac{(r - c_{ij}^{\mathrm{c}})^2}{w_{ij}}\right). \tag{19}$$

Parameters c_{ij}^{c} and w_{ij}, $i = 1, 2, 3$, $j = 1, 2, \ldots, 5$ are defined by the values in Table 1. The estimated possible range S of scenario variable is a triangular fuzzy number depicted in Figure 1.

Let $\alpha = 0.7$. We obtain $\bar{s}^{\mathrm{L}} = 3.45$ and $\bar{s}^{\mathrm{R}} = 6.36$. From the necessity fractile optimization model, we obtain the optimal solution shown in Table 2. On the other hand, from the minimum regret model with $\alpha = 0.7$, we obtain the optimal solution shown in Table 2.

Comparing the solutions in Table 2, we observe that the solution of the necessity fractile optimization model suggests the investment to bond whose estimated return rates are small and their variations are small while the solution to the minimax regret model suggests the investment to bond whose estimated return rates are large and their variations are large. In the necessity fractile optimization model, only the minimal return rates of bonds are used to estimate the worst case and thus a pessimistic solution is obtained. On the other hand, in the minimax regret model, the maximal return rates of bonds are also used to estimate the worst regret and thus the solution is not very pessimistic. Finally, we notes that the solutions are easily changed by a small change of parameters because the five bonds are comparable.

Example 2. As an example of the oblique fuzzy vector, we consider a case

Table 2: Necessity fractile optim. and minimax regret models - solution

Model	x_1	x_2	x_3	x_4	x_5	z/q
Necessity fractile	0	0	0.33456	0.54488	0.12056	0.07513
Minimax regret	0.42766	0.32750	0.22807	0.01677	0	0.210966

where D is given by

$$D = \begin{pmatrix} 1 & 7 & -1.5 & 1 & -6 \\ 0 & 20 & 20 & 10 & 3 \\ 0 & 0 & 0.5 & 3 & 3 \\ 0 & 0 & 0 & 3 & 2 \\ 4 & 0 & 6 & -6.5 & 3 \end{pmatrix}, \tag{20}$$

and B_i, $i = 1, 2, \ldots, 5$ is defined by membership function,

$$\mu_{B_i}(r) = \exp\left(-\frac{(r - b_i^{\mathrm{c}})^2}{s_i}\right), \tag{21}$$

with parameters: $b_1^{\mathrm{c}} = 0.396$, $b_2^{\mathrm{c}} = 11.194$, $b_3^{\mathrm{c}} = 1.396$, $b_4^{\mathrm{c}} = 1.078$, $b_5^{\mathrm{c}} = 1.463$, $s_1 = 0.008$, $s_2 = 0.0025$, $s_3 = 0.0036$, $s_4 = 0.0009$, $s_5 = 0.006$. In this case,

$$D^{-1} = \begin{pmatrix} 0.091265 & -0.031943 & -1.174707 & 1.742992 & 0.227184 \\ 0.170795 & -0.009778 & 1.415906 & -1.532757 & -0.042699 \\ -0.156454 & 0.054759 & -1.129074 & 1.083442 & 0.039113 \\ -0.052151 & 0.018253 & -1.043025 & 1.361147 & 0.013038 \\ 0.078227 & -0.027379 & 1.564537 & -1.541721 & -0.019557 \end{pmatrix} \tag{22}$$

and the membership function of marginal fuzzy set C_{i0}, $i = 1, 2, \ldots, 5$ is obtained as (19) with parameters: $c_{10}^{\mathrm{c}} = 0.25$, $c_{20}^{\mathrm{c}} = 0.22$, $c_{30}^{\mathrm{c}} = 0.2$, $c_{40}^{\mathrm{c}} = 0.214$, $c_{50}^{\mathrm{c}} = 0.218$, $w_{10} = 0.022539$, $w_{20} = 0.022503$, $w_{30} = 0.014402$, $w_{40} = 0.012101$, $w_{50} = 0.022501$.

Applying the necessity fractile model with $\alpha = 0.7$, we obtain an optimal solution as $y_1^+ = 0$, $y_2^+ = 0.015873$, $y_3^+ = 0$, $y_4^+ = 0$, $y_5^+ = 0.024420$, $y_1^- = 0$, $y_2^- = 0$, $y_3^- = 0$, $y_4^- = 0$, $y_5^- = 0$. Then the optimal investment rate shares $\boldsymbol{x}$ is obtained by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = D^{\mathrm{T}} \left(\begin{pmatrix} y_1^+ \\ y_2^+ \\ y_3^+ \\ y_4^+ \\ y_5^+ \end{pmatrix} - \begin{pmatrix} y_1^- \\ y_2^- \\ y_3^- \\ y_4^- \\ y_5^- \end{pmatrix} \right) = \begin{pmatrix} 0.097680 \\ 0.317460 \\ 0.463981 \\ 0 \\ 0.120879 \end{pmatrix}.$$

On the other hand, applying minimax regret model with $\alpha = 0.7$, we obtain an optimal solution q, $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)^{\mathrm{T}}$, $Y^+ = (y_{ij}^+)$ and $Y^- = (y_{ij}^-)$ as

$$q = 0.440294, \quad \boldsymbol{x} = (0.159177,\ 0.328070,\ 0.043885,\ 0.061670,\ 0.407198)^{\mathrm{T}},$$

$$Y^+ = \begin{pmatrix} 0.154542 & 0 & 0 & 0.466399 & 0.171570 \\ 0.151061 & 0 & 1.287608 & 0 & 0.002029 \\ 0 & 0.034040 & 0 & 0.266643 & 0.052392 \\ 0.046613 & 0 & 0 & 0.471559 & 0.028141 \\ 0.106851 & 0 & 1.091117 & 0 & 0.013082 \end{pmatrix},$$

$$Y^- = \begin{pmatrix} 0 & 0.037686 & 0 & 0 & 0 \\ 0 & 0.019275 & 0 & 1.248760 & 0 \\ 0.050390 & 0 & 0.229241 & 0 & 0 \\ 0 & 0 & 0.113644 & 0 & 0 \\ 0 & 0.020994 & 0 & 1.007853 & 0 \end{pmatrix}.$$

We observe the difference between those solutions: the solution of the necessity fractile optimization model suggests large amounts of investment to x_2 and x_3 while the solution of the minimax regret model suggests large amounts of investment to x_2 and x_5.

Finally to see the significance of the interaction, we solve the problem with non-interactive fuzzy numbers having the marginal membership functions. Namely, this problem can be seen as the problem discarding the interaction. We obtain $\boldsymbol{x} = (0, 0, 0, 1, 0)^{\mathrm{T}}$ from necessity fractile optimization model, and

$$q = 0.4640220,\ \boldsymbol{x} = (0.255508, 0.190326, 0.179810, 0.155279, 0.219077)^{\mathrm{T}} \quad (23)$$

from the conventional minimax regret model. We observe the big differences of solutions between problems with oblique fuzzy vector and with non-interactive fuzzy numbers. In necessity fractile optimization model, while the solution to the problem discarding the interaction suggests the concentrate investment on the fourth bond, the solution to the problem taking care of the interaction suggests no investment on the fourth bond. Similarly, in the conventional minimax regret model, while the solution to the problem discarding the interaction suggests the investment of more than 15% of the fund on the fourth bond, the solution to the problem taking care of the interaction suggests the investment of less than 7% of the fund. These facts show that the fourth bond is not very attractive one if we consider the interaction among the return rates of five bonds. Moreover, the distributions of the fund are significantly different between the solutions of the two conventional minimax regret models. By these results, we understand the significance of the interaction.

Acknowledgement. This work was supported by JSPS KAKENHI No.23510169

References

[1] Inuiguchi, M., Tanino, T. (2000), Portfolio selection under independent possibilistic information, *Fuzzy Set. Syst.* **115**(1), 83–92.

[2] Inuiguchi, M., Tanino, T. (2002), Possibilistic linear programming with fuzzy if-then rule coefficients, *Fuzzy Optim. Decis. Ma.* **1**(1), 65–91.

[3] Inuiguchi, M., Ramík, J., Tanino, T. (2003), Oblique fuzzy vectors and their use in possibilistic linear programming, *Fuzzy Set. Syst.* **137**(1), 123–150.

Decision Making in MAX-PROD Algebra

Hana Tomášková
Faculty of Informatics and Management
University of Hradec Králové
hana.tomaskova@uhk.cz

Martin Gavalec
Faculty of Informatics and Management
University of Hradec Králové
martin.gavalec@uhk.cz

Abstract

In AHP approach to multi-criteria decision problem, the relative importance of alternatives is computed from preference matrices, which come from experience and can possibly be inconsistent. Standardly, the preference vector is computed as the eigenvector of the preference matrix by methods of linear algebra.

Alternative use of non-standard methods in tropical algebra is considered in this paper. Two most frequently used tropical algebras are the max-plus and the max-prod algebra. The preference matrix will be processed by the methods used in max-prod algebra. By max-prod algebra we understand a linear structure on a linearly ordered set R of real numbers together with the binary operations $\oplus$ = maximum and $\otimes$ = product, similarly as the ordinary addition and multiplication operations are used in the classical linear algebra. The operations $\oplus$ and $\otimes$ are extended to matrices and vectors in a natural way. We should remark that the max-prod algebra is isomorphic to max-plus algebra, with the operations maximum and addition. The eigenvalue of a given max-plus or max-prod matrix and the eigenvectors can be efficiently described by considering cycles in specifically evaluated directed graphs.

Given preference matrix will be transformed by the tropical operations, until a steady state is reached. The eigenvector of the matrix then describes the steady state preferences and respects all preference relations contained in the original matrix.

Efficient algorithms for computing eigenvectors in the tropical algebra are described. The method is illustrated by numerical examples and compared with the linear algebra approach. The consistent and inconsistent cases are considered.

1 Introduction

Analytic Hierarchy Process (AHP) is a method developed for creating structured models of multi-criteria decision problems. The method helps to find an alternative which suits best the given needs of the deciding person. Analyzing the set of possible alternatives, the AHP method finds the one with the best rating, based on the structure of the problem and given preferences. Saaty formulated the principles of AHP in late 1970s [Saaty, 1980], and the method has been broadly studied and applied in many cases since the time, [Dytczak & Szklennik, 2011, Mls & Gavalec, 2009, Ramík & Perzina, 2010]. The method combines mathematical and psychological aspects, starting with defining the structure of the problem, then quantifying the relative preferences, computing the priorities and finally computing the evaluation of all considered alternatives. First of all, the multi-criteria decision problem is converted into a hierarchy of sub-problems and each of the sub-problems is then independently analyzed. The criteria of the sub-problems in the hierarchy may have very heterogeneous nature, they may be precisely or vaguely defined, with crisp or fuzzy parameters, formal or intuitive, etc. The relative preferences of heterogeneous criteria are then quantified by human decision-maker using the ability to compare various aspects of the problem. The decision maker systematically compares the criteria in pairs and quantifies the relative importance either by available data or by intuitive judgment. The relative preferences found by pairwise comparisons are then used to compute weights (priorities) for every part of the hierarchy model. The evaluation computed for all decision alternatives then shows their relative strength from the point of view of the entire problem. It is the advantage of AHP that even considerably diverse criteria can be used in the model, and that not only exact data but also human judgments can be applied to describe various aspects of the problem, [Saaty, 1994].

Formally, AHP is expressed by matrices and matrix operations are used to find and evaluate the best alternative:

Let $A_1, A_2, .., A_n$ be a set of variables. The quantified judgments on pairs of variables A_i, A_j are represented by an $n \times n$ matrix $A = (a_{ij})$, $i, j = 1, 2, ..., n$. Entries a_{ij} are defined as follows: if $a_{ij} = a$, then $a_{ji} = 1/a$; $a \neq 0$ for all i, j. As A_i is considered to be of equal relative intensity to itself, then $a_{ii} = 1$ for all i.

$$A = \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{pmatrix}$$

Then, relative rankings of variables will be obtained by computing an eigenvector X of the matrix A.

$$AX = \lambda X$$

Finally, by matrix multiplication of matrices of relative rankings of criteria and alternatives relative rankings matrix, total ranking of particular alternatives is obtained.

2 Max-Prod algebra

In this paper we work with preference matrices in the multiplicative form. It is natural, therefore, to process the preference matrix in the max-prod algebra. By max-prod algebra we understand a linear structure on a linearly ordered set $\mathcal{R}$ of real numbers together with the binary operations $\oplus$ = maximum and $\otimes$ = product, similarly as the ordinary addition and multiplication operations are used in the classical linear algebra. The operations $\oplus$ and $\otimes$ are extended to matrices and vectors in a natural way. We should remark that the max-prod algebra is isomorphic to max-plus algebra, with the operations maximum and addition. The eigenvalue of a given max-plus or max-prod matrix and the eigenvectors can be efficiently described by considering cycles in specifically evaluated directed graphs.

The eigenproblem in the max-prod algebra $(\mathcal{R}, \oplus, \otimes)$ can be formulated as follows. Given matrix $A \in \mathcal{R}^{n\times n}$, find $\lambda \in \mathcal{R}$ and $X \in \mathcal{R}^n$ such that

$$A \otimes X = \lambda \otimes X$$

It is a well-known fact that the eigenvalue λ can be computed as the maximal geometric cycle mean in the complete directed graph $G(A)$ with n nodes, in which the edges are evaluated by the matrix inputs, e.i. the weight of the edge (i, j) is $w(i, j) = a_{ij}$ for every i, j. If $C = (v_0, v_1, v_2, \ldots, v_k)$ with $v_0 = v_k$ is a cycle of length $k > 0$ in $G(A)$, then the weight of C is the value $w(C) = \prod_{i=1}^{k} w(v_{i-1}, w_i)$, and the geometric cycle mean $\overline{w}(C) = w(C)^{1/k}$, [Ramík & Korviny, 2010]. Hence $\lambda = \max_{C \text{ cycle in } G(A)} \overline{w}(C)$. The maximal geometric cycle mean can be computed in $O(n^3)$ time be slightly modified Karp's method, see [Karp, 1978].

When the eigenvalue λ has been computed, then we denote $B = A \otimes \lambda^{-1}$ and compute the matrix $B^\star = I \oplus B \oplus B^2 \oplus \ldots \oplus B^{n-1}$, which is called Kleene star. The columns of the Kleene star matrix with diagonal value 1 are the fundamental eigenvectors of B with eigenvalue value 1, and they are also the eigenvectors of the original matrix A with eigenvalue λ.

3 Matrices

Consistent matrix is a form of preference matrix where the following conditions are satisfied.

1. $a_{ii} = 1$
2. $a_{ij} = 1/a_{ji}$
3. $a_{ij} * a_{jk} = a_{ik}$

It should be remarked that Kleene star matrix $B^\star = (A \otimes \lambda^{-1})^\star$ is consistent. Conversely, if a preference matrix A is consistent, then it is equal to its Kleene star matrix, moreover, all eigenvectors are multiples of each other. Thus, for

a consistent matrix there exists only one fundamental eigenvector, which gives the priority vector, after normalization.

For an inconsistent matrix A the priority vector is found in the following steps.

1. Find the eigenvalue λ of matrix A. The eigenvalue is maximal geometric mean weight of cycle in the corresponding graph, computation is done by the modified Karp's algorithm
2. Converte A into matrix $B = A \otimes \lambda^{-1}$ to get the eigenvalue equal to one.
3. Compute the Kleene star matrix $B^\star = I \oplus B \oplus B^2 \oplus \ldots \oplus B^{n-1}$, a consistent matrix.
4. Normalize the fundamental eigenvector, any column from $B^\star$. The normalized column is the priority vector.

4 Example

The example was published in [Gavalec & Tomášková] and is based on an actual problem: how to choose a conference for presentation of research results. We prepare a 4 attributes and 3 alternatives, as is shown in a picture bellow.

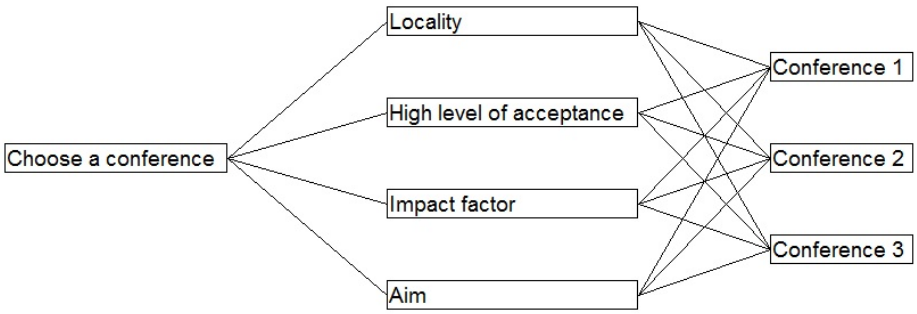

Figure 1: Hierarchy graph

The problem is situated as a way, how to choose a conference. We defined a four levels

- Locality - exotic, holiday areas, etc.,

- High level of acceptance,
- Impact factor - means the outputs of the conference, from nonindexed proceedings to journal with impact factor,
- Aim - if the conference is very close to my work.

We anonymised the conference, but there is a list of short information which will described the instance.

Conference 1 is located on very interesting places, in most cases oceanic places, the level of acceptance is almost 95%, the outputs have almost zero value for current scientific scoring system, the conferences are wide, with many scientific areas.

Conference 2 is located on interesting places, not necessary close to ocean, the level of acceptance is high in case of not so wide target scientific population. The outputs have moderate importance, but the level of papers and discussions on the conference allows you to prepare a journal paper. The aim is very close.

Conference 3 is located at normal places and it is easy to travel on it. The level of acceptance is higher. The outputs of the conference are indexed in required databases and the aim is close.

In the example are used two programs Criterium Decision Plus and our LibNOM working with the extremal algebra max-prod.

Both approaches use a pairwise comparisons, where the priorities will be derived from a series of measurements: pairwise comparisons involving all the nodes. The nodes at each level will be compared, two by two, with respect to their contribution to the nodes above them. The results of these comparisons will be entered into a matrix which is processed mathematically to derive the priorities for all the nodes on the level.

Intensity of Importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance

Locality matrix has a $\lambda = 1.1587$.
High level of acceptability matrix has a $\lambda = 1.3867$.
Impact factor matrix has a $\lambda = 1.3867$
Aim matrix has a $\lambda = 1.1587$.

$$\text{Loc} = \begin{pmatrix} 1 & 7 & 9 \\ 1/7 & 1 & 2 \\ 1/9 & 1/2 & 1 \end{pmatrix} \quad \text{HLA} = \begin{pmatrix} 1 & 9 & 8 \\ 1/9 & 1 & 1/3 \\ 1/8 & 3 & 1 \end{pmatrix}$$

$$\text{IF} = \begin{pmatrix} 1 & 1/9 & 1/8 \\ 9 & 1 & 3 \\ 8 & 1/3 & 1 \end{pmatrix} \quad \text{Aim} = \begin{pmatrix} 1 & 1/9 & 1/7 \\ 9 & 1 & 2 \\ 7 & 1/2 & 1 \end{pmatrix}$$

	Loc	HLA	IF	Aim	Criteria
Crit.Des.	$\begin{pmatrix} 0.793 \\ 0.131 \\ 0.076 \end{pmatrix}$	$\begin{pmatrix} 0.798 \\ 0.064 \\ 0.138 \end{pmatrix}$	$\begin{pmatrix} 0.052 \\ 0.648 \\ 0.300 \end{pmatrix}$	$\begin{pmatrix} 0.057 \\ 0.597 \\ 0.346 \end{pmatrix}$	$\begin{pmatrix} 0.051 \\ 0.282 \\ 0.526 \\ 0.141 \end{pmatrix}$
LibNOM	$\begin{pmatrix} 0.7928 \\ 0.1312 \\ 0.0760 \end{pmatrix}$	$\begin{pmatrix} 0.7978 \\ 0.0639 \\ 0.1383 \end{pmatrix}$	$\begin{pmatrix} 0.0519 \\ 0.6484 \\ 0.2997 \end{pmatrix}$	$\begin{pmatrix} 0.0572 \\ 0.5969 \\ 0.3458 \end{pmatrix}$	$\begin{pmatrix} 0.0524 \\ 0.2857 \\ 0.5191 \\ 0.1428 \end{pmatrix}$

5 Acknowledgements

The support of Czech Science Foundation GAČR #402/09/0405, project CZ.1.07 / 2.3.00 /20.0001 INKOV and Grant Agency of Excellence UHK FIM #2214 is kindly acknowledged.

References

[Brualdi & Ryser, 1991] Brualdi, R. A.,Ryser, H. J. (1991). *Combinatorial Matrix Theory*, in: Encyclopaedia of Mathematics and its Applications, vol. 39, Cambridge Univ. Press, Cambridge, 1991.

[Butkovič, 2010] Butkovič, P. (2010). *Max-linear Systems: Theory and Algorithms*, Springer Monographs in Mathematics, Springer-Verlag, 2010.

[Carlsson & Fullér, 1996] Carlsson, C., Fullér, R. (1996). *Fuzzy multiple criteria decision making: Recent developments*, Fuzzy Sets and Systems, Volume 78, Issue 2, 11 March 1996, Pages 139-153.

[Cuninghame-Green, 1976] Cuninghame-Green, R. A. (1976). *Minimax algebra*, Lecture Notes in Econom. and Math. Systems **166**, Springer-Verlag, Berlin, 1979.

[Dytczak & Szklennik, 2011] Dytczak, M., Szklennik, N. (2011). *Principles and applications of AHP/ANP - based multiple MCDA methods approach*, in: ISAHP 2011. Proceedings, Sorrento, 15-18 June 2011.

[Rashid et al., 2012] Rashid, I., Gavalec, M., Sergeev, S. (2012). *Eigenspace of a three-dimensional max-Lukasiewicz fuzzy matrix*, Kybernetika **48** (2012), 309–328.

[Gavalec et al., submitted] Gavalec, M., Rashid, I., Cimler, R. (submitted). *Eigenspace structure of a max-drast fuzzy matrix*, Fuzzy Sets and Systems (submitted).

[Gavalec & Tomášková] Gavalec, M., Tomášková, H. (submitted). *Preference matrices in tropical algebra*, ISAHP 2013 (submitted).

[Heidergott et al., 2006] Heidergott, B., Olsder, G.J., van der Woude, J. (2006). *Max Plus at Work*, Princeton University Press, Princeton, New Jersey, 2006.

[Karp, 1978] Karp, R.M. (1978). *A characterization of the minimum cycle mean in a digraph*, Discrete Math. **23** (1978), 309–311.

[Leung & Cao, 2000] Leung, L.C., Cao, D. (2000). *On consistency and ranking of alternatives in fuzzy AHP*, European Journal of Operational Research, Volume 124, Issue 1, 1 July 2000, Pages 102-113.

[Mls & Gavalec, 2009] Mls, K. and Gavalec, M. (2009). *Multi-criteria models in autonomous decision making systems.* In Proceedings of the 10th International Symposium on the Analytic Hierarchy/Network Process. pp 1-8, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.

[Olsder, 1991] Olsder, G. (1991). *Eigenvalues of dynamic max-min systems*, in Discrete Events Dynamic Systems 1, Kluwer Academic Publishers, 1991, 177-201.

[Ramík & Korviny, 2010] Ramík, J., Korviny, P. (2010). *Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean*, Internat. J. Fuzzy Sets and Systems 161(2010), 1604-1613.

[Ramík & Perzina, 2010] Ramík, J., Perzina, R. (2010). *A method for solving fuzzy multicriteria decision problems with dependent criteria*, Fuzzy Optimization and Decision Making, Springer Verlag,Vol.9, No.2, 123-141.

[Saaty, 1980] Saaty, T.L. (1980). *The analytic hierarchy process.* McGraw-Hill, New York.

[Saaty, 1994] Saaty, T.L. (1994). *Fundamentals of decision making.* Pittsburgh: RWS Publications.

[Saaty, 2003] Saaty, T.L. (2003). *Decision-making with the AHP: Why is the principal eigenvector necessary*, European Journal of Operational Research, Volume 145, Issue 1, 16 February 2003, Pages 85-91.

Decision Making Process in Agent-based Roundabout Simulation

Richard Cimler
Faculty of Informatics and Management
University of Hradec Králové
richard.cimler@uhk.cz

Abstract

The objective of the paper is to present the agent-based model of roundabout. The model is implemented in NetLogo - agent-based programming language with integrated modeling environment. Agents correspond to drivers. Each driver has got its own parameters and behavior. The behavior is based on multi-criteria decision making. Driver has to be careful and make many decisions during driving by the crossing or the roundabout. Decision of the driver can be expressed using the analytic hierarchy process.

There are few types of similar drivers however each driver has its own parameters. Ratio of the driver's types can be changed by the control elements and the impact of this change on the fluency of the traffic can be explored.

Research is based on a real traffic problem. In the city Hradec Kralove is a roundabout instead of the heavily used crossing planned to be built. Model of the roundabout is created due to the size and layout of the planned building. Traffic density can be changed by the control elements and it enables to explore different traffic situation and compare advantages and disadvantages of the current crossing with the planned roundabout.

1 Introduction

Agent-based simulations help to explore complex systems and in some cases also predict behavior of whole system after interferences. One of the problems which can be explored by agent-based simulation is traffic problem.[1] Our project is based on a real traffic problem in city Hradec Kralove in Czech Republic. Two roundabouts instead of the heavily used crossing are planned to be build. We would like to explore impacts of this change on the traffic situation in this location. Due to different types of experiments which we would like to conduct we have decided to create our own simulation instead of using one of the existing programs. Simulation is created in the programming environment NetLogo.[2]

Figure 1: Visualization of the planned system of two roundabouts

2 Agent-based model

Each car with driver is in our simulation represented by an agent. There are a lot of parameters of each car such as maximal speed of the car or acceleration as well as characteristics of the driver such as tendency to go fast. There are few different types of vehicles except normal cars there are for example trucks or ambulances. Each agent has destination point which is trying to reach by driving through the simulated area. Agents behave due to traffic rules. Logic of the driving is based on heading to the waypoints.

There are six types of waypoint:

- Entering the area near roundabout - agent should decrease its speed
- Start of roundabout - agent has to decide here whether enter the roundabout or wait
- End point of entrance to the roundabout - agent changes type of his movement due to algorithms created for moving on the roundabout.
- Exit - is the waypoint near the exit of the roundabout and contains information about destinations in the direction of this exit. If waypoint contains information about drivers destination, driver leaves the roundabout through this exit.
- Destination - this waypoint shows driver that he reached his destination. Statistics about this agent are saved and then is agent removed from the simulation.
- Tunel - This waypoint doesnt affects agents behavior. If the agent reaches this waypoint car starts to be invisible for other agents as well as for user. Car becomes visible as it reaches the other tunel mark.

Figure 2: Simulation

3 Driver's decision making

In a real life a lot of different factors affects drivers decision making process. Driver has to watch carefully his surroundings and due to the aquired information make a decision. Each driver can percive information in a different way. In an agent-based simulation each agent has its sensors which enables to aquire information from the surroundings. Advantage of the agent simulation is that new sensors can be added easily and each agent can process information and decide autonomously. We have decided to use analytic hierarchy process (AHP) in agents decision making process.

3.1 AHP

Principles of AHP were formulated by Saaty in late 1970s [3]. This method helps to create structured models of multi-criteria decision problems. Problem is divided into the sub problems and each sub problem is analyzed separately. There can be many criteria connected to the decision making process. To create ranking of these criteria pair-wise comparison is used. Every two criteria are compared and relative importance of each is set. Result values are in the comparison matrix A.

In AHP approach an nxn pair-wise comparison matrix A (criteria matrix) with positive elements a_{ij} is considered [4](Mls, 2013)

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ a_{21} & 1 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & 1 \end{pmatrix}$$

This matrix is reciprocal, so $a_{ij} = 1/a_{ji}$ for each $1 \leq i, j \leq n$. Normalised principal eigenvector of the matrix A is computed to get priorities of respective alternatives. We say that A is consistent, if $a_{ij} \times a_{jk} = a_{ik}$ for each $1 \leq i, j, k \leq n$.. If $a_{ij} \times a_{jk} \neq a_{ik}$ for some i, j, k, than A is said to be inconsistent. In AHP, it is assumed that $1/9 \leq a_{ij} \leq 9.$, for all $1 \leq i, j \leq n$. The inconsistency of A is measured by the consistency index CI_n defined as

$$CI_n = \frac{\lambda_{\max} - n}{n - 1}$$

where $\lambda_{\max}$ is the principle eigenvalue of A. If A is an nxn positive reciprocal matrix, then A is consistent if $CI_n = 0$. Inconsistency index is discussed also in [5] .

By computing an eigenvector X of the matrix A relative rankings of variables can be obtained.

$$AX = \lambda X$$

Final ranking of particular alternatives is computed as matrix multiplication of eigenvector of X and matrix describing possible alternatives.

Many factors affect driver's decision. In our previous basic model of the traffic only proximity to the other car on the roundabout arbitrates driver's actions. All other factors which in a real life influence driver's decision were neglected. Driver, as he reaches start of roundabout mark, has to decide if he wants to enter the roundabout or wait. There was a simple rule: if is distance to the other car (which is already on the roundabout and can cross driver's trajectory) smaller than a safe distance (set in the start of the model) driver waits and do not continue in his driving. Agent based model was extend by multi-criteria decision making.

Each driver has its own criteria matrix and in the moment, he is deciding about his next action, situation vector is created. Using the decision making process, described below, driver chooses the action which should be best for him. Our model is presented on the following example. Number of criteria can increase as well as situation vector can be more comprehensive.

4 Example

Simulation created in NetLogo allows us to obtain a lot of information about drivers status and its surroundings. Some of the information which is possible to obtain is number and speed of other cars, proximity to crossing, proximity to other car etc. Based on this information driver can decide about his next action.

4.1 Criteria matrix

Criteria matrix is based on the attributes of drivers. Criteria matrix consists of the 3 criteria - number of the criteria will increase while the model is extended.

Driver's character is divided into two parts. Driver can be patient or impatient and experienced or inexperienced - see fig.2.

Criteria matrix consists of:

- Proximity: In this criterion, reaction on the car which is already on the roundabout and can cross driver's trajectory is described. Cautious drivers tend to stop if the other car is approaching. In contrast to rude drivers who will in the same situation continue driving or increase their speed. Proximity is very important for inexperienced drivers.
- Rush: This criterion shows if the driver is in a hurry. This criterion is important for impatient drivers.
- Environment: Expresses caution of the drivers if weather conditions and conditions of the road are not good. For inexperienced drivers is this criterion important.

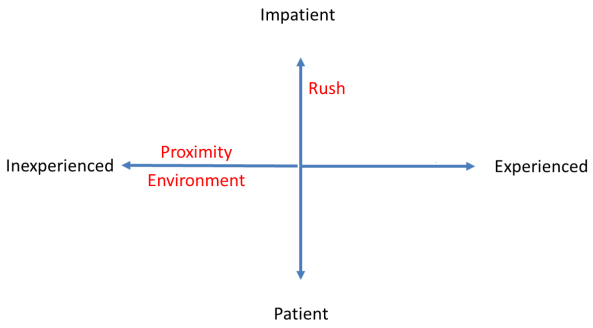

Figure 3: Driver's attributes

4.2 Situation vectors

Situation vector, written as [stop; go normal; go fast], represents current situation of the driver. There are three types of situation vectors in our example: environment vector, rush vector and proximity vector. Vector doesn't describe driver's character, it only describes his current situation and is recommendation how to behave.

4.2.1 Environment

At the begging of the simulation random environment status is generated. In our example environment status is static for the entire simulation. The future enhancements will include the possibility of modifying the environment status, which is going to enable other scenarios such as alternation of day and night or changing of the weather.

4.2.2 Proximity

Based on the proximity of the other car on the roundabout this vector is created. If the other car is far, driver can continue driving and enter the roundabout. If other car is near, driver should increase his speed and enter the roundabout fast or can also stop and wait.

4.2.3 Rush

As the agent is created, his rush vector is generated. This vector represents how this particular driver hurries. It doesn't tell anything about how driver behaves if he is in a hurry.

4.3 Final decisions

While driver approaching to the roundabout he has to choose from three possible actions. Decision depends on driver's character (criteria - comparison matrix) and the current situation on the roundabout. In previous section driver's criteria matrix and situation vector has been introduced. Possible actions are to continuing driving normally, stopping or increasing speed and enter quickly the roundabout. Result of the computation is final, driver will do the most recommended action.

4.3.1 Computation

Driver is entering to the simulated area and his criteria (comparison) matrix $A(nxn)$ and eigen-vector x, where the x_i respresents the input of $i-$th row, of this matrix is created. While is driver deciding about his next action, situation column vectors $v(i)$, where the i is a number of type of the situation, $i = 1$ to n, are created. The column inputs of result matrix B are created as: for every j from 1 to n and for every i from 1 to $n : B_j = v(i) * x_i$ hold true. Column vector $s(j) : j = 1..n$, for final decision f is $\sum$ for every i from 1 to n of b_{ij}.

5 Experiment

One of many possible settings of the model is shown in this experiment. We have decided to have two types of drivers. First type is patient and inexperienced driver and second type is impatient and experienced driver. For this example, criteria matrices have been chosen as following:

Patient and inexperienced

	Proximity	Rush	Environment
Proximity	1	8	2
Rush	1/8	1	1/7
Environment	1/2	7	1

Impatient and experienced

	Proximity	Rush	Environment
Proximity	1	2	2
Rush	1/2	1	2
Environment	1/2	1/2	1

Situation vector is created as [stop] [normal] [fast] (sum is equal to 1) in the lines of following matrices. Appropriateness of each action (stop, go normal, go fast) depends on the current situation. For example in table "Proximity vectors": if other cars proximity is moderate it is appropriate to stop and do not enter the roundabout or increase speed and enter it fast before other car can cross our trajectory. It is not appropriate to enter the roundabout normal speed because there is a bigger chance of crash.

Proximity vectors

	Stop	Normal	Fast
Far	0,05	0,8	0,15
Moderate	0,7	0	0,3
Small	1	0	0

Rush vectors

	Stop	Normal	Fast
Far	0	0,3	0,7
Moderate	0,2	0,4	0,4
Small	0.3	0,7	0

Proximity vectors

	Stop	Normal	Fast
Far	0	0,5	0,5
Moderate	0,3	0,5	0,2
Small	0,5	0,4	0,1

5.1 Results

Each setup was run for ten times and results are average values. Impact of the environment status has been tested on patient-inexperienced drivers and impatient-experienced drivers. Environment status has been set to bad, moderate and good. Rush status has been set to moderate for all of the drivers.

All drivers do not risk getting fast to the roundabout if the environment is not good. Experienced and impatient drivers made decision go fast in 5.5% of decision making situations in a good environment. Results of driver's decisions can be seen in the following table.

Drivers decisions

	Fast	Normal	Stop
Patient and inexperienced	0%	6%	94%
Impatient and experienced	7%	28%	65%

6 Conclusion

Agent based model of the roundabout containing agents using multi-criteria decision has been introduced. Model is prepared for the discussion about setting different types of drivers and their criteria matrices. Model enables to gain a lot of information from the surroundings of the agent which can be used for decision making. New features will be added such as a changing of day and night or changing of the weather which affects driving conditions. Condition of the vehicle can be also considered in the decision making process. Agent based simulation can also provide useful information about problematic locations and creation of the traffic jams on the crossings and roundabouts.

7 Acknowledgement

This work was supported by the project No. CZ.1.07/2.2.00/28.0327 Innovation and support of doctoral study program (INDOP), financed from EU and Czech Republic funds, GACR 402/09/0405, University Hradec Kralove specific research and program Excelence.

References

[1] Cimler, R., Kautzká, E., Olševičová, K., Gavalec, M. (2012) *Agent-based Model for Comparison of Aircraft Boarding Methods.* Mathematical Methods in Economy.

[2] Wilensky, U. (1999). *NetLogo* . http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

[3] Saaty, T.L. (1980). *The analytic hierarchy process.* McGraw-Hill, New York.

[4] Mls, K., Otenkov, T. (2013) *Analysis of Complex Decisional Situations in Companies with the Support of AHP Extension of Vroom-Yetton Contingency Model.* IFAC MIM '2013 International Conference, Sain Petersburg, Russia

[5] Ramk, J. and Korviny, P. (2009). *Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean.* Fuzzy Sets and Systems, 161, pp.1604-1613.

Location, Allocation and Capacity Design for Emergency Medical Centers

Young Dae Ko
Deloitte Analytics
Deloitte Anjin LLC
ydko@kaist.ac.kr

Byung Duk Song
Ind. and Systems Eng.
KAIST
bdsong@kaist.ac.kr

Hark Hwang
Ind. and Systems Eng.
KAIST
hhwang@kaist.ac.kr

Abstract

This paper studies the location, allocation and capacity design of emergency medical centers (EMC) in a given region under the closest assignment rule. It is assumed that the capability, initial capacity, and cost associated with unit increase in the capacity of treating various categories of treatable medical diseases are provided for each candidate medical center. Once a candidate center is selected, it will receive subsidies from the government to support the offering of medical services. It is further assumed that the number of patients occurring at each patient group node during a unit time is known along with the categories of their diseases. With the objective of minimizing the total subsidies paid, we select EMCs from among the candidates and also determine the desirable level of capacity in each category of the diseases subject to a minimum desired survival rate constraints. The CPLEX version 12.1 solver is used to obtain an optimal solution.

1 Introduction

1.1 Background

A medical emergency is an injury or illness that is acute and poses an immediate risk to a person's life or long term health. For emergencies starting outside of medical care, two key components of providing proper care are to summon the emergency medical services and to arrive at an emergency medical center where the necessary medical care is available. To facilitate this process, each country provides its own national emergency telephone number (c.f., 911 in the USA, 119 in Korea) that connects a caller to the appropriate local emergency service provider. Appropriate transportation, such as an ambulance, will be dispatched to deliver the emergency patient from the site of the medical emergency to an available emergency medical center. A flow chart depicting the Korean

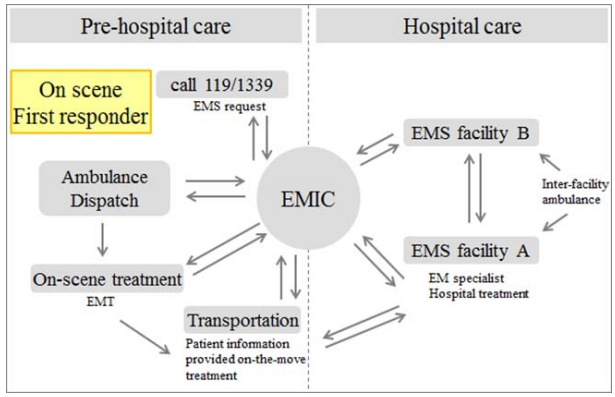

Figure 1: Flowchart of Emergency Medical Information Center

emergency medical procedure is provided in Figure 1; it is from the National Emergency Medical Center of Korea. Initially, the victim(s) or a first responder calls 119 to request emergency medical service. An ambulance will then be dispatched to the scene. When an ambulance arrives to the scene, on-scene treatment is first performed, followed by transportation of the patient(s) to an emergency medical service facility. During transport, information on the patient's condition may be communicated to the emergency medical information center. Inter-facility ambulance transport may be conducted if the patient(s) should be treated at different facilities. According to statistics from the Korean government, the mortality rate of emergency patients has increased continuously from 2009 through 2012. One of the major factors causing this trend is an increase in the duration of time between emergency events and time-of-treatment. The connection is simple: emergency medical conditions are time-critical events. Two major factors contribute to the increase in this time-to-treatment duration. First is an increase in the time from the request of an ambulance to the arrival of the vehicle. Second is an increase in the time from arrival of the ambulance to delivery of the patient to an appropriate emergency medical center. In this paper, we study the facility location problem in an effort to reduce the second factor. Transport time of patients is largely dictated by the distance of the nearest emergency medical center with the capability to treat the appropriate category of treatable medical disease and sufficient capacity of medical staff to do so. The capability and capacity varies by medical center. A mathematical model is developed with the objective of minimizing the amount of subsidies the government expends. We select emergency medical centers from among candidate centers and determine desirable level of capacity in each category of the diseases subject to minimum desired survival rate constraints and the closest capable facility assignment rule. We use CPLEX version 12.1 to obtain an optimal solution of the model. Simulation studies are conducted to assess the quality of the solution suggested by our deterministic model in the context of more realistic stochastic problem situations.

1.2 Literature Survey

Since emergency medical centers provide a critical life-saving service, they have been studied for several decades. The typical approaches to locate emergency medical centers include use of the set covering problem and the maximal covering problem. The set covering problem seeks to position the least number of facilities while covering all points of demand subject to certain distance or time constraints. In 1971, Toregas et al.[12] presented a linear programming set covering problem for locating emergency service facilities. Daskin and Stern [4] presented a hierarchical objective set covering problem for emergency medical services in order to find the minimum number of vehicles that are required to cover all demand areas while simultaneously maximizing multiple coverage. Church and Revelle [2] and White and Case [14] developed a maximal covering location problem that does not require full coverage to all demand points. Instead, the model seeks the maximal coverage with a given number of facilities. Another important metric to measure the effectiveness of facility location is the average (total) distance between demand points and facilities. To address this issue, the P-median location model has been applied for locating emergency medical centers. The P-median problem was introduced by Hakimi [6] and determines the locations of P facilities in order to minimize the average (total) travelling distance between demand points and facilities. Carbone [1] suggested a P-median model with the objective of minimizing the travelling distance between users and facilities such as medical or day care centers. Paluzzi [11] discussed and tested a P-median based heuristic location model for siting emergency service facilities in Carbondale. Jia et al. [7] analyzed the characteristics of large scale emergencies and proposed a P-median location model for emergency facilities. Several researchers have proposed facility location problems which guarantee high accessibility via closest assignment constraints. Closest assignment constraints force demands to be served by the nearest open facility. Church and Roberts [3] introduced a public sector location model to maximize the total public benefit which was assumed to be a function of the demand populations and their proximity to a facility. The demand nodes were assigned to a closest facility to obtain the highest public benefit. Recently, Kim and Kim [8] employed closest allocation constraints to determine the locations of public long term care facilities. They considered uncapacitated long term care facilities of a single type and tried to balance the numbers of patients assigned to the facilities. In addition to the concepts of coverage and travelling distance, an important measure for locating emergency medical centers is survival rate. Erkut et al. [5] introduced the concept of using a survival function to evaluate the performance of the covering facility location models with a focus on emergency medical service systems. Mclay and Mayorga [9] also proposed a methodology to evaluate the performance of emergency medical service systems using a survival probability function which is piecewise in the distance to the facility.

2 Mathematical Formulation

2.1 Problem description

We now describe the nature of the problem in this study. In a given region, there are I patient group nodes each exhibiting up to K categories of treatable emergency diseases. There are J candidate medical centers with given initial capacity and cost associated with unit increase of capacity in treating various categories of treatable medical diseases. The capacity of each candidate facility and unit capacity incremental cost are not necessarily identical. From among these facilities, we will determine which will be declared as government sponsored emergency medical centers (EMCs) along with their desired capacities. The selected centers will receive subsidies from the government in order to keep the medical services competitive. The required amount of subsidies may be different among the centers and affected by the amount of capacity increased. It is further assumed that the number of patients occurring at each patient group node during a unit time is known along with the categories of their diseases. The survival rate of patients is expressed as a function of patient transportation time and category of disease. Through the development of a mathematical model, we will select from among the candidate centers which are to be declared as government sponsored. The objective function is to minimize the government subsidies expended.

2.2 Assumptions

1. J existing hospitals are considered as candidates to be declared as government sponsored EMCs. Their locations are known.
2. There are K types of treatable medical diseases that each emergency medical center may have the capacity to serve.
3. There are I patient group nodes.
4. Initial capacity of each candidate center to serve a treatable medical disease (in units of man-hours) is known and they are not necessarily identical among candidate centers. The capacity can be expanded to a certain extent by an additional amount of government subsidies. In this case, unit capacity expansion cost is known which may not be the same among candidate centers and disease categories.
5. The man-hour required to treat an incoming emergency patient is dependent on the disease category and capability of candidate center.
6. The expected number of emergency patients with disease k at patient group node i can be expressed as the product of its population (p_i) and the occurrence ratio (o_{ik}) per unit time of emergency disease k.
7. Patients with disease k at node i are served by the nearest possible emergency medical center with the capacity to do so.

8. The elapsed time from the request of an ambulance to the arrival of the vehicle is ignored.

9. A minimum target survival rate is prescribed for patients with disease k.

2.3 Notations

- i : Index for patient group node, i = 1,2,...,I
- j : Index for candidate medical center, j = 1,2,...,J
- k : Index for category of treatable medical disease, k =1,2,...,K
- c_j : Amount of governmental subsidy required for maintaining initial capacity by candidate center j
- c_{jk} : Amount of governmental subsidy required by candidate center j for a unit capacity expansion to treat disease k
- L_{jk} : Capacity in [man-hour] of emergency medical center j for the treatment of patients with category k disease
- d_{ij} : Distance between patient group node i and medical center j
- h_{jk} : Treatment time required in [man-hour] for a patient with category k disease at medical center j
- a_{ik} : Number of patients with category k disease in patient group node j = p_i x o_{ik}
- $f_k(d_{ij})$: Survival rate of group node i patient with category k disease when served by candidate medical center j
- b_{jk} : Initial capacity in [man-hour] for treating emergency patient with category k disease at candidate medical center j
- e_{jk} : Conversion factor: [man-hour] associated with one unit increase of the treatment capacity for category k disease at candidate center j
- lu_{jk} : Upper limit on the capacity expansion units at candidate medical center j for category k disease
- sr_{ijk} : Survival rate of patient with category k disease in patient group node i if the patient is treated in emergency medical center j
- sr_k : Minimum desired level of survival rate of category k patient
- t_{jk} : Treatment capability of patient with category k disease at medical center j. It will be equal to 1 if candidate medical center j can treat the patient. Otherwise, 0.

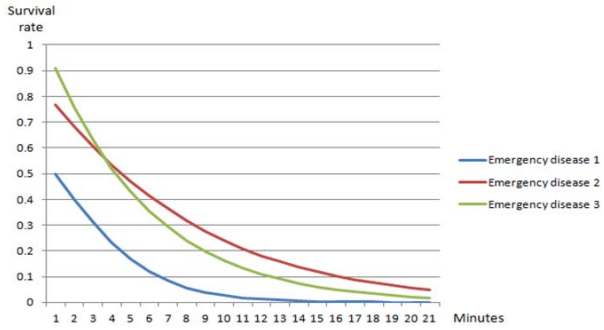

Figure 2: Survival rate functions for each disease type

- y_j : Location decision variable. It will be equal to 1 if candidate medical center in location j is selected as candidate medical center. Otherwise, 0.
- x_{ijk} : Allocation decision variable. It will be equal to 1 if patients with category k disease in location i are served by candidate medical center j. Otherwise, 0.
- u_{jk} : Number of capacity expansion units at candidate medical center j for category k disease (integer decision variables)
- MU : Monetary Unit

2.4 Survival rate function

Valenzuela et al. [13] developed a survival function to predict human life as a function of two variables: collapse to cardiopulmonary resuscitation and collapse to defibrillation intervals. In this study, we use the same general form as their model to express the survival rate as a function of category of emergency disease and transportation time of patient from patient group node i to the nearest medical center j. Let d_{ij} and v_{ij} denote the distance and average velocity of ambulance from location i to j, respectively. The survival rate function adopted in this study is

$$f_k(d_{ij}) = (a_k + e^{b_k \cdot d_{ij}/v_{ij}})^{-1} \tag{1}$$

Note that a_k characterizes the survival rate of patients with category k disease and is chosen such that $1/(a_k+1)$ is the probability of survival if they receive immediate treatment, i.e., zero distance to the medical facility. The parameter b_k characterizes the decreasing survival rate of patients with disease k as time passes. Figure 2 depicts the survival rate functions; they are dependent on the transportation time of the patient as well as the type of disease.

2.5 Mathematical formulation

$$Minimize \qquad \sum_{j=1}^{J}(c_j \cdot y_j) + \sum_{j=1}^{J}\sum_{k=1}^{K}(c_{jk} \cdot u_{jk}) \tag{2}$$

Subject to

$$\sum_{j=1}^{J} x_{ijk} = 1 \qquad for\ all\ i,\ k \tag{3}$$

$$x_{ijk} \leq y_j \cdot t_{jk} \qquad for\ all\ i,\ j,\ k \tag{4}$$

$$\sum_{j=1}^{J} d_{ij} \cdot x_{ijk} \leq d_{ij} + M(1 - y_j \cdot t_{jk}) \qquad for\ all\ i,\ j,\ k \tag{5}$$

$$\sum_{i=1}^{I} a_{ik} \cdot h_{jk} \cdot x_{ijk} \leq L_{jk} \qquad for\ all\ j,\ k \tag{6}$$

$$b_{jk} + e_{jk} \cdot u_{jk} \geq L_{jk} \qquad for\ all\ j,\ k \tag{7}$$

$$u_{jk} \leq lu_{jk} \qquad for\ all\ j,\ k \tag{8}$$

$$sr_{ijk} = f_k(d_{ij}) \qquad for\ all\ i,\ j,\ k \tag{9}$$

$$\sum_{j=1}^{J} sr_{ijk} \cdot x_{ijk} \geq sr_k \qquad for\ all\ i,\ k \tag{10}$$

$$y_j \in [0, 1] \qquad for\ all\ j \tag{11}$$

$$x_{ijk} \in [0, 1] \qquad for\ all\ i,\ j,\ k \tag{12}$$

$$u_{jk}\ is\ nonnegative\ integer \qquad for\ all\ j,\ k \tag{13}$$

The objective function (2) of is to minimize the government's subsidy to emergency medical centers. It has two components, one for maintaining initial capacities of emergency medical centers selected from among candidate medical centers and the other for the compensation of capacity expansion encouraged. Constraint (3) ensures that the patient group with category k disease in node i can be assigned to only one emergency medical center (a patient group cannot be split among centers). Constraint (4) ensures that patient group with category k disease in node i can receive medical service from emergency medical center j only if it has the capability to treat disease category k. Constraint (5) requires that each patient group be assigned to the closest open emergency medical

center. Constraint (6) sees to it that the capacity of each emergency medical center has to be sufficient enough to treat patients with disease category k coming from the assigned patient nodes. Constraints (7) and (8) are for capacity expansion. According to Constraint (7), the actual capacity of an emergency medical center consists of an initial capacity and expanded capacity encouraged by the government through subsidy and it has to be larger than the required capacity. Constraint (8) represents the upper limit on the capacity expansion units at candidate medical center j for category k disease. Constraints (9) and (10) are related with the survival rate. Constraint (9) determines the survival rate of patient group i with category k disease using the survival rate function discussed in Section 2.4. Constraint (10) guarantees a minimum desired survival rate of patients with category k disease. Finally, constraints (11) and (12) restrict our location and allocation decision variables to be binary variables and constraint (13) restricts capacity extension decision variables to be nonnegative integer. To find an optimal solution of the proposed mathematical model, we used CPLEX version 12.1. Since the model is an integer program, CPLEX can generate a globally optimal solution. We carried out simulation studies via Excel by generating stochastic occurrences of emergency patients. Based on the results of the simulation, the solutions from CPLEX are evaluated to determine whether or not the minimum desired level of survival rate at the nearest emergency center can be satisfied. Also, the effects of stochastic data on the system performance measures are examined.

