Advances in Intelligent Systems and Computing 462

Piotr Kulczycki
László T. Kóczy
Radko Mesiar
Janusz Kacprzyk *Editors*

Information Technology and Computational Physics

Springer

Advances in Intelligent Systems and Computing

Volume 462

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

More information about this series at http://www.springer.com/series/11156

Piotr Kulczycki · László T. Kóczy
Radko Mesiar · Janusz Kacprzyk
Editors

Information Technology and Computational Physics

Springer

Editors
Piotr Kulczycki
Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
Kraków
Poland

and

Systems Research Institute
Polish Academy of Sciences
Warsaw
Poland

László T. Kóczy
Department of Information Technology
Széchenyi István University
Győr
Hungary

and

Faculty of Electrical Engineering
and Informatics
Budapest University of Technology
and Economics
Budapest
Hungary

Radko Mesiar
Faculty of Civil Engineering
Slovak University of Technology
Bratislava
Slovakia

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warsaw
Poland

Associate Editors
Piotr A. Kowalski
Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
Kraków
Poland

and

Systems Research Institute
Polish Academy of Sciences
Warsaw
Poland

Szymon Łukasik
Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
Kraków
Poland

and

Systems Research Institute
Polish Academy of Sciences
Warsaw
Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-44259-4 ISBN 978-3-319-44260-0 (eBook)
DOI 10.1007/978-3-319-44260-0

Library of Congress Control Number: 2016961343

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Information Technology (IT)—highly sophisticated information processing—forms the intellectual basis at the forefront of the current, third scientific–technical revolution. The first, generally placed at the turn of the eighteenth and nineteenth centuries, created the foundations for replacing muscle power with that of steam, and—as a consequence—manual production started to be displaced by industrial mass production. The second, which took place at the turn of the nineteenth and twentieth centuries, was brought about by the large number of groundbreaking concepts and inventions occurring, thanks to the new energy carrier—electricity. This third scientific–technical revolution, rooted in the fifties, is of a different character from the previous ones in that it is nonmaterial. In essence it constitutes the collection, treatment, and transmission of data, and so the subject of research and operation is here abstract, unreal objects. The dominant discipline for innovative development and progress became Information Technology as it is widely understood.

If therefore the crux of the changes does not consist of creating new machines or devices, but of the radical transformation of preexisting essence and character, then the spectrum of research and practical interests is unusually broad, even unlimited in the framework of contemporary science and applicational fields. Techniques used can be divided into different, partially intersecting groups. Generally, the first group consists of disciplines, which actually originated within the context of and for the needs of IT: computational intelligence and data analysis (especially exploratory). The second is the application of new technologies for tasks appropriate to distinct practical problems; a typical example of this is image processing. The third comprises support for basic sciences, dedicated to describing the world's reality, mainly physics and mathematics. This subject of this edited book has a similar division; each of its parts represents one of these groups.

The opening part concerns Intelligent Computing and Data Analysis. The first (Zadrożny, Kacprzyk, Gajewski) and second (Kulczycki, Kowalski) chapters deal with the classification of text and interval *data*, respectively. In turn, fuzzy logic was used in the third chapter (Pósfai, Magyar, Kóczy) to synthesize a recommender system for social networks, and in the fourth (Nicolau, Andrei) for predictive

diagnosis via sophisticated clustering. Finally, the last chapter (Hudec) investigates quality measure of data summaries, related to outliers.

The second part is devoted to Information Systems and Image Processing as they are broadly understood. The opening sixth chapter (Rolik, Halushko, Kolesnik) concerns management of service of corporate IT infrastructures. Next, the authors of the seventh chapter (López de Luise, Bel, Mansilla, Lobatos, Blanc, Malca la Rosa) use statistical and heuristic tools to investigate the prediction of risk associated with traffic accidents. The subject of considerations of the next chapter (Andrei, Nicolau) is the task of robustness of wireless communication systems. The subject of the ninth chapter (Krivá, Handlovičová) is evaluation of gradient norms on a deformed quadtree grid. In the tenth chapter (Grigorescu, Macesanu) a robust facial features detector is considered. And finally, the last in this part, the eleventh chapter (Świebocka-Więk), is devoted to aspects of medical diagnosis based on analysis of tomographic images.

Finally, the subject of the third part constitutes tasks of basic sciences—computational physics and applied mathematics. In the twelfth chapter (Kozik, Łużny) of this edited book, genetic algorithms were used to determine the structure of crystals. The next text (Poliński, Stęgowski) considers the computational fluid dynamics method to model flow behavior in a photoreactor. In turn, the authors of the fourteenth chapter (Palutkiewicz, Wołoszyn, Spisak) deal with transport characteristics of semiconductor nanowire transistors. Closing this book, the material of the fifteenth chapter (Mesiar, Kolesárová) investigates new methods for constructing bivariate copulas which provide a mathematical apparatus to aggregate available knowledge.

The subject choice of particular parts of this edited book was determined through discussions and reflection arising during the *Congress on Information Technology, Computational and Experimental Physics* (CITCEP 2015), 18–20 December 2015, Kraków, Poland. The authors of selected papers were invited to present extended descriptions of their research as part of this post-conference publication.

We would hereby like to express our heartfelt thanks to the technical associate editors of this book, Dr. Piotr A. Kowalski and Dr. Szymon Łukasik, as well as the co-organizers of the above conference, Dr. Joanna Świebocka-Więk and Artur Nowosielski, as well as all participants of this interesting interdisciplinary event.

Kraków/Warsaw, Poland — Piotr Kulczycki
Győr/Budapest, Hungary — László T. Kóczy
Bratislava, Slovakia — Radko Mesiar
Warsaw, Poland — Janusz Kacprzyk
April 2016

Contents

Part I
Intelligent Computing and Data Analysis

The Problem of First Story Detection in Multiaspect Text Categorization

Sławomir Zadrożny, Janusz Kacprzyk and Marek Gajewski

Abstract The new concept of *multiaspect text categorization* (MTC), recently introduced in a series of our papers, may be viewed as a combination of the classic and well-known text categorization (TC) and some kind of sequential data classification. The first aspect of the problem, i.e., the assignment of a document to a *category*, may be addressed using one of the well-known techniques such as, e.g., the *k*-nearest neighbors method. The second aspect is, however, less standard and boils down to the assignment of a document to one of the sequences, called *cases*, of documents maintained within a category. Cases cannot be treated in the same way as categories as, first, they contain an ordered—by the time of arrival—set of documents, and second, they are usually represented in a training dataset by a (relatively) small number of documents. Moreover, it is assumed that new cases can emerge during the document collection lifetime. Hence, the assignment of a document to a case is a challenging task by itself, and then the deciding if a document starts a new case is even more difficult. In this paper, we deal with the latter problem, discussing it in the broader perspective of sequential data mining and comparing a number of approaches to solve it.

1 Introduction

Text categorization (TC) [1] is among the most important tasks defined in the framework of the class (textual) information retrieval (IR) [2]. It plays an important role in the automatic handling of large document collections as it, basically, boils down

S. Zadrożny (✉) · J. Kacprzyk · M. Gajewski
Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01-447 Warszawa, Poland
e-mail: Slawomir.Zadrozny@ibspan.waw.pl

J. Kacprzyk
e-mail: Janusz.Kacprzyk@ibspan.waw.pl

M. Gajewski
e-mail: Marek.Gajewski@ibspan.waw.pl

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_1

to the assignment of documents to some predefined categories. There are many different variants of this basic task which may be distinguished depending, e.g., on following aspects (cf., e.g., [1]). Documents to be classified may be available all at once (off-line categorization) or they are to be classified one by one when they appear on the input of the system (on-line categorization). The structure of categories may be flat (one level) or there may be a hierarchy of them (hierarchical text categorization). Each document may be assigned to at most one category (single-label categorization) or to many categories (multi-label categorization). Analogously to the general task of classification, a special case is when there are just two categories among which usually one is of interest (single-class categorization)—this may be confronted with the general case when the number of categories is a larger number (multi-class categorization), of course of a reasonable value, for comprehensiveness. The very nature of categories also makes a difference: a canonical variant of text categorization refers to the thematic (topical) categories while in other cases there may be essentially different criteria deciding on the grouping of documents into categories (e.g., the type of a business document in a company meant as one of the memo, advertisement brochure, meeting announcement etc.) Finally, the text categorization may be carried out: manually by an expert or a group of experts (an option viable only for small document collections); automatically, using some hand crafted rules (in the vein of knowledge engineering) or the whole process can be fully automated, i.e., some machine learning techniques can be used to automatically derive classifiers based on the set of training data.

In a number of papers [3–9] we have introduced the new concept of the *multiaspect text categorization* (MTC) and some approaches to solve it. The MTC problem may be seen as a special case of the general text categorization problem. Referring to the aspects of the TC mentioned earlier, it may be characterized as a TC problem which is on-line, hierarchical, single-label, multi-class, and with mixed types of categories, and for which a fully automatic solution is sought. The formulation of the MTC is motivated by a practical problem of managing collections of documents dealt with within an organization, notably a public institution in Poland, which has to follow some formal legal regulations. Namely, on the first level, the documents have to be arranged according to a hierarchy of prespecified thematic/topical categories. On the second level, each document has to be assigned within its category to a sequence of documents, referred to as a *case*. The cases will usually correspond to some business processes carried out by a company. For example, the process of the purchase of some accessories will be usually initialized with a formal request from a department in need of them, which may be followed by a call for tender, in turn followed by the offers from prospective suppliers, etc. Thus, besides a hierarchy of thematic categories we have to deal with a different kind of hierarchy relating a thematic category and cases belonging to this category.

The classification of documents on the first level alone may be directly dealt with using techniques known in the classic *text categorization* (TC) [1]. The second-level classification is, however, more difficult. The cases may be considered as categories, similarly to the situation at the first level, but the problem is implied, basically, by a limited number of training documents representing such a category and, moreover,

by the fact that only a part of such categories (cases) is known in advance. Any incoming document may turn out to be initiating a new case. Thus, an important part of a successful solution to the MTC problem is the detection if this takes place. In this paper we deal exactly with this problem.

In the next section, we remind the formal statement of the multiaspect text categorization problem. Then, we review the related work. Next, we propose the use of a number of techniques to solve the problem of the first story detection and, finally, we report the results of the computational experiments aimed at comparing these techniques.

2 The MTC Problem

The multiaspect text categorization (MTC) problem may be illustrated with an example of a public administration institution dealing with various affairs. One of the aspects of its activity is a proper organization of documents concerning particular efforts and yielded in the course of a business process carried out by this institution. An example of such a business process may be arranging a public tender for the purchase of office equipment. The related documents include the announcement of the tender, offers incoming from companies responding to the tender and offering the equipment, minutes of meetings of a committee responsible for carrying out the process, etc. Usually, there is specified a list of *categories* of affairs which are dealt with. Sometimes these categories are arranged in a hierarchy (in this paper we assume a flat, one level, list of categories but an interested reader may consult our another paper [10] as well as the literature therein on the hierarchical text categorization). The accomplishment of an instance of a business process will be referred to as a *case* and the institution has to store together and in a proper order all documents related to a given case.

Thus, we consider the MTC problem in the following context. There are many *on-going* cases belonging to various categories and our aim is to build an automatic system which will assist a human operator in assigning a new incoming document to a proper case, i.e., to a case which is related (possibly to a high extent) to a business process instance to which this document actually belongs. We assume that the system takes into account only the content of the document and does not use, e.g., metadata accompanying this document. Such an assumption may be more or less justified in various practical scenarios but it guarantees a broader applicability of the designed system.

What is very important is the fact that, if it is justified, a new document can initiate a new instance of a business process and thus can originate a case of which it becomes the first document. In this paper, we are interested in finding a way to automatically decide if a new incoming document really starts a new case in view of its contents and the contents of all documents stored so far, and their organization in categories and cases.

Let us now formally describe the above characterized problem. We assume that a collection D of documents is given:

$$D = \{d_1, \ldots, d_n\} \tag{1}$$

These documents are assigned to some predefined *categories* from the set C:

$$C = \{c_1, \ldots, c_m\} \tag{2}$$

in such a way that each document $d \in D$ is assigned to exactly one category $c \in C$. The documents are further arranged within each category into sequences $\sigma \in \Sigma$, rank ordered with respect to the time of arrival, which are referred to as *cases*:

$$\sigma_k = < d_{k_1}, \ldots, d_{k_l} > \tag{3}$$

$$\Sigma = \{\sigma_1, \ldots, \sigma_p\} \tag{4}$$

Again, each document $d \in D$ belongs to exactly one case $\sigma \in \Sigma$.

The goal is to build a system, using D as the training collection, which will support a human user in deciding how to add a new incoming document d^* to the collection D. Thus, a document d^* has to be assigned to a category $c \in C$ and to a case $\sigma \in \Sigma$ within this category.

Various strategies may be adopted to obtain a proper classification. A two-level approach may be applied in which, first, a category is assigned and then the case. The motivation is that the classification to a category may be relatively easier and the classic text categorization techniques should be effective and efficient enough to do this. Then, when a category is already selected, one can expect that it should be easier to assign the document to a proper case within this category. The reason is that local characteristic features of cases in a given category may be employed and, moreover, the number of candidate cases will be much lower in such a scenario; cf., e.g., our papers [6, 8, 9] for examples of such an approach in the framework of the MTC or the paper by Yang et al. [11] for a related approach in another context. It is worth noting that it is also possible to skip the assignment of a category to a document d^* and to focus on the choice of a proper case as such a choice directly implies also a category c to which the case σ belongs. However, this way the extra information on the category of the document d^* is ignored when choosing the case, provided that the category assignment is successful. Finally, the decisions concerning the assignment of d^* to a category and to a case may be combined with the hope that both decisions will mutually support each other. An example of such an approach is given in our paper [5].

To summarize, the MTC problem may be characterized as a text categorization problem with two levels of broadly defined categories. At the upper level, these may be assumed to be typical prespecified thematic categories, represented in the training collection d with a sufficiently large number of examples. At the lower level, these are cases the number of which is dynamically changing and which may be poorly, or even not at all, represented in the training collection of the documents, D.

3 Related Works

The task considered in this paper refers mainly to the context of the multiaspect text categorization problem (MTC) recently proposed in our earlier work; cf., e.g., [4], and formally presented in the previous section. A similar problem known in literature is the *Topic Detection and Tracking* (TDT) [12].

The TDT was a part of the DARPA Translingual Information Detection, Extraction, and Summarization (TIDES) program, closely related to the well-known Text REtrieval Conferences (TREC). Research on TDT started in 1997 [13] and was followed by regular workshops during the next 7 years. The topic of detection and tracking is considered in the context of processing of a stream of news coming from various sources and concerning some events/topics. It is assumed that events evolve over time and some new news stories related to them are incoming. However, new events are happening which are also represented in the stream of incoming news stories. The basic task is here to group together news stories concerning the same events and describing their development over time, various aspects etc.

An individual piece of news in TDT is referred to as a *story* and corresponds to a document in our new MTC problem definition. Stories in TDT describe *events* and some major events together with interrelated minor events are referred to as *topics* and correspond to both categories and cases in MTC with an emphasis on the latter. Topics, similarly to cases, are not predefined and new topics have to be *detected* in the stream of stories and then *tracked*, i.e., all subsequent stories dealing with the same major event have to be recognized and classified to a topic detected earlier. A number of specific tasks are distinguished within the TDT. From our perspective the most important are *topic detection* and *first story detection*. The former may be identified with the classification of documents to the cases in our MTC: starting with a set of groups of stories forming particular topics—which may be empty in the beginning—a new incoming document has to be assigned to one of these topics or to form a new topic. The latter task is, in fact, a part of the former and consists in recognizing if a document belongs to one of the earlier detected topics or is the first story of a new topic. It is however distinguished due to its importance and difficulty [14].

The main differences between the TDT and MTC may be briefly stated as:

1. categories and cases are considered in the MTC as opposed to topics only in the TDT,
2. cases are sequences of documents while topics are basically just sets of stories; even if stories are timestamped, their possible temporal type relations are not analyzed and the timestamps are only used to discount the information related to older stories,
3. there is a different practical inspiration for the TDT and MTC which implies further differences in assumptions adopted in both cases (besides the two aspects mentioned above).

For a further analysis of relations of the multiaspect categorization problem and the topic detection and tracking problem the reader is referred to our earlier paper [7].

In the current paper, we are concerned with a counterpart of the first story detection (FSD) task present in our multiaspect text categorization problem. Thus, it is worthwhile to briefly review a few techniques that have been proposed to solve the FSD in the framework of the TDT.

Approaches to first story detection are often based on the similarity of the incoming document d^* with respect to all or a part of the documents collected earlier. A threshold is assumed and if the similarity of a document d^* with respect to, e.g.:

- any of the earlier collected documents, or
- any of k recently collected documents, or
- any centroid of documents belonging to particular topics earlier recognized,

exceeds the assumed threshold, then document d^* is deemed to be related to some earlier seen topic and is assigned to it. Otherwise it is treated as starting a new topic and becomes its first story.

Another idea consists in the monitoring of term distribution over time and a new topic is recognized in case an abrupt change in this distribution is detected for a given story. Another strategy in this vein refers to the solution of another TDT task, namely that of *topic tracking*. This task consists in deciding if an incoming story d^* belongs to a given topic represented by a (small) number of stories. Assuming that the tool for topics tracking is available it may be immediately used to solve also first story detection problem. Namely, if an incoming document d^* is not indicated as belonging to any tracked topic then it has to be the first story of a new topic.

Yang et al. [11] propose a relatively simple, and effective and efficient approach to the FSD problem which is, moreover, very relevant to our MCT problem and the detection of the first document of a new case. Namely, they group the topics into a higher level of categories (originally, in [11], topics are referred to as events and categories as topics but we will keep here the terminology consistent with the previously introduced one for the TDT problem). An incoming story is first classified to a category and only then to a topic within that category. If the latter classification fails, i.e., the similarity of a new document to its closest neighbor within given category is lower than some threshold value, then such a story is qualified as a first story of a new topic. This resembles very much our two-level approach to assigning a document to a case within an earlier chosen category; cf., e.g., our [6]. The motivation for the Yang et al. [11] approach is that it makes it possible to use different features (keywords) while classifying a story to a category and to a topic. Thus, the features shared by first stories of topics with all other stories belonging to a given category may be taken into account while classifying a document to a category but they may be ignored (treated as stopwords) while classifying this document to a topic. Thanks to that, an actual first story may be properly recognized as such because it may turn out not to be similar to other stories in a given category even if it shares a number of features with them. The important points of this approach are the following:

- different representation of stories for their classification at the levels of categories and topics via a separate feature selection at each level;
- enhancing the basic vector space model representation of stories with named entities.

For the top-level classification of the stories Yang et al. [11] use the Rocchio-style classifier, popular in the framework of the TDT [15]. The recognition of the first story at the lower level of classification is based on the similarity to a nearest neighbor, as mentioned earlier.

The problem of first story detection may be considered in a broader framework of the *novelty detection* problem [16]. The concept of novelty detection, in general, refers to recognizing that an object under consideration belongs to a class which has not yet been represented in a training dataset. In the MTC context we face this problem in an even more intense form as if we treat cases as classes then:

- evidently, we should expect incoming documents belonging to new cases, i.e., new classes, and moreover,
- even for cases (classes) represented in the training dataset, very often this representation will be very limited, i.e., a case will be often 1–2 document long.

The novelty detection approaches address directly the first of the above problems but usually take into account also the sparsity of the training data in which the examples of novel data are scarce.

The statistical approaches to novelty detection often consider the problem as a binary classification task aiming at distinguishing novel data from the rest, "normal" data [16, 17]. Popular solutions are based on estimating the probability density for the data belonging to the "normal" class and deciding on the novelty of incoming data objects if they fall in regions of low density. Examples of the approaches in this vein, specifically meant for the novelty detection in the textual information processing context, include those presented in the papers by Hofmann et al. or Hansen et al. [18, 19]. Recent surveys on novelty detection are papers by de Faria et al. [17, 20].

Some other related concepts discussed in the literature are also relevant for the FSD problem definition and solution. These include anomaly detection and rare events mining; cf., e.g., [21]. This is due to the fact that if a standard binary classification approach is adopted to solve the FSD problem, then one class, of the first stories, will be usually an order of magnitude smaller than another class of non-first stories.

Our task may also be studied from a broader perspective of applying machine learning methods to the sequential data [22]. Namely, the main idea of an intelligent approach to classifying a document to a proper sequence calls for understanding the mechanism behind the forming of document sequences within a given collection or a part of it (category). Knowing this mechanism, we can decide if a document under consideration fits an existing sequence or rather should be treated as starting a new sequence. In our earlier works [4, 23], we propose to employ, first of all, tools and technique of the hidden Markov models (HMMs) to get such an understanding of sequences of documents within categories. Hidden states may then be identified with stages of a business process which produce a given sequence of documents (a case). If these stages may be explicitly identified, then a broader repertoire of models/techniques for sequential data processing may be considered as helpful [22] such as, e.g., the conditional random fields (CRF).

4 A Direct Approach for Solving the FSD as a Classification Problem

In our approach reported in this paper we adopt an approach to the FSD task solving differently from most of those proposed in the framework of the TDT. Namely, the latter approaches employ a topic tracking technique and declare a story as a first story when it does not fit any of the topics recognized so far; cf. Sect. 3. This observation applies also to the approach proposed by Yang et al. [11], even if it introduces a two-level classification schema. Our approach is an attempt to solve the FSD problem using directly a binary classification. Thus, we start with a collection of training documents which are organized in cases (sequences) according to the definition of the MTC problem. The first documents of all cases present in the collection are positive examples while the remaining documents form the set of negative examples. Then, we employ some variants of a number of well-known machine learning algorithms and compare their effectiveness.

The algorithms we started with are the following:

1. a variant of the *k*-nearest neighbors algorithm,
2. the random forests,
3. logistic regression,
4. linear discriminant analysis,
5. an approach based on modeling the probability density of selected keywords in positive and negative examples,
6. support vector machine.

We have also tested a feature selection technique as well as two schemes of training the classifiers:

- locally, a separate individual classifier for each category,
- globally, one classifier for the whole collection.

In the tests we carried out, the feature selection techniques did not improve the results for most of the considered algorithms. On the other hand, most of the algorithms produced much better results for the global variant mentioned above, i.e., when one classifier is constructed for recognizing first stories based on the whole training collection. Thus, in the next section we report the results of our experiments only for the case where stories are represented using all considered features (keywords) and for the global approach. Now, we will briefly describe the algorithms and justify their use in our experiments.

The *k*-nearest neighbors technique (*k*-nn)-based algorithms proved to be effective and efficient in our earlier approaches to the MTC problem. In [6, 9] we use a variant of *k*-nn to assign a category and a case to a document. It is also widely used by the TDT community (cf., e.g., [24]). We use it here in the version proposed by Yang et al. [24] (cf. also [9]) which is referred to as kNN.avg2. It is based on a function defined as follows:

$$r(d^*, k_p, k_n, D) = \frac{1}{|U_{k_p}|} \sum_{d \in U_{k_p}} sim(d^*, d) - \frac{1}{|V_{k_n}|} \sum_{d \in V_{k_n}} sim(d^*, d) \quad (5)$$

The value of this function is computed for the document d^* to be checked for being a first story with respect to the training dataset D. There are two parameters k_p and k_n which determine the cardinality of the sets U_{k_p} and V_{k_n}, respectively. The former set comprises the k_p positive examples (i.e., first stories in the training data set D) most similar to the document d^* while the latter set comprises k_n negative examples (i.e., non-first stories in the training data set D) most similar to d^*. In our experiments we use the kNN.avg2 algorithm with the parameters set as follows: $k_p = k_n = 1$. The similarity is computed using a function denoted by *sim* which is identified with the classic cosine measure in [24] and with the complement to the Euclidean distance in [9] (vectors representing documents are assumed to be normalized and, thus, there is a well-defined maximal possible Euclidean distance between two documents). Thus, the value of the function r for a document d^* is its average similarity to k_p most similar first stories reduced by its average similarity to the k_n most similar non-first stories. If there are less than k_p positive documents in D then all positive documents in D are employed. The same applies to the negative documents (even if this can rarely happen).

The document d^* is recognized as the first story if:

$$r(d^*, k_p, k_n, D) > 0$$

and as the non-first story otherwise, for the chosen values of the parameters k_p and k_n. As compared to the standard k-nn technique, the kNN.avg2 version is more suitable for the first story detection problem solving as it addresses the problem of imbalance between the positive and negative classes (usually there will be much less first stories than non-first stories in the training data set) using a fixed number of the nearest positive and negative examples. At the same time it also takes into account how similar the nearest examples actually are.

The algorithm of random forests [25] is used in the experiments in the version implemented as the `randomForest` function in the package `randomForest` using standard parameters. In particular, the number of trees to grow is set to 500. The model is constructed to distinguish first stories using all keywords used to represent the documents in the collection. This is one of the algorithms deemed to be highly effective and efficient in the machine learning community. It has a sophisticated built in mechanism for feature selection which should help recognize the first stories which, as argued earlier, are by definition very similar to other documents in a given category but are expected to be less similar with respect to some subset of specific keywords.

The logistic regression algorithm [26] is used in the experiments using the `glm` standard function of the R environment. The binomial distribution over the positive (first stories) and negative (non-first stories) classes is modeled via the logit link function using again the weights of all keywords as independent variables in the

linear regression analysis. This algorithm also belongs to the most popular discriminative classification techniques [27].

The fourth algorithm employed is the one based on linear discriminant analysis [28]. One of the classic techniques which, basically, requires class conditional normal distributions. In our case, the multidimensional distributions of the documents characterized by keywords weights are far from normal within the classes of both the first stories and non-first stories due to, e.g., a high sparsity of the document-term matrices. However, this technique is known to be robust and in our computational experiments it has also proved to be good.

The fifth algorithm used is a simple attempt to apply an aggressive dimension reduction technique combined with a straightforward probabilistic approach which boils down to the naive Bayes approach with the kernel density estimation [28]. Namely, first a number of keywords $t \in T$ with the highest mean in the representations of the first stories present in a training dataset are selected. Then, those whose mean is significantly higher than in the non-first stories are preserved and form a set $T_s \subset T$. The t-test is used to assess the significance with p-value equal 0.05. Their standard deviation in the first stories is also recorded. Then, two probability density functions are constructed using a kernel-based method [29], for each of these keywords: in the first stories and in the non-first stories of the training dataset. Then, the following function is used to discriminate first stories from non-first stories:

$$g(d) = \log(\frac{P(fs|d)}{P(nfs|d)}) = \\ = \log(\frac{P(fs)}{1 - P(nfs)}) + \sum_{t \in T_s} (\log(f^t_{fs}(d[t])) - \log(f^t_{nfs}(d[t])) \quad (6)$$

where $d[t]$ denotes the weight of a keyword t in the representation of a document d, f^t_{fs} and f^t_{nfs} are approximated conditional probability density functions of the particular keywords $t \in T_s$ in the first stories and non-first stories of the training dataset, respectively, and $P(fs)$, $P(nfs)$ are a priori probabilities of a document being a first story and non-first story, respectively. The latter a priori probabilities are estimated on the training data set. However, in our experiments reported in Sect. 5 the assumption of the a priori probability equal 0.5 for both classes produced much better results and, thus, we adopted this strategy. The function $g(d)$ corresponds therefore to the logarithm of the odds of a document to be the first story, assuming the conditional independence of the keywords in documents of both classes. Thus, if $g(d) > 0$, then the document d is classified as the first story and, otherwise, as the non-first story.

A variant of formula (6) is also employed:

$$g'(d) = \sum_{t \in T_s} \sigma'(t)(\log(f^t_{fs}(d[t])) - \log(f^t_{nfs}(d[t])) \quad (7)$$

where $\sigma'(t) = 1 - \frac{\sigma(t)}{\max_s \sigma(s)}$, $\sigma(t)$ denotes the standard deviation of the weights of keywords in the first stories of the training dataset. This variant is meant to differentiate

the influence of particular keywords $t \in T_s$ on the classification decision, i.e., the keywords with a higher standard deviation have a lower influence. In the formula (7) we assume the a priori probabilities equal 0.5, as mentioned earlier, and thus here we dropped the first component of the formula (6).

The purpose of this approach is a direct selection of keywords that are relatively highly frequent in the first stories and less frequent in the non-first stories. In our experiments, it was most often possible to spot such keywords. In case it was not possible, all keywords were taken into account. The approach is similar to the linear discriminant analysis (LDA) but explicitly drops the assumption on the normality of distributions required by the LDA.

Finally, the sixth algorithm employed is a powerful and very popular support vector machine (SVM) method [30]. We experimented with various kernels and other parameters and finally decided to use the RBF kernel. The SVM technique perfectly fits in its original form the binary problem of distinguishing first stories from non-first stories.

5 Computational Experiments

We have tested the approaches presented in the previous section using a data collection which we have used also in our previous work [5, 6, 8, 9]. The starting point is the set of articles on computational linguistics available in the framework of the ACL Anthology Reference Corpus (ACL ARC) [31]. We use a subset of 113 papers. The papers are originally partitioned into sections and the idea is to treat each article as a case and its sections as documents of such a case. What is missing is the grouping of documents/cases into categories. Thus, to do this, we first use the k-means algorithm to partition the set of articles into 7 clusters which play the role of categories. This number of clusters has been chosen experimentally in order to secure a reasonable number of categories and their cardinalities.

The articles and, later on, the resulting documents are represented using the vector space model (cf., e.g., [2]) and standard preprocessing techniques such as the removal of the punctuation, numbers, and multiple white spaces; stemming; changing all characters to the lower case; dropping stopwords and words shorter than 3 characters. The $tf \times IDF$ scheme is employed to compute the weights of particular keywords in documents. The obtained document-term matrix is very sparse and, thus, the keywords present in less than 10% of the papers are further removed. As a result 125 keywords are employed. The vectors representing particular documents are normalized by dividing each coordinate by the Euclidean norm of the whole vector and thus the Euclidean norm of each vector equals 1.

Finally, we obtain a collection of 113 cases comprising 1453 documents which is then split into the training and testing datasets. A number of cases are randomly chosen and a cut-off point in each of them is again randomly selected. All documents at positions starting from the cut-off point are removed from the case and form the test dataset (in the experiments reported here we used the datasets composed of only

the documents located at the cut-off points). All remaining documents from the collection serve as the training dataset. This way, if a cut-off point corresponds to the first position in a case we obtain a first story in the test data set.

All computations are carried out using the R platform [32] with the help of the packages: `tm` [33], `FNN` [34], `randomForest` [35], `kernlab` [30], `MASS` [36] and our own R scripts.

We have tested the algorithms for first story detection mentioned in Sect. 4 in two configurations:

1. for each category separately, i.e., we assumed that the incoming document d^* has been first properly assigned a category and only then it is checked as a candidate for being a first story based on the training data set confined to this category; this is the case referred by Yang et al. as the *simple case* (cf Table 2 in [11];
2. for the whole collection at once, i.e., the incoming document d^* is first checked for being a first story using the whole training data set; this is referred as the *baseline* case in [11].

All algorithms have proved to give better results for the second configuration. We have expected that for different categories different keywords may be better at distinguishing between first stories and non-first stories. However, this potential cannot be exploited due to a limited number of training documents representing first stories when considered for each category separately. The similar conclusions follow from the experiments reported in [11], even if a slightly different context of the TDT is considered there. Thus, in what follows we will present the results only for the second strategy.