3 Numerical Examples

The proposed model was tested with randomly generated data for 20 patient demand nodes and 10 candidate medical centers in a (10km, 10km) area. We consider K=3 categories of diseases. The number of patients with type k disease was obtained by multiplying the population of each patient group node by the occurrence rate of each type of disease, ($o_{i1} = 0.005$, $o_{i2} = 0.008$, $o_{i3} = 0.012$) for each i. For the survival rate functions, the constants (a_1, a_2, a_3) and (b_1, b_2, b_3) were set to (1, 0.3, 0.1) and (0.4, 0.2, 0.15), respectively. The transportation velocity, v_{ij}, was set to 0.5km/minute and the minimum desired levels of survival rate of category (sr_1, sr_2, sr_3) were set to (0.071, 0.212, 0.307). The location, population and number of patients at each patient group node were given as shown at Table (1).

Table 2 lists additional system parameter values, e_{jk}, lu_{jk} and h_{jk}.

Also, the location, capacity and capability, unit capacity expansion cost and amount of basic government subsidies required by each candidate medical center are listed at Table 3.

With the above data, the mathematical model was solved using CPLEX 12.1. The numerical experiments were conducted on an Intel(R) Core(TM)2 Quad CPU Q8400, 2.66GHz and 4.00GB RAM personal computer. The optimal solution from CPLEX located emergency medical centers at candidate centers 1, 3, 5, 7 and 9 with the government subsidy of 82,610 MU. It consists

Table 1: The Parameter values for patient group nodes

P. g. node	Loc. x (km)	Loc. y (km)	Population [person]	Number of patients[person]		
				Disease 1	Disease 2	Disease 3
1	3.37	9.13	1,048	5	8	13
2	7.58	8.43	895	4	7	11
3	8.36	7.18	1,124	6	9	13
4	2.06	8.21	1,382	7	11	17
5	3.37	3.02	950	5	8	11
6	8.80	4.23	961	5	8	12
7	4.60	6.02	806	4	6	10
8	8.48	5.75	1,210	6	10	15
9	8.53	6.82	687	3	5	8
10	3.63	3.31	1,464	7	12	18
11	3.02	8.65	532	3	4	6
12	1.35	4.17	1,436	7	11	17
13	4.89	5.55	1,239	6	10	15
14	2.14	3.53	682	3	5	8
15	2.82	5.58	1,005	5	8	12
16	5.40	8.68	1,025	5	8	12
17	6.44	2.04	1,038	5	8	12
18	3.25	8.06	1,198	6	10	14
19	1.66	4.27	1,094	5	9	13
20	2.88	5.42	828	4	7	10

of 74,000 MU for keeping the initial capacities of emergency medical centers selected and 8,610 MU for the capacity expansion. To satisfy the patient demand, the capacity for the treatment of category 2 disease at emergency medical center 3 and the capacity for the treatment of categories 2 and 3 diseases at emergency medical center 5 are increased by 1unit, 1unit and 2 units, respectively. And at emergency medical center 5 the amount of idle capacities for categories 2 and 3 diseases are relatively low with 18 and 6.1 units, respectively. The result of the numerical example is shown at Figure 3 and Table 4.

4 Conclusions

To provide a reference model for the location design problem of emergency medical centers, this paper proposed a mathematical model reflecting the real operating situation of emergency medical systems. In the model, we considered the category of treatable medical diseases of each candidate medical center and its capability. In reality, selected medical centers are provided with a certain amount of subsidies from the government in order to keep the medical services competitive. Also, we assumed that the government can encourage emergency

Table 2: The parameter values for e_{jk}, lu_{jk} and h_{jk}

Candidate	e_{jk}			lu_{jk}			h_{jk}		
med. center	k=1	k=2	k=3	k=1	k=2	k=3	k=1	k=2	k=3
1	24	24	24	1	1	3	1.8	1.5	2.0
2	24	0	24	1	0	1	1.0	1.6	1.4
3	0	24	24	0	2	1	1.5	1.6	1.8
4	24	24	0	1	2	0	1.4	1.9	1.2
5	24	24	24	1	1	2	1.2	1.5	1.7
6	24	24	24	3	1	1	1.4	1.3	1.8
7	24	24	24	1	2	2	1.4	2.0	1.3
8	24	24	0	1	3	0	1.6	1.6	1.6
9	24	0	24	1	0	3	1.2	1.6	1.1
10	0	24	24	0	3	1	1.4	1.9	1.7

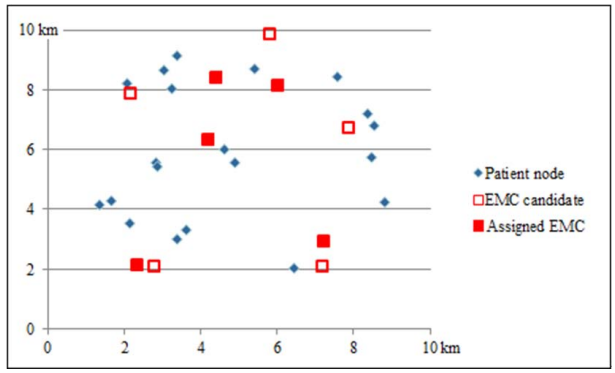

Figure 3: Locations for EMC

medical center to increase treatment capacity to some extent through the government subsidy. In the model patients are transported to the closest emergency medical center where they can receive a required medical treatment. In addition, the concept of minimum desired survival rate was considered. The proposed mathematical model was solved using CPLEX. We hope that our research results can be useful as guidance for government officials in charge of the design of emergency medical system or other similar systems. At the present time, we are studying simulation procedures to assess the performance of our deterministic models in a stochastic context.

5 Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Edu-

Table 3: The parameter values for candidate medical centers

id	Location (km)		Initial/Maximum capacity [man * hour] and (unit expansion cost[MU])			Subs. for init. capac. [MU]
	x	y	Category 1	Category 2	Category 3	
1	6.02	8.11	72/96(1,300)	144/168(2,250)	144/216(1,800)	20,000
2	2.17	7.85	48/72(1,200)	N/A	120/144(1,700)	15,000
3	4.41	8.39	N/A	72/120(2,550)	96/120(1,900)	11,000
4	7.21	2.10	48/72(1,400)	48/96(2,700)	N/A	10,000
5	2.34	2.13	96/120(1,450)	72/96(2,560)	72/120(1,750)	15,000
6	2.81	2.10	24/96(1,380)	120/144(2,800)	96/120(1,650)	25,000
7	7.22	2.90	72/96(1,270)	120/168(2,670)	144/192(1,500)	16,000
8	5.82	9.83	48/72(1,390)	96/168(3,050)	N/A	8,000
9	4.19	6.32	72/96(1,540)	N/A	144/216(1,860)	12,000
10	7.87	6.70	N/A	120/192(2,390)	144/168(2,050)	13,000

cation, Science and Technology (2012-009448).

References

[1] Carbone, R. (1974), Public facility location under stochastic demand, *INFOR*, 12: 261-270.

[2] Church, R.L., ReVelle, C. (1974), The maximal covering location problem, *Papers of the Regional Science Association*, 32: 101-118.

[3] Church, R.L., Roberts, K.L. (1983), Generalized coverage models and public facility location, *Papers of the Regional Science Association*, 53: 117-135.

[4] Daskin, M.S., Stern, E.H. (1981), A hierarchical objective set covering model for emergency medical service vehicle deployment, *Transportation Science*, 15: 137-152.

[5] Erkut, E., Ingolfsson, A., Erdogan, G. (2008), Ambulance location for maximum survival, *Naval Research Logistics*, 55: 42-58.

[6] Hakimi, S.L. (1964), Optimum locations of switching centers and the absolute centers and medians of a graph, *Operations Research*, 12: 450-459.

[7] Jia, H., Ordonez, F., Dessouky, M. (2007), A modeling framework for facility location of medical services for large-scale emergencies, *IIE Transactions*, 39: 41-55.

[8] Kim, D.G., Kim, Y.D. (2010), A branch and bound algorithm for determining locations of long-term care facilities, *European Journal of Operational Research*, 206: 168-177.

Table 4: Emergency medical center assigned for each patient group node and category of disease

Patient group node	No. of assigned emergency medical center		
	Category 1	Category 2	Category 3
1	1	3	3
2	1	1	1
3	1	1	1
4	9	3	3
5	5	5	5
6	7	7	7
7	9	3	9
8	7	7	7
9	1	1	1
10	5	5	5
11	9	3	3
12	5	5	5
13	9	1	9
14	5	5	5
15	9	3	9
16	1	1	1
17	7	7	7
18	9	3	3
19	5	5	5
20	9	5	9

[9] McLay, L.A., Mayorga, M.E. (2010), Evaluating emergency medical service performance measures, *Health Care Management Science*, 13: 124-136.

[10] National Emergency Medical Center: NEMC Home Page, http://nemc.or.kr/.

[11] Paluzzi, M. (2004), Testing a heuristic P-median location allocation model for siting emergency service facilities, *The Annual Meeting of the Association of American Geographers*, Philadelphia, PA.

[12] Toregas, C., Swain, R., ReVelle, C., Bergman, L. (1971), The location of emergency service facility, *Operations Research*, 19: 1363-1373.

[13] Valenzuela, T.D., Roe, D.J., Cretin, S., Spaite, D.W., Larsen, M.P. (1997), Estimating effectiveness of cardiac arrest interventions a logistic regression survival model, *Circulation*, 96: 3308-3313.

[14] White, J., Case, K. (1974), On covering problems and the central facility location problem, *Geographical Analysis*, 81: 281-293.

Credal Networks and Compositional Models: Preliminary Considerations

Jiřina Vejnarová
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
vejnar@utia.cas.cz

Abstract

In the paper we present basic concepts concerning credal networks and compositional models for credal sets and describe the problem of imprecision increase in the first type of these models.

Keywords: Credal sets, credal networks, compositional models, strong independence.

1 Introduction

The most widely used models managing uncertainty and multidimensionality are, at present, so-called *probabilistic graphical Markov models*. The problem of multidimensionality is solved in these models with the help of the concept of conditional independence, which enables factorization of a multidimensional probability distribution into small parts (marginals, conditionals or just factors). Among them, the most popular are Bayesian networks. Therefore, it is not very surprising, that analogous models have been studied also in several theories of imprecise probability [1, 2, 3].

Credal networks are a generalization of Bayesian networks, able to deal with imprecision. Compositional models for credal sets, on the other hand, are intended to be generalization of compositional models for precise probabilities [6, 7, 8]. As the equivalence between Bayesian networks and precise compositional models is well known, it seems quite natural to ask a similar question also in this, more general case.

Compositional models were introduced also in possibility theory [12, 13] (here the models are parameterized by a continuous t-norm) and a few years ago also in evidence theory [9, 10]. In all these frameworks the original idea is kept, but there exist some slight differences among them.

Although Bayesian networks and compositional models represent the same class of distributions, they do not make it in the same way. Bayesian networks

use *conditional distributions* whereas compositional models consist of *unconditional distributions*. Naturally, both types of models contain the same information but while some marginal distributions are explicitly expressed in compositional models, it may happen that their computation from a corresponding Bayesian network is rather computationally expensive. Therefore it appears that some of computational procedures designed for compositional models are (algorithmically) simpler than their Bayesian network counterparts.

Furthermore, the research concerning relationship between compositional models in evidence theory and evidential networks [14] revealed probably a more important thing. Even though any evidential network (with proper conditioning rule and conditional independence concept) can be expressed as a compositional model, if we do it in the opposite way and transform a compositional model into an evidential network, we may realize, that the model is more imprecise than the original one. It is caused by the fact that conditioning increases imprecision.

The goal of this paper is twofold. First, we want to show that the operator of composition can also be defined for credal sets (at least under specific conditions). Second, we want to argue that it is worth-developing, as conditioning in the framework of credal sets also increases imprecision.

The contribution is organized as follows. In Section 2 we summarize basic concepts and notation. Definition of the operator of composition is introduced in Section 3, where also its basic properties can be found. Finally, in Section 4 we recall the concept of credal networks and demonstrate how conditioning increases imprecision of the resulting model.

2 Basic Concepts and Notation

In this section we will recall basic concepts and notation necessary for understanding the contribution.

2.1 Variables and Distributions

For an index set $N = \{1, 2, \ldots, n\}$ let $\{X_i\}_{i \in N}$ be a system of variables, each X_i having its values in a finite set $\mathbf{X}_i$ and $\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n$ be the Cartesian product of these sets.

In this paper we will deal with groups of variables on its subspaces. Let us note that X_K will denote a group of variables $\{X_i\}_{i \in K}$ with values in

$$\mathbf{X}_K = \bigtimes_{i \in K} \mathbf{X}_i$$

throughout the paper.

Having two probability distributions P_1 and P_2 of X_K we say that P_1 is *absolutely continuous* with respect to P_2 (and denote $P_1 \ll P_2$) if for any $x_K \in \mathbf{X}_K$

$$P_2(x_K) = 0 \Longrightarrow P_1(x_K) = 0.$$

This concept plays an important role in the definition of the operator of composition.

2.2 Credal Sets

A *credal set* $\mathcal{M}(X_K)$ about a group of variables X_K is defined as a closed convex set of probability measures about the values of this variable.

In order to simplify the expression of operations with credal sets, it is often considered [11] that a credal set is the set of probability distributions associated to the probability measures in it. Under such consideration a credal set can be expressed as a *convex hull* of its extreme distributions

$$\mathcal{M}(X_K) = \mathrm{CH}\{\mathrm{ext}(\mathcal{M}(X_K))\}.$$

Consider a credal set about X_K, i.e. $\mathcal{M}(X_K)$. For each $L \subset K$ its *marginal credal set* $\mathcal{M}(X_L)$ is obtained by element-wise marginalization, i.e.

$$\mathcal{M}(X_L) = \mathrm{CH}\{P^{\downarrow L} : P \in \mathrm{ext}(\mathcal{M}(X_K))\}, \tag{1}$$

where $P^{\downarrow L}$ denotes the marginal distribution of P on $\mathbf{X}_L$.

Having two credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ about X_K and X_L, respectively (assuming that $K, L \subseteq N$), we say that these credal sets are *projective* if their marginals about common variables coincide, i.e. if

$$\mathcal{M}_1(X_{K\cap L}) = \mathcal{M}_2(X_{K\cap L}).$$

Let us note that if K and L are disjoint, then $\mathcal{M}_1$ and $\mathcal{M}_2$ are projective, as $\mathcal{M}(X_\emptyset) = 1$.

Conditional credal sets are obtained from the joint ones by point-wise conditioning of the extreme points and subsequent linear combination of the resulting conditional distributions. More formally: Let $\mathcal{M}(X_I X_J)$ be a credal set about (groups of) variables $X_I X_J$. Then for any $x_J \in \mathbf{X}_J$

$$\mathcal{M}(X_I|x_J) = \mathrm{CH}\{P(X_I|x_J) : P \in \mathrm{ext}(\mathcal{M}(X_I X_J))\}, \tag{2}$$

is a *conditional credal set* about X_I given $X_J = x_J$.

2.3 Strong Independence

Among numerous definitions of independence for credal sets [4] we have chosen strong independence, as it seems to be the most appropriate for multidimensional models.

We say that (groups of) variables X_K and X_L (K and L disjoint) are *strongly independent* with respect to $\mathcal{M}(X_{K\cup L})$ iff (in terms of probability distributions)

$$\mathcal{M}(X_{K\cup L}) \quad = \quad \mathrm{CH}\{P_1 \cdot P_2 : P_1 \in ext(\mathcal{M}(X_K)), P_2 \in ext(\mathcal{M}(X_L))\}. \tag{3}$$

Again, there exist several generalizations of this notion to conditional independence, see e.g. [11], but since the following definition is suggested by the authors as the most appropriate for the marginal problem, it seems to be a suitable concept also in our case, since the operator of composition can also be

used as a tool for solution of a marginal problem, as shown (in the framework of possibility theory) e.g. in [13].

Given three groups of variables X_K, X_L and X_M (K, L, M be mutually disjoint subsets of N, such that K and L are nonempty), we say analogous[1] to [11] that X_K and X_L are *conditionally independent on the distribution* given X_M under global set $\mathcal{M}(X_{K\cup L\cup M})$ (in symbols $K \perp\!\!\!\perp L|M$) iff

$$\begin{aligned}\mathcal{M}(X_{K\cup L\cup M}) \;=\; \mathrm{CH}\{(P_1 \cdot P_2)/P_1^{\downarrow M} : P_1 \in ext(\mathcal{M}(X_{K\cup M})),\\ P_2 \in ext(\mathcal{M}(X_{L\cup M})), P_1^{\downarrow M} = P_2^{\downarrow M}\}\,. \qquad (4)\end{aligned}$$

This definition is a generalization of stochastic conditional independence: if $\mathcal{M}(X_{K\cup L\cup M})$ is a singleton, then also $\mathcal{M}(X_{K\cup M})$ and $\mathcal{M}(X_{L\cup M})$ are (projective) singletons and the definition collapses into definition of stochastic conditional independence.

3 Operator of Composition and Its Properties

Now, let us start considering how to define composition of two credal sets. Consider two index sets $K, L \subset N$. At this moment we do not put any restrictions on K and L; they may be but need not be disjoint, one may be subset of the other.

In order to enable the reader the understanding of this concept, let us first present the definition of composition for precise probabilities [6]. Let P_1 and P_2 be two probability distributions of (groups of) variables X_K and X_L. Then

$$(P_1 \triangleright P_2)(X_{K\cup L}) = \frac{P_1(X_K) \cdot P_2(X_L)}{P_2(X_{K\cap L})}, \qquad (5)$$

whenever $P_1(X_{K\cap L}) \ll P_2((X_{K\cap L})$. Otherwise, it remains undefined.

Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be credal sets about X_K and X_L, respectively. Our goal is to define a new credal set, denoted by $\mathcal{M}_1 \triangleright \mathcal{M}_2$, which will be about $X_{K\cup L}$ and will contain all of the information contained in $\mathcal{M}_1$ and as much as possible of information of $\mathcal{M}_2$.

The required properties are met by Definition 1[2] in [15]. However, the definition exhibits a kind of discontinuity and should be reconsidered. In this paper we will deal only with the composition of projective credal sets.

Definition 1 For two projective credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ about X_K and X_L, a *composition* $\mathcal{M}_1 \triangleright \mathcal{M}_2$ is defined by the following expression:

$$\begin{aligned}(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L}) \;=\; \mathrm{CH}\{(P_1 \cdot P_2)/P_2^{\downarrow K\cap L} : P_1 \in ext(\mathcal{M}_1(X_K)),\\ P_2 \in ext(\mathcal{M}_2(X_L)), P_1^{\downarrow K\cap L} = P_2^{\downarrow K\cap L}\}.\end{aligned}$$

[1] Let us note that our definition differs somehow from that presented in [11]: the authors require point-wise satisfaction in (3) and (4), which leads to non-convexity. In [5] this type of independence is called *complete*.

[2] Let us note that the definition was based on Moral's concept of conditional independence with relaxing convexity.

The following lemma presents basic properties possessed by this operator of composition.

Lemma 1 *For two projective credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ about X_K and X_L, respectively, the following properties hold true:*

(i) $\mathcal{M}_1 \triangleright \mathcal{M}_2$ *is a credal set about* $X_{K\cup L}$.

(ii) $\mathcal{M}_1 \triangleright \mathcal{M}_2 = \mathcal{M}_2 \triangleright \mathcal{M}_1$.

Proof.

(i) To prove that $\mathcal{M}_1 \triangleright \mathcal{M}_2$ is a credal set about $X_{K\cup L}$ it is enough to show that any $P \in \{ext(\mathcal{M}_1 \triangleright \mathcal{M}_2)\}$ is a probability distribution on $\mathbf{X}_{K\cup L}$, as the convexity of $\mathcal{M}_1 \triangleright \mathcal{M}_2$ is evident. But it is obvious, as any $P \in \{ext(\mathcal{M}_1 \triangleright \mathcal{M}_2)\}$ is obtained via formula for composition of probability distributions (5).

(ii) For any distribution P of $\{ext(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L})\}$ there exist $P_1 \in \{ext(\mathcal{M}_1(X_K))\}$ and $P_2 \in \{ext(\mathcal{M}_2(X_L))\}$ such that $P_1^{\downarrow K\cap L} = P_2^{\downarrow K\cap L}$ and $P = (P_1 \cdot P_2)/P_2^{\downarrow K\cap L}$. But simultaneously (due to projectivity) $P = (P_1 \cdot P_2)/P_1^{\downarrow K\cap L}$, which is an element of $(\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cup L})$. Hence

$$(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_{K\cup L}) = (\mathcal{M}_2 \triangleright \mathcal{M}_1)(X_{K\cup L}),$$

as desired. □

Let us now illustrate the application of the operator of composition and its properties by two examples. The first shows what happens when $K \cap L = \emptyset$.

Let use note the all variables in the examples in this paper are binary.

Example 1 Let

$$\mathcal{M}_1(X_1) = \mathrm{CH}\{[0.2, 0.8], [0.7, 0.3]\}$$

and

$$\mathcal{M}_2(X_2) = \mathrm{CH}\{[0.6, 0.4], [1, 0]\}$$

be two credal sets about X_1 and X_2, respectively. Then, as mentioned above, $\mathcal{M}_1(X_1)$ and $\mathcal{M}_2(X_2)$ are projective, and therefore $\mathcal{M}_1 \triangleright \mathcal{M}_2$ is obtained via Definition 1:

$$\begin{aligned}(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_1X_2) \quad = \quad & \mathrm{CH}\{[0.12, 0.08, 0.48, 0.32], [0.2, 0, 0.8, 0],\\ & [0.42, 0.28, 0.18, 0.12], [0.7, 0, 0.3, 0]\}.\end{aligned}$$

It is evident, that not every element of $\mathcal{M}_1 \triangleright \mathcal{M}_2$ can be expressed as a product of its marginals, as e.g.

$$\begin{aligned}[0.41, 0.04, 0.39, 0.16] \quad \in \quad & \mathrm{CH}\{[0.12, 0.08, 0.48, 0.32], [0.2, 0, 0.8, 0],\\ & [0.42, 0.28, 0.18, 0.12], [0.7, 0, 0.3, 0]\},\end{aligned}$$

but $[0.41, 0.04, 0.39, 0.16] \notin \{P_1 \cdot P_2 : P_1 \in \mathcal{M}(X_K), P_2 \in \mathcal{M}(X_L)\}$. ♢

The following example is devoted to the case, when $K \cap L \neq \emptyset$.

Example 2 Let

$$\begin{aligned}\mathcal{M}_1(X_1X_2) &= \mathrm{CH}\{[0.2, 0.2, 0, 0.6], [0.1, 0.4, 0.1, 0.4],\\ &\qquad [0.25, 0.25, 0.25, 0.25], [0.2, 0.3, 0.3, 0.2]\}.\end{aligned}$$

be a credal set about variables X_1X_2 and

$$\begin{aligned}\mathcal{M}_2(X_2X_3) = \mathrm{CH}\{&[0.2, 0, 0.3, 0.5], [0, 0.2, 0, 0.8],\\ &[0.5, 0, 0.5, 0], [0.2, 0.3, 0.2, 0.3]\},\end{aligned}$$

be a credal set about variables X_2X_3. These two credal sets are projective, as

$$\mathcal{M}_1(X_2) = \mathrm{CH}\{[0.2, 0.8], [0.5, 0.5]\} = \mathcal{M}_2(\mathrm{X}_2),$$

therefore Definition 1 can be applied:

$$\begin{aligned}&(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_1X_2X_3)\\ &\quad = \mathrm{CH}\{[0.2, 0, 0.075, 0.125, 0, 0, 0.275, 0.375], [0, 0.2, 0, 0.2, 0, 0, 0, 0.6],\\ &\qquad [0.1, 0, 0.15, 0.25, 0.1, 0, 0.15, 0.25], [0, 0.1, 0, 0.4, 0, 0.1, 0, 0.4],\\ &\qquad [0.25, 0, 0.25, 0, 0.25, 0, 0.25, 0], [0.1, 0.15, 0.1, 0.15, 0.1, 0.15, 0.1, 0.15],\\ &\qquad [0.2, 0, 0.2, 0, 0.3, 0, 0.3, 0], [0.08, 0.12, 0.08, 0.12, 0.12, 0.18, 0.12, 0.18]\}.\end{aligned}$$

It can easily be checked that both $(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_1X_2) = \mathcal{M}_1(X_1X_2)$ and $(\mathcal{M}_1 \triangleright \mathcal{M}_2)(X_2X_3) = \mathcal{M}_2(X_2X_3)$. ♢

The following theorem reveals the relationship between strong independence and the operator of composition. It is, together with Lemma 1, the most important assertion enabling us to introduce multidimensional models.

Theorem 1 *Let $\mathcal{M}$ be a credal set about $X_{K\cup L}$ with marginals $\mathcal{M}(X_K)$ and $\mathcal{M}(X_L)$. Then*

$$\mathcal{M}(X_{K\cup L}) = (\mathcal{M}^{\downarrow K} \triangleright \mathcal{M}^{\downarrow L})(X_{K\cup L}) \tag{6}$$

iff

$$(K \setminus L) \perp\!\!\!\perp (L \setminus K) | (K \cap L). \tag{7}$$

Proof. Let us suppose that (6) holds. Since $\mathcal{M}_1(X_K)$ and $\mathcal{M}_2(X_L)$ are projective, Definition 1 can be applied and therefore

$$\begin{aligned}\mathcal{M}(X_{K\cup L}) &= \mathrm{CH}\{(P_1 \cdot P_2)/P_2^{\downarrow K\cap L} : P_1 \in \mathcal{M}(X_K),\\ &\qquad P_2 \in \mathcal{M}(X_L), P_1^{\downarrow K\cap L} = P_2^{\downarrow K\cap L}\}).\end{aligned}$$

To prove (7) means to find for any P from $\mathcal{M}(X_{K\cup L})$ a pair of projective distributions P_1 and P_2 from $\mathcal{M}(X_K)$ and $\mathcal{M}(X_L)$, respectively, such that $P = (P_1 \cdot P_2)/P_1^{\downarrow K\cap L}$. But due to condition of projectivity, $\mathcal{M}(X_{K\cup L})$ consists of exactly this type of distributions.

Let on the other hand (7) be satisfied. Then any P from $\mathcal{M}(X_{K\cup L})$ can be expressed as conditional product of its marginals, namely

$$P = (P^{\downarrow K} \cdot P^{\downarrow K})/P^{\downarrow K\cap L},$$

$P^{\downarrow K} \in \mathcal{M}(X_K)$ and $P^{\downarrow L} \in \mathcal{M}(X_L)$. Therefore,

$$\mathcal{M}(X_{K\cup L}) \quad = \quad \{(P^{\downarrow K} \cdot P^{\downarrow K})/P^{\downarrow K\cap L} : P^{\downarrow K} \in \mathcal{M}_1(X_K), P^{\downarrow L} \in \mathcal{M}_2(X_L))\},$$

which concludes the proof. □

4 Credal Networks

A *credal network* [1] over X_N is (analogous to Bayesian networks) a pair $(\mathcal{G}, \{\mathbf{P}_1, \ldots, \mathbf{P}_k\})$ such that for any $i = 1, \ldots, k$ $(\mathcal{G}, \mathbf{P}_i)$ is a Bayesian network over X_N.

The resulting model is a credal set, which is the convex hull of the Bayesian networks, i.e.

$$\mathrm{CH}\{P_1(X_N), \ldots P_k(X_N)\}.$$

It is evident, that this definition looses the attractiveness of Bayesian networks, where the overall information is computed from the local pieces of information.

The most popular (and also effective) type of credal networks are those called separately specified. A *separately specified credal networks* over X_N is a pair $(\mathcal{G}, \mathbf{M})$, where $\mathbf{M})$ is a set of conditional credal sets $\mathcal{M}(X_i|pa(X_i))$ for each $X_i \in X_N$.

Here the overall model is obtained analogous to Bayesian networks as the strong extension of the $\mathcal{M}(X_i|pa(X_i)), i \in N$.

Nevertheless, the reverse side of this nice property is the imprecision increase of this type models, as can be seen even from the following simple example.

Example 3 Let $\mathcal{M}(X_1X_2)$ be a credal set about variables X_1 and X_2 with values in $\mathbf{X}_1$ and $\mathbf{X}_2$ ($\mathbf{X}_i = \{x_i, \bar{x}_i\}$), respectively, defined as in Example 2.

From its extreme points we obtain the following distributions:

$$\begin{array}{lll} P_1(x_2) = 0.2 & P_1(x_1|x_2) = 1 & P_1(x_1|\bar{x}_2) = 0.25 \\ P_2(x_2) = 0.2 & P_2(x_1|x_2) = 0.5 & P_2(x_1|\bar{x}_2) = 0.5 \\ P_3(x_2) = 0.5 & P_3(x_1|x_2) = 0.5 & P_3(x_1|\bar{x}_2) = 0.5 \\ P_4(x_2) = 0.5 & P_4(x_1|x_2) = 0.4 & P_2(x_1|\bar{x}_2) = 0.6, \end{array}$$

These are, together with the graph $X_2 \longrightarrow X_1$ four Bayesian networks. Their convex hull is exactly the set $\mathcal{M}_1(X_1X_2)$. Nevertheless, it is not separably specified credal network. To obtain it we need the credal sets $\mathcal{M}(X_2)$, $\mathcal{M}(X_1|x_2)$ and $\mathcal{M}(X_1|\bar{x}_2)$

From the above values one will get the "extreme" points of $\mathcal{M}(X_1|x_2)$ and $\mathcal{M}(X_1|\bar{x}_2)$:

$$[1, 0], [0.5, 0.5], [0.5, 0.5], [0.4, 0.6],$$

and

$$[0.25, 0.75], [0.5, 0.5], [0.5, 0.5], [0.6, 0.4],$$

respectively.

As $[0.5, 0.5]$ is a linear combination of both [1,0] and [0.4,0.6], and [0.25,0.75] and [0.6,0.4], the resulting (conditional) credal sets are

$$\begin{aligned} \mathcal{M}(X_2) &= \mathrm{CH}\{[0.2, 0.8], [0.5, 0.5]\}, \\ \mathcal{M}(X_1|x_2) &= \mathrm{CH}\{[1, 0], [0.4, 0.6]\}, \\ \mathcal{M}(X_1|\bar{x}_2) &= \mathrm{CH}\{[0.25, 0.75], [0.6, 0.4]\}. \end{aligned}$$

The strong extension of these credal sets is

$$\begin{aligned} \mathcal{M}_1(\tilde{X}_1 X_2) = \mathrm{CH}\{&[0.2, 0.2, 0, 0.6], [0.2, 0.48, 0, 0.32], [0.08, 0.2, 0.12, 0.6], \\ &[0.08, 0.48, 0.12, 0.32], [0.5, 0.125, 0, 0.375], \\ &[0.5, 0.3, 0, 0.2], [0.2, 0.125, 0.3, 0.375], [0.2, 0.3, 0.3, 0.2]\}. \end{aligned}$$

which is less precise than the original model. ♢

5 Conclusions

We introduced an operator of composition of projective credal sets — a generalization of that introduced about 15 years ago in (precise) probability framework. The operator satisfies the basic properties necessary for the introduction of compositional models of credal sets. Nevertheless, the definition must be extended to non-projective credal sets, which seems to be the most important problem to be solved in the near future.

We also recalled the concept of credal networks and we suggested that compositional models of credal sets are potentially good counterpart of these models, which are either not separately specified (contrary to our expectation concerning compositional models), or more imprecise.

Acknowledgements

The support of Grant GAČR P402/11/0378 is gratefully acknowledged.

References

[1] A. Antonucci, M. Zaffalon, Decision-Theoretic Specification of Credal Networks: A Unified Language for Uncertain Modeling with Sets of Bayesian Networks. *Int. J. Approx. Reasoning,* **49** (2008), 345–361.

[2] S. Benferhat, D. Dubois, L. Gracia and H. Prade, Directed possibilistic graphs and possibilistic logic. In: B. Bouchon-Meunier, R.R. Yager, (eds.) *Proc. of the 7th Int. Conf. on Information Processing and Management of*

Uncertainty in Knowledge-based Systems IPMU'98, Editions E.D.K. Paris, pp. 1470–1477.

[3] B. Ben Yaghlane, Ph. Smets and K. Mellouli, Directed evidential networks with conditional belief functions. *Proceedings of ECSQARU 2003,* eds. T. D. Nielsen, N. L. Zhang, 291–305.

[4] I. Couso, S. Moral and P. Walley, Examples of independence for imprecise probabilities, *Proceedings of ISIPTA'99,* eds. G. de Cooman, F. G. Cozman, S. Moral, P. Walley, Ghent, 1999, pp. 121–130.

[5] F. G. Cozman, Sets of probability distributions, independence, and convexity, *Synthese*, **186** (2012), pp. 577–600.

[6] R. Jiroušek. Composition of probability measures on finite spaces. *Proc. of UAI'97*, (D. Geiger and P. P. Shenoy, eds.). Morgan Kaufmann Publ., San Francisco, California, pp. 274–281, 1997.

[7] R. Jiroušek. Graph modelling without graphs. *Proc. of IPMU'98*, (B. Bouchon-Meunier, R.R. Yager, eds.). Editions E.D.K. Paris, pp. 809–816, 1988.

[8] R. Jiroušek. Marginalization in composed probabilistic models. *Proc. of UAI'00* (C. Boutilier and M. Goldszmidt eds.), Morgan Kaufmann Publ., San Francisco, California, pp. 301–308, 2000.

[9] R. Jiroušek, J. Vejnarová and M. Daniel, Compositional models for belief functions. *Proceedings of 5th International Symposium on Imprecise Probability: Theories and Applications ISIPTA'07*, eds. G. De Cooman, J. Vejnarová, M. Zaffalon, Praha, 2007, pp. 243-252.

[10] R. Jiroušek and J. Vejnarová, Compositional models and conditional independence in Evidence Theory, *Int. J. Approx. Reasoning,* **52** (2011), pp. 316-334.

[11] S. Moral and A. Cano, Strong conditional independence for credal sets, *Ann. of Math. and Artif. Intell.*, **35** (2002), pp. 295–321.

[12] J. Vejnarová, Composition of possibility measures on finite spaces: preliminary results. In: *Proc. of 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems IPMU'98,* (B. Bouchon-Meunier, R.R. Yager, eds.). Editions E.D.K. Paris, 1998, pp. 25–30.

[13] J. Vejnarová, On possibilistic marginal problem, *Kybernetika* **43**, 5 (2007), pp. 657–674.

[14] J. Vejnarová, Evidential networks from a different perspective. In: *Synergies of Soft Computing and Statistics for Intelligent Data Analysis,* Soft Methods In Probability and Statistics, (2012). pp. 429–436.

[15] J. Vejnarová, Operator of composition for credal sets, *ISIPTA'13: 8th International Symposium on Imprecise Probability: Theories and Applications*, pp. 355–364.

Hierarchical Clustering Using Proximity Measures Based on Discernibility of Clusters

Yoshifumi Kusunoki
Division of Electrical Electronic and Information Engineering
Osaka University
kusunoki@eei.eng.osaka-u.ac.jp

Tetsuzo Tanino
Division of Electrical Electronic and Information Engineering
Osaka University
tanino@eei.eng.osaka-u.ac.jp

Abstract

Clustering is a method to divide given data or a set of objects into clusters such that similar objects are grouped in the same cluster and dissimilar objects in different clusters. Similarity or dissimilarity of objects are given by a predefined proximity measure. In this research, we focus on data sets with nominal attributes, namely, we can only consider that two values are equal or not. For nominal data sets, logical expressions such as "attribute a equals to v" or "a is not less than v" are suitable to describe clusters. To obtain clusters with simple expressions, we propose a new proximity measure of clusters based on discernibility of clusters. Moreover, we apply the proposed proximity measure to agglomerative hierarchical clustering, and examine their characteristics by numerical experiments.

1 Introduction

Clustering is a method to divide given data or a set of objects into clusters such that similar objects are grouped in the same cluster and dissimilar objects in different clusters. Similarity or dissimilarity of objects are given by a predefined proximity measure.

Proximity measures are defined based on attribute values (descriptions) of objects. Attributes can be divided to two types: numerical and nominal. In basic clustering methods for data sets with numerical attributes, a proximity measure for clusters of objects is given by our geometrical intuition, for example

Euclidean distance. Therefore, representative of a clusters is naturally defined by the centroid or the medoid of the cluster.

In this research, we focus on data sets with nominal attributes, namely, we can only consider that two values are equal or not. For nominal data sets, logical expressions such as "attribute a equals to v" is suitable to describe clusters. However, clustering methods based on conventional proximity measures do not necessarily output clusters of simple and compact logical expressions. Several authors [4, 6, 3] have addressed to this problem.

In this paper, we propose a new proximity measure for pairs of clusters, which directly reflect discernibility of clusters, where given an attribute subsets, clusters are discernible if no pair of objects in different clusters has the same profile on the attribute subset. First, we define an discernibility or indiscernibility relation on clusters. Then, the proposed proximity measure is defined by counting attribute subsets which can discern clusters. Using the fact that the discernibility relation is closely related to Boolean algebra, we propose a method to compute the proximity measure using Boolean formulas. Moreover, we apply the proposed proximity measure to agglomerative hierarchical clustering, and examine their characteristics by numerical experiments.

This paper organized as follows. In Section 2, we introduce several materials such as an information table, indiscernibility and disparateness relations, hierarchical clustering, and Boolean functions, which is used in the following sections. In Section 3, we propose a new proximity measure based on the indiscernibility and disparateness relations. Moreover, we propose a computational method of the proximity measure. In Section 4, we conduct numerical experiments to examine hierarchical clustering with the proposed proximity measure comparing with single-linkage and complete-linkage. Concluding remarks are given in Section 5.

2 Preliminaries

2.1 Information Tables and Indiscernibility and Disparateness Relations

In this research, data sets are modeled by information tables. An information table is defined by $\langle U, AT, V\rangle$. $U = \{u_1, u_2, ..., u_n\}$ is a finite set of objects. $AT = \{a_1, a_2, ..., a_m\}$ is a finite set of attributes. V is a set of attribute values. Every attribute $a \in AT$ is a mapping whose domain and range are U and V_a, respectively, where V_a is the set of values which attribute a takes.

Given an attribute subset $A \subseteq AT$, if objects have the same attribute values of A, then they are indiscernible to each other. The indiscernibility relation R_A with respect to A is defined by

$$R_A = \{(u, u') \in U^2 \mid \forall a \in A,\ a(u) = a(u')\}. \tag{1}$$

Contrary, when $(u, u') \notin R_A$ the pair u, u' is said to be discernible, because at least one of attributes in A can distinguish the pair. We define $R_\emptyset = U^2$. On the

other hand, we define a disparateness relation S_A with respect to A as follows:

$$S_A = \{(u, u') \in U^2 \mid \forall a \in A,\ a(u) \neq a(u')\}. \tag{2}$$

The pair $(u, u') \in S_A$ are completely different because no attribute in A takes the same value for the pair. We define $S_\emptyset = U^2$.

These relations can be extended to relations on object subsets, conjunctively or distinctively. Let $\mathcal{P}(U)$ be the set of all nonempty subsets of U. For an attribute subset $A \subseteq AT$, they are defined as follows:

$$[R]_A = \{(X, Y) \in \mathcal{P}(U)^2 \mid \forall (x, y) \in X \times Y,\ (x, y) \in R_A\}, \tag{3}$$

$$\langle R \rangle_A = \{(X, Y) \in \mathcal{P}(U)^2 \mid \exists (x, y) \in X \times Y,\ (x, y) \in R_A\}, \tag{4}$$

$$[S]_A = \{(X, Y) \in \mathcal{P}(U)^2 \mid \forall (x, y) \in X \times Y,\ (x, y) \in S_A\}, \tag{5}$$

$$\langle S \rangle_A = \{(X, Y) \in \mathcal{P}(U)^2 \mid \exists (x, y) \in X \times Y,\ (x, y) \in S_A\}. \tag{6}$$

Obviously, $(X, Y) \in [R]_A$ (resp. $(X, Y) \in [S]_A$) implies $(X, Y) \in \langle R \rangle_A$ (resp. $(X, Y) \in \langle S \rangle_A$). Additionally, when $(X, Y) \in [R]_A$, all objects in X or Y have the same value v for each $a \in A$, namely, for any $x, x' \in X$ we have $a(x) = a(y) = v$. Similarly, when $(X, Y) \in [S]_A$, the sets of all values for each $a \in A$ in X and Y does not have intersection, namely, $\{a(x) \mid x \in X\} \cap \{a(y) \mid y \in Y\} = \emptyset$. Especially when a is binary, i.e., $V_a = \{0, 1\}$, all objects in X or Y have the same attribute value for a. R_A is related to definability of rough set theory [5].

Because relations $[R]_A$ and $[S]_A$ are too strict to use for proximity of clusters, we use only $\langle R \rangle_A$ and $\langle S \rangle_A$. For simplicity, we denote $\langle R \rangle_A$ and $\langle S \rangle_A$ by R_A and S_A, respectively. Moreover, for an object pair $u, u' \in U$, $(u, u') \in R_A$ (resp. $(u, u') \in S_A$) means $(\{u\}, \{u'\}) \in \langle R \rangle_A$ (resp. $(\{u\}, \{u'\}) \in \langle S \rangle_A$).

2.2 Hierarchical Clustering

Hierarchical clustering groups objects by hierarchical classification which is a series of partitions from a single cluster containing all of objects to all singleton clusters of objects. A tree diagram representing the hierarchical classification is called a dendrogram. Hierarchical clustering methods are divided into agglomerative and divisive approaches. In this paper, we use the agglomerative approach. The agglomerative approach successively merges objects into groups, while a divisive method successively separates clusters into finer ones. To conduct the agglomerative hierarchical clustering, a proximity measure between pair of nonempty clusters $X, Y \subseteq U$ is needed. Several definition of proximity measures between clusters have been proposed, for instance single-linkage, complete-linkage and so on.

A general procedure of the agglomerative hierarchical clustering is described as follows. Now, we use dissimilarity measure d.

(1) Initial clusters are set to be the singleton clusters of objects: $C_1 = \{u_1\}, C_2 = \{u_2\}, ..., C_n = \{u_n\}$.

(2) Marge a pair C_i and C_j whose dissimilarity $d(C_i, C_j)$ is minimum among all pairs of current clusters, and rearrange the index.

(3) If all clusters are merged into one cluster then the procedure ends, otherwise go to (2).

2.3 Boolean Functions

Boolean functions play a key role for computation of a proposed proximity measure. A Boolean function is a mapping $f : \{0,1\}^q \to \{0,1\}$, where $w \in \{0,1\}^q$ is called a Boolean vector whose ith component is w_i. For two Boolean function f and g, $g \leq f$ means that f and g satisfy $g(w) \leq f(w)$ for all $w \in \{0,1\}^q$, and $g < f$ means that $g \leq f$ and $g \neq f$. Boolean variables $x_1, x_2, ...$ and the complements $\bar{x}_1, \bar{x}_2, ...$ are called literals. A clause (resp., term) is a disjunction (resp., conjunction) of at most one of x_i and $\bar{x}_i$ for each variable. The empty disjunction (resp., conjunction) is denoted by $\bot$ (resp., $\top$).

A clause c (resp., term t) is an implicate (resp., implicant) of a function f, if $f \leq c$ (resp. $t \leq f$). Moreover, it is prime if there is no implicate $c' < c$ (resp., no implicant $t' > t$) of f, and positive (monotone) if it consists of positive literals only. A conjunction normal form (CNF) (resp., disjunction normal form (DNF)) is a conjunction of clauses (resp. disjunction of terms), and it is prime if all its members are prime.

3 Proximity Based on Discernibility

3.1 Definition of Proximity Measure

In this paper, we aim to propose a new proximity measure for the agglomerative hierarchical clustering such that the obtained clusters can be described simple logical expressions. To describe clusters by logical forms, it is important that a cluster can be discernible from other clusters on some attribute subset. Namely, a nonempty cluster $X \in \mathcal{P}(U)$ and another nonempty cluster $Y \in \mathcal{P}(U)$ have the discernibility relation R, i.e., no pair of objects in the different clusters has the same attribute profile. Moreover, if X and Y are discernible on many attribute subsets, we can say that X and Y are dissimilar. Therefore, we can define a dissimilarity measure $d_R(X, Y)$ by the cardinality of the attribute subsets which discern X and Y.

$$d_R(X, Y) = |\{A \subseteq AT \mid (X, Y) \notin R_A\}|. \tag{7}$$

The goal of our clustering is to obtain clusters each pair of which has a large value of d_R. However, applying d_R directly to the hierarchical clustering has difficulties.

- The measure d_R is monotonic with the sizes (cardinalities) of the input clusters. Large clusters have small values of d_R.

- Let $X, Y, Z \subset U$ be nonempty clusters. When $d_R(X, Y)$ is smaller than the other pairs, X and Y are merged in the agglomerative hierarchical clustering. Nevertheless, $d_R(X \cup Y, Z)$ may be smaller than the other possibilities $d_R(X \cup Z, Y)$ and $d_R(Y \cup Z, X)$.
- The measure d_R is insensitive to the input clusters. Some pairs of clusters have the same values of d_R.

To overcome these difficulties, we modify the dissimilarity measure:

$$d(X, Y) = |\{A \subseteq AT \mid (X, Y) \notin R_A \text{ and } (X, Y) \in S_A\}|. \tag{8}$$

Attribute subsets A such that $(X, Y) \notin S_A$ may distinguish the merged cluster $X \cup Y$ from the other clusters, because objects in X and Y possibly have the same attribute values on A. Hence, $d(X, Y)$ estimates how many attribute subsets which discern X or Y from the other clusters are reduced after merging X and Y.

For $x, y \in U$, let $d'(x, y)$ be the conventional distance of x and y: $d'(x, y) = |\{a \in AT \mid a(x) \neq a(y)\}|$. The proposed dissimilarity of the singleton clusters $\{x\}$ and $\{y\}$ can be expressed by d':

$$d(\{x\}, \{y\}) = 2^{d'(x,y)} - 1. \tag{9}$$

We can rescale $d(X, Y)$ to $d'(X, Y)$ by the above equation:

$$d'(X, Y) = \log_2(d(X, Y) + 1). \tag{10}$$

The rest of the paper, we use d' instead of d, and use the same symbol d instead of d'.

3.2 On Computation for Dissimilarity Measure

To use the proposed dissimilarity measure, we need to calculate the number of attribute subsets which satisfy the condition of (8). It is hard task to search all of subsets of attributes, but we have an explicit formula of a Boolean function corresponding to the condition. For $X, Y \in \mathcal{P}(U)$, these formulas are

$$\varphi_{X,Y}(a_1^*, a_2^*, \ldots, a_m^*) = \varphi^R_{X,Y}(a_1^*, a_2^*, \ldots, a_m^*) \wedge \bar{\varphi}^S_{X,Y}(a_1^*, a_2^*, \ldots, a_m^*), \tag{11}$$

where,

$$\varphi^R_{X,Y}(a_1^*, a_2^*, \ldots, a_m^*) = \bigwedge_{u \in X, u' \in Y} \bigvee \{a_i^* \mid a_i \in A, a_i(u) \neq a_i(u')\},$$
$$\varphi^S_{X,Y}(a_1^*, a_2^*, \ldots, a_m^*) = \bigwedge_{u \in X, u' \in Y} \bigvee \{a_i^* \mid a_i \in A, a_i(u) = a_i(u')\},$$

a_i^* is the positive literals corresponding to attributes a_i, $\bar{\varphi}^S_{X,Y}$ is the negation of $\varphi^S_{X,Y}$, and m is the number of attributes.

There is the one-to-one correspondence between attribute subsets A and m dimensional Boolean vectors $A^* = (A_1^*, A_2^*, ..., A_m^*)$ such that,

$$a_i \in A \Leftrightarrow A_i^* = 1, \text{ for } a_i \in AT. \tag{12}$$

We have the following proposition.

$$(X, Y) \notin R_A \text{ and } (X, Y) \in S_A \Leftrightarrow \varphi_{X,Y}(A^*) = 1, \text{ for } A \subseteq AT. \tag{13}$$

Now obtaining the value of $d(X, Y)$ transforms to calculating the number of true vectors of $\varphi_{X,Y}$. For this purpose, we generate a sequence of positive Boolean formulas from $\varphi_{X,Y}$. Let $\varphi_{X,Y}^S$ be denoted by $c_1^S \wedge \cdots \wedge c_l^S$, where $c_i^S, i = 1, \ldots, l$ are clauses. First, we define a positive Boolean formula $\varphi_{X,Y}^1 = (\varphi_{X,Y}^R)_{\bar{c}_1^S=1}$, where the subscript $\bar{c}_1^S = 1$ means the negation of c_1^S is set to 1. It implies that all literals appear in c_1^S are set to 0. Second, we define $\varphi_{X,Y}^2 = (\varphi_{X,Y}^R \wedge c_1^S)_{\bar{c}_2^S=1}$. Finally, we obtain positive Boolean formulas $\varphi_{X,Y}^1, \varphi_{X,Y}^2, \ldots, \varphi_{X,Y}^l$, and we can see that the number of true vectors of $\varphi_{X,Y}$ is the sum of those of $\varphi_{X,Y}^1, \varphi_{X,Y}^2, \ldots, \varphi_{X,Y}^l$.

Then, we calculate the number of true vectors of the positive Boolean formulas $\varphi_{X,Y}^i$ by a branching method with a recursive formula. Let $\mathcal{W}$ and φ be a set of Boolean variables and a Boolean formula. Let $\mathcal{N}(\varphi, \mathcal{W})$ be the number of true vectors of φ with the input variables $\mathcal{W}$ Moreover, Let $\mathcal{V}(\varphi)$ be the set of Boolean variables which actually appear in φ. We have the following equation,

$$\mathcal{N}(\varphi, \mathcal{W}) = (\mathcal{N}(\varphi_{x=1}, \mathcal{V}(\varphi) - \{x\}) + \mathcal{N}(\varphi_{x=0}, \mathcal{V}(\varphi) - \{x\})) \times 2^{|\mathcal{W} - \mathcal{V}(\varphi)|}, \tag{14}$$

where x is arbitrary variable in $\mathcal{W}$, $\varphi_{x=1}$ and $\varphi_{x=0}$ are the formulas with fixing $x = 1$ and $x = 0$, respectively. We define $\mathcal{N}(\top, \mathcal{W}) = 2^{|\mathcal{W}|}$ and $\mathcal{N}(\bot, \mathcal{W}) = 0$. Because at least one variable is removed from the formula at each branching $x = 1$ and $x = 0$, this procedure necessarily reaches either $\mathcal{N}(\top, \mathcal{W})$ or $\mathcal{N}(\bot, \mathcal{W})$ and stops within finite steps. Set $\mathcal{W} = \{a_1^*, a_2^*, \ldots a_m^*\}$, and applying (14) to $(\varphi_{X,Y}^i, \mathcal{W})$ recursively, we obtain the number of its true vectors.

4 Numerical Experiments

We examine characteristics of the proposed dissimilarity measure by numerical experiments. As mentioned above, we apply it to agglomerative hierarchical clustering.

Data sets are obtained from UCI Machine Learning Repository [1]. We use two data sets "breast-cancer" and "zoo", whose numbers of objects and attributes are shown in Table 1. Moreover duplicated objects were deleted to simplify results.

For reference methods, we use agglomerative hierarchical clustering with single-linkage and complete-linkage. At single-linkage method, dissimilarity between a pair of clusters is defined by the minimum dissimilarity of a pair of

Table 1: Data Sets

Data set	Num of objects	Num of attributes
breast-cancer	286	10
zoo	58	18

Table 2: Results for breast-cancer

	Proposed	Single-linkage	Complete-linkage
Num clusters	30	33	32
Max size	116	247	26
Ave size	9.5	8.9	8.6
Intra distance	0.39	0.51	0.30
Num rules	64	44	104

Table 3: Results for zoo

	Proposed	Single-linkage	Complete-linkage
Num clusters	7	8	7
Max size	14	19	18
Ave size	8.3	7.3	8.3
Intra distance	0.20	0.24	0.22
Num rules	7	8	9

objects in different clusters. Contrary, at complete-linkage method, it is defined by the maximum dissimilarity.

The experimental results are shown in Table 2 and 3. The values in the tables are characteristics of clusters for each data set computed by the proposed method, and single- and complete-linkage methods. Clusters are obtained from a dendrogram, which is the output of hierarchical clustering, by cutting it at some level. The cutting level for each method is chosen so that the numbers of clusters of three methods are almost same.

In the table, "Max size" and "Ave size" mean the maximum and average cardinality of clusters. "Intra distance" means the average of dissimilarities d of pairs x, y in the same clusters. "Num rules" means the number of rules representing clusters, which are induced by the LEM2 algorithm [2].

Comparing the single-linkage method, the clusters of the proposed method have smaller intra distance. On the other hand, comparing the complete-linkage method, the clusters of the proposed method can be represented by the smaller number of rules. Namely, the proposed method compromises homogeneity and simple expression of obtained clusters.

5 Concluding Remarks

In this paper, we have studied proximity based on discernibility of clusters on attribute subsets. We have proposed a dissimilarity measure based on indiscernibility and disparateness relations R and S. For two nonempty clusters X and Y, the proposed dissimilarity $d(X, Y)$ is defined by the number of attribute subsets A such that $(X, Y) \notin R_A$ and $(X, Y) \in S_A$. Because the indiscernibility and disparateness relations are closely related to Boolean algebra, we can compute the proposed dissimilarity $d(X, Y)$ by counting the number of true vectors of Boolean formula $\varphi_{X,Y}$.

To examine the proposed dissimilarity measure, we have conducted numerical experiments. We have applied the dissimilarity measure to agglomerative hierarchical clustering, and compared it with agglomerative clustering with single- and complete-linkage. The results show that the proposed method compromises homogeneity and simple expression of obtained clusters.

To examine the proposed dissimilarity measure using more benchmark data sets is one of future topics.

Acknowledgment We acknowledge that this work has been supported by the Grant-in-Aid for Young Scientists (B) No. 23700265.

References

[1] Frank, A. and Asuncion, A. (2010), UCI Machine Learning Repository [`http://archive.ics.uci.edu/ml`], Irvine, CA: University of California, School of Information and Computer Science

[2] Grzymala-Busse, J. W. (1992), LERS – A system for learning from examples based on rough sets, in *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, R. Słowínski (ed.), Kluwer Academic Publishers, Dordrecht, 3–18

[3] Honda, K., Notsu, A., Ichihashi, H. (2009), Collaborative filtering by sequential extraction of user-item clusters based on structural balancing approach, *IEEE International Conference on Fuzzy Systems*, art. no. 5277251, 1540–1545

[4] Michalski, R. S. and Stepp, R. E. (1983), Learning from observation: Conceptual Clustering, *Machine Learning: An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell and T. M. Mitchell (eds.), 331–363, Tioga Publishing Co.

[5] Pawlak, Z. and Skowron, A. (2007), Rudiments of rough sets, *Inform. Sci.*, 177, 3–27

[6] San, O. M., Huynh, V.-N and Nakamori, Y. (2004), An alternative extension of the k-means algorithm for clustering categorical data, *Int. J. Appl. Math. Comput. Sci.*, 14(2), 241–247

Rule Based Systems for Monitoring and Controller Re-adaptation

Jana Nowaková
Department of Cybernetics and Biomedical Engineering
VŠB-Technical University of Ostrava
jana.nowakova@vsb.cz

Miroslav Pokorný
Department of Informatics and Applied Mathematics
Moravian University College Olomouc
miroslav.pokorny@mvso.cz

Abstract

Finding of monitoring systems for deciding if or how re-adapt a PID controller in literature is not so complicated. These monitoring systems are also widely used in industry. But monitoring system which is based on non-conventional methods for deciding, takes into account the non-numeric terms and it is open for adding more rules, is not so common. In the paper the procedures of control quality monitoring and necessary re-adaptation of PID controller is solved using the fuzzy-logic principle through the rule-based expert systems. There are used two fuzzy expert systems. The first one for monitoring of quality of regulation process is of Mamdani type, its rule base is created within two input linguistic variables - namely the relative settling time and relative overshoot are mentioned. The second expert system for designing of parameters of classic PID controller is of Takagi-Sugeno type, whose knowledge base is built on know-how obtained from the combination of the frequency response method and the step response Ziegler-Nichols design method. The proof of efficiency of the proposed method and a numerical experiment is presented by the simulation in the software environment Matlab-Simulink. It is shown that presented monitoring system with following design of PID controller is useful for family of controlled systems of second order. The both described knowledge-based systems are open and can be widen any time.

1 Introduction

The paper is focused on adjustment of the monitoring system for deciding when re-adapt the classic PID controller. Presented monitoring system has some

common elements with [1]. It is created fuzzy expert system of Mamdani type (ES1) with two inputs - overshoot and settling time, but settling time is not defined as a classic time, but as the part or multiple of previous measured settling time (relative settling time) and one output - score. So there are monitored simply obtained process parameters. The score determines if is necessary to re-adapt the controller.

The following design of parameters of classic PID controller is done by the second fuzzy expert (ES2) system with a knowledge base is built on know-how obtained from the combination of the frequency response method and the step response Ziegler-Nichols design method [2].

2 Monitoring System

The monitoring system is fuzzy expert system [4], [5], [6] of Mamdani type with two inputs, knowledge base with linguistic rules and one output and it has been created and the efficiency is proofed for controlled systems of the second order with transfer function in the form

$$G_S(s) = \frac{1}{a_2 s^2 + a_1 s + a_0}. \tag{1}$$

2.1 Inputs - Relative Overshoot and Relative Settling Time

The first input is the linguistic variable relative overshoot (RO) - the difference between the controlled value (CV) and the required value (RV) is rated relatively to the required value (2).

$$RO = \frac{|CV - RV|}{RV} \tag{2}$$

The maximal overshoot of the time response of the system is detected after the fast step change of timing, the fast step changes in timing could be caused by the e.g. change of the controlled system or change of the required value. As it is expressed in percentage, it is relative overshoot (RO). The value of the overshoot is stored in the memory and then used with the settling time for determining the score.

The second input is the linguistic variable relative settling time (RST). As the name says, it is not the classic settling time (ST_k), it is defined as the part or multiple of previous settling time ($ST_k - 1$)(3). So the classic settling time is also stored in the memory as the overshoot, but the linguistic variable relative settling time is defined as the ratio of current settling time and previous settling time.

$$RST_k = \frac{ST_k}{ST_{k-1}} \tag{3}$$

For evaluation of the settling time the 3 % standard deviation from steady-state value [7]. The linguistic values of both linguistic variables are expressed by fuzzy sets, for each linguistic variable by three linguistic values (Figure 1,2).

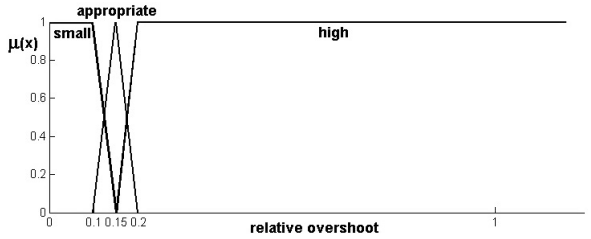

Figure 1: The shape of the membership functions of linguistic values for input linguistic variable Relative Overshoot (RO)

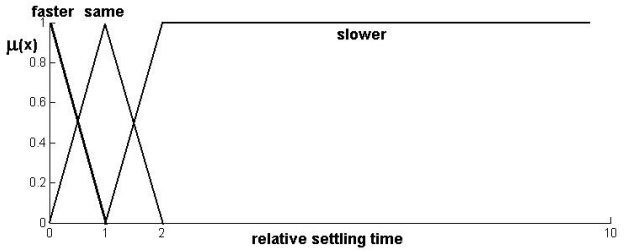

Figure 2: The shape of the membership functions of linguistic values for input linguistic variable Relative Settling Time (RST)

2.2 Output - Score

The output of the monitoring fuzzy expert system is the score, which is also the linguistic variable and its linguistic values are expressed by fuzzy sets (Figure 3). As the fuzzy expert system of Mamdani type is used [8], the linguistic variable score must be defuzzificated. For defuzzification it is used the COA method (Center of Area) [9]. The score more than 2 means that the time response with the current controller can be considered as appropriate. The score less than 2 is considered as not satisfactory and the controller has to be re-adapted [1].

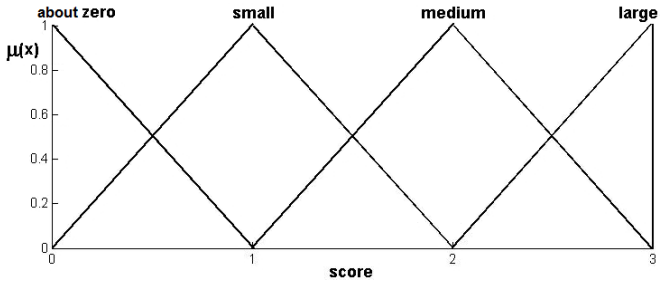

Figure 3: The shape of the membership functions of linguistic values for output linguistic variable Score

2.3 Knowledge Base

The knowledge base is formed by nine linguistic IF-THEN rules of the Mamdani type [9]:

1. If (RO is $Small$) & (RST is $Slower$) then ($Score$ is $Medium$)
2. If (RO is $Small$) & (RST is $Same$) then ($Score$ is $Large$)
3. If (RO is $Small$) & (RST is $Faster$) then ($Score$ is $Large$)
4. If (RO is $Appropriate$) & (RST is $Slower$) then ($Score$ is $Small$)
5. If (RO is $Appropriate$) & (RST is $Same$) then ($Score$ is $Medium$)
6. If (RO is $Appropriate$) & (RST is $Faster$) then ($Score$ is $Large$)
7. If (RO is $High$) & (RST is $Slower$) then ($Score$ is $About\ zero$)
8. If (RO is $High$) & (RST is $Same$) then ($Score$ is $Small$)
9. If (RO is $High$) & (RST is $Faster$) then ($Score$ is $Medium$)

The shape of membership function of output variable is inferred using the Mamdani method. The crisp value of the output score is determined using the defuzzification method Center of Area [9].

3 PID Parameter Design System

As it was mentioned also for design parameters of conventional PID controller fuzzy expert system (ES2) is used [2], [3]. It uses know-how obtained from the combination of the frequency response method and the step response Ziegler-Nichols design method. [7], [10]

Expert design system is model of Takagi-Sugeno type [11] so it does not require defuzzification. The knowledge base of the ES2 is consisted of 27 linguistic rules. For detailed information see [2].

4 The Description of Implemented Algorithm

It is important to define the algorithm of monitoring and following re-adaptation of the controller. The relative overshoot is monitored and stored in memory after every step change of controlled value. The settling time is measured also after every step change of controlled value and is assessed to the previous settling time. If these two monitored parameters are obtained the score is assessed. According to the value of the score, the identification of the system starts and the controller is re-adapted (section 2.2). The simplified model in Matlab-Simulink [12] is depicted in Figure 4.

$$Score \quad \left\{ \begin{array}{ll} \geq 2 & \text{do not re-adapt} \\ < 2 & \text{re-adapt} \end{array} \right\}. \tag{4}$$

For identification the stochastic identification - ARMAX method is used [13]. The re-adaptation (change of parameters of controller) procedure ES2 is done only after the change of required value.

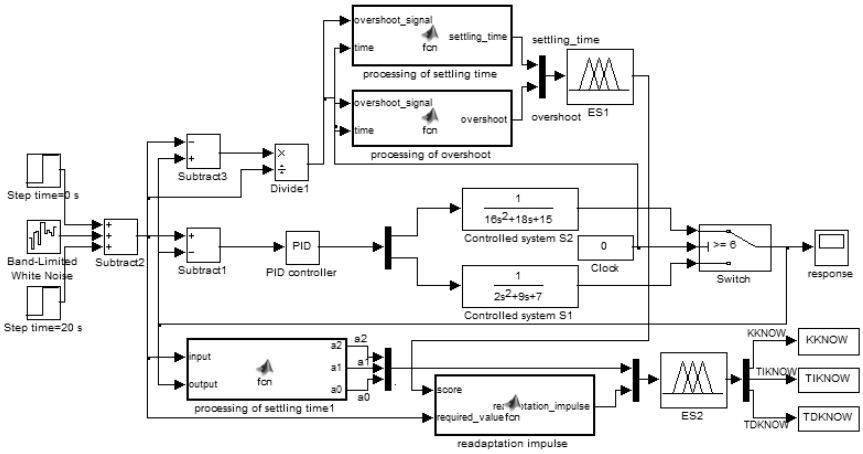

Figure 4: Simplified model in Matlab-Simulink

5 Verification of Created System

The verification was done in Matlab-Simulink [12], the timing with description of important moments is depicted in Figure 5. Verification of the re-adaptation procedure proposed above is started using the controlled system (S1) with transfer function

$$G_{S1}(s) = \frac{1}{2s^2 + 9s + 7} \tag{5}$$

for which the controller with transfer function

$$G_{R1}(s) = 6.1 \left(1 + \frac{1}{0.58s} + 0.14s\right) \tag{6}$$

designed through the identification system ES2 is used.

At the time t_A the unit step of required value is introduced. Therefore, the control process is carried out with 14%-overshoot and settling time $t_{st1} = 4.3\ sec$. The appropriate calculated score by ES1 is

$$score_1 = 2.10 > 2, \tag{7}$$

which corresponds to the satisfactory control course.

At the time t_C a sudden change of the controlled system (to controlled system S2) is simulated from the transfer function $G_{S1}(s)$ to the transfer function

$$G_{S2}(s) = \frac{1}{16s^2 + 18s + 15}. \tag{8}$$

Thus, a non-zero control deviation appeared which is compensated by the original controller $G_{R1}(s)$. The oscillating control course appeared with 12%-overshoot and settling time $t_{st2} = 8.4$ *sec.* Now, the calculated appropriate score is

$$score_2 = 1.66 < 2 \tag{9}$$

and insufficient control course is now indicated.

Therefore, when the nearest change of the deviation appeared at the time t_E (the unit step of reguired value is introduced) the re-adaptive process ES2 is initialized and it is designed a new controller with transfer function

$$G_{R2}(s) = 5.1\left(1 + \frac{1}{0.80s} + 0.20s\right). \tag{10}$$

Now, the control process is carried out without any overshoot and with settling time $t_{st3} = 5.5$ *sec.* The calculated appropriate score value is

$$score_3 = 2.65 > 2 \tag{11}$$

and the satisfactory control course is restored again.

6 Conclusion

The procedures of control quality monitoring and necessary re-adaptation of PID controller is solved using the fuzzy-logic principle through the rule-based expert systems. The first one concludes the initial impulse for controller adaptation. The rule base is created within two input linguistic variables - namely the relative settling time and relative overshoot are mentioned. The following design of parameters of classic PID controller is done by the second fuzzy expert system with a knowledge base which is built on know-how obtained from the combination of the frequency response method and the step response Ziegler-Nichols design method. The proof of efficiency was done using simulations in Matlab-Simulink. It is shown that presented monitoring system with following design of PID controller is useful for family of controlled systems of second order. The both described knowledge-based systems are open. Therefore, next time authors think of adding more monitored parameters and widening of family of controlled systems.

Acknowledgements

This work has been supported partly by Project SP2013/168, "Methods of Acquisition and Transmission of Data in Distributed Systems", of the Student Grant System, VŠB - Technical University of Ostrava and partly by Project GAČR P403-12-1811: Unconventional Managerial Decision making Methods Development in Enterprise Economics and Public Economy.

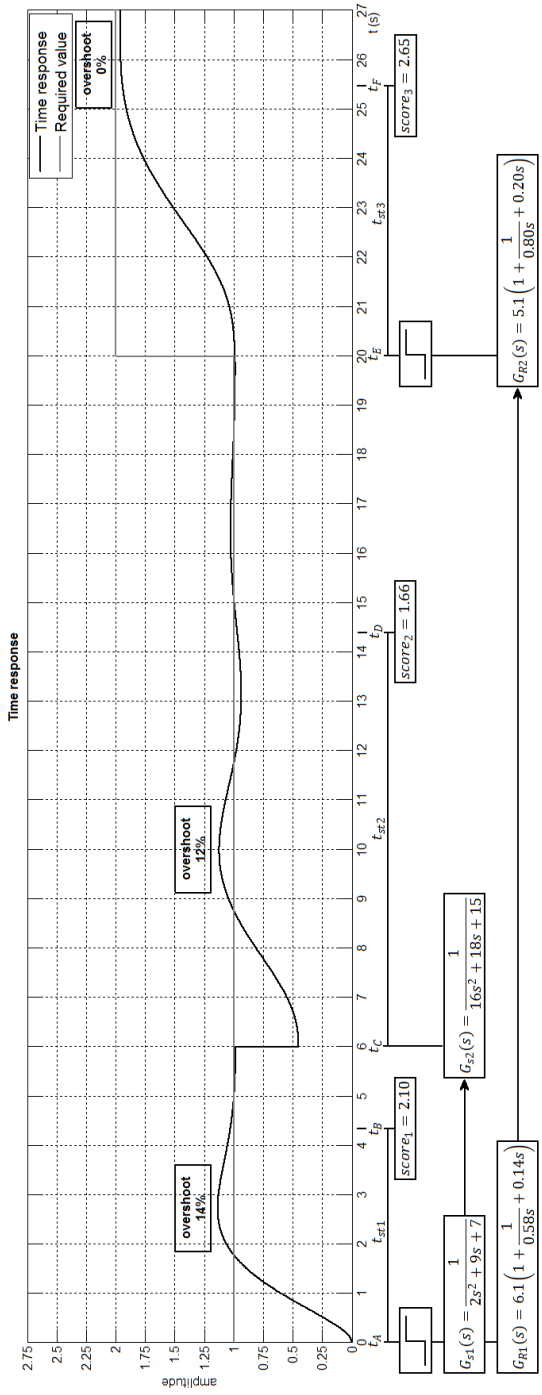

Figure 5: Time response

References

[1] Farsi, M. - Karam, K.Z. - Abdalla, H.H. (1996), *Intelligent multi-controller assessment using fuzzy logic.*, Fuzzy Sets and Systems. **79**, 25-41.

[2] Nowaková, J. - Pokorný, M. (2012), *On PID Controller Design Using Knowledge Based Fuzzy System.*, Advances in Electrical and Electronic Engineering, **10(1)**, 18-27.

[3] Nowaková, J. - Pokorný, M. (2012), *Fuzzy rule based optimization of systems control.*, 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty Osaka, September 2012.

[4] Jacskon, P. (1998), *Introduction To Expert Systems (3 ed.).*, Addison Wesley

[5] Siler,W., Buckley, J.J. (2004), *Fuzzy Expert Systems and Fuzzy Reasoning.*, Wiley-Interscience.

[6] Jones, Ch.H. (1995), *Knowledge Based Systems Methods: A Practitioners' Guide.*, Prentice Hall PTR.

[7] Astrom, K. J., Hagglund, T. (1995), *PID Controllers: theory, design, and tuning.*, USA.

[8] Novák, V., Perfilieva, I., Močkoř,J. (1999), *Mathematical Principles of Fuzzy Logic.*, Kluwer, Boston.

[9] Jager, R. (1995), *Fuzzy Logic in Control.*, Delft.

[10] Kilian, Ch,.g (2005), *Modern Control Technology.*, Thompson Delmar Learning.

[11] Klir, B.J., YUAN, B. (1995),*Fuzzy Sets and Fuzzy Logic: Theory and Application.*, Prentice Hall, New Jersey.

[12] MATLAB - The MathWorks-MATLAB and Simulink for Technical Computing. [cit. 2012-07-10]. http://www.mahworks.com

[13] Noskievič, P. (1999), *Modelování a identifikace systémů.*, Montanex, Ostrava.

Triads of implicational, double-implicational and equivalency data-mining quantifiers

Jiří Iványek
Department of Information and Knowledge Engineering
University of Economics, Prague
e-mail: ivanek@vse.cz

Abstract

Implicational $\Rightarrow^*$, double-implicational $\Leftrightarrow^*$ and equivalency $\equiv^*$ data-mining quantifiers defined on four-fold tables can form logical triads connected by the set of inequalities among their values (which are treated within fuzzy logic as truth values in the unit interval). In the contribution, properties of those triads will be described and illustrated on examples of truth-configurations of the set of seven formulae:

$$\varphi \Rightarrow^* \psi, \psi \Rightarrow^* \varphi, \neg\psi \Rightarrow^* \neg\varphi, \neg\varphi \Rightarrow^* \neg\psi, \varphi \Leftrightarrow^* \psi, \neg\varphi \Leftrightarrow^* \neg\psi, \varphi \equiv^* \psi$$

1 Introduction

The notion of generalized observational quantifiers was introduced in the framework of logical theory of the GUHA method of mechanized hypothesis formation [2], [3]. It should be stressed that this method is one of the earliest methods of data mining. The method was during years developed. Further investigations of its mathematical and logical foundations can be found e.g. in [15], [14].

In the contribution, several classes of the most widely used four-fold table quantifiers with truth values in the unit interval are investigated. The definition of triads of quantifiers provides a logically strong one-to-one correspondence among classes of implicational, Σ-double implicational, and Σ-equivalency quantifiers. Possible truth configurations of formulas with ratio-quantifiers (given data and some threshold) are discussed in the last section.

2 Four-fold table generalized quantifiers

Assume having a data file and consider two Boolean (binary, dichotomic) attributes φ and ψ. A four-fold table $< a, b, c, d >$ corresponding to these attributes is composed from numbers of objects in data satisfying four different

Boolean combinations of attributes:

	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

a - satisfying φ and ψ,
b - satisfying φ and $\neg\psi$,
c - satisfying $\neg\varphi$ and ψ,
d - satisfying $\neg\varphi$ and $\neg\psi$.

To avoid degenerated situations, we shall assume that all marginals of the four-fold table are non-zero: $a+b>0, \quad c+d>0, \quad a+c>0, \quad b+d>0$.

Various relations between φ and ψ can be measured in given data by different four-fold table generalized quantifiers $\sim(a,b,c,d)$ which will be understood here as functions with values in the interval $[0,1]$ like in fuzzy logic with evaluated formulae [1], [12].

A **four-fold table generalized quantifier** $\sim$ is a $[0,1]$-valued function defined for all four-fold tables $<a,b,c,d>$.

We shall write $\sim(a,b)$ if the value of the quantifier $\sim$ depends only on a,b; $\sim(a,b,c)$ if the value of the quantifier $\sim$ depends only on a,b,c; $\sim(a,b,c,d)$ if the value of the quantifier $\sim$ depends on all a,b,c,d. For brevity, we shall call in this paper the "four-fold table generalized quantifiers" simply "quantifiers".

The most general class of quantifiers originally introduced in two-valued logic in [2] and called there **associational** is reflecting the following property: If the four-fold table $<a,b,c,d>$ represents a behavior of the derived attributes φ and ψ in the given data, then the numbers a,d are supporting correlation of φ and ψ but the numbers b,c are against. This property can be formulated in fuzzy logic approach by: The higher are a,d and the smaller are b,c, the better or at least not worse is truth-value of association of φ and ψ in given data.

The most common examples of such associational quantifiers are:

- quantifier $\Rightarrow_{\oslash}$ of **basic implication** (or **confidence** of association rules):

$$\Rightarrow_{\oslash}(a,b)=\frac{a}{a+b}.$$

- quantifier $\Leftrightarrow_{\oslash}$ of **basic double implication** (Jaccard 1900):

$$\Leftrightarrow_{\oslash}(a,b,c)=\frac{a}{a+b+c}.$$

- quantifier $\equiv_{\oslash}$ of **basic equivalence** (Kendall, Sokal-Michener 1958):

$$\equiv_{\oslash}(a,b,c,d)=\frac{a+d}{a+b+c+d}.$$

Properties of the basic quantifiers are in the core of definitions of several useful classes of quantifiers (introduced originally in two-valued logic in [2], [3]) which can be naturally given in fuzzy logic as follows (a,b,c,d,a',b',c',d' mean frequencies from arbitrary pairs of four-fold tables $<a,b,c,d>$ and $<a',b',c',d'>$, respectively):

1. A quantifier $\sim(a,b)$ is **implicational** if
   $\sim(a',b') \geq \sim(a,b)$ when $a' \geq a, b' \leq b$.

2. A quantifier $\sim(a,b,c)$ is **Σ-double implicational** if
   $\sim(a',b',c') \geq \sim(a,b,c)$ when $a' \geq a, b'+c' \leq b+c$.

3. A quantifier $\sim(a,b,c,d)$ is **Σ-equivalence** if
   $\sim(a',b',c',d') \geq \sim(a,b,c,d)$ when $a'+d' \geq a+d, b'+c' \leq b+c$.

3 Affiliated double-implication and equivalency quantifiers

In the paper [7], the method of construction of **triads of quantifiers** is described. Starting from an **implicational quantifier** $\Rightarrow^*$, **affiliated double-implicational quantifier**
$\Leftrightarrow^*$ is given by the formula

$$\Leftrightarrow^*(a,b,c) = \Rightarrow^*(a,b+c),$$

and **affiliated equivalency quantifier** $\equiv^*$ is given by the formula

$$\equiv^*(a,b,c,d) = \Rightarrow^*(a+d,b+c).$$

Double-implicational quantifier $\Leftrightarrow^*$ measures the validity of bi-implication $(\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \varphi)$ in data taking into account only cases where φ or ψ is satisfied. Equivalency quantifier $\equiv^*$ measures the validity of equivalency $\varphi \equiv \psi$ in the whole data. Both affiliated quantifiers $\Leftrightarrow^*, \equiv^*$ naturally extend quantification of implication (given by a definition of particual implicational quantifier $\Rightarrow^*$) for covering also two types of symmetric relations between φ and ψ in data.

It is proved [7] that the above constructed double-implicational quantifier $\Leftrightarrow^*$ is in some sense the least strict one out of the class of Σ-double implication quantifiers satisfying the following inequality:

$$\Leftrightarrow^*(a,b,c) \leq \min(\Rightarrow^*(a,b), \Rightarrow^*(a,c)).$$

Analogically, the above constructed equivalency quantifier $\equiv^*$ is in some sense the most strict one out of the class of Σ-equivalency quantifiers satisfying the following inequality:

$$\equiv^*(a,b,c,d) \geq \max(\Leftrightarrow^*(a,b,c), \Leftrightarrow^*(d,b,c)).$$

From the fuzzy logic point of view, the following deduction rules are correct for such triads of quantifiers:

$$\frac{\varphi \Leftrightarrow^* \psi}{\varphi \Rightarrow^* \psi}, \quad \frac{\varphi \Leftrightarrow^* \psi}{\psi \Rightarrow^* \varphi}, \quad \frac{\varphi \Leftrightarrow^* \psi}{\varphi \equiv^* \psi}, \quad \frac{\neg\varphi \Leftrightarrow^* \neg\psi}{\varphi \equiv^* \psi}.$$

The most common example of a triad of quantifiers is the triad of basic quantifiers given in the previous section.

Another triad of quantifiers can be obtained from the statistically motivated quantifier of *upper critical implication* (see [2], [15]) $\Rightarrow_p^?$ (where p is a parameter, $0 < p < 1$):

$$\Rightarrow_p^? (a,b) = \sum_{i=0}^{a} \frac{(a+b)!}{i!\,(a+b-i)!}\, p^i\, (1-p)^{a+b-i}$$

Quantifier $\Leftrightarrow_p^?$ of *upper critical double implication*

$$\Leftrightarrow_p^? (a,b,c) = \sum_{i=0}^{a} \frac{(a+b+c)!}{i!(a+b+c-i)!} p^i (1-p)^{a+b+c-i}$$

Quantifier $\equiv_p^?$ of *upper critical equivalence*

$$\equiv_p^? (a,b,c,d) = \sum_{i=0}^{a+d} \frac{(a+b+c+d)!}{i!(a+b+c+d-i)!} p^i (1-p)^{a+b+c+d-i}$$

There was proved [7] that each Σ-double implication quantifier $\Leftrightarrow^*$ or Σ-equivalency quantifier $\equiv^*$ is a member of some triad of quantifiers.

4 Triads of ratio-quantifiers

This is one of the main properties of the basic implicational quantifier: the greater the ratio a/b, the greater the value of the quantifier. This property is stronger than that used in the definition of implicational quantifiers. Therefore we introduced a subclass of implicational quantifiers with this property [6]:

A quantifier $\sim (a,b)$ is **ratio-implicational**, if $\sim (a',b') \geq \sim (a,b)$ when $a'b \geq ab'$. For any $\theta > 0$ the following quantifier is ratio-implicational:

$$\Rightarrow_\theta (a,b) = \frac{a}{a+\theta b}.$$

It is clear that each ratio-implicational quantifier is also implicational, but the class of ratio-implicational quantifiers is a proper subclass of the class of implicational quantifiers (the quantifier $\Rightarrow_p^?$ is a counterexample).

In the case when the starting quantifier $\Rightarrow^*$ is ratio-implicational, we obtain the corresponding **triad of ratio-quantifiers** $\Rightarrow^*, \Leftrightarrow^*, \equiv^*$. There are some further useful connections inside the triad of ratio-quantifiers, namely the following inequalities (proved in [9]):

For all a, b, c, d the value $\equiv^* (a,b,c,d)$ lies both

1. between the values $\Rightarrow^* (a,b)$, and $\Rightarrow^* (d,c)$;
2. between the values $\Rightarrow^* (a,c)$, and $\Rightarrow^* (d,b)$.

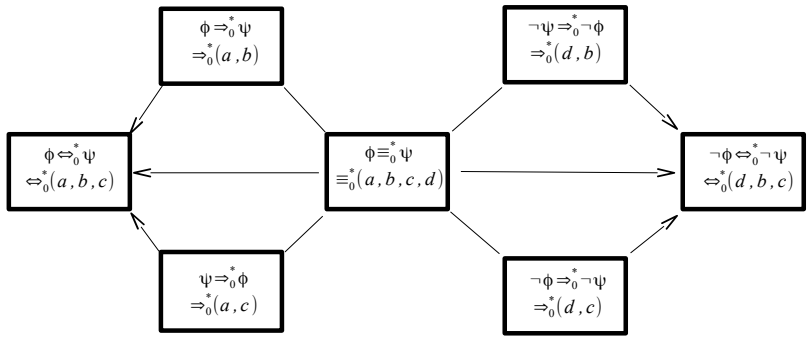

Figure 1: Relations among values of a triad of ratio-quantifiers on a four-fold table $< a, b, c, d >$.

5 Discussion of possible truth configurations

Let $\Rightarrow^*, \Leftrightarrow^*, \equiv^*$ be a triad of ratio-quantifiers. Assume some truth threshold t from $[0, 1]$ is given. A formula $\alpha \sim \beta$ is treated as true in data if the value of the quantifier $\sim$ in the four-fold table $< a, b, c, d >$ corresponding to the attributes α, β is greater or equal to t. Using inequalities presented in the previous text (see Figure 1), we discussed in [9] possible truth configurations of the set of formulae

Implications:	$\varphi \Rightarrow^* \psi$,	$\psi \Rightarrow^* \varphi$,	$\neg\varphi \Rightarrow^* \neg\psi$,	$\neg\psi \Rightarrow^* \neg\varphi$;
Double-implications:	$\varphi \Leftrightarrow^* \psi$,	$\neg\varphi \Leftrightarrow^* \neg\psi$;		
Equivalency:	$\varphi \equiv^* \psi$.			

There are formally $2^7 = 128$ configurations, but we concluded that most of them are not possible in any data (there are only 27 possible truth configurations).

6 Conclusions

Implicational, Σ-double implicational, and Σ-equivalence quantifiers compose logically affiliated triads $\Rightarrow^*$, $\Leftrightarrow^*$, $\equiv^*$, where $\Rightarrow^*$ is some implicational quantifier, $\Leftrightarrow^*$ is the least strict Σ-double implicational quantifier corresponding to $\Rightarrow^*$, and $\equiv^*$ is the most strict Σ-equivalence quantifier corresponding to $\Leftrightarrow^*$.

The best known example is the triad of basic quantifiers $\Rightarrow_{\oslash}$, $\Leftrightarrow_{\oslash}$, $\equiv_{\oslash}$.

Let us stress that to each given implicational quantifier, such a triad can be constructed. This can naturally extend the particular quantifier's definition

for covering all three types of relations (implication, double implication, equivalence). Moreover, proven deduction rules and inequalities among the values of affiliated quantifiers in data can be used in various ways in data-mining procedures oriented to association rules. For the subclass of ratio-implicational quantifiers, there was demonstrated that a number of possible truth-configurations of implication, double implication, equivalence in data (given some threshold) is significantly reduced. The approach presented in the paper can serve to formulate new data-mining tasks seeking for both asymmetric and symmetric association rules in an unified way or to filter sets of association rules for more user-oriented outputs of data-mining procedures.

References

[1] Hájek, P. (1998): *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht 1998. 297 p.

[2] Hájek, P., Havránek, T. (1978): *Mechanising Hypothesis Formation - Mathematical Foundations for a General Theory*. Springer-Verlag, Berlin 1978, 396 p.

[3] Hájek, P., Havránek, T., Chytil M. (1983): *Metoda GUHA*. Academia, Praha 1983, 314 p. (in Czech)

[4] Iványek, J. (1999): On the Correspondence between Classes of Implicational and Equivalence Quantifiers. In: *Principles of Data Mining and Knowledge Discovery*. Proc. PKDD'99 Prague (Zytkow,J. and Rauch,J., eds.), Springer-Verlag, Berlin 1999, pp. 116-124.

[5] Iványek, J. (2004): Using Fuzzy Logic Operators for Construction of Data Mining Quantifiers. *Neural Network World* 5/04, 2004, pp. 403-410.

[6] Iványek, J. (2005): Construction of Implicational Quantifiers from Fuzzy Implications. *Fuzzy Sets and Systems* 151, 2005, pp. 381-391

[7] Iványek, J. (2005): Strictness of Four-fold Table Quantifiers in Fuzzy Logic. In: *Fuzzy Logic, Soft Computing, and Computational Intelligence II. 11th IFSA World Congress*. Tsinghua University Press and Springer, Beijing, China 2005, pp. 1239-1243

[8] Iványek, J. (2006): Combining Implicational Quantifiers for Equivalence Ones by Fuzzy Connectives. *International Journal of Intelligent Systems* 21, 2006, pp. 325-334.

[9] Iványek, J. (2012): Affiliated Ratio-implicational and Equivalency Data-mining Quantifiers and their Truth Configurations. In: *9th Workshop on Uncertainty Processing*. Eds. Kroupa, T. - Vejnarová, J., Univ. of Economics, Prague 2012, pp. 82-89

[10] Iványek, J. (2012): Some Properties of Evaluated Implications Used in Knowledge-based Systems and Data-mining. *Journal of Systems Integration* 3, No 3, 2012, pp. 17-23.

[11] Iványek, J. (2012): Compound Quantification of Implications, Double-implications and Equivalency in Four-fold Tables. *IBCSS* No 2012/4, pp. 1-14.

[12] Pavelka, J. (1979): On Fuzzy Logic I, II, III. *Zeischr. f. Math. Logik und Grundl. der Math.* 25, 1979, pp. 45-52, 119-134, 447-464

[13] Rauch, J. (1998): Classes of Four-Fold Table Quantifiers. In: *Principles of Data Mining and Knowledge Discovery* (Quafafou,M. and Zytkow,J., eds.), Springer Verlag, Berlin 1998, pp. 203-211.

[14] Rauch, J. (2005): Logic of Association Rules. *Applied Intelligence* 22 (1), 2005, pp. 9-28

[15] Rauch, J. (2013): *Observational Calculi and Association Rules.* Studies in Computational Inteligence 469. Springer 2013, 296 p.

Three-Person Bargaining Problems

Milan Vlach
Theoretical Computer Science and Mathematical Logic
Faculty of Mathematics and Physics, Charles University
Malostranské náměstí 25, 180 00 Prague 1, Czech Republic
e-mail: mvlach@ksi.ms.mff.cuni.cz
and
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Prague 6, Czech Republic

Abstract

The paper deals with conflict situations in which a finite number of individuals (players) can reach a consensus about choosing an element from a fixed set of feasible alternatives. If no unanimous agreement is reached, then some specified (disagreement) alternative, not necessarily belonging to the set of feasible alternatives, is the result of bargaining. In other words, every player can veto any alternative that is different from the disagreement one.

There are two main streams of research in the literature: one approach develops axiomatizations that define a unique solution, while the second one constructs a non-cooperative game or a sequence of such games whose solutions (usually Nash's equilibria) are related to solutions provided by axiomatization. We show that, in the three-player situation, some of the standard axiomatizations have several distinctive features by which they essentially differs from n-player situations with $n \neq 3$. Then we conclude with presenting problems for further research.

Keywords: Bargaining problems, point-solutions, cooperative games, non-transferable utility

1 Introduction

Throughout the paper, we use the following notation. For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ from $\mathbb{R}^n$, we write $x < y$ and $x \leq y$ if, respectively, $x_i < y_i$ and $x_i \leq y_i$ for each i from $\{1, 2, \ldots, n\}$. If $x \leq y$ and $x \neq y$, then we write $x \prec y$. The relations $>, \geq$, and $\succ$ between elements of $\mathbb{R}^n$ are defined analogously. The sets $\{x \in \mathbb{R}^n : x \geq 0\}$ and $\{x \in \mathbb{R}^n : x > 0\}$ are denoted by $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$, respectively. If A is a subset of $\mathbb{R}^n$ and x is a point in $\mathbb{R}^n$, then we denote the sets $\{a + x : a \in A\}$ and $\{a - x : a \in A\}$ by $A + x$ and $A - x$, respectively. Similarly,

if λ is a real number, we define λA as the set $\{\lambda a : a \in A\}$. Moreover, we define the sets A_x^+ and A_x^{++} by $A_x^+ = \{y \in A : y \geq x\}$ and $A_x^{++} = \{y \in A : y > x\}$, respectively.

Formally, by an *n-player bargaining problem* we understand a nonempty collection of pairs (S, d) where S is a nonempty subset of $\mathbb{R}^n$ and d is a point in $\mathbb{R}^n$. Often the elements of S are interpreted as utility n-tuples that the players can obtain by reaching a unanimous agreement, and d as the outcome when the players do not reach agreement. Following this interpretation we call elements of S *feasible alternatives* and refer to d as the *disagreement alternative*. Notice that if d belongs to S, then the disagreement alternative is feasible. In other words, the players can agree to disagree.

Let $\mathcal{B}$ be an n-player bargaining problem. A *bargaining solution* for $\mathcal{B}$ is a mapping f from $\mathcal{B}$ to $\mathbb{R}^n$ such that, for each instance (S, d) of $\mathcal{B}$, the value[1] $f(S, d)$ of f belongs to $S \cup \{d\}$. The value $f(S, d)$ of bargaining solution f is called *f-solution* (or simply *solution*) for instance (S, d). If (S, d) is an instance of a bargaining problem, then we say that S_d^+ is the *individually rational part* of (S, d), and S_d^{++} is the *strict individually rational part* of (S, d). An instance (S, d) is said to be *comprehensive* if, for each $x \in S$, the set $\{y : y \leq x\}$ is included in S. For each i, the maximum of the function $(x_1, \ldots, x_n) \mapsto x_i$ over the individually rational part of (S, d) will be denoted by $m_i(S, d)$, and the point $m(S, d) = (m_1(S, d), \ldots, m_n(S, d))$ is called the *utopia point* (or *bliss point* or *ideal point*) for instance (S, d).

Roughly speaking, there are two basic streams of research in the literature: cooperative and non-cooperative. In the former, one develops an axiomatization that defines a unique bargaining solution for $\mathcal{B}$ implicitly by requiring fulfillment of a set of properties, while in the latter, one formalizes a bargaining process (for each instance of $\mathcal{B}$) explicitly through a non-cooperative game in the extensive form.

Here, we shall deal only with the cooperative approach. First we recall the classical bargaining problem and bargaining solution proposed by Nash [2] for two-player problems. Then, for the same problem, we present some of other bargaining solutions together with their axiomatizations. In particular, we introduce the Raiffa solution and Kalai-Smorodinsky solution. We then continue by discussing questions related to bargaining problems involving more than two players. It turns out that a satisfactory extensions of some solution concepts to problems with more than two players require modification of original Nash's problem. We focus attention to the three-player problems because usually, but not always, the essential differences clearly appear already in the three-player problems. We conclude with discussing problems for further research.

2 Two Players

In the cooperative approach, a particular bargaining solution of a bargaining problem can be defined either explicitly (for example, as a result of some clearly

[1] We write $f(S, d)$ instead of $f((S, d))$.

stated procedure) or implicitly (for example, by a system of clearly stated requirements or axioms). To illustrate the difference between these two basic ways, we first consider Nash's two player bargaining problem.

2.1 The Nash bargaining problem

The two-player bargaining problem introduced by Nash [2], denoted here by $\overline{\mathcal{B}}$, is composed of instances (S, d) such that:

- The set S of feasible alternatives is a nonempty convex compact subset of $\mathbb{R}^2$.
- The disagreement alternative $d = (d_1, d_2)$ belongs to S.
- There is at least one alternative $x = (x_1, x_2)$ in S such that $x_1 > d_1$ and $x_2 > d_2$.

The *Nash bargaining solution* for the Nash bargaining problem $\overline{\mathcal{B}}$ is defined as the mapping f from $\overline{\mathcal{B}}$ to $\mathbb{R}^2$ whose value $f(S, d)$ is the maximizer of the product $(x_1 - d_1)(x_2 - d_2)$ over the individually rational part of (S, d). It is easy to see that the Nash solution for $\overline{\mathcal{B}}$ is well defined:

(i) The existence of a maximizer is guaranteed by the continuity of function $(x_1, x_2) \mapsto (x_1 - d_1)(x_2 - d_2)$ and compactness of the individually rational part of (S, d).

(ii) The uniqueness is guaranteed by convexity of S and strict quasi-concavity of function $(x_1, x_2) \mapsto (x_1 - d_1)(x_2 - d_2)$ on the strict individually rational part of (S, d).

As an illustration consider the instance (S, d) depicted in Figure 1, where $d = (0, 0, 0)$ and $S = \{(x_1, x_2) \in \mathbb{R}^2_+ : (x_1, x_2) \leq (x_1, \sqrt{(1 - x_1)}, 0 \leq x_1 \leq 1\}$. See [3] for an interpretation in terms of utility functions.

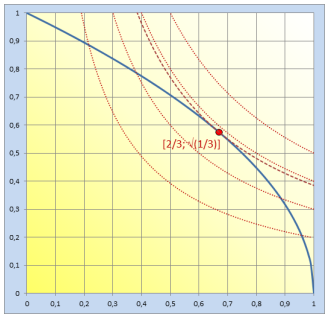

Figure 1: The Nash solution

Nash characterized this solution by proving that it is the only bargaining solution for $\overline{\mathcal{B}}$ satisfying the following four conditions.

1. WEAK PARETO OPTIMALITY. For each instance (S, d) of $\overline{\mathcal{B}}$, if x and y are in S and $y > x$, then $f(S, d) \neq x$.

2. SYMMETRY. For each instance (S, d) of $\overline{\mathcal{B}}$ such that $d_1 = d_2$ and (x_2, x_1) is in S whenever (x_1, x_2) is in S, we have: $f_1(S, d) = f_2(S, d)$.

3. SCALE INVARIANCE. If the instance (T, b) is obtained from (S, a) by the transformations $x_1 \mapsto \alpha_1 x_1 + \beta_1$ and $x_2 \mapsto \alpha_1 x_2 + \beta_2$ with positive α_1 and α_2, then $f_1(T, b) = \alpha_1 f_1(S, a) + \beta_1$ and $f_2(T, b) = \alpha_2 f_2(S, a) + \beta_2$.

4. INDEPENDENCE OF IRRELEVANT ALTERNATIVES. For every pair (S, d) and (T, d) of instances in $\overline{\mathcal{B}}$ such that $S \subseteq T$, we have: if $f(T, d)$ belongs to S, then $f(T, d) = f(S, d)$.

It is worth noting the difference between the first two conditions and remaining conditions. The conditions of Pareto optimality and Symmetry act on single instances, while Scale invariance and Independence of irrelevant alternatives require some kind of consistency across instances. In addition, if we modify the problem by reducing $\overline{\mathcal{B}}$ to the subset of $\overline{\mathcal{B}}$ consisting only from the symmetric instances[2], then every bargaining solution satisfying Pareto optimality and Symmetry coincides with the Nash bargaining solution for the reduced problem.

2.2 Other Solutions

It is well known that none of Nash's four requirements is superfluous in the sense that, for each trio of these conditions, there exists a bargaining solution for $\overline{\mathcal{B}}$ which is different from the Nash solution. In this subsection we consider two solutions for $\overline{\mathcal{B}}$ different from the Nash solution that satisfy all Nash's conditions except Independence of irrelevant alternatives. For a detailed survey of solutions for the Nash bargaining problem, see [6] or [11].

2.2.1 The Kalai-Smorodinsky solution

The *Kalai-Smorodinsky bargaining solution* for $\overline{\mathcal{B}}$ is the function that assigns to instance (S, d) from $\overline{\mathcal{B}}$ the maximal point (with respect to $\geq$ in $\mathbb{R}^2$) on the straight line joining d and the utopia point for (S, d); see Figure 2. It turns out that the Kalai-Smorodinsky solution is the only bargaining solution for $\overline{\mathcal{B}}$ satisfying Pareto optimality, Symmetry, Scale invariance and the following condition of monotonicity.

5. INDIVIDUAL MONOTONICITY. For every pair $(S, d), (T, d)$ of instances in $\overline{\mathcal{B}}$ with the same utopia point, we have: if $S \subseteq T$, then $f(S, d) \leq f(T, d)$.

Note that, for the domain $\overline{\mathcal{B}}$, this result implies that the Nash solution does not satisfy the condition of Individual monotonicity, and the Kalai-Smorodinsky solution does not satisfy Independence of irrelevant alternatives.

[2] An instance (S, d) is *symmetric* if $d_1 = d_2$ and $(x_1, x_2) \in S$ if and only if $(x_2, x_1) \in S$.

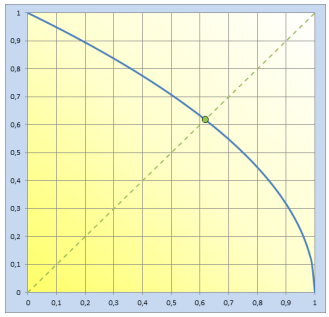

Figure 2: The Kalai-Smorodinsky solution

2.2.2 The discrete Raiffa solution

Raiffa [4] and [5] (see also [1]) proposed several bargaining solutions for the two-person Nash bargaining problem. One of them, called the *discrete (or sequential) Raiffa solution*, is defined as the limit of sequence $\{x^k\}$ of points from S generated as follows: Let (S, d) be an instance of $\overline{\mathcal{B}}$. Set $x^0 = d$, and continue by defining x^{k+1} as the middle point of the line segment connecting the points $(x_1^k, m_2(S, x^k))$ and $(m_1(S, x^k), x_2^k)$. See Figure 3, which is taken from [12].