We have run a series of 200 experiments and their results are presented in Table 1. In each run we randomly select 56 cases (i.e., 50% of all cases) as on-going, i.e., those in which randomly a cut-off point is selected as earlier described. In order to evaluate the results obtained using particular approaches we use the F1 measure and the cost-based measure CO_{fsd} (in its normalized version) employed by Yang et al. [11]. These

Table 1 The results of 200 runs of the compared algorithms given in terms of the F1 and CO_{fsd} measures. The mean values and standard deviations of both measures are reported. Notice that for the second measure lower values indicate better results

The algorithm	F1		CO_{fsd}	
	Mean	sd	Mean	sd
k-nn	0.1705	0.1349	0.9912	0.2425
Random forests	0.2319	0.2278	0.9426	0.1890
Logistic regression	**0.4063**	0.1847	**0.6789**	0.2483
Linear discriminant analysis	**0.4427**	0.2094	**0.6571**	0.2684
Naive Bayes with kernel density estimation	0.26	0.1727	0.8735	0.2320
As above with standard deviation based keywords weighting	0.3384	0.1730	0.7356	0.2636
Support vector machine	0.3441	0.2195	0.8169	0.2407

measures are defined as follows, denoting the elements of the standard contingency table as TP, FP, TN, and FN, i.e., the true positives, false positives, true negatives, and false negatives, respectively:

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{8}$$

$$CO_{fsd} = \frac{CO_m * P_m * P_t + CO_f * P_f * P_{nt}}{\min(CO_m * P_t, CO_f * P_{nt})} \tag{9}$$

where CO_m and CO_f are costs of the *miss* and *false alarm*, respectively, i.e., the former is the cost of classifying a first story as the non-first story while the latter is the cost of classifying a non-first story as the first story; $P_m = \frac{FN}{TP+FN}$ and $P_f = \frac{FP}{TN+FP}$, i.e., are the *false negative rate* (*miss*) and the *false positive rate* (*fall-out*), respectively; P_t and $P_{nt} = 1 - P_t$ are probabilities of a first story and non-first story occurrence, respectively.

Thus, (9) expresses the expected cost of the error to be made by the first story detection system. It is normalized by the cost of the better of two trivial algorithms which would classify all stories as first stories or as non-first stories, respectively. We set $CO_m = 1.0$ and $CO_f = 0.1$ after [11], adopting the justification given there that a miss may be easier recognized as a mistake by the human operator assisted by our system. On the other hand, we set $P_T = 0.1$ for the whole collection and for each category separately as this is the average frequency of first stories therein.

The results shown in Table 1 indicate the linear discriminative analysis and logistic regression as the best algorithms in detecting first stories. Due to the Wilcoxon two-sided test the former is significantly better than the latter in terms of both the F1 and CO_{fsd} measures. The second group form the Algorithms 6 (naive Bayes with kernel density estimation and standard deviation based keywords weighting) and 7 (support vector machines). Both are not significantly different concerning their effectiveness in terms of F1 measure but the latter is better in terms of CO_{fsd} measure. The latter effect is due to a very high number of false alarms (false positives) produced by Algorithm 6 compared to Algorithm 7, even if the former produced also slightly more true positives than the latter.

Summarizing, the effectiveness of the best of the tested methods is not fully satisfactory but taking into account the well-known difficulty of the first story detection problem it is not that bad. Yang et al. [11] report better results in terms of the CO_{fsd} measure but for a different dataset and using a richer representation of documents.

The fact that we obtained the best results using linear discriminant analysis is interesting in itself. The method is based on a rather strong assumptions which are not satisfied in our experiments and is fairly simple at the same time. It could be expected that Algorithm 5 should better fit the problem in question. However, it turns out that it performs rather poorly and only if combined with an extra weighting of the keywords produces relatively as good results as Algorithm 6. It should be however noted that both Algorithms 5 and 6 operate on a highly reduced set of keywords.

6 Conclusion

We have addressed the crucial problem of first story detection (FSD) in the framework of the multiaspect text categorization (MTC), a new problem class introduced in our former papers. The adopted approach is a rather straightforward one and boils down to formulating the FSD as a classic binary classification problem. Then, we have employed a number of standard classification algorithms, proposing some extensions in case of some of them. The best results have been obtained using the standard linear discriminant analysis. It should be noted that we have used a simple vector space model based representation of the documents. Our conclusions are based on computational experiments carried out on a dataset used in our previous work. Thus, our plans for a further research comprise both the search for a more sophisticated documents representation and more extensive tests on larger and more numerous datasets.

Our approach, though relatively straightforward and intuitively appealing, is still not that popular in the context of topic detection and tracking (TDT) in which the FSD problem is quite similar to the FSD considered in the context of our MTC problem. The approaches proposed by the TDT community usually base their approaches to the FSD problem on the same algorithm which is used for the TDT. Namely, a document is classified as the first story if it does not qualify as belonging to one of the recognized topics so far. The latter decision is in turn based on checking its similarity to the previously seen documents or their representatives (e.g., centroids of the documents related to the same topic) against some threshold value. Such an approach has been failing so far in case of our algorithms for the MTC problem and that is the motivation for our search for another solution.

Acknowledgements This work is partially supported by the National Science Centre (contract no. UMO-2011/01/B/ST6/06908).

References

1. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. **34**(1), 1–47 (2002)
2. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press and Addison Wesley (1999)
3. Zadrożny, S., Kacprzyk, J., Gajewski, M., Wysocki, M.: A novel text classification problem and two approaches to its solution. In: Proceedings of the International Congress on Control and Information Processing 2013. Cracow University of Technology (2013)
4. Zadrożny, S., Kacprzyk, J., Gajewski, M., Wysocki, M.: A novel text classification problem and its solution. Tech. Trans. Autom. Control 4-AC, 7–16 (2013)
5. Zadrożny, S., Kacprzyk, J., Gajewski, M.: A novel approach to sequence-of-documents focused text categorization using the concept of a degree of fuzzy set subsethood. In: Proceedings of the Annual Conference of the North American Fuzzy Information processing Society NAFIPS'2015 and 5th World Conference on Soft Computing 2015, Redmond, WA, USA, August 17–19, 2015 (2015)

6. Zadrożny, S., Kacprzyk, J., Gajewski, M.: A new two-stage approach to the multiaspect text categorization. In: IEEE Symposium on Computational Intelligence for Human-like Intelligence, CIHLI 2015, Cape Town, South Africa, December 8–10, 2015. IEEE 2015, pp. 1484–1490 (2015)
7. Gajewski, M., Kacprzyk, J., Zadrożny, S.: Topic detection and tracking: a focused survey and a new variant. Informatyka Stosowana **2014**(1), 133–147 (2014)
8. Zadrożny, S., Kacprzyk, J., Gajewski, M.: A new approach to the multiaspect text categorization by using the support vector machines. In: De Tré, G., Grzegorzewski, P., Kacprzyk, J., Owsiński, J.W., Penczek, W., Zadrożny, S. (eds.) Challenging problems and solutions in intelligent systems, pp. 261–277. Springer International Publishing, Heidelberg (2016)
9. Zadrożny, S., Kacprzyk, J., Gajewski, M.: Multiaspect text categorization problem solving: a nearest neighbours classifier based approaches and beyond. J. Autom. Mob. Rob. Intell. Syst. **9**, 58–70 (2015)
10. Zadrożny, S., Kacprzyk, J., Gajewski, M.: A hierarchy-aware approach to the multiaspect text categorization problem. In: Proceedings of the World Conference on Soft Computing, Berkeley, CA, US (2016, in press)
11. Yang, Y., Zhang, J., Carbonell, J., Jin, C.: Topic-conditioned novelty detection. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, pp. 688–693 (2002)
12. Allan, J. (ed.) Topic Detection and Tracking: Event-based Information. Kluwer Academic Publishers (2002)
13. Allan, J., Carbonell, J., Doddington, G., Yamron, J., Yang, Y.: Topic detection and tracking pilot study: final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop (1998)
14. Allan, J., Lavrenko, V., Jin, H.: First story detection in TDT is hard. In: Proceedings of the Ninth International Conference on Information and Knowledge Management, CIKM '00, pp. 374–381. ACM, New York, NY, USA (2000)
15. Yang, Y.: An evaluation of statistical approaches to text categorization. Inf. Retriev. **1**(1–2), 69–90 (1999)
16. Markou, M., Singh, S.: Novelty detection: a review—part 1: statistical approaches. Signal Process. **83**(12), 2481–2497 (2003)
17. De Faria, E., Gonçalves, I., Gama, J., De Leon Ferreira Carvalho, A.: Evaluation of multiclass novelty detection algorithms for data streams. IEEE Trans. Knowl. Data Eng. **27**(11), 2961–2973 (2015)
18. Hofmann, D.B.T., Baker, L.D., Hofmann, T., Mccallum, A.K., Yang, Y.: A hierarchical probabilistic model for novelty detection in text (1999)
19. Hansen, L.K., Sigurdsson, S., Kolenda, T., Nielsen, F.A., Kjems, U., Larsen, J.: Modeling text with generalizable gaussian mixtures. In: Proceedings of ICASSP'2000, pp. 3494–3497. IEEE (1999)
20. De Faria, E., Gonçalves, I., De Leon Ferreira Carvalho, A., Gama, J.: Novelty detection in data streams. Artif. Intell. Rev. **45**(2), 235–269 (2016)
21. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**(3) (2009)
22. Dietterich, T.G.: Machine learning for sequential data: a review. In: Caelli, T., Amin, A., Duin, R.P.W., Kamel, M.S., de Ridder, D. (eds.) Structural, Syntactic, and Statistical Pattern Recognition, Joint IAPR International Workshops SSPR 2002 and SPR 2002, Windsor, Ontario, Canada, August 6–9, 2002, Proceedings. Lecture Notes in Computer Science, vol. 2396, pp. 15–30. Springer (2002)
23. Zadrożny, S., Kacprzyk, J., Gajewski, M.: A solution of the multiaspect text categorization problem by a hybrid HMM and LDA based technique. In: 16th International Conference Information Processing and Management of Uncertainty in Knowledge-Based Systems, Eindhoven, The Netherlands (2016, in press)
24. Yang, Y., Ault, T., Pierce, T., Lattimer, C.W.: Improving text categorization methods for event tracking. In: SIGIR, pp. 65–72 (2000)

25. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
26. McCullagh, P., Nelder, J.: Generalized Linear Models, 2nd edn. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis (1989)
27. Ng, A.Y., Jordan, M.I.: On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3–8, 2001). Vancouver, British Columbia, Canada], pp. 841–848. MIT Press (2001)
28. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA (2001)
29. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
30. Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A.: kernlab-an S4 package for kernel methods in R. J. Stat. Softw. **11**(9), 1–20 (2004)
31. Bird, S., et al.: The ACL anthology reference corpus: a reference dataset for bibliographic research in computational linguistics. In: Proceedings of Language Resources and Evaluation Conference (LREC 08), Marrakesh, Morocco, pp. 1755–1759
32. R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2014). http://www.R-project.org
33. Feinerer, I., Hornik, K., Meyer, D.: Text mining infrastructure in R. J. Stat. Softw. **25**(5), 1–54 (2008)
34. Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S.: FNN: Fast Nearest Neighbor Search Algorithms and Applications, R package version 1.1 (2013). http://CRAN.R-project.org/package=FNN
35. Liaw, A., Wiener, M.: Classification and regression by randomforest. R News, vol. 2, no. 3, pp. 18–22 (2002). http://CRAN.R-project.org/doc/Rnews/
36. Venables, W.N., Ripley, B.D.: Modern Applied Statistics with S, 4th edn. Springer, New York (2002)

A Metaheuristic for Classification of Interval Data in Changing Environments

Piotr Kulczycki and Piotr A. Kowalski

Abstract The Bayes approach is arguably the classification method most used in unspecialized applications, thanks to its robustness, simplicity, and interpretability. The main problem here is establishing proper probability values. This paper deals with adapting the above method for cases where the classified data is of interval type, with changing environments (evolving data stream, concept drift, nonstationarity). The probability values are estimated using nonparametric methods, thanks to which the procedure becomes independent of characteristics of learning subsets representing particular classes. They can also be supplemented with new, current observations, added while performing the algorithm. The investigated process also removes elements with negligible or even negative impact on accuracy of results, which increases the effectiveness of adaptation in conditions of changing reality. It is possible to differentiate the meanings of particular classes. The method allows any number of them. The particular attributes of data elements may be continuous, categorical, or both.

Keywords Data analysis · Classification · Interval data · Changing environment · Adaptation

1 Introduction

One of the main tasks of contemporary data analysis is classification [2, 5]. Suppose that we have a data set, whose particular elements are assigned labels explicitly, indicating membership of particular, previously defined subsets, constituting

P. Kulczycki (✉) · P.A. Kowalski
AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland
e-mail: kulczycki@agh.edu.pl; kulczycki@ibspan.waw.pl

P.A. Kowalski
e-mail: pkowal@agh.edu.pl; pakowal@ibspan.waw.pl

P. Kulczycki · P.A. Kowalski
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_2

specific classes. Such a label should be forecast for another element submitted for testing which does not already have one. This procedure of mapping a label suggesting membership to a class, to an investigated element is called a classifier.[1] If the concept of the classifier is based on a rough method, giving no strict guarantee of finding the best or even a correct solution, it can be categorized as heuristic [23], while if a few different concepts combine, where some act as servants to others, then it becomes metaheuristic. Finally, when computational intelligence methodology [11] is used, the data set mentioned at the beginning becomes a learning set. Its subsets assigned to particular classes are referred to as patterns.

This publication concerns the classification of data given in interval form [10], including also the multidimensional case. The fundamental benefit of this type of data is its simplicity, transparency, and possibility of using well-developed mathematical apparatus. Besides actual interval analysis, the case investigated here also includes a probabilistic approach with uniform distribution as well as fuzzy logic for a rectangular membership function. On the other hand in this publication, patterns consist of elements which are uniquely determined (including single-point distribution or crisp numbers for probabilistic and fuzzy approaches, respectively). This corresponds to many situations occurring in practice, for instance when patterns are formed from elements precisely measured some time ago (e.g., exchange rates, outside temperature), but the forecast, ambiguous in nature, is classified and presented in interval form [17].

Changeability in time of analyzing data is assumed here. Literature terms this a changing environment [21], occasionally also evolving data stream [3], concept drift [29], nonstationarity [19], or relates it with the adaptation process [4]. Such a problem is most commonly connected to permanent supplementation of a data set with new elements, which are naturally the most up to date and therefore the most valuable. In the methodology presented below, each of the patterns' element receives coefficients proportional to their influence on correct results. Those elements with smallest coefficients are removed, although an exception is made for those with successively growing values, as their character is in accordance with the trend of changes in the environment.

The metaheuristic proposed here will construct Bayes classifier [5], with a deservedly high opinion among researchers. It possesses a range of advantages, both theoretical (ensuring minimum expectation value of losses resulting from classification errors, albeit for incompletely fulfilled assumption of the attributes' independence) and practical (the idea is simple, robust, and being easy to interpret, is easy to modify). This method allows any number of classes and enables to differentiate their meaning from a practical perspective. The probability values existing in the classifier will be established by means of the nonparametric kernel estimators methodology [16]. Patterns can therefore be of any shape, including consisting of separate parts. Particular attributes of processed data may be

[1]Sometimes this procedure performs the function of reflecting reality with mathematics and information technology, which explains why it is occasionally called a model.

continuous, categorical, or a combination of both. It is worth noting that, thanks to the correctly chosen measure of similarity, it is possible to treat categorical variables as multivalued, including binary. The fixing and adaptation of estimators' parameters are carried out based on optimization procedures [12] and a sensitivity analysis known from the artificial neural networks technique [30].

The initial sections, Sects. 2–5, shortly present a theoretical basis applied later in the Sect. 6, the main section, to create the classification procedure for use in changing environments. Conclusions with numerical verification, followed by final comments, are the subject of Sect. 7.

The concept worked out here connects research for the interval stationary case with the deterministic nonstationary, which are accessible in the papers [18, 19], respectively. Initial results were described in the publication [20]. The specific aspects of using neural networks in the methodology proposed here are the subject of the articles [14, 15], currently in press.

2 Kernel Estimators

The nonparametric method of statistical kernel estimators enables the establishment of characteristics—mainly density of distribution—without any prior knowledge concerning its type. Thus, let an n-dimensional continuous random variable be given. Suppose that its distribution has a density, denoted by f. Having the random sample

$$x_1, x_2, \ldots, x_m \tag{1}$$

one can obtain its kernel estimator [16, 26, 28] defined as

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{2}$$

whereas the function $K: \mathrm{R}^n \to [0, \infty)$, named a kernel, is measurable, symmetrical with respect to zero, has a weak global maximum at this point, and fulfills the condition $\int_{\mathrm{R}^n} K(x)\,\mathrm{d}x = 1$; the constant $h > 0$ is called a smoothing parameter.

The generalized one-dimensional Cauchy kernel

$$K(x) = \frac{2}{\pi\left(x^2 + 1\right)^2}, \tag{3}$$

will be used in the following. This type of kernel lends itself especially well to the classification problem, thanks to the presence of so-called "heavy tails", valuable in areas of potential division into particular classes, actually lying on peripheries of distributions associated with them. For the multidimensional case, the product approach will be used. The kernel is then defined as

$$K(x) = K\left(\begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_n \end{bmatrix}\right) = K_1(x_1)\, K_2(\mathrm{x}_2)\, \ldots\, K_n(\mathrm{x}_n), \tag{4}$$

where K_1, K_2, $\ldots$, K_n represent one-dimensional kernels (3). Note that the expression h^n must be substituted in definition (2) by $h_1 \cdot h_2 \cdot \ldots \cdot h_n$, i.e., the product of smoothing parameters for consecutive coordinates. Observe also that thanks to the continuity of the kernel (3)–(4), the estimator $\hat{f}$ defined by equality (2) is also continuous.

Due to the planned correction in the smoothing parameter h, for calculation of its value the so-called simplified method is enough [16—Sect. 3.1.5; 28—Sect. 3.2.1]. In the one-dimensional case, as well as for particular coordinates in the multidimensional case, the smoothing parameter can be then calculated from a simple formula:

$$h = \left(\frac{W(K)}{U(K)^2}\frac{8\sqrt{\pi}}{3\,m}\right)^{1/5} \widehat{\sigma}, \tag{5}$$

while $W(K) = \int_{\mathrm{R}} K(x)^2\, \mathrm{d}x$, $U(K) = \int_{\mathrm{R}} x^2 K(x)\, \mathrm{d}x$, and $\widehat{\sigma}$ is an (one-dimensional) estimator of standard deviation obtained on the basis of sample (1). For the Cauchy kernel (3) one has $W(K) = 1$ and $U(K) = 5/4\pi$.

Kernel estimators are fully presented in the classic monographs [16, 26, 28], also including among others comments on the choice of kernel type [16—Sect. 3.1.3; 28—Sects. 2.7 and 4.5], algorithms for calculation of the smoothing parameter [16—Sect. 3.1.5; 28—Chap. 3 and Sect. 4.7], and additional concepts for fitting this type of estimator to specific conditions (e.g., boundary of random variable support) and procedures generally increasing its quality. In this latter group, it is worth highlighting the procedure for a smoothing parameter modification [16—Sect. 3.1.6; 26—Sect. 5.3.1], narrowing of particular kernels in dense areas (which enables better characterization of individual features of distribution), and also "flattening" them in sparse regions to additionally smooth the estimator on the peripheries ("tails") of distribution. The potential addition of this aspect to the material presented below is obvious and has been described in detail in the paper [19].

Kernel estimators can also be constructed for different than continuous types of attributes, in particular categorical (nominal and ordered), which through the appropriate selection of similarity measure offers a wide range of generalizations to multivalued variables, including binary. Various compositions of the above types are also possible. The explanations for this topic can be found in the publications [7, 22, 24]. The supplementation of this aspect to the considerations presented in this work is obvious.

3 Bayes Classification

The classification process consists of creating a decision rule, which will map to the tested element an additional label, demonstrating supposed membership to one of the earlier defined classes. These classes are represented by patterns, i.e., sets of elements already possessing such labels. At the beginning consider a continuous random variable. First, the one-dimensional case (relating to the previous section: $n = 1$) will be investigated. Consider therefore the tested quantity, given in the form of the interval

$$[\underline{x}, \overline{x}], \tag{6}$$

while $\underline{x} \leq \overline{x}$. Note that when $\underline{x} = \overline{x}$, it becomes precise (i.e., deterministic or sharp). Let also J classes of the sizes $m_1, m_2, \ldots, m_J$ be represented by patterns composed of real numbers:

$$x_1^1, x_2^1, \ldots, x_{m_1}^1 \tag{7}$$

$$x_1^2, x_2^2, \ldots, x_{m_2}^2 \tag{8}$$

$$\vdots$$

$$x_1^J, x_2^J, \ldots, x_{m_J}^J. \tag{9}$$

(Note that the upper index in the notations (7)–(9) denotes membership to a fixed class). Bayes classification consists of mapping the tested element (6) to the j-class ($j = 1, 2, \ldots, J$) if the largest is the j-th value among

$$m_1 f_1(\tilde{x}), m_2 f_2(\tilde{x}), \ldots, m_J f_J(\tilde{x}), \tag{10}$$

where $f_1, f_2, \ldots, f_J$ denote probability density with the condition of its membership to the class 1, 2, …, J, respectively. In the metaheuristic investigated here, these densities will be defined by kernel estimators methodology, described in Sect. 2, where successive patterns (7)–(9) will be used as samples (1). Suppose therefore such estimators of the above densities as $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_J$. Then expressions (10) take the form

$$m_1 \hat{f}_1(\tilde{x}), m_2 \hat{f}_2(\tilde{x}), \ldots, m_J \hat{f}_J(\tilde{x}). \tag{11}$$

In turn for interval type of data, denoted in the form of element (6), one can conclude that it belongs to the j-class when the biggest is the j-th value from among

$$\frac{m_1}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\,\mathrm{d}x, \frac{m_2}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\,\mathrm{d}x, \ldots, \frac{m_J}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\,\mathrm{d}x. \tag{12}$$

If one uses the continuous kernel K, then formula (12) becomes the generalization of (11). In fact, here the kernel estimator $\hat{f}_j$ is also continuous, therefore for any fixed $\tilde{x} \in [\underline{x}, \overline{x}]$, if the length of interval (6) is reduced to 0 by $\underline{x} \to \tilde{x}$ and $\overline{x} \to \tilde{x}$, then one obtains

$$\lim_{\substack{\underline{x} \to \tilde{x} \\ \overline{x} \to \tilde{x}}} \frac{1}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_j(x)\,\mathrm{d}x = \hat{f}_j(\tilde{x}) \quad \text{for } j = 1, 2, \ldots, J. \tag{13}$$

The expressions (12) transform into (11).

Furthermore, the positive expression $1/(\overline{x} - \underline{x})$ can be removed as having no influence on which factor in formula (12) is the largest. Then it becomes equivalent to

$$m_1 \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\,\mathrm{d}x,\ m_2 \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\,\mathrm{d}x,\ \ldots,\ m_J \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\,\mathrm{d}x. \tag{14}$$

Moreover, for every $j = 1, 2, \ldots, J$ we have

$$\int_{\underline{x}}^{\overline{x}} \hat{f}(x)\,\mathrm{d}x = \widehat{F}(\overline{x}) - \widehat{F}(\underline{x}) \tag{15}$$

with

$$\widehat{F}(x) = \int_{-\infty}^{x} \hat{f}(y)\,\mathrm{d}y. \tag{16}$$

Substituting to the above dependency the definition for kernel estimator (2) (for $n = 1$) with Cauchy kernel (3) and removing once again the positive constant $1/m\pi$ irrelevant here, one can obtain the following analytical formula:

$$\widehat{F}(x) = \sum_{i=1}^{m} \left[\frac{(x^2 - 2xx_i + x_i^2 + h^2)\ \operatorname{arctg}\left(\frac{x - x_i}{h}\right) + h(x - x_i)}{x^2 - 2xx_i + x_i^2 + h^2} + \frac{\pi}{2} \right]. \tag{17}$$

In summary: the tested element (6) should be mapped to the j-class ($j = 1, 2, \ldots, J$) if the j-th value is the largest from expressions (14). The integrals appearing there can be calculated using formula (15) with substitution of dependence (17). This completes the classification algorithm in the one-dimensional case.

Now consider the multidimensional case, i.e., $n > 1$, when the interval vector

$$\begin{bmatrix} [\underline{x_1}, \overline{x_1}] \\ [\underline{x_2}, \overline{x_2}] \\ \vdots \\ [\underline{x_n}, \overline{x_n}] \end{bmatrix} \tag{18}$$

is tested, while elements of patterns (7)–(9) belong to the space R^n. Then expressions (14) are

$$m_1 \int_E \hat{f}_1(x)\,\mathrm{d}x,\ m_2 \int_E \hat{f}_2(x)\,\mathrm{d}x,\ \ldots,\ m_J \int_E \hat{f}_J(x)\,\mathrm{d}x, \tag{19}$$

where $E = [\underline{x_1}, \overline{x_1}] \times [\underline{x_2}, \overline{x_2}] \times \cdots \times [\underline{x_n}, \overline{x_n}]$. To calculate the above integrals, observe that for the product kernel (4), the following is true:

$$\int_E K(x)\,\mathrm{d}x = [I_1(\overline{x_1}) - I_1(\underline{x_1})][I_2(\overline{x_2}) - I_2(\underline{x_2})]\ \ldots\ [I_n(\overline{x_n}) - I_n(\underline{x_n})], \tag{20}$$

where I_i means the primitive function of the one-dimensional kernel K_i for $i = 1, 2, \ldots, n$. Equalities (15) and (17) provide analytical formulas for obtaining the values of these integrals, which completes the procedure for classification of interval data in the continuous random variable case.

The above material can be easily transposed from continuous to categorical variables. Here, an interval element should be understood to be the set sum of several categories. In this situation, testing an element of such type, one should add the kernel estimators values for all categories belonging to the created sum (or their combinations if there are a number of categorical attributes), and then apply criterion (11). The procedure is similar for a combination of continuous and categorical attributes: for fixed categories belonging to the set one should—using the above-presented methodology—calculate kernel estimator values for continuous attributes, add them, and finally apply criterion (11).

Finally, generalize expressions existing in (11) and (19), introducing the coefficients $z_1, z_2, \ldots, z_J > 0$ in the following manner:

$$z_1 m_1 \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\,\mathrm{d}x,\ z_2 m_2 \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\,\mathrm{d}x,\ \ldots,\ z_J m_J \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\,\mathrm{d}x \tag{21}$$

$$z_1 m_1 \int_E \hat{f}_1(x)\,\mathrm{d}x,\ z_2 m_2 \int_E \hat{f}_2(x)\,\mathrm{d}x,\ \ldots,\ z_J m_J \int_E \hat{f}_J(x)\,\mathrm{d}x, \tag{22}$$

respectively. Taking as standard values $z_1 = z_2 = \ldots = z_J = 1$, formula (21) brings us to (14), and (22) to (19). By appropriately changing the value z_i, one can appropriately influence the probability of assigning elements from the i-th class to other wrong classes, although potentially at the cost of increasing the total number of misclassifications. This concept can be applied in such situations where particular classes are associated with phenomena of different significance to the investigated task, or diverse conditioning. In the case of changing environments, moving patterns represent a much more difficult scenario. They may contain elements which are no longer current, or have already appeared, but will only become typical in the future. The adaptation procedure for such patterns is significantly less efficient than for unchanging patterns, where instead of the necessity for updating they can be successively improved by removing less effective elements. In the presented problem, the coefficient z_i values should be, respectively, proportional to the speed of changes of the i-th classes. The value, 1.25 can be proposed as initial; generally for the most applicational tasks $z_1, z_2, \ldots, z_J \in [1,\ 1.5]$.

Bayes classification is highly regarded among practitioners. It is uncomplicated, easily interpretable, and often provides results better than many more refined procedures. Together with kernel estimators, with a very small value of the smoothing parameter, it is reminiscent of the nearest neighbor algorithm, whereas when it is large, it is similar to average (mean) linkage. Thanks to the proper choice of the smoothing parameter, it seems possible to obtain better results than in the case of those two effective methods. Within the proposed metaheuristic, this aspect is reflected in the optimal correction of the above parameter, presented in the next section.

More details concerning Bayes classification is included in the publications [1, 5]; see also [9, 13]. A somewhat broader presentation of the material of the above section can be found in the paper [18].

4 Correction for Smoothing Parameters

With the aim of improving quality of results as well as creating the possibility of keeping up with environment changes, the metaheuristic investigated here applies a correction procedure to the smoothing parameters values, using optimizing algorithms, suiting the value (5) to the classification problem.

Thus, suppose n correcting coefficients $b_1,\ b_2,\ \ldots,\ b_n > 0$, which will be used to multiply the particular smoothing parameters $h_1,\ h_2,\ \ldots, h_n$ calculated using formula (5), respectively. Note that the case $b_1 = b_2 = \ldots = b_n = 1$ means a lack of correction. Assume the natural performance index

$$J(b_1, b_2,\ \ldots, b_n) = \#\{\text{incorrect classifications}\}, \tag{23}$$

where # denotes here the number of elements, and the task of minimization of its value. First, on the grid created for the values $b_j = 0.25,\ 0.5,\ \ldots,\ 1.75$ for every

coordinate $j = 1, 2, \ldots, n$, one should calculate the values of the above index, and then choose the best five. Next, treating these points as initial, static optimization methods in the space R^n ought to be used. The value of index (23) can be calculated by the classic leave-one-out method. Due to these values being integers, a modified Hook–Jeeves procedure [12], with initial step taken as 0.2, was applied. Other conceptions are described in the survey paper [27]. After finishing the above five "runs" of the Hook–Jeeves procedure, one should select one of these values of the correcting coefficients $b_1, b_2, \ldots, b_n$ for which functional (23) value for the end point is the smallest.

However, the above-presented correction of the smoothing parameters procedure is not necessary, it increases classification accuracy, enhances adaptation, and furthermore enables the use of a simplified method for calculating smoothing parameters values (5), based on the square criterion, which is not always beneficial to the classification task [8]. Its influence could have particular significance in abrupt or atypical changes of environment. When applying the modification procedure for the smoothing parameter (see the penultimate paragraph of Sect. 2), the above action undergoes moderate generalization in accordance with the concept described in the paper [19].