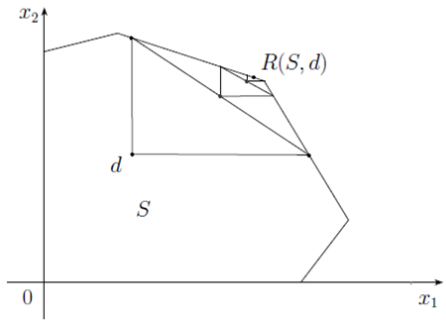

Figure 3: The discrete Raiffa solution

The convexity of S and the definitions of functions m_1, m_2 guarantee that, for each k, the point x^k belongs to S and $x^{k+1} \geq x^k$. Because S is compact, we know that the sequence $\{x^k\}$ is convergent and its limit belongs to the Pareto subset[3] of S.

[3]If the Pareto subset of S is piecewise linear, the convergence may be finite.

3 Three Players

It is not difficult to verify that the two-player Nash bargaining problem and solution can easily be extended to situations involving more than two players. By means of an there-player example (see Figures 4 and 5) we show that the situation is not so clear for the Kalai-Smorodinsky and discrete Raiffa solution.

The Nash solution If there are three players, then the Nash bargaining problem $\overline{\mathcal{B}}$ consists of ordered (S, d) pairs in which S is a nonempty compact convex subset of $\mathbb{R}^3$, $d \in S$, and there is $x \in S$ such that $x > d$. The Nash bargaining solution is then defined as the mapping that assigns to each instance (S, d) the maximizer of the product $(x_1 - d_1)(x_2 - d_2)(x_3 - d_3)$ over the individually rational part of (S, d). It can be shown that this solution is the only one that satisfies the three-dimensional analogues of Nash's conditions.

The Kalai-Smorodinsky solution Figure 4 shows that there exists an instance in the three-dimensional domain $\overline{\mathcal{B}}$ for which the direct extension of Kalai-Smorodinsky bargaining solution is quite uninteresting: it is not Pareto optimal, it is equal to the disagreement alternative.

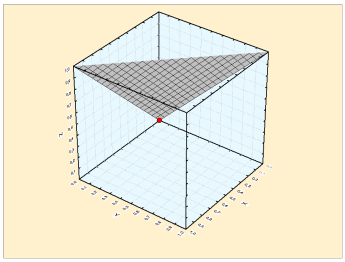

Figure 4: The Kalai-Smorodinsky solution

Discrete Raiffa Solution Figure 5 shows that there exists an instance in the three-dimensional domain $\overline{\mathcal{B}}$ for which the direct extension of discrete Raiffa solution is not a single-point bargaining solution. each point in the depicted segment in the Pareto subset of S can be a solution.

We can remedy the problem with the Kalai-Smorodinsky and discrete Raiffa solution by changing the domain $\overline{\mathcal{B}}$ of the bargaining problem to problems consisting from instances (S, d) which have the following properties: The disagreement alternative belongs to S, and S is closed, convex, comprehensive, positively bounded and such that all its boundary points are Pareto optimal. To distinguish this domain from $\overline{\mathcal{B}}$, we shall denote it by $\mathcal{B}^*$. Then the natural extensions of the Kalai-Smorodinsky solution and the discrete Raiffa solution satisfy Pareto optimality, Symmetry and Scale invariance, and their axiomatizations differ from the Nash one only in the requirement of Independence of irrelevant

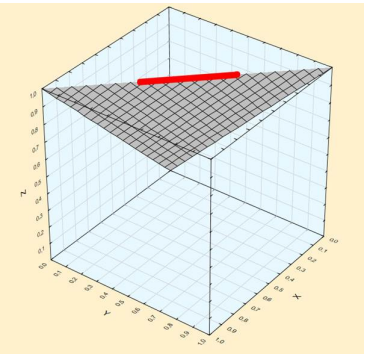

Figure 5: The discrete Raiffa solution

alternatives. Moreover, this modification allows for new class of bargaining solutions that are impossible in two-player problems. Namely, the solutions which satisfy the strengthening of the condition of Scale invariance to so called Ordinal invariance; that is, the invariance with respect strictly increasing transformation of individual utilities. This is certainly of interest because of attempts to build economic theory on ordinal preferences.

The Shapley-Shubik solution We mentioned in the previous section that there are no interesting ordinal solutions for the two-player Nash problem even if $\overline{\mathcal{B}}$ is changed to $\mathcal{B}^*$. This is easy to show by considering the instance (S, d) of $\mathcal{B}^*$ with $d = 0$ and $S = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$. Let T be the transformation (taken from [6]) of the individual utilities defined by $T(x_1, x_2) = \left(\frac{2x_1}{1+x_1}, \frac{x_2}{2-x_2}\right)$. It can easily be verified that T preserves utility orderings of both players on the unit square $Q = \{(x_1, x_2) : 0 \leq x_1 \leq 1,\ 0 \leq x_2 \leq 1\}$, and that it maps the set S onto itself. It follows that each solution function f on $\mathcal{B}^*$ which is invariant with respect to ordinal transformations must assign to the instance $(S, 0)$ an alternative in S which is also a fixed point of mapping T on Q. However, the only fixed points of T on Q are points $(0, 0), (0, 1), (1, 0), (1, 1)$. The alternative $(1, 1)$ is infeasible and the remaining points are uninteresting: $(0, 0)$ is the disagreement alternative, and alternatives $(0, 1)$ and $(1, 0)$, are so called dictatorial solutions.

The main argument of this proof cannot be extended to Nash's bargaining problem with three players, and non-dictatorial ordinally invariant solution exist for problems with more than two players. One such solution has been proposed by Shapley and Shubik already in [9], see also [10]. Recently Safra and Samet [7], [8] have generalized the Shapley-Shubik solution for problems with more than three players and demonstrated that there is even a continuum of ordinal solutions that are simultaneously Pareto optimal and symmetric.

The construction of ordinal solutions for three player problems is based

on the observation that if Q is a Pareto surface in $\mathbb{R}^3$ and (a_1, a_2, a_3) is a point in $\mathbb{R}^3 \setminus Q$, then there is a unique point (b_1, b_2, b_3) such that the points $(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)$ belong to S. Using this fact, one can define the Shapley-Shubik solution outcome for an instance (S, d) as the limit of sequence $\{x^k\}$ of points defined by setting $x^0 = d$ and defining x^k to be the unique point $x^{k+1} = (x_1^{k+1}, x_2^{k+1}, x_3^{k+1})$ determined by the property that the points $(x_1^k, x_2^{k+1}, x_3^{k+1}), (x_1^{k+1}, x_2^k, x_3^{k+1}), (x_1^{k+1}, x_2^{k+1}, x_3^k)$ belong to the Pareto surface of S.

Modification of the discrete Raiffa solution It is of interest to combine the idea of the Shapley-Shubik solution and the discrete Raiffa solution to obtain the solution to an instances (S, d) from $\mathcal{B}^*$ as the limit of the sequence $\{y^k\}$ of points from S defined as follows. Again we set $y^0 = d$. Let (x_1, x_2, x_3) be the point obtained from y^0 by one step of the Shapley-Shubik procedure. Then we construct next point y^1 by the same averaging that is used in the Sequential Raiffa procedure, but using points $(y_1^0, x_2, x_3), (x_1, y_2^0, x_3), (x_1, x_2, y_3^0)$ instead of using the points $(m_1(S, y^0), y_2^0, y_3^0), (y_1^0, m_2(S, y^0), y_3^0), (y_1^0, y_2^0, m_3(S, y^0))$. Then we continue in the same way, that is, we construct y^{k+1} from y^k as follows. First we use the fact that there is a unique point (x_1, x_2, x_3) such that the points $(y_1^k, x_2, x_3), (x_1, y_2^k, x_3), (x_1, x_2, y_3^3)$ belong to S and set $y^{k+1} = \frac{1}{3}((y_1^k, x_2, x_3) + (x_1, y_2^k, x_3) + (x_1, x_2, y_3^k))$.

Again the convexity of S guarantee that, for each k, the point y^k belongs to S and $y^{k+1} \geq y^k$. Because the set S is compact, we know that the sequence $\{y^k\}$ converges to a point in S; in fact to a point on the Pareto surface of S. However, this bargaining solution no more satisfies ordinal invariance.

Notes and Comments. It would be interesting to compare the proposed modification of the Raiffa procedure with other available procedures on some standard classes of problems. For example, our experience suggests it is substantially faster than the Shapley-Shubik and discrete Raiffa procedures when applied to instances from $\mathcal{B}^*$. Constructing extensions to problems with more than three players[4] would also be of interest. However, as the main open question we consider the problem of establishing systems of axioms that define the proposed solution uniquely on reasonable classes of bargaining problems.

4 Acknowledgment

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 žDYME – Dynamic Models in Economics.

[4]The procedure can be modified to be applicable to the two player problems, but then it is identical to the Raiffa procedure.

References

[1] Luce, R. D., Raiffa, H. (1957) *Games and Decisions: Introduction and Critical Survey*, Wiley, New York.

[2] Nash, J. (1950), The Bargaining Problem, *Econometrica* 18, 155–162.

[3] Peters, H. (2008) *Game Theory: A multi-leveled approach*, Springer, Berlin.

[4] Raiffa. M. (1950), Arbitration schemes for generalized two-person games. *Report No. M720-1, R30*, Engineering Research Institute, University of Michigan, Ann Arbor.

[5] Raiffa. H. (1953), Arbitration Schemes for Generalized Two Person Games, *Contributions to the Theory of Games II*, (H.W. Kuhn and A.W. Tucker eds.), 361–387, Princeton University Press, Princeton, New Jersey.

[6] Roth, A. E. (1979) *Axiomatic models of bargaining*, Springer-Verlag, Berlin.

[7] Safra, Z., Samet, D. (2004), An ordinal solution to bargaining problems with many players, *Games and Economic Behavior*, 46, 129–142.

[8] Samet, D., Safra, Z. (2005), A family of ordinal solutions to bargaining problems with many players, *Games and Economic Behavior* 50, 89–106.

[9] Shapley, L.S., Shubik, M. (1974), Game Theory in Economics – Chapter 4: Preferences and Utility, *R-904/4-NSF*, The Rand Corporation.

[10] Shubik, M. (1982, 1984) *Game Theory in the Social Sciences*, Cambridge, Mass. The MIT Press.

[11] Thomson, W. (2009), Bargaining and the theory of cooperative games: John Nash and beyond, *Working Paper* No. 554, University of Rochester.

[12] Trockel, W. (2009), An axiomatization of the sequential Raiffa solution, *IMW working paper* No. 425, Bielefeld University.

Role of Pavlov-like Strategy on Emergence of Cooperation in Demographic Donor-Recipient Game

Tsuneyuki Namekata
Information and Management Science
Otaru University of Commerce
e-mail: namekata@res.otaru-uc.ac.jp

Yoko Namekata
e-mail: PallaYoko@namekata.org

Abstract

We deal with Pavlov-like strategy as well as Tit for Tat-like strategy in Demographic Donor-Recipient (DR) game. We study the role of Pavlov-like strategy on the emergence of cooperation by Agent-Based Simulation.

We extend Tit for Tat (TFT) and Pavlov (Pav) up to three states from two and call them TFT-like and Pavlov-like strategy, respectively. Unlike TFT-like, Pav-like has the following feature: Pav-like changes to using C from using D or remains in using D if he is using D and experiences opponents' D's or C's, respectively. Thus we expect that some Pavlov-like strategies in the population may soften the tendency toward defection of the whole population and also the tendency toward full cooperation of the whole population. Although sole Pavlov-like strategy is not so effective to promote the cooperation, we found case where the cooperation emerges more frequently with both TFT-like and Pavlov-like strategy than with sole TFT-like (or Pav-like) strategy.

Keywords: Pavlov, Donor-Recipient game, emergence of cooperation, generalized reciprocity, Agent-Based Simulation

1 Introduction

This paper investigates the role of Pavlov-like strategy on the emergence of cooperation in Demographic DR game.

Epstein[1] introduces demographic model. He shows the emergence of cooperation where AllC and AllD are initially randomly distributed in a square lattice of cells. Here AllC always Cooperates and AllD always Defects. In each period, players move locally and play Prisoner's Dilemma (PD) game against local player(s). If wealth (accumulated payoff) of a player becomes negative or his age becomes greater than his lifetime, he dies. If his wealth becomes greater

than some amount and there is an unoccupied local cell, he has an offspring and gives the offspring some amount from his wealth.

Namekata and Namekata[2] extend Epstein's original model discussed above by introducing global move, global play, Reluctant players, who delay replying to changes and use extended forms of TFT, into demographic PD game and consider the effect of Reluctant players on the emergence of cooperation, and show cases where the reluctance promotes the emergence of cooperation. Here TFT Cooperates at first encounter and at later encounters uses the same move as the opponent did in the previous encounter. Namekata and Namekata[3] examine the effect of move-play pattern on the emergence of cooperation and the distribution of strategies. They restrict patters of move and play of a player to simple structure; local or global, where local or global means that with high probability the player moves (plays) locally or globally, respectively. For example, a player with global move and local play (abbreviated as gl) moves globally with high probability and plays DR games against (possibly different) local opponents with high probability at each period. They show that cooperative strategies evolutionarily tend to move and play locally, defective strategies do not, and AllC and AllD are abundant unless all strategies initially play locally.

Nowak and Sigmund[4] consider the emergence of cooperation in infinitely repeated PD game. Population consists of strategies that depend on one's own move as well as the opponent's at the last encounter, i.e., $(p_{\mathrm{CC}}, p_{\mathrm{CD}}, p_{\mathrm{DC}}, p_{\mathrm{DD}})$ where p_{XY} is the probability with which C is used at this encounter given that the outcome of the last encounter is XY. They do not use Demographic model. Players play *infinitely* repeated PD game at each period instead of one-shot PD game against randomly selected opponent. The frequency of each strategy in population at the next period is proportional to its payoff at this period. They show that not TFT but Pavlov $(0.999, 0.001, 0.007, 0.946)$ is most abundant strategy in the population in the long run. They argue that Pavlov's success is based on the following two advantageous features compared with TFT in *infinitely* repeated PD game: (1) Pavlovs can correct inadvertent defection and return to mutual cooperation. (2) Pavlov can exploit AllC.

We deal with Pavlov-like strategy as well as Tit for Tat-like strategy in *finitely* repeated Demographic Donor-Recipient (DR) game. Pavlov (Pav) is known to be one of the basic strategies in dilemma situations as well as Tit for Tat (TFT). TFT and Pav have two inner states whose label C(ooperate) or D(efect) indicates their current move. The state in the next encounter is determined based on the current opponent's move, differently between TFT and Pav. The next state of TFT is the immediate neighbor of the current state toward C or D (if possible) in case of the current opponent's C or D, respectively. On the other hand, that of Pav remains the same if the current opponent uses C or is changed from C to D (or from D to C) if the current opponent uses D, respectively. Alternatively Pav is described as follows: Win Stay, Lose Shift, that is, Pav remains in the same move if he feels comfortable, whereas Pav changes his move if he feels uncomfortable, because we configure the payoff matrix so that the payoff is positive if the opponent uses C or negative if the opponent uses D.

In this paper, we extend TFT and Pav up to three states and call them TFT-like and Pavlov-like strategy, respectively. Pavlov-like changes to using C from using D or remains in using D if he is using D and experiences opponents' D's or C's, respectively. Thus we expect that some Pavlov-like strategies in the population may soften the tendency toward defection of the whole population and also the tendency toward full cooperation of the whole population. We examine initial distribution of strategies that promote the emergence of cooperation and study the role of Pav-like strategy on the emergence of cooperation.

2 Model

We start with extending TFT and Pav as follows in order to introduce TFT-like and Pavlov-like (Pav-like) strategy. The idea is to introduce reluctance to immediate reply to its opponent's change: Let $m = 0, \ldots, n; t = 0, \ldots, m+1; s = 0, \ldots, m$. Strategy $(m, t; s)X$ is illustrated in Figure 1 where X is T for TFT-like or P for Pav-like. It has $m+1$ inner states. The inner states are numbered $0, \ldots, m$; thus m is the largest state number. State i is labeled D_i if $i < t$ or C_i if not. If current state is labeled C or D, then the strategy prescribes using C or D, respectively. In other words, the strategy prescribes using D if the current state $i < t$ and using C if not; thus the value t is the threshold which determines the move of a player. Initial state in period 0 is state s; its label is D_s if $s < t$ or C_s if not. If current state is i, then the next state of TFT-like is $\min\{i+1, m\}$ or $\max\{i-1, 0\}$ given that the opponent uses C or D, respectively, in this encounter. If current state is i and $i \geq t$, then the next state of Pav-like is $\min\{i+1, m\}$ or $\max\{i-1, 0\}$ given that the opponent uses C or D, respectively, in this encounter. If current state is i and $i < t$, then the next state of Pav-like is $\max\{i-1, 0\}$ or $\min\{i+1, m\}$ given that the opponent uses C or D, respectively, in this encounter. Thus TFT-like and Pav-like strategies act differently if their current state $i < t$; TFT-like strategy in Defective state $(i < t)$ tends to use the same move as the opponents, whereas Pav-like in Defective state $(i < t)$ tends to use the opposite move as the opponents. If $m > 1$, then the strategy may delay replying to its opponent's change.

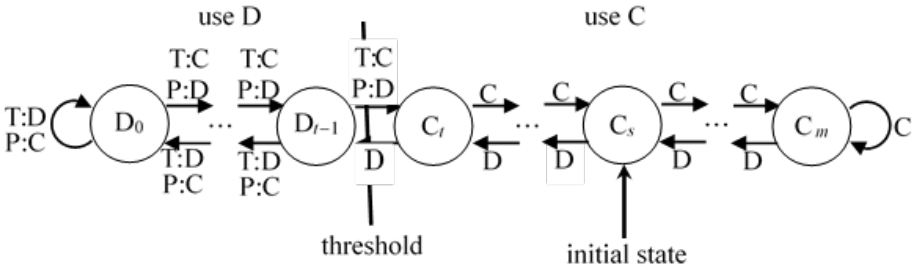

Figure 1: TFT-like and Pav-like strategies

Note that TFT or Pav is expressed as $(1,1;1)$T or $(1,1;1)$P, respectively, in this notation. Thus strategy $(m,t;s)X$ is an extended form of TFT or Pav. To sum up, our strategies are expressed as $(m,t;s)X$; m is the largest state number, t is the threshold, and s is the initial state number, X denotes TFT-like or Pav-like. We omit the initial state like $(m,t;*)X$ if it is determined randomly. We also omit the initial state like $(m,t)X$ if we have no need to specify it.

Table 1: Payoff Matrix of DR game ($b = 4.5$ and $c = 1$)

		Recipient
Donor	C	$-c-x, b-x$
	D	$-x, -x$

Table 2: Initial distribution of inheriting properties

property:	initial distribution
strategy:	We deal with 3 populations, T(x,m), TP(x,m), and P(x,m) for $x = 0.05$, $1/4$ or $1/6$ and $m = 4$ or ∞ as follows: T(x,m):= $\{x$AllD(m), $\frac{1-x}{2}(2,2;*)$T, $\frac{1-x}{2}(2,1;*)$T, xAllC$(m)\}$, TP(x,m):= $\{x$AllD(m), $\frac{1-x}{4}(2,2;*)$P, $\frac{1-x}{4}(2,2;*)$T, $\frac{1-x}{4}(2,1;*)$P, $\frac{1-x}{4}(2,1;*)$T, xAllC$(m)\}$, P(x,m):=$\{x$AllD(m), $\frac{1-x}{2}(2,2;*)$P, $\frac{1-x}{2}(2,1;*)$P, xAllC$(m)\}$, where AllC$(m) = (2,0)$T for $m = \infty$, AllC$(m) = (4,1;4)$T for $m = 4$, and AllD$(m) = (2,3)$T for $m = \infty$, AllD$(m) = (4,4;0)$T for $m = 4$. The notation, for example, of T(x,m), means that with probability x strategy AllC(m) is selected, with probability $\frac{1-x}{2}$ strategy $(2,1;*)$T is selected, and so on, where $*$ indicates that initial state is selected randomly. Note that initially 50% of players use C on the average since both AllC(m) and AllD(m) are included with the same probability and so are both $(m,t;*)X$ and $(m,m-t+1;*)X$. As reference populations, we also deal with All:=$\{0.5$AllD$(\infty), 0.5$AllC$(\infty)\}$ and All4:=$\{0.5$AllD$(4), 0.5$AllC$(4)\}$.
(rGM, rGP)**:**	We deal with distribution $\{0.25ll, 0.25lg, 0.25gl, 0.25gg\}$. For example, gl means rGM is distributed in interval g and rGP in interval l, where $l := (0.05, 0.2)$ and $g := (0.8, 0.95)$. $\{0.25ll, 0.25lg, 0.25gl, 0.25gg\}$ means rGM and rGP are selected randomly among ll, lg, gl, and gg.

Note that AllC is denoted by $(m,0)$T and AllD by $(m,m+1)$T. If m is large, $(m,1;m)$T and $(m,m;0)$T are very close to AllC and AllD, respectively. We use these pseudo-AllC $(m,1;m)$T and pseudo-AllD $(m,m;0)$T for $m = 4$ later in this paper because we want to relax unrealistic fixed move strategy.

We deal with Donor-Recipient (DR) game as a stage game. DR game is a two-person game where one player is randomly selected as Donor and the other as Recipient. Donor has two moves, Cooperate (C) and Defect (D). C means Donor pays cost c in order for Recipient to receive benefit b ($b > c > 0$). Defect means Donor does nothing. Recipient has no move. Since it is common in demographic dilemma game that the sum of payoffs of a player, in two successive games once as Donor and once as Recipient, to be positive if the opponent uses C and negative if D and the worst sum of a player is equal to the best sum

Table 3: Detailed Description of (1) Move and (2) Play

(1)	With probability rateOfGlobalMove (abbreviated as rGM), a player moves to random unoccupied cell in the whole lattice. If there is no such cell, he stays at the current cell. Or with probability $1 - rGM$, a player moves to random cell in von Neumann neighbors if it is unoccupied. If there is no such cell, he stays at the current cell.
(2)	With probability rateOfGlobalPlay (abbreviated as rGP), the opponent against whom a player plays dilemma game is selected at random from all players (except himself) in the whole lattice. Or with probability $1-rGP$, the opponent is selected at random from von Neumann neighbors (no interaction if none in the neighbors). This process is repeated 8 times. (Opponents are possibly different.)

in absolute value, we transform the original payoffs to new ones by subtracting constant x. Constant x is given by $x = \frac{b-c}{4}$. We set $b = 4.5$ and $c = 1$ in this paper. Table 1 shows the transformed payoff matrix of DR game. We assume that each player plays 8 games against (possibly different) players at each period.

In period $0, N(= 100)$ players (agents) are randomly located in 30-by-30 lattice of cells. The left and right borders of the lattice are connected. If a player moves outside, for example, from the right border, then he comes inside from the left border. So are the upper and lower borders. Players use strategies of $(m, t; s)X$ form. Initial wealth of every player is 6. Their initial (integer valued) age is randomly distributed between 0 and deathAge (= 50). In each period, each player (1st) moves, and (2nd) plays DR games given by Table 1 against other players. Positive payoff needs opponent's C. (The detailed description of (1st) move and (2nd) play is given in Table 3.) The payoff of the game is added to his wealth. If the resultant wealth is greater than fissionWealth (= 10) and there is an unoccupied cell in von Neumann neighbors, the player has an offspring and gives the offspring 6 units from his wealth. His age is increased by one. If the resultant wealth becomes negative or his age is greater than deathAge (= 50), then he dies. Then next period starts.

In our simulation we use synchronous updating, that is, in each period, all players move, then all players play, then all players have an offspring if possible. We remark that the initial state of the offspring's strategy is set to the current state of the parent's strategy. There is a small mutationRate (= 0.05) with which inheriting properties are not inherited. Initial distributions of inheriting properties given in Table 2 are also used when mutation occurs. We assume that with errorRate (= 0.05) a player makes mistake when he makes his move. Thus AllC may defect sometime. If population consists of AllC and AllD, $rGM = 0$, and $rGP = 0$, then our model is similar to that of Epstein[1]. His model uses asynchronous updating while our model uses synchronous updating.

3 Simulation and Results

We use Ascape (http://sourceforge.net/projects/ascape/) to simulate our model. We execute 300 runs of simulations in each different setting. We judge that the cooperation emerges in a run if there are more than 100 players and the average C rate (average Cr) is greater than 0.2 at period 500, where the average Cr at a period is the average of the player's Cooperation rate (Cr) at the period over *all* players and the player's Cr at the period is defined as the number of move C used by the player divided by the number of games played as Donor at the period. (We interpret 0/0 as 0.) This average Cr is the rate at which we see cooperative move C as an outside observer. Since negative wealth of a player means his death in our model and he has a lifetime, it is necessary for many players to use C in order that the population is not extinct. We focus on emergence rate of cooperation that is rate at which the cooperation emerges.

We are interested in cases where the cooperation emerges more frequently with both TFT- and Pavlov-like strategy than with sole Pavlov-like strategy and then than with sole TFT-like strategy. We examine how often the cooperation emerges in Demographic DR game with several different initial distributions of strategies. Pure AllC and AllD, and even with their low frequency 0.05 at period 0 prevent Pav-like strategy from promoting cooperation as shown in Table 4, e.g., .557 < .630 and .420 < .657. Table 4 shows emergence rate of cooperation Ce's for equal frequency at period 0 in the second row and for low 0.05 frequency in the third row, in pure AllC and AllD population. All column in Table 4 indicates pure AllC and AllD population defined in Table 2, T and P columns indicate the corresponding $x = 1/4$ and $x = 0.05$ population defined in Table 2 in case of $m = \infty$, and TP column indicates $x = 1/6$ and $x = 0.05$ population defined in Table 2 in case of $m = \infty$. In place of pure AllC and AllD, we use pseudo-AllC $(4, 1; 4)$T and pseudo-AllD $(4, 4; 0)$T with initial low frequency 0.05 at period 0. The emergence rate of cooperation Ce's and other related data are summarized in Table 5. Table 5 shows that Pav-like (.590) and TFT-like + Pav-like (.703) promote the cooperation in this order compared with All4 (.513). The third and fourth column in Table 5, aaCr's are the *actual* average Cr of pseudo-AllC $(4, 1)$T and pseudo-AllD $(4, 4)$T. aaCr of a strategy is defined as the average of players' Cr over all player using the strategy and playing *at least one* game as Donor. The sum of aaCr's of pseudo-AllC and pseudo-AllD, for example, 1.46 for All4, is much larger than 1 for All4 and T, but is almost equal to 1 for TP and P. We conclude that introducing pseudo-AllC and pseudo-AllD in place of pure AllC and AllD is reasonable modelling if there exists Pav-like strategy in

Table 4: Ce for pure AllC and AllD

$m = \infty$	All	T	TP	P
Ce(equal)	.473	.630	.557	.447
Ce($x = .05$)	.473	.657	.717	.420

Table 5: Ce, average actual Cr

$m = 4$	All4	T	TP	P
Ce($x = .05$)	.513	.487	.703	.590
aaCr$(4, 1)$T	.900	.931	.805	.804
aaCr$(4, 4)$T	.560	.636	.214	.222

the population.

Next we investigate the role of Pav-like strategy. We select 35 successful runs of T, TP and P populations, respectively, from data in Table 5. We trace average Cr from period 1 to period 500 at each successful run. We judge average Cr at a period is High(> 0.7) if it is greater than 0.7, or is Low(≤ 0.2) if it is less than or equal to 0.2, or is Middle otherwise. We see in Figure 2 that average Cr is almost High in T population, whereas it is mostly Middle in TP population and it is almost Middle in P population. Thus Pav-like makes average Cr Middle. We want to evaluate easily the change of average Cr over periods. We assign Low to 0 as a new vertical value different from the original value of average Cr, Middle to 1, and High to 2. Then we focus only on their local maximums and minimums. A transition of local optimums is classified into one of $\{-2, -1, 1, 2\}$. Suppose, for example, that local maximum is 2 at some period and the nearest local minimum is 1 at some later period, then the transition is evaluated as -1. We count all transitions of these local optimums over periods in each run. We show average of these number of transitions over 35 runs in Figure 3. We conclude that population T, TP, and P decreases the number of transitions in this order.

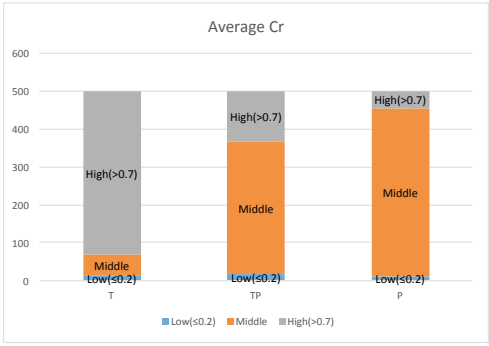

Figure 2: Average Cr

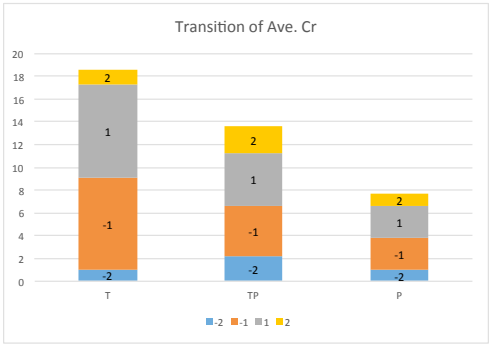

Figure 3: Transition of AveCr

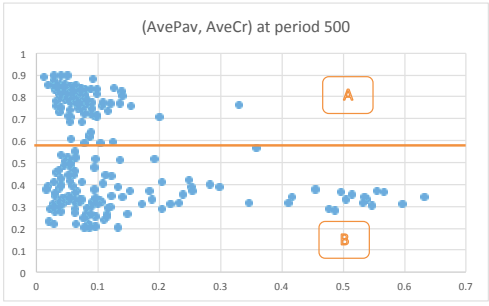

Figure 4: (AvePav,AveCr)

Next we concentrate on the average frequency of Pav-like strategy and the average Cr at period 500 in TP population. Figure 4 is scatter diagram of (average frequency of Pav-like strategy, average Cr) at period 500 of all successful runs in TP population. For convenience sake, let us divide all successful runs into two cases, A and B; A for average Cr $>= 0.57$, B for average Cr < 0.57. Figure 4 shows that the larger the average frequency of Pav-like strategy the smaller the average Cr at period 500 in TP population.

Figure 5 and 6 show the average distributions of strategies at period 500 for case A and B, respectively. We see that average frequency of Pav-like strategies, $(2,2)$P and $(2,1)$P, is not so large, around 0.15 even in case B.

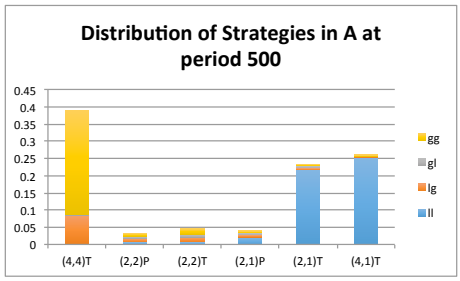

Figure 5: Distribution in A

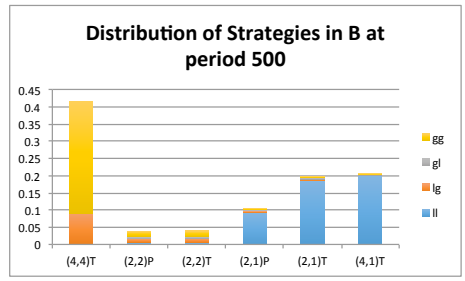

Figure 6: Distribution in B

4 Conclusion

We examine the role of Pav-like strategy on the emergence of cooperation in Demographic DR game by Agent-Based Simulation. We show that some Pav-like strategies promote cooperation and soften the tendency toward defection and toward full cooperation in whole population if there initially are low frequent pseudo-AllC and pseudo-AllD in stead of equal pure AllC and AllD.

References

[1] Epstein, J. M. (2006). "Zones of Cooperation in Demographic Prisoner's Dilemma." In: *Generative Social Science*. Princeton University Press, 199-221.

[2] Namekata, T. and Namekata, Y. 2011. "Effect of Reluctant Players in Demographic Prisoner's Dilemma Game." In: R. Bartak (ed.): *Proceedings of the 14th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty* (held in September 18-21, 2011, Hejnice, Czech Republic), 102-109.

[3] Namekata, T. and Namekata, Y. 2012. "Emergence of cooperation and patterns of move-play in Demographic Donor-Recipient Game." In: Masahiro Inuiguchi, Yoshifumi Kusunoki and Hirosaki Seki (eds.): *Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty* (held in September 24-27, 2012 Osaka, Japan), 51-58.

[4] Nowak, M. A. and Sigmund, K. (1993). "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game." *Nature*, No. 364, 56-58.

Additively and Multiplicatively Transitive Fuzzy Relations in Ranking of Alternatives

Jaroslav Ramík
School of Business Administration in Karviná
Silesian University in Opava
ramik@opf.slu.cz

Petr Korviny
School of Business Administration in Karviná
Silesian University in Opava
korviny@opf.slu.cz

Abstract

A fuzzy preference matrix is the result of pair-wise comparison a powerful method in multi-criteria optimization. When comparing two elements, the decision maker assigns the value between 0 and 1 to any pair of alternatives representing the element of the fuzzy preference matrix. Here, we investigate relations between transitivity and consistency of fuzzy preference matrices and multiplicative preference ones. The obtained results are applied to situations where some elements of the fuzzy preference matrix are missing. We propose a new method for completing fuzzy matrix with missing elements called the extension of the fuzzy preference matrix. We investigate some important particular cases of fuzzy preference matrix with missing elements. Consequently, by the eigenvector of the transformed matrix we obtain the corresponding priority vector. Illustrative numerical examples are supplemented.

1 Introduction

In many DM problems, procedures have been established to combine opinions about alternatives related to different points of view. These procedures are often based on pair-wise comparisons, in the sense that processes are linked to some degree of preference of one alternative over another, [3], [4], [6], [7], or [11]. The aim of this contribution is to investigate some important particular cases of fuzzy preference matrix with missing elements.

The paper is organized as follows. Multiplicative and additive preference relations (i.e. fuzzy preference relations) and their properties are introduced

in Section 2. In Section 3, based on inconsistency measurement of multiplicative preference matrix known from AHP, the inconsistency of fuzzy preference matrix is measured by the a-inconsistency index and a-inconsistency ratio as well as a-intransitivity index and a-intransitivity ratio. The priority vector for ranking the alternatives is derived by the classical Perron-Frobenius theory. In Secton 4, a special notation for the matrix with missing elements is introduced and the concept of the extension of fuzzy preference matrix with missing elements is defined. This concept is based on a particular representation of an a-consistent/a-transitive matrix and the elements of the extended matrix are calculated by the least squares method. In Section 5, two special cases of fuzzy preference matrix with missing elements are investigated. Here, the expert evaluates only $n-1$ pairs of alternatives. In this section, two numerical examples illustrating the necessary and sufficient conditions for elements to be evaluated in the pairwise comparison matrix are presented. In Section 6, some concluding considerations and remarks are presented.

2 Multiplicative and additive preferences

The DM problem can be formulated as follows. Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of alternatives. These alternatives have to be classified from best to worst, using the information given by a DM in the form of pairwise comparison matrix.

The preferences over the set of alternatives, X, may be represented in the following two ways: multiplicative and additive. Let us assume that the preferences on X are described by a preference relation on X given by a positive $n \times n$ matrix $A = \{a_{ij}\}$, where $a_{ij} > 0$ for all i, j indicates a preference intensity for alternative x_i to that of x_j, i.e. it is interpreted as "x_i is a_{ij} times better than x_j". The elements of $A = \{a_{ij}\}$ satisfy the following reciprocity condition [9].

A positive $n \times n$ matrix $A = \{a_{ij}\}$ is *multiplicative-reciprocal* (*m-reciprocal*), if

$$a_{ij}.a_{ji} = 1 \text{ for all } i, j. \tag{1}$$

A positive $n \times n$ matrix $A = \{a_{ij}\}$ is *multiplicative-consistent* (or, *m-consistent*) [9], if

$$a_{ik} = a_{ij}.a_{jk} \text{ for all } i, j, k \tag{2}$$

Here, $a_{ii} = 1$ for all i, and also (2) implies (1), i.e. an m-consistent matrix is m-reciprocal (however, not vice-versa).

Notice that $a_{ij} > 0$ and m-consistency is not restricted to the Saaty's scale $\{1/9, 1/8, ..., 1/2, 1, 2, ..., 8, 9\}$. Here, we extend this scale to the closed interval $[1/\sigma; \sigma]$, where $\sigma > 1$.

Sometimes it is more natural, when comparing x_i to x_j, that the decision maker assigns the value b_{ij} to x_i and b_{ji} to x_j, where $b_{ij} + b_{ji} = 1$. With this interpretation, the preferences on X can be understood as a *fuzzy preference relation*, or, *valued relation*, with membership function $\mu_R : X \times X \to [0; 1]$, where $\mu_R(x_i, x_j) = b_{ij}$ denotes the preference of the alternative x_i over x_j [6],

[9]. Hence, the fuzzy preference relation on X can be understood as a *fuzzy preference matrix*. Important properties of fuzzy preference matrix $B = \{b_{ij}\}$, can be summarized as follows.

An $n \times n$ matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j is *additive-reciprocal* (*a-reciprocal*) [3], if

$$b_{ij} + b_{ji} = 1 \text{ for all } i, j. \tag{3}$$

Evidently, if (3) holds, then $b_{ii} = 0.5$ for all i.

For making a coherent choice (when assuming fuzzy preference relations) a set of properties to be satisfied by such relations have been suggested in the literature [10].

The nomenclature of properties of relations is, however, not stabilized yet, compare e.g. [3], [10], [5], [11]. Here, we use the usual nomenclature which is as close as possible to the one used in the literature.

Transitivity is one of the most important properties concerning preferences, and it represents the idea that the preference intensity obtained by comparing directly two alternatives should be equal to or greater than the preference intensity between those two alternatives x and y obtained using an indirect chain of alternatives x and z, z and y.

Let $B = \{b_{ij}\}$ be an $n \times n$ a-reciprocal matrix with $0 < b_{ij} < 1$ for all i and j.

We say that $B = \{b_{ij}\}$ is *multiplicative-transitive (m-transitive)* [10], if

$$\frac{b_{ik}}{b_{ki}} = \frac{b_{ij}}{b_{ji}} . \frac{b_{jk}}{b_{kj}} \text{ for all } i, j, k. \tag{4}$$

Notice that if B is m-consistent then B is m-transitive. Moreover, if $B = \{b_{ij}\}$ is m-reciprocal, then B is m-transitive if and only if B is m-consistent.

We say that $B = \{b_{ij}\}$ is *additive-transitive* (*a-transitive*) [3], if

$$b_{ik} - 0.5 = (b_{ij} - 0.5) + (b_{jk} - 0.5) \text{ for all } i, j, k. \tag{5}$$

Equation (5) can be equivalently rewritten as (see [10]):

$$b_{ik} = 0.5 + b_{ij} - b_{kj} \text{ for all } i, j, k. \tag{6}$$

or, see [3],

$$b_{ij} + b_{jk} + b_{ki} = 1.5 \text{ for all } i, j, k. \tag{7}$$

In [5] this property is called additive consistency, here, we reserve this name for different concept, see below.

Moreover, we shall investigate some relationships between a-reciprocal and m-reciprocal pairwise comparison matrices. We start with extension of the result published by E. Herrera-Viedma et al. [3]. For this purpose, given $\sigma > 1$, we define the following function φ_σ and its inverse function φ_σ^{-1} as

$$\varphi_\sigma(t) = \frac{1}{2}(1 + \frac{\ln t}{\ln \sigma}) \text{ for } t \in [1/\sigma; \sigma], \tag{8}$$

$$\varphi_\sigma^{-1}(t) = \sigma^{2t-1} \text{ for } t \in [0; 1]. \tag{9}$$

We obtain the following results, characterizing a-transitive and m-consistent matrices, see [3], [6]. By $\sigma > 1$ evaluation scale $[1/\sigma; \sigma]$ is defined.

Proposition 1. Let $\sigma > 1$, $A = \{a_{ij}\}$ be an $n \times n$ matrix with $\frac{1}{\sigma} \leq a_{ij} \leq \sigma$ for all i and j. If $A = \{a_{ij}\}$ is m-consistent then $B = \{\varphi_\sigma(a_{ij})\}$ is a-transitive.

Proposition 2. Let $\sigma > 1$, $B = \{b_{ij}\}$ be an $n \times n$ matrix with $0 \leq b_{ij} \leq 1$ for all i and j. If $B = \{b_{ij}\}$ is a-transitive then $A = \{\varphi_\sigma^{-1}(b_{ij})\}$ is m-consistent.

Now, let us define the function ϕ and its inverse function ϕ^{-1} as follows

$$\phi(t) = \frac{t}{1+t} \text{ for } t > 0, \phi^{-1}(t) = \frac{t}{1-t} \text{ for } 0 < t < 1. \tag{10}$$

We obtain the following results, see [6].

Proposition 3. Let $A = \{a_{ij}\}$ be an $n \times n$ matrix with $0 < a_{ij}$ for all i and j. If $A = \{a_{ij}\}$ is m-consistent then $B = \{b_{ij}\} = \{\phi(a_{ij})\}$ is m-transitive.

Proposition 4. Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. If $B = \{b_{ij}\}$ is m-transitive then $A = \{a_{ij}\} = \{\phi^{-1}(b_{ij})\}$ is m-consistent.

From Proposition 2 it is clear that the concept of m-transitivity plays a similar role for a-reciprocal fuzzy preference matrices as the concept of m-consistency does for m-reciprocal matrices. That is why it is reasonable to introduce the following definition:

Definition. An a-reciprocal and m-transitive matrix $B = \{b_{ij}\}$ with $0 < b_{ij} < 1$ is called *additive-consistent* (*a-consistent*).

Proposition 4 can be reformulated accordingly: If $B = \{b_{ij}\}$ is a-consistent, then $A = \{\phi^{-1}(b_{ij})\}$ is m-consistent.

In practice, perfect consistency/transitivity is difficult to obtain, particularly when evaluating preferences on a set with a large number of alternatives.

3 Inconsistency of pairwise comparison matrices, priority vectors

If for some positive $n \times n$ matrix $A = \{a_{ij}\}$ and for some $i, j, k = 1, 2, ..., n$, multiplicative consistency condition (2) does not hold, then A is said to be

multiplicative-inconsistent (or, *m-inconsistent*). Eventually, if for some $n \times n$ fuzzy preference matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j, and for some tripple of indices i, j, k, (4) does not hold, then B is said to be *additive-inconsistent* (or, *a-inconsistent*) . Finally, if for some $n \times n$ fuzzy matrix $B = \{b_{ij}\}$ with $0 \leq b_{ij} \leq 1$ for all i and j, and for some tripple of indices i, j, k, (5) does not hold, then B is said to be *additive-intransitive* (*a-intransitive*). In order to measure the grade of inconsistency/intransitivity of a given matrix several instruments have been proposed in the literature. Recall that in AHP, multiplicative reciprocal matrices have been investigated, see [9].

As far as additive-reciprocal matrices are concerned, some instruments for measuring a-inconsistency/a-intransitivity are proposed here. Instead of positive matrices we consider matrices with nonnegative elements, i.e. some elements could be eventually zeros. Inconsistency of such matrix is based on Perron-Frobenius theory. The Perron-Frobenius theorem describes some of the remarkable properties enjoyed by the eigenvalues and eigenvectors of irreducible nonnegative matrices (e.g. positive matrices).

Theorem. (Perron-Frobenius, [2]) Let A be an irreducible nonnegative $n \times n$ matrix. Then the spectral radius, $\rho(A)$, is a real eigenvalue, which has a positive (real) eigenvector. This eigenvalue (called the principal eigenvalue of A) is simple, and its eigenvector (called priority vector) is unique up to a multiplicative constant.

The m-consistency of a nonnegative m-reciprocal $n \times n$ matrix A is given by the *m-consistency index* $I_{mc}(A)$ defined in [9] as

$$I_{mc}(A) = \frac{\rho(A) - n}{n - 1}, \tag{11}$$

where $\rho(A)$ is the spectral radius of A (particularly, *the principal eigenvalue* of A).

The rank of the alternatives in X is determined by the vector of weights $w = (w_1, w_2, ..., w_n)$, with $w_i > 0$, for all $i = 1, 2, ..., n$, such that $\sum_{i=1}^{n} w_i = 1$, satisfying $Aw = \rho(A)w$. This vector is called the (*normalized*) *principal eigenvector of* A, or, *the priority vector of* A. Since the element of the priority vector w_i is interpreted as the relative importance of alternative x_i, the alternatives $x_1, x_2, ..., x_n$ in X are ranked by their relative importance. The following result has been derived in [9].

Proposition 5. If $A = \{a_{ij}\}$ is an $n \times n$ positive m-reciprocal matrix, then $I_{mc}(A) \geq 0$. Moreover, A is m-consistent if and only if $I_{mc}(A) = 0$.

To provide a consistency measure independently of the dimension n of the matrix A, T. Saaty in [9] proposed the consistency ratio. In order to distinguish it here from the other consistency measures, we shall call it *m-consistency ratio*. This is obtained by taking the ratio I_{mc} to its mean value R_{mc}, i.e. the mean

value of $I_{mc}(A)$ of positive m-reciprocal matrices A of dimension n, whose entries a_{ij} are uniformly distributed random variables on $[1/9; 9]$, i.e.

$$CR_{mc} = \frac{I_{mc}}{R_{mc}}. \tag{12}$$

For this consistency measure it was proposed an estimation of 10% threshold of CR_{mc}. In other words, a pairwise comparison matrix could be accepted (in a DM process) if its m-consistency ratio does not exceed 0.1, see [9]. The m-consistency index I_{mc} has been defined by (11) for m-reciprocal matrices, now, we shall investigate inconsistency/intransitivity property also for a-reciprocal matrices. For this purpose we use relations between m-consistent and a-transitive/a-consistent matrices derived in Propositions 1 to 4. Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. We define the *a-consistency index* $I_{ac}(B)$ of $B = \{b_{ij}\}$ as

$$I_{ac}(B) = I_{mc}(A), \text{where } A = \{\phi^{-1}(b_{ij})\}. \tag{13}$$

From (13) we obtain the following result which is parallel to Proposition 5.

Proposition 6. If $B = \{b_{ij}\}$ is an a-reciprocal $n \times n$ fuzzy matrix with $0 < b_{ij} < 1$ for all i and j, then $I_{ac}(B) \geq 0$. Moreover, B is a-consistent if and only if $I_{ac}(B) = 0$.

The proof of Proposition 6 follows directly from Proposition 4 and Proposition 5.

Now, we shall deal with measuring a-intransitivity of a-reciprocal matrices. Recall transformation functions φ_σ and φ_σ^{-1} defined by (8), (9), where $\sigma > 1$ is a given value. Let $B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. We define the *a-transitivity index* $I_{at}^{\sigma}(B)$ of $B = \{b_{ij}\}$ as

$$I_{at}^{\sigma}(B) = I_{mc}(A_\sigma), \text{where } A_\sigma = \{\varphi_\sigma^{-1}(b_{ij})\}. \tag{14}$$

From (9), (14) we obtain the following result which is parallel to Propositions 5 and 6.

Proposition 7. If $B = \{b_{ij}\}$ is an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j, then $I_{at}^{\sigma}(B) \geq 0$. Moreover, B is a-transitive if and only if $I_{at}^{\sigma}(B) = 0$.

The proof of Proposition 7 follows directly from Proposition 2 and Proposition 5.

Let $A = \{a_{ij}\}$ be an a-reciprocal $n \times n$ matrix. In (12), the *m-consistency ratio of* A denoted by $CR_{mc}(A)$ is obtained by taking the ratio $I_{mc}(A)$ to its mean value $R_{mc}(n)$. i.e.

$$CR_{mc}(A) = \frac{I_{mc}(A)}{R_{mc}(n)}. \tag{15}$$

The table that gives the function values of $R_{mc}(n)$, $n = 3, 4, ..., 15$, can be found e.g. in [9]. Similarly, we define a-consistency ratio and a-transitivity ratio. Let

$B = \{b_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < b_{ij} < 1$ for all i and j. We define the *a-consistency ratio* CR_{ac} of B as follows

$$CR_{ac}(B) = \frac{I_{ac}(B)}{R_{mc}(n)}. \tag{16}$$

The corresponding priority vector $w^{ac} = (w_1^{ac}, w_2^{ac}, ..., w_n^{ac})$ is given by the characteristic equation $\phi^{-1}(B)w^{ac} = \rho(\phi^{-1}(B))w^{ac}$, where $\phi^{-1}(B) = \{\phi^{-1}(b_{ij}))\}$.

Moreover, given $\sigma > 1$, we define *a-transitivity ratio* CR_{at}^{σ} of B as

$$CR_{at}^{\sigma}(B) = \frac{I_{at}^{\sigma}(B)}{R_{mc}(n)}. \tag{17}$$

The corresponding priority vector $w^{at} = (w_1^{at}, w_2^{at}, ..., w_n^{at})$ is given by

$$\varphi_{\sigma}^{-1}(B)w^{at} = \rho(\varphi_{\sigma}^{-1}(B)w^{at},$$

where $\varphi_{\sigma}^{-1}(B) = \{\varphi_{\sigma}^{-1}(b_{ij})\}$. In practical DM situations a-inconsistency of a positive a-reciprocal pairwise comparison matrix B is "acceptable" if $CR_{ac}(B) < 0.1$. Also, a-intransitivity of a positive a-reciprocal pairwise comparison matrix B is "acceptable" if $CR_{at}^{\sigma}(B) < 0.1$. The final ranking of alternatives is given by the corresponding priority vector.

The following two results give a characterization of m-consistent matrix as well as a-consistent one by the vectors of weights, i.e. positive vectors with sum of elements equal to one, see [6].

Proposition 8. Let $A = \{a_{ij}\}$ be a positive $n \times n$ matrix. A is m-consistent if and only if there exists a vector $w = (w_1, w_2, ..., w_n)$ with $w_i > 0$ for all i=1,2,...,n, and $\sum_{j=1}^{n} w_j = 1$ such that

$$a_{ij} = \frac{w_i}{w_j} \text{ for all } i, j = 1, 2, ..., n. \tag{18}$$

Proposition 9. Let $A = \{a_{ij}\}$ be an a-reciprocal $n \times n$ matrix with $0 < a_{ij} < 1$ for all i and j. $A = \{a_{ij}\}$ is a-consistent if and only if there exists a vector $v = (v_1, v_2, ..., v_n)$ with $v_i > 0$ for all i=1,2,...,n, and $\sum_{j=1}^{n} v_j = 1$ such that

$$a_{ij} = \frac{v_i}{v_i + v_j} \text{ for all } i, j = 1, 2, ..., n. \tag{19}$$

A parallel result can be derived for a-transitive matrices.

Proposition 10. Let $A = \{a_{ij}\}$ be an a-reciprocal $n \times n$matrix with $0 < a_{ij} < 1$ for all i and j. $A = \{a_{ij}\}$ is a-transitive if and only if there exists a vector $u = (u_1, u_2, ..., u_n)$ with $u_i > 0$ for all $i = 1, 2, ..., n$, and $\sum_{j=1}^{n} u_j = 1$ such that

$$a_{ij} = \frac{1}{2}(1 + nu_i - nu_j) \text{ for all } i, j = 1, 2, ..., n. \tag{20}$$

Example 1. Let $X = \{x_1, x_2, x_3, x_4\}$ be the set of 4 alternatives. The preferences on X are described by the positive matrix $B = \{b_{ij}\}$,

$$B = \begin{pmatrix} 0.5 & 0.6 & 0.6 & 0.9 \\ 0.4 & 0.5 & 0.6 & 0.7 \\ 0.4 & 0.4 & 0.5 & 0.5 \\ 0.1 & 0.3 & 0.5 & 0.5 \end{pmatrix}. \tag{21}$$

Here, $B = \{b_{ij}\}$ is a-reciprocal and a-inconsistent, as it may be directly verified by (4), e.g. $b_{12}.b_{23}.b_{31} \neq b_{32}.b_{21}.b_{13}$. At the same time, B is a-intransitive as $b_{12} + b_{23} + b_{31} = 1.9 \neq 1.5$. Now, consider $\sigma = 9$. Then we calculate

$$E = \{\phi^{-1}(b_{ij})\} = \begin{pmatrix} 1 & 1.50 & 1.50 & 9.00 \\ 0.67 & 1 & 1.5 & 2.33 \\ 0.67 & 0.67 & 1 & 1 \\ 0.11 & 0.43 & 1 & 1 \end{pmatrix},$$

$$F = \{\varphi_9^{-1}(b_{ij})\} = \begin{pmatrix} 1 & 1.55 & 1.55 & 5.80 \\ 0.64 & 1 & 1.55 & 2.41 \\ 0.64 & 0.64 & 1 & 1 \\ 0.17 & 0.42 & 1 & 1 \end{pmatrix}.$$

Further, we calculate the maximal eigenvalues $\rho(E) = 4.29$ and $\rho(F) = 4.15$. By (11), (16) and (17) we obtain $CR_{ac}(B) = 0.11 > 0.1$ with the priority vector $w^{ac} = (0.47, 0.25, 0.18, 0.10)$, which gives the ranking of the alternatives as $x_1 > x_2 > x_3 > x_4$. Similarly, $CR^9_{at}(B) = 0.056 < 0.1$ with the priority vector $w^{at} = (0.44, 0.27, 0.18, 0.12)$, giving the same ranking of alternatives $x_1 > x_2 > x_3 > x_4$.

As it is evident, a-consistency ratio $CR_{ac}(B)$ is too high that matrix B is considered a-consistent. On the other hand, a-transitivity ratio $CR^9_{at}(B)$ is sufficiently low that matrix B is considered a-transitive. The ranking of alternatives given by both methods remains, however, the same.

In this example it is evident that the values of $CR_{ac}(B)$ and $CR_{at}(B)$ can be different values for the a-reciprocal matrix B. In order to investigate a possible relationship between the consistency and transitivity indexes of a-reciprocal matrices, we performed a simulation experiment with randomly generated 1000 a-reciprocal matrices, ($n = 4$ and $n = 15$), then we calculated corresponding indexes I_{ac} and I^{σ}_{at} , with $\sigma = 9$. Numerical experiments have shown that there is no particular relationship between a-consistency and a-transitivity.

4 Fuzzy preference matrix with missing elements

In many decision-making procedures we assume that experts are capable of providing preference degrees between any pair of possible alternatives. However,

this may not be always true, which makes a missing information problem. A missing value in the fuzzy preference matrix is not equivalent to a lack of preference of one alternative over another. A missing value can be the result of the incapacity of an expert to quantify the degree of preference of one alternative over another. In this case he/she may decide not to guess the preference degree between some pairs of alternatives. It must be clear that when an expert is not able to express a particular value b_{ij}, because he/she does not have a clear idea of how the alternative x_i is better than alternative x_j , this does not mean that he/she prefers both options with the same intensity. To model these situations, in the following we introduce the incomplete preference relation matrix. Here, we use different approach and notation comparing to e.g. [1].

Now, we are going to define the fuzzy preference matrix with missing elements. For the sake of simplicity of presentation we identify the alternatives $x_1, x_2, ..., x_n$ with integers $1, 2, ..., n$, i.e. by $X = \{1, 2, ..., n\}$ we denote the set of alternatives, $n > 1$. Moreover, let, $X \times X = X^2$ be the Cartesian product of X, i.e. $X^2 = \{(i, j)|i, j \in X\}$. Let $K \subset X^2$, $K \neq X^2$ and $\mathcal{B}$ be the fuzzy preference relation on K given by the membership function $\mu_{\mathcal{B}} : K \to [0; 1]$. The fuzzy preference relation $\mathcal{B}$ is represented by the $n \times n$ *fuzy preference matrix* $B(K) = \{b_{ij}\}_K$ *with missing elements depending on* K as follows

$$b_{ij} = \begin{cases} \mu_{\mathcal{B}}(i, j) & \text{if} \quad (i, j) \in K, \\ - & \text{if} \quad (i, j) \notin K. \end{cases}$$

In what follows we shall assume that each fuzzy preference matrix $B(K) = \{b_{ij}\}_K$ with missing elements is a-reciprocal, i.e.

$$b_{ij} + b_{ji} = 1 \text{ for all } (i, j) \in K.$$

If $L \subset X^2$, and $L = \{(i_1, j_1), (i_2, j_2), ..., (i_q, j_q)\}$ is a set of couples (i, j) of alternatives such that there exist evaluations b_{ij}, with $0 \leq b_{ij} \leq 1$ for all $(i, j) \in L$, then the symmetric subset L' to L, i.e. $L' = \{(j_1, i_1), (j_2, i_2), ..., (j_q, i_q)\}$ is also a subset of K, i.e. $L' \subset K$. It is clear that by reciprocity each subset K of X^2 can be represented as follows: $K = L \cup L' \cup D$, where L is the set of couples of alternatives (i, j) of given preference degrees b_{ij} of the fuzzy preference matrix $B(K)$ and D is the diagonal of this matrix, i.e. $D = \{(1, 1), (2, 2), ..., (n, n)\}$, where $b_{ii} = 0.5$ for all $i \in X$. The reciprocity property means that the expert is able to quantify both b_{ij} and b_{ji} as well as b_{ii}. The elements b_{ij} with $(i, j) \in X^2$ - K are called *the missing elements of matrix* $B(K)$. Notice that the missing elements of $B(K)$ are denoted by symbol "-" ("dash"). On the other hand, the elements - preference degrees given by the experts are denoted by b_{ij} where $(i, j) \in K$. By reciprocity it is sufficient that the expert will quantify only the elements b_{ij}, where $(i, j) \in L$, such that $K = L \cup L' \cup D$. In what follows we shall investigate special important situations of L, particularly, $L = \{(1, 2), (2, 3), ..., (n - 1, n)\}$, or, $L = \{(1, 2), (1, 3), ..., (1, n)\}$.

Now, we shall deal with the problem of finding the values of missing elements of a given fuzzy preference matrix so that the extended matrix is as much a-consistent as possible. In the ideal case the extended matrix would become a-consistent.

Let $K \subset X^2$, let $B(K) = \{b_{ij}\}_K$ be a fuzzy preference matrix with missing elements. The matrix $B^{ac}(K) = \{b_{ij}^{ac}\}_K$ called the *ac-extension of* $B(K)$ is defined as follows

$$b_{ij}^{ac} = \begin{cases} b_{ij} & \text{if} \quad (i,j) \in K, \\ \frac{v_i^*}{v_i^* + v_j^*} & \text{if} \quad (i,j) \notin K. \end{cases}$$

Here, $v^* = (v_1^*, v_2^*, ..., v_n^*)$ called the *ac-priority vector with respect to* K is the optimal solution of the following problem

$$(P_{ac}) \qquad d_{ac}(v, K) = \sum_{(i,j)\in K} \left(b_{ij} - \frac{v_i}{v_i + v_j} \right)^2 \longrightarrow min;$$

subject to

$$\sum_{j=1}^n v_j = 1,\ v_i \geq \epsilon > 0 \text{ for all } i = 1, 2, ..., n.$$

(ϵ is a preselected sufficiently small positive number.)

Notice, that a-consistency index of the matrix $B^{ac}(K) = \{b_{ij}^{ac}\}_K$ is defined by (13) as $I_{ac}(B^{ac}(K))$. The proof of the following proposition follows directly from Proposition 9.

Proposition 11. $B^{ac}(K) = \{b_{ij}^{ac}\}_K$ is a-consistent, (i.e. $I_{ac}(B^{ac}(K)) = 0$) if and only if $d_{ac}(v^*, K) = 0$.

Now, we would like to find the values of missing elements of a given fuzzy preference matrix so that the extended matrix is as much a-transitive as possible. In the ideal case the extended matrix would become a-transitive.
Again, let $K \subset I^2$, let $B(K) = \{b_{ij}\}_K$ be a fuzy preference matrix with missing elements, $K = L \cup L' \cup D$ as before.

The matrix $B^{at}(K) = \{b_{ij}^{at}\}_K$ called an *at-extension of* $B(K)$ *with respect to* K is defined as follows

$$b_{ij}^{at} = \begin{cases} b_{ij} & \text{if} \quad (i,j) \in K, \\ max\{0, min\{1, \frac{1}{2}(1 + nu_i^* - nu_j^*)\}\} & \text{if} \quad (i,j) \notin K. \end{cases}$$

Here, $u^* = (u_1^*, u_2^*, ..., u_n^*)$ called the *at-priority vector with respect to* K is the optimal solution of the following problem

$$(P_{at}) \qquad d_{at}(v, K) = \sum_{(i,j)\in K} \left(b_{ij} - \frac{1}{2}(1 + nu_i - nu_j) \right)^2 \longrightarrow min;$$

subject to

$$\sum_{j=1}^n u_j = 1,\ u_i \geq \epsilon > 0 \text{ for all } i = 1, 2, ..., n.$$

Now, let a-transitivity index $I_{at}^{\sigma}(B^{at}(K))$ of the matrix $B^{at}(K) = \{b_{ij}^{at}\}_K$ is defined by (14) with a given $\sigma > 0$. The next proposition follows directly from Proposition 10.

Proposition 12. If $B^{at}(K) = \{b_{ij}^{at}\}_K$ is a-transitive (i.e. $I_{at}^{\sigma}(B^{at}(K)) = 0$), then $d_{at}(u^*, K) = 0$.

5 Special cases of fuzzy preference matrix with missing elements

For a complete definition of a reciprocal fuzzy preference nn matrix we need pairs of elements to be evaluated by an expert. For example, if $n = 10$, then $N = 45$, which is a considerable amount of pairwise comparisons. We ask that the expert would evaluate only around n pairwise comparisons of alternatives which seems to be a reasonable amount. In this section we shall deal with two important particular cases of fuzzy preference matrix with missing elements where the expert will evaluate only $n-1$ pairwise comparisons of alternatives. Let $K \subset I^2$ be a set of indexes given by an expert, $B(K) = \{b_{ij}\}_K$ be a fuzzy preference matrix with missing elements. Moreover, let $K = L \cup L' \cup D$. In fact, it is sufficient to assume that the expert will evaluate only matrix elements of L, i.e. $b_{12}, b_{23}, b_{34}, ..., b_{n-1,n}$.

5.1 Case $L = \{(1,2),(2,3),...,(n-1,n)\}$

Here, we assume that the expert evaluates $n-1$ elements of the fuzzy preference matrix $B(K)$, $b_{12}, b_{23}, b_{34}, ..., b_{n-1,n}$. First, we investigate the ac-extension of $B(K)$. We derive the following result.

Proposition 13. Let $L = \{(1,2),(2,3),...,(n-1,n)\}$, $0 < b_{ij} < 1$ with $b_{ij} + b_{ji} = 1$ for all $(i,j) \in K$, $K = L \cup L' \cup D$, and $L' = \{(2,1),(3,2),...,(n,n-1)\}, D = \{(1,1),...,(n,n)\}$. Then ac-priority vector $v^* = (v_1^*, v_2^*, ..., v_n^*)$ with respect to K is given as

$$v_1^* = \frac{1}{S} \text{ and } v_{i+1}^* = a_{i,i+1} v_i^* \text{ for } i = 1, 2, ..., n-1, \tag{22}$$

where

$$S = 1 + \sum_{i=1}^{n-1} a_{i,i+1} a_{i+1,i+2} ... a_{n-1,n} \text{ and } a_{ij} = \frac{1 - b_{ij}}{b_{ij}} \text{ for all } (i,j) \in K. \tag{23}$$

By (13) it follows that $B^{ac}(K) = \{b_{ij}^{ac}\}_K$ is a-consistent. Now, we investigate the at-extension $B^{at}(K)$ of$B(K)$. We obtain the following result.

Proposition 14. Let $L = \{(1,2),(2,3),...,(n-1,n)\}$, $0 < b_{ij} < 1$ with $b_{ij} + b_{ji} = 1$ for all $(i,j) \in K$, $K = L \cup L' \cup D$, and $L' = \{(2,1),(3,2),...,(n,n-1)\}, D = \{(1,1),...,(n,n)\}$. Let $u^* = (u_1^*, u_2^*, ..., u_n^*)$ be defined as

$$u_i^* = \frac{2}{n^2} \sum_{j=1}^{n-1} \alpha_j - \frac{2}{n} \alpha_{i-1} - \frac{n-i-1}{n} \text{ for } i = 1, 2, ..., n, \tag{24}$$

where

$$\alpha_0 = 0, \alpha_j = \sum_{i=1}^{j} b_{i,i+1} \text{ for } j = 1, 2, ..., n-1. \tag{25}$$

If $u^* = (u_1^*, ..., u_n^*)$ is a vector with positive elements, then u^* is an at-priority vector with respect to K.

Remark. In general, the optimal solution $u^* = (u_1^*, u_2^*, ..., u_n^*)$ of (P_{at}) does not satisfy condition

$$0 \leq \frac{1}{2}(1 + nu_i^* - nu_j^*) \leq 1, \text{ for all } i, j = 1, 2, ..., n, \tag{26}$$

i.e. $B = \{b_{ij}\} = \{\frac{1}{2}(1 + nu_i^* - nu_j^*)\}$ is not a fuzzy preference matrix. We can easily prove the necessary and sufficient condition for satisfying (26) based on evaluations $b_{i,i+1}$.

Proposition 15. Let $L = \{(1,2), (2,3), ..., (n-1, n)\}$, $0 \leq b_{ij} \leq 1$ with $b_{ij} + b_{ji} = 1$ for all $(i,j) \in K$, $K = L \cup L' \cup D$, and $L' = \{(2,1), (3,2), ..., (n, n-1)\}, D = \{(1,1), ..., (n,n)\}$. Then the at-extension $B^{at}(K) = \{b_{ij}^{at}\}_K$ is a-transitive if and only if

$$\left| \sum_{k=i}^{j-1} b_{k,k+1} - \frac{j-i}{2} \right| \leq \frac{1}{2} \text{ for } i = 1, 2, ..., n-1, j = i+1, ..., n. \tag{27}$$

Example 2. Let $L = \{(1,2), (2,3), (3,4)\}$, let the expert evaluations be $b_{12} = 0.9, b_{23} = 0.8, b_{34} = 0.6$, with $b_{ij} + b_{ji} = 1$ for all $(i,j) \in L$, let $K = L \cup L' \cup D$. Hence $B(K) = \{b_{ij}\}_K$ is a fuzzy preference matrix with missing elements as follows

$$B(K) = \begin{pmatrix} 0.5 & 0.9 & - & - \\ 0.1 & 0.5 & 0.8 & - \\ - & 0.2 & 0.5 & 0.5 \\ - & - & 0.4 & 0.5 \end{pmatrix}. \tag{28}$$

Solving (P_{ac}) we obtain ac-priority vector v^* with respect to K, particularly, $v^* = (0.864, 0.096, 0.024, 0.016)$. By (24) we obtain $B^{ac}(K)$ - ac-extension of $B(K)$ as follows

$$B^{ac}(K) = \begin{pmatrix} 0.5 & 0.9 & 0.97 & 0.98 \\ 0.1 & 0.5 & 0.8 & 0.86 \\ 0.03 & 0.2 & 0.5 & 0.6 \\ 0.02 & 0.14 & 0.4 & 0.5 \end{pmatrix}, \tag{29}$$

where, $B^{ac}(K)$ is a-consistent, as $d_{ac}(v, B(K)) = 0$, hence $I_{ac}(B^{ac}(K)) = 0$.
Solving (P_{at}) we obtain at-priority vector u^* with respect to K,
$u^* = (0.487, 0.287, 0.137, 0.088)$. Then $B^{at}(K)$ is an at-extension of $B(K)$ as

$$B^{at}(K) = \begin{pmatrix} 0.5 & 0.9 & 1.0 & 1.0 \\ 0.1 & 0.5 & 0.8 & 0.9 \\ 0.0 & 0.2 & 0.5 & 0.6 \\ 0.0 & 0.1 & 0.4 & 0.5 \end{pmatrix}, \tag{30}$$

where, $B^{at}(K)$ is not a-transitive, as $d_{ac}(v, B(K)) > 0$. It can be easily verified as $I^9_{at}(B^{at}(K)) = 0.057$.

5.2 Case $L = \{(1,2),(1,3),...,(1,n)\}$

Now, we assume that the expert evaluates the pairs of a fixed element with the remaining $n-1$ elements, i.e. the fuzzy preference matrix $B(K)$ is given by $b_{12}, b_{13}, ..., b_{1n}$. We investigate the ac-extension of $B(K)$ and obtain the following result.

Proposition 16. Let $L = \{(1,2),(1,3),...,(1,n)\}$, $0 < b_{ij} < 1$ with $b_{ij}+b_{ji} = 1$ for all $(i,j) \in K$, $K = L \cup L' \cup D$, and $L' = \{(2,1),(3,2),...,(n,n-1)\}, D = \{(1,1),...,(n,n)\}$. Then ac-priority vector $v^* = (v^*_1, v^*_2, ..., v^*_n)$ with respect to K is given as

$$v^*_1 = \frac{1}{V} \text{ and } v^*_{i+1} = a_{1,i+1} v^*_i \text{ for } i = 1,2,...,n-1, \tag{31}$$

where

$$V = 1 + \sum_{i=1}^{n-1} a_{1,i+1} \text{ and } a_{ij} = \frac{1-b_{ij}}{b_{ij}} \text{ for all } (i,j) \in K. \tag{32}$$

We conclude that the ac-extension of $B(K)$, i.e. matrix $B^{ac}(K) = \{b^{ac}_{ij}\}_K$ is a-consistent. Now, we investigate the at-extension matrix $B^{at}(K)$ of $B(K)$. We can prove the following result.

Proposition 17. Let $L = \{(1,2),(1,3),...,(1,n)\}$, $0 < b_{ij} < 1$ with $b_{ij}+b_{ji} = 1$ for all $(i,j) \in K$, $K = L \cup L' \cup D$, and $L' = \{(2,1),(3,2),...,(n,n-1)\}, D = \{(1,1),...,(n,n)\}$. Let $u^* = (u^*_1, u^*_2, ..., u^*_n)$ be defined as follows

$$u^*_1 = \frac{2}{n^2} \sum_{j=1}^{n-1} b_{1,j+1} + \frac{1}{n^2} \text{ and } u^*_{i+1} = u^*_1 + \frac{1-2b_{1,i+1}}{n} \text{ for } i = 1,2,...,n-1. \tag{33}$$

If $u^* = (u^*_1, ..., u^*_n)$ is a vector with positive elements, then u^* is an at-priority vector with respect to K.

Remark. In general, the optimal solution $u^* = (u^*_1, u^*_2, ..., u^*_n)$ of (P_{at}) does not satisfy condition (26), i.e. $B = \{b^*_{ij}\} = \{\frac{1}{2}(1 + nu^*_i - nu^*_j)\}$ is not a fuzzy preference matrix. By a similar way we can prove the result which is parallel to Proposition 15.

Proposition 18. Let $L = \{(1,2),(2,3),...,(n-1,n)\}$, $0 \le b_{ij} \le 1$ with $b_{ij}+b_{ji} = 1$ for all $(i,j) \in K$, $K = L \cup L' \cup D$, and $L' = \{(2,1),(3,2),...,(n,n-1)\}, D = \{(1,1),...,(n,n)\}$. Then the at-extension $B^{at}(K) = \{b^{at}_{ij}\}_K$ is a-transitive if and only if

$$|b_{1j} - b_{1i}| \le \frac{1}{2} \text{ for } i,j = 1,2,...,n. \tag{34}$$

Example 3. Let $L = \{(1,2),(1,3),(1,4)\}$, let the expert evaluations be $b_{12} = 0.9, b_{13} = 0.8, b_{14} = 0.3$, with $b_{ij} + b_{ji} = 1$ for all $(i,j) \in L$, let $K = L \cup L' \cup D$. Hence $B(K) = \{b_{ij}\}_K$ is a fuzzy preference matrix with missing elements as follows

$$B(K) = \begin{pmatrix} 0.5 & 0.9 & 0.8 & 0.3 \\ 0.1 & 0.5 & - & - \\ 0.2 & - & 0.5 & - \\ 0.7 & - & - & 0.5 \end{pmatrix}. \tag{35}$$

Solving (P_{ac}) we obtain ac-priority vector v^* with respect to K, particularly, $v^* = (0.271, 0.030, 0.068, 0.632)$. Then we obtain $B^{ac}(K)$ - ac-extension of $B(K)$ as

$$B^{ac}(K) = \begin{pmatrix} 0.5 & 0.9 & 0.80 & 0.30 \\ 0.10 & 0.5 & 0.30 & 0.04 \\ 0.20 & 0.70 & 0.5 & 0.10 \\ 0.70 & 0.96 & 0.90 & 0.5 \end{pmatrix}. \tag{36}$$

where, $B^{ac}(K)$ is a-consistent, as $d_{ac}(v, B(K)) = 0$, hence $I_{ac}(B^{ac}(K)) = 0$. Solving (P_{at}) we obtain at-priority vector u^* with respect to K,
$u^* = (0.312, 0.113, 0.162, 0.412)$. Then $B^{at}(K)$ is an at-extension of $B(K)$ as

$$B^{at}(K) = \begin{pmatrix} 0.5 & 0.90 & 0.80 & 0.30 \\ 0.10 & 0.5 & 0.40 & 0.00 \\ 0.20 & 1.00 & 0.5 & 0.00 \\ 0.70 & 1.00 & 1.00 & 0.5 \end{pmatrix}, \tag{37}$$

where $B^{at}(K)$ is not a-transitive, as $|b_{12} - b_{14}| > 0.6 > \frac{1}{2}$.

6 Conclusions

In this paper we have dealt with some properties of fuzzy preference relations, particularly reciprocity, consistency and transitivity of relations given in the form of square matrices with the entries from the unit interval. We have shown how to measure the grade of a-consistency and/or a-transitivity, and also how to evaluate the pairs of elements by fuzzy values. Also, we have proposed a new method for measuring of inconsistency based on Saatys principal eigenvector method. Moreover, we have dealt with two particular cases of the incomplete fuzzy preference matrix, where some elements of pairwise comparison are missing. We have proposed some special methods for dealing with these cases. Two illustrating examples have been presented to clarify our approach.

References

[1] Alonso, S., Chiclana, F., Herrera. F., Herrera-Viedma, E., Alcala-Fdes, J., Porcel, C.: A consistency-based procedure to estimate missing pairwise preference values. *Internat. J. Intelligent Syst.*, 23 (2008), 155-175.

[2] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J., Zimmermann, K.: *Linear Optimization Problems with Inexact Data*. Berlin - Heidelberg - New York - Hong Kong - London - Milan, Tokyo: Springer, 2006, 229 p.

[3] Chiclana, F. Herrera, F., Herrera-Viedma, E.: Integrating Multiplicative Preference Relations in a Multipurpose Decision Making Model Based on Fuzzy Preference Relations. *Fuzzy Sets and Systems* 112 (2001) 277-291.

[4] Herrera-Viedma, E., Herrera, F., Chiclana, F., Luque, M.: Some issues on consistency of fuzzy preference relations. *European Journal of Operat. Res.* 154 (2004) 98-109.

[5] Ma, J. et al.: A method for repairing the inconsistency of fuzzy preference relations. *Fuzzy Sets and Systems* 157 (2006) 20–33.

[6] Ramík, J., Vlach, M.: Measuring consistency and inconsistency of pair comparison systems. In: *Kybernetika* 49 (2013), 3, 465-486.

[7] Ramík, J., Korviny, P.: Inconsistency of pairwise comparison matrix with fuzzy elements based on geo-metric mean. *Fuzzy Sets and Systems* 161 (2010), 1604-1613.

[8] Saaty, T. L.: *The analytic hierarchy process*. McGraw-Hill, New York, 1980.

[9] Saaty, T. L.: *Fundamentals of Decision Making and Priority Theory with the AHP*. RWS Publications, Pittsburgh, 1994.

[10] Tanino, T.: Fuzzy Preference Relations in Group Decision Making. In: J. Kacprzyk, M. Roubens (Eds.), *Non-Conventional Preference Relations in Decision Making*. Springer-Verlag, Berlin, 1988, 54-71.

[11] Xu, Z.S., Da, Q.L.: A least deviation method to obtain a priority vector of a fuzzy preference relation. *European Journal of Operational Research* 164 (2005) 206-216.

Acknowledgement.
This research has been supported by GACR project No. 402090405.

An Interpretation of Conflicting Parts of Belief Functions on Two-Element Frame of Discernment

Milan Daniel*
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, CZ – 182 07 Prague 8, Czech Republic
e-mail: milan.daniel@cs.cas.cz

Abstract

When combining belief functions by the conjunctive rules of combination, conflicts often appear, which are assigned to $\emptyset$ by un-normalized conjunctive rule $\textcircled{\cap}$ or normalized by Dempster's rule of combination $\oplus$.

This contribution is devoted to an interpretation of the conflicting part of a belief function on a two-element frame of discernment. It is based on the author's idea of the unique decomposition of such function into its conflicting and non-conflicting part (CJS 2010, Otaru). A relation of conflicting part of a belief function to internal conflict of the function is also studied and a new definition of internal conflict is introduced. New internal conflict is compared with the previous approaches.

Keywords: Belief function, Dempster-Shafer theory, uncertainty, Dempster's semigroup, internal conflict, conflict between belief functions, non-conflicting part of belief function, conflicting part of belief function.

1 Introduction

Belief functions are one of the widely used formalisms for uncertainty representation and processing that enable representation of incomplete and uncertain knowledge, belief updating, and combination of evidence. They were originally introduced as a principal notion of the Dempster-Shafer Theory or the Mathematical Theory of Evidence [16].

When combining belief functions (BFs) by the conjunctive rules of combination, conflicts often appear, which are assigned to $\emptyset$ by un-normalized conjunctive rule $\textcircled{\cap}$ or normalized by Dempster's rule of combination $\oplus$. Combination

*This research is supported by the grant P202/10/1826 of the Grant Agency of the Czech Republic. The partial institutional support RVO: 67985807 is also acknowledged.

of conflicting belief functions and interpretation of conflicts is often questionable in real applications, thus a series of alternative combination rules was suggested and a series of papers on conflicting belief functions was published, e.g. [4, 10, 14, 15, 18]. Unfortunately, a complete nature of the conflicts is still not satisfactorily captured.

The new measure of conflict between BFs on 2-element frame was introduced at CJS 2012 (Osaka [7]). It is based on the important author's idea of distinguishing of internal conflicts of individual belief functions from a conflict between the belief functions (IPMU 2010, Dortmund [4]), and on the idea of the unique decomposition of a belief function on 2-element frame of discernment into its conflicting and non-conflicting part (CJS 2010, Otaru [5]).

This contribution is devoted to an interpretation of the conflicting part of a belief function on 2-element frame of discernment and to its relation to the internal conflict of the belief function. An analysis of interpretation of the conflicting part of a BF is followed by a new definition of internal conflict of a BF on 2-element frame and by its comparison with previous approaches: plausibility and combinational internal conflicts [4] and auto-conflict [15].

2 Preliminaries

2.1 General primer on belief functions

We assume classic definitions of basic notions from theory of *belief functions* (BFs) [16] on finite frames of discernment $\Omega_n = \{\omega_1, \omega_2, ..., \omega_n\}$, see also [2, 3]. A *basic belief assignment (bba)* is a mapping $m : \mathcal{P}(\Omega) \longrightarrow [0,1]$ such that $\sum_{A\subseteq\Omega} m(A) = 1$; the values of the bba are called *basic belief masses (bbm)*. $m(\emptyset) = 0$ is usually assumed, then we speak about *normalized bba*. A *belief function* is a mapping $Bel : \mathcal{P}(\Omega) \longrightarrow [0,1]$, $Bel(A) = \sum_{\emptyset\neq X\subseteq A} m(X)$. A *plausibility function* $Pl(A) = \sum_{\emptyset\neq A\cap X} m(X)$. Due to a unique correspondence among m and corresponding Bel and Pl, we often speak about m as about BF.

A *focal element* is a subset X of the frame of discernment, such that $m(X) > 0$. If all the focal elements are *singletons* (i.e. one-element subsets of Ω), then we speak about a *Bayesian belief function* (BBF). An *indecisive BF* is a BF, which does not prefer any $\omega_i \in \Omega_n$, it gives no decisional support for any ω_i.

Dempster's (conjunctive) rule of combination $\oplus$ is given as $(m_1 \oplus m_2)(A) = \sum_{X\cap Y=A} K m_1(X) m_2(Y)$ for $A\neq\emptyset$, where $K = \frac{1}{1-\kappa}$, $\kappa = \sum_{X\cap Y=\emptyset} m_1(X)m_2(Y)$, and $(m_1 \oplus m_2)(\emptyset) = 0$, see [16]; putting $K = 1$ and $(m_1 \oplus m_2)(\emptyset) = \kappa$ we obtain the *non-normalized conjunctive rule of combination* $\textcircled{\cap}$, see e. g. [17].

We say that BF Bel is *non-conflicting* when conjunctive combination of Bel with itself does not produce any conflicting belief masses (when $(Bel \textcircled{\cap} Bel)(\emptyset) = 0$. Otherwise, BF is *conflicting*, i.e., it contains some internal conflict [4].

Normalized plausibility of singletons[1] is probability distribution $Pl\_P(\omega_i) = \frac{Pl(\{\omega_i\})}{\sum_{\omega\in\Omega} Pl(\{\omega\})}$ [1, 3]; Smets' *pignistic probability* $BetP(\omega_i) = \sum_{\omega_i\in X} \frac{m(X)}{|X|}$ [17].

[1] Plausibility of singletons is called *contour function* by Shafer in [16], thus $Pl\_P(Bel)$ is a normalization of contour function in fact.

2.2 Belief functions on two-element frame of discernment; Dempster's semigroup

We assume $\Omega_2 = \{\omega_1, \omega_2\}$, in this study. Representing a (normalized) belief function by enumeration of bbms we obtain a pair $(a, b) = (m(\{\omega_1\}), m(\{\omega_2\}))$ as $m(\{\omega_1, \omega_2\}) = 1 - a - b$; this is called *Dempster's pair* or simply *d-pair* in [2, 12, 13] (it is a pair of reals such that $0 \leq a, b \leq 1$, $a + b \leq 1$).

Extremal d-pairs are the pairs corresponding to belief functions for which either $m(\{\omega_1\}) = 1$ or $m(\{\omega_2\}) = 1$, i.e., $\top = (1, 0)$ and $\perp = (0, 1)$. The set of all non-extremal d-pairs is denoted as D_0; the set of all non-extremal *Bayesian d-pairs* (i.e. d-pairs corresponding to Bayesian BFs, where $a + b = 1$) is denoted as G; the set of d-pairs such that $a = b$ is denoted as S (set of indecisive[2] d-pairs), the set where $b = 0$ as S_1, and analogically, the set where $a = 0$ as S_2 (simple support BFs, simple *d*-pairs). Vacuous BF is denoted as $0 = (0, 0)$ and there is a special BF (d-pair) $0' = (\frac{1}{2}, \frac{1}{2})$, see Figure 1.

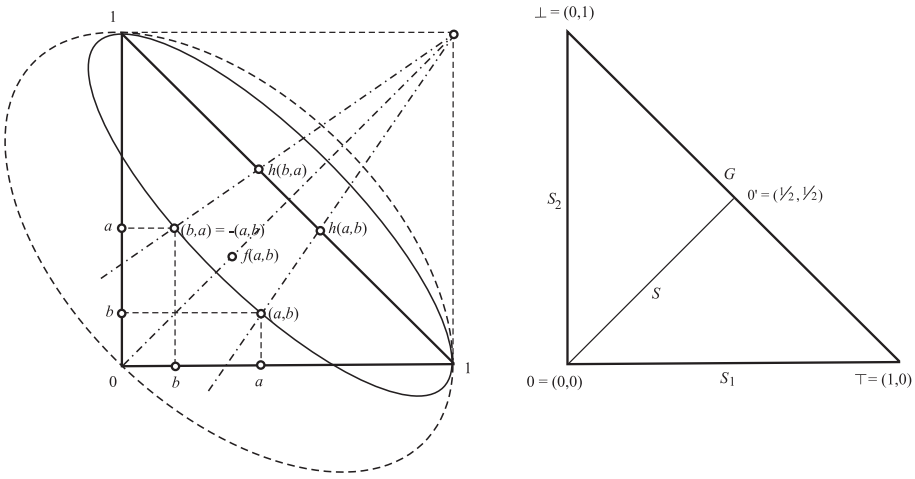

Figure 1: Dempster's semigroup $\mathbf{D}_0$. Homomorphism h is in this representation a projection of $\mathbf{D}_0$ to group $\mathbf{G}$ along the straight lines (*h-lines*) running through the point $(1, 1)$. All the Dempster's pairs lying on the same ellipse (*f-ellipse*, running through points $(0, 1)$ and $(1, 0)$) are mapped by homomorphism f to the same *d*-pair in semigroup S.

The *(conjunctive) Dempster's semigroup* $\mathbf{D}_0 = (D_0, \oplus, 0, 0')$ is the set D_0 endowed with the binary operation $\oplus$ (i.e. with the Dempster's rule) and two distinguished elements 0 and $0'$. Dempster's rule can be expressed by the formula $(a, b) \oplus (c, d) = (1 - \frac{(1-a)(1-c)}{1-(ad+bc)},\ 1 - \frac{(1-b)(1-d)}{1-(ad+bc)})$ for *d*-pairs [12, 13]. In D_0 it is defined further: $-(a, b) = (b, a)$, $h(a, b) = (a, b) \oplus 0' = (\frac{1-b}{2-a-b},\ \frac{1-a}{2-a-b})$, $h_1(a, b) = \frac{1-b}{2-a-b}$, $f(a, b) = (a, b) \oplus -(a, b) = (a, b) \oplus (b, a) =$

[2]BFs (a, a) from S are called *indifferent* BFs by Haenni [11].

$(\frac{a+b-a^2-b^2-ab}{1-a^2-b^2},\ \frac{a+b-a^2-b^2-ab}{1-a^2-b^2})$; $(a,b) \leq (c,d)$ iff $[h_1(a,b) < h_1(c,d)$ or $h_1(a,b) = h_1(c,d)$ and $a \leq c]$ [3]. For principal properties of $\mathbf{D}_0$ and its subalgebras see e.g. [2, 12, 13].

Let us denote $h^{-1}(a,b) = \{(u,v) \,|\, h(u,v) = (a,b)\}$ and similarly $f^{-1}(a,b) = \{(u,v) \,|\, f(u,v) = (a,b)\}$. Using h^{-1} and f^{-1}, we can express $\oplus$ as:

$$((a,b) \oplus (c,d)) \ = \ h^{-1}(h(a,b) \oplus h(c,d)) \cap f^{-1}(f(a,b) \oplus f(c,d)).$$

On Ω_2 we have further, $Bel(\{\omega_1\}) = a$, $Bel(\{\omega_2\}) = b$, $Pl(\{\omega_1\}) = 1-b$, $Pl(\{\omega_2\}) = 1-a$, $Pl\_P(\omega_1) = \frac{1-b}{2-a-b}$, $Pl\_P(\omega_2) = \frac{1-a}{2-a-b}$. $(a,b) \circledcirc (c,d) = (\, a(1-d)+c(1-a-b),\ b(1-c)+d(1-a-b);\ (1-a-b)(1-c-d))$, $m_{(a,b)\circledcirc(c,d)}(\emptyset) = ad+bc$, $BetP(\omega_1) = \frac{1+a-b}{2}$, $BetP(\omega_2) = \frac{1+b-a}{2}$.

2.3 Conflicts of belief functions

Internal conflicts $IntC(m_i)$ which are included in particular individual BFs are distinguished from *conflict between BFs* $C(m_1, m_2)$ in [4]; the entire sum of conflicting masses is called *total conflict* $TotC(m_1, m_2) = (m_1 \circledcirc m_2)(\emptyset)$; and three approaches to conflicts were introduced: combinational, plausibility and comparative.

Unfortunately, there are not yet any precise formulas, but only bounding inequalities for *combinational conflicts*: ,

$$\frac{1}{2} TotC(m,m)) \leq cb\text{-}IntC(m) \leq TotC(m,m)$$

$$TotC(m_1,m_2) - (cb\text{-}IntC(m_1) + cb\text{-}IntC(m_2)) \ \leq \ cb\text{-}C(m_1,m_2) \ \leq \ TotC(m_1,m_2).$$

Internal plausibility conflict of BF Bel is defined as $Pl\text{-}IntC(Bel) = 1 - max_{\omega\in\Omega} Pl(\{\omega\})$, where Pl is the plausibility equivalent to Bel. On Ω_2 we have $Pl\text{-}IntC(a,b) = 1 - max(1-b, 1-a) = min(a,b)$.

Plausibility conflict between BFs Bel_1 and Bel_2 is defined by the formula $Pl\text{-}C(Bel_1, Bel_2) = min(\ \sum_{\omega\in\Omega_{PlC}(Bel_1,Bel_2)} \frac{1}{2} |Pl\_P(Bel_1)(\omega) - Pl\_P(Bel_2)(\omega)|,\ (m_1 \circledcirc m_2)(\emptyset)\ )$, where $\Omega_{PlC}(Bel_1, Bel_2)$ is the set of elements $\omega \in \Omega$ with conflicting Pl_P masses [4, 8, 9]. BFs (a,b) and (c,d) on Ω_2 are mutually non-conflicting when $a \geq b$ & $c \geq d$ or $a \leq b$ & $c \leq d$ (i.e., $Pl\text{-}C = 0$); otherwise, we have $Pl\text{-}C((a,b),(c,d)) = |\frac{1-b}{2-a-b} - \frac{1-d}{2-c-d}| = |\frac{1-a}{2-a-b} - \frac{1-c}{2-c-d}|$.

The idea of comparative conflictness / non-conflictness is a specification of bbms to smaller focal elements such that fit to focal elements of the other BF as much as possible. *The comparative conflict between BFs* Bel_1 *and* Bel_2 $cp\text{-}C(m_1, m_2)$ is defined as the least difference of such more specified bbms derived from the input m_1 and m_2. On Ω_2 it is a (partial) specification of $m_i(\{\omega_1, \omega_2\})$ to singletons, i.e., specification of $1-a-b$ to a if $a<c$ and/or to b if $b<d$ and analogically specification of $1-c-d$ to c if $c<a$ and/or to d if $d<b$. There is no internal conflict defined in the case of comparative conflict.

Analogy of internal conflict is Martin's auto-conflict [15] defined as $(m \circledcirc m)(\emptyset)$, i.e., it is $TotC(m,m)$ in fact. On Ω_2 we have $((a,b) \circledcirc (a,b))(\emptyset) = 2ab$.

[3] Note, that $h(a,b)$ is an abbreviation for $h((a,b))$, similarly for $h_1(a,b)$ and $f(a,b)$.

3 State of the Art

3.1 Decomposition of a belief function to its non-conflicting and conflicting part

We will use the important property of Dempster's sum, which is respecting the homomorphisms h and f, i.e., respecting the h-lines and f-ellipses, when two BFs are combined on two-element frame of discernment [2, 12, 13], see Figure 1. Using this property we obtain the following statement, see Figure 2.

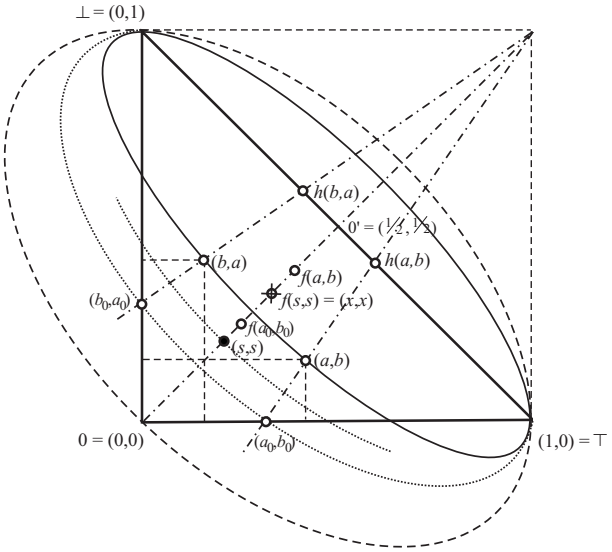

Figure 2: Conflicting and non-conflicting part of a BF (a, b) on 2-element frame of discernment.

Theorem 1 *Any BF* (a, b) *on 2-element frame of discernment* Ω_2 *is Dempster's sum of its unique* non-conflicting part $(a_0, b_0) \in S_1 \cup S_2$ *and of its unique* conflicting part $(s, s) \in S$, *which does not prefer any element of* Ω_2, *i.e.* $(a, b) = (a_0, b_0) \oplus (s, s)$. *It holds true that* $s = \frac{b(1-a)}{1-2a+b-ab+a^2} = \frac{b(1-b)}{1-a+ab-b^2}$ *and* $(a_0, b_0) = (\frac{a-b}{1-b}, 0)$, $(a, b) = (\frac{a-b}{1-b}, 0) \oplus (s, s)$ *for* $a \geq b$; *and similarly that* $s = \frac{a(1-b)}{1+a-2b-ab+b^2} = \frac{a(1-a)}{1-b+ab-a^2}$ *and* $(a_0, b_0) = (0, \frac{b-a}{1-a})$ *for* $a \leq b$.

For proof see [6].

3.2 Conflict between belief functions

Using of the unique decomposition of a BF (a, b) on Ω_2: $(a, b) = (a_0, b_0) \oplus (s, s)$, a new definition of conflict between BFs on 2-element frame of discernment was defined at CJS 2012, Osaka [7]:

Definition 1 *Let Bel', Bel'' be two belief functions on 2-element frame of discernment $\Omega_2 = \{\omega_1, \omega_2\}$. Let $Bel'_0 \oplus Bel'_S = Bel'$ and $Bel''_0 \oplus Bel''_S = Bel''$ be their decompositions into their non-conflicting and conflicting parts. We define* conflict between BFs Bel' and Bel'' *as $Conf(Bel', Bel'') = m_{Bel'_0 \textcircled{\cap} Bel''_0}(\emptyset)$.*

We can compute conflict between any 2 BFs from D_0^+ as it follows:

Theorem 2 *Let $(a, b), (c, d)$ be arbitrary BFs defined on 2-element frame of discernment. Conflict between (a, b) and (c, d) is given by the following expression:*
$Conf((a, b), (c, d)) = \frac{a-b}{1-b} \cdot \frac{d-c}{1-c}$ *if* $a > b$ & $c < d$,
$Conf((a, b), (c, d)) = \frac{b-a}{1-a} \cdot \frac{c-d}{1-d}$ *if* $a < b$ & $c > d$, $Conf((a, b), (c, d)) = 0$ *otherwise.*

For proof and other properties of $Conf$ see [7].

4 Interpretation of a Conflicting Part

Summarizing the present state of the art, according to CJS 2010, Otaru [5], we have a decomposition $Bel = Bel_0 \oplus Bel_S$ ($(a, b) = (a_0, b_0) \oplus (s, s)$) of any BF $Bel = (a, b)$ on a 2-element frame of discernment. $Bel_0 = (a_0, b_0)$ is its non-conflicting part, it has no internal conflict, it is a simple BF (simple support BF) in S_1 or S_2. On the other hand $Bel_S = (a_S, b_S) = (s, s)$, which is indecisive (i.e., it has the same decisional supports for both the elements of the frame of discernment), bears all internal conflict of Bel. The importance of non-conflicting Bel_0 is its decisional support for one of the elements and also its principal role in the new definition of conflict between two BFs on a 2-element frame of discernment, see CJS 2012, Osaka [7]. What is an interpretation of Bel_S? We try to answer this question in this section.

4.1 Conflicting part as an internal conflict?

The simplest interpretation of conflicting part of a belief function (or of its components) is internal conflict of the BF.

Let us start with a simple case of indecisive belief functions from semigroup S. We have $Bel_0 = 0$ and $Bel_S = Bel = (s, s)$ in this case, thus the decomposition is trivial $Bel = 0 \oplus Bel$ in this case. There is no problem with interpretation in this case. $0 = (0, 0)$ is non-conflicting, thus its conflicting part is $0 = (0, 0)$ with both components equal to zero and internal conflict is also 0. Increasing s, the same support to both elements of the frame of discernment and also both components of conflicting part increase up to $\frac{1}{2}$ for $0' = (\frac{1}{2}, \frac{1}{2})$ with the greatest conflicting part $0'$. $0'$ has also the greatest internal conflict of BFs on S. Thus s nicely represents the internal conflict of a BF on S.

$0 = (0, 0)$ is also conflicting part of any non-conflicting BF from S_1 and S_2 with no (i.e., zero) internal conflict.

The problems start for Bayesian BFs from G: all BBFs have the same conflicting part $0' = (\frac{1}{2}, \frac{1}{2})$, thus if we keep the above interpretation internal conflict of all BBFs should be same as that of $0'$. But this does not hold true, especially

for categorical BFs $(0,1)$ and $(1,0)$, which are both (individually) non-conflicting (i.e. they have zero internal conflicts; of course, there is full conflict between them).

Analogously for any other set of BFs with the same conflicting part, where internal conflict should decrease towards categorical BFs and it should be close to zero in surrounding of categorical BFs.

Thus we cannot interpret conflicting part of a BF (or component(s) of the conflicting part) to be an internal conflict of the BF.

4.2 Conflicting part as a relative internal conflict

The nice property of conflicting parts of BFs from S holds true also for any set of BFs with the same non-conflicting part, thus with the same Pl_P, i.e. for BFs on an intersection of some h-line with the triangle of BFs on Ω_2. Both internal conflict and the conflicting part are 0 (0 resp. $0 = (0,0)$) for the intersection of h-line with S_i; both internal conflict and the conflicting part are maximal for the intersection of h-line with G; and both internal conflict ($Pl\text{-}IntC$, $m(\emptyset)$, thus also both bounds of $cb\text{-}IntC$) and the conflicting part increase between the above two intersections. Hence, we can interpret conflicting part Bel_S of BF Bel as its relative internal conflict (a relative internal conflict of the BF $Bel = Bel_0 \oplus Bel_S$ with non-conflicting part Bel_0 and conflicting part Bel_S).

5 A New Definition of Internal Conflict

5.1 Definition 2

Using the above interpretation of conflicting part of BFs on 2-element frame of discernment $\Omega_2 = \{\omega_1, \omega_2\}$ (and using Pl_P, which is constant for BFs with a fixed non-conflicting part), we can define internal conflict of a BF Bel defined on Ω_2 as it follows:

$$IntC(Bel) = (Bel_S(\{\omega_1\}) + Bel_S(\{\omega_2\})) \cdot min(Pl\_P(\omega_1), Pl\_P(\omega_2)), \quad (1)$$

$$IntC(a,b) = (a_S + b_S)\ min(\frac{1-b}{2-a-b}, \frac{1-a}{2-a-b}) = 2s\ min(\frac{1-b}{2-a-b}, \frac{1-a}{2-a-b}).$$

5.2 Properties and comparison with previous approaches

For BFs from $D_0^{\geq 0}$, where $a \geq b$, we have simply $IntC(a,b) = 2s\frac{1-a}{2-a-b}$, for BFs from $D_0^{\leq 0'}$, where $a \leq b$, we have $IntC(a,b) = 2s\frac{1-b}{2-a-b}$, and specially for BFs from S, where $a = b$, we have simply $IntC(a,b) = 2s\frac{1-a}{2-a-b} = s$. Thus $IntC$ has the nice property described in Subsection 4.1 for BFs from S.

From the last expression we, further, obtain the following observation.

Observation 1 *IntC defined by (1) coincides with Pl-IntC for BFs from S.*

For simple BFs from S_1 and S_2 we observe the following:

Observation 2 *IntC defined by (1) is equal to* 0 *for all simple (simple support) BFs, thus it coincides with Pl-IntC, cb-IntC and with auto-conflict for all BFs from* S_1 *and* S_2.

Comparing *IntC* with the previous definitions of internal conflict we obtain the following lemmata.