5 Pattern Size Reduction

In practical tasks, several elements of patterns (7)–(9) might be unimportant, and in some cases may even have negative influence for classification quality. Their proper selection and removal can improve the correctness of results, and also—thanks to a reduction in pattern sizes—significantly accelerate calculations. To this end, we shall generalize the definition of kernel estimator (2) to the following form:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} w_i K\left(\frac{x - x_i}{h}\right), \tag{24}$$

where the coefficients $w_1, w_2, \ldots, w_m \geq 0$ introduced above are normed such that

$$\sum_{i=1}^{m} w_i = m. \tag{25}$$

In the special case $w_i \equiv 1$, formula (24) reduces to its initial definition (2). The parameters w_i are intended to characterize the influence of the respective i-th elements of the patterns on the accuracy of results. In order to calculate their values, the sensitivity analysis, familiar from the theory of artificial neural networks [6, 30], will be applied. Its aim is to define—after the learning phase—the influence of the particular inputs u_i of a neural network on its output value y, described in the natural way by the quantity

$$S_i = \frac{\partial y(x_1, x_2, \ldots, x_m)}{\partial x_i} \text{ for } i = 1, 2, \ldots, m, \tag{26}$$

and then to aggregate information in the form of the coefficients

$$\overline{S}_i = \sqrt{\frac{\sum_{p=1}^{P} (S_i^{(p)})^2}{P}} \text{ for } i = 1, 2, \ldots, m, \tag{27}$$

where $S_i^{(p)}$ with $p = 1, 2, \ldots, P$ denotes the value (26) for particular iterations. A detailed description of the sensitivity method, together with the appropriate formulas, is presented in the publications [6, 30]. The configuration of neural networks and specific aspects associated with this topic are presented in the separate papers [14, 15]. To every class characterized by patterns (7)–(9) an individual network is assigned. For the sake of simplified notation, the index $j = 1, 2, \ldots, J$ of particular classes will be fixed hereinafter.

In order to define the values of the parameters introduced in definition (24), first calculate auxiliary quantities

$$\tilde{w}_i = \left(1 - \frac{\overline{S}_i}{\sum_{j=1}^{m} \overline{S}_j} \right), \tag{28}$$

finally normed—in consideration of condition (25)—to

$$w_i = m \frac{\tilde{w}_i}{\sum_{i=1}^{m} \tilde{w}_i}. \tag{29}$$

The concept of the above formulas stems from the fact that neural networks are most sensitive to redundant and atypical elements which, from a classification point of view, are mainly of negative significance, therefore they receive the values $\tilde{w}_i$ and in consequence w_i should be proportionately small. Note also that due to the shape of formulas (26)–(27), in practice not all coefficients $\overline{S}_i$ are equal to zero, which guarantees the nominator in dependence (28) is not equal to zero.

Finally, those elements of patterns (7)–(9) for which $w_i < 1$ are removed. The limit value 1 results from the fact that, thanks to the form of normalization (29), the arithmetic mean of parameters equals 1. Empirical research carried out confirmed this theoretically conditioned point of view [14, 15].

6 Classification Metaheuristic

This crucial section collates the material presented in this paper. Procedures presented earlier in Sects. 2–5, will be joined in the classifying metaheuristic designed for the changing environment case. An illustration is provided in Fig. 1. Blocks drawn with a continuous line denote operations performed on all elements of patterns, with a dashed line—on particular classes, while a dotted line symbolizes operations for each element of those patterns.

To start, one should fix the so-called reference sizes of patterns (7)–(9), denoted hereinafter as $m_1^*, m_2^*, \ldots, m_J^*$. They are the sizes of patterns defined during the reduction procedure presented in Sect. 5. Of course, initial patterns must be of a size no smaller than the reference ones. These values may be changed, with the natural boundary that their increase cannot be smaller than the amount of new elements. To begin one can propose $m_1^* = m_2^* = \cdots = m_J^* = 25 \cdot 2^n$. Greater values may cause an increase in calculation time, while smaller a drop in accuracy of results.

Initial patterns (7)–(9) constitute preliminary data submitted for investigated procedure. First, the values of the smoothing parameters $h_1, h_2, \ldots, h_n$ are calculated according to the material of Sect. 2. This action is denoted in Fig. 1 as block A. The subsequent block B symbolizes computation for the coefficients $b_1, b_2, \ldots, b_n$ values, realizing a correction of the smoothing parameters, worked out in Sect. 4.

The next step, described in Sect. 5 (block C in Fig. 1), consists of the calculation of the parameters w_i values, carried out separately for particular classes. After that, these parameters are sorted within each class (block D in Fig. 1). Any sorting procedure [25] can be used here. Following this, shown in Fig. 1 as block E, the $m_1^*, m_2^*, \ldots, m_J^*$ elements corresponding to the largest values w_i are the basis of the principal phase of the investigated procedure—Bayes classification (block F in Fig. 1), which will be discussed in the subsequent paragraph. On the other hand, elements corresponding to smaller values w_i are sent to block U, during which the derivative w_i' is calculated individually for each of them. Newton's interpolation polynomial for the last three observations can be proposed here; its description, together with formulas as well as similar methods are presented in the survey paper [27]. (If for some element, three previous values w_i are not available, then they can be filled with zeroes, artificially increasing a derivative, while at the same time securing such elements against premature removal.) Later the values w_i' are sorted separately for specific classes (block V in Fig. 1), after which—within block W—elements of each pattern in the number

$$qm_1^*, qm_2^*, \ldots, qm_J^*, \tag{30}$$

respectively, with the largest positive derivative values, return to block A at the beginning. The leftover elements are finally removed, as is shown in block Z.

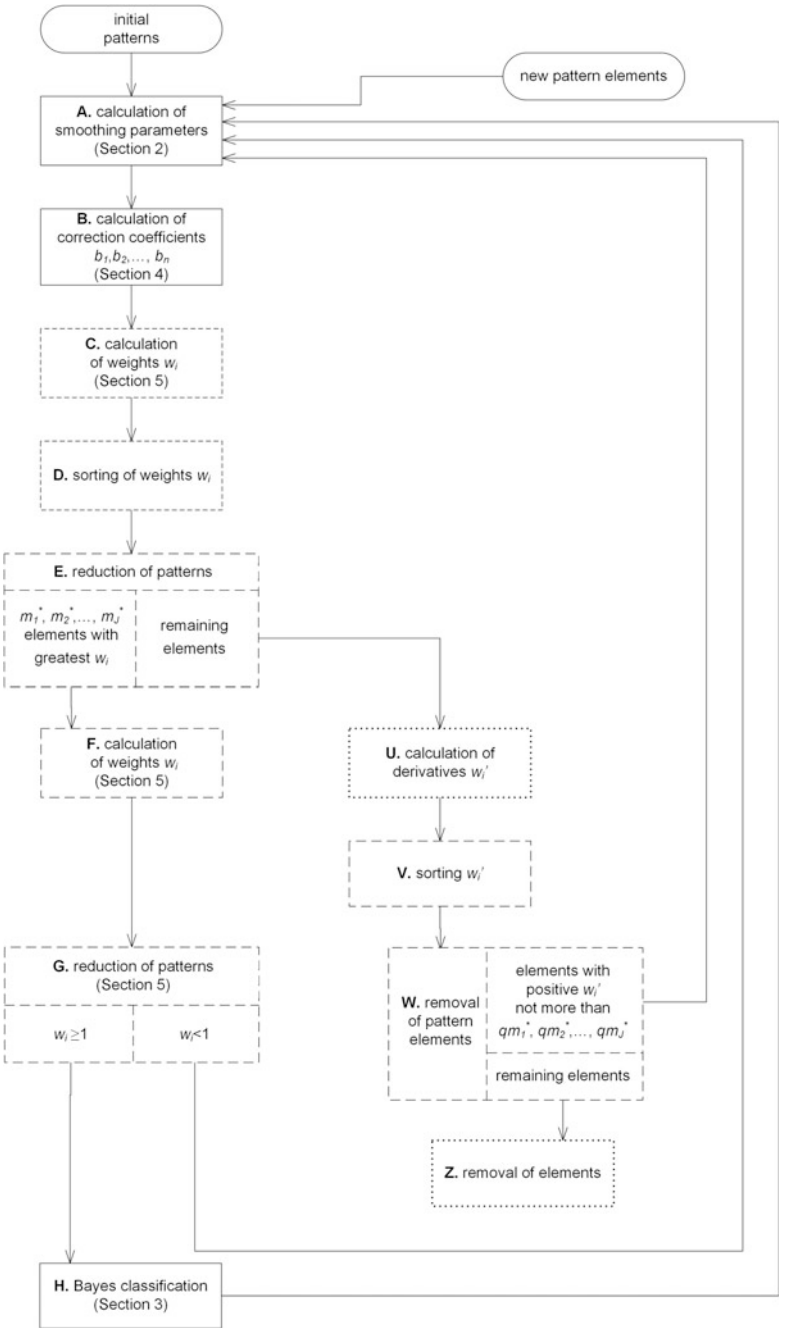


Fig. 1 Classification metaheuristic

The positive parameter q introduced above in formula (30) implies the part played in further tests of elements with small, but successively increasing significance, therefore preceding trends of environment changes, as it were. The initial value $q = 0.2$ is proposed; generally $q \in [0.1,\ 0.25]$ depending on intensity and uniformity of changes. Bigger values may improve the adaptation process but lengthen calculation time, while smaller ones bring contrary effects.

Let us return to Bayes classification, the essence of the procedure presented here. As mentioned at the top of the previous paragraph, this stage sees the arrival of those patterns' elements which have the greatest influence on accurate results. First the parameters' w_i values are once more calculated, in accordance with Sect. 5 (block F in Fig. 1). Then within block G those elements for which $w_i < 1$ are excluded from further processing and sent at the beginning to block A, while those with $w_i \geq 1$ are prescribed to block H, where they form the basis for Bayes classification, described in Sect. 3 (block H in Fig. 1). Testing can be performed on many interval data of type (6) or (18). Next all patterns' elements join block A at the beginning.

The presented procedure can be repeated as soon as new elements are provided to block A. In addition, there are also applied the previously used $m_1^*,\ m_2^*,\ \ldots, m_J^*$ elements with the largest values w_i as the most valuable for accuracy of results, as well as approximately $qm_1^*,\ qm_2^*,\ \ldots, qm_J^*$ ones having the greatest positive derivative w_i', as not having yet big influence but successively increasing their significance as the environment changes.

The expanded description of the procedure presented above can be found in the paper [19].

7 Verification and Final Comments

The correctness of the method described in this paper underwent comprehensive numerical verification. In particular, it was shown that the classification developed here offers correct results also in cases of nonseparated classes with composite multisegment and multimodal patterns. The character of changing environment may increase successively, abruptly, or also periodically, although the best results are found in the first case. The standard values proposed in this text for the parameters used were obtained as deductions from simulations carried out.

The results differed little in nature from those obtained in the basic case where an element which is uniquely defined, e.g., deterministic or crisp, undergoes testing. It proves proper averaging introduced by formulas (14) and (19).

As an example, presented in Fig. 2, let us consider the illustratory two-dimensional case with two classes, one of which is invariable, with the other also unchanging at the beginning, after the 18th step it starts to change its place, and then—after describing a full orbit around the first class—stops in the 54th step at its initial location. The remaining parameters are accepted in the form proposed above

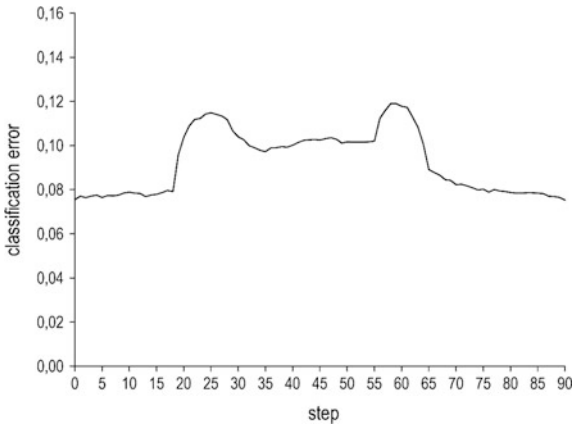


Fig. 2 Number of misclassifications at particular steps of the representative run

in this text. One can see in Fig. 2 that the number of misclassifications increases sharply at times when the environment changes its character, i.e., in steps 18 and 54. The prediction function is then ineffective by nature. In the periods of non-stationarity, i.e., before the 18th and after the 54th step, the rate of errors stabilizes at a value of 0.08, whereas in the period of constant changes between the 18th and 54th steps, at the higher 0.105. This is still lower than the maximums values 0.12, which would be maintained without the influence of the adaptation function designed here.

Further research was undertaken on the influence of size of imprecision of classified data—represented by the length of intervals—on accuracy of results. In this aspect also the effects showed themselves to be fully satisfactory. If the interval length was less than the generally understood distance between centers of specific patterns (a condition usually fulfilled in practice), then its growth did not cause an increase in the mean value of incorrect classifications, but in fact the results underwent some stabilization—the variance of misclassifications decreased. Again averaging, introduced by formulas (14) and (19), proves to have a positive influence.

A broader description of particular aspects of the above simulations can be found in the papers [14, 15, 18, 19].

The metaheuristic proposed in this paper was compared with other classification methods based on computational intelligence, e.g., Support Vector Machine, as well as natural, e.g., counting components of patterns which are included in the tested element. Unfortunately, no method has been found to allow exactly the same conditionings: uniquely defined patterns elements, interval form of tested element, changing environment, any number of classes and patterns shapes, categorical attributes. For this reason, it was possible only to compare with simplifications fitting suitable methodologies, and so offer the results presented below purely in a qualitative aspect. The advantage of the metaheuristic proposed in this paper mainly lies in the smaller number of misclassifications for stabilized variability of environment, which in Fig. 2 appears as a significant decrease in errors between 30 and

55 steps. Better results are also achieved here in areas between particular patterns, which are always troublesome for classification, as well as for long intervals representing specific attributes of tested elements. Thanks to the calculational complexity of particular procedures of the metaheuristic under investigation, the proposed method is especially destined for those cases where slow learning is permitted, but the classification process itself must be fast. This is achieved in great part by obtaining an analytical form of formulas (15)–(17). The computational complexity of the classification phase alone amounts to $O(nJm)$, and therefore is linear with respect to dimensionality of space, number of classes, and size of their patterns.

References

1. Aggarwal, C.C.: Data classification: algorithms and applications. Chapman & Hall/CRC, London (2014)
2. Aggarwal, C.C.: Data mining. The textbook. Springer, Cham (2015)
3. Aggarwal, C.C., Han, J., Wang, J., Yu, P.S.: A framework for on-demand classification of evolving data streams. IEEE Trans. Knowl. Data Eng. **18**, 577–589 (2006)
4. Bouchachia, A.: Adaptation in classification systems. In: Hassanien, A.E., Abraham, A., Herrera, F. (eds.) Foundations of Computational Intelligence, vol. 2, pp. 237–258. Springer, Berlin (2009)
5. Duda, R.O., Hart, P.E., Storck, D.G.: Pattern classification. Wiley, New York (2001)
6. Engelbrecht, A.P., Cloete, I., Zurada, J.: Determining the significance of input parameters using sensitivity analysis. In: Mira, J., Sandoval F. (eds.) From Natural to Artificial Neural Computation. Lecture Notes in Computer Science, pp. 382–388. Springer, Berlin (1995)
7. Gaosheng, J., Rui, L., Zhongwen, L.: Nonparametric estimation of multivariate CDF with categorical and continuous data. Adv. Econom. **25**, 291–318 (2009)
8. Ghosh, A.K., Chaudhuri, P., Sengupta, D.: Classification using kernel density estimation: multiscale analysis and visualization. Technometrics **48**, 120–132 (2006)
9. Hryniewicz, O., Kaczmarek, K., Nowak, P.: Bayes statistical decisions with random fuzzy data—an application for the Weibull distribution. Maint. Reliab. **17**, 610–616 (2015)
10. Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied interval analysis. Springer, Berlin (2001)
11. Kacprzyk, J., Pedrycz, W. (eds.): Springer handbook of computational intelligence. Springer, Dordrecht (2015)
12. Kelley, C.T.: Iterative methods for optimization. SIAM, Philadelphia (1999)
13. Kobos, M., Mandziuk, J.: Multiple-resolution classification with combination of density estimators. Connect. Sci. **23**, 219–237 (2011)
14. Kowalski, P.A., Kulczycki, P.: A complete algorithm for the reduction of pattern data in the classification of interval information. Int. J. Comput. Methods. **13**(1650018) (2016)
15. Kowalski, P.A., Kulczycki, P.: Interval probabilistic neural network. Neural Comput. Appl. (2017, in press)
16. Kulczycki, P.: Estymatory jadrowe w analizie systemowej. WNT, Warsaw (2005)
17. Kulczycki, P., Hryniewicz, O., Kacprzyk, J. (eds.): Techniki informacyjne w badaniach systemowych. WNT, Warsaw (2007)
18. Kulczycki, P., Kowalski, P.A.: Bayes classification of imprecise information of interval type. Control Cybern. **40**, 101–123 (2011)
19. Kulczycki, P., Kowalski, P.A.: Bayes classification for nonstationary patterns. Int. J. Comput. Methods **12**(1550008) (19 pages) (2015a)

20. Kulczycki, P., Kowalski, P.A.: Classification of interval information with data drift. In: Christiansen, H., Stojanovic, I., Papadopoulos, G.A. (eds.) Modeling and Using Context. Lecture Notes in Computer Science, pp. 495–500. Springer, Berlin (2015b)
21. Kuncheva, L.I.: Classifier ensembles for changing environments. In: Roli, F., Kittler, J., Windeatt, T. (eds.) Multiple Classifier Systems. Lecture Notes in Computer Science, pp. 1–15. Springer, Berlin (2004)
22. Li, Q., Racine, J.S.: Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. J. Bus. Econ. Stat. **26**, 423–434 (2008)
23. Michalewicz, Z., Fogel, D.B.: How to Solve It: Modern Heuristics. Springer, New York (2004)
24. Ouyang, D., Li, Q., Racine, J.: Cross-validation and the estimation of probability distributions with categorical data. J. Nonparametric Stat. **18**, 69–100 (2006)
25. Sedgewick, R., Wayne, K.: Algorithms. Addison-Wesley, Upper Saddle River (2011)
26. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
27. Venter, G.: Review of optimization techniques. Encyclopedia of Aerospace Engineering, pp. 5229–5238. Wiley, New York (2010)
28. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1995)
29. Zlobaite, I.: *Learning under Concept Drift: an Overview*, Technical report, Faculty of Mathematics and Informatics, Vilnius University (2009)
30. Zurada, J.: Introduction to Artificial Neural Network Systems. West Publishing, St. Paul (1992)

A Fuzzy Information Propagation Algorithm for Social Network Based Recommender Systems

Gergely Posfai, Gabor Magyar and Laszlo T. Koczy

Abstract Web-based services that have become prevalent in people's everyday life generate huge amounts of data, which makes it hard for the users to search and discover interesting information. Therefore, tools for selecting and delivering personalized contents for users are crucial components of modern web applications. Social recommender systems suggest items to users assuming the knowledge of the users' social network. This new approach can alleviate the common weaknesses of traditional recommender systems, which completely ignore the users' personal relationships in the recommendation process. In this paper, a social network based fuzzy recommendation technique is presented, which propagates information through the users' social network and predicts how users would probably like a certain product in the future. Experimental results on a public dataset show that the proposed method can significantly outperform popular and widely used recommendation system methods in terms of recommendation coverage while maintaining prediction accuracy and performs especially well for cold start users, that have only rated a few items or no item at all previously.

1 Introduction

Nowadays, the growing popularity of web-based services and applications resulted in the generation of enormous amounts of data. The dramatically increasing volume of information makes it difficult for the applications to provide relevant, personalized,

G. Posfai (✉) · G. Magyar · L.T. Koczy
Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics, Budapest, Hungary
e-mail: posfai@tmit.bme.hu

G. Magyar
e-mail: magyar@tmit.bme.hu

L.T. Koczy
Szechenyi Istvan University, Gyor, Hungary
e-mail: koczy@sze.hu

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_3

engaging contents for their users. Recommender systems emerged to deal with this information overload problem and have become extremely common in recent years. Recommender systems are software tools and techniques that seek to predict the rating that a user would give to an item [1]. That is, typically, in a recommender system, we have a set of users and a set of items, and each user rates a subset of items on a certain scale of rating values. In order to suggest items to users they would probably like, a recommender system tries to predict how users would rate the non-rated items. Applications of recommender systems include recommender systems for movies (imdb), music (last.fm), jobs (LinkedIn), friends (Facebook), search queries (Google), products in general (Amazon, Ebay), etc.

Traditional recommender systems typically can be classified into two groups—content-based and collaborative filtering recommender systems [2]. Content-based methods recommend items to users that are similar to the ones that the user has already liked [3]. Similarity between items is computed based on the set of features associated with items. Collaborative filtering techniques leave the properties of items out of the consideration and only take the users' past ratings into account. Such methods recommend items to users that similar users have already liked [4–6]. The underlying idea of collaborative filtering is that if a user in the past agreed with other users, then other recommendations coming from these similar users should be relevant as well. Similarity between users is calculated based on the users' rating history.

Collaborative filtering is the most widely used recommender system method and has been successfully employed in many applications. In most cases it is very efficient and can be applied in any domain as it relies only on the users' past ratings and does not use the features of items as opposed to content-based recommender systems. Despite the prevalence and popularity of collaborative filtering systems, they have some significant limitations [7–9]. Collaborative filtering performs very poorly for so-called cold start users, who only rated a few items in the past and therefore their similarity with other users cannot be estimated accurately. Additionally, traditional recommender systems does not incorporate information from the users' personal relationships in the recommendation process, although, in real life, when people make decisions—e.g., choose a restaurant to go to, a movie to watch, a product to buy—they often rely on the opinion of friends, family, or colleagues.

The widespread expansion of social network based services propelled the explosion of data available regarding the users' personal relationships, which provides possibility for deeper analysis and exploitation of such information. Social and trust-aware recommender systems aim to overcome the weaknesses of collaborative filtering and content-based recommendations by leveraging the users' social network and producing recommendations for users based on the preferences of their neighbors [7–16]. Therefore, in a social recommender system setup the users' social network is assumed to be known explicitly or implicitly. We focus on the case when users explicitly establish friendships between each other, however, in many cases such information is not available and relationships have to be inferred implicitly.

In order to solve the mentioned problems of collaborative filtering we introduce a novel, social network based fuzzy recommender system method called SNF,

which produces recommendations by propagating known ratings along the friendships of the users' social network and uses fuzzy sets and fuzzy operations to take into account the uncertain nature of how friends influence each other. In this paper, we focus on the rarely studied recommender system setup, when there is only a single item of interest that users rate (e.g., a service, a political candidate, a new product), and given an initial set of known ratings on the item and the friendship network of users, the recommender system provides rating predictions for other users.

The remainder of this paper is organized as follows. Section 2 provides an overview of existing approaches to social and trust-aware recommender systems. In Sect. 3, we introduce our novel social network based fuzzy recommender system method. The results of an experimental evaluation and comparisons with other methods are presented in Sect. 4, followed by the conclusion and possible directions of future research in Sect. 5.

2 Related Work

Social and trust-aware recommender systems both aim to leverage the users' personal relations in the recommendation process in order to improve recommendations [7–16]. Trust-aware methods utilize the users' web of trust, while social recommender systems utilize the users' friendship network during recommendation. In the web of trust edges indicate directed trust relationships, while the edges of friendship networks represent bidirectional friendships. As a consequence trust-aware and social recommender systems have a few differences [15], however, they are essentially the same, and often they are not even distinguished [9]. In this paper, we introduce a method which relies on the users' friendship network, therefore we propose a social recommender system method. However, as many other social recommender system method, ours can also be applied to trust networks too by assigning directions to the edges of the friendship network in both directions.

Many of the earliest trust-aware recommendation methods belong to Masa and Avesani. In [7] they provide a trust-aware method, which computes recommendations in two steps. First, they extend the existing direct trust relationships, and compute indirect trust values between distant users by propagating trust values along the edges of the web of trust. Then they use the direct and inferred trust values in the second step to give more weight to trusted users in the calculation of recommendations. Their method also uses the mean values of the users' previous ratings, which is not available in our setting, since we focus on the case when there is only a single item that users rate, however, the dataset used in our experiments (see Sect. 4.1) makes it possible to retrieve the users' mean rating values, therefore we included the method proposed by Masa and Avesani as a trust-aware benchmark in our experiments.

Matrix factorization is extensively used for model based collaborative filtering. These methods transform both items and users to the same latent factor space in a way to minimize an objective function, then the computed latent feature vectors are used to compute the final recommendations [1]. Social and trust-aware methods are

also frequently built on matrix factorization [9, 13–15] as it can be efficiently used to minimize the distance between the latent feature vectors of connected users. Additionally, more refined recent models also incorporate the notion of trust propagation in the recommendation process [9, 15]. A significant drawback of matrix factorization models is that the latent feature vectors have to be recomputed when a new rating becomes available, which can be very resource demanding [9]. Since matrix factorization-based approaches heavily rely on the user-item matrix, they cannot be applied in our setup.

In [17], Andersen et al. pose a set of axioms of desirable characteristics for trust-aware recommender systems and introduce a few recommendation strategies that are in compliance with certain subsets of the introduced axioms. One of their recommendation methods is based on random walks and works by propagating known ratings from source users to others along the edges of random walks in the users' web of trust. In accordance with our experiments, they also focus on the case, when there is only one single item that users rate. In their model ratings can be either positive (voters), negative (nonvoters) or neutral. Unfortunately, the paper lacks of empirical evaluations, therefore it is difficult to determine the efficiency of the proposed methods, so we included a slightly modified version of their random walk-based method in our analysis (see Sect. 4.2).

In [16], Bharadwajk and Al-Shamri propose a fuzzy computational trust and reputation model. While trust describes an individual relation between two users, reputation is a global attribute of users that can be defined as what is generally said or believed about a person's standing. In their model Bharadwajk and Al-Shamri compute reputation values in two steps. First, individual reputation scores are computed according to the beta reputation model [18], and second, the OWA [19] fuzzy operator is used to compute the users' overall reputation scores by aggregating the individual scores. In order to assess trust values two fuzzy subsets are applied (satisfied and unsatisfied) to describe the users' global trust values from which they infer individual trust values for pairs of users. They also propose a way to incorporate the assessed reputation and trust values in a recommender system in order to restrict the set of participating users in the recommendation process to the most trustworthy and most similar users. An empirical analysis showed that their procedure outperformed other well-known methods like the eBay reputation model. Unfortunately their method is not appropriate for our setup as it leverages the user-item matrix and uses a symmetric trust scale where distrust can also be expressed along with trust, whereas in our analysis we only have friendships that can be considered as fully positive trust statements.

3 Methodology

In this section, we present our novel social network based fuzzy recommender system called SNF, which computes personalized recommendations for users by propagating information along the edges of the users' friendship network. The proposed

method is an enhanced and optimized version of the preliminary recommender system introduced in [20].

We denote the users' social network by the undirected graph $G < U, E >$, where $U = \{u_1, u_2, \ldots, u_N\}$ is the set of users and $E \subseteq U \times U$ is the set of friendships. The discrete set of possible rating values is denoted by $V = \{v_1, v_2, \ldots, v_M\}$ and r_u represents the rating value given by user u. We have an initial set of users $U_0 = \{u_{r_1}, u_{r_2}, \ldots, u_{r_P}\}$ with observed ratings $R_0 = \{r_{u_{r_1}}, r_{u_{r_2}}, \ldots, r_{u_{r_P}}\}$ and another set of users $U_1 \subseteq U \setminus U_0$ with unknown ratings R_1. Our goal is to determine as many elements of R_1 as accurately as possible based on knowledge of R_0 and $G < U, E >$.

For each possible rating value we define a fuzzy subset on the set of users:

$$W_{v_i} : U \to [0, 1] \, , i = 1, 2, \ldots, M, \tag{1}$$

where $W_{v_i}(u_j)$ is the membership value of user j in W_{v_i} which is the fuzzy subset corresponding to the v_i rating value. The membership values of a certain user describe the user's—known or predicted—rating, that is, the membership value of a given rating value's fuzzy subset represents that how likely the user would give that rating value for the item. Obviously, for known ratings the membership degree of the actual rating value's fuzzy subset is 1 and all the other membership values are 0. All the membership degrees of the predicted ratings can take any value from the $[0, 1]$ interval. The final prediction value for a user is computed by weighting the possible rating values by the user's membership values in the corresponding fuzzy subsets and calculating their weighted arithmetic mean. That is, the prediction for user u is given as

$$p_u = \frac{\sum_{i \in M} W_{v_i}(u) v_i}{\sum_{i \in M} W_{v_i}(u)} \, . \tag{2}$$

The membership values of users who do not have known rating are determined by an information propagation procedure, which propagates the membership values of known ratings from the users who issued the rating over all paths of the friendship network to a given maximum path length. That is, we propagate the membership values that belong to users with known ratings (U_0) to users who do not have known rating $((U \setminus U_0))$. The membership values of user $u \in (U \setminus U_0)$ are calculated according to the following formula:

$$W_{v_i}(u) = \vee^1_{q \in U_{v_i}} s(\vee^2_{p \in P_{q,u,d}} w(l_p)) \, , \tag{3}$$

where $U_{v_i} \subseteq U_0$ is the set users who rated the item with value v_i, while $P_{v,u,d}$ denotes the set of paths from user v to user u with length equal or less than d, which is an independent parameter determining the maximum distance of the propagation. Additionally, l_p denotes the length of path p and w is the weight function, which lowers the propagated membership value as the length of the path increases, s denotes a transformation function which rescales the membership value gained from a single

source user before it is aggregated with the membership values gained from other source users, $\vee^1$ and $\vee^2$ denote fuzzy t-conorm functions used to aggregate membership values gained from different source users and different propagation paths coming from the same source user, respectively.

The following weight function was applied in (3):

$$w(l) = \wedge_{i=1}^{l} \alpha \ , \tag{4}$$

where $\wedge$ denotes a fuzzy t-norm operation, while α is an independent parameter. The underlying idea of the weight function is that in real life people usually rely more on the opinion of close friends, than the opinion of barely known or unknown people. Therefore, the effect of known ratings should decrease as they are propagated far away from the source users [21]. Hence, the more far a rating is propagated the less it will influence the predictions.

As the ratings might be propagated along a huge number of paths in the social network the aggregated membership values retrieved by applying the $\vee^1$ fuzzy t-conorm can be very high and can easily reach 1.0, which would deteriorate the effectiveness of the $\vee^2$ operation. Therefore, it is necessary to rescale the propagated membership values between the application of the $\vee^1$ and $\vee^2$ fuzzy t-conorms, so we defined the following scaling function which was applied in the formula of (3):

$$s(x) = \beta x \ , \beta \in [0, 1], \tag{5}$$

where β is an independent parameter.

We also define a confidence value c for each prediction, which is obtained by taking the fuzzy t-conorm of the user's membership values, that is, the confidence of the rating prediction for user u is calculated as:

$$c_u = \vee_{i \in M}^{3} W_{v_i}(u) \ , \tag{6}$$

where $\vee^3$ is an independent parameter.

Finally, at the end of the recommendation process we keep only the recommendations that have confidence values greater than or equal to θ, which is an independent parameter functioning as a confidence threshold. That is, the final set of rating predictions and their corresponding set of users are given as

$$R_p = \{p_u | u \in (U - U_0) \wedge c_u \geq \theta\} \ , \tag{7}$$

$$U_p = \{u | u \in (U - U_0) \wedge c_u \geq \theta\} \ , \tag{8}$$

where R_p is the set of produced rating predictions and U_p is the set of users for which we have rating predictions.