Lemma 1 *IntC defined by (1) is greater or equal to cb-IntC and auto-conflict for all indecisive BFs (BFs from* S*). Equality holds true just for* 0 *in the case of cb-IntC; IntC is further compatible with the definition of cb-IntC for* $0'$ *(it coincides with the upper bound of cb-IntC for* $0'$*. Equality holds true just for* 0 *and* $0'$ *in the case of auto-conflict.*

Proof. A proof follows the expressions for *cb-IntC* and auto-conflict on S: $s^2 \leq$ *cb-IntC*$(s,s) \leq 2s^2$ and $m_{(s,s)\textcircled{\cap}(s,s)}(\emptyset) = 2s^2$.

Lemma 2 *(i) IntC defined by (1) coincides with Pl-IntC for BFs from* G*. (ii) IntC defined by (1) is compatible with the bounds of cb-IntC for BFs from* G*. It is equal to the cb-IntC's upper bound for* $0'$ *and coincides with cb-IntC (it is equal to both bounds of cb-IntC) for* $(0,1)$ *and* $(1,0)$.

Proof. A proof follows the expressions for *Pl-IntC* and *cb-IntC* on G: *Pl-IntC*$(a, 1-a) = min(a, 1-a)$, $a(1-a) \leq$ *cb-IntC*$(s,s) \leq 2a(1-a)$ and $max(a, 1-a) \geq \frac{1}{2}$, $2max(a, 1-a) \geq 1$.

Corollary 1 *IntC defined by (1) is less or equal to auto-conflict for all Bayesian BFs (BFs from* G*). Equality holds true just for* $0' = (\frac{1}{2}, \frac{1}{2})$*,* $(0,1)$ *and for* $(1,0)$.

Proof. A proof follows the expression for auto-conflict on G: $m_{(a,1-a)\textcircled{\cap}(a,1-a)}(\emptyset) = 2a(1-a)$.

We have seen that *IntC* coincides with *Pl-IntC* on the border of D_0 (S_1, S_2, G) and on S. It does not hold true for all BFs from D_0, see e.g. $(\frac{1}{2}, \frac{1}{4})$: *IntC*$(\frac{1}{2}, \frac{1}{4}) = \frac{4}{15}$, whereas *Pl-IntC*$(\frac{1}{2}, \frac{1}{4}) = \frac{4}{16}$. Equality of *IntC* and *Pl-IntC* for fixed $b = \frac{1}{4}$ holds true only for $a = \frac{1}{4}$ and $a = \frac{3}{4}$ in the case of $a \geq b$; similarly for b fixed to $\frac{1}{10}$, $\frac{4}{10}$ or to another values. (*IntC*$(\frac{1}{2}, \frac{1}{10}) = \frac{5}{42} >$ *Pl-IntC*$(\frac{1}{2}, \frac{1}{10}) = \frac{1}{10}$, *IntC*$(\frac{1}{2}, \frac{4}{10}) = \frac{40}{99} >$ *Pl-IntC*$(\frac{1}{2}, \frac{4}{10}) = \frac{4}{10}$.) We can summarize a relation of *IntC* and *Pl-IntC* as it follows.

Theorem 3 *IntC defined by (1) coincides with Pl-IntC on the border of* D_0 *(on its subalgeras* S_1, S_2, G*) and on* S.

Hypothesis 1 *IntC*$(Bel) \geq$ *Pl-IntC*(Bel) *for any belief function defined on a 2-element frame of discernment* Ω_2*, where equality holds true for any Bel defined on* S, S_1, S_2, G.

Analogically, we can summarize a relations of *IntC* to *cb-IntC* and to auto-conflict. A general relation of *IntC* to auto-conflict is more complicated than its relation to *Pl-IntC* and *cb-IntC*.

Theorem 4 *IntC defined by (1) is compatible with cb-IntC on the border of D_0 (on its subalgeras S_1, S_2, G) and IntC $\geq$ cb-IntC on S.*

Hypothesis 2 *IntC(Bel) is compatible with limits of cb-IntC(Bel) or it is greater than cb-IntC(Bel) for any belief function defined on a 2-element frame of discernment Ω_2.*

Theorem 5 *IntC defined by (1) is less or equal to auto-conflict on the border of D_0 (on its subalgeras S_1, S_2, G), equality holds true for S_1, S_2 and $0'$;*
IntC is greater or equal to auto-conflict on S, equality holds true for 0 and $0'$.

Statements of the Theorems 3 – 5 follow Observations 1, 2, Lemmata 1, 2 and Corollary 1.

6 Conclusion

This contribution presents theoretical results regarding conflicts of belief functions defined on two-element frame of discernment. Interpretation of conflicting part of belief functions is analysed here. Based on this analysis a new definition of internal conflict of belief functions is introduced. Basic properties of internal conflict according to the new definition and its comparison with previous approaches (plausibility internal conflict, combinational internal conflict and auto-conflict) are presented, especially for simple belief functions, indecisive belief functions, and Bayesian belief functions.

For general belief functions on two-element frame the comparison is partly sketched in form of hypotheses and remains as a topic for further research.

This theoretical contribution improves general understanding of conflict of belief functions and entire nature of belief functions. Correct understanding of conflicts may, consequently, improve a combination of conflicting belief functions in practical applications of belief functions.

References

[1] Cobb B. R., Shenoy P. P. (2003), A Comparison of Methods for Transforming Belief Functions Models to Probability Models. In: Nielsen T.D., Zhang N.L. (eds.), ECSQARU 2003. LNAI, vol.2711, pp.255–266, Springer.

[2] Daniel M. (1995), Algebraic structures related to Dempster-Shafer theory. In: Bouchon-Meunier B., Yager R. R., Zadeh L. A. (eds.), *Advances in Intelligent Computing - IPMU'94*. LNCS, vol. 945, pp. 51–61, Springer.

[3] Daniel M. (2005), Probabilistic Transformations of Belief Functions. In: Godo L. (ed.), ECSQARU 2005. LNAI, vol. 3571, pp. 539–551. Springer.

[4] Daniel M. (2010), Conflicts within and between Belief Functions. In: Hüllermeier E., Kruse R., Hoffmann E. (eds.), *IPMU 2010*. LNAI, vol. 6178, pp. 696–705. Springer-Verlag, Berlin Heidelberg.

[5] Daniel M. (2010), Properties of Conflicts of Belief Functions on Two-Element Frames of Discernment. In: Itoh T., Suzuki K. (eds.), *Proceedings of 13th Czech-Japan Seminar on Data Analysis and Decision Making in Service Science,* pp. 73–78. University of Commerce, Otaru.

[6] Daniel M. (2011), Non-conflicting and Conflicting Parts of Belief Functions. In: Coolen F., de Cooman G., Fetz T., Oberguggenberger M. (eds.), *ISIPTA'11; Proceedings of the 7th ISIPTA*, pp. 149–158. Innsbruck

[7] Daniel M. (2012), Conflict between Belief Functions on Two-Element Frames of Discernment. In: Inuiguchi M. et al. (eds.), *Proceedings of 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty,* pp. 29–36. Osaka University, Osaka.

[8] Daniel M. (2013), *Properties of Plausibility Conflict of Belief Functions.* In: Rutkowski L. et al. (eds.), *ICAISC 2013, Part I.* LNAI 7894: 235-246. Springer-Verlag Berlin Heidelberg.

[9] Daniel M. (2013), Belief Functions: a Revision of Plausibility Conflict and Pignistic Conflict. In: In: Liu W. et al. (eds.), *SUM 2013.* LNAI 8078: 190–203. Springer-Verlag Berlin Heidelberg.

[10] Destercke S. Burger T. (2012), Revisiting the Notion of Conflicting Belief Functions. In: Denœux T., Masson M. H. (eds.), *Belief Functions: Theory and Applications*, AISC 164, pp 153–160. Springer-Verlag.

[11] Haenni R. (2008), Aggregating Referee Scores: an Algebraic Approach. *COMSOC'08, 2nd International Workshop on Computational Social Choice*, Liverpool, UK.

[12] Hájek P., Havránek T., Jiroušek R. (1992), *Uncertain Information Processing in Expert Systems.* CRC Press, Boca Raton, Florida.

[13] Hájek P., Valdés J. J. (1991), Generalized algebraic foundations of uncertainty processing in rule-based expert systems (dempsteroids). *Computers and Artificial Intelligence* **10** (1) 29–42.

[14] Liu W. (2006), Analysing the degree of conflict among belief functions. *Artificial Intelligence* **170**, 909–924.

[15] Martin A. (2012), About Conflict in the Theofy of Belief Functions. In: Denœux T., Masson M. H. (eds.), *Belief Functions: Theory and Applications*, Proc. of 2nd Int. Conf. on Belief Functions, AISC 164, pp 161–168, Springer.

[16] Shafer G. (1976), *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey.

[17] Smets P. (1990), The combination of evidence in the transferable belief model. IEEE-Pattern analysis and Machine Intelligence 12, 447–458.

[18] Smets P. (2007), Analyzing the combination of conflicting belief functions. *Information Fusion* **8**, 387–412.

The Linear Space of Transferable Utility Games

Tetsuzo Tanino
Division of Electrical, Electronic and Information Engineering
Osaka University
tanino@eei.eng.osaka-u.ac.jp

Yoshifumi Kusunoki
Division of Electrical, Electronic and Information Engineering
Osaka University
kusunoki@eei.eng.osaka-u.ac.jp

Abstract

In this paper we deal with transferable utility games (TU-games). Each TU-game is characterized by a set of players and a real-valued function on the power set of the player set. We fix the player set and identify a TU-game with a function in this paper. Then the set of TU-games forms a linear space with respect to ordinary function addition and scalar multiplication. We consider some bases in this space, one of which is new. The following topics are investigated: Relationships among the coefficients of linear combination representation of a game by the bases, characterization of additive games and convex games, description of the Shapley value of a game in terms of the coefficients by the bases, and the expressions of the multilinear extension and the Lovász extension of a game.

1 Introduction

Among cooperative games, TU-games are fundamental and important. In this paper we fix the player set and identify a TU-game with a function defined on the power set of the player set. The value of a function for each subset is called the worth. Then the set of TU-games forms a linear space with respect to ordinary function addition and scalar multiplication.

When we introduce the unanimity games, they are linearly independent and forms a basis in the linear space. The coefficients of the linear combination representation of a game are called the Harsanyi dividends [4, 5] and provide a number of interesting results. We may introduce other bases, and in fact Driessen, Khmelnitskaya and Sales introduced the basis consisting of complementary unanimity games. In this paper we introduce the basis consisting of

support games. If we regard the unanimity game as a conjunctive support game, the support game is a disjunctive support game. Moreover it is closely related to the complementary unanimity game.

In this paper we investigate the following topics. 1) Relationships among the coefficients of linear combination representation of a game by the above bases. 2) Characterization of important classes of games such as additive games and convex games. 3) Description of the Shapley value of a game in terms of the coefficients by the above bases. 4) Expressions of multilinear extensions and Lovász extensions of games by the coefficients.

Throughout this paper we distinguish proper inclusion $S \subset N$ from ordinary inclusion $S \subseteq N$. We denote the number of elements in a set by its corresponding small letter, i.e., $s = |S|$, $t = |T|$, $s' = |S'|$, and so on. We also use some simplified notations such as $v(i) = v(\{i\})$, $S \cup i = S \cup \{i\}$, and so on.

2 The linear space of transferable utility games and some bases

Let $N = \{1, 2, \ldots, n\}$ be a finite set of players. A transferable utility game (TU-game) on N is a function $v : 2^N \to \mathbf{R}$ satisfying that $v(\emptyset) = 0$. We denote the set of all TU-games on N by $\mathcal{G}^N$ or simply $\mathcal{G}$, because N is fixed throughout this paper.

It is clear that $\mathcal{G}$ is a $2^n - 1$ dimensional linear space with ordinary addition and scalar multiplication of functions. We can consider some bases in $\mathcal{G}$.

1) Identity games (standard basis games): For any nonempty $T \subseteq N$,

$$e_T(S) = \begin{cases} 1 & \text{if} \quad S = T, \\ 0 & \text{if} \quad S \neq T. \end{cases}$$

This basis corresponds to the standard basis in the Euclidean space and

$$v = \sum_{\emptyset \neq T \subseteq N} v(T) e_T.$$

2) Unanimity games: For any nonempty $T \subseteq N$

$$u_T(S) = \begin{cases} 1 & \text{if } S \supseteq T, \\ 0 & \text{otherwise.} \end{cases}$$

The worth $u_T(S)$ is 1 if all the players in T are included in S and is 0 otherwise. In this sense the players in T support the game conjunctively.

$$v = \sum_{\emptyset \neq T \subseteq N} d(T) u_T.$$

The coefficients $d(T)$ are called the Harsanyi dividends (Harsanyi [4, 5]) or the Möbius transform (Grabisch, Marischal and Roubens [3]) of v.

3) Complementary unanimity games: For any $T \subset N$

$$\bar{u}_T(S) = \begin{cases} 1 & \text{if } S \subseteq N \setminus T, S \neq \emptyset, \text{ i.e., if } S \cap T = \emptyset, S \neq \emptyset, \\ 0 & \text{otherwise} \end{cases}$$

This basis was introduced by Driessen, Khmelnitskaya and Sales [2]. The worth $\bar{u}_T(S)$ is 1 if no player in T is contained in a nonempty set S and is 0 otherwise. In this sense any player in T is a veto player.

$$v = \sum_{T \subset N} \bar{d}(T)\bar{u}_T.$$

4) Support games: For any nonempty $T \subseteq N$

$$w_T(S) = \begin{cases} 1 & \text{if } S \cap T \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

The worth $w_T(S)$ is 1 if some player in T is contained in S and is 0 otherwise. In this sense any player in T can support the game disjunctively and therefore we call this game w_T a support game by T.

$$v = \sum_{\emptyset \neq T \subseteq N} \gamma(T) w_T.$$

Definition 1 *A game $v \in \mathcal{G}$ is convex if*

$$v(S \cup T) + v(S \cap T) \geq v(S) + v(T)$$

for all $S, T \subseteq N$. If $-v$ is convex, v is said to be concave.

Proposition 1 *1) For each nonempty $T \subseteq N$, the unanimity game u_T is convex.*
2) For each $T \subset N$, the complementary unanimity game $\bar{u}_T$ is subadditive. However it is generally neither convex nor concave.
3) For each nonempty $T \subseteq N$, the support game w_T is concave.

Proposition 2 *For each nonempty $T \subset N$*

$$w_T(S) = 1 - \bar{u}_T(S)$$

for any nonempty subset S of N.

Remark 1 *In the complementary unanimity game basis, the game $\bar{u}_\emptyset$ is added to the set $\{\bar{u}_T | \emptyset \neq T \subset N\}$. On the other hand in the support game basis, the game w_N is added to the set $\{w_T | \emptyset \neq T \subset N\}$. We should note that $\bar{u}_\emptyset(S) = w_N(S) = 1$ for any nonempty $S \subseteq N$.*

3 Relationships among the representations by the bases

In this section we consider relationships among the representations by the bases. Namely we try to represent the coefficients by other coefficients.

$$v(S) = \sum_{\emptyset \neq T \subseteq N} d(T) u_T(S) = \sum_{\emptyset \neq T \subseteq S} d(T)$$

$$v(S) = \sum_{T \subset N} \bar{d}(T) \bar{u}_T(S) = \sum_{T \cap S = \emptyset} \bar{d}(T)$$

$$v(S) = \sum_{\emptyset \neq T \subseteq N} \gamma(T) w_T(S) = \sum_{T \cap S \neq \emptyset} \gamma(T)$$

The coefficients $\bar{d}(T)$ and $\gamma(T)$ are closely related to the co-Möbius transform $b(T)$ of v given by

$$b(T) = \sum_{S \supseteq N \setminus T} (-1)^{n-s} v(S)$$

in Grabisch et al. [3]. In view of the result in [3], for any nonempty subset S of N

$$v(S) = \sum_{T \subseteq N \setminus S} (-1)^t b(T) = \sum_{T \subset N, T \cap S = \emptyset} (-1)^t b(T)$$

Therefore,

$$\bar{d}(T) = (-1)^t b(T), \ \ \forall T \subset N, \text{ and } \gamma(T) = (-1)^{t-1} b(T), \ \ \emptyset \neq \forall T \subseteq N.$$

Lemma 1 *For $T \subset S \subseteq N$, it holds that* $\sum_{T \subseteq R \subseteq S} (-1)^r = 0.$

Proposition 3 (e.g. Grabisch [3]) *For a nonempty subset S of N,*

$$d(S) = \sum_{\emptyset \neq T \subseteq S} (-1)^{s-t} v(T)$$

Proposition 4 (Driessen et al. [2]) *For a proper subset S of N,*

$$\bar{d}(S) = \sum_{T \supseteq N \setminus S} (-1)^{s-n+t} v(T).$$

Theorem 1 *For each nonempty $S \subseteq N$*

$$\gamma(S) = \sum_{T \subseteq S} (-1)^{t-1} v((N \setminus S) \cup T) = \sum_{T \supseteq N \setminus S} (-1)^{s-n+t-1} v(T)$$

Theorem 2 *For a nonempty $S \subseteq N$,*

$$d(S) = (-1)^{s-1} \sum_{T \supseteq S} \gamma(T)$$

Theorem 3 *For a nonempty $S \subseteq N$,*

$$\gamma(S) = (-1)^{s-1} \sum_{T \supseteq S} d(T)$$

Theorem 4 *For each $S \subset N$*

$$\bar{d}(S) = \begin{cases} -\gamma(S) & \text{if} \quad \emptyset \neq S \subset N, \\ \displaystyle\sum_{\emptyset \neq T \subseteq N} \gamma(T) & \text{if} \quad S = \emptyset. \end{cases}$$

Conversely, for each nonempty $S \subseteq N$

$$\gamma(S) = \begin{cases} -\bar{d}(S) & \text{if} \quad \emptyset \neq S \subset N, \\ \displaystyle\sum_{T \subset N} \bar{d}(T) & \text{if} \quad S = N. \end{cases}$$

Theorem 5 *For each nonempty $S \subseteq N$*

$$d(S) = (-1)^{s-1} \sum_{S \subsetneq T \subset N} \bar{d}(T).$$

Conversely, for each $S \subset N$

$$\bar{d}(S) = (-1)^{s} \sum_{\emptyset \neq T \supseteq S} d(T).$$

4 Characterization of some classes of games

In this section we characterize two important classes of games, i.e. additive games and convex games, in terms of the coefficients of the basis representations.

4.1 Additive games

Definition 2 *A game $v \in \mathcal{G}$ is said to be additive if*

$$v(S \cup T) = v(S) + v(T), \ \ \forall S, T \subseteq N, \ S \cap T = \emptyset$$

The following proposition is known well and the two theorems are new.

Proposition 5 *A game v is additive if and only if $d(S) = 0$ for all $S \subseteq N$ with $s > 1$. In this case $d(i) = v(i)$ for all $i \in N$.*

Theorem 6 *A game v is additive if and only if $\bar{d}(S) = 0$ for all $S \subset N$ with $s > 1$ and $\bar{d}(\emptyset) + \displaystyle\sum_{i \in N} \bar{d}(i) = 0$. In this case $\bar{d}(i) = -v(i)$ for all $i \in N$.*

Theorem 7 *A game v is additive if and only if $\gamma(S) = 0$ for all $S \subseteq N$ with $s > 1$. In this case $\gamma(i) = v(i)$ for all $i \in N$.*

4.2 Convex games

The following two equivalent propositions provide the necessary and sufficient conditions for the convexity of games in terms of the dividends.

Proposition 6 (Kuipers, Vermeulen and Voorneveld [6]) *A game $v \in \mathcal{G}$ is convex if and only if*

$$\sum_{T \subseteq S} d(T \cup \{i,j\}) \geq 0$$

for all $i,j \in N$ $(i \neq j)$ and all $S \subseteq N \setminus \{i,j\}$.

Proposition 7 (Chateauneuf and Jaffray [1]) *A game $v \in \mathcal{G}$ is convex if and only if*

$$\sum_{i,j \in R \subseteq S} d(R) \geq 0,$$

for all $S \subseteq N$ with $i \neq j \in S$.

We can also characterize convex games as follows.

Theorem 8 *A game $v \in \mathcal{G}$ is convex if and only if*

$$\sum_{T \subseteq S} \gamma(T \cup \{i,j\}) \leq 0$$

for all $i,j \in N$ $(i \neq j)$ and all $S \subseteq N \setminus \{i,j\}$.

Theorem 9 *A game $v \in \mathcal{G}$ is convex if and only if*

$$\sum_{\{i,j\} \subseteq R \subseteq S} \bar{d}(R) \geq 0,$$

for all $S \subset N$ with $i \neq j \in S$ and

$$\sum_{\{i,j\} \not\subseteq R \subset N} \bar{d}(R) \leq 0$$

for all $i,j \in N$ with $i \neq j$.

5 Shapley Values

The Shapley value for $v \in \mathcal{G}$ is the most fundamental single-valued solution (Shapley [9]). Let π be a permutation on N and define the marginal contribution vector $m^\pi(v) = (m_1^\pi(v), \ldots, m_n^\pi(v))$ at π for v by

$$m_i^\pi(v) = v(P(\pi,i) \cup \{i\}) - v(P(\pi,i))$$

where $P(\pi, i) = \{j \in N | \pi(j) < \pi(i)\}$. Then the Shapley value for v is defined by

$$\varphi(v) = \frac{1}{n!} \sum_{\pi \in \Pi_N} m^\pi(v)$$

where Π_N denoted the set of all permutations on N.

The Shapley values can be obtained from the coefficients by the bases. As for the formula by the Harsanyi dividends is known well. The other two representations are obtained in this paper.

Theorem 10 *The Shapley value of a game $v \in \mathcal{G}$ is given by*

$$\varphi_i(v) = \sum_{T \ni i} \frac{d(T)}{t} = \sum_{T \ni i} \frac{\gamma(T)}{t} = \frac{1}{n} \sum_{T \subset N} \bar{d}(T) - \sum_{N \supset T \ni i} \frac{\bar{d}(T)}{t}.$$

6 Extensions

A game $v \in \mathcal{G}$ is a function from 2^N to $\mathbf{R}$. Since each coalition $S \in 2^N$ can be identified with the vector $e^S \in \{0,1\}^n$ defined by $e_i^S = 1$ if $i \in S$ and $e_i^S = 0$ otherwise. Therefore v can be regarded as a function from $\{0,1\}^n$ to $\mathbf{R}$. Then we may consider to extend the domain to $[0,1]^n$. Namely we may extend v to a function $\xi_v : [0,1]^n \to \mathbf{R}$. Two typical examples of extensions are the multilinear extension and the Lovász extension. In this section we consider the expressions of these extensions in terms of the coefficients of the basis representations. As for the results concerning the Harsanyi dividends are known well.

6.1 Multilinear extensions

The multilinear extension of v is defined by the following formula (Owen [8]).

$$m_v(x) = \sum_{S \subseteq N} \left(\prod_{i \in S} x_i \prod_{i \notin S} (1 - x_i) \right) v(S).$$

Theorem 11 *The multilinear extension of $v \in \mathcal{G}$ is given by*

$$\begin{aligned} m_v(x) &= \sum_{\emptyset \neq T \subseteq N} \left(\prod_{i \in T} x_i \right) d(T) = \sum_{\emptyset \neq T \subseteq N} \left(1 - \prod_{i \in T} (1 - x_i) \right) \gamma(T) \\ &= \sum_{T \subset N} \left(\prod_{i \in T} (1 - x_i) - \prod_{i \in N} (1 - x_i) \right) \bar{d}(T), \end{aligned}$$

where $\prod_{i \in \emptyset} (1 - x_i)$ *is supposed to be* 1.

6.2 Lovász extensions

Given a vector $x \in [0,1]^n$, we define the following level sets:

$$[x]_h = \{i \in N \mid x_i \geq h\}, \quad h \in [0,1].$$

Then the Lovász extension of v can be defined as follows (Lovász [7]):

$$l_v(x) = \int_0^1 v([x]_h)dh.$$

Theorem 12 *The Lovász extension of $v \in \mathcal{G}$ is given by*

$$\begin{aligned} l_v(x) &= \sum_{\emptyset \neq T \subseteq N} \left(\min_{i \in T} x_i \right) d(T) = \sum_{\emptyset \neq T \subseteq N} \left(\max_{i \in T} x_i \right) \gamma(T) \\ &= \sum_{T \subset N} \left(\max_{i \in N} x_i - \max_{i \in T} x_i \right) \bar{d}(T), \end{aligned}$$

where $\max_{i \in \emptyset} x_i$ is supposed to be 0.

References

[1] Chateauneuf A., J.-Y. Jaffray (1989), Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, *Math. Soc. Sci.* **17**, 263–283

[2] Driessen T.S.H., Khmelnitskaya A.B., Sales J. (2012), 1-concave basis for TU games and the library game, *TOP* **20**, 578–591

[3] Grabisch M., Marichal J.-L., Roubens M. (2000), Equivalent representations of set functions, *Math. Methods Oper. Res.* **25**, 157–178

[4] Harsanyi J.C. (1959) A bargaining model for cooperative n-person games, in Tucker A.W., Luce R.D. (eds.) *Contributions to the Theory of Games*, Princeton Univ. Press, 325–355

[5] Harsanyi J.C. (1963), A simplified bargaining model for the n-person cooperative game, *Int. Econ. Rev.* **4**, 194–220

[6] Kuipers J., Vermeulen D., Voorneveld M. (2010), A generalization of the Shapley-Ichiishi result, *Int. J. Game Theory* **39**, 585–602

[7] Lovász L. (1983), Submodular functions and convexity, in Bachem A. et al. (eds.), *Mathematical Programming: The State of the Art*, Springer-Verlag, 235–257

[8] Owen G. (1972), Multilinear extensions of games, *Man. Sci.* **18**, 64–79

[9] Shapley L.S. (1953), A value for n-person games, in: Tucker A.W., Kuhn H.W. (eds.) *Contributions to the Theory of Games*, Princeton Univ. Press, 307–317

F-Transform for Image Reduction

Irina Perfilieva, Petr Hurtík
Institute for Research and Applications of Fuzzy Modeling
University of Ostrava
e-mail: Irina.Perfilieva,Petr.Hurtik@osu.cz

Abstract

We present a new method of color images reduction, which is based on the direct F-transform with a generalized fuzzy partition. We show that complexity of this method is linear. We also analyze measures (MSE and SSIM) that are commonly used for estimation a quality of reduced images and see that all these measures have better values on the newly proposed method.

Keywords: F-transform, generalized partition, image reduction, MSE, SSIM

1 Introduction

Various signal processing methods have been developed rapidly in recent decades to address numerous problems of multimedia and communications applications, as well as of ever increasing signal processing software. This work is focused on the issue of *image reduction* which is connected with a compact visual representation of an image required by mobile phones, photo cameras, tablets, etc.

There are at least two different meanings of image reduction. In [3], image reduction is considered as a (shrinking) operator which reduces the resolution of an image in order to speed up computations. Usually a low-pass filter is used for this purpose. In [1], image reduction is understood as a technique which similar to the image compression, aims at

(i) minimization of the number of bits required to represent an image,

(ii) maintaining acceptable quality of a reduced image.

In our paper, we consider the problem of image reduction in the second meaning.

In this paper, we present a new method based on the direct F-transform for color image reduction. We justify its suitability for image reduction proving that the sequential application of the direct and inverse F-transform to an image works as an approximator. We analyze how a quality of approximation by the inverse F-transform depends on the choice of basic functions in a fuzzy

partition. Moreover, we estimate complexity of a corresponding algorithm for image reduction.

We compare the results obtained using the new F-transform based reduction algorithm with the ones obtained using interpolation and aggregation. The comparison is performed on the basis of (*i*) two quality measures: MSE and SSIM, (*ii*) computation time and (*iii*) noise removing ability. The measure SSIM is used to estimate a quality of compression made by JPEG and JPEG2000, and this is the reason why it was selected. Moreover, we discuss what actually is measured by quality measures MSE and SSIM.

2 Image reduction methods

2.1 Preliminaries: Image reduction and quality measures

This work is focused on the issue of *image reduction* – a technique which aims at a compact representation, while maintaining acceptable quality. Let us explain this characterization using denotation that will be kept throughout this paper.

A (gray-scale) image is identified with a representing it (intensity) function $u : [1, N] \times [1, M] \to [0, 255]$ where the domain $[1, N] \times [1, M] = \{(i, j) \mid i = 1, \ldots, N; j = 1, \ldots, M\}$ and the range $[0, 255]$ contain only natural numbers. A color RGB-image is represented by three intensity functions u^R, u^G and u^B, each in the respective color band. If not explicitly mentioned, we assume that the image is gray-scale. A *reduced (compact) representation* $\bar{u} : n \times m \to [0, 255]$ of u is determined by the *reduction ratio* $\rho = \frac{NM}{nm}$ where $n < N$, $m < M$, and N, M (n, m) are respective sizes of u and its reduction $\bar{u}$. Usually, the reduction ratio is written in the form $\rho : 1$.

In this paper, we assume that reduction is performed on the basis of the "block-to-pixel" scheme. This means that the domain $[1, N]\times[1, M]$ of the image function u is divided into $N_b \times M_b$-sized blocks $B_{1,1}, \ldots, B_{n,m}$, where $N_b \cdot M_b = \rho$, each block $B_{i,j}$ is replaced by one pixel (i, j) and this pixel is assigned a new intensity value $\bar{u}(i, j)$. The way how this value is computed specifies the method: subsampling, interpolation [3], aggregation [1], F-transform (below).

The following two criteria: MSE or SSIM are chosen for estimation quality of a reconstructed image and comparison between various reduction methods including the one proposed in this paper. Below, we characterize each criterion and explain why it has been selected.

Mean Square Error (shortly, MSE)

$$\mathrm{MSE}(u, \hat{u}) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (u(i, j) - \hat{u}(i, j))^2}{NM}, \tag{1}$$

where u denotes an original image and $\hat{u}$ its reconstruction. It is easy to see that MSE is a squared Euclidean distance between u and $\hat{u}$ in the corresponding vector space.

Structure Similarity Index (shortly SSIM)

$$\mathrm{SSIM}(u, \hat{u}) = f(\ell(u, \hat{u}), c(u, \hat{u}), s(u, \hat{u})) \tag{2}$$

is a compound function that includes measures of luminance ℓ, contrast c and structure s. The detailed expression of SSIM requires extended explanations and thus, is omitted. SSIM was introduced with the purpose to measure the quality of a single (gray-scale) image by comparing it with the ideal representation of the same image. Similarly to MSE, the SSIM value of a color image can be obtained as an arithmetic mean of three SSIM values in each color band. Opposite to MSE, the SSIM value measures a similarity of two images according to the principle: the higher, the better. We chose SSIM because it shows to be consistent with a human eye perception and moreover, it is used to estimate a quality of compression made by JPEG and JPEG2000.

3 F-transform for Functions of One Variable

In this section, we introduce a new modification of the method of F-transform and explain how it differs from the original version in [4]. We remind that according to [4], the F-transform of a function f of one variable is determined by a fuzzy partition of the domain of f. A fuzzy partition consists of a finite set of basic functions (fuzzy sets) that are characterized axiomatically.

The proposed modification consists in using less axioms in the definition of fuzzy partition. Especially, we drop three of them: normality, convexity and orthogonality (the latter is also known as the Ruspini condition). As a result, a newly defined generalized fuzzy partition has additional degrees of freedom that can be tuned. In the case of image reduction, this fact leads to better results than were announced in [1].

3.1 Generalized fuzzy partitions

A *generalized fuzzy partition* appeared in [6] in connection with the notion of the higher-degree F-transform. Its even weaker version was implicitly introduced in [2] for the purpose of meeting the requirements of image compression. We summarize both these notions and propose the following definition.

Definition 1 *Let $[a, b]$ be an interval on the real line $\mathbb{R}$, $n \geq 2$, and let $x_1, \ldots, x_n$ be nodes such that $a \leq x_1 < \ldots < x_n \leq b$. Let $[a, b]$ be covered by the intervals $[x_k - h'_k, x_k + h''_k] \subseteq [a, b]$, $k = 1, \ldots, n$, such that their left and right margins $h'_k, h''_k \geq 0$ fulfill $h'_k + h''_k > 0$.*

We say that basic functions $A_1, \ldots, A_n : [a, b] \to [0, 1]$ constitute a generalized fuzzy partition *of $[a, b]$ (with nodes $x_1, \ldots, x_n$ and margins h'_k, h''_k, $k = 1, \ldots, n$), if for every $k = 1, \ldots, n$, the following three conditions are fulfilled:*

1. *(locality) — $A_k(x) > 0$ if $x \in (x_k - h'_k, x_k + h''_k)$, and $A_k(x) = 0$ if $x \in [a, b] \setminus (x_k - h'_k, x_k + h''_k)$;*

2. *(continuity)* — A_k *is continuous on* $[x_k - h'_k, x_k + h''_k]$;

3. *(covering)* — *for* $x \in [a, b]$, $\sum_{k=1}^{n} A_k(x) > 0$.

It is important to remark that by conditions of *locality* and *continuity*,

$$\int_a^b A_k(x)dx > 0.$$

In what follows, we will omit the word "generalized" whenever we refer to a fuzzy partition. Moreover, we assume that in every below considered partition, basic functions $A_1, \ldots, A_n$ are *normalized* in the sense that $A_k(x_k) = 1$, $k = 1, \ldots, n$.

The illustration of (h, h')-uniform fuzzy partitions where $h = 3$ and $h' = 2$ is in Fig. 1.

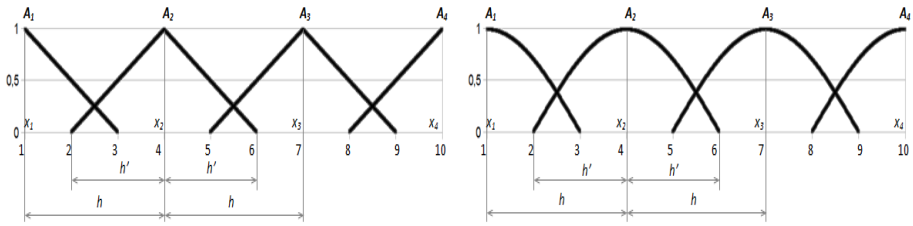

Figure 1: Generalized $(3, 2)$-uniform fuzzy partitions: triangular-shaped (left) and sinusoidal-shaped (right).

3.2 Direct F-transform

In this Section, we give a definition of the integral and discrete (direct) F-transform according to [4] and recall those properties of it that will be used for image reduction. We assume that the universe is an interval $[a, b]$ and $x_1 < \ldots < x_n$ are fixed nodes from $[a, b]$ such that $x_1 = a$, $x_n = b$ and $n \geq 2$. Let $A_1, \ldots, A_n$ be basic functions that form a fuzzy partition of $[a, b]$ according to Definition 1. The latter will be fixed throughout this Section. Let $C([a, b])$ be the set of continuous functions on the interval $[a, b]$. The following definition introduces the integral F-transform of a function $f \in C([a, b])$.

Definition 2 *Let* $A_1, \ldots, A_n$ *be basic functions that form a fuzzy partition of* $[a, b]$ *and* f *be any function from* $C([a, b])$. *We say that the* n*-tuple of real numbers* $\mathbf{F}[f] = (F_1, \ldots, F_n)$ *given by*

$$F_k = \frac{\int_a^b f(x)A_k(x)dx}{\int_a^b A_k(x)dx}, \qquad k = 1, \ldots, n, \tag{3}$$

is the integral F-transform of f *with respect to* $A_1, \ldots, A_n$.

The discrete form of the F-transform is applied to functions f that are defined on a finite set $P = \{p_1, \ldots, p_l\} \subseteq [a, b]$. We assume that the set P is *sufficiently dense with respect to the fixed partition*, i.e.,

$$(\forall k)(\exists j) A_k(p_j) > 0.$$

Then, the discrete F-transform $\mathbf{F}[f] = (F_1, \ldots, F_n)$ of f is defined as follows:

$$F_k = \frac{\sum_{j=1}^{l} f(p_j) A_k(p_j)}{\sum_{j=1}^{l} A_k(p_j)}, \qquad k = 1, \ldots, n. \tag{4}$$

3.3 Inverse F-transform

The inverse F-transform establishes a backward correspondence from the set of n-dimensional vectors to the set of continuous/discrete functions. This correspondence is not inverse with respect to the direct F-transform, but if both are applied sequentially then the result approximates the original function.

Definition 3 *Let $A_1, \ldots, A_n$ be basic functions that form a generalized fuzzy partition of $[a, b]$ and f be a function from $C([a, b])$. Let $\mathbf{F}[f] = (F_1, \ldots, F_n)$ be the F-transform of f with respect to $A_1, \ldots, A_n$. Then, the function $\hat{f} : [a, b] \to \mathbb{R}$ represented by*

$$\hat{f}(x) = \frac{\sum_{k=1}^{n} F_k A_k(x)}{\sum_{k=1}^{n} A_k(x)}, \qquad x \in [a, b], \tag{5}$$

is called the inverse F-transform.

In the discrete case, the inverse F-transform $\hat{f}$ is defined using the same expression (5) that is applied to set P where the original discrete function was defined.

The below given Theorem 4 demonstrates that the inverse F-transform $\hat{f}$ approximates a continuous function f with arbitrary precision. Thus, it explains why the F-transform has convincing applications in various fields, including image processing.

Theorem 4 *Let f be a continuous function on $[a, b]$. Then, for any $\varepsilon > 0$, there exist h_ε such that for any $h_\varepsilon/2 < h' \leq h_\varepsilon$ and any (h_ε, h')-uniform generalized fuzzy partition of $[a, b]$, the corresponding inverse F-transform $\hat{f}_\varepsilon$ of f fulfills*

$$|f(x) - \hat{f}_\varepsilon(x)| \leq \varepsilon, \; x \in [a, b]. \tag{6}$$

4 New F-transform Based Image Reduction and Its Effectiveness

Image compression was the first application of the F-transform to image processing. In [4], we proposed to represent a compressed image by a matrix of

F-transform components computed over a uniform Ruspini partition of the image domain. The reconstruction to a full-size image was done using the inverse F-transform. This method (we will call it the " simple F-transform based compression") does not take advantage of any property of the original image and therefore, its quality is not very high. In [2], we proposed another compression methods and proved that a proper choice of a fuzzy partition improves a quality of the reconstructed image.

In this Section, we introduce a new F-transform based image reduction that is based on a generalized fuzzy partition. We will see that similar to the case of compression, a tuning of fuzzy partition leads to better results in reduction. We will see that the method of F-transform that is based on a specially designed (generalized) fuzzy partition is the most suitable reduction method from both quality (measured by MSE or SSIM) and complexity points of view.

4.1 Proposed algorithm and its complexity

Below, we introduce the new F-transform based reduction algorithm. It is applied to an image function $u : [1, N] \times [1, M] \to [0, 255]$ where the domain and the range contain only natural numbers. The following expression for the F-transform components U_{kl}, $k = 1, \ldots, n$, $l = 1, \ldots, m$, of u is a direct generalization of (4):

$$U_{kl} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} f(i,j) A_k(i) B_l(j)}{\sum_{i=1}^{N} \sum_{j=1}^{M} A_k(i) B_l(j)}. \tag{7}$$

In (7), it is assumed that basic functions $A_1, \ldots, A_n$ ($B_1, \ldots, B_m$) establish a fuzzy partition of $[1, N]$ ($[1, M]$) and that the set $[1, N]$ ($[1, M]$) is sufficiently dense with respect to $A_1, \ldots, A_n$ ($B_1, \ldots, B_m$).

The description of the algorithm is given in terms of procedures, i.e., without unnecessary technical details.

Algorithm *FT* of image reduction on the basis of the F-transform with generalized fuzzy partition

Inputs: $N \times M$ image u, reduction ratio ρ.
Output: Reduced image $\bar{u}$.

Step 1. Find values $n, m \geq 2$ such that $\frac{NM}{nm} = \rho$. Let $h_x = \frac{N-1}{n-1}$, $h_y = \frac{M-1}{m-1}$.
Step 2. Choose n h_x-equidistant nodes $x_1, \ldots, x_n \in [1, N]$ and m h_y-equidistant nodes $y_1, \ldots, y_m \in [1, M]$.
Step 3. Choose margins h'_x and h'_y, generating functions A_{0x} and A_{0y} and establish (h_x, h'_x)- and (h_y, h'_y)-uniform fuzzy partitions of $[1, N]$ and $[1, M]$, respectively.
Step 4. Compute the F-transform components U_{kl}, $k = 1, \ldots, n$, $l = 1, \ldots, m$, of u on the basis of (7) and arrange them into matrix $\mathbf{F}[u]$. Take $\mathbf{F}[u]$ as the output reduced image $\bar{u}$.

We claim that the complexity of Algorithm *FT* is *linear* with respect to the length of the input. In order to justify the claim, we estimate the complexity of the main *Step 4*. According to (7), a computation of the F-transform component U_{kl} can be taken over those pixels (i, j) that are "covered by" the product $A_k B_l$, i.e. fulfill $A_k(i)B_l(j) > 0$. Due to the uniformity of partition, for all k, l, $k = 1, \ldots, n$, $l = 1, \ldots, m$, a number of such pixels depends on the initial choice of margins h'_x and h'_y and it is a constant characteristic of the partition. Therefore, there is a constant number, say C, of operations involved into the computation of each component U_{kl}. Consequently, the total number of operations required by the *Step 4* is equal to Cnm or to CNM/ρ. Thus, the complexity of the *Step 4* is linear with respect to the product NM or with respect to the length of the input.

4.2 Optimal choice of parameters

In this Section, we analyze parameters of Algorithm *FT* in order to choose their optimal values that minimize MSE and maximize SSIM. Let us remark that for any input image, the output of the algorithm (reduced image comprised by the F-transform components) is fully determined by the choice of a fuzzy partition. This means that the parameters of the latter will be analyzed in this Section.

In Tables 2, we show quality measures MSE and SSIM for Algorithm *FT* where the reduction ratio $\rho = 9 : 1$ and the reconstruction is "pixel-to-block". Other parameters are as follows: distance between nodes $h = 3$ and margins $h' = 2, 3, 4$. Let us remark that except for $h' = 3$ (Ruspini partition), fuzzy partitions with $h' = 2, 4$ are of the generalized type. The optimal values of quality measures are highlighted.

Table 2. MSE and SSIM for Algorithm FT

MSE				**SSIM**			
Img No.	$h'=2$	$h'=3$	$h'=4$	Img No.	$h'=2$	$h'=3$	$h'=4$
1	**241**	262	307	1	**0.95**	0.94	0.93
2	**45**	48	56	2	**0.98**	0.98	0.97
3	**46**	49	57	3	**0.98**	0.98	0.97
4	**164**	178	210	4	**0.93**	0.93	0.91
5	**123**	131	154	5	**0.98**	0.97	0.97
6	**191**	205	237	6	**0.95**	0.95	0.94
7	**77**	83	97	7	**0.98**	0.97	0.97
8	**121**	131	158	8	**0.98**	0.98	0.98
9	**290**	307	348	9	**0.94**	0.93	0.92
10	**414**	445	523	10	**0.94**	0.93	0.92
11	**343**	367	414	11	**0.88**	0.87	0.85
Mean	**186.82**	200.55	232.82	Mean	**0.95**	0.94	0.93

It is immediate from Table 2, that the uniform $(3,2)$-fuzzy partition (where $h=3$ and $h'=2$) is the optimal with respect to the chosen input images and quality measures MSE and SSIM.

In Figure 2, the two color image, numbered by 5, illustrates reductions produced by the Algorithm FT where $h'=2,3,4$.

Figure 2: Image 5 (left) and its 9 : 1-reductions corresponding to $h'=2$ (top-right), $h'=3$ (middle-right), $h'=2$ (bottom-right).

It is visible that the sharpest reduction corresponds to the value of margin $h'=2$.

4.3 Comparison with interpolation and aggregation based image reduction algorithms

In this Section, various reduction algorithms are compared on a data set that contains 53 color images taken from

http://sipi.usc.edu/database/database.php/volume=textures.

The images are of different resolutions: 512×512 px (20 pieces), 256×256 px (8 pieces), 1024×1024 px (24 pieces) and 2250×2250 px (one piece). The reduction ratio $\rho = 9 : 1$ and the reconstruction is performed on the basis of the "pixel-to-block" scheme. For the sake of brevity, we will display standard statistics of MSE for all five algorithms: interpolation *INT* (bilinear *bl*, bicubic *bc* and Lanczos *Lns*), 2*AGG* and the F-transform *FT* with the optimal setting values: $h = 3$ and $h' = 2$.

In Table 5, we see that the F-transform based reduction is slightly better than Algorithm *AGG* of image reduction via aggregation and visibly better than interpolation methods (the best quality measures are printed in bold).

Table 5. MSE for FT, INT_{bl}, INT_{bc}, INT_{Lns}, AGG

Stat	FT	INT_{bl}	INT_{bc}	INT_{Lns}	AGG
Min	**32.6**	43.6	44.7	41.9	**32.6**
Q_1	**81.2**	113.7	113.8	121.4	82.0
Median	**102.1**	145.6	150.0	170.8	103.7
Mean	**146.5**	197.3	199.6	212.0	152.4
Q_3	**163.0**	246.3	242.0	262.6	167.3
Max	**517.0**	606.0	607.0	626.7	527.3

5 Conclusion

In this contribution, we introduced the F-transform with a generalized fuzzy partition and proved that the corresponding inverse F-transform works as an approximator. We showed that reduction of color images which is based on this type of F-transform has better quality (measured by criteria MSE and SSIM) than that obtained using the aggregation based image reduction algorithms from [1]. We estimated the complexity of the F-transform based reduction and proved that it is linear with respect to the length of the input. We showed that reduction of color images which is based on this type of F-transform has better quality (measured by the criterion MSE) than that obtained using the interpolation and aggregation based image reduction algorithms.

Acknowledgement

This work relates to Department of the Navy Grant N62909-12-1-7039 issued by Office of Naval Research Global. The United States Government has a royalty-

free license throughout the world in all copyrightable material contained herein. Additional support was given also by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] G. Beliakov, H. Bustince, and D. Paternain, "Image reductions using means of discrete product lattices," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1070–1083, march 2012.

[2] P. Hurtik and I. Perfilieva, "Image compression methodology based on fuzzy transform," in *Advances in Intelligent and Soft Computing. Proc. Intern. Conf. on Soft Computing Models in Industrial and Environmental Applications (SoCo2012)*, 2012, pp. 525–532.

[3] E. Karabassis and M. E. Spetsakis, "An analysis of image interpolation, differentiation, and reduction using local polynomial fits," *Graphical Models and Image Processing*, vol. 57, pp. 183–196, 1995.

[4] I. Perfilieva, "Fuzzy transforms: Theory and applications," *Fuzzy Sets and Systems*, vol. 157, pp. 993–1023, 2006.

[5] I. Perfilieva, "Fuzzy transforms and their applications to image compression," in *LNAI 3849. Fuzzy Logic and Applications*, I. Bloch, A. Petrosino, and A. Tettamanzi, Eds. Heidelberg: Springer, 2006, pp. 19–31.

[6] I. Perfilieva, M. Daňková, and B. Bede, "Towards F-transform of a higher degree," *Fuzzy Sets and Systems*, vol. 180, pp. 3–19, 2011.

Discrete Representation of Stationary Random Processes Using Fuzzy Transform

Michal Holčapek **Vilém Novák** **Irina Perfilieva**
CE IT4Innovations
Division of the University of Ostrava
Michal.Holcapek@osu.cz Vilem.Novak@osu.cz Irina.Perfilieva@osu.cz

Abstract

In this paper, we show that fuzzy transform originally introduced for a transformation of complex spaces of functions to simpler ones can be used in the analysis of real stationary random processes. We will show that under certain assumptions the fuzzy transform may be used for an approximation of this type of stationary processes. The obtained results could help researches to understand better the analysis of time series based on fuzzy transform.

1 Introduction

In time series analysis, random noise component is assumed in many cases to be a stationary process for its valuable properties. Fuzzy transform (F-transform for short) is a technique based on a partitioning of a real interval using fuzzy sets that generally transforms complex spaces of functions to simpler ones. By setting of fuzzy partition parameters the F-transform can be used for approximation and smoothing of original functions. The latter has been used among others in time series analysis for a trend extraction and a reduction of seasonal components (see [1], [2]).

In this paper, we focus on the random noise component of time series described by a weakly stationary process (with zero mean value). We will show that under specific assumptions the F-transform can be used for an approximation of this type of stationary process.

The result is motivated by the idea to represent complex stationary processes using processes with a discrete spectrum. It can be shown (see, e.g., [3]) that each stationary process $\xi(t)$ defined on a wide interval $[-T, T]$ (it means for a large T) can be approximated arbitrarily closely by a linear combination of harmonic oscillations of the form $\sum_{j=1}^{n} \xi_j e^{i\lambda_j t}$, where $\xi_1, \ldots, \xi_n$ are pairwise uncorrelated random variables with mean zero independent on time t, i is the

imaginary unit and $\lambda_1, \dots, \lambda_n$ are real constant. The representation of real stationary processes by F-transform, however, keeps a different idea than the previous one. The linear combination of harmonic oscillations is here replaced by the combination of basic functions which uniformly partition the real line and the closeness of frequencies λ_j by closeness of nodes over which fuzzy partitions are built.

The presented results have to be considered as preliminary ones justifying the investigation of stationary processes using F-transform. Another argument supporting the investigation of F-transform in the area of stationary processes comes from the computation complexity $O(n)$ of F-transform in contrast to the computational complexity $O(n \log n)$ that holds for the fast Fourier transform.

The paper is structured as follows. A necessary background for the analysis of stationary processes is provided in next section. The third section is devoted to the basic F-transform concepts translated into the language of stochastic processes. The main results are presented in the fourth section. The last section is a conclusion.

2 Stationary processes

In this section, we provide a necessary background for our analysis of weakly stationary process $\xi(t)$ by fuzzy transform.

2.1 Assumptions

In what follows, we assume that a probability space $(\Omega, \mathcal{F}, P)$ is fixed and we consider a real random process $\xi(t)$ (defined for any real number t) such that for any finite sequence $t_1, \dots, t_n$ $(n = 1, 2, \dots)$ of times there is a joint distribution function given by

$$F_{t_1,\dots,t_n}(x_1, \dots, x_n) = P(\{\xi(t_1) \le x_1, \dots, \xi(t_n) \le x_n\}). \tag{1}$$

The distribution functions (1) must satisfy the following two conditions:

(D1) *The symmetry condition*, according to which

$$F_{t_{i_1},\dots,t_{i_n}}(x_{i_1}, \dots, x_{i_n}) = F_{t_1,\dots,t_n}(x_1, \dots, x_n),$$

where $i_1, \dots, i_n$ is a permutation of the indices $1, \dots, n$;

(D2) *The compatibility condition*, according to which

$$F_{t_1,\dots,t_m,t_{m+1},\dots,t_n}(x_1, \dots, x_m, \infty, \dots, \infty) = F_{t_1,\dots,t_m}(x_1, \dots, x_m)$$

for any $t_{m+1}, \dots, t_n$ if $m < n$.

We use **E**, **Var**, **Cov** to denote the expected value, variance and covariance of random variables, respectively. Further, let us define a real function $B(t,s)$ called *covariance function* of $\xi(t)$ (see [4]) by

$$B(t,s) = \mathbf{E}[\xi(t)\xi(s)]. \tag{2}$$

We assume that $\xi(t)$ satisfies the following conditions for any t:

(i) $\mathbf{E}[\xi(t)] = 0$;

(ii) $B(t, t+\tau)$ is independent of t for each τ;

(iii) $B(\tau) = B(0,\tau)$ is Lebesgue integrable.

The first two conditions says that $\xi(t)$ is a (weakly) stationary process, the latter is a necessary condition for our analysis. Obviously, the covariance of random variables $\xi(t)$ and $xi(s)$ is equal to $B(t-s)$, i.e.,

$$\mathbf{Cov}(\xi(t), \xi(s)) = B(t-s). \tag{3}$$

Specifically, we have $\mathbf{Var}(\xi(t)) = B(0) = \sigma^2$. Then $|B(\tau)| \leq B(0) = \sigma^2$ for any t, s.[1]

Note that the previous inequality says that the random variables $\xi(t)$ and $\xi(s)$ are dependent to each other in a degree which absolute value is at most equal to the variance of $\xi(t)$. Sometimes, it seems to be natural to assume that higher difference between t and s causes lower dependence. The assumption on integrability of $B(\tau)$ is a necessary condition for our analysis of approximation and variability reduction of $\xi(t)$ using the fuzzy transform.

2.2 Limit of sequences of random variables

Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables. We say that a random variable ξ is a *limit in the mean square* of the sequence of random variables $\xi_1, \xi_2, \ldots$ and denote it by

$$\operatorname*{l.i.m}_{n\to\infty} \xi_n = \xi \tag{4}$$

if

$$\lim_{n\to\infty} \mathbf{E}[(\xi_n - \xi)^2] = 0,$$

i.e., for any $\varepsilon > 0$ there exists a natural number n_0 such that $\mathbf{E}[(\xi_n - \xi)^2] < \varepsilon$ for any $n > n_0$.[2] Let us show two important properties of limit in the mean square which will be used later (see [5]).

Theorem 1 *Let $(\xi_n)_{n=1}^{\infty}$, $(\psi_n)_{n=1}^{\infty}$ be two sequences of random variables and let us suppose that $\operatorname{l.i.m}_{n\to\infty} \xi_n = \xi$ and $\operatorname{l.i.m}_{n\to\infty} \psi_n = \psi$. Then,*

(i) $\mathbf{E}[\xi] = \lim_{n\to\infty} \mathbf{E}[\xi_n]$,

(ii) $\mathbf{E}[\xi\psi] = \lim_{n\to\infty} \mathbf{E}[\xi_n\psi_n]$.

[1] It follows from the Cauchy-Schwartz's inequality.

[2] Note that we use the symbol l.i.m to distinguish the limit in the mean square of a sequence of random variables and the common limit of a sequence of numbers (cf., [5]).

2.3 Integral of stationary process

Let $f(t)$ be an arbitrary real function and $\xi(t)$ a stationary random process. The integral

$$\int_c^d \xi(t)f(t)dt \tag{5}$$

is defined as the limit (in the mean square) of random variables

$$\sum_{j=2}^{n} \xi(t'_j)f(t'_j)(t_j - t_{j-1}),$$

where $c = t_1 < t_2 < \cdots < t_n = d$ and $t_{j-1} \leq t'_j \leq t_j$ holds for any $j = 2, \ldots, n$. Of course, this integral does not exist for all pairs of real functions and stationary processes. For details, we refer to [3]. In what follows, we will assume only such real functions and stationary processes that are integrable with respect to the integral (5).

In order to show the approximation of stationary processes using the F-transform, we will need the following special case of Hölder's inequality for integrals which holds in the mean.

Theorem 2 *Let $\xi(t)$ be a stationary process and $f(t)$ be a real function defined on $[c, d]$. Then,*

$$\mathbf{E}\left[\left(\int_c^d |\xi(t)f(t)|dt\right)^2\right] \leq \mathbf{E}\left[\int_c^d |\xi(t)|^2 dt \int_c^d |f(t)|^2 dt\right] \tag{6}$$

3 F-transform of stationary process

In this section, we will briefly review the main principles of the fuzzy transform. Detailed explanation of the general theory can be found in [6, 7, 8].

Let U be an arbitrary (nonempty) set called a *universe*. By a fuzzy set in the universe U we will understand a function $A : U \to [0, 1]$. The F-transform is a special technique that can be applied to real continuous functions f, defined on an interval $[a, b] \subset \mathbb{R}$. The essential idea is to transfer f into another, simpler space, and then to transfer the respective image back. The latter space consists of finite vectors that are obtained on the basis of the well formed *fuzzy partitions* of the domain of the given function. Thus, the first step called *direct F-transform* results in the vector of averaged functional values. The second step called *inverse F-transform* converts this vector into another continuous function $\hat{f}$, which approximately reconstructs the original f.

3.1 Uniform fuzzy partition

Let $\mathbb{Z}$ denote the set of integers. It is well-known that a uniform fuzzy partition is defined using a generating function K which is modified by a parameter

h expressing the required spread. Each basic function of the uniform fuzzy partition is then constructed using a suitable shift of the modified generating function K, where the uniformity for all shifts is supposed. The generating function is defined as follows.

Definition 1 *A function $K : \mathbb{R} \to [0,1]$ is said to be a* generating function *if K is an even Lebesgue integrable function (fuzzy set) which is non-increasing in $[0,\infty)$ and*

$$K(x) \begin{cases} > 0, & \text{if } x \in (-1,1); \\ = 0, & \text{otherwise.} \end{cases} \tag{7}$$

A generating function K is said to be normal *if $K(0) = 1$.*

It should be noted that the previous definition is more general than the analogous definition of a generating function in [9], because the continuity of K is replaced by its integrability and the normality of K is considered as an additional condition.[3] Uniform fuzzy partitions of the real line are defined as follows (cf., [11]).

Definition 2 *Let K be a normal generating function, h be a positive real number and $c_0 \in \mathbb{R}$. A system of fuzzy sets defined by*

$$A_k(x) = K\left(\frac{x - c_0}{h} - k\right) \tag{8}$$

for any $k \in \mathbb{Z}$ is said to be a uniform fuzzy partition (UFP) of the real line determined by the triplet (K, h, c_0) *if the Ruspini's condition is satisfied, i.e.,*

$$S(x) = \sum_{k \in \mathbb{Z}} A_k(x) = 1 \tag{9}$$

holds for any $x \in \mathbb{R}$.

In the sequel, the parameters h and c_0 are called a *spread* and a *central node*, respectively. The fuzzy sets A_k defined by (8) that form a uniform fuzzy partition of the real line are called *basic functions*. A simple consequence of (8) is the formula $A_k(x) = A_0(x - hk)$ that holds for any $x \in \mathbb{R}$ and $k \in \mathbb{Z}$. Putting $c_k = c_0 + kh$ one can simply check that $A_k(c_k) = 1$ and A_k is centered around the node c_k.

3.2 Direct and inverse F-transform

We use $\mathbf{A}_h = (A_k)_{k \in \mathbb{Z}}$ to denote a uniform fuzzy partition of real line determined by (K, h, x_0) and denote $(c_k)_{k \in \mathbb{Z}}$ their corresponding nodes, i.e., $A_k(x) = A_0(x - c_k) = K(\frac{x - c_k}{h})$.

[3] In [10], a generating function was called a *basal function*.

Definition 3 *Let $\xi(t)$ be a stationary process, $\mathbf{A}_h$ be a uniform fuzzy partition and $(c_k)_{k\in\mathbb{Z}}$ denote the respective nodes. An infinite vector of random variables $(\xi_k)_{k\in\mathbb{Z}}$ is called a* direct fuzzy transform (F-transform) *of $\xi(t)$ with respect to $\mathbf{A}_h$ if*

$$\xi_k = \frac{1}{h}\int_{c_{k-1}}^{c_{k+1}} \xi(t)A_k(t)dt, \quad k \in \mathbb{Z}.$$

The random variable ξ_k is called a component of F-transform.

It is easy to show that the linearity of F-transform is preserved for stationary processes, i.e., if $\xi(t) = a\eta(t) + b\zeta(t)$, $a, b \in \mathbb{R}$, then $\xi_k = a\eta_k + b\zeta_k$. Note that the linearity belongs among the most valuable properties of the F-transform often used in proofs. In this paper, we use $\xi_{k,h}$ to denote the k-th F-transform component at the node c_k with respect to $\mathbf{A}_h$ and suppose only such stationary processes $\xi(t)$ for which $\xi_{k,h}$ can be found for any $k \in \mathbb{Z}$ and $h > 0$.

The inverse F-transform is defined as the linear combination of components and basic functions. We use a slight modification of the original definition in [9] as follows.

Definition 4 *Let $\xi(t)$ be a stationary process and $(\xi_k)_{k\in\mathbb{Z}}$ be the direct F-transform of $\xi(t)$ with respect to $\mathbf{A}_h$. Then,*

$$\hat{\xi}(t) = \sum_{k\in\mathbb{Z}} \xi_k A_k(t) \tag{10}$$

is called an inverse F-transform *of $\xi(t)$ with respect to $\mathbf{A}_h$.*

4 Approximation of $\xi(t)$ by F-transform

In order to investigate the approximation of stationary processes $\xi(t)$ by the F-transform, let $B^*(\tau) = B(0) - B(\tau)$ and suppose that

$$\lim_{h\to 0} \frac{1}{h}\int_0^h B^*(\tau)d\tau = 0. \tag{11}$$

It is easy to see that the function $B^*(\tau)$ is a non-negative real function and the assumption (11) on $\xi(t)$ says that random variables $\xi(t)$ and $\xi(s)$ are very strongly dependent for small differences between t and s, in other words, $B(\tau)$ converges to $B(0)$ for $\tau \to 0$.[4]

In what follows, we provide several theorems demonstrating how the F-transform can approximate stationary process satisfying (11). In the original paper ([6]) on the F-transform, the author shows that a twice continuously differentiable function differs from the F-transform components at nodes c_k up

[4]The latter follows from the integral mean value theorem saying that, for each $h > 0$, there is $\tau' \in (0, h)$ such that $1/h \int_0^h B^*(\tau)d\tau = B^*(\tau')$, i.e., $B^*(\tau)$ converges to 0.

to h^2. As a consequence we obtain that the F-transform components converge to the values of original function at nodes c_k for $h \to 0$. The following theorem shows an analogous property for stationary stochastic process under the assumption (11).

Theorem 3 *Let $\xi(t)$ satisfy (11) and c_k be a fixed node. Then, there exists a sequence of F-transform components $\xi_{k,h_1}, \xi_{k,h_2}, \ldots$ at the node c_k w.r.t. $\mathbf{A}_{h_1}, \mathbf{A}_{h_2}, \ldots$, respectively, such that $\mathrm{l.i.m}_{n\to\infty} \xi_{k,h_n} = \xi(c_k)$.*

The following two theorems show that each stationary process $\xi(t)$ can be approximate arbitrarily closely by the F-transformed components belonging to the nearest neighborhood of t (i.e., $[t-h, t+h]$).

Theorem 4 *Let $\xi(t)$ satisfy (11) and $(\mathbf{A}_{h_n})_{n=1}^{\infty}$ be a sequence of UFPs such that $\lim_{n\to\infty} h_n = 0$. Then, for any $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for any $n > n_0$ it holds $\mathbf{E}[(\xi(t) - \xi_{k,h_n})^2] < \varepsilon$ for any F-transform component ξ_{k,h_n} w.r.t. $\mathbf{A}_{h_n}$ and $t \in \mathbb{R}$ such that $|c_k - t| \leq h_n$.*

Theorem 5 *Let $\xi(t)$ satisfy (11) and $(\mathbf{A}_{h_n})_{n=1}^{\infty}$ be a sequence of UFPs such that $\lim_{n\to\infty} h_n = 0$. Then, for any $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for any $n > n_0$ it hods $\mathbf{E}[(\xi(t) - \xi_{k,h_n})(\xi(t) - \xi_{k+1,h_n})] < \varepsilon$ for any F-transform components ε_{k,h_n} and ε_{k+1,h_n} w.r.t. $\mathbf{A}_{h_n}$ and $t \in \mathbb{R}$ such that $|c_k - t_0| \leq h_n$ and $|c_{k+1} - t_0| \leq h_n$.*

The last theorem is a version of Theorem 2 in [6] for stationary processes.

Theorem 6 *Let $\xi(t)$ satisfy (11) and $(\mathbf{A}_{h_n})_{n=1}^{\infty}$ be a sequence of UFPs such that $\lim_{n\to\infty} h_n = 0$. Then, the corresponding sequence of inverse F-transforms $\hat{\xi}_{h_1}(t), \hat{\xi}_{h_2}(t), \ldots$ converges in the mean square to $\xi(t)$, i.e., $\mathrm{l.i.m}_{n\to\infty} \hat{\xi}_{h_n}(t) = \xi(t)$ for any $t \in \mathbb{R}$.*

5 Conclusion

In this paper, we analyzed the (weakly) stationary processes. We showed that under certain assumptions the fuzzy transform can be used for an approximation of this type of stationary process. Although, our results are only preliminary ones and they are far from the known results on the discrete representation of stochastic processes (see, e.g., [12]), we believe that they support and give us rational arguments to continue in the investigation of stationary processes in the context of the F-transform.

Acknowledgment

The authors would like to thank to Department of the Navy Grant N62909-12-1-7039 issued by Office of Naval Research Global. The United States Government has a royalty-free license throughout the world in all copyrightable

material contained herein. Additional support was given also by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

References

[1] V. Novák, I. Perfilieva, and V. Kreinovich. F-transform in the analysis of periodic signals. In M. Inuiguchi, Y. Kusunoki, and M. Seki, editors, *Proc. 15^{th} Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, Osaka, Japan, 2012. Osaka University.

[2] V. Novák, I. Perfilieva, M. Holčapek, and V. Kreinovich. Filtering out high frequencies using F-transform. *Submitted to Information Sciences.*

[3] A.M. Yaglom. *An introduction to the theory of stationary random functions. Revised English ed. Translated and edited by Richard A. Silverman.* Englewood Cliffs, NJ: Prentice-Hall, Inc. XIII, 235 p. , 1962.

[4] P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting.* Springer-Verlag New York, 2002.

[5] A.H. Jazwinski. *Stochastic processes and filtering theory.* Mineola, NY: Dover Publications. xiv, 376 p. , 2007.

[6] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy sets syst.*, 157(8):993–1023, 2006.

[7] I. Perfilieva and M. Daňková. Towards F-transform of a higher degree. In *Proceedings of IFSA/EUSFLAT 2009*, pages 585–588, Lisbon, Portugal, 2009.

[8] I. Perfilieva. Fuzzy transforms: A challenge to conventional transforms. In P.W. Hawkes, editor, *Advances in Images and Electron Physics, 147*, pages 137–196. Elsevier Academic Press, San Diego, 2007.

[9] I. Perfilieva, M. Daňková, and B. Bede. Towards a higher degree F-transform. *Fuzzy Sets Syst.*, 180(1):3–19, 2011.

[10] M. Holčapek and T. Tichý. A smoothing filter based on fuzzy transform. *Fuzzy Sets Syst.*, 180(1):69–97, 2011.

[11] K. Loquin and O. Strauss. Histogram density estimators based upon a fuzzy partition. *Stat. Probab. Lett.*, 78(13):1863–1868, 2008.

[12] K. L. Jordan. Discrete representation of random signals. Technical Report 378, Massachusetts Institute of Technology Research Laboratory of Electronics, 1961.

UAV Environment recognition for better horizon detection with usage of FUZZY IS

David Novák
Department of Informatics
Silesian university in Opava
frvolny@centrum.cz

Petr Čermák
Department of Informatics
Silesian university in Opava
petr.cermak@fpf.slu.cz

Abstract

In this paper the design of the environment detection method is presented. By environment we mean surroundings in which autonomous robot operates, especially the UAV. This method is suitable for the application in UAV systems for the horizon detection, but also in the other artificial intelligences applications, which require recognition of the environments character in which autonomous entity operates. Part of the document includes also an overview of methods used for the horizon detection and problems which may arise during the detection. Environment detection algorithm is based on the principles of fuzzy sets and Sugeno - type fuzzy inference systems. The detected straight line segments are the basic input algorithm, segment detection method is subject of the further development. Environment detection method is computationally undiscerning and therefore suitable for implementation on programmable microcontrollers, which are often used to control the UAV devices. Environment detection and horizon recognition will be used to deal with position sensors failures and replace their functionality in image recognition with previously proposed algorithms in UAV control method using the principles of multi-agent systems.

1 UAV multiagent – based control system

In Figure 1, we can see the proposal of UAVcontrol system, based on multiagent principles [3]. The main reason for choosing this approach is the fact that during the decomposition of the function of the UAV system, specific modules are proposed and each module is responsible for the implementation of certain features of the UAV system. These modules have to work, if it is possible, as separate, autonomous entities. It is obvious, that individual modules must cooperate, so that the whole system is working. If we are aware of the principles of multi-agent systems, we find that the above modules meet the criteria of

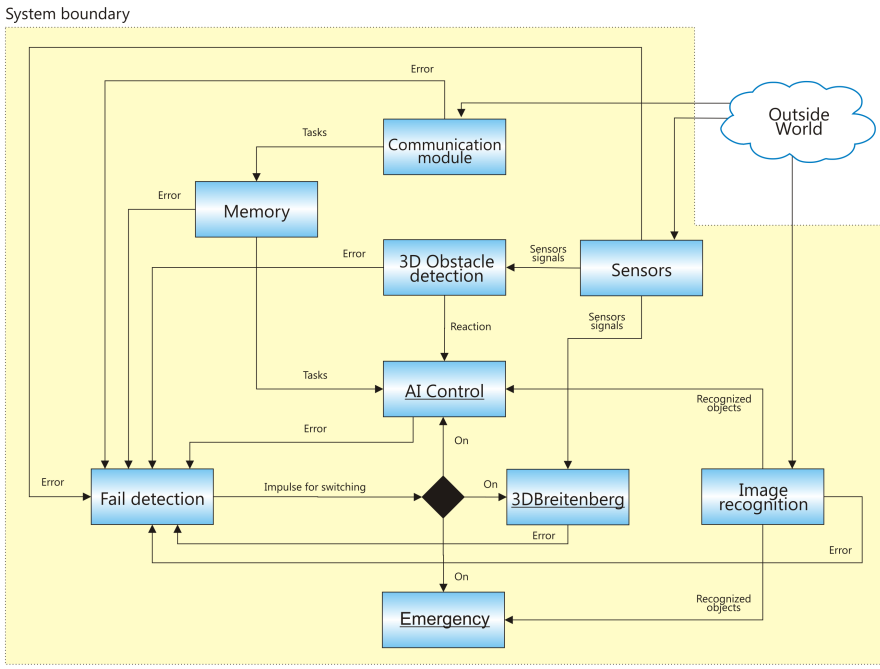

Figure 1: Multiagent - based UAV control system [3].

agents in the multi-agent community.

Some modules of the proposed system have high demands on the reaction rate (in particular parts of the system, ensuring obstacle avoidance), while other modules (such as planning and decision making) require the ability to handle complex tasks at the expense of reaction response. If we refer again to the theory of agent and multi-agent systems, we find that the criterion of a high response time meet the so-called reactive agents. On the contrary, thinking skills meet the criterion of deliberative agents. Other modules can be created using a combination of both approaches. The cooperation of the modules must benefit from mutual interaction, which is another problem that can be solved using the principles of multi-agent communication, cooperation and coordination.

The basic idea of the proposed UAV system is to switch three basic modes of flight. We define a higher control mode as flight control using AI algorithms and low control mode based on simpler, reactive algorithms. Switching is carried out by the module named "error detection". Flight modes are:

- Autonomous (AI) control flight mode
- 3D Breitenberg flight mode
- Emergency flight mode

"Emergency" module and flight mode is activated on a signal from the module "fail detection" and where there is a shortage of modules through which you can no longer continue in flight. Module " Emergency flight" is mostly a reactive character of the agent. The main priorities and objectives of vehicle behavior is to reduce the speed and direction of the device falling out of the area, where it could result in damage to property or threat to humans.

To fulfill this task with the help of image recognition, we need a fixed point in a scene. As a suitable candidate for such point can be a horizon.

2 Horizon detection

The horizon recognition is most widely solved problem related to the development of the autonomous control of UAVs. Correctly recognized horizon and its angle enable you to obtain additional information that may be important for the controls, such as an estimate of altitude, speed, and so on. The horizon is also often the only permanent landmark in the scene, so it can be used for UAV stabilization.

There are a number of horizons detection solutions with different percentage success of the detection. For small UAVs, which reach maximum heights of hundreds of meters, the horizon detection is not a trivial task though. At this point it is necessary to define the difference between the horizon and the visible horizon. We consider as the horizon the dividing line between sky and earth, that is, a horizontal line that is level to the ground. We understand the visible horizon as the real dividing line between sky and earth, which is visible from a given location in the scene. Visible horizon may therefore not always be horizontal with the ground plane. Another problem arises naturally in urban areas, where horizon must be detected for example from vanishing points of detected building outlines.

The ideal objective is to find a universal approach how to detect the actual position of the horizon and any environment. Such procedure would not fail in any area, which the UAV can get in.

2.1 Usual problems with the horizon detection algorithms

Currently, there are some general aproaches for horizon detection. Each of them is able to detect the horizon under certain conditions from the detection of horizon using Hough transformation to more complex methods. There is an overview of some of the techniques and problems depending on the conditions In Table 1. Intentionally I dont mention the detection horizon in urban areas here, because in such conditions all listed procedures fail virtually. (The Table 1.)

To detect horizon in urban areas there are several methods that are based on the monitoring of main lines in the image and the detection of the vanishing points, from which the likely position of horizon and its angle can be calculated.

The method	The best results	The possible weaknesses
Hough transformation + edge detectors.	Great heights, straight horizon.	Fails at low altitudes and indented areas.
Searching of the centers of gravity of the ground and sky.	Equally contoured horizon	In the indented surroundings and a higher cloudiness.
Optical flow.	Great heights, a little contoured horizon.	May fail at low altitudes and indented surroundings.

Table 1: Problems of standard horizon detectors depending on conditions

```
1: for ever do
2:   if (detectedenviromentconstant) > urban then
3:     return horizon detection algorithm for urban enviroment
4:   else
5:     return horizon detection algorithm for open enviroment
6:   end if
7: end for
```

Figure 2: Environment detection pseudocode

Algorithms that work with edges however fail in open landscape, because there is no correct detection of horizontal lines and the subsequent derivation of the vanishing points because they are virtually absent in the scene.

3 The detection of the environment

For the problems described in the previous chapters a simple solution can be used, based on the idea of detecting the environment, in which the UAV device is moving, and the subsequent correct choice of the algorithm for detecting the horizon. At this point, it is irrelevant what kind of algorithms for horizon recognition we will use; the important thing is to choose at least one for landscape without build-up area and one for urban areas. The pseudo-code on Figure 2. represents the basic algorithm for environment detection:

Detected_environment_constant is a number which expresses the probability that it is the urban environment.

However, the choice of correct algorithm which correctly calculates the degree of probability based on the detected parameters, which will lead to the correct environment identification, remains the problem. Another problem is, of course, selecting the right input parameters.

$$seg = \frac{image_{width} + image_{height}}{2 * c_{long}}$$

Figure 3: Minimum leght of straight lines.

3.1 The choice of detection parameters

The information listed in the text above implies that one of the key parameter for the decision-making algorithm will be straight lines detection and their characteristics. Quantity of detected straight lines with specific lengths is quite different in open scenery and urban development. However, this may not always be decisive. The straight lines, which can provide false information, such as location of the vanishing point (power lines, fences, roads), can be detected even in open landscape. Therefore, it is necessary to provide additional input parameters. We must include only the lines that have a certain length to the calculations, though. It is necessary to exclude short lines, which may occur in both types of environments. We can select minimum length of straight lines using for example the size of the processed (see Fig. 3), Clong is a constant, which defines the minimum size of the detected lines. It is appropriate to use the direction of detected lines to calculate the next parameter. It is specifically the number of lines that have the same direction as this attribute corresponds to a high degree of urban environment. It is appropriate to use the number of detected rectangles is the last parameter. It is essential to use an algorithm that is able to detect rectangles with perspective. Representation of rectangles in the images from urban enviroment is usually high. Because we detect the straight lines and their directions, there is no problem to find the lines that form the edges of detected rectangles. For detection of environment it is necessary to count with the fact that quadrangles are also detected outside urban areas, but their number will be low. Therefore we have three input parameters:

- Number of line segments
- Number of parallel line clusters
- Number of detected rectangles

Output parameters for decision making are two:

- Probability of urban environment
- Probability of non urban environment

3.2 The proposed method

On Figure 4. the proposed method scheme can be seen. The choice of the decision-making algorithm is the last step needed for the realization of environment detection. Since this is a relatively small number of input parameters,

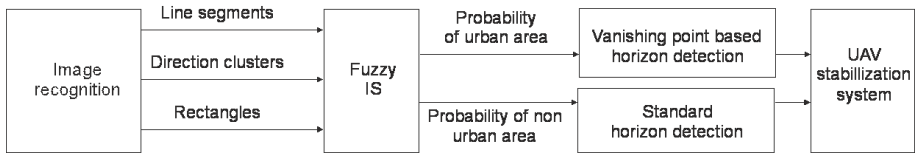

Figure 4: Scheme of proposed system.

it seems appropriate to use the principles of Fuzzy inference systems. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. Because this is a UAV and there is a presumption of embeded hardware that has limited computing power, we suggest using Sugeno model [4].

A typical rule in a Sugeno fuzzy model has the form [2]:

IF Input 1 = x AND Input 2 = y, THEN Output is z = ax + by + ci

For a zero-order Sugeno model, the output level z is a constant (a=b =0). The output level zi of each rule is weighted by the firing strength wi of the rule. For example, for an AND rule with Input 1 = x and Input 2 = y,
the firing strength is:
wi = AndMethod (F1(x), F2(y))
where F1,2 (.) are the membership functions for Inputs 1 and 2. The final output of the system can be the weighted average of all rule outputs More about the Sugeno Fuzzy inference systems can be found in [3] or [4]. Let us denote the input and output parameters as:

- SEG number of line segments
- CL number of parallel lines clusters
- RC number of detected rectangles
- URB probability of urban environment
- NURB probability of non urban environment

URB and NURB variables are used for final decision proces. Probability in this case does not mean mathematical probability. In some cases, both can be at the same level. Inference system works in the basic version with 4 fundamental rules:

- IF SEG = HIGH AND CL = HIGH AND RC = HIGH
  THEN URB = SEG + CL + RC
  This rule corresponds to a situation where there are entire buildings, including the windows, etc. represented in the scene.
- IF SEG = HIGH AND CL = HIGH AND RC = LOW
  THEN URB = RC This rule corresponds to a situation with lots of lines,

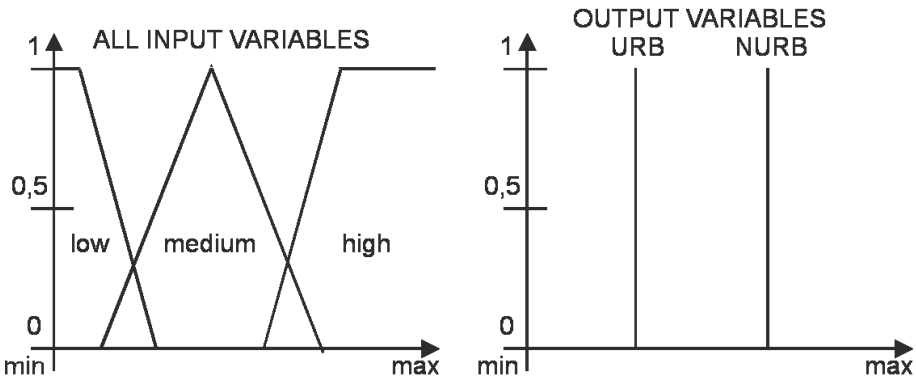

Figure 5: Input and output FUZZY variables.

Rectangles	Lines	Segments	Urban environment
HIGH	HIGH	HIGH	HIGH
LOW	HIGH	HIGH	LOW
LOW	HIGH	LOW	LOW
LOW	LOW	LOW	LOW

Table 2: Environment detection rules behavior

but lack of geometric entities so there is a low probability of urban environment.

- IF SEG = LOW AND CL = HIGH AND RC = HIGH
  THEN URB = SEG This rule corresponds to the situation of detection of ornaments or decoration. Can be in a case when building is too close to UAV. In this scenario, usage of urban scene horizon detection is not recommended.

- IF SEG = LOW AND CL = LOW AND RC = LOW
  THEN URB = RC This rule corresponds to a situation or environmet without detected rectangles, or line clusters, which means very low probability of urban surroundings.

Other rules are not fundamentally necessary, only the detection of the urban environment is required for the needs of the decision-making process.

4 Conclusion and future work

The ability to detect the character of the environment in which autonomous entity operates should be an important feature for future autonomous systems, especially for UAVs. The use of fuzzy inference system for such detection is

probably an appropriate approach as it corresponds to the character of the input parameters that have a high degree of uncertainty. The method proposed in this paper will be implemented within the UAV system, which is based on the principles of multi-agent systems. The ARM architecture was chosen as the target hardware. Programming language for FUZZY inference system will be an offshoot of the LISP programming language designed for ARM hardware (armpit SCHEME). Further work on the project includes testing the accuracy of detection using existing algorithms to detect the horizon and design custom algorithms for both urban zone and the open landscape.

4.1 Acknowledgements

This paper has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. No. CZ.1.05/1.1.00/02.0070 supported by Operational Program 'Research and Development for Innovations' funded by Structural Funds of the European Union and state budget of the Czech Republic. We would like to thank to for SGS/6/2011 "Artificial intelligence in Robotics and Medical Informatics, application of artificial intelligence algorithms in design of UAV systems".

References

[1] SHAW D, BARNES N, Perspective rectangle detection, *Research School of Information Science and Engineering, Australian National University, URL: http://users.cecs.anu.edu.au/ nmb/papers/06-PRD.pdf, [cit. 2013-20-08].*

[2] D. NOVAK, P. CERMAK, Advantages of using multi-agent principles in FAIL - SAFE UAV system design, *MMAR2011, 2011, ISBN 978-1-4577-0912-8, p. 162-167..*

[3] Kevin M. PASSINO, Stephen YURKOVICH, Fuzzy Control, *Addison Wesley Longman, Menlo Park, CA, 1998, ISBN-10: 020118074X.*

[4] T. TAKAGI and M. SUGENO, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans, on Systems, Man and Cybernetics, 15, 1985, pp. 116-132. .*

On Stable Solutions of Atomic Low-Carbon Growth Partnership Formation

Mayuko Miyata
Azbil Corporation

Shao-Chin Sung
Aoyama Gakuin University
son@ise.aoyama.ac.jp

Abstract

East Asia Low Carbon Growth Partnership is a concept advocated by Japanese Prime Minister Noda as a measure against global warming, which is an idea is to build supplier-client relationships for sharing both advanced low-carbon technologies and the benefit of greenhouse gas emission reduction between countries. We first introduce our game theoretical framework, and show existence results of its stable solutions.

1 Introduction

Low-carbon growth partnership, *LCG-partnership* in short, is a concept advocated in 2011 by Japanese Prime Minister Noda as a measure against global warming which is one of the most serious environmental problem. The idea is to build a supplier-client relationship of sharing advanced low-carbon technologies between countries, and the incentive for both suppliers and clients are provided by sharing the benefit of reduction of greenhouse gas emission by introducing the technologies.

The global objective is to maximize the total reduction of greenhouse gas emission among all possible LCG-partnerships, and the problem for finding such an *optimal* LCG-partnership is shown to be $\mathcal{NP}$-hard in the strong sense [4]. Obviously, countries make their decision based on their incentives, i.e., the benefit of reduction of greenhouse gas emission by introducing advanced low-carbon technologies. Hence, it is natural to handle LCG-partnership with game theoretical frameworks. As the first attempt [4], a game model based on network formation game [5] is introduced, and for such a model, some rules are imposed for avoiding unnecessary conflict between countries. Such a model is a kind of cooperative games, and it can be considered as a restricted version of hedonic games [1] as well. Later, in order to increase the reduction of greenhouse gas emission, the model is extended in [2] by introducing a requirement, based on the concept of *market price*, for forming a new link and deleting a existing link in an LCG-partnership and some numerical results are provided as well.

In this paper, we provide a formulation of LCG-partnership formation based on investment functions as its outcomes, each of which describes the amount of investment between each pair of countries, and we consider LCG-partnership formation in non-atomic case and atomic case. In addition, we introduce the concept of separateness, which requires each country to be either client, supplier, or not involved. Moreover, we define utilities of countries based on investment functions, and introduce stability concepts based on positional difference of suppliers and clients. We show the complexity results for finding an investment function which maximizes the total reduction of greenhouse gas emission, and show the existence results of stable investment functions.

2 Preliminaries

Let us introduce instances of LCG-partnership formation. Let $N = \{1, 2, \ldots, n\}$ be a finite set of its countries. For each country $i \in N$,

- its greenhouse gas emission GGE_i,
- its GDP[1] at purchasing power parity (PPP) [2] exchange rates PPP_i,
- its industrial scale SCL_i,
- its capacity CAP_i,
- its payback rate $\alpha_i \in (0, 1)$ for its support

are provided, where capacity CAP_i is the amount of resource (in the unit of industrial scale) prepared for supporting other countries. In our numerical analysis, the values of GGE_is and PPP_is are provided by [6, 7].

Moreover, we denote the economic efficiency of greenhouse gas emission (i.e., GDP PPP per a unit of greenhouse gas emission) of country i by EFF_i, i.e.,

$$\mathrm{EFF}_i = \frac{\mathrm{PPP}_i}{\mathrm{GGE}_i} \quad \text{for each } i \in N.$$

Outcomes of LCG-partnership formation are defined as follows. Let φ be a non-negative function over N^2, where $\varphi(i, j)$ denotes the amount of investment by country i for supporting country j. In other words, φ determines the amounts of investments all pairs of countries.

Definition 1 *A function $\varphi : N^2 \to \mathbf{R}_{\geq 0}$ is an investment function if the following constraints are satisfied.*

- *for each $i, j \in N$, $\varphi(i, j) = 0$ if $\mathrm{EFF}_i \leq \mathrm{EFF}_j$,*

[1] *Gross Domestic Products (GDP)* is a value which strongly reflects the development power of a country, and its the market value of all goods and services produced in a country within a given period.

[2] *Purchasing Power Parity (PPP)* is a technique to provide relative values of different currencies, and the relative values are calculated based on the amount of a certain item each currency can purchase.

- $\sum_{j \in N} \varphi(i,j) \leq \mathrm{CAP}_i$ *for each* $i \in N$, *and*
- $\sum_{i \in N} \varphi(i,j) \leq \mathrm{SCL}_j$ *for each* $j \in N$.

■

In general, each country is allowed to have more than one supplier. When a country is supported by more than one country, more negotiations between countries are required to determine the investment ratio, which may slow down the progress of LCG-partnership formation. In order to reduce the number of negotiations between countries, we introduce the concept of atomicity.

Definition 2 *An investment function* φ *is called atomic if* $\varphi(i,j) \in \{0, \mathrm{SCL}_j\}$ *for each* $i, j \in N$. *We say that an LCG-partnership formation is atomic if all of its non-atomic outcomes (i.e., non-atomic investment functions) are considered to be infeasible.* ■

When an atomic LCG-partnership formation is under consideration, the description of its outcomes can be simplified, namely, each investment function can be characterized by a binary relation. In order to distinguish atomic LCG-partnership formation with other LCG-partnership formation, we say that an LCG-partnership formation is *non-atomic* if atomicity is not imposed. Observe that a non-atomic LCG-partnership formation may have atomic outcomes.

For each $i \in N$, we define $S_i(\varphi)$ and $C_i(\varphi)$ as follows.

$$\begin{aligned} S_i(\varphi) &= \{j \in N \mid \varphi(j,i) > 0\}, \\ C_i(\varphi) &= \{j \in N \mid \varphi(i,j) > 0\}. \end{aligned}$$

Here, $S_i(\varphi)$ is the set of all suppliers of country i under φ, and similarly $C_i(\varphi)$ is the set of all clients of country i under φ. By definition, we have $\sum_{j \in S_i(\varphi)} \varphi(j,i) = \sum_{j \in N} \varphi(j,i)$ and $\sum_{j \in C_i(\varphi)} \varphi(i,j) = \sum_{j \in N} \varphi(i,j)$.

In our models, players' utilities are defined as follows. For each pair $(i,j) \in N^2$, let $\mathrm{RDT}_{(i,j)}(\varphi)$ be the reduction of greenhouse gas emission obtained by implementing advanced low-carbon technology of country i into country j (with investment $\varphi(i,j)$) under φ.

$$\mathrm{RDT}_{(i,j)}(\varphi) = \frac{\varphi(i,j)}{\mathrm{SCL}_j} \left(\frac{\mathrm{PPP}_j}{\mathrm{EFF}_j} - \frac{\mathrm{PPP}_j}{\mathrm{EFF}_i} \right)$$

In other words, when country j is fully supported by country i with advanced low-carbon technologies and PPP of country j is kept unchanged,

- the investment $\varphi(i,j)$ of country i to country j is SCL_j,
- economic efficiency of country j is improved from EFF_j to EFF_i, and
- the greenhouse gas emission improved from $\mathrm{GGE}_i = \mathrm{PPP}_j/\mathrm{EFF}_j$ to $\mathrm{PPP}_j/\mathrm{EFF}_i$,

and hence the reduction of greenhouse gas emission is $\text{PPP}_j/\text{EFF}_j - \text{PPP}_j/\text{EFF}_i$. Moreover, when country j is partially supported by country i (i.e., $\varphi(i,j) < \text{SCL}_j$), the reduction of greenhouse gas emission becomes $\varphi(i,j)/\text{SCL}_j$ times of that with full support.

Utilities of countries are defined as follows.

$$u_i(\varphi) = \alpha_i \sum_{j \in C_i(\varphi)} \text{RDT}_{(i,j)}(\varphi) + \sum_{j \in S_i(\varphi)} (1-\alpha_j)\text{RDT}_{(j,i)}(\varphi)$$

Here,

- the first term $\alpha_i \sum_{j \in C_i(\varphi)} \text{RDT}_{(i,j)}(\varphi)$ is total payback reduction of greenhouse gas emission from country i's clients, and
- the second term $\sum_{j \in S_i(\varphi)} (1-\alpha_j)\text{RDT}_{(j,i)}(\varphi)$ is the remaining reduction of country i's own greenhouse gas emission after paying back to its suppliers.

Now let us introduce the concept of stability.

Definition 3 *Let φ and φ' be investment functions. We say that φ is dominated by φ' if any one of the following conditions holds.*

Client Switch: *There exists a pair $(i,j) \in N^2$ such that*

- $u_i(\varphi') > u_i(\varphi)$,
- $\varphi'(i,j) > \varphi(i,j)$,
- $\varphi'(i,\ell) \leq \varphi(i,\ell)$ *for each* $\ell \in N \setminus \{j\}$,
- $\varphi'(k,\ell) = \varphi(k,\ell)$ *for each* $(k,\ell) \in N^2$ *with* $k \neq i$,

Supplier Switch: *There exists a pair $(i,j) \in N^2$ such that*

- $u_j(\varphi') > u_j(\varphi)$,
- $\varphi'(i,j) > \varphi(i,j)$,
- $\varphi'(k,j) \leq \varphi(k,j)$ *for each* $k \in N \setminus \{i\}$,
- $\varphi'(k,\ell) = \varphi(k,\ell)$ *for each* $(k,\ell) \in N^2$ *with* $\ell \neq j$,

We say that an investment function φ is stable if it cannot be dominated by any investment function. ■

3 Optimal Outcomes and Stable Outcomes

Let φ be an investment function. By $\text{RDT}(\varphi)$ we denote the total reduction of greenhouse gas emission under φ, i.e.,

$$\text{RDT}(\varphi) = \sum_{(i,j) \in N} \text{RDT}_{(i,j)}(\varphi).$$

We say that an investment function φ is *optimal* if it maximizes the total reduction of greenhouse gas emission among all feasible investment functions, i.e., $\text{RDT}(\varphi) \geq \text{RDT}(\varphi')$ for each feasible investment function φ'. Observe that all atomic investment functions are feasible outcomes of non-atomic LCG partnership formation. Hence, stable outcomes of non-atomic LCG partnership formation may have larger greenhouse gas emission than that of atomic LCG partnership.

Lemma 1 *An optimal investment function always exists*

For non-atomic LCG-partnership formation, finding an optimal investment function is a linear programming problem, which can be formulated as follows.

$$\begin{array}{lrcl}
\text{Maximize} & \displaystyle\sum_{i,j\in N} \varphi(i,j) \times \frac{1}{\text{SCL}_j}\left(\frac{\text{PPP}_j}{\text{EFF}_j} - \frac{\text{PPP}_j}{\text{EFF}_i}\right) & & \\
\text{Subject to} & \displaystyle\sum_{j\in N} \varphi(i,j) & \leq & \text{CAP}_i \quad \text{for each } i \in N, \\
& \displaystyle\sum_{i\in N} \varphi(i,j) & \leq & \text{SCL}_j \quad \text{for each } j \in N. \\
& \varphi(i,j) & = & 0 \quad \text{for each } i,j \in N \text{ s.t. } \text{EFF}_i \leq \text{EFF}_j, \\
& \varphi(i,j) & \geq & 0 \quad \text{for each } i,j \in N.
\end{array}$$

Observe that the value $\frac{1}{\text{SCL}_j}\left(\frac{\text{PPP}_j}{\text{EFF}_j} - \frac{\text{PPP}_j}{\text{EFF}_i}\right)$ is determined by input data. Obviously, the total reduction of greenhouse gas emission is bounded from above, and thus, an optimal solution exists. Moreover, this problem can be formulated as a minimum cost flow problem, which can be solved in polynomial time.

For atomic LCG-partnership, finding an optimal investment function is a combinatorial optimization problem with a finite number of solutions, and obviously an optimal solution exists. This problem is NP-hard in the strong sense shown by a polynomial time reduction from BIN-PACKING problem, where each bin is associated to a supplier country (a country with high economic efficiency), and each item is associated to a client country (a country with low economic efficiency).

Lemma 2 *An optimal investment function is stable when all α_is have the same value.*

Proof. Let φ and φ' be investment functions. Suppose φ is dominated by φ' with client switch via pair (i,j). By definition,

$$u_i(\varphi) = \alpha_i \sum_{\ell\in C_i(\varphi)} \text{RDT}_\varphi(i,\ell) + \sum_{k\in S_i(\varphi)} (1-\alpha_k)\text{RDT}_\varphi(k,i),$$

and from $\varphi'(k,\ell) = \varphi(k,\ell)$ for each $(k,\ell) \in N^2$ with $k \neq i$,

$$u_i(\varphi') = \alpha_i \sum_{\ell\in C_i(\varphi')} \text{RDT}_{\varphi'}(i,\ell) + \sum_{k\in S_i(\varphi)} (1-\alpha_k)\text{RDT}_\varphi(k,i).$$

From $u_i(\varphi') > u_i(\varphi)$, we have

$$\sum_{\ell \in C_i(\varphi')} \text{RDT}_{\varphi'}(i,\ell) > \sum_{\ell \in C_i(\varphi)} \text{RDT}_{\varphi}(i,\ell)$$

Moreover, again from $\varphi'(k,\ell) = \varphi(k,\ell)$ for each $(k,\ell) \in N^2$ with $k \neq i$,

$$\sum_{\ell \in C_k(\varphi')} \text{RDT}_{\varphi'}(k,\ell) = \sum_{\ell \in C_k(\varphi)} \text{RDT}_{\varphi}(k,\ell).$$

Hence,

$$\begin{aligned}
\sum_{k,\ell \in N} \text{RDT}_{\varphi'}(k,\ell) &= \sum_{k \in N} \sum_{\ell \in C_k(\varphi')} \text{RDT}_{\varphi'}(k,\ell) \\
&= \sum_{\ell \in C_i(\varphi')} \text{RDT}_{\varphi'}(i,\ell) + \sum_{k \in N \setminus \{i\}} \sum_{\ell \in C_k(\varphi)} \text{RDT}_{\varphi}(k,\ell) \\
&> \sum_{k \in N} \sum_{\ell \in C_k(\varphi)} \text{RDT}_{\varphi}(k,\ell) \\
&= \sum_{k,\ell \in N} \text{RDT}_{\varphi}(k,\ell).
\end{aligned}$$

Suppose φ is dominated by φ' with supplier switch via pair (i,j). From $\varphi'(k,\ell) = \varphi(k,\ell)$ for each $(k,\ell) \in N^2$ with $\ell \neq j$,

$$u_j(\varphi') = \alpha_j \sum_{\ell \in C_j(\varphi)} \text{RDT}_{\varphi}(j,\ell) + \sum_{k \in S_j(\varphi')} (1-\alpha_k) \text{RDT}_{\varphi'}(k,j).$$

From $u_j(\varphi') > u_j(\varphi)$, we have

$$\sum_{k \in S_j(\varphi')} (1-\alpha_k) \text{RDT}_{\varphi'}(k,j) > \sum_{k \in S_j(\varphi)} (1-\alpha_k) \text{RDT}_{\varphi}(k,j),$$

and since all α_ks have the same value (less than 1),

$$\sum_{k \in S_j(\varphi')} \text{RDT}_{\varphi'}(k,j) > \sum_{k \in S_j(\varphi)} \text{RDT}_{\varphi}(k,j).$$

Moreover, again from $\varphi'(k,\ell) = \varphi(k,\ell)$ if $\ell \neq j$,

$$\sum_{k \in S_\ell(\varphi')} \text{RDT}_{\varphi'}(k,\ell) = \sum_{k \in S_\ell(\varphi)} \text{RDT}_{\varphi}(k,\ell).$$

Hence,

$$\sum_{k,\ell \in N} \text{RDT}_{\varphi'}(k,\ell) = \sum_{k \in N} \sum_{k \in S_\ell(\varphi')} \text{RDT}_{\varphi'}(k,\ell)$$

$$
\begin{aligned}
&= \sum_{k \in S_j(\varphi')} \mathrm{RDT}_{\varphi'}(k, j) + \sum_{\ell \in N \setminus \{j\}} \sum_{k \in S_\ell(\varphi)} \mathrm{RDT}_{\varphi}(k, \ell) \\
&> \sum_{\ell \in N} \sum_{k \in S_\ell(\varphi)} \mathrm{RDT}_{\varphi}(k, \ell) \\
&= \sum_{k, \ell \in N} \mathrm{RDT}_{\varphi}(k, \ell).
\end{aligned}
$$

Therefore, every investment function which maximizes the total reduction of greenhouse gas emission, i.e., an optimal investment function, cannot be dominated, which is a stable investment function. ■

References

[1] J. Drèze and J. Greenberg (1980). Hedonic coalitions: optimality and stability. *Econometrica* 48, pp 987-1003.

[2] M. Miyata and S. Sung (2012). A Market Price Model for Low Carbon Growth Partnership. The 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty 2012, Osaka.

[3] M.R. Garey, D.S. Johnson (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Co., San Francisco, MA.

[4] M. Miyata, T. Tsubota and S. Sung (2012). A Game Theoretical Framework for Low Carbon Growth Partnership. International Symposium on Management Engineering 2012, Kitakyushu.

[5] E. Tardos and T. Wexler (2007). Network Formation Games and the Potential Function Method. in *Algorithmic Game Theory* (Edited by N. Nisan, T. Roughgarden, E. Tardos, V.V. Vazirani), Cambridge university press.

[6] International Money Fund (IMF) : Data and Statistics (2012). Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP, URL: http://www.imf.org/external/index.htm.

[7] U.S. Energy Information Administration (EIA): International Energy Statistics (2012). "Total Carbon Dioxide Emissions from the Consumption of Energy (Million Metric Tons)", URL: http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm.

Notes on Solving the Pipelined Set Cover Problem

Václav Lín
Department of Exact Methods
Faculty of Management, Prague University of Economics
Czech Republic
e-mail: lin@utia.cas.cz

Abstract

We describe a our work on exact algorithms for solving a combinatorial problem called *pipelined set cover*. The problem is *NP*-hard and combines set covering with sequencing. We describe an A^* search algorithm as well as an integer linear programming formulation of the problem.

1 Description of the problem

Min-sum set cover is a minimum latency version of the well known set cover problem: we have a finite set U and a collection of subsets $\mathcal{C} \subseteq 2^U$. Our task is to find a linear ordering $S_{\pi(1)}, S_{\pi(2)}, \ldots, S_{\pi(|\mathcal{C}|)}$ of the sets in $\mathcal{C}$ minimizing

$$\sum_{u \in U} i_\pi(u)$$

where $i_\pi(u) = \min\{i : u \in S_{\pi(i)}, 1 \leq i \leq |\mathcal{C}|\}$. That is, $i_\pi(u)$ is the index of the first set $S_{\pi(i)} \in \mathcal{C}$ containing ("covering") element u under the ordering π.

The *pipelined set cover* problem is a generalization of min-sum set cover where we have

- a "cost" function $c : \mathcal{C} \to \mathbb{R}^+$,
- and a "weight" function $w : U \to \mathbb{R}^+$. In this paper we do assume (without loosing generality) $w : U \to [0, 1]$, and $\sum_{u \in U} w(u) = 1$.