Table 1 Selected parameter values for the SNF method

Parameter	SNF
d	3
α	0.2
β	0.01
$\wedge$	Algebraic product t-norm [22]
$\vee^1$	Yager t-conorm [23] (2.5)
$\vee^2$	Dombi t-conorm [24] (0.8)
$\vee^3$	Zadeh t-conorm [25]
θ	0.004

Based on extensive experiments we used the parameter settings presented in Table 1 in our final model. The values enclosed in parentheses describe the chosen hyperparameter values.

4 Experimental Results

This section describes the empirical evaluation of our method. First, we introduce the dataset used for the analysis. Then we present a set of benchmark methods to which we compared the performance of the SNF method, and also describe the metrics we used to evaluate the performance of recommendation techniques. Then we assess the impact of various parameters of SNF on the recommendation performance, and finally, we present the comparisons with the benchmark methods.

4.1 Dataset

We used the Flixster dataset [9] in our empirical analysis, which is publicly available at http://www.cs.ubc.ca/~jamalim/datasets/. Flixster[1] is a social movie rating website, where users can rate movies on a $[0.5, 5]$ discrete scale with step size 0.5, and also can establish friendships with other users. The dataset consists of ratings and friendships. Each rating includes a user and a movie id, a rating value and a timestamp, while the undirected friendships are identified by two user ids.

As mentioned earlier, we focus on the recommender system setup where there is only a single item that users rate, therefore, when producing recommendations for a movie the SNF method does not rely on ratings given to other movies. On the other hand, since the Flixster dataset contains ratings on many movies, it is possible to evaluate other types of recommendation techniques such as ones that leverage the

[1] http://www.flixster.com.

Table 2 Datasets used in the empirical analysis

Dataset	Statistic	Value
Original ratings	# Users with rating	147,612
	# Items	48,794
	# Ratings	8,196,077
Filtered observed ratings	# Users with rating	19,407
	# Items	1,500
	# Ratings	84,212
Extended observed ratings	# Users with rating	21,286
	# Items	39,310
	# Ratings	1,759,244
Filtered test ratings	# Users with rating	10,147
	# Items	1,500
	# Ratings	31,215
Social network	# Users	787,213
	# Users with friend	1,500
	# Undirected friendships	5,897,324
	Network density	0.000019
	Clustering coefficient	0.03299

user-item matrix. These methods then can be used as reference methods to which we can compare the SNF method.

We created a known and a test dataset for the evaluation. Ratings between 06/01/2008 and 06/01/2009 constitute the known or observed dataset, while ratings after 06/01/2009 constitute the test set. Since the original dataset is huge, we randomly selected 1,500 items for our analysis, which make up the filtered observed and filtered test datasets. We also created an extended version of the filtered observed dataset by adding the remaining ratings of both the observed and test datasets' users from 06/01/2008 to 06/01/2009. This additional dataset only serves for the evaluation of benchmark methods that use the user-item matrix.

A few statistics of the various datasets are shown in Table 2.

4.2 Benchmarks

To assess the performance of the SNF method compared to well-known, popular recommendation methods, we included the following reference methods in our analysis:

- **Item Mean (IM)**: As a rudimentary method, we included the mean value of the movies' observed ratings for each movie.
- **Collaborative Filtering (CF)**: We also evaluated Resnick's standard formula [4], which is an extremely popular collaborative filtering method [7–9, 11, 12]. We

used the Pearson correlation coefficient for the similarity metric between users. Note that this approach utilizes the user-item matrix.

- **Trust-aware (TA)**: As a trust-aware technique, we evaluated the method presented by Massa and Avesani in [7] using the users' friendship network as the web of trust and assigning directions for every edge in both ways. Note that the method also uses the users' mean rating value on other items. We set the maximum distance of trust propagation to 3, which is the same value we used in our method to restrict the maximum distance of information propagation (see Table 1).
- **Random Walk (RW)**: As a social recommender system method, we included the random walk-based method presented in [17]. We slightly altered their algorithm to process ratings from the [0.5, 5.0] scale, and calculated the prediction for a user as the mean of the rating values that were reached by the random walks started from the current user. We generated 20,000 random walks for each user without known rating that were terminated after either reaching a user with known rating or reaching a maximum path length of 5,000 steps. Note that the RW method is the only benchmark that utilizes exactly the same input data as the SNF method.

4.3 Metrics

The following two metrics were used in our analysis to assess the performance of various recommendation strategies:

- **Root Mean Squared Error (RMSE)**: Frequently used metric [8, 9, 15] to measure the error in recommendations, defined as:

$$RMSE = \frac{\sum_{u \in (U_p \cap U_1)} |p_u - r_u|}{|U_p \cap U_1|} . \tag{9}$$

- **Coverage**: The percentage of unknown ratings in the test dataset for which prediction was provided:

$$Coverage = \frac{|U_p \cap U_1|}{|U_1|} \cdot 100\% . \tag{10}$$

4.4 Impact of Confidence Parameters

In this subsection, we examine how the value of θ and $\vee^3$ influence the prediction error and coverage. While θ controls how confidence values are used, $\vee^3$ controls how the confidence values are computed. Figure 1 shows the impact of the two parameters. Each line corresponds to a fuzzy operation used as the value of $\vee^3$ and describes how the RMSE and coverage change if we vary the value of θ.

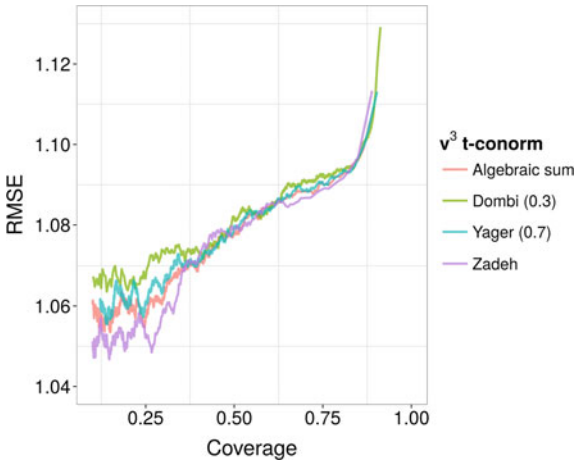


Fig. 1 The impact of the indepent parameter θ on the MAE and coverage. By choosing the value of θ, we determine the trade-off between MAE and coverage (quality and quantity)

The values in parentheses after the name of the operators indicate the used hyperparameter value of the corresponding fuzzy operator.

θ is used as a confidence threshold to filter the final set of predictions. Increasing the value of θ decreases coverage, as less predictions will be accepted. However, it also improves accuracy (i.e. RMSE) as higher θ means we accept only higher confidence values. Hence by setting the value of θ we can adjust the trade-off between RMSE and coverage, i.e., prediction quality and quantity. In our final model we set θ to 0.004, which results in obtaining 70% coverage and a reasonable RMSE.

We tried four different fuzzy t-conorm operators for $\vee^3$ in our experiments: algebraic sum [22], Dombi [24], Yager [23] and Zadeh [25] t-conorms. As can be seen on Fig. 1 the performance of the various operators were close to each other, however, the Zadeh t-conorm apparently outperformed the other operators under 30% coverage in terms of RMSE, so in the final model we used the Zadeh t-conorm as $\vee^3$.

4.5 Impact of Information Propagation Fuzzy Operators

The process of information propagation in the SNF method is vitally affected by the three fuzzy operators $\wedge$, $\vee^1$ and $\vee^2$, which determine how ratings are transitioned and aggregated during the propagation, therefore it is very important to choose these parameters carefully. We investigated the performance of the following fuzzy operators in our experiments:

- **t-norms**: algebraic product [22], bounded subtraction [22], Dombi [24], drastic [22], Dubois [26], Hamacher [27], Schweizer and Sklar [28], Yager [23], Zadeh [25],
- **t-conorms**: algebraic sum [22], bounded sum [22], Dombi [24], drastic [22], Dubois [26], Hamacher [27], Schweizer and Sklar [28], Yager [23], Zadeh [25].

Table 3 Most efficient combinations of fuzzy operators

$\wedge$t-norm	$\vee^1$t-conorm	$\vee^2$t-conorm	RMSE at 70% coverage
Algebraic product	Yager (2.5)	Dombi (0.8)	**1.0867**
Algebraic product	Schweizer and Sklar (-5.0)	Dombi (0.8)	**1.0884**
Schweizer and Sklar (-2.0)	Yager (2.5)	Dombi (0.8)	**1.0893**
Dombi (2.0)	Yager (2.5)	Dombi (0.8)	**1.0899**
Schweizer and Sklar (-2.0)	Schweizer and Sklar (-5.0)	Dombi (0.8)	**1.0906**

Table 3 shows the most efficient combinations of the $\wedge$, $\vee^1$ and $\vee^2$ operators regarding the achieved RMSE at 70% coverage based on our empirical analysis. The values in parentheses indicate the applied hyperparameter of the corresponding parameter. The combination of the algebraic product t-norm, and the Yager and Dombi t-conorms obtained the lowest RMSE, although, other combinations were also able to perform almost as well. An interesting outcome of the experiments is the notable prevalence of the Dombi t-conorm, which was unexceptionally used as the $\vee^2$ parameter in the top combinations, which might be due to the peculiarity that with the proper choice of its hyperparameter the characteristics of the Dombi t-conorm operator is quite unique and significantly differ from all the other tested t-conorms.

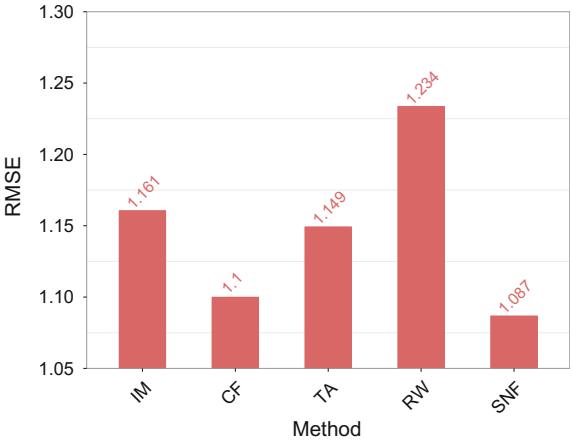
4.6 Comparisons

Figure 2 shows the obtained RMSE and coverage values of the SNF and the reference methods.

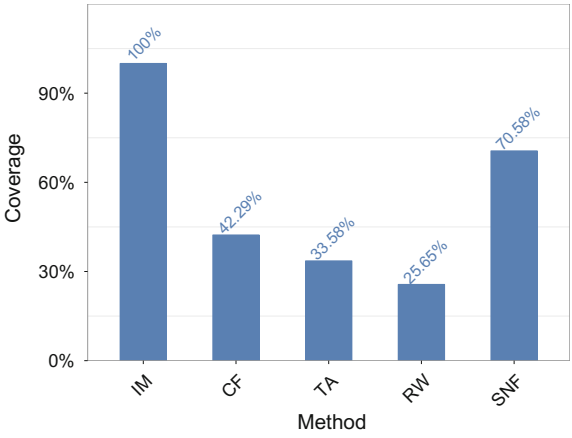
As can be seen, the SNF method obtained the lowest RMSE noticeably outperforming all the benchmark methods. It is followed by the CF method which was able to approximate the performance of SNF. The remainder methods performed substantially poorer. The TA method attained only slightly lower RMSE than the rudimentary IM method, while the RW method attained by far the worst accuracy.

The IM method achieved 100% due to the data sampling methodology (see Sect. 4.1), during which only those items were selected into the test dataset that had at least one rating in the observed dataset. Despite the IM method being able to produce recommendations for every item that has at least one observed rating, it has very poor accuracy; it obtained the second worst RMSE. Besides the IM method, the SNF method had the highest coverage obtaining more than 70% and dramatically exceeding the other reference methods, including the CF method. The RW method performed worst in terms of coverage too surprisingly underachieving the other methods.

Fig. 2 Comparison of RMSE values



(a) Comparison of RMSE values.



(b) Comparison of coverages.

Based on the empirical analysis it could be inferred that with the exception of the IM method's coverage the SNF method was able to considerably outperform all the reference methods both in terms of prediction quality and quantity, i.e., RMSE and coverage. In terms of RMSE only the performance of the CF method was comparable to the SNF method, however, it is important to note that the CF method relies on a completely different type of input data, as it uses the user-item matrix to produce recommendations. Only the RW method uses exactly the same input data as the SNF method, but that method achieved very poor results, and was significantly exceeded by the SNF method.

5 Conclusion

Nowadays recommender systems have become a fundamental part of online applications and are used extensively to handle the information overload problem posed by the huge amounts of generated data. The most popular and widely used technique for recommender systems is collaborative filtering, which is extremely efficient in many cases, although, it has a few major drawbacks, therefore new approaches required to alleviate these deficiencies. Social recommender systems represent a new type of strategy that aims to incorporate the users' friendships in the recommendation process to improve recommendation accuracy and coverage.

In this paper, we proposed a novel information propagation based fuzzy recommender system for social networks called SNF. We focused on the recommender system setup when there is only a single item for users to rate. An empirical analysis showed that the SNF was able to significantly outperform the popular and widely used standard collaborative filtering technique and also other reference methods both in terms of prediction accuracy and coverage. Since the SNF method does not rely on ratings of more than one item it can be of great use especially for cold-start users who have issued only very few or no ratings at all in the past.

Our empirical analysis showed compelling results and allows for many directions for further research. In future work, we will investigate how the dynamics of the friendship network influence the users' ratings, since the people's friendships usually change as time passes and are not static, therefore friendship dynamics should be accounted for in the recommendation process.

We also want to explore how the SNF method can be enhanced by incorporating additional data in the recommendation process. That is, e.g., how we can improve predictions by leveraging the user-item matrix as many other recommender systems do, or how we can cluster the users by taking into consideration their position in the social network and then use these clusters to produce more accurate recommendations. Finally, as many social applications allow users to express negative reviews on other users, i.e., express distrust, it also seems to be worth to extend the SNF method to be able to handle distrust relationships between users.

Acknowledgements The research was supported by National Research, Development and Innovation Office (NKFIH) K105529, K108405.

References

1. Ricci, F., Rokach, L., Shapira, B.: Introduction to Recommender Systems Handbook. Springer (2011)
2. Jafarkarimi, H., Sim, A.T.H., Saadatdoost, R.: A naive recommendation model for large databases. Int. J. Inf. Educ. Technol. **2**(3), 216–219 (2012)
3. Lops, P., De Gemmis, M., Semeraro, G.: Content-based recommender systems: State of the art and trends. In: Recommender Systems Handbook, pp. 73–105. Springer (2011)

4. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference On Computer Supported Cooperative Work, pp. 175–186. ACM (1994)
5. Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Commun. ACM **35**(12), 61–70 (1992)
6. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
7. Massa, P., Avesani, P.: Trust-aware collaborative filtering for recommender systems. In: On the Move to Meaningful Internet Systems, : CoopIS, DOA, and ODBASE pp. 492–508. Springer (2004)
8. Jamali, M., Ester, M.: Trustwalker: a random walk model for combining trust-based and item-based recommendation. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 397–406. ACM (2009)
9. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 135–142. ACM (2010)
10. Avesani, P., Massa, P., Tiella, R.: A trust-enhanced recommender system application: Moleskiing. In: Proceedings of the 2005 ACM Symposium on Applied Computing, pp. 1589–1593. ACM (2005)
11. O'Donovan, J., Smyth, B.: Trust in recommender systems. In: Proceedings of the 10th International Conference on Intelligent User Interfaces, pp. 167–174. ACM (2005)
12. Massa, P., Avesani, P.: Trust-aware recommender systems,. In: Proceedings of the 2007 Acm Conference on Recommender Systems, pp. 17–24. ACM (2007)
13. Ma, H., Yang, H., Lyu, M.R., King, I.: Sorec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM Conference On Information And Knowledge Management, pp. 931–940. ACM (2008)
14. Ma, H., King, I., Lyu, M.R.: Learning to recommend with social trust ensemble. In: Proceedings of the 32nd International ACM SIGIR Conference On Research And Development in Information Retrieval, pp. 203–210. ACM (2009)
15. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I., Recommender systems with social regularization. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, pp. 287–296. ACM (2011)
16. Bharadwaj, K.K., Al-Shamri, M.Y.H.: Fuzzy computational models for trust and reputation systems. Electron. Commer. Res. Appl. **8**(1), 37–47 (2009)
17. Andersen, R., Borgs, C., Chayes, J., Feige, U., Flaxman, A., Kalai, A., Mirrokni, V., Tennenholtz, M.: Trust-based recommendation systems: an axiomatic approach. In: Proceedings of the 17th International Conference on World Wide Web, pp. 199–208. ACM (2008)
18. Jsang, A., Ismail, R.: The beta reputation system. In: Proceedings of the 15th Bled Electronic Commerce Conference, vol. 5, pp. 2502–2511 (2002)
19. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans. Syst. Man Cybern. **18**(1), 183–190 (1988)
20. Pósfai, M., Kóczy.: Idf-social: an information diffusion-based fuzzy model for social recommender systems. In: The Congress on Information Technology, Computational and Experimental Physics (CITCEP 2015), pp. 106–112. (2015)
21. Guha, R., Kumar, R., Raghavan, P., Tomkins,A.: Propagation of trust and distrust. In: Proceedings of the 13th International Conference on World Wide Web, pp. 403–412. ACM (2004)
22. Kóczy, L.T., Tikk, D.: Fuzzy rendszerek. Typotex, Budapest (2000)
23. Yager, R.R.: On the measure of fuzziness and negation II. Lattices. Inf. Control **44**(3), 236–260 (1980)
24. Dombi, J.: A general class of fuzzy operators, the demorgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. Fuzzy Sets Syst. **8**(2), 149–163 (1982)
25. Zadeh, L.A.: Fuzzy sets. Inf. control **8**(3), 338–353 (1965)
26. Dubois, D.J.: Fuzzy Sets And Systems: theory And Applications. Academic pres. vol. 144 (1980)

27. Hamacher, H.: Über logische Verknüpfungen unscharfer Aussagen und deren zugehörige Bewertungsfunktionen. (1975)
28. Schweizer, B., Sklar, A.: Associative functions and abstract semigroups. Publ. Math. Debrecen **10**, 69–81 (1963)

Fuzzy System for Parameter Estimation of Complex Shape Clustering Algorithms in Predictive Diagnosis

Viorel Nicolau and Mihaela Andrei

Abstract In predictive diagnosis of a system, the operating point clusters, which characterize different functioning states, can have more complex nonspherical shapes, especially for systems functioning in outdoor industrial environments. In such cases, clustering algorithms based on potential functions, using a measure of similarity to characterize the membership of a point to a group of points, are more suitable than algorithms based on the distance of a point to the prototype vectors, such as k-means and ISODATA. Potential function-based algorithms (PFBA) make no implicit assumptions on the cluster shapes and do not use any prototype vectors of the clusters. In addition, the parameter of potential function can be used to generic characterize the shape of the clusters: more compact, oblong, or irregular. Hence, the selecting process of the function parameter values has direct influence on clustering performance, and can reduce the seeking process. In this paper, aspects of function parameter estimation using fuzzy logic are presented, so that the best clustering to be obtained with less seeking efforts. A fuzzy parameter estimator is proposed, based on intrinsic properties and clustering tendency of PFBA. Comparative simulation results are presented for oblong and irregular clusters.

Keywords Fuzzy · Clustering · Potential function · Estimation

V. Nicolau (✉) · M. Andrei
Department of Electronics and Telecommunications, "Dunarea de Jos"
University of Galati, Galati, Romania
e-mail: viorel.nicolau@ugal.ro

M. Andrei
e-mail: mihaela.andrei@ugal.ro

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_4

1 Introduction

Diagnostic and monitoring techniques can be used to optimize maintenance practices and normal operation. The diagnosis must be predictive in order to follow the evolution of the system from one mode to another one [1].

In predictive diagnosis of a system, the operating point clusters, which characterize different functioning states, can have more complex nonspherical shapes, especially for systems functioning in outdoor industrial environments.

For example, oil motor pumps working outdoor are affected by perturbations with different influences on their operating point, depending on the time horizon which is considered: long-, medium-, or short-term time period [2]. The most important factor is the temperature, which affects the oil viscosity and mechanical friction. On medium-term, during a one-year period, the mean value of normal operating point has a cyclic movement. In addition, during the life cycle, the mean value of operating point is slowly changing from normal to abnormal states. As a result, the shapes of operating point clusters characterizing functioning states can be irregular.

In such cases, clustering algorithms based on potential functions, using a measure of similarity to characterize the membership of a point to a group of points, are more suitable than algorithms based on the distance of a point to the prototype vectors, such as k-means and ISODATA.

Clustering a set of experimental data can be done in two main ways: hierarchical and partitive approaches. Density-based methods and distance-based methods are the most important classes of partitive algorithms. These methods are better than hierarchical ones in the sense that they do not depend on previously found clusters.

In general, the distance-based methods are used for unsupervised clustering problems. The methods include algorithms based on distance of a point to the prototype vectors and algorithms based on potential functions. The first type of algorithms, such as k-means and ISODATA, uses a measure of dissimilarity, which is the distance between the points of the data set and the prototype vectors. The main drawback of these algorithms is that they need prototype vectors and they try to find spherical clusters, when Euclidean distance is chosen. In addition, for better performance, the number of clusters is usually predefined.

The potential function-based algorithms (PFBA) use a measure of similarity created with a function between two points of the data set, called potential function [3]. PBFA are capable of clustering a set of data, making no implicit assumptions on the cluster shapes and without knowing in advance the number of clusters. In addition, they do not use any prototype vectors of the clusters, and the computing time is predictable, depending on the number of vectors in the data set. Although PFBA are known for a long time, they were used recently, due to their intensive computational need, which is not a real problem for modern computers.

The parameter of potential function can be used to generic characterize the shape of the clusters: more compact, oblong, or irregular. Hence, the selecting process of the function parameter has direct influence on clustering performance, and can

reduce the efforts of seeking process. It is important to understand the essence of PBFA and to generate expert rules for developing more efficient algorithms. A knowledge base can be used in selecting process of clustering parameter values or to generate a fuzzy classifier.

Fuzzy logic has proved its efficiency in system monitoring and diagnosis [4]. Fuzzy clustering techniques have been used extensively, with different algorithms such as fuzzy c-means (FCM) [5], kernel-based methods [6], fuzzy support vector clustering [7], volume criteria [8], fuzzy covariance matrix [9], and modified distance measures [10].

In this paper, aspects of function parameter estimation using fuzzy logic are presented, so that the best clustering to be obtained with less seeking efforts. A fuzzy estimator for parameter tuning of PFBA is proposed, based on intrinsic properties and clustering tendency of potential function-based algorithms. Comparative simulation results are presented for oblong and irregular operating point clusters.

The paper is organized as follows. Section 2 describes the potential function-based algorithms. In Sect. 3, selection aspects of potential function parameter are presented, based on intrinsic properties of PFBA. In Sect. 4, a fuzzy system for parameter tuning is proposed. Simulation results are illustrated in Sect. 5 and conclusions are presented in Sect. 6.

2 Potential Function-Based Clustering Algorithms

The potential function-based clustering algorithms work well for complex cluster shapes. In contrast, the algorithms based on the distance of a point to the prototype vectors, such as k-means and ISODATA, are sensitive to the cluster shapes and give good results just for spherical well-separated clusters.

Consider a data set S of N input vectors into a d-dimensional space:

$$S = \left\{ x_i | x_i = (x_{1i},\ x_{2i},\ \ldots,\ x_{di})^T \in \Re^d,\ i = \overline{1, N} \right\} \tag{1}$$

A potential function $K(x_i, x_k)$ associated with the vector $x_i \in S$ defines a positive value, called potential of the point x_i to the reference point $x_k \in R^d$. The potential depends on distance between the points x_i and x_k, denoted $d_{ik} = d(x_i, x_k)$, and it is a nonincreasing function with d_{ik}.

Two potential functions are commonly used:

$$K_1(x_i,\ x_k) = \frac{1}{1 + \alpha \cdot d_{ik}^2}, K_2(x_i,\ x_k) = \exp\left(-\alpha \cdot d_{ik}^2 \right) \tag{2}$$

where parameter α controls the slope of the function. The potential values belong to range (0, 1] and the maximum value is obtained for $d_{ik} = 0$.

The distance d_{ik} can be the general Minkovski distance:

$$d(x,y) = \sqrt[p]{\sum_{i=1}^{d} |x_i - y_i|^p},\ x, y \in \Re^d \tag{3}$$

where for p = 2 Euclidean distance is obtained, which is considered in this paper.

The function variations with d_{ik} for different values of α-parameter are illustrated in Fig. 1. The potential functions K_2 are represented with continuous lines, and K_1 are represented with dotted lines.

It can be observed that for the same constant value of α-parameter, if distances between the points are small, the two potential functions have similar values. In this case, the potential values are high and the clustering performances are similarly for both functions. Therefore, if the input data sets are normalized, the distances are smaller than unity value and the potential functions produce similar clustering performances.

A constant potential value to a reference point $x_k \in R^d$ is obtained by the potential function $K(x_i, x_k)$ associated with the points $x_i \in R^d$ for which the distance d_{ik} is constant. All points x_i generate a constant potential surface, whose shape depends on distance definition. For Euclidean distance, the constant potential surface has spherical shape around the reference point $x_k \in R^d$. The α-parameter affects the constant potential surface, different α values generating different potential surfaces, but their shapes are similar around the reference point. If α value increases, the potential surface is moving nearer to the reference point.

Similar, a potential value of a point x_i to a group of reference points $M = \{x_{k1}, x_{k2}, \ldots, x_{km}\}$ can be defined as the average of the potential values of the point x_i to all reference points x_{kj}. In this case, a constant potential value to the group M generates a potential surface, which also depends on distance definition. The constant potential surfaces surround the reference points, but their shapes are affected by α-values and reference point positions.

A potential function-based clustering algorithm uses a measure of similarity, which characterizes the membership of a point to a group of points, based on a potential function [3]. Consider a group of points M from S, $M \subset S$ and a point

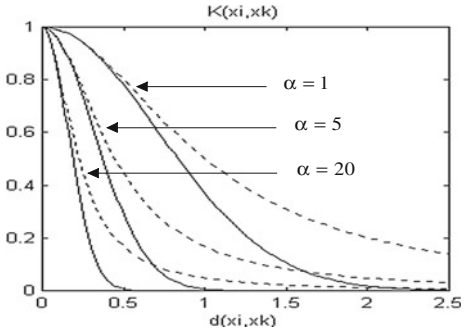


Fig. 1 Potential functions for three values of α

xi ∈ S, xi ∉ M. A similarity measure of xi to M can be defined as the average Ai of the potential values of the point xi to all points of the group M:

$$A_i = A(x_i, M) = \frac{1}{N_M} \cdot \sum_{x_j \in M} K(x_i, x_j) \tag{4}$$

where N_M represents the number of points in M.

Using this measure of similarity, the points of the data set S can be arranged in a certain order, starting from a specified starting point, pursuant to the following rule [11]:

• select the starting point, let it be x1 ∈ S, form the first group $M_1 = \{x1\}$ and denote A1 = 1, which represents the maximum potential value;

• find in S/M1 the point x2 with the maximum measure of similarity to M_1 in the meaning of (4):

$$A_2 = A\left(x^2, M_1\right) = \max_{x \in S/M_1} \left(A(x, M_1)\right) \tag{5}$$

Form a new group $M_2 = \{M_1, x2\} = \{x1, x2\}$.

• repeat the previous step until all the points of the data set S are assigned:

$$A_k = A\left(x^k, M_{k-1}\right) = \max_{x \in S/M_{k-1}} \left(A(x, M_{k-1})\right) \tag{6}$$

Form the groups $M_k = \{M_{k-1}, x^k\}$. In this way, the set S is ordered, $S = \{x^1, x^2, \ldots, x^N\}$ and a new series with N elements is obtained: $A_1, A_2, \ldots, A_N$.

All potential function-based algorithms compute the new series $A_1 \ldots A_N$, which contains the necessary information for clustering. The analysis of this series differs from algorithm to algorithm. An example with PFBA stages are presented in [12].

For example, the algorithm which is considered in this paper [11] has the following stages:

- Select the starting point. It can be arbitrarily selected;
- Arrange the points of the data set S, using the rule described above;
- Compute the ratios $R_1, \ldots, R_N$, where

$$R_1 = 1,\ R_k = \frac{A_{k-1}}{A_k},\ k = \overline{2, N} \tag{7}$$

- Compute the mean value m_R and the standard deviation σ_R of the ratios R_k;
- Consider a threshold $p = r \cdot c \cdot \sigma_R$, where r = 1…20 and c ∈ [0.3, 1];
- The clustering decision is made comparing the difference R_k-R_{k-1} with p and a new cluster begin if R_k-$R_{k-1} > p$;
- Compute new partitions for different threshold values, by increasing *r*, until $p > R_k$-R_{k-1} for all differences.

The clustering result is considered the partition, which remains unchanged for the greatest number of *r*-values.

In general, optimal clustering means partitioning a data set into a set of clusters, which minimizes distances within and maximizes distances between clusters. However, within- and between-cluster distances can be defined in several ways.

To select the best one from many partitions, a validity index can be used to evaluate them. Different validity indices can be defined, depending on which distances are considered [13]. For example, the Davies–Bouldin index uses centroid distance d_C as within-cluster, and centroid linkage D_C as between-cluster distance:

$$\frac{1}{C} \cdot \sum_{i=1}^{C} \max_{i \neq k} \left(\frac{d_c(Q_i) + d_c(Q_k)}{D_c(Q_i, Q_k)} \right) \tag{8}$$

where C is the number of clusters.

3 Selecting Process of Function Parameter

To define the distance range where the two potential functions are similarly, the relative variation between them (ΔK) is computed, by imposing a maximum value of its tolerance (tol):

$$\Delta K(d_{ik}) = \frac{K_1(d_{ik}) - K_2(d_{ik})}{K_1(d_{ik})} \leq tol \tag{9}$$

An analytical condition of (9) can be obtained if the exponential function K_2 is approximated by a rational function, such as Pade approximation. For practical situations, the approximation relative error is smaller than 1%.

Computing the inequality (9) with K_2 replaced by its Pade approximation, it results the condition for the distance d_{ik}:

$$\alpha \cdot d_{ik}^2 \leq \frac{tol + \sqrt{tol \cdot (8 + tol)}}{2}. \tag{10}$$

It can be observed that the distance range which verifies the condition (10) depends on the value of α parameter. If all the distances between the points (d_{ik}) are small enough to verify the condition (10), then the relative variation of potential functions is smaller than selected tolerance (tol) and the clustering performances are similarly for both functions.

Clustering results depend on the α-parameter value of potential functions. For optimum value of α-parameter, the distance d_{ik} corresponding to the maximum variation of potential functions must be computed. For the two potential functions, it results:

$$K_1^{''}(d_{ik}) = 0 \Rightarrow d_1 = \sqrt{\frac{1}{3 \cdot \alpha}} \Rightarrow K_1(d_1) = 0.75, \tag{11}$$

$$K_2^{''}(d_{ik}) = 0 \Rightarrow d_2 = \sqrt{\frac{1}{2 \cdot \alpha}} \Rightarrow K_2(d_2) = 0.6. \tag{12}$$

It is desired to obtain maximum variation of potential functions for most distances between the points.