Our task is to find an ordering π of the sets in $\mathcal{C}$ minimizing

$$\sum_{u \in U} w(u) \cdot \sum_{j=1}^{i_\pi(u)} c\big(S_{\pi(j)}\big) \tag{1}$$

Note that when $c = 1$ and $w = 1/|U|$, the pipelined set cover coincides with min-sum set cover.

Applications and references. The min-sum set cover problem is defined and analyzed in [Feige et al., 2004]. The pipelined set cover problem is studied by Munagala et al. [2005] and Kaplan et al. [2005]. Applications mentioned in the cited papers range from speeding up matrix computations to database query optimization and machine learning.

As observed in [Lín, 2013], an application of the pipelined set cover arises in the field of *decision-theoretic troubleshooting* (or sequential fault diagnosis):

- Let U be a set of faults and $w(u)$ be a probability of occurence of fault u.
- Let $\mathcal{C}$ be a collection of tests. Each test $S \in \mathcal{C}$ bears a cost $c(S)$ and can reliably detect all the faults $u \in S$, it cannot detect any other faults. By reliable detection is meant that there is zero probability of error.

Then (1) is the expected cost of detecting a fault by a sequence of tests. In the terminology of troubleshooting [Ottosen, 2012], this diagnostic problem is a special case of single-fault troubleshooting with dependent actions.

2 Integer linear programming (ILP)

We give an original ILP formulation of the pipelined set-cover problem. For the ease of presentation, we assume that the elements of U and $\mathcal{C}$ are indexed as $u_1, \ldots, u_m$ and $S_1, \ldots, S_n$. The corresponding costs and weights are denoted $c_j = c(S_j)$ and $w_j = w(u_j)$. We assume $\bigcup_j S_j = U$ and $\forall_j\ S_j \neq \emptyset$.

Our linear model contains two sets of binary variables:

- x_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$.
- δ_{jk}, $j = 1, \ldots, n$, $k = 1, \ldots, n$.

Given a strict linear ordering π of the sets in $\mathcal{C}$, we have $x_{ij} = 1$ if element u_i is covered by some set preceeding S_j under π, otherwise $x_{ij} = 0$. Variables δ_{jk} are used to encode the ordering π. We have $\delta_{jk} = 1$ if set S_j precedes S_k under the ordering π, otherwise $\delta_{jk} = 0$.

The function (1) can be written as $\sum_{i=1}^{m} w_i \cdot \sum_{j=1}^{n} c_j(1 - x_{ij})$. Minimizing (1) is therefore equivalent to

$$\max \sum_{i,j} w_i \cdot c_j \cdot x_{ij} \ , \tag{2}$$

which is the *objective function* of our linear program. Following are the *constraints*.

For every u_i and every pair of distinct sets S_j, S_k covering u_i, we require $\delta_{jk} \Rightarrow x_{ik}$:

$$\delta_{jk} \leq x_{ik} \tag{3}$$

We denote $J(u_i) = \{j : S_j \ni u_i\}$. We require $x_{ij} \Rightarrow \bigvee_{k \in J(u_i) \setminus \{j\}} \delta_{kj}$ for every u_i and S_j:

$$x_{ij} \leq \sum_{k \in J(u_i) \setminus \{j\}} \delta_{kj} \tag{4}$$

Constr.	count
(5)	$\binom{n}{2}$
(6)	n
(7)	$\binom{n}{3}$
(3)	$2\sum_{i=1}^{m}\binom{\lvert J(u_i)\rvert}{2}$
(4)	mn

Table 1: Numbers of constraints

The strict linear ordering π of sets is asymmetric ($\delta_{jk} \Leftrightarrow \neg\delta_{kj}$), irreflexive ($\delta_{jj} = 0$) and transitive ($\delta_{jk} \wedge \delta_{k\ell} \Rightarrow \delta_{j\ell}$). We enforce these properties by the following (in)equalities:
Asymmetry:

$$(\forall_{j\neq k})\ \delta_{jk} + \delta_{kj} = 1 \tag{5}$$

Irreflexivity:

$$(\forall_j)\ \delta_{jj} = 0 \tag{6}$$

Transitivity:

$$\delta_{jk} + \delta_{k\ell} - \delta_{j\ell} \leq 1 \text{ for all distinct } j, k, \ell \tag{7}$$

For all the variables, we set bounds $0 \leq x_{ij} \leq 1$, $0 \leq \delta_{jk} \leq 1$. We require that all the variables are *integral*, that is, they either equal 0 or 1. However, for computatinal reasons it is desirable to declare as few integral variables as possible. In our model, it is sufficient to explicitely declare as integral only a subset of the δ-variables:

- When any $\delta_{jk} \in \{0, 1\}$, then δ_{kj} is "forced" to be integral by constraint (5). Hence only δ_{jk} needs to be explicitely declared as integral.
- All the δ_{jj} variables are "forced" to be integral by constraint (6).
- When all the δ's attain integral values, then so do all the x's due to constraints (3) and (4).

The linear model contains $mn + n^2$ variables of which $\frac{n^2-n}{2}$ are integral, and $O(n^3 + mn)$ constraints as indicated in Table 1.

3 A^* search

Vomlelová and Vomlel [2003] designed an AO^* algorithm for quite general troubleshooting problem. The algorithm and its guiding heuristic function was later applied in a less general setting by Ottosen [2012], who analyzed and slightly modified the heuristic function. We use basically the same algorithm as [Ottosen, 2012], but our formulation and analysis is simpler because we consider a special case of the troubleshooting problem.

Heuristic function. Let $\mathbf{s} = S_{\pi(1)}, S_{\pi(2)}, \ldots$ be a sequence of some sets from $\mathcal{C}$ and let $\cup\mathbf{s} \subseteq U$ be the set of elements of U covered by $\mathbf{s}$. We define three functions:

$$\begin{aligned} g(\mathbf{s}) &= \sum_{u \in \cup\mathbf{s}} w(u) \cdot \sum_{j=1}^{i_\pi(u)} c\big(S_{\pi(j)}\big), \\ h(\mathbf{s}) &= \sum_{u \notin \cup\mathbf{s}} w(u) \cdot \min_{S \in \mathcal{C} : u \in S} c(S), \\ f(\mathbf{s}) &= g(\mathbf{s}) + h(\mathbf{s}). \end{aligned}$$

Value $g(\mathbf{s})$ is the weighted cost of covering $\cup\mathbf{s}$, value $h(\mathbf{s})$ is a lower bound of the weighted cost required to cover $U \setminus \cup\mathbf{s}$, and value $f(\mathbf{s})$ is a lower bound of total weighted cost of any sequence starting with $\mathbf{s}$.

Function h is the heuristic we use to guide A^* search. Ottosen [2012] proved that it has the desirable property of being admissible and monotonic. *Monotonicity* means that for any sequence $\mathbf{s}$ and any set $S \notin \mathbf{s}$, we have

$$h(\mathbf{s}) \leq c(S) + h(\mathbf{s}, S).$$

Monotonicity is easy to verify for our simplified formulation of h:

$$h(\mathbf{s}) - h(\mathbf{s}, S) = \sum_{u \in S \setminus \cup\mathbf{s}} w(u) \cdot \min_{S' \ni u} c(S') \leq \sum_{u \in S \setminus \cup\mathbf{s}} w(u) \cdot c(S) \leq c(S) \cdot \underbrace{\sum_{u \in U} w(u)}_{=1}$$

Let $h^*(\mathbf{s})$ be the optimal cost of covering $U \setminus \cup\mathbf{s}$. Function h is *admissible*:

$$(\forall \mathbf{s}) \quad h(\mathbf{s}) \leq h^*(\mathbf{s}) \ .$$

Admissibility follows from the monotonicity of h by a simple inductive argument.

Algorithm. The A^* algorithm generates sequences of sets by successively appending sets from $\mathcal{C}$ in a best-first fashion. The full algorithm listing is shown at page 5.

The candidate sequences to be expanded are stored in `OPEN`, a priority queue. Priority in the queue is given by the function f: operation `OPEN.pop()` always returns a sequence $\mathbf{s}$ with the minimal value $f(\mathbf{s})$, and removes $\mathbf{s}$ from the queue. We use associative array `coalesceMap` for pruning. The array is indexed by (a representation of) subsets of U. When we store a sequence $\mathbf{s}$ in `coalesceMap`, we store it at index corresponding to $\cup\mathbf{s}$. Obviously. efficient implementation of data structures `OPEN` and `coalesceMap` is important for performance of the algorithm.

We use a greedy algorithm[1] to find an initial feasible solution (line 5 of the algorithm listing 1). This allows us to prune costly sequences early on. We prune sequences that cannot be optimal at several places in the algorithm:

[1] Known as *Updating P-over-C* in troubleshooting, cf. Ottosen [2012].

- At lines 8 and 15, we prune any sequence for which the lower bound is worse than the lowest currently known cost.
- At line 10, we avoid appending an action that covers only a subset of U of zero weight.
- At line 17, we prune any sequence that is more costly than some other already constructed sequence covering the same subset of elements.

We refer to [Ottosen, 2012] for further analysis and pointers to literature.

Algorithm 1: A^* algorithm for the pipelined set cover

input : U, $\mathcal{C}$, w, c
output: An optimal sequence of sets covering U

```
   /* initialization */
 1 coalesceMap := {} ;                                  // associative array
 2 OPEN := {} ;                                         // priority queue
 3 s ← ∅;                                               // empty sequence
 4 OPEN.push(s);
   /* find initial feasible solution */
 5 bestSeq ← Greedy(U,C) ;
   /* main loop - expand partial sequences */
 6 while OPEN ≠ ∅ do
 7     s ← OPEN.pop() ;                          // s = arg min_{t∈OPEN} f(t)
 8     if f(s) ≥ g(bestSeq) then continue;
 9     ;
       /* try expanding s with unused sets */
10     for A ∈ {S ∈ C : S ∉ s} do
11         if 0 = Σ_{u∈A\∪s} w(u) then continue;
12         ;
13         s' ← s‖A ;                                  // append A to s
14         if ∪s' = U then
               /* Is s' better than the best known solution?  */
15             if g(s')≤ g(bestSeq) then bestSeq ← s' ;
16             ;
17         else sequence s' is a partial solution
18             if f(s') ≥ g(bestSeq) then continue;
19             ;
20             if coalesceMap(∪s') ≠ ∅ then
21                 if g(coalesceMap(∪s')) ≤ g(s') then continue;
22                 ;
               /* s' is a candidate for further expanding */
23             coalesceMap(∪s') ← s';
24             OPEN.push(s');
25 return bestSeq
```

Function Greedy(U, $\mathcal{C}$)

input : U, $\mathcal{C}$, w, c
output: A sequence $\mathbf{s}$ of sets covering U

```
1 s ← ∅ ;                                   // initialize with empty sequence
2 while U ≠ ∅ do
3     A ← arg max_{S∈C} Σ_{u∈S∩U} w(u)/c(S);
4     s ← s‖A;                                          // append A to s
5     C ← C \ {A};
6     U ← U \ A;
7 return s
```

4 Discussion

In the talk, we will discuss some of the computatinal experiences with the above two algorithms. We have used the GLPSOL/GLPK solver for integer programming. It should be noted that solving the pipelined set cover with a general purpose ILP solver is at the moment inferior to using the A^* algorithm. The author is currently working on a specific branch&bound algorithm that takes into account the specific form of the pipelined set cover problem.

References

Uriel Feige, László Lovász, and Prasad Tetali. Approximating Min Sum Set Cover. *Algorithmica*, 40(4):219–234, 2004.

Haim Kaplan, Eyal Kushilevitz, and Yishay Mansour. Learning with attribute costs. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 356–365. ACM, 2005.

Václav Lín. Decision-theoretic Troubleshooting: Hardness of approximation. *International Journal of Approximate Reasoning*, 2013.

Kamesh Munagala, Shivnath Babu, Rajeev Motwani, and Jennifer Widom. The pipelined set cover problem. In Thomas Eiter and Leonid Libkin, editors, *ICDT*, volume 3363 of *Lecture Notes in Computer Science*, pages 83–98. Springer, 2005. ISBN 3-540-24288-0.

Thorsten J. Ottosen. *Solutions and Heuristics for Troubleshooting with Dependent Actions and Conditional Costs (PhD thesis)*. Aalborg University, Denmark, 2012.

Marta Vomlelová and Jiří Vomlel. Troubleshooting: NP-hardness and solution methods. *Soft Comput.*, 7(5):357–368, 2003.

Fuzzy rule based behavioral models in boarding a train

Petr Čermák
Research Institute of the IT4Innovations Centre of Excellence
Silesian university in Opava
petr.cermak@fpf.slu.cz

Martin Morawiec
Institute of Computer Science
Silesian University in Opava, Czech Republic
olda220@gmail.com

August 29, 2013

Abstract

This paper describes behavioral models in a team of agents. These models are applied in the simulation of the ingoing crowd to a train. To find the optimal behavior of agents we use genetic algorithms. Behavior is encoded by setting of fuzzy rule-base system. The theoretical part describes genetic algorithms, linguistic modeling and fuzzy sets. In the practical part we create genetic algorithm in language Scheme and also we define the inputs and outputs of the fuzzy system. We will design rules of fuzzy system for speed and linguistic variables for speed and direction. Then program application to simulate the ingoing crowd to a train.

1 Introduction

This paper presents an experiment for ingoing crowd to a train with using genetic algorithms, fuzzy rules Takagi-Sugeno (TSK) and Mamdani and their comparison. This solution has been implemented by using of Fuznet which solves structural and parametric identification of used fuzzy models. Fuznet has been implemented in programming language Scheme[1] and subsequently were developed the application SimDav programmed in C#, inspired by. Firstly were generated a finite set of individuals which represent a set of fuzzy rules (TSK and/or Mamdani model). The individual represent evaluation of truth of each rule.

The first theoretical part introduces theoretical background of used methods such as genetic algorithms and TSK and Mamdani fuzzy rules. Introduces how to select an optimal rules to ingoing[5].

The practical part of this paper discusses reached results from practical experiments with my developed software using Scheme. In addition has been developed a GUI application in C# which communicates over UDP with Scheme libraries[5].

2 Modeling of crowd behavior

Whole system is implemented as a multi-agent system. In this system, an agent represents a person from the crowd. Each agent is defined by position, direction, age and state of health. The following figure demonstrates mapping of rules and behavior patterns which are used in simulation (Fig. 1). The following formulas describe fuzzy modeling by using TSK and Mamdani models[3]. Combination of linguistic values for the speed and direction determines rules for used models. By using genetic algorithms (GA) we obtain an ideal set of rules to fast ingoing of agents to the train.

$$\begin{array}{c} IF\ (x_1\ is\ A_{1,1})\ AND\ \ldots\ (x\ is\ A_{n,1})\ THEN\ (y_1\ is\ C_1) \\ IF\ (x_1\ is\ A_{1,1})\ AND\ \ldots\ (x\ is\ A_{n,1})\ THEN\ (y_2\ is\ C_2) \\ \vdots \\ IF\ (x_1\ is\ A_{1,r})\ AND\ \ldots\ (x\ is\ A_{n,r})\ THEN\ (y_r\ is\ C_r) \end{array} \tag{1}$$

$$\begin{array}{c} IF\ (x_1\ is\ A_{1,1})\ AND\ \ldots\ (x_n\ is\ A_{n,1})\ THEN\ (y_1\ is\ f\,(x_1,x_2,x_3,...,x_n)) \\ IF\ (x_1\ is\ A_{1,1})\ AND\ \ldots\ (x_n\ is\ A_{n,1})\ THEN\ (y_2\ is\ f\,(x_1,x_2,x_3,...,x_n)) \\ \vdots \\ IF\ (x_1\ is\ A_{1,r})\ AND\ \ldots\ (x_n\ is\ A_{n,r})\ THEN\ (y_r\ is\ f\,(x_1,x_2,x_3,...,x_n)) \end{array} \tag{2}$$

3 Movement speed of each agent

Inputs for the movement speed are linguistic variables of *Distance from the obstacle in direction*, *Age of the agent*, *Crowd density*, *State of health of the agent* and *Remaining time to departure of the train*. Output of the model is given by speed of the agent. Rules are not generated by genetic algorithm but are created a priori.

Linguistic variables and rules are selected according to real life. We selected 5 linguistics variables which could take account of influence of speed. The first used variable is *Age of the agent* which represents real age of a person. For this variable we defined three linguistic values – *youth*, *adult* and *senior*.

Speed also depends on crowd density. In addition, age depends on state of health which is an another variable. The *State of health* represents its movement disorders, e.g. wheelchair user, broken limb, mental disorders, etc. Linguistic values for this variable are *Healthy user* and *Wheelchair user*.

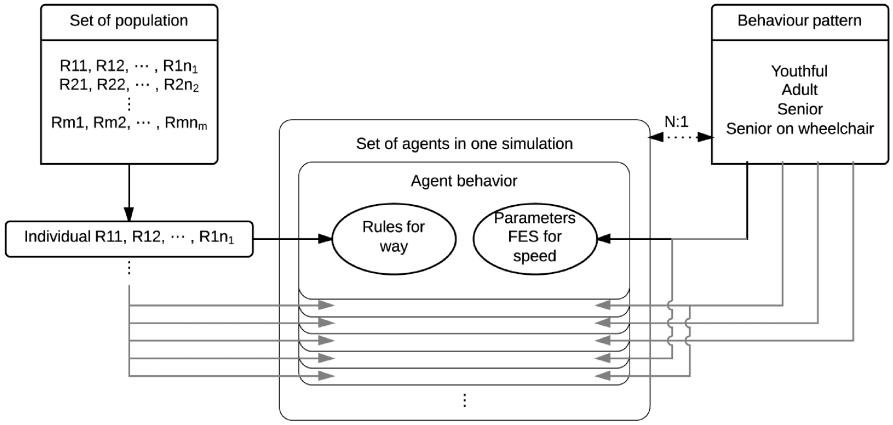

Figure 1: Rules and behavior patterns mapping of agents

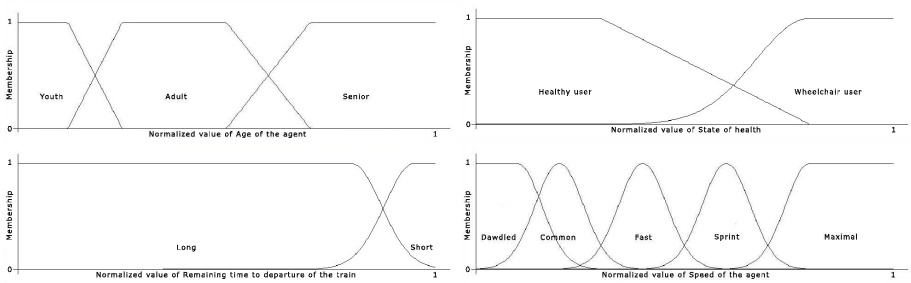

Figure 2: Age of the agent, State of health, Remaining time to departure of the train and Speed of the agent

Last variable *Remaining time to departure of the train* describes hurrying of the agent. Values are *Long* and *Short*.

All sets of antecedent are defined. So that, we can define consequent (output) which represents *Speed of the agent*. Maximal speed is given by running world record. Names of linguistics values are *Dawdled*, *Common*, *Fast*, *Sprint* and *Maximal*.

4 Movement direction of each agent

In this phase we have created model of speed movement of each agent. Self-evaluation, fear, social norms and cultural patterns are also known factors influence of the behavior[2]. Direct relation among these factors and behavior does not exist. In other words, we cannot with absolute certainty know how people will act. Subsequently, we must create the second model which represents move-

Age	Time train dep.	Crowd density	State of health of the agent	Speed
youthful	long	free passage	healthy user	common
youthful	long	free passage	wheelchair user	dawdled
youthful	long	limited move	healthy user	common
youthful	long	limited move	wheelchair user	dawdled
youthful	short	free passage	healthy user	fast *and* sprint
youthful	short	free passage	wheelchair user	common
youthful	short	limited move	healthy user	fast
youthful	short	limited move	wheelchair user	common
adult	long	free passage	healthy user	common *and* fast
adult	long	free passage	wheelchair user	dawdled
adult	long	limited move	healthy user	dawdled *and* com.
adult	long	limited move	wheelchair user	dawdled
adult	short	free passage	healthy user	maximal
adult	short	free passage	wheelchair user	common
adult	short	limited move	healthy user	fast
adult	short	limited move	wheelchair user	common
senior	long	free passage	healthy user	common
senior	long	free passage	wheelchair user	common
senior	long	limited move	healthy user	dawdled
senior	long	limited move	wheelchair user	dawdled
senior	short	free passage	healthy user	fast *and* sprint
senior	short	free passage	wheelchair user	dawdled
senior	short	free passage	wheelchair user	common
senior	short	limited move	healthy user	dawdled *and* com.
senior	short	limited move	wheelchair user	dawdled

Table 1: Linguistic model for speed

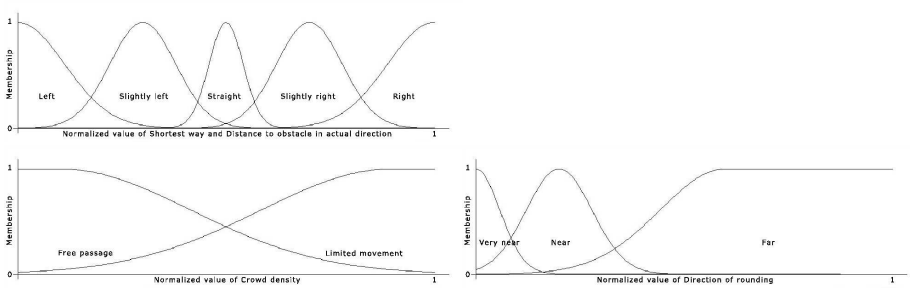

Figure 3: The shortest way and Distance to obstacle in actual direction, Crowd density and Direction of rounding

ment direction which is influenced by different parameters such as *The shortest way*, *Distance to obstacle in actual direction*, *Crowd density in actual direction*, *Direction of rounding*. For the variable *The shortest way* and *Distance to obstacle in actual direction* are linguistic values named as *Left*, *Slightly left*, *Straight*, *Slightly right* and *Right*.

Linguistic values for *Crowd density in actual direction* are *Free passage* and *Limited movement*. For *Direction of rounding* are linguistic values named as *Very near*, *Near* and *Far*.

4.1 Using genetic algorithms for this simulation

This phase is solved by using genetic algorithms (GA) inspired by biological genetic processes of cells. The GA is defined by 7-tuple as follows[4]:

$$GA{=}(N, P, f, \Theta, \Omega, \Psi, \tau), \tag{3}$$

where P is population containing N elements, Θrepresents parent selection operator, which select u elements from P, $\Omega$is set of genetic operators, which include crossover operator Ω_c, mutation operator Ω_m and others problem-oriented or implementation-oriented specific operators, which all together generate v offspring from u parents. $\Psi$is deletion operator, which removes v selected elements from actual population $P(t)$. v elements is add to new population $P(t{+}1)$ after it, $\tau$is stop-criterion. Parent selection operator $\Theta$and genetic operators Ωhave stochastic character, deletion operator $\Psi$is generally deterministic. Chromosome is compound of genes; each gene represents certain step of the Scenario. Stop criterion of the evolution process is given by ingoing of all agents to the train.

To individual selection we use Weighted Roulette algorithm. Size of parts on the roulette of each individual is given by its fitness value.

$$\frac{f_i}{\sum_i^N f_i}, \tag{4}$$

probability of selection of i-th individual is given by the formula (4) where $f_i$is the fitness value of i-th individual and N is the number of individuals in population. The following formula describes fitness function

$$f = \frac{60}{c}g + t, \tag{5}$$

where g is number of agents in goal, t remaining time and c total number of all agents.

5 Simulation experiments with SimDav

In this section we will present practical experimental simulation which uses Scheme with library Fuznet which is able to create a structure of a TSK and Mamdani model based on previously discussed theoretical background.

Within the practical experiment has been generated population for TSK. We used 10 agents and the simulation ran for 3 days a has been generated 3767 generations, each of them contains 8 individuals though all agents did not get to the train. Number of combination of linguistic values in antecedent was of 150 and output linguistic variable have 5 values, so that totally 750 combinations (genes in individuals). This simulation was very slow and did not fulfill expected goal (all agents got to the train). Due to these limitations has been optimized genetic operators of crossover and mutation to optimal solution. This non-optimal version of GA used one point crossover and mutated only on one position in chromosome. Structure of each individual that each 5 rules (genes in line) have the same antecedent part.

New version of GA is based on generation of population for both of models – TSK an Mamdani. In case of new optimized version of GA it was configured with probability of 50 % of crossovered selected individuals. Amount of selection was 5 genes. Mutation was changed similarly. In other words, each individual can mutates each 5 genes. Position of the mutation is random and their number is various. The following figure shows increasing of the fitness values of individuals depending on number of generations. Highlighted line of fitness value of 60 means that all agents got to the train. For both of models were used the same genetic operators settings, number of individuals, agents and their parameters. Initial population was generated randomly and because could be make more favorable one of used models.

Parameters and number of agents is equal as for previous experiment. We investigate that population size influences time of finding an ideal individual. As before we compared both of models, TSK and Mamdani. With larger population we found an optimal solution more rapidly.

In addition, we changes number of agents to 20 from 10 and we created of 15 combinations of frequency of types of behavior patterns for agents, see Table 2. In this experiment we used TSK only with the best individual from generations.

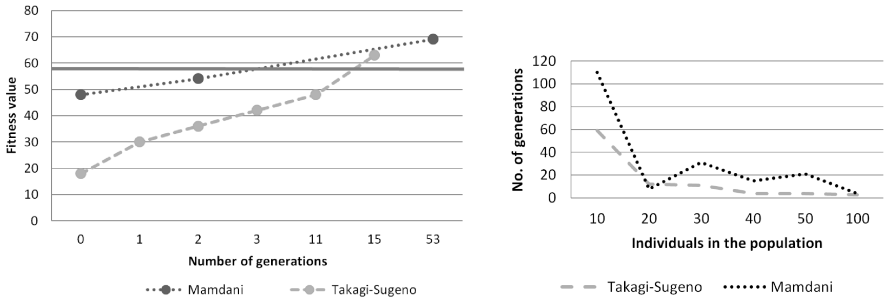

Figure 4: Dependency number of generation on their size (left), Dependency number of generation on number of individuals in generation (right)

Binary combination	Youthful	Adult	Senior	Senior in a wheelchair	Time
0001	4	4	4	8	180
0010	4	4	8	4	175
0011	3	3	7	7	134
0100	4	8	4	4	134
0101	3	7	3	7	140
0110	3	7	7	3	140
0111	2	6	6	6	142
1000	8	4	4	4	128
1001	7	3	3	7	159
1010	7	3	7	3	159
1011	6	2	6	6	161
1100	7	7	3	3	134
1101	6	6	2	6	118
1110	6	6	6	2	116
1111	5	5	5	5	134

Table 2: Time dependency on combination of Agents behavior patterns

6 Conclusion

The main aim of the presented paper was to simulate of ingoing crowd at the railway station with creating behavior patterns, an application, data processing from the simulation and experiments. We created behavior patterns using TSK and Mamdani. In this developed system we use rules generated by genetic algorithm. Also we designed linguistic variables and simulation environment and also we improved using GA for this specific situation.

We investigated generation of populations by using both of models and subsequently we proved that TSK is better for this situation. Experiments with

different population size proved that larger population resulted to faster convergence in time to finding an optimal individual. The last experiment investigated dependency of ingoing time on combination of Agents behavior patterns.

In future, the application will be improved with environment editor and we also use different inputs into Fuznet. This work only simulates presumable behavior of the crowd but we do not known what people will do in different situations, conditions and unexpected occurrences.

Acknowledgments

This work was supported by the European Social Fund and the Czech Republic State Budget in the Operational program Education for Competitiveness project namely "ICT Interdisciplinary education with foreign language ability" (CZ.1.07/2.2.00/28.0014) and the IT4Innovations Centre of Excellence project, reg. No. CZ.1.05/1.1.00/02.0070 supported by Operational Program 'Research and Development for Innovations' funded by Structural Funds of the European Union and state budget of the Czech Republic.

References

[1] Higher-order and symbolic computation: Revised5 Report on the Algorithmic Language Scheme. Hingham, MA, USA: Kluwer Academic Publishers, 1998, 11, No. 1. ISSN 1388-3690. http://cs.au.dk/~hosc/11-1/sR5.pdf

[2] Ajzen I., Fishbein M.: Understanding attitudes and pre- dicting social behavior. Englewood Cliffs, Prentice-Hall, 1980, 280 s. ISBN 9780139364433

[3] Cermak, P., Pokorny, M.: An Improvement of Non-Linear Neuro-Fuzzy Model Properties, In: Neural Network World, ICS AV CR, Prague, CZ, Vol 11, No. 5 (2001) 503 - 523

[4] Cermak, P. and Chmiel, P.: Parameters optimization of fuzzy-neural dynamic model, NAFIPS2004, Banff, Canada Vol 2, (2004) 762 - 767

[5] Morawiec, M.: The team of agents with elements of behaviors, MSc. thesis in czech, Silesian university in Opava, Opava, CZ (2013)

[6] SAVANNAH SIMULATIONS AG. SIMWALK: the simulation software for pedestrian flows [online]. 2013 [cit. 2013-04-04]. http://www.simwalk.com

[7] THUNDERHEAD ENGINEERING: Pathfinder [online]. 2011 [cit. 2013-04-04], http://www.thunderheadeng.com/pathfinder

On the Practical Issues of Document Context Extraction And Similarity Measuring in Multilingual Environment

Jiří Přibil
Department of Exact Methods, Faculty of Management
University of Economics, Prague
pribil@fm.vse.cz

Abstract

The process of document context extracting, which expresses the basic topic (or topics) of a text, is unusually difficult algorithmic task. In our previous work we were dealing mainly with the question of the basic definition of document similarity and process of searching of (similar) sections of text, regardless of the language of the document. This uninformed approach is algorithmically more complex than the classic approach, which does use different linguistic tools of the given language (thesauri, stemmers, etc.).
In practical experiments on a large-scale extraction storage it is necessary to solve problems that arise from the nature of the information processed of document - each language has its own dictionary and the authors use only a subset; documents must be acquired somehow, must be processes and stored for later measurements.
Furthermore, the described issues are a small selection from the experience gained from document context extraction experiments in the corporate sector - especially managers are exposed to the high information pressure and is why tools of business intelligence are of great importance.

1 Introduction and Motivation

In [1] and [2] we discuss general procedures for basic conformity and similarity characteristics of the documents for the purpose of plagiarism detection, in [3] are introduced some basic algorithmic procedures to calculate and visualize (plagiarized) context of the document and [4] shows some application in business decision supporting systems.

In the terms of the practical usability, the most important area of interest is just the area of corporate management - corporate managers I both the profit

and non-profit companies have to monitor customer reviews on the internet, working with positive (but mostly negative) reactions, trying to act proactively and strengthen customer relationships. In this paper we describe a case where there is an automatic system that scans (multilingual) website containing customer references and must be able to correctly categorize the various posts, evaluate and extract their context.

Experiments described below were performed on a corpus of 2,305 theses in Czech and Slovak qualification works defended at the University of Economics in Prague.

2 Problems of Natural Text Processing

Analysis of the natural text compared to any artificial language (such as any programming language) is a much more difficult task - any artificial language has a clear and unequivocal rules for subsequent analysis of the content or at least structure; any deviations are identified as errors and the analysis of non-valid text is not possible. A set of rules for such a language is finite and relatively small and contains maximally thousands of rules.

For documents written in any in natural language, there is a set of rules to follow in each language. The main problem is that every speaker adapts these common rules to their own needs and each language changes and adapts over time. For example, if we consider one hundred customers describing the same (good or bad) experience with the product, we get almost one hundred of (very) different specific expressions, which, however, are identical in content.

Because the language in which customers write their assessments is very often colloquial, expressive and of different "linguistic quality" (spelling, grammar), the monitoring system that must contain specific degree of tolerance to errors. This system must thus include elements of fuzzy logic and practical experiments show that the key importance of a correct set of system parameters.

Statistical models [5] often use additional vocabularies (often synonymous) to improve their function. These dictionaries are not an essential element of these systems the analysis of customer references has to identify all texts regardless of single word or its synonym. The main problem, however, is that these dictionaries are not freely available and introduces more expenses to any system.

The larger corpus of documents on which we build a statistical model we have, the more it is possible to do this without the dictionary methods. It is also the aim of our experiments

3 Document Preprocessing

The basic steps of pre-processing of the respective text before the context analysis must be performed first. A few general steps are adequately described and for the purpose of customer reviews analysis are reduced to basic conversion to

plain text; these documents need not contain only text, but can enclose pictures, graphs, diagrams, formulas, tables, and other objects. These objects are often difficult to represent by plain text.

Even the text itself can be formatted using a different styles and fonts. Text can be variously divided (into sentences, paragraphs and chapters). The additional formatting enhances information beyond simple text - this added value is also interesting from the view of document context extraction. Two words (terms, topics) which are located next to each other in the same sentence have "stronger bond" than the words that are adjacent, but stands at the end of first and beginning of the next sentence.

3.1 Tokenization

As a token we denote [6] such an identification of lexical units, which have separate meaning and function. Token may yet be made up of one word or a steady phrase; thus the set of words is in any language only a subset of its tokens. Tokens are composed of one or more characters - eg. word, number, date, website address, expressions containing mixed letters, numbers and other characters, sentence delimiters, separators and more complex sentences.

For example, [5] provides two different ways of creating tokens from the document - top-down and bottom-up algorithms. The need for a set of tokenizing rules for each language is the reason why it is not usable in any uninformed and multi-language system built only on statistical analysis.

The output of the tokenization is then a list of tokens, de facto the gross structure of the text converted to basic tokens (Figure 1).

```
WORD(the) WORD(quick) WORD(brown) WORD(fox) COMMA(,) WORD(grey) WORD(wolf)
WORD(and) WORD(big) WORD(cat) WORD(jump) WORD(over) WORD(the) WORD(lazy)
WORD(dog) DOT(.)
```

Figure 1: Textual representation of the output of basic lexical analyzer.

Also, sentences may contain significantly more complicated tokens. Complex tokens, such as abbreviations or phrases, are worth to analyze and detect at other layers of the analyzer, which stands immediately above the lexical analyzer (Figure 2).

```
STOPWORD(the) WORD(quick) WORD(brown) ANIMAL(fox) COMMA(,) WORD(grey)
ANIMAL(wolf) WORD(and) WORD(big) ANIMAL(cat) WORD(jump) WORD(over)
STOPWORD(the) WORD(lazy) ANIMAL(dog) DOT(.)
```

Figure 2: Textual representation of the output of advanced lexical analyzer.

3.2 Stop words and document language detection

Stop word is a word or phrase (thus token) that has a high incidence among documents, and at the same time it is a word that is often used within a single document - frequent pronouns, prepositions, conjunctions and other words. Stop words for text have a very low added information value, and can be ruled out of context analysis completely; procedure to find stop words in the single-language corpus is very easy and well described [7].

3.3 Stemming

Many languages (especially Slavic) can generate dozens of different forms of the same word. For further analysis is necessary to find a common morphological root of each word - lemma.

As stemming we call the process of searching for the base of the word for further processing. The result of this adjustment is the stem, which may not coincide with linguistic root-words, but for many tasks of text mining this simplification is sufficient.

In practice, the multilingual system must use algorithmic stemmer. These are facing with a problem of many exceptions in the language and the stemming procedure may result in the implementation of two types of errors - Over-stemming and Under-stemming [8].

4 Proposed solution: The algorithm based on identifiers

The basic task of document context extraction is the searching (and subsequent comparison and statistical analysis) of exactly defined sections of the text. At this point it will be introduced and discussed the possibility of using algorithms based on wildcard identifiers.

Out of this option we may consider eg. vector algorithms (text or its parts is converted to vectors, where each position in this vector corresponds to a word in the dictionary and the value at this position corresponds to the frequency of the word in a document) or plain text processing algorithms (algorithms are based on repeated re-scanning of the sample text in the document, which is a very time-consuming and inefficient method).

Experiments show that text identifiers are the fastest method that enables any further document processing. The main attention is paid to methods that use n-grams - especially bigrams - and their unordered modifications [9].

4.1 Algorithms working with text identifiers

The algorithm in this case does not work with tokens directly in their text representation, but with the numeric identifiers for each token. Characters that

are not part of words (commas, dots, dashes, parentheses) are eliminated or their presence in the text is represented by other means.

A necessary condition is that the identifiers used in this transformation match the identifiers that were used in the transformation of the documents in the document repository. Identifiers can be assigned almost arbitrarily, but different words have to be always represented by different identifier. In practice, assignment of incremented numbers seems to be suitable.

This approach of text representation highlights the importance of document normalization and good stemming of individual words. Two different stems, which differ from each other even by a single letter, will receive a totally different identifier. The algorithm then has only to compare those identifiers and check their identity. Faulty stemming can generate the same root word for many different identifiers - this will result in errors in calculation of the document context.

The cardinality of the set of identifiers must completely cover all the possible words that can come across in the documents. The above declared test corpus of 2,305 documents in Czech and Slovak languages after verbal standardization reached the cardinality of vocabulary over 200,000 unique words. The size of vocabulary grows almost continuously with the increasing number of processed documents - but the Czech and Slovak vocabulary will probably not outgrow the number of 500,000 words.

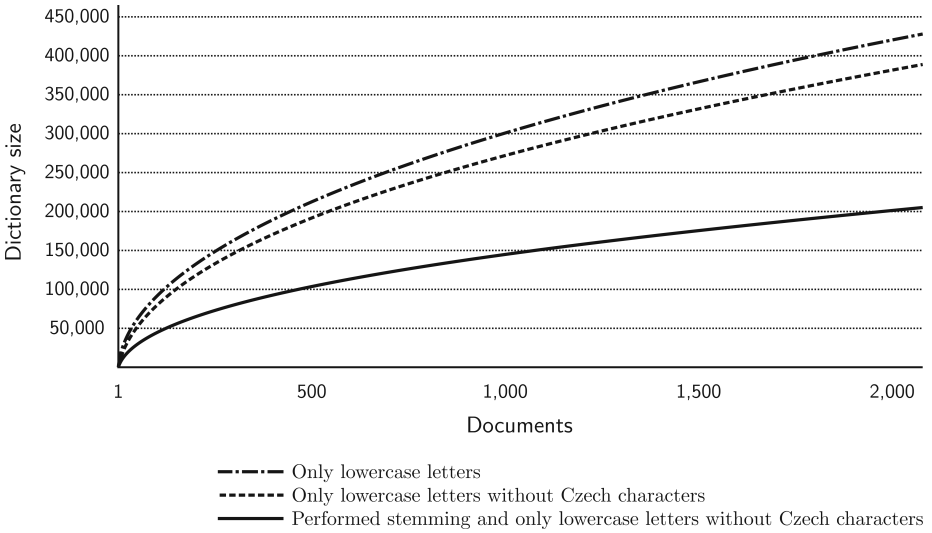

Figure 3: The relation between the dictionary size and the method of word normalization.

This approximately corresponds to the identifier of at least 19 bits - in practice it is appropriate to use the next common length of 32 bits; this represents a total of 4,294,967,296 different tokens. From this perspective, there is con-

siderable potential for reserve growth in vocabulary and it is possible to assign identifiers for many other languages at once.

Numeric representation of the tokens has two major advantages:

- higher efficiency than string representation,
- faster implementation of atomic operations on numbers than on strings.

4.2 Sentence breakdown

The documents consist of a set of sentences that are further comprised of individual words. Any human reader does not perceive the words as isolated, but always in the context of other words in a sentence; if the document does not contain uppercase and lowercase letters, commas or periods, it would be very difficult for the reader to get orientation in the information.

In pursuit of the most accurate document context extraction it is very important to observe the quality of the tokenization. Unfortunately, the sentences cannot be determined only on the basis of dot and comma separators (or other characters, such as parentheses, dash, ellipsis, and more). These symbols are commonly part of tokens - the best examples of such a behavior include shortcuts, web and email addresses, initials and more. To distinguish the real ends of phrases or sentences these tokens should be interpreted.

4.3 Bigram document representation

Bigram is a word pair and we always assume unordered 2-gram introduced in [9]. From this viewpoint, the words A and B will always generate bigram AB, regardless of their sequence in the sentence.

The total amount of bigram matches is enormous and without some prefiltering of documents the system using only bigrams will face huge performance problems. On the other hand, this search is the most tolerant of all possibilities of text modifications. The benefits of using bigrams proves Figure 4.

Operation	Sentence diagram	1-1 bigram	2-1 bigram	3-2 bigram	4-2 bigram	5-3 bigram	3-gram	4-gram	5-gram
remove one word	ABDEFGHI | XYZ	77,8	76,5	80,8	81,8	**83,3**	66,7	62,5	57,1
remove two words from different locations	ABDEFHI | XYZ	55,6	52,9	57,7	60,6	**61,9**	44,4	25,0	14,3
remove two following words	ABEFGHI | XYZ	**66,7**	58,8	61,5	63,6	**66,7**	55,6	50,0	42,9
swapping the order of two words	ACBDEFGHI | XYZ	77,8	94,1	96,2	97,0	**97,6**	88,9	87,5	85,7
swapping the order of two pairs of words	AEFDBCGHI | XYZ	55,6	64,7	69,2	78,8	**85,7**	55,6	37,5	42,9
swapping the order of half of the sentence	EFGHIABCD | XYZ	**77,8**	70,6	57,7	48,5	59,5	44,4	12,5	0,0
add a single word	ABMCDEFGHI | XYZ	88,9	88,2	92,3	93,9	**95,2**	77,8	75,0	71,4
add two subsequent words	ABMNCDEFGHI | XYZ	88,9	82,4	84,6	87,9	**90,5**	77,8	75,0	71,4
add two words in two different locations	ABNCDEFMGHI | XYZ	77,8	76,5	80,8	81,8	**85,7**	55,6	50,0	28,6
swap the order of sentences	ABCDEFGHI | RST	**88,9**	88,2	80,8	78,8	71,4	66,7	62,5	57,1
substitute of one word in a sentence	AMCDEFGHI | XYZ	77,8	82,4	84,6	84,8	**85,7**	77,8	75,0	71,4
substitute two subsequent words	AMNDEFGHI | XYZ	66,7	64,7	69,2	69,7	**71,4**	66,7	62,5	57,1

Figure 4: Bigram representation and the resistance to changes in the text

Bigram may be defined as two adjacent tokens (Figure 5). Then, five token (A, B, C, D E) sentence is represented the following set of word pairs: AB, BC, CD, DE.

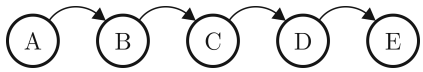

Figure 5: Bigrams based only on adjacent tokens.

For bigrams is not strictly necessary to control only the direct neighbors. It is also possible to check words in the sentence which are distant from each other over a few words. Figure 5 shows the pairing of words in a sentence to create bigrams at a distance of three. In to the further process of context extraction enter following bigrams: AB, AC, AD, BC, BD, BE, CD, CE, DE. It can be seen that the number bigrams increased from four to nine, which is more than double the number of classic bigrams (Figure 6).

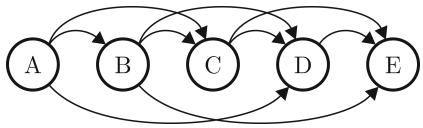

Figure 6: Bigrams overlooking three tokens.

4.3.1 Bigrams sentence overlap

To fit the sentence into the context of the text it is possible to choose another parameter - the length of overlook to the next sentence. This view may be shorter than the standard of overlook within the sentence.

Methods working with bigrams are specially labeled according to the value of both perspectives. Schematic entry is the X-Y method, where the value of X marks the length of the overlook into a sentence and the value of Y marks the length of the overlook into the next sentence. Thus, Figure 5 shows a 1-0 system, Figure 6 shows a 3-0 system.

4.3.2 Some bigram characteristics

As well as the number of unique words in the dictionary grows, it is expected that the number of unique bigrams also increase with the number of processed documents. The number of unique bigrams however grow significantly faster than the size of vocabulary and the slowdown in the growth will occur later - the choice of bigram generation method will have significant influence. Figure 6 indicates the growth in the number of unique bigrams depending on the number of processed documents from the corpus.

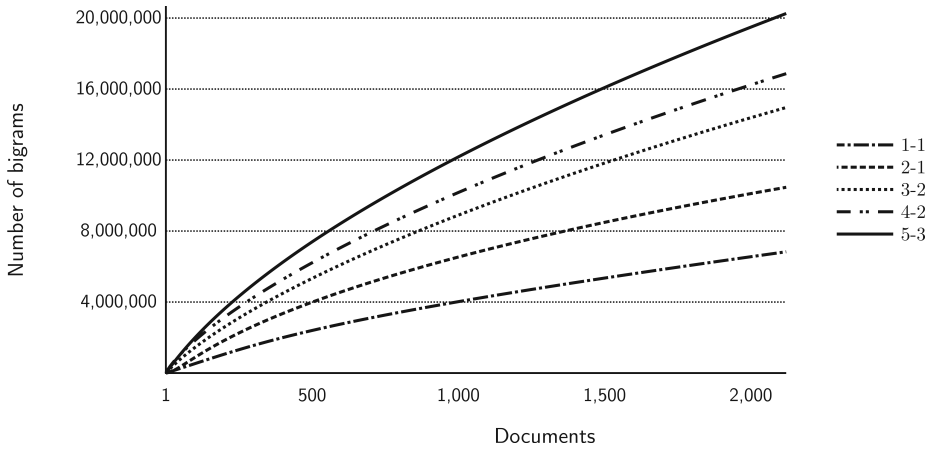

Figure 7: Number of unique bigrams.

The extension of overlook leads to an increase in the number of unique bigrams. We can also notice a trend of slowing down the growth of new bigrams when processing large quantities of documents. This decline in growth is different for each method of bigrams calculation, but the percentage drop of the number of unique bigrams growth over the first 500 processed documents is very similar for all methods (see Figure 8).

	Growth rate				
Number of documents	1-1	2-1	3-2	4-2	5-3
500	100 %	100 %	100 %	100 %	100 %
1,000	65 %	64 %	63 %	63 %	63 %
1,500	57 %	55 %	54 %	54 %	53 %
2,000	45 %	44 %	43 %	43 %	43 %

Figure 8: The growth rate of the number of unique bigrams depending on the method and the number of documents.

For the 1-1 bigrams we can estimate the growth of the number of unique bigrams as approximately 20 % for 4,000 processed documents. This means 500,000 new unique bigrams that are obtained by analysis of the documents in the order of 3,501 to 4,000. The growth will drop to 4 % after processing 7,000 documents.

5 Conclusion

In this paper we have discussed possible approaches to some aspects of the process of the document context extraction. The main attention was paid to

methods that work with text identifiers. The size of blocks of few words, in other words n-gram, proved to be a good strategy for searching in large text corpora with a reasonable compromise between accuracy and detection performance. Also a discussion about the pros and cons of using different bigrams was provided.

Acknowledgement

This work has been supported by funds from grant GACR 403/12/2175.

References

[1] Pribil, J., Kincl, T., Bina, V., Novak, M. (2011). Language Independent System for Document Context Extraction. In *Proceedings of the World Congress on Engineering and Computer Science 2011*, vol. 1, WCECS 2011, October 19-21, 2011, San Francisco, USA. pp. 51-55.

[2] Pribil, J., Leseticky, O., Bina, V. *Plagiarism at Universities - How to Fight it? The Case of the Czech Republic.* UFV Research Review: A Special Topics Journal, 2012, ro. 2, . 4, pp. 37-48.

[3] Bina V., Pribil, J. (2012). Preliminary steps in uninformed approach to the extraction of context in multilingual corpora. In *Proceedings of the 9th Workshop on Uncertainty Processing*, Marianske Lazne, Czech Republic. pp 13-21.

[4] Bina V., Pribil, J. (2012). Preliminary steps in uninformed approach to the extraction of context in multilingual corpora. In *Proceedings of the 9th Workshop on Uncertainty Processing*, Marianske Lazne, Czech Republic. pp 13-21.

[5] Konchady, Manu. Text Mining Application Programing. Boston: Charles River Media, 2006.

[6] Miner, G. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications.* 1 st en. Elsevier Science & Technology Books, 2012.

[7] Czech stop words. In: *Ranks.NL* [online]. 2013 Available online: http://www.ranks.nl/stopwords/czech.html

[8] Chmelar, P., Hellebrand, D., Hrusecky, M., Bartik, V.: Nalezeni slovnich korenu v cestine, In: *Znalosti 2011: Sbornik prispevku 10. rocniku konference*, Stara Lesna, SK, VSB TU, 2011, pp. 66-77.

[9] Pribil, J., Kalinova, H. Types of text plagiarism detectable by unoriented n-grams. Otaru 03.11.2010 - 05.11.2011. In: *13th Czech-Japan Seminar on Data Analysis and Decision Making in Service Sciences.* Otaru : University Hall, 2010, pp. 211-218.

Index of Authors

Barták, Roman, 7
Bína, Vladislav, 65

Cimler, Richard, 99
Čermák, Petr, 223, 245

Daniel, Milan, 187

Gavalec, Martin, 91

Holčapek, Michal, 29, 215
Hurtík, Petr, 205
Hwang, Hark, 107

Inuiguchi, Masahiro, 83
Ishii, Hiroaki, 1
Ivánek, Jiří, 145

Jiroušek, Radim, 73

Ko, Young Dae, 107
Korviny, Petr, 171
Kusunoki, Yoshifumi, 129, 197

Lín, Václav, 239

Miyata, Mayuko, 231
Morawiec, Martin, 245

Namekata, Tsuneyuki, 163
Namekata, Yoko, 163
Novák, David, 223
Novák, Vilém, 43, 215
Nowaková, Jana, 137

Přibil, Jiří, 253
Perfilieva, Irina, 205, 215
Pokorný, Miroslav, 137

Ramík, Jaroslav, 171

Song, Byung Duk, 107
Sung, Shao-Chin, 231
Štěpnička, Martin, 29

Tanino, Tetsuzo, 129, 197
Tomášková, Hana, 91

Vejnarová, Jiřina, 119
Vlach, Milan, 153
Vomlel, Jiří, 19

Wang, Chenxue, 51
Watada, Junzo, 51

Zhou, Neng-Fa, 7