If the input data set is a priori known, an optimum α-parameter value can be computed, based on average distance between data points (d_{av}), so that the sensibility of potential function with distance to be maximized. For that, the average distance is used in (11) and (12) instead of d_1 and d_2, resulting in:

$$\alpha_1 = \frac{1}{3 \cdot d_{av}^2}, \alpha_2 = \frac{1}{2 \cdot d_{av}^2} \tag{13}$$

If the input data set is not a priori known, as in many practical clustering applications, some expert rules of clustering algorithms can be used instead, based on their intrinsic properties and clustering tendency.

In addition, if a priori information about clusters in the data set is known, it can be used to adjust the parameters of clustering algorithms, so that the best clustering to be obtained faster and with less seeking efforts. The clustering process is more reliable and it can be automated.

For PFBA, the constant potential surface to a group of points M tends to take similar shape as the one of the cluster when α increases, even for more complex clusters. As potential value decreases, the shape tends to the one generated by the chosen Minkovski distance.

The influence of α values on constant potential surface is illustrated for two different α values, using the potential function K_2. Increasing α value, the constant potential surface will be closely to the cluster points and the new cluster will be oblong. Thus, the parameter α can be used to characterize the shape of the clusters: more compact or oblong.

Consider a complex cluster M with 199 points and a new point x_{200}, which has the measure of similarity to M denoted A_{200}. The value of the constant potential surface was chosen equal to A_{200}, which is useful to compare new additional points with x_{200}. For $\alpha = 25$, the constant potential value is $A_{200} = 0.055$ and the constant potential surface is illustrated with gray color in Fig. 2.

Similarly, for $\alpha = 80$, the constant potential value is $A_{200} = 0.029$ and the constant potential surface is illustrated with gray color in Fig. 3.

The points of the cluster are marked with '+' and the last point placed on the constant potential surface is marked with 'o'. Additional points placed outer potential surface have measure of similarity to M smaller than A_{200} and the points are ordered after x_{200}. By contrary, any additional point placed into potential

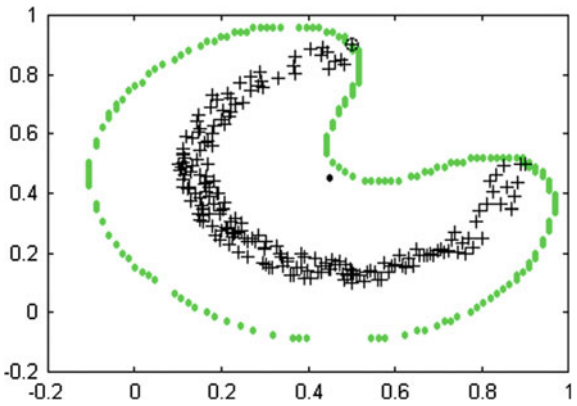


Fig. 2 Constant potential surface for $\alpha = 25$

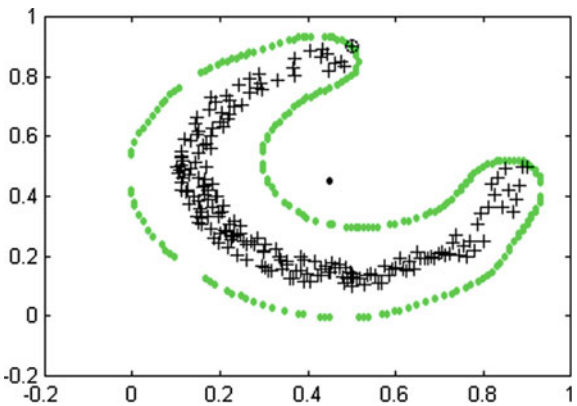


Fig. 3 Constant potential surface for $\alpha = 80$

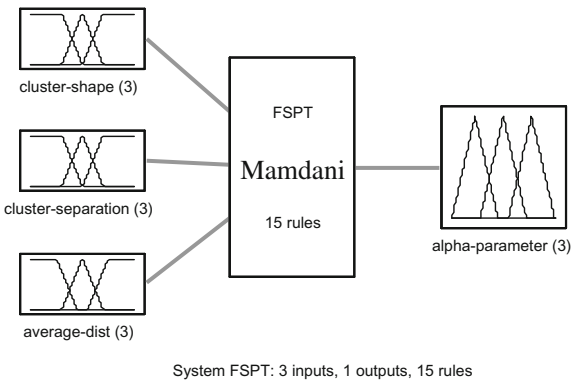


Fig. 4 Structure of fuzzy system for parameter tuning

surface is ordered before x_{200}. For example, in Figs. 2 and 3, the point at the (0.45, 0.45) coordinates is marked with '.'. This point is ordered before x_{200} if $\alpha = 25$, and it is ordered after x_{200} if $\alpha = 80$.

Concluding, the cluster boundaries depend on the α-parameter of the potential function, and the α-parameter can be used to characterize the shape of the clusters: more compact or oblong. The function parameter must be selected at the beginning of each clustering process.

4 Fuzzy System for Parameter Tuning

In this paper, a fuzzy system for α-parameter tuning (FSPT) is proposed. The proposed fuzzy system, with three inputs and one output, is represented in Fig. 4. It is a Mamdani type fuzzy system, and it generates the α parameter for PFBA. Each input of FSPT and the output have three membership functions, which are illustrated in Fig. 5.

The inputs give information about cluster shape (cluster shape), separation possibilities of clusters (cluster separation), and average distance between first data point and the other points of data set (average dist). The output of FSPT represents function parameter (alpha parameter).

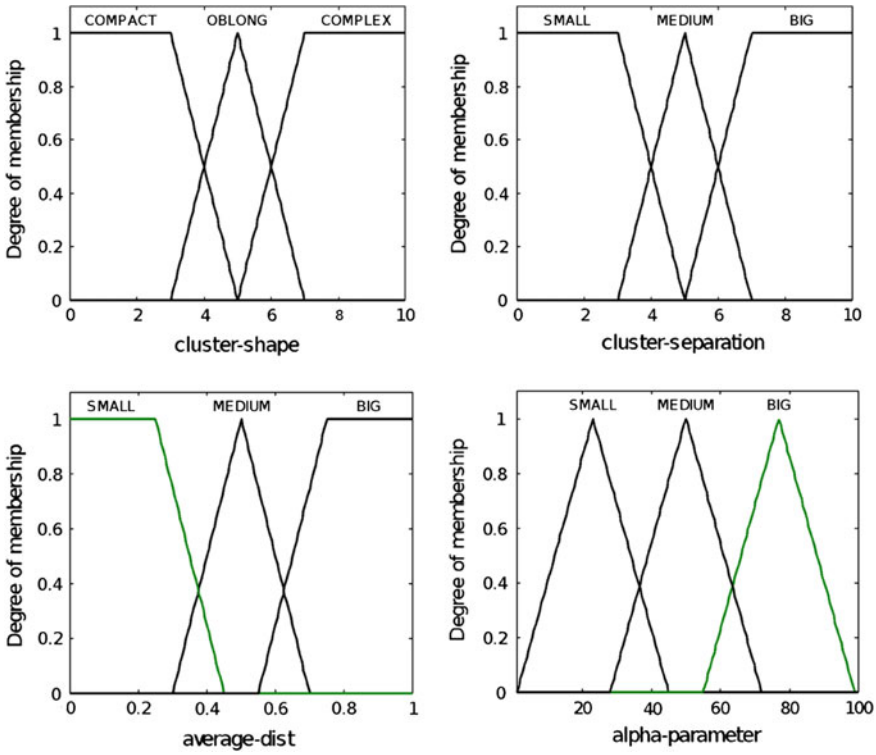


Fig. 5 Membership functions of FSPT inputs and output

The ranges of membership functions are chosen based on examples in the next section, for oblong and irregular clusters, respectively.

The first input (cluster shape) characterizes the shape of the clusters, on a scale from 0 to 10. It has three membership functions, denoted: compact, oblong, and complex.

The second input (cluster separation) characterizes how far away are the clusters to each other, on a scale from 0 to 10. It has three membership functions, denoted: small (S), medium (M) and big (B).

The third input (average dist) characterizes how far away to each other are the points of the data set. The membership functions are also denoted: small, medium, and big.

The output of FSPT (alpha parameter) indicates the α-value, on a scale from 0 to 100, with the same type of membership functions: small, medium, and big.

The knowledge base is generated using the expert rules, which describes the influence of fuzzy inputs over function parameter. The complete rule base has 27 fuzzy rules, and it is represented in Table 1.

The output of FSPT is moving on surfaces generated by combination of inputs. Projections of output surface on 2-dimensional input space with medium values of the third input are represented in Fig. 6.

5 Simulation Results

The potential function-based algorithms work well for complex cluster shapes. By contrast, the algorithms based on distance to the prototype vectors are sensitive to the cluster shapes and give good results just for spherical, well-separated and comparable in dimension clusters.

Without FSPT, the best clustering can be obtained by running the potential function-based clustering algorithm several times, with different values of α-parameter.

Table 1 Rule base of the fuzzy system

Cluster-separation	Cluster shape			Average dist
	Compact	Oblong	Complex	
Small	B	B	B	Small
	B	B	B	Medium
	M	M	B	Big
Medium	M	M	B	Small
	M	M	M	Medium
	S	M	M	Big
Big	S	M	M	Small
	S	S	S	Medium
	S	S	S	Big

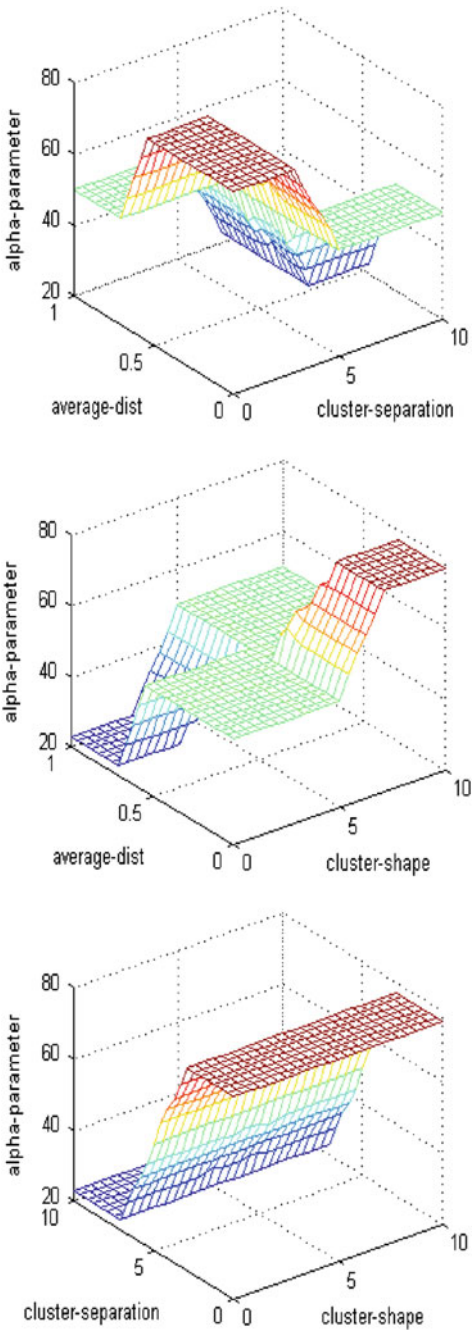


Fig. 6 Projections of FSPT output surface on 2-dimensional input space

By using fuzzy system for α-parameter estimation, a good clustering is obtained by running PFBA only once. In this way, the solution may not be the optimal one, but it is obtained faster and with less seeking efforts.

To illustrate the importance of parameter selection, different data sets are used in clustering process with PFBA. The data sets are generated in 2-dimensional space with normalized input vectors. Knowing a priori information about clusters in data set, the best clustering can be obtained with less seeking efforts. The cluster boundaries depend on α value, which is generated by FSPT.

First data set contains 150 points, which are arranged into two oblong clusters with parallel main directions, and it is illustrated in Fig. 7. The clusters have 50 and 100 points, which are marked with '+' and 'o', respectively.

Using the FSPT, the α-parameter value is computed, resulting $\alpha = 60$. Applying the value in potential function K2 and running the potential function-based clustering algorithm, the ordered set of the original data set I is obtained, as illustrated in Fig. 8. The ordered data set is obtained starting arbitrary from the point x_{70} in the original data set I, which is marked distinctly.

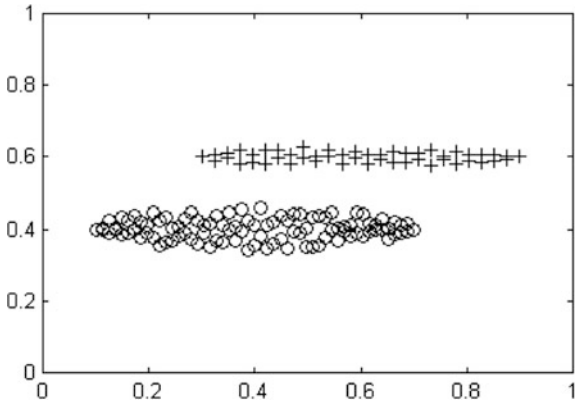


Fig. 7 Data set I: two oblong clusters

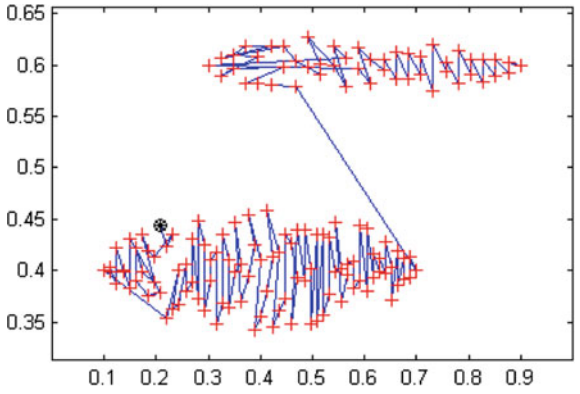


Fig. 8 The ordered data set I, starting from x_{70}

In Fig. 8, the points are marked with '+'. In addition, lines are drawn between every two consecutive points in the ordered data set, to highlight the arrangement tendency of the points. It can be observed that the points are ordered in successive layers around the first-ordered points.

After analyzing the A_k series, the clusters are well identified with PFBA, as shown in Fig. 9.

The cluster boundaries are illustrated, as constant potential surface. They correspond to constant potential surfaces of last point in the ordered data set in each cluster. It can be observed that the clusters are well identified along with their boundaries. In addition, the computed separation line between clusters is represented. It contains points with the same potential value to the clusters, which are determined for the input data space.

Without fuzzy system parameter tuning, the potential function-based clustering algorithm must be run many times, by manually selecting of α-parameter value. The best clustering results are obtained empirically.

The clustering results of data set I, for manual selection of α-parameter using 10 different values, are represented in Table 2. For every value of α-parameter, the final number of clusters (correct or erroneous) considered by PFBA to be the best results is represented, along with their probability of good clustering.

By using fuzzy system for α-parameter estimation, the FSPT result is $\alpha = 60$, and a good clustering is obtained by running PFBA only once, marked in gray color in Table 2.

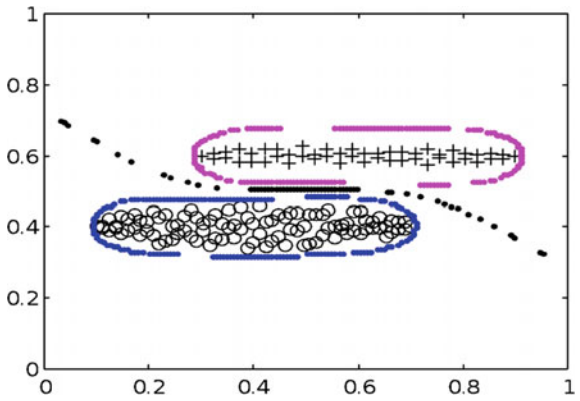


Fig. 9 Clustering results using PFBA for data set I

Table 2 Manual clustering results of data set I

Data set I—manual selection of α-parameter										
α	10	20	30	40	50	60	70	80	90	100
No. of clusters	3-err	2-err	2	2	2	2	2	2	2	2
Probability	0.09	0.15	0.25	0.68	0.82	0.87	0.88	0.91	0.92	0.92

Using ISODATA, the clusters cannot be well identified, because the algorithm attempts to separate spherical clusters. The best clustering results are illustrated in Fig. 10.

To avoid point mixture, the cluster dimensions must be reduced. As a result, the number of clusters is bigger. It can be observed that the number of cluster is five to avoid clusters with points from the two original ones. The points from first cluster are considered included into three smaller clusters, and for the second cluster, the points are included into two smaller clusters. It should be mentioned that the best clustering is obtained empirically, by running the algorithm many times with different values of parameters.

The second data set contains two irregular clusters, which are more difficult to separate due to their positions, as shown in Fig. 11. In data set II, the clusters have also 50 and 100 points, which are marked with '+' and 'o', respectively.

Using the FSPT, the α-parameter value is computed, resulting $\alpha = 80$. Applying the value in potential function K2 and running the potential function-based clustering algorithm, the ordered set of the original data set II is obtained, as illustrated

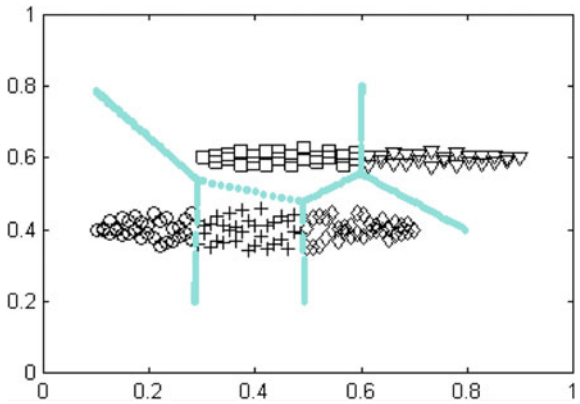


Fig. 10 Best clustering of data set I using ISODATA

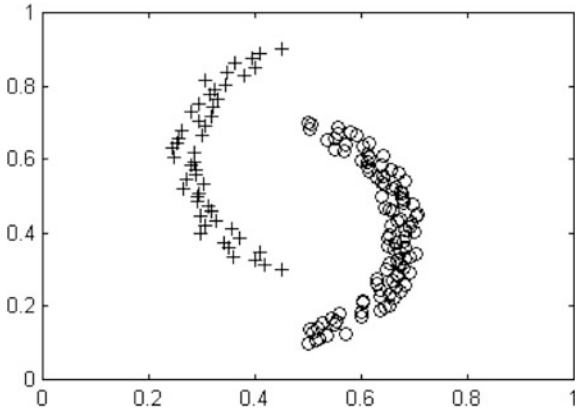


Fig. 11 Data set II: two irregular clusters

in Fig. 12. The ordered data set is obtained starting from the first point x_1, which is marked distinctly.

Again, after analyzing the A_k series, the clusters are well identified with PFBA, as shown in Fig. 13. The cluster boundaries are illustrated, as constant potential surface. It can be observed that the clusters are well identified along with their boundaries. Also, the computed separation line between clusters is represented.

Without fuzzy system parameter tuning, the best clustering results are obtained empirically. The clustering results of data set II for manual selection of α-parameter using 10 different values are represented in Table 3. For every value of α-parameter, the final number of clusters (correct or erroneous) considered by PFBA to be the best results is represented, along with their probability of good clustering.

By using fuzzy system for α-parameter estimation, the FSPT result is $\alpha = 80$. Therefore, a good clustering is obtained by running potential function-based clustering algorithm only once, marked in gray color in Table 3.

Using ISODATA, the clusters cannot be well identified, because the algorithm attempts to separate spherical clusters. To avoid point mixture, the cluster

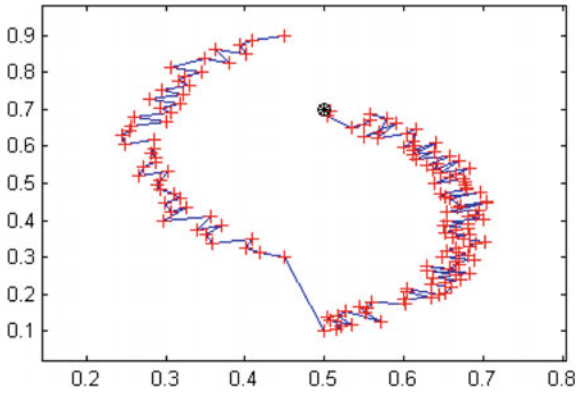


Fig. 12 The ordered data set II, starting from x_1

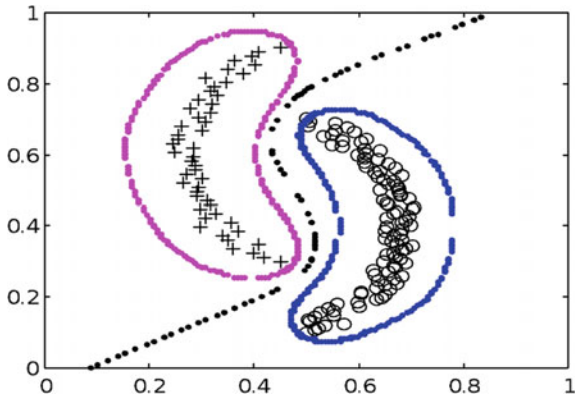


Fig. 13 Clustering results using PFBA with α-parameter estimated by FSPT

Table 3 Manual clustering results of data set II

Data set II—manual selection of α-parameter										
α	10	20	30	40	50	60	70	80	90	100
No. of clusters	4-err	3-err	2-err	2-err	2	2	2	2	2	2
Probability	0.09	0.15	0.23	0.25	0.33	0.65	0.80	0.86	0.91	0.92

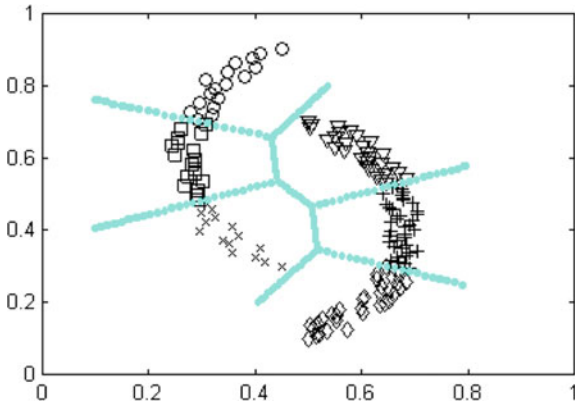


Fig. 14 Best clustering results using ISODATA

dimensions must be reduced. As a result, the number of clusters is bigger. The best clustering results are illustrated in Fig. 14.

It can be observed that the number of cluster is six to avoid clusters with mixed points from both original clusters. The points from every original cluster are considered included into three smaller clusters. Again, the best clustering is obtained empirically. As the algorithm is running, in every stage, the prototype vectors of the considered clusters are shifted dynamically in the input data space.

An application example of using PFBA for complex clustering tasks is the printed circuit board (PCB) quality control, for online diagnosis of good or faulty PCBs copper tracks.

In quality control of PCBs, an automated system based on video inspection should be able to identify different states of good or faulty PCBs [14]. Copper tracks have complex shapes, which are close to each other, making analyzing process more difficult. In addition, clustering techniques must deal with unwanted copper material between conductive tracks, which affects dynamic behavior of the circuit.

Simulation results are presented for two circuit segments with parallel main directions, obtained from image processing system. They are two oblong clusters of copper tracks, which are not well separated. Simulations take into account also the measurement noise, referring to camera resolution, segmentation performance, etc.

A faulty PCB is analyzed, with some copper material placed between the tracks, which is close enough but contactless with them. PFBA are able to identify the complex shapes of copper tracks in good PCBs, and also work well in heavy

classifying process of the unwanted copper clusters placed very close to conductive tracks.

The data set III refers to a faulty PCB. It contains two oblong clusters of copper tracks with a faulty drop of copper material, between but contactless with the conductive tracks, as shown in Fig. 15. The goal is to correctly identify the unwanted drop of copper material with an automated inspection system.

The data set III has 520 points. The two oblong clusters of copper tracks have 250 points each. The third cluster has 20 points, and it is close enough but contactless with the each copper track.

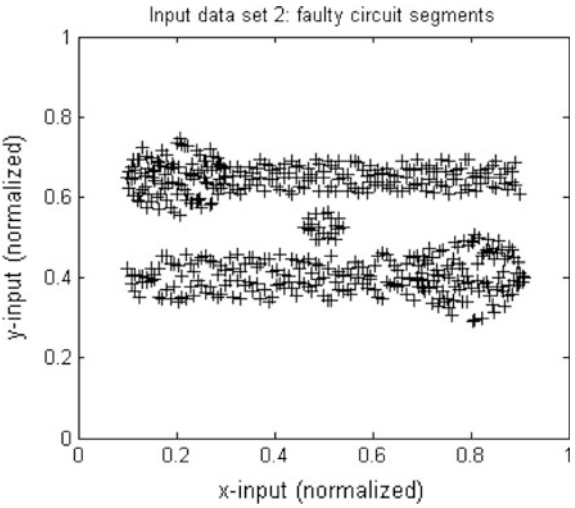


Fig. 15 Data set III with a faulty drop of copper material

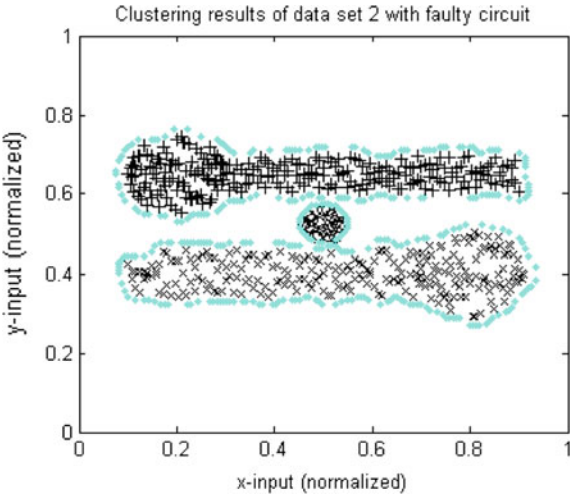


Fig. 16 Clustering results of data set III using FSPT and PFBA

The clustering process for the data set III is far more complicated than the first two data sets, so that the membership functions of the output of FSPT should have much wider value ranges. For this case, the maximum value for FSPT output was selected 1000.

Using fuzzy system for α-parameter estimation, the FSPT result is $\alpha = 700$. The parameter of potential function is selected with big value, so that the constant potential surfaces to be close to cluster points. With this value, the PFBA determines the correct number of clusters, and cluster points are well identified, by running the clustering algorithm only once. The results include three different clusters, which are illustrated in Fig. 16. Their points are marked with '+', 'x', and 'o'. Also, the cluster boundaries, as constant potential surfaces, are represented with gray color.

6 Conclusions

The algorithms based on potential functions are intensive computational. The selecting process of clustering parameter values is very important, having direct influence on clustering performance. A knowledge base is generated and a fuzzy system is proposed for parameter tuning of PFBA, so that the best clustering to be obtained with less seeking efforts. The fuzzy system gives good results, according with cluster shapes. The complex shape clusters are well identified along with their boundaries by running PFBA only once. An application example of using FSPT and PFBA is presented for automated video inspection and online diagnosis of good or faulty of PCBs copper tracks.

References

1. Bouguelid, M.S., Mouchaweh, M.S., Billaudel, P.: Adaptive and predictive diagnosis based on pattern recognition. 11th International Conference on Intelligent Engineering Systems (INES 2007), pp. 139−144.DOI:10.1109/INES.2007.4283687. 2007
2. Nicolau, V.: Fuzzy diagnostic system for oleo-pneumatic drive mechanism of high-voltage circuit breakers. In: the Scientific World Journal, vol. 2013, Article ID: 248487, Hindawi, ISSN: 1537-744X, DOI:10.1155/2013/248487, WOS:000327148000001, (2013)
3. Dorofeyuk, A.A.: Algorithms of teaching the machine the pattern recognition without teacher based on the method of potential functions. Avtomatika i Telemechanica **10**, 78–87 (1966)
4. Peltier, M.A., Dubuisson, B.: A fuzzy monitoring and diagnosis process to detect evolutions of a car driver's behaviour. 2nd IFAC Symposium SICICA 1994, pp. 340–345, Budapesta, (1994)
5. Bezdek, J., Keller, J., Krishnapuram, R., Pal, T.: Fuzzy models and algorithms for pattern recognition and image processing. Kluwer, Norwell, MA (1999)
6. Girolami, M.: Mercer kernel-based clustering in feature space. IEEE Trans. Neural Networks **13**, 780–784 (2002)

7. Lin, C.-F., Wang, S.-D.: Fuzzy support vector machines. IEEE Trans. Neural Networks **13**, 464–471 (2002)
8. Krishnapuram, R., Kim, J.: Clustering algorithms based on volume criteria. IEEE Trans. Fuzzy Syst. **8**, 228–236 (2000)
9. Gustafson, G., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix, pp. 761–766. IEEE Conf. Decision, Control (1979)
10. Klawonn, F., Keller, A.: Fuzzy clustering based on modified distance measures, 3rd Int. Symp. Adv. Intell. Data Anal. **1642**, 291–301 (1999)
11. Bumbaru, S., Ceanga, E., Bivol, I.: Self-learning classifier for data analysis. Proceedings of the VI-th Iugoslav International Symposium on Information Processing, vol. H5, pp. 1–6. Iugoslavia, (1970)
12. Nicolau, V., Ceangă, E., Puşcaşu, G.: Advantages of partitive clustering algorithms based on potential functions. 13th European Simulation Symposium, pp. 675–679, Marseille, ISBN 90-77039-02-3 (2001)
13. Bezdek, J.C.: Some new indexes of cluster validity. IEEE Trans. Syst. Cybern. **28**:301–315 (1998)
14. Zhang, F., Luk T.: A data mining algorithm for monitoring PCB assembly quality. IEEE Trans. on Electron. Packag. Manuf. **3**(4):299–305, ISSN: 1521-334X, 2007

Merging Validity and Coverage for Measuring Quality of Data Summaries

Miroslav Hudec

Abstract Data summarization by quantified sentences of natural language simulates human reasoning in summing up from the data. Linguistic summaries are focused either on a whole data set, or on a part of a data set delimited by the flexible restrictions expressed as fuzzy sets. First, the paper examines influences of t-norms in compound predicates merged by the *and* connective and constructed fuzzy sets on the validity (truth value) of summaries. Further, linguistic summaries with restriction may express mined knowledge from the outliers and therefore be of low quality, even though the validity of summary could be high. The main aim of this paper is building a quality measure based on validity and coverage. Finally, additional possibilities related to the suggested measure and perspective topics for future research are outlined.

Keywords Linguistic summaries · Validity · Quality · Outliers · T-norms · Fuzzy sets

1 Introduction

Nowadays, mining summarized information from data sets is a topic of interest for researchers and practitioners. Data summarization can be efficiently realized by statistical methods which however, are understandable for rather small group of specialists. This observation is expressed in [1] as: "summarization would be especially practicable if it could provide us with summaries that are not as terse as the mean". Graphical interpretation is a valuable way of summarization but cannot be always effective [2]. Linguistics is an interesting alternative when data is hard to show graphically [3]. A linguistically summarized sentence can be read out by a text-to-speech synthesis system. It especially holds when the visual attention should

M. Hudec (✉)
Faculty of Economic Informatics, University of Economics in Bratislava, Bratislava, Slovakia
e-mail: miroslav.hudec@euba.sk

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*, Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_5

not be disturbed [4]. These advantages hold when the resulting summarization is of a high quality.

People tend to summarize by terms of natural language. But, literally unlimited variations of linguistic terms and their modifications for expressing summaries exist. In order to put together mathematical formalization and people's preferred way, quantified sentences of natural language, i.e. Linguistic Summaries (LSs) were introduced in [5]. Since then LSs have been intensively researched in, e.g. [6–15].

Generally, LSs summarize the whole data set or a restricted part. In the former, LSs are of the structure *Q entities are (have) S*, where *Q* is a quantifier, and *S* is a summarizer. One example of such a summary is: most of houses have high gas consumption. In the latter, LSs are of the structure *Q R entities are (have) S*, where *R* puts some restriction on data sets. One example of such a summary is: most of old houses have high gas consumption. In addition, *R* and *S* can be consisted of several atomic predicates merged by the *and* connective [7, 8] which is usually modelled by t-norms [16]. The truth value of LSs (also called validity) gets value form the [0, 1] interval by agreement. Hence, validity is influenced by selected t-norm and constructed fuzzy sets.

LSs with restriction may be trapped into outliers due to possible very low coverage of tuples in *R* and *S* parts, even though the validity is high. Hence, this problem of the LSs quality should not be neglected. Hirota and Pedrycz [17] suggested five quality measures: validity, generality, usefulness, simplicity and novelty. These measures are further examined for LSs with the restriction part in [15] for the purpose of converting mined summaries into fuzzy rules. Further set of measures was introduced in [18, 19].

The main goal of this paper is focused on building outlier measure expressed by coverage and validity [15, 17]. Preliminary results in this direction were published in [20]. Furthermore, this paper extends discussion to the influence of t-norms and fuzzy sets to the validity of LSs. The reminder of this chapter is organized as follows. Section 2 gives some preliminaries of LSs which are used as a basis for the next sections. In Sect. 3 influences of different t-norms in *R* and *S* parts on validity are examined. Impact of constructed fuzzy sets is examined in Sect. 4. Section 5 is devoted to building a new quality measure related to outliers, discussion supported by illustrative example and future challenges. Section 6 gives a short note to different applications. Finally, Sect. 7 concludes this work.

2 Linguistic Summaries in Brief

LSs summarize knowledge from the data into the concise and easily understandable way for people. LS for summarizing the whole data set is of the structure *Q entities in database are (have) S*, where *Q* is a relative quantifier and *S* is a summarizer. Both are expressed by linguistic terms (fuzzy sets). The validity of summary is computed in the following way [5]:

$$v(Qx(Px)) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_S(x_i)\right) \tag{1}$$

where n is the number of tuples in a data set (cardinality), $\frac{1}{n}\sum_{i=1}^{n}\mu_S(x_i)$is the proportion of tuples in a data set that satisfy predicate S and μ_Q is the membership function of chosen relative quantifier.

LS with restriction has the form *Q R entities in database are (have) S*, where R is a restriction (expressed by fuzzy set) focusing on a part of data set relevant for the summarization task. The validity is computed in the following way [14]:

$$v(Qx(Px)) = \mu_Q\frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{\sum_{i=1}^{n}(\mu_R(x_i)} \tag{2}$$

where $\frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{\sum_{i=1}^{n}\mu_R(x_i)}$ is the proportion of tuples in a data set that satisfy S and belong to R, t is a t-norm and μ_Q is the membership function of chosen relative quantifier.

Linguistic terms such as *medium* (*around m*), *small* and *high* used in S and R can be expressed by triangular or trapezoidal fuzzy sets, L fuzzy set and linear gamma fuzzy set consequently (Fig. 1) ensuring the smooth transition between relevant and non-relevant tuples.

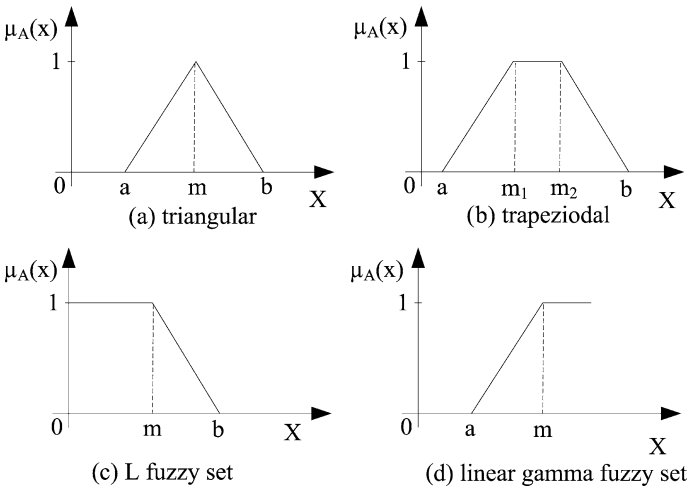


Fig. 1 Fuzzy sets for restrictions and summarizers

Summarizer and restriction may contain several atomic predicates merged by the *and* connective [7, 8]. These connectives are usually modelled by t-norms [16]. Four basic t-norms are:

- minimum t–norm:

$$t_m(\mu_{A_1}(x), \mu_{A_2}(x)) = \min(\mu_{A_1}(x),\ \mu_{A_2}(x)) \tag{3}$$

- product t-norm:

$$t_p(\mu_{A_1}(x), \mu_{A_2}(x)) = \mu_{A_1}(x) \cdot \mu_{A_2}(x) \tag{4}$$

- Łukasiewicz t-norm:

$$t_L(\mu_{A_1}(x), \mu_{A_2}(x)) = \max(\mu_{A_1}(x),\ \mu_{A_2}(x) - 1,\ 0) \tag{5}$$

- drastic product

$$t_d(\mu_{A_1}(x), \mu_{A_2}(x)) = \begin{cases} \min(\mu_{A_1}(x), \mu_{A_2}(x)) & \max(\mu_{A_1}(x), \mu_{A_2}(x)) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\mu_{Aj}(x)$ ($j = 1, 2$) denotes the membership degree to the j-th fuzzy set for element x.

The validity of LS is computed by the relative quantifiers such as *few, about, half, most of*. The *most of* quantifier, plotted in Fig. 2, is often used because users are interested to see which summaries are met by the majority of tuples.

We can say that the linguistic summary is

a more or less accurate textual description (summary) of a data set

This simple definition hides many challenges: construction of fuzzy sets for summarizers, restrictions and quantifiers, selecting appropriate t-norms, sufficient coverage of data, simplicity, usefulness, accuracy, summarizing from the outliers instead from the regular data and the like. The influences of t-norms, construction of fuzzy sets, coverage and outliers to quality of LSs are examined in the next sections.

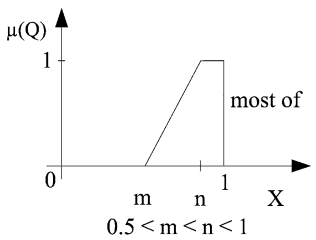


Fig. 2 Relative quantifier *most of*

3 Impact of T-Norms in Restriction and Summarizer to Validity

For LSs with restriction the quality is measured for each data point x_i ($i = 1, \ldots, n$) by t-norm in the numerator of (2) [2]. This section is focused on searching for suitable t-norms not only for merging restriction and summarizer but also for conjunction of atomic predicates inside restriction and summarizer. Two examples of such queries are: most high polluted and low situated (altitude) municipalities have a high number of respiratory diseases, and most middle aged customers have high turnover and small payment delays.

When restriction or summarizer consists of several atomic conditions (predicates P) connected by the *and* operator, t-norms come to the stage. All t-norms meet all axiomatic properties explained in e.g. [22], but differ in satisfying algebraic properties. Let us recall the following three algebraic properties [16]:

- The t-norm is an idempotent one if for $\forall a \in [0, 1],\ t(a, a) = a$
- The t-norm is a nilpotent one if there exists some $n \in N$ such that $t^{(n)}(a) = 0$
- The t-norm has a limit property if for $\forall a \in (0, 1),\ \lim_{n \to \infty} t^{(n)}(a) = 0$

LSs express proportion of tuples which meet atomic or compound predicate in S and/or R. For instance, when each atomic predicate P_j ($j = 1, \ldots, n$) is satisfied with degree of 0.48, then the tuple should participate in S with degree of 0.48. This requirement meets idempotent t-norm. The only idempotent t-norm is the minimum one (3). Furthermore, this t-norm is not nilpotent and does not have limit property. Łukasiewicz t-norm (5) meets the second property causing that tuple participates in proportion with value of 0. Product t-norm (4) meets third property causing decreasing tuples participation in the proportion, when the number of atomic predicates increases. When $j = 2$, tuple participates with degree of 0.2304; but when $j = 4$, tuple participates in summary with degree of 0.05308.

For the basic structure of LSs (1) when S is a compound predicate selecting the suitable t-norm is a pivotal task for obtaining LS of a high quality. Concerning the LS with restriction (2), selecting appropriate t-norm influences quality but further quality aspects should be considered.

To summarize, the only suitable t-norm is the minimum one (3), because it does not unnaturally reduce the proportion of tuples in a data set that satisfy LS. Interestingly, in the Sect. 5 the situation regarding suitable t-norms is opposite.

4 Influence of Constructed Fuzzy Sets to Validity and Coverage

The subjectivity in constructing fuzzy sets may influence quality of summarized information. It especially holds for the sufficient coverage and outliers which are examined later on.

The domains of attributes are, during the database design phase, defined in a way that all theoretically possible values can be stored. For instance, for the attribute monitoring frequency of an activity during a year the domain is the [0, 365] interval of integers. In practice, collected values can be far from the lower and upper limits of the domain. In the constructing fuzzy sets this fact should be considered [23], because users are not always aware of collected attributes' values. The situation plotted in Fig. 3a, where *L* and *H* are the lowest and the highest values in the current content of attributes, respectively, and *Dmin* and *Dmax* are the lower and upper limit of domains, respectively, might appear. The truth value equal to 1 in Fig. 3a may express summary on outliers and therefore, is of a low quality.

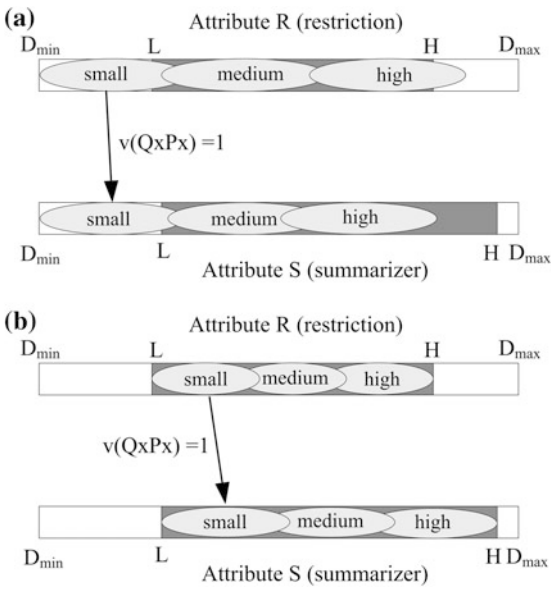


Fig. 3 Fuzzy sets for restriction and summarizer: **a** fuzzy sets do not reflect stored data; **b** fuzzy sets reflect stored data

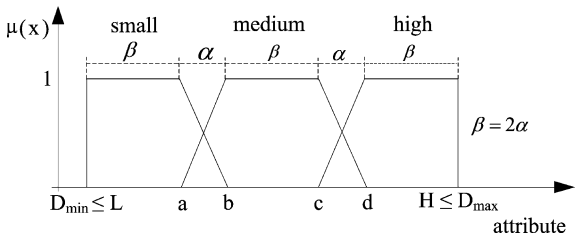


Fig. 4 Fuzzy sets uniformly distributed in the part of attribute domain covered by data

Moreover, one should be very careful when no tuple meets the *R* part because it leads to dividing by zero in (2).

In order to mitigate this problem, we should construct membership functions considering only parts of domains that contain data [23]. The validity equal to 1 in Fig. 3b can be relevant summary. But it does not hold automatically.

The shapes of membership functions have adopted several conventions [24]. Generally, the membership functions are convex and normalized piecewise linear functions. Figure 4 shows the situation where the family of fuzzy sets consists of three sets to cover terms *small*, *medium* and *high*. The flat segments of these fuzzy sets (β) express no uncertainty in belonging to sets, whereas parameter α expresses the uncertainty in belonging to a set. When $\alpha = 0$, the domain is partitioned into crisp sets. If a requirement for finer granulation exists, more fuzzy sets (e.g. five sets: very small, small, medium, high, very high) can be straightforwardly constructed adjusting parameters α and β. These concepts can be defined by nonlinear functions as well. Concerning practical applications and the simplicity for end users, linear functions are often preferable.

Even though fuzzy sets are constructed on parts of domains where data are recorded, the data distribution far from the uniform one might cause that LSs express relations detected in outliers. For example, let only 20 of $5 \cdot 10^6$ tuples fully meet the *R* and the same tuples fully meet the *S*, then the validity (2) gets the value of 1, leading us to the false conclusion.

5 Quality Measure Focused on Outliers and Coverage

Keeping the aforementioned in mind, we can say that if LSs with restriction have high validity *v* (2), it does not straightforwardly mean that these LSs are suitable for expressing summarized information, even though suitable t-norm is applied and care was taken during the construction of fuzzy sets. Thus, quality measures should be applied in order to mitigate vagueness of calculated validity. Five quality measures: validity, generality, usefulness, novelty and simplicity were suggested in [17] and further examined in [15]. Four measures: coverage, brevity (or shortness), specificity and accuracy mainly for non-quantified linguistic summaries are examined in [18]. All these measures get values from the [0, 1] interval.

The novelty measure means that unexpected summaries represent valuable knowledge, if they do not express knowledge mined from the outliers [15] (errors in observations or existence of few very different tuples). Therefore, for calculating the novelty measure outliers should be recognized and measured. Furthermore, outliers and coverage are related. The outlier's measure is examined in this section.

5.1 Outliers

Wu et al. [15] explained that outliers appear if the validity degree v is very small or very high and the sufficient coverage C must be very small. Therefore this measure can be expressed as

$$O = \min(\max(v, 1-v), (1-C)) \tag{7}$$

where C is the coverage, which is defined later. If coverage is small ($C \rightarrow 0$), then outlier measure O is near the value of 1 (if v gets value near 1 or 0). If coverage is high ($C \rightarrow 1$), then the outlier measure is near the value of 0. In a general way (7) can be expressed as:

$$O = t(s(v, 1-v), (1-C)) \tag{8}$$

where t is a t-norm and s is a s-norm.

The non-outlier measure is calculated as the negation of (8) by De Morgan's law, i.e.:

$$1 - O = s(t(1-v, v), C) \tag{9}$$

when the standard fuzzy negation is used.

We can say that LSs are of a high quality if validity and non-outliers are high. This observation is formally written as

$$Q_c = t(v, 1-O) = t(v, s(t(1-v, v), C)) \tag{10}$$

From the properties of t-norms holds: $t(1 - v, v) \leq 0.5$. If we define quality as significant, when coverage is higher or equal 0.5, then from (10) yields:

$$Q_c = \begin{cases} t(v, C) & C \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $C = 0.5$ is considered as a threshold value of coverage.

The next task is calculating coverage in a way that it meets the requirement (11).

Coverage

Concerning the basic structure of LS (1), the whole data set is covered due to appearance of the variable n (cardinality of a data set) in the denominator, i.e. coverage is implicitly calculated. If the coverage is low, then it directly influences the validity. Regarding the LS with restriction (2), coverage should be calculated explicitly. The following coverage index for LSs of structure (2) is created [25]:

$$i_C = \frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{n} \tag{12}$$

where n is the number of tuples in a data set. Other variables have the same meaning as in (2). The coverage index i_c explains how many records' membership degrees influence the validity of a LS. In practice, the coverage index is a small number, because LSs with restriction usually cover relatively small subset of the considered data set [15]. Therefore, the mapping which converts i_c (12) into the coverage C (used in (7–11)) yields [15]:

$$C = f(i_C) = \begin{cases} 0, & i_c \le r_1 \\ 2\left(\frac{i_c - r_1}{r_2 - r_1}\right)^2, & r_1 \le i_c < \frac{r_1 + r_2}{2} \\ 1 - 2\left(\frac{r_2 - i_c}{r_2 - r_1}\right)^2, & \frac{r_1 + r_2}{2} \le i_c < r_2 \\ 1, & i_c \ge r_2 \end{cases} \tag{13}$$

where $r_1 = 0.02$ and $r_2 = 0.15$. Anyway, parameters r_1 and r_2 can be set according to user preferences in a same way as for other fuzzy sets: for S and R (Figs. 1 and 4) and quantifiers (Fig. 2). When R is more restrictive, i.e. several atomic predicates merged by the *and* connective, then parameters r_1 and r_2 can be smaller.

Naturally, the question which t-norm in (11) is the suitable one appears. Let us have calculated values of validity and coverage for two LSs shown in Table 1.

The minimum t-norm (3) says that *ls1* and *ls2* are indistinguishable. Hence, we need t-norm which considers all attributes, not only attributes bearing minimal value. The solution provides product t-norm (4) stating that *ls1* is of higher quality than *ls2* (Table 1). It is the opposite observation than for aggregating atomic predicates by the *and* connective in S (1), (2) and R (2) parts of LSs (Sect. 3). Instead of product t-norm, we can apply another non-idempotent t-norm: a Łukasiewicz one, but it further decreases measure (11).

Table 1 Merging validity and coverage of LSs by product and minimum t-norms in (11)

LS	Validity	Coverage	$Q_c(v, C)$ by (4)	$Q_c(v, C)$ by (3)
ls1	0.75	0.95	0.7125	0.75
ls2	0.75	0.75	0.5625	0.75

The quality measure regarding the outliers can be also expressed as

$$Q_c = f(v, 1 - C) \tag{14}$$

In this case, we do not consider coverage (13) but its negation. This equation represents a bipolar relation because v is a positive predicate and $(1 - C)$ is a negative one.

Furthermore, if the requirement for a high quality is the full coverage (13), i.e. $C = 1$, then the non-continuous drastic t-norm (6) is a rational option for merging validity and coverage:

$$Q_c = d_p(v, C) = \begin{cases} v, & C = 1 \\ C, & v = 1 \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

The validity of the LS is taken into account only if $C = 1$. All summaries which pass this filter can be ranked downwards form the best one according to the validity degree. Summaries are evaluated by their respective validities, as is the case in (2), but only when they pass this simple filter.

Illustrative Example

In order to mine all relevant summaries, user has defined set of attributes, quantifiers and linguistic terms for attributes appearing in restriction and summarizer. For simplicity, mined LSs are written as *ls1* (i=1,…,9) and shown in Table 2.

When coverage is fully satisfied the same result is obtained by (11) and (15). The latter is a filter and expressed as first meet coverage and then validity. This approach is suitable when coverage is a sharp condition. Otherwise, product t-norm is the option. The drawback of drastic product is in its sharpness. When both validity and coverage are close to 1, the result is 0.

However, in measuring quality by drastic product (15) we have the second option: when $v = 1$, the result is C (the last record in Table 2). This option may be either excluded or used as an alternative: if validity is fully satisfied, then preferable summary is one with higher coverage degree.

Table 2 Quality of mined LSs

LS	Validity	Coverage	$Q_c(v, C)$ (11) by product t-norm	Q_c (15)
ls1	0.80	0.80	0.6400	0.00
ls2	0.75	0.85	0.6375	0.00
ls3	0.65	1.00	0.6500	0.65
ls4	0.93	1.00	0.9300	0.93
ls5	0.81	0.24	0.0000	0.00
ls6	0.12	1.00	0.1200	0.12
ls7	0.23	0.14	0.0000	0.00
ls8	0.95	0.95	0.9025	0.00
ls9	1.00	0.58	0.5800	0.58

Mining LSs from the Data

Generally, two ways for mining summaries exist:

- User defines all relevant linguistic terms for quantifiers, restrictions and summarizers and all attributes of interest.
- User defines term sets for quantifiers, summarizers and restrictions without selecting relevant attributes.

In both cases an application reveals all summaries for which validity (1) or (2) is higher than 0, or higher than the defined threshold value. The difference is in mined summaries. In the first way, only summaries of a clear interest are mined. In the next step quality measure (11) can be applied. In the second way, the usefulness of mined summaries is a further measure which should be considered, i.e. high validity and coverage of a quantified sentence: most territorial units with high percentage of public greenery have small unemployment, presumably is irrelevant for analysing reasons for high unemployment and building related rule base.

Some Perspectives for Further Research

The first perspective is aggregating quality measures mentioned in [15, 18] and measure suggested in this work. But it is not an easy task because we need to aggregate several measures which may be partially redundant and conflicting [25].

For instance, the simplicity measure [15] concerns the syntactic and semantic complexity of the LSs. This measure expresses how many attributes in restriction and summarizer in a summary exist. Complex summaries are less legible for users. Hence, the simplicity measure can be expressed as [15]:

$$S_{im} = 2^{2-l} \tag{16}$$

where l is a total number of atomic predicates in restriction and summarizer. Evidently, Sim gets values from the unit interval. The example of a summary having $Sim = 1$ is: most young customers have a small payment delay.

Regarding the basic structure of LS (1), Eq. (16) yields:

$$S_{im} = 2^{1-l} \tag{17}$$

ensuring that the simplest structure (one atomic predicate inside the summarizer, e.g. most customers are middle aged) has simplicity equal to 1.

The second perspective is the focus on quantified restrictions and summarizers. A structure of LSs with restriction (2) can be also expressed as

$$Q\left(\wedge_{i=1}^{n} R_i(x) \right) \text{are} \left(\wedge_{j=1}^{m} S_j(x) \right) \tag{18}$$

where R_i and S_j are atomic predicates in restriction and summarizer consequently.

When $i = j = 1$ we obtained the structure frequently examined in the literature. It is obvious that when n and m are larger numbers the sentence becomes very restrictive. The structure (18) can be relaxed to the following structure:

$$Q\left(\text{most of } R_i(x), i = 1, \ldots, n\right)\text{are}\left(\text{most of } S_j(x), j = 1, \ldots, m\right) \tag{19}$$

This structure corresponds with the structure of quantified queries [26], where tuples which meet the majority of atomic predicates are selected.

The benefit is a less restrictive summary concerning all atomic predicates. A tuple which meets four atomic predicates with degrees 0.2, 0.1, 0.25, 0.2 has a lower impact than a tuple which meets these predicates with degrees 1, 0.95, 0.9, 0. Drawback lies in the fully non-satisfied predicate. Attribute's value might be very far from the acceptable value or very close. Apparently, this is a challenge for future research where the cardinalities of tuples which are in predicates' neighbourhoods should be measured. The calculation of validity is not as complex task as coverage, because validity is directly calculated from (19).

6 Short Note to Applications

LSs are applicable in a variety of tasks. Three of them are mentioned in this section. Presumably, the first attempt to apply LSs with restriction in data imputation related to the item non-response was discussed in [27]. For this purpose we need to calculate validity (2) by the more restrictive quantifier *most of*. The restriction is realized by adjusting parameters of the quantifier shown in Fig. 2 in the following way: $m > 0.5$ and $n = 1$ yielding the quantifier *almost all*. Further, when validity is significant but not sufficiently high we should focus on a more restrictive part of a database. One option is the conjunction of initial and additional atomic predicates in the R part. Hence, the care should be taken when constructing fuzzy sets. Further, a minimum t-norm should be used for merging atomic predicates. Finally, quality measures should be applied. Regarding quality measures, validity and coverage are more important than simplicity. A more restrictive part of a database may have strong relation between attributes (high values of validity and coverage) but the simplicity measure (16) is low. Therefore, in the terms of bipolar approaches validity and coverage (11) are restrictions and simplicity is desire. In this way LSs might be competitive to other data imputation approaches but definitely further research is required.

The second method of application is converting mined LSs into fuzzy if-then rules [15, 23]. The research of quality measures was influenced by this task, because fuzzy rules should be of a high quality due to their broad applicability in, e.g. control and classification. Therefore, the aforementioned aspects of an LSs quality should not be neglected. Furthermore, less complex rules are preferred. Hence, the simplicity measure (16) has in this field higher importance than in the data imputation field.

The third kind of applications is mining "abstracts" from the data for informative purposes and to support decision and policy making processes. In the former, the less restrictive quantifier *majority of* can be applied. It corresponds with the *most of* quantifier defined in [9] as $m = 0.3$ and $n = 0.85$ (Fig. 2). Other quality aspects should be also considered depending of the type of LS. Contrary to the two aforementioned types of tasks, in these tasks both types of LSs (basic structure and structure with restriction) are applicable. In this field reading LSs by a text-to-speech synthesis system is a suitable way for distributing mined information to users. Thus, the simplicity is a measure which should have similar importance as validity and coverage.

Although these three kinds of tasks are used for different purposes (from data collection through data analysis to data dissemination), they share quality issues but different relevance of particular quality measures.

7 Concluding Remarks

LSs play a pivotal role in summarizing information from the data when uncertainty related to the semantic meaning of the phenomena (fuzziness) is included in the task. The validity of the LS may be influenced by constructed fuzzy sets, or selected t-norm function, or may explain relational knowledge in outliers. The last observation holds for LSs with restriction part (2). Outliers appear due to the measurement and observational errors and when very few tuples has significantly different values than the high majority of tuples. At any rate, before accepting LSs, it is advisable to filter them by quality measure(s).

In this chapter, we have created a simplified outlier measure that consists of coverage and validity merged by t-norm. LS is of a sufficient quality if it has high validity and high coverage. The suggested quality measure (11) can be used as a standalone one when non-outlier coverage and validity are sufficient. Furthermore, this measure can be part of the set of quality measures. As a connective in this measure the minimum t-norm should be avoided. Suitable t-norms are those which take into consideration both attributes and do not meet idempotency property. Hence, the option is product t-norm. In cases when the full coverage ($C = 1$) is required, the suitable connective can be obtained by drastic t-norm. In this case all summaries which pass this simple filter can be ranked according to the validity degree.

Concerning the *and* connective in compound restriction and summarizer, we believe that the only suitable t-norm is the minimum t-norm, because the proportion of tuples which contribute to the summary is not unnaturally decreased.

In the future activities, we will focus on aggregating quality measures into the compound one and on developing quality measures for summaries consisted of quantified restriction and summarizer.

References

1. Yager, R.R., Ford, M., Cañas, A.J.: An approach to the linguistic summarization of data. In: 3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU '90), Paris, France, July 2–6, pp. 456–468 (1990)
2. Lesot, M.-J., Moyse G., Bouchon-Meunier, B.: Interpretability of fuzzy linguistic summaries. Fuzzy Sets Syst. (In press) **292**(1), 307–317 (2016)
3. Yu, J., Reiter, E., Hunter, J., Sripada, S.: Sumtime-turbine: a knowledge-based system to communicate gas turbine time-series data. In: Chung, P.W.H., Hinde, C.J., Ali, M. (eds.) Lecture Notes in Computer Science, LNAI, vol. 2718, pp. 379–384. Springer, Berlin, Heidelberg (2003)
4. Arguelles, L., Triviño, G.: I-struve: automatic linguistic descriptions of visual double stars. Eng. Appl. Artif. Intell. **26**(9), 2083–2092 (2013)
5. Yager, R.R.: A new approach to the summarization of data. Inf. Sci. **28**(1), 69–86 (1982)
6. Bouchon-Meunier, B., Moyse, G.: Fuzzy linguistic summaries: where are we, where can we go? In: 2012 IEEE Conference on Computational Intelligence for Financial Engineering and Economics (CIFEr 2012), New York, USA, March 29–30, pp. 1–8 (2012)
7. George, R., Srikanth, R.: Data summarization using genetic algorithms and fuzzy logic. In: Herrera, F., Verdegay, J.L. (eds.) Genetic Algorithms and Soft Computing, pp. 599–611. PhysicaVerlag, Heidelberg (1996)
8. Hudec, M.: Issues in construction of linguistic summaries. In: Mesiar, R., Bacigál, T. (eds.) Proceedings of Uncertainty Modelling 2013, pp. 35–44. STU, Bratislava (2013)
9. Kacprzyk, J., Zadrożny, S.: Protoforms of linguistic database summaries as a human consistent tool for using natural language in data mining. Int. J. Software Sci. Comput. Intell. **1**(1), 1–11 (2009)
10. Kacprzyk, J., Yager, R.R.: Linguistic summaries of data using fuzzy logic. Int. J. General Syst. **30**(2), 133–154 (2001)
11. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summarization of time series using a fuzzy quantifier driven aggregation. Fuzzy Sets Syst. **159**(12), 1485–1499 (2008)
12. Niewiadomski, A., Ochelska, J., Szczepaniak, P.S.: Interval-valued linguistic summaries of databases. Control Cybern. **35**, 415–443 (2006)
13. Raschia, G., Mouaddib, N.: SAINTETIQ: a fuzzy set-based approach to database summarization. Fuzzy Sets Syst. **129**(2), 137–162 (2002)
14. Rasmussen, D., Yager, R.R.: Summary SQL—A fuzzy tool for data mining. Intell. Data Anal. **1**(1–4), 49–58 (1997)
15. Wu, D., Mendel, J.M., Joo, J.: Linguistic summarization using if-then rules. In: 2010 IEEE International Conference on Fuzzy Systems, Barcelona, Spain, July 18–23, pp. 1–8 (2010)
16. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms: basic notions and properties. In: Klement, E.P., Mesiar, R. (eds.) Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms, pp. 17–60. Elsevier, Amsterdam (2005)
17. Hirota, K., Pedrycz, W.: Fuzzy computing for data mining. Proc. IEEE **87**(9), 1575–1600 (1999)
18. Castillo-Ortega, R., Marín, N., Sánchez, D., Tettamanzi, A.: Quality assessment in linguistic summaries of data. In: 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2012), Catania, Italy, July 9–13, pp. 285–294 (2012)
19. Pereira-Fariña, M., Eciolaza, L., Triviño, G.: Quality assessment of linguistic description of data. In: ESTYLF, Valladolid, Spain, February 1–3, pp. 608–612 (2012)
20. Hudec, M.: Merging validity and coverage for measuring quality of data summaries. In: Congress on Information Technology, Computational and Experimental Physics, Cracow, Poland, December 18–20, pp. 149–153 (2015)
21. Zadrożny, S., Kacprzyk, J.: Issues in the practical use of the OWA operators in fuzzy querying. J. Intell. Inf. Syst. **33**(3), 307–325 (2009)

22. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
23. Hudec, M., Vučetić, M., Vujošević, M.: Synergy of linguistic summaries and fuzzy functional dependencies for mining knowledge in the data. In: 18th International Conference on System Theory, Control and Computing (IEEE ICSTCC), Sinaia, Romaina, October 17–19, pp. 335–340 (2014)
24. Garibaldi, J.M., John, R.I.: Choosing membership functions of linguistic terms. In: 12th IEEE International Conference on Fuzzy Systems (FUZZ '03), St. Louis, USA, May 25–28, pp. 578–583 (2003)
25. Hudec, M.: Linguistically summarizing hierarchical data. In: 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI 2015), Budapest, Hungary, November 19–21, pp. 141–145 (2015)
26. Kacprzyk, J., Ziółkowski, A.: Database queries with fuzzy linguistic quantifiers. IEEE Trans. Syst. Man Cyber. SMC-**16**(3):pp. 474–479 (1986)
27. Hudec, M.: Linguistic summaries applied on statistics—case of municipal statistics. Austrian J. Stat. **43**(1), 63–75 (2014)

Part II
Information Systems and Image Processing

Decomposition-Compensation Method for IT Service Management

Oleksandr Rolik, Valerii Kolesnik and Dmytro Halushko

Abstract A novel approach for service level management of corporate IT infrastructures is considered. Decomposition-compensation method of service level management of corporate IT infrastructures is proposed in this work. The method assumes the decomposition of tasks related to service level management and the compensation of negative impact of various factors by allocating extra resources for critical applications. The approach is based on the interaction of three integrated hierarchical processes—matching the level of services, resource planning, and service level management.

Keywords IT infrastructure · Cloud management · Resource allocation · Two-level management systems · Service level management · Resource planning · SLA management

1 Introduction

Business considers the information technologies (IT) toolkit as a means for improvement their productivity and competitiveness. The efficiency of business processes significantly depends on the IT operation services. An increasing number of IT services required for business technology automation, complexity of applications and the increasing number of IT infrastructure components leads to a decrease in effectiveness of IT departments and increasing the cost of maintaining a regular operation mode of the IT infrastructure. The IT department provides the

O. Rolik (✉) · V. Kolesnik · D. Halushko
Department of Automation and Control in Technical Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
e-mail: o.rolik@kpi.ua

V. Kolesnik
e-mail: kolesnik.valerii@gmail.com

D. Halushko
e-mail: dima.halushko@ukr.net

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_6

maintenance of IT services, included to the catalog. Provision of IT services is regulated by the service level agreement (SLA), signed between the business units and the IT department. The SLA defines key performance indicators (KPI) and key quality indicators (KQI) which are a limited set of objectively measured parameters for assessing the quality of IT services. For support of KPI values and KQI level, agreed in the SLA, administrators provide continuous operation of the IT infrastructure; carry out maintenance and repairs using the automatic, automated, and manual control [1].

To improve performance of the IT infrastructure and automation of the serving processes there are created the IT departments of IT infrastructure (DITI), which are used for development of IT infrastructure management systems (MS) [2]. Increasing business demand for IT services, a variety of IT, as well as the continuous improvement of business processes and the consequent need for the development and implementation of new IT lead to excessive complexity of MS, which is the product of system integration of different approaches and incompatible solutions from different manufacturers. Increasing complexity of managing IT infrastructure, accompanied by increase on operations costs, makes it necessary to search for new approaches to the management of IT infrastructure.

Currently, the development of IT is strongly influenced by factors such as consolidation and virtualization of resources, cloud computing revolution, convergence of telecommunication technologies and services. Due to influence of these factors in the informational industry can be seen intense process of globalization of information and communication technologies, which leads to new IT environment, that provides a promising means for doing business. Revolutionary transformations in IT contribute to the progress of the business technologies; however, the effectiveness of IT brings the delay in the progress on management technologies of IT infrastructure [3].

The high cost of ownership in the functional IT infrastructure at corporate-level and significant dependence of business success on the quality of IT services make important scientific and applied problems of creation the information technology for management of the corporate IT infrastructure. The complexity here is connected with the constraints put on IT infrastructure: IT infrastructure aims are the maintaining of agreed level of IT services when resources should be used rationally in terms of virtualization, clustering, and consolidation with taking into account significant dynamic of user requests.

2 General Problem Statement

Significant influence of IT to achieve business goals not only underlines the importance of IT services, but also emphasizes the management of these services. Leading position in the area of IT service management belongs to the ITSM [4], what is recognized around the world approach that is also implemented everywhere. This approach also evidenced by the emergence of ISO/IEC 20000 [5, 6]

international standard. The standard ISO/IEC 20000 is the first international standard on IT services management. It includes the requirements for management, documenting, competence, awareness and training of staff; requirements to the monitoring, measurement, evaluation and improvement of processes; principles of the management plan for services and the application of the Deming Cycle for IT service management. That standard had made changes and additions to the process model by moving from the set of processes to creation of an integrated IT service management system. Moreover, it had provided 13 processes, divided into 5 groups and 2 domains of top-level management.

Requirements for IT service management and management system are defined in ISO/IEC 20000-1, and the guidelines for the organization of activity on IT service management are in ISO/IEC 20000-2. In accordance with the standard, MS with policies and structured approach should implement embedding and effective management of all IT services. In this case, it deals only with process management, when operational issues on IT infrastructure, whose state and functioning has the most impact on service level, are not considered.

However, processes of operational management of level of IT services in IT infrastructure are missing among ITSM processes and in the ISO/IEC 20000. The reason for this is the management of IT services is relatively new and actively developing field of management. The possibility and necessity of IT service management is recognized throughout the world and as evidenced by the emergence of the standard, while the consideration of IT infrastructure as a control object during management of IT services is still not fully realized.

At the same time, the consideration of IT infrastructure as a control object to maintain the quality of IT services provisioning at an acceptable level within the change of an IT infrastructure components' state and taking into account dynamics of the user requests is not only possible, but also necessary.

Thus, questions of process management of service level are well-researched and standardized, and issues on management of services level through operational management of IT infrastructure is still not giving sufficient attention. Therefore, in this work attention is paid to the aspect of management of IT infrastructure, which connected with the importance of receiving IT services with a consistently high quality and without irregularities in the work.

The aim of this work is to develop such approach to management of IT infrastructure, which guarantee that the quality of provided IT services correspond to the specified value, which was agreed with the business-department level.

3 Basic Management Model

From the point of view of IT infrastructure management, three management loops can be distinguished: outer, inner, and operational. Figure 1 depicts these management loops, which shows the basic model of management of the IT infrastructure.

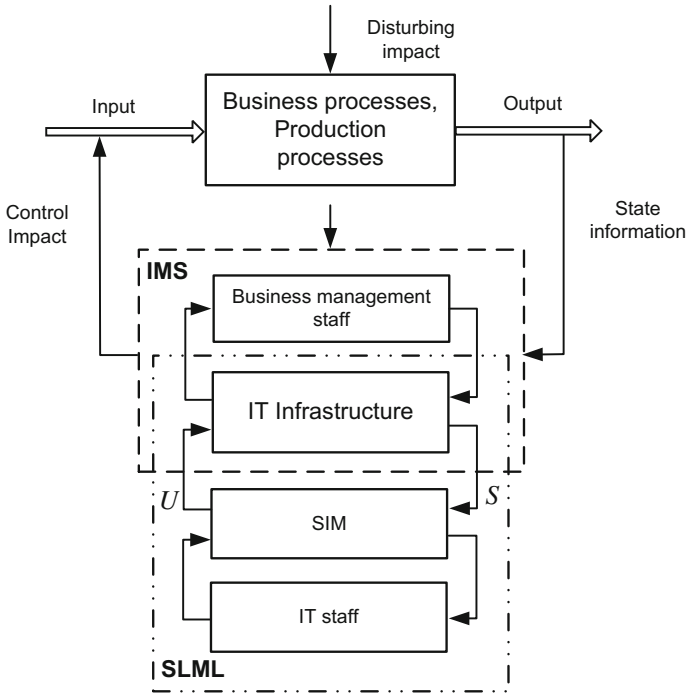


Fig. 1 Basic management model of IT infrastructure

The outer loop provides designation and implementation of business processes with following control on efficiency of the processes execution. Thus, informational management system (IMS) is the basis for effective running of business. Inner loop provides a mapping of business processes to IT services with the definition of target values, established in the SLA, and the control of the quality of IT services. Decomposition of IT services defines the tasks of operational loop, which are reduced to the continuous maintenance of the level of services on agreed level with minimal effort.

Business processes (management, operation, and support) or production processes are using criteria of business efficiency during decision-making.

IMS is a system, which combines business management staff and IT infrastructure. IMS is responsible for making decisions, which ensures the fulfillment of business criteria.

At the same time, service level management loop (SLML) is a union of IT infrastructure, its support staff (IT staff), and technical, software and informational tools for maintenance in the form of system for IT infrastructure management (SIM). SLML ensures the provision of IT service for business units on an agreed level of quality with a reasonable usage of resources.

As the input signal, IT infrastructure receives a flow of request. The IT infrastructure responds with the flow of results. Disturbing impacts affect IT infrastructure from outside. The SIM analyzes current state of IT infrastructure and selects a management impact, at which the maximum management efficiency is achieved.

In accordance to Fig. 1. IT infrastructure is a common component of the interaction-based components: the IMS and the SLML. Respectively, there are two large-scale management objects: business processes (production processes) with IT infrastructure, and two control systems: the IMS and the SIM. In management loops of these systems, there are business management staff and IT infrastructure management staff respectively.

For effective business, there are should be formed coordinated integral system of the processes of the three management loops with differentiating following process. In the outer loop: forecasting → development planning → accounting → estimation by the criterion yield/quality → control → development plans correction. On inner loop, as a rule, the Deming cycle is used: planning → execution → verification → actions for the implementation of the plan. On operational management: monitoring → state analysis → management → system optimization → development planning. Timeliness and accuracy of execution of integrated complex of processes on three management loops achieves business goals.

For support of closed management loops, in IT DITI operational management aspect integrates with an aspect of the coordinated interaction of the components of the management system of a company. Coordination contributes to reaching the goal of functioning of the IT DITI as a component of company's management system. Thus, there are should be ensured coordinated interaction and exchange of information between the management loops with the corresponding generalization of information in each of them to make decisions. In this case, the role of IT DITI is enhanced, which resulted in a shift of emphasis of IT infrastructure management from support of informational and communication technologies and equipment to the maintenance of the level of IT services.

Thus, the management of IT infrastructure cannot be considered in separately from business management, while a central role for improving the efficiency of the functioning of the SLML with orientation on maintenance an agreed level of IT services belongs to IT DITI.

In operational loop there appears measurement tasks, analysis, assessment, accounting, control, forecasting, planning and management tasks, quality evaluation tasks [7], what raises the need for development of the relevant models and methods for solving these problems. Before developing these models and methods, should be decided the general approach to solving problems of the operational loop with consideration of its complex nature and need of interaction and coordination with outer and inner management loops [8].

Managing of the IT department, which focuses IT staff in it, means first organizational forms of management, which are well developed in the ITSM [9] on the ISO/IEC 20000, whereas the IT DITI presenting tools to automate management of IT department.

Management mechanisms of the SIM are laid at design stage and are mainly including a change in configuration and self-diagnostics.

IT department performs IT infrastructure management with the use of automatic, automated, and manual control.

4 The Problem Definition of IT Infrastructure Management Conducted With Service Level Management

For such a control object as an IT infrastructure, it is very difficult to define and formalize a unified management task. That is why the decomposition of the general problem of IT infrastructure management is made, which is based on the choice of such permissible control, that maximizes the value of the control effectiveness $(K(U) \to \max\limits_{U \in \boldsymbol{U}})$. In the result separate tasks with further formalization is obtained.

IT infrastructure intends to be a provider of IT services for users. In this case, management efficiency can be evaluated on the quality $\boldsymbol{Q}$ of provided services and management costs. With the operational management of IT infrastructure, the quality management task is to maintain a given level of quality of the services with a minimal amount of used resource. Then the maximum management efficiency achieved by the choice of such control, in which the actual level of service corresponds to the agreed level with a business department $\boldsymbol{Q}_{agr}$ and is achieved with minimal effort:

$$\text{Expenses}(\boldsymbol{Q}(U) \to \boldsymbol{Q}_{agr}) \to \min_{U \in \boldsymbol{U}}. \tag{1}$$

The quality $\boldsymbol{Q}$ of services is determined by the quality $Q_j, j = \overline{1,N}$ of all IT services:

$$\boldsymbol{Q} = f(Q_1, \ldots, Q_N), \tag{2}$$

Therefore, control impacts should maintain a specific level of the quality of each service using for this purpose the minimum number of resources:

$$\text{Expenses}(Q_j(U) \to Q_j^{\text{agr}}) \to \min_{U \in \boldsymbol{U}}, \quad j = \overline{1,N}, \tag{3}$$

where Q_j^{agr}, $j = \overline{1,N}$ is the agreed level of the j-th service.

In contrast to the process management, whose aim is constant improvement of service quality, operational management aimed to maintain quality of service on the agreed level by the cheapest way, while management should be such that be ensured next

$$q_{k,j} - q^{*}_{k,j} \to 0, \ \forall j,k, \tag{4}$$

where $q_{k,j}$ and $q^{*}_{k,j}$ are respectively, target and actual values of the k-th indicator of quality of j-th service.

From the point of view of the criterion of effectiveness of business-efficiency management of IT infrastructure, while ensuring the quality of the j-th service may be the choice of the management impact $U \in \boldsymbol{U}$, at which the minimum actual processing time of i-th request to the application $A_j, j = \overline{1,N}$ is reached, N—number of applications

$$\min_{\forall i,j}(T_{\mathrm{Ac}_{i,j}} = (t_{R_{i,j}} - t_{A_{i,j}})), \tag{5}$$

where $t_{A_{i,j}}$ is the admission time of i-th user request of j-th service $t_{R_{i,j}}$ is the admission time of response to i-th user request to the application Aj.

With use of the criterion (5) from IT DITI, it would be impossible to achieve the processing time, equals to 0. As for the quality of services provisioning, the IT DITI has to manage and strive to minimize the difference between the target T_{T_j} and the actual time T_{Ac_j} of the query to the j-th application, measured on the user side using the least amount of resources

$$\min_{\forall j}(T_{\mathrm{T}_j} - T_{\mathrm{Ac}_j}) = \min_{\forall j}(\Delta T_j), \text{ when } T_{\mathrm{Ac}_j} > T_{\mathrm{T}_j}. \tag{6}$$

In case when $T_{\mathrm{Ac}_j} - T_{\mathrm{T}_j} > 0$, the cause of which may be an increase in the number of user requests, fulfilling the criteria (5) or (6) is possible by allocating additional information and computing resources to the application Aj, and/or prioritized transfer of user data of application Aj over the telecommunication network.

5 The Approach to Operational Management of the Level of IT Services

The essence of the approach to the operative management of the level of IT services. The main purpose of business is getting the maximum profit. Maximum profit due to IT is accessible when a business offers a set $\boldsymbol{S} = \{s_i\}$, $i = \overline{1,K}$ of required IT services with maximum quality $\boldsymbol{Q}$ and a minimum expense $\boldsymbol{C}$.

Service level management in corporate IT infrastructures is implemented by integrated interaction of three processes: the agreement of services level, the planning of resources and management of service level (Fig. 2).

Business managers initiates the launch of the process of agreement the level of service, and this process ends by the creation or updating of the elements of the set $\boldsymbol{S}$ and the matrix $Q = \|q_{ki}\|$, element q_{ki}, $k = \overline{1,M_i}$, $i = \overline{1,K}$, which corresponds to an agreed value of k-th indicator of quality of the i-th service. Business allocates

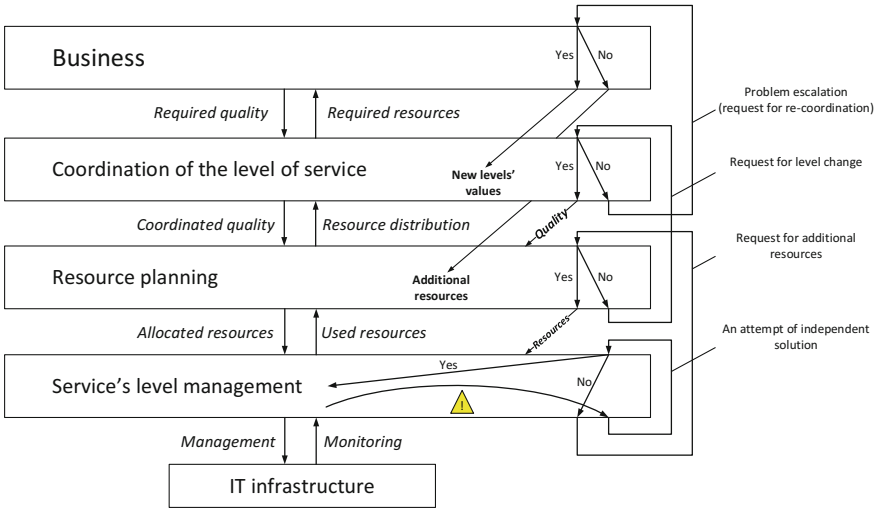


Fig. 2 Interaction of processes in the management of service level

r amount of resources to the services $\boldsymbol{S}$. The resulting resource support in the form of system

$$< Q, r > \tag{7}$$

is the basis for solving the problems in the lower located level.

The planning process is based on allocation and consolidation for each service s_i, $i = \overline{1, K}$ part of the resources R_1, …, R_m of IT infrastructure. Such allocation of resources allows to maintain the services, while r_1,…,r_m are amount of resource R_1, …, R_m, and c_1, …, c_m are unit cost of the resource R_1, …, R_m respectively.

Then the amount of total resource is calculated as follows:

$$r = \sum_{j=1}^{m} r_j. \tag{8}$$

In addition, the cost c of resources is determined by next equation:

$$c = \sum_{j=1}^{m} r_j \cdot c_j. \tag{9}$$

Using the resources by services is defined by a matrix $\mathrm{P} = \left\| \rho_{ij} \right\|$, where ρ_{ij} is equal to the number of dedicated resource R_j, $j = \overline{1, m}$ to service s_i, or 0 if the resource is not required.

The process of management of the service level is managing the IT infrastructure, so that the actual values q_{ki}^{*}, $k = \overline{1, M_i}$, $i = \overline{1, K}$, of indicators of values

respectively, would correspond to agreed values of matrix Q, so that the equality fulfills.

$$q_{ki} - q_{ki}^{*} = 0, \quad k = \overline{1, M_i}, i = \overline{1, K}. \tag{10}$$

The essence of the approach to the management of level of services is the next.

In terms of non-compliance, condition (10) identifies the elements of the matrix of the actual values of the quality indicators $Q^{*} = \left\| q_{ki}^{*} \right\|$ for which $q_{ki}^{*} < q_{ki}$, $k = \overline{1, M_i}, i = \overline{1, K}$. IT DITI trying to solve a problem on the lower level (see Fig. 1) by changing the values of the functioning parameters of IT infrastructure elements or by redistributing the resources between applications so to increase the value q_{ki}^{*} in the result.

If in the result of the recovery measures, it was possible to ensure the equality (10) is true, then functioning of IT infrastructure continues with the new settings. If the authority of lower level is not sufficient for the achieving (10), the escalation of the problem is carried out at the level of resource planning.

During the process of resource planning are made attempts to solve the problem of allocation of additional resources $R_1, \ldots, R_m$ for s_i service, for which the condition $q_{ki}^{*} < q_{ki}$. is true. If additional resources are allocated, then the matrix $P' = \left\| \rho_{ij}' \right\|$ formed with new values of elements, moreover $\rho_{ij}' > \rho_{ij}$, $j = \overline{1, m}$ or result is equal to zero if the j-th resource is not required. If additional resources are missing, then at the level of resource planning there are conducts the attempts to perform a redistribution of resources between the services by allocating the resources to more important services due to less important. If the problem is solved, the values of the matrix $P' = \left\| \rho_{ij}' \right\|$, with a new plan of resource allocation go to the lower level. If it is unable to resolve the problem at the level of planning the resources then the escalation of the problem to a higher level is done.

The planning process initiates the process of agreement of services level to review first the value q_{ki} for which $q_{ki}^{*} < q_{ki}$, and then, perhaps, the values of all elements q_{ki}, $k = \overline{1, M_i}$, $i = \overline{1, K}$ of the matrix of service quality Q in descending order. If it is possible to form a matrix $Q' = \left\| q_{ki}' \right\|$ with the new values of quality indicators, then it is transferred to the lower level, which produces the release of resources and the allocation of them to services with following condition $q_{ki}^{*} < q_{ki}$, $k = \overline{1, M_i}$, $i = \overline{1, K}$. If the process of agreement the service level does not have permissions to the procedure of forming the matrix $Q' = \left\| q_{ki}' \right\|$, then produced an escalation of the problem to the level of business. This level must either generate a matrix $Q' = \left\| q_{ki}' \right\|$ with the new values, or increase the total amount of resources, which leads to an increase of the values of $r_1, \ldots, r_m$. Otherwise accept the actual level of services is the last option.

Let us consider the processes implemented through agreement of service level, planning the resources and the management of service level.

The process of agreement of service level in the corporate IT infrastructures. Requirements for maximizing $(\max \boldsymbol{Q})$ the quality of services and minimizing the associated expenses $(\min \boldsymbol{C})$ can be extracted from a usual conflict, which leads to the need of establishing an economically reasonable level of services with taking into account the company's capabilities, and achieved level and customer expectation in the industry.

If $\hat{D}$ are business revenues from IT, and $\hat{C}$ are expenses on IT infrastructure, then the business together with the process of agreement about service level is trying to achieve

$$\max(\hat{D} - \hat{C}). \tag{11}$$

In an effort to continually improve the quality, business and IT within ITSM must execute an analytical assessment of dependency of the quality of service from the cost of resources $\hat{Q} = f_1(c)$; business losses $\hat{L}$ from poor quality of service with taking into account the risks M_0 and uncertainties N_0 (Fig. 3). Right after that, the departments should take into account the level of quality achieved by the industry and determine accessible values of the level of service $Q = \|q_{ki}\|$, which have not yet led to losses at the runtime of business operations. Thus, the system

$$< \hat{D}, \hat{C}, \hat{L}, \hat{Q}, M_o, N_0 > \tag{12}$$

defines the task of the formation of the pair (7).

If the received values, with considering risks, exceed the achieved level by industry, then there is a cause for optimism, and business development. Otherwise,

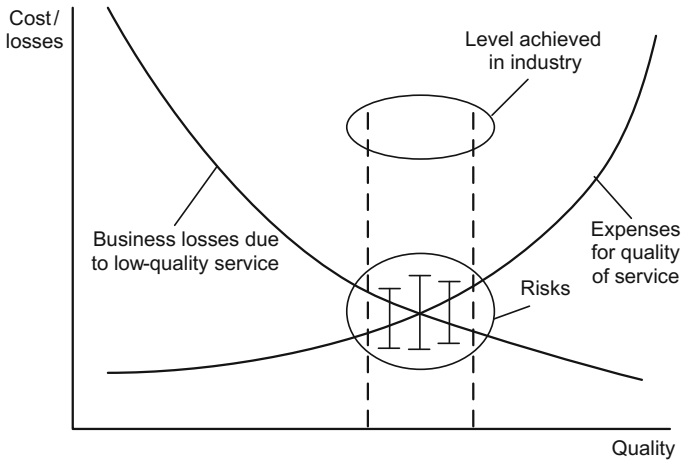


Fig. 3 Search for the optimal relation of quality level of service and a cost of achieving this quality

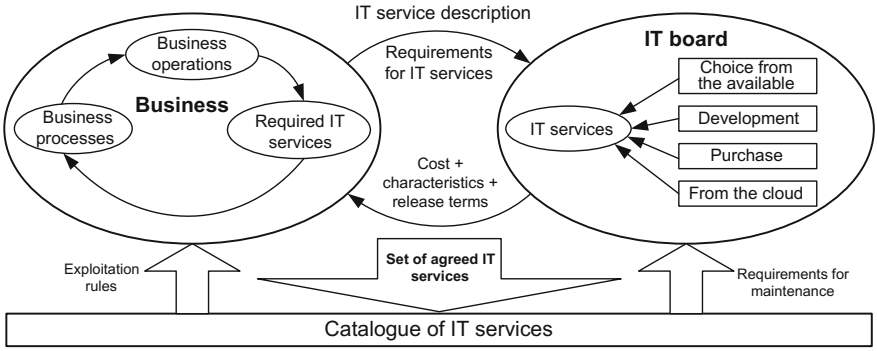


Fig. 4 The process of forming the catalog of IT services

business faced up with a problem of the delivered quality what can be the basis for the collapse of business in their relative areas.

Business is delivered with options for service implementation with cost, technical, organizational, and time indicators (Fig. 4). Business determines the optimum ratio of profitability/(quality of IT services). For such case losses are estimated depending on the reduction of the level of services and the cost of delivering of a high quality services, then the point of the minimum accessible quality is found. After that the business endorses one of the variants or adjusts the requirements for a new IT service, basing on the importance of service, benchmarking, experience and so on. The agreed services' levels are scripted in the SLA or service catalog.

IT board prepares the plan of the implementation of new services considering the financial and resource support for their provision and management. This problem is more preferable to solve with the use of methods for effective management of resources of IT infrastructure.

Coordination of service level is running under conditions of excess or shortage of resources. The excess of resources initiates the iterative procedure for agreement of the level of service. At the same time on the basis of (8) the planning process determines the values q_{ki}, $k = \overline{1, M_i}$, $i = \overline{1, K}$, then the resulted indicators in the form of matrix Q are presented to business. If values of matrix elements do not satisfy the business, then made redistribution of resources between services and starts a new cycle of defining the q_{ki}, $k = \overline{1, M_i}$ values, when for a given resource consolidation plan of R_1, …, R_m resources for services s_i, $i = \overline{1, K}$ is being determined the expected quality Q.

During solving the problems of coordination the level of service can be used methods of mathematical programming, operations research, decision theory, methods of game theory, methods of solution the tasks with multiple criteria, decision-making under conditions of non-certainty and risk, methods of artificial intelligence, balance models, and so on.

The process of resource planning. At the level of resource planning problems are solved in the interests of the process of service level agreement, also is made a

determination of required amount of resources, and distribution with binding of resources for service s_i, $i = \overline{1, K}$.

Services s_i, $i = \overline{1, K}$ are supported by applications from the set $A = \{A_l\}$, $l = \overline{1, I}$, where I is the number of applications, where each service s_i, $i = \overline{1, K}$ is supported by one or several applications, while each application A_l, $l = \overline{1, I}$ supports one or more services.

We introduce the concept of the degree of service's support. Service support s_i means the interaction of multiple resources assured applications from the set A, directed to achieving a common result that is workability of service s_i. Let suppose that resources R_j, $j = \overline{1, m}$ are used to support s_i, $i = \overline{1, K}$ services [3]. First, let us consider the boolean variable x_i, $i = \overline{1, K}$ which defines the level of support for the i-th service and takes the following values:

$$x_i = \begin{cases} 1, \text{ if } i \text{ } -\text{th service gets entire support;} \\ 0, \text{ if } i \text{ } -\text{th service gets no support.} \end{cases} \tag{13}$$

Let the users of s_i service are forming an average number of requests a_{li} to the A_l application for time unit. Then the number of client requests to application A_l is determined as follows:

$$a_l = \sum_{i=1}^{K} a_{li} \cdot x_i, \tag{14}$$

and the number of requests to applications A_l, $l = \overline{1, I}$ we will present as vector $\hat{a} = \{a_l, l = \overline{1, I}\}$.

Application A_l needs resources of type j which is given by

$$r_{jl} = b_{jl} a_l + d_{jl} \tag{15}$$

where b_{jl} is the average amount of resources of the type j, used by application A_l to process one client request; d_{jl} is the amount of resources of the type j, used by application A_l regardless to the number of client requests.

Then the total amount of resources r, which are necessary to maintain all the services from the $\boldsymbol{S}$, defining by the expression:

$$r = \sum_{j=1}^{m} \sum_{l=1}^{I} r_{jl} = \sum_{j=1}^{m} \sum_{l=1}^{I} (b_{jl} a_l + d_{jl}). \tag{16}$$

Planning tasks essentially depend on the constraints on the budget of the IT department. With the absence of financial constraints, when the criterion min$\boldsymbol{C}$ is not taken into account, it is planned and projected in the IT infrastructure, whose resources would be sufficient for maintenance of applications $\{A_l\}$ with the required values of quality indicators of services q_{ki}, $k = \overline{1, M_i}$, $i = \overline{1, K}$ with maximally allowed values of vector $\hat{a}$. Moreover, the most important resources are duplicating

or reserving $(r_r > 0)$ and accounted possible increase of the number of requests a_{li} beyond the bounds.

With a limited budget for the creation and development of IT infrastructure the lack of resources can be laid at the stage of design. It can even arise during the usage, and developed IT infrastructure cannot ensure the fulfillment of condition (10). Despite this, the IT infrastructure, which was designed with a lack of resources, can deliver services effectively if provided the support for the most critical applications a priori. To do this, IT DITI redistributes resources on the runtime in accordance to the defined regulations of resource usage.

Thus, planning and resource management problems need to be solved within both conditions whether excess or lack of resources. In each case, some of the specific features of the allocation of resources must be considered.

After calculating the need of resource with the expression (16) and comparing it with the available resources we will get the problem of managing the service level with a lack of resources if emergencies in the IT infrastructure, increase of the intensity of client requests and other factors lead to failure of equality (10).

In this case, for making decisions for resource allocation should be considered the additional information.

We introduce the concept of the importance w_i of service s_i, $i = \overline{1, K}$ which will be used in the solution of services level coordination tasks in conditions when there is a lack of resources.

The problem of service level agreement is reduced to the definition of values of the matrix Q within an assigned amount of resources:

$$Q = \boldsymbol{F}_1(\boldsymbol{S}, r, W_p, Z_s), \tag{17}$$

where $W_p = \{w_i | i = \overline{1, K}\}$; $Z_s = \{z_i | i = \overline{1, K}\}$; z_i is planned level of support for the i-th service. Here, the standard value of client requests number considered in the value of r, and the value z_i, $i = \overline{1, K}$, in contrast to (13), is a continuous variable, that takes values from the interval [0, 1].

Nevertheless, if after the calculation of the need for resources and further comparison with the available resources, the amount of available resources is greater than needed, then we get the problem of managing service level in resource abundance. Particularly, can be allocated a reserved amount of resources r_r and the number of resources $\hat{r}$ increases. These resources $\hat{r}$ are allocated to support all of the services $\boldsymbol{S}$, and are defined by the expression

$$\hat{r} = r + r_r. \tag{18}$$

Then the resources r is determined by the values of the elements of the matrix Q, and the value of r_r is determined by the probability of occurrences of emergency situations and bound values of a_{li}.

In this case, there are two distinct problems of service level agreement. The problem of the first kind is similar to (17) and is to determining the values of quality indicators for a known amount of resources allocated:

$$Q = \boldsymbol{F}_2(\boldsymbol{S}, \hat{r}, r_r, \boldsymbol{C}). \tag{19}$$

The objective of the second kind is to determine the necessary resources to ensure the specified values of quality indicators:

$$r = \boldsymbol{F}_3(\boldsymbol{S}, Q, \boldsymbol{C}). \tag{20}$$

During solving the problems (17), (19), and (20) may use the methods of queuing theory [10], reliability theory, the theory of fractals, simulation modeling and analysis, particularly with application of queuing theory methods and artificial intelligence methods.

The process of management of the service level. After agreement of the level of services and planning of resources, the process of management is carried out so that the following criteria is fulfilled:

$$\min(q_{ki} - q_{ki}^*),\ k = \overline{1, M_i}, i = \overline{1, K},\ \text{when}\ q_{ki}^* < q_{ki} \tag{21}$$

or

$$\min \hat{C},\ \text{when}\ q_{ki}^* > q_{ki}, \quad k = \overline{1, M_i}, i = \overline{1, K}. \tag{22}$$

In this case, for the cost savings, there are made reductions of allocated resources to applications $\{A_l\}$ and unused resources are released. Such problems are solved in [11]. Moreover, the paper [12] describes possible methods and detailed problem statement for related issues.

Criteria (21) and (22) are applied only when $k = \overline{1, M_i}$ and $i = \overline{1, K}$, and by comparing the values q_{ki}^* and q_{ki} the condition "only not more" or "only not less" fulfills. Otherwise, resources between applications $\{A_l\}$, can be redistributed so that for the applications for which next is correct $q_{ki}^* < q_{ki}$, there would be allocated the resources by the applications, for which fulfills next $q_{ki}^* > q_{ki}$, $k = \overline{1, M_i}$, $i = \overline{1, K}$.

If by use of lower level facilities, it is impossible to ensure equality $q_{ki}^* = q_{ki}$ when $q_{ki}^* < q_{ki}$, then is executed an iterative procedure in which the upper levels are utilized (see Fig. 2). In this case, on the upper levels the management system allocates additional quantum of resources Δr for application $A_l^* \in A$, which fulfills next condition $q_{ki}^* < q_{ki}$. Then execute the check of fulfillment the condition (10). If everything remains the same $q_{ki}^* < q_{ki}$, then infrastructure provides another quantum of resources Δr. The procedure repeats until the condition (10) fulfills. In the

absence of resources in IT infrastructure, there are two possible situations during the management of service level:

(1) Beginning of the revision of values in matrix Q;
(2) Allocation of quantum of the resource Δr by application from the set $\{A_l\}$ with consideration of importance W_p of services.

The dependence of the values of quality indicators q_{ki}, $k=\overline{1,M_i}$, $i=\overline{1,K}$, from the resources r without the loss of generality can be provided as follows:

$$q=f_{qr}(r), \tag{23}$$

where q—quality of services. For increasing the value of q to the corresponding application from the set, $\{A_l\}$ it is necessary to allocate additional resources. Then

$$q^{'}=f_{qr}(r+\Delta r). \tag{24}$$

If $\Delta r>0$, then $q^{'}\geq q$, what allows to make the assumption of monotonous character of the function f_{qr}.

Similarly, it can be assumed that the function

$$q=f_{qa}(\hat{a}), \tag{25}$$

is also monotonous.

Then, if the functions (23) and (25) are monotonous, then the function

$$q=f_q(r,\hat{a}) \tag{26}$$

will also be monotonous [13].

Let the control $u^+\in U$, where the U is set of control impacts, is the allocation of additional resources to application $A_l^*\in A$. Furthermore, for this application the actual quality q_a is worse than target one q_t, $q_a<q_t$, and $u^-\in U$ is the management impact which withdraws resources from the application $A_l^*\in A$ if $q_a>q_t$.

With taking into account the monotonous nature of the dependence between q_{ki}, $k=\overline{1,M_i}$, $i=\overline{1,K}$ and r, lets prove necessary statements, but making the following suggestions.

The use of resources by applications is set as a matrix $\mathrm{P}=\left\|\hat{\rho}_{lj}\right\|$, $l=\overline{1,I}$, $j=\overline{1,m}$, where

$$\hat{\rho}_{lj}=\begin{cases} n_{lj}\Delta r_j, \text{if } l \text{ – th appliction uses } j \text{ – th resource;} \\ 0, \text{if } l \text{ – th application does not use } j \text{ – th resource,} \end{cases} \tag{27}$$

where n_{lj}—the number of quants of j-th resource allocated to l-th application; Δr_j is the size of the quantum of j-th resource. In such case, the next restrictions must be fulfilled:

$$\Delta r_j \sum_{l=1}^{I} n_{lj} \le r_j, \quad j = \overline{1,m}. \tag{28}$$

Then we can define the following mapping:

$$\boldsymbol{Q} = U \times \mathrm{P} \times \bar{\hat{a}} \to Q, \tag{29}$$

where $\bar{\hat{a}} = \{\hat{a}\}$ is the set of vectors $\hat{a} \in \bar{\hat{a}}$.

Whereas

$$\boldsymbol{U} = Q^* \times \mathrm{P} \times \bar{\hat{a}} \to U. \tag{30}$$

Proposition 1 *For a given value* q_{ki}, $k = \overline{1,M_i}$, $i = \overline{1,K}$ *in the case when* $q_{ki}^* < q_{ki}$, *there are exists such management impact* $u^+ \in U$, *which allows to provide* $q_{ki}^* = q_{ki}$ *at the* min*r for maintenance the level of service.*

This follows from the monotony of the functions (23)–(26), a finiteness of sets Q^*, P and $\bar{\hat{a}}$, and comparability of processes' objectives in Fig. 2.

Management $u^+ \in U$ is found iteratively.

To prove the following statement, we introduce a mapping:

$$\boldsymbol{F} = Q \times \bar{\hat{a}} \to R. \tag{31}$$

Proposition 2 *If the condition* (18) *fulfilled when* $q_{ki}^* < q_{ki}$, *if values* $\hat{a}$ *are known, then control impact* $u^+ \in U$, *allowing to restore equality* $q_{ki}^* = q_{ki}$, $k = \overline{1,M_i}$, $i = \overline{1,K}$, *can be found without the use of iterative procedures.*

Proof For fixed values of the vector $\hat{a}$, basing on (31), there are dependencies $q_t \to r$, $q_1 \to r_1$ and $q_2 \to r_2$. Then, similarly to $q_1 - q_2 \to r_1 - r_2 = \Delta r_{12}$ when $q_{ki}^* < q_{ki}$ to enforce the fulfillment of equality $q_{ki}^* = q_{ki}$ it is necessary to allocate additionally $\Delta\hat{\rho}_{lj}$, $i = \overline{1,K}$, $l = \overline{1,I}$, $j = \overline{1,m}$ to the l-th application, maintaining the i-th service with taking into account (27). Moreover, value $\Delta\hat{\rho}_{lj}$ can be defined based on $\Delta q_{ki} = q_{ki} - q_{ki}^* \to \Delta\hat{\rho}_{lj}$. In general, the allocation of additional resources with consideration of (28) without iterations may only be done if condition (18) is fulfilled, which proves the proposition.

Consequence. If the $q_{ki}^* < q_{ki}$ and i-th service is supported by the application $A_l^* \in A$, and due to applications from the set A, for which $q_a > q_t$, there are resources which can be freed $\Delta\hat{\rho}_{в}$ then the management impact $u^+ \in U$ within the resource deficit conditions allows to restore the level of the i-th service, maintained with l-th application in one pass. Also, $\Delta\hat{\rho}_{в} \ge \Delta\hat{\rho}_{lj}$, where $\Delta\hat{\rho}_{lj}$ is additional amount of

resources, which is needed by l-th application for fulfillment of the equality $q_{ki}^{*} = q_{ki}$.

The implementation of the lower level, which is directly carried out the operational management of the level of services, is advisable to run based on the coordinator [13, 14]. Also, we propose to use neural network with the coordinator to achieve better results [15].

The research of proposed decomposition-compensation approach to managing the service levels not only confirmed its performance, but also showed the ability to increase the effectiveness of using the resources. The proposed approach was compared with the method of full reservation of resources and the method of node extension by resource efficiency q_E, defined as:

$$q_E = \frac{r_u}{r_e}, \tag{32}$$

where r_u is the amount of actually used resources, and r_e is the amount of reserved and allocated resources.

The most common practical method of the full reservation involves the determining the amount of resources $r_{\max,l}$ with expression (15) for the maximum number of user $a_{\max,l}$ of l-th service agreed in the SLA. The resources $r_{\max,l}$ assigned to the application A_l does not change during the operation. If the actual number of client requests $a_l < a_{\max,l}$ then resources are used inefficiently. When $a_l > a_{\max,l}$ quality of services is reduced, the management impact to improve the level of services is not carried out, and users should be satisfied with the actual quality of services.

The method of maintaining the quality of services by increasing the node during the horizontal scale tracks not the level of services, but the percentage of utilization of allocated resources. If you exceed the degree of involvement of the individual resources of a predetermined threshold then happens the increase of the amount of resources allocated to the application by one node. During this process, the increase of amount of all kinds of resources is performed, regardless of the actual need in increasing the amount of only some of the resources. In this case, the unused resources cannot be used by other applications.

The dependence of the efficiency of resource use q_E on the ratio $a_l/a_{\max,l}$ are shown on the graphs, Fig. 5 shows the proposed approach (curve 1), the method of full reservation of resources (curve 2) and the method of increasing the nodes (curve 3) with different amount d_l of resources used by the application A_l, is independent of the number of user requests. For research, the value of the quantum of resources was set at the level of 10% from node size.

Analysis of the graphs in Fig. 5 shows that the proposed approach allows using the resources of corporate IT infrastructure much more efficiently, whereas efficiency of resource use increases due to decreasing value of the relation $a_l/a_{\max,l}$.

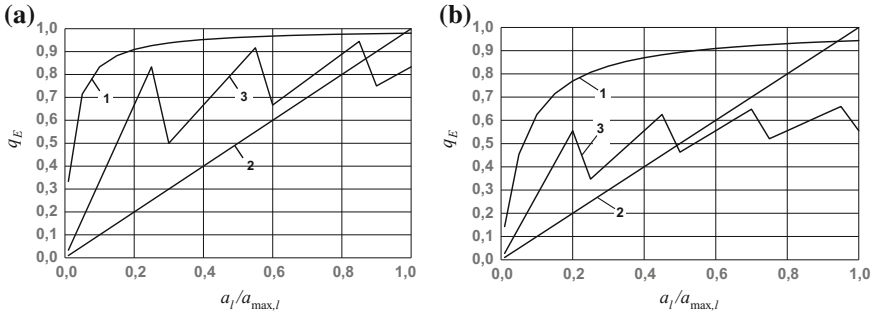


Fig. 5 Dependence q_E from $a_l/a_{\max,l}$ for the cases: **a** proportional and **b** the disproportional resource needs

6 Conclusion

Effective management of the level of service for corporate IT infrastructures is possible with application of the proposed decomposition-compensation approach, involving decomposition of tasks for service level management and compensation of negative impact of different factors by allocation of additional resources to critical applications. The approach is based on the integrated interaction of three hierarchical processes—agreement of the level of service, resource planning and the management of the level of services with consideration of a hierarchy of IT infrastructure. It makes possible to create a hierarchy of decisions for management or maintenance on agreed service level by taking into account the existing resource restrictions and levels of authority. This is possible by the use of mechanisms and capacities of higher levels of the hierarchy for selection of the control impact which makes it impossible to implement management at lower levels.

References

1. Brooks, P.: Metrics for IT Service Management. Van Haren Publication, Zaltbommel (2006)
2. Rolik, A.I., Voloshin, A.V., Galushko, D.O., Mozharovsky, P.F., Pokotilo O.O.: Agent-based corporative information-telecommunication infrastructure control system. Visnyk NTUU "KPI" Informatics, Operation and Computer Science, vol. 52, pp. 39–52 (2010)
3. Rolik, A.I.: Decomposition-compensation method of service level management of corporate IT infrastructures. Visnyk NTUU "KPI" Informatics, Operation and Computer Science, vol. 58, pp. 78–88 (2013)
4. Hubbert, E.: TechRadar™ for I&O professionals: IT service management processes, Q1 2012/E. In: Hubbert, J.P., Garbani, G., O'Donnell, S., Mann, J. (eds.) Rakowski—Forrester Research, Inc. Feb 7, p. 44 (2012)
5. Information Technology. Service management. Part 1: Specification: ISO/IEC 20000-1:2005. ISO/IEC, p. 16 (2005)

6. Information Technology. Service management. Part 2: Code of practice: ISO/IEC 20000-1:2005. ISO/IEC, p. 34 (2005)
7. Telenyk, S., Rolick, O., Dorogiy, Y., Halushko, D., Bukasov, M., Pysarenko, A.: Qualitative evaluation method of IT-infrastructure elements functioning. In: Proceedings of 2014 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom 2014), Chisinau, Moldova, May 27–30, pp. 165–169 (2014)
8. Rolik, O.I.: The complex method of service level management in the corporate IT-infrastructure. Visnyk NTUU "KPI" Informatics, Operation and Computer Science, vol. 61, pp. 148–161 (2014)
9. IT Service Management: An Introduction// J.V. Bon, G. Kemmerling, D. Pondman, p. 217. Publisher: Van Haren Publishing (2002)
10. Kleinrock, L.: Queueing Systems, vol. I: Theory. New York, Wiley Interscience (1975)
11. Telenik, S.F., Rolik, A.I., Savchenko, P.S.: Adaptive genetic algorithm for data center resource distribution. Visnyk NTUU "KPI" Informatics, Operation and Computer Science, vol. 54, pp. 164–174 (2011)
12. Telenyk, S. Rolik, O., Bukasov, M., Halushko, D.: Models and methods of resource management for VPS hosting. Technical Transaction. Automatic Control. Politechnica Krakowska, vol. 4-AC, pp. 41–52 (2013)
13. Mesarovic, M.D., Macko, D., Takahara, Y.: Theory of Hierarchical, p. 283. Multilevel Systems. Academic Press, New York (1970)
14. Rolik, A.I.: Service level management of corporate IT infrastructure based on the coordinator. Visnyk NTUU "KPI" Informatics, Operation and Computer Science, vol. 59, pp. 98–105 (2013)
15. Rolik, O., Kolesnik, V., Halushko, D.: Neural network approach for resource allocation in IT-infrastructure management system. In: Proceedings of the Congress on Information Technology, Computational and Experimental Physics 2013 (CITCEP'15) 18–20 December, Cracow, Poland, pp. 176–179 (2015)

Risk Prediction Based on Time and GPS Patterns

Daniela López De Luise, Walter Bel, Diego Mansilla, Alberto Lobatos, Lucía Blanc and Rigoberto Malca la Rosa

Abstract Traffic produces not only pollution but also many incidents resulting in material lost and human injuries or even persons dead. But not all the incidents involve two cars, it may be pedestrian and any type of cycles (motorcycles, bicycles, tricycles, etc.). Most of the approaches try to model traffic accidents using traditional information and avoiding others, such as environmental elements, driver profile, weather, regulations, eventual circumstances like strikes with roadblocks, street reparations, railroads crossings, etc. This paper presents a model for risk prediction, and the impact of varying geographical information details on the precision of the underlying Inference System (a Soft Computing model with a ruled Expert System and a Harmonic System focused on time patterns of events). Its flexibility and robustness has a price: certainly minimal to apriori knowledge. This work outlines the working model implemented as a prototype named KRONOS, and a statistical evaluation of its sensibility to dynamic GPS information. Traffic risk requires this type of flexible and adaptive model due to the high number of alternatives to consider. The model would also be improved by adding certain specific Fuzzy Logic for pattern management during the matching process. The model would also be improved by adding certain specific Fuzzy Logic for pattern management during the matching process.

Keywords Risk prediction · Time mining · Machine learning · Expert systems · Harmonic systems · Pedestrian risk · Traffic risk

D. López De Luise (✉) · R. Malca la Rosa
Computational Intelligence and Information Systems Lab, Ciudad de Buenos Aires, Argentina
e-mail: daniela_ldl@ieee.org

R. Malca la Rosa
e-mail: daniela_ldl@ieee.org

W. Bel · D. Mansilla · A. Lobatos · L. Blanc
Universidad Autónoma de Entre Ríos, Concepción Del Uruguay, Argentina

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_7

1 Introduction

One of the most compelling problems in the current world is traffic accidents, and the consequences [1, 2] involve materials lost, injuries and even death [3, 4]. But an effective evaluation of the risk must include information on the environment, context details, vulnerability of the individual, biomechanical resistance to sudden forces [5], etc.

Although there are many studies and statistics, any model representing this type of risk become apparent unless it covers a representative number of hidden factors that are indirect cause or bias. For instance, pedestrian injuries increase with a higher speed of traffic, status of footpath, availability of adequate crossing facilities, pedestrian crossing opportunities, number of lanes to cross, complexity of traffic movements at intersections, etc.

Furthermore, current and past statistics and proposals are statically defined in advance, considering just most meaningful and logical variables, but there is a lack of flexibility to append new factors dynamically. This is important, because technology evolves, society changes and therefore variables change as well. Just to mention a few of the variables, there may be the age of the pedestrians [6], subtle tips like crowd management [6], pedestrian attitudes [7], pedestrian crossings [8], etc. The prototype in this paper was tested with an initial knowledge that combine, many of these mentioned items and others. A few traffic risk models are in this line, like the Traffic Management Hazard Identification & Risk Assessment Control Form [9], that checks relevant causes and related events and G20/OECD, that is a framework for risk assessment [3]. Other proposals are still waiting for implementation and evaluation.

For pedestrian risk there are also some further alternatives. Among others can be mentioned the proposal to assess the risk of collision related to a pedestrian-based scenario [10], a Case-control approach [11], Micro-simulation Model with SSAM (Surrogate Safety Assessment Model, developed by FHWA, US) [12], a tailor-made statistical tool [13], Journey Risk Management [14], etc.

The authors in [15, 16] suggests modifying the physical environment, but to be aware of how this should be accomplished it is necessary to understand better which are all the main factors in most of the cases.

Although many experts in the field [17, 18] consider education and prevention initiatives is the most effective way to decrease mortality, it is still necessary to develop a tuned and dynamic model to keep track of and overcome statistical obsolescence.

From a Data Mining (DM) perspective, risk can be thought as a derivation of a set of variables heuristically selected as the best describing accident origin. When this is properly studied it is possible to predict not only a disaster but also its characteristics [19–21].

Many approaches in DM are used to predict events and find out its current and/or subsequent facts, like in [22, 23], etc.

Instead of that, this paper combines two reasoning systems: Expert Systems (ES) and Harmonics Systems (HS). The first one derives from the well-known technology started in 1980s but the second is quite a new technology presented in [24].

While Expert Systems (ES) remains focused on explicit rules of expert knowledge related to statistical prediction of the risk, the Harmonics System (HS) takes a heuristic data-driven approach. In HS, the information of interest is not the complex data produced along the development of an event, but only its timing patterns as a consequence of deep variables relationship during the process of an accident. This perspective is here performed by Harmonics Systems (HS), using combinations of variables (selected by an Expert System) as patterns. HS is a type of mining focused on rhythm, accelerations, static periods, and others aspects related to time features of selected patterns. HS also allow real-time processing, which is well fitted for applications that require prompt answers upon data collecting (that is the case of a driver o pedestrian collecting environmental data during displacement from one point to another). It is also included preliminary statistical results from real cases taken from [25]. Taking an Expert Systems in combination with HS [1], the traditional risk knowledge from the typical problems can be enhanced with dynamical timing patterns derived from previous activity and its variables. This kind of plastic, flexible, and self-trained learning model may serve from data. A resonance in this context can be thought of as a pattern matching with weighted features and chaining patterns. This modeling approach is being applied as the KRONOS prototype, to evaluate pedestrian and car risk. As a prototype is partially implemented, statical evaluation does not fully include HS add-ons or Fuzzy Logic at the pattern's matching process.

In the following, we shall present the basics for ES (Sect. 2), Harmonic systems (Sect. 3), the global architecture of KRONOS prototype that implements it as the core of its Expert System (Sect. 4) and a test application with real data (Sect. 5) followed by Conclusions and future work.

2 Expert System (ES)

This section presents a summary of ES as used in this project and its main characteristics.

2.1 ES Goal

The ES is used to reflect long-term expert knowledge and reasoning. The rules in KRONOS are defined using the variables and knowledge presented in [26]. Table 1 shows a reduced set of rules.

It is important to note that rules are not probabilistic or fuzzy. They fire just using the traditional approach.

Table 1 Some of the rules in the ES

IF	THEN
(individual.alcohol < 0.15) and (car.time > weather.sunset)	*car.risk.level = LOW* *individual.alcohol.level = LOW* *individual.risk.description = "You're perfects conditions"* *individual.risk.type = 0*
(individual.alcohol > 0.15) and (car.time > weather.sunset)	*car.risk.level = LOW* *individual.alcohol.level = LOW* *individual.risk.description = "You experimented a decreased reflexes"* *individual.risk.type = 1*
(individual.alcohol > 0.20) and (car.time > weather.sunset)	*car.alcohol.level = LOW* *individual.alcohol.level = LOW* *individual.alcohol.description = "Decreased reflexes, dysmetria and velocity underestimation"* *car.alcohol.type = 1*

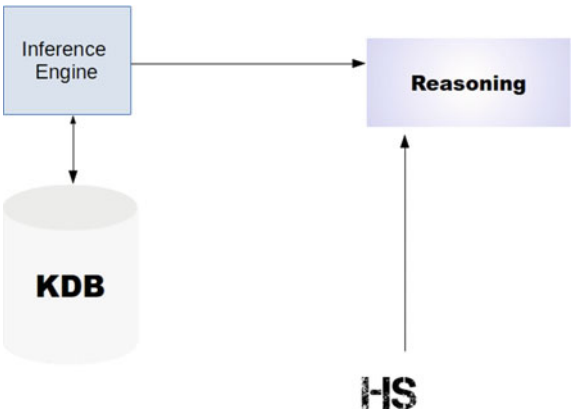


Fig. 1 ES architecture

2.2 ES Architecture

As any other ES, the prototype implements a core using the modules shown in Fig. 1.

3 Harmonic Systems (HS)

This section presents a summary of HS presented previously in [24, 26–28].

3.1 HS Goal

Once a problem has two characteristics: Real-time requirement and complex time behavior, HS can be applied. One reason is its lightweight algorithm and simple evaluation.

The other main reason is the production of meta-data generated as a part of model, reflecting change in time (as opposed to those models created through mining) while preserving other information related to the identity of the problem. When data results are processed it can be done in one of more patterns with the same or a different time variation.

An extra practical feature is the optional preprocessing with filters, to select specific time subsequence and ignore the rest of the data. This reduces significantly the amount of data being processed.

3.2 HS Problems

As the focus of HS is time and its change, it is suitable for problems that require a model of changes in time. Of course it demands one or more variables with specific patterns: co-occurrences, mutual exclusions, sequences, etc. One restriction in applicability is that variables must have a numerable finite data domain.

Taking into account the mentioned tips, HS can be used to test specific patterns of interest (for instance production failures, software/hardware faults, hang out of processes, deadlocks, etc.) while the main system works. As a consequence it can react to changes of behavior upon those patterns.

3.3 HS Functioning

Here there is a very short description of HS functioning. Since HS is out of the scope of this paper, readers interested in details may find them in [24].

Let a problem $\boldsymbol{R}$ consisting of a set of variables $\{v_i\}$ each one with a specific numerable finite data domain $\boldsymbol{D}_i$.

Let any relevant event $\boldsymbol{e}$ represented by one o more patterns $\boldsymbol{M}_j$, each one a combination of any subset of $\{v_i\}$ with specific values $\{v'_i\}$:$v_i \in \boldsymbol{D}_i$.

Then, the HS model to approximate consists of the union $\cup_i\{\boldsymbol{M}_i\}$, for all the events $\boldsymbol{j}$ whose patterns are being analyzed. And $\boldsymbol{M}_i$ defined as

$$U_{\mathbf{k}} = U_{\mathbf{k}} + \boldsymbol{\eta}_{\mathbf{u}}[U_{\mathbf{k}} - \boldsymbol{\Pi}_{\mathbf{l}} P_{\mathbf{0}}(t_{\mathbf{l}}|M_{\mathbf{i}})]$$

Table 2 Pattern 1

T	Property-1	Property-1
t1 = λ1	PROC = A	USR = 034
t2 = λ2	PROC = C	USR = 035
t3 = λ3	PROC = A	USR = 035

Where $\boldsymbol{U}_{\mathbf{k}}$ stands for time behavior model for pattern $\boldsymbol{M}_{\mathrm{i}}$, $\boldsymbol{t}_{\mathrm{l}}$ is the time elapsed from a previous occurrence of patterns feature $\boldsymbol{l}$, $\boldsymbol{\eta}_{\mathbf{u}}$ is a elasticity parameter for model $\boldsymbol{U}_{\mathrm{k}}$, $\boldsymbol{P}_0(\boldsymbol{t}_{\mathrm{l}}|\boldsymbol{M}_{\mathrm{i}})$ is the Poisson distribution probability for $(\boldsymbol{t}_{\mathrm{l}}|\boldsymbol{M}_{\mathrm{i}})$.

A set of additional parameters for $\boldsymbol{M}_{\mathrm{i}}$ are $\{\lambda_{\mathrm{l}}\}$ where

$$\boldsymbol{\lambda}_l = \boldsymbol{\lambda}_l + \boldsymbol{\eta}(t_l - \boldsymbol{\lambda}_l)$$

with η being a global parameter for all the model that represents a global adaptation coefficient for all the λ_{l} parameters.

In this context, an harmonic is the occurrence in time t_{l} of certain combination of properties that are of interest, and is referred to as pattern $\boldsymbol{M}_{\mathrm{i}}$. For example a pattern may be the one represented in Table 2.

In the table, PROC is a variable representing a software process of a complex system, A is a specific procedure, USR is an user ID that is being running that procedure, and t_1, t_2, t_3 are the typical time elapsed between them (in this example they are set of λ_1, λ_2, λ_3, respectively, as an initialization procedure). Another point of view of this problem is to model the sequence t_1, t_2, t_3, occurrence and variations. It may be used for instance to trigger actions while the sequence is happening or after it. When events match the pattern (a harmonic is found) there is a resonance, and the model may learn any variation in critical parameters.

A resonance has the following steps:

- Pattern detection: Patterns are evaluated against current data (In example 1: Property-1 = A, Property-2 = 034), compare the probability of the pattern against its threshold U (0.3 for example).
- Resonance: when there is resonance, the model parameters are updated.
- Fire an activity (optional) to produce meta-data and tracking data.
- Time information is processed (t_1, t_2, and t_3 in the example) as time-stamp shifts of the events.
- The size n is compared against a certain cut-off threshold nc (i.e., nc = 80). When n < nc, small (n) is true, otherwise it is false. When small (n) gives true, the Binomial dispersion of harmonics is assumed, otherwise it is considered to be Poisson.

3.4 HS Combined with FL

Harmonics may be implemented as data vectors with a predefined timing. But those times are the leading factor for the pattern to be in resonance or not. Thus, whenever

time may relax and the relevant feature of the pattern is the set and organization of variables, then time may be considered not as sharp, but as a fuzzy number.

This Fuzzy Harmonic System (HFS) may be considered as a new approach, and takes traditional Fuzzy Logic (FL) as a shortcut to improve model stability when the patterns have many fluctuations in time. That way, for a narrow set of problems, it is possible to define a self-tuning set of λ_i parameters that converge asymptotically to a static value.

The reason to consider FL is historical. FL is usually considered an extension of classical logic. It can also be thought from the set theory as a sharp set with a fuzzy boundary. In NLP it is usually applied to model semantics and subjective information [29]. Computational Intelligence usually applies FL to a variety of problems, usually with complex and imprecise values.

Among others, additional benefits of fuzzy logic are its simplicity and its flexibility. The main reason to choose Fuzzy logic is not its ability to handle incomplete data, but the possibility to undertake problems with imprecise data, to model nonlinear functions of arbitrary complexity.

Fuzzy logic models are usually called fuzzy inference systems. They consist of a number of conditional "if-then" rules. For the designer who understands the system, these rules are easy to write, and as many rules as necessary can be supplied to describe the system adequately.

The main characteristic in fuzzy logic, unlike standard conditional logic, is that the truth of any statement has a degree. The conditions are usually coded as inference rules of the form A - > B (A implies B). But in FL, it can be said as (0.2 * A) - > (0.5 * B).

For example: the rule

$$\mathbf{A -> B}$$

with

A: module A takes 34 Mb
B: the weather daemon is on

can be restated as

if ***(module A takes 34 Mb)***
then (*the weather daemon must be started*)

Here are two variables: memory consumption for module A, and weather daemon status.

Both can relate to ranges of values (the first in Megabytes and the second a set of possible status).

Fuzzy inference systems rely on membership functions that represents to the computer how to calculate the correct value, between 0 and 1. It is often said that the degree to which any fuzzy statement is true is the denoted by a value between 0 and 1.

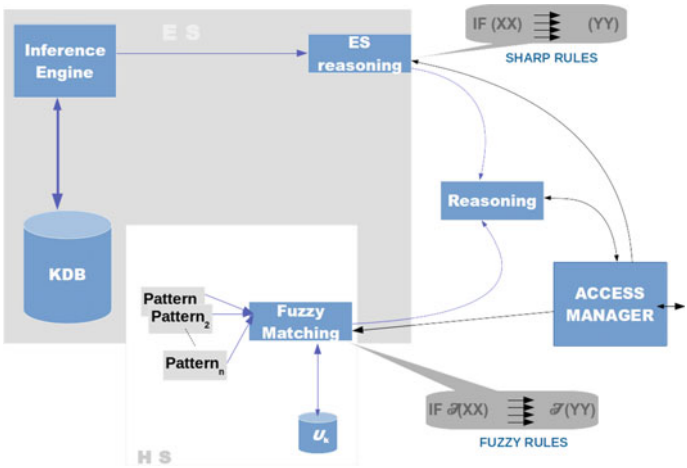


Fig. 2 ES architecture with Module Fuzzy Logic

Another perspective for the same approach is the Fuzzy Set Theory. It was formalized by Professor Lofti Zadeh at the University of California (1965). Zadeh proposed this paradigm with successful application worldwide. In this case, there is a set of rules and regulations that define boundaries and depict the best solution to solve problems restricted to those boundaries. The use of Fuzzy Set Theory from conventional bivalent, sharp sets theory is also considered a paradigm shift.

The use of FL will allow the system to pay attention to boundary events and contexts that have complex resolution. This extra flexibility will pay the cost known as the Non-Free-Lunch theorem [30]. Although there are many speculations regarding it, in general sense it describes that the best an algorithm solves a type of problems, the worse it performs in general. Taking that into account, FL processes facts in a proper way to overcome transient alterations in the system [31, 32].

Figure 2 shows the architecture of the FHS proposal. It has the previous ES with the same components, coordinated with HS. As earlier, both receive the input information from an Access Manager (a module to interface a global controller focused on compatibility with external devices and systems.

The key difference is the set of furry rules that are now biasing the pattern matching process inside the HS. Thus the entire process' performance is being altered.

3.5 HS Specific Features

Harmonics may be implemented as data vectors with a predefined threshold of tolerance for diverges. But some other characteristics of this approach are

- There is no precise time but relative: Time is the duration of certain event, opposed to classic techniques [33] where the value of a certain property is compared at time t_i respective to t_{i-1}, and the magnitude of a property associated. As a consequence there is no comparison between length series or corrections in them due to a different length. Therefore there is no normalization.
- No corrections required: Since no component alignment is required between patterns, distance has no need for corrective techniques such as dynamic time warping, longest common subsequence similarity, local scaling functions, global scaling function, etc.
- Flexibility: HS manages properties being measured as a pattern, which identifies the components in a time series, and models the time dispersion instead of the set of properties inside the pattern.

KRONOS models patterns' time features, and could be analogous to probabilistic similarity measure where methods are model based (they can incorporate prior knowledge into the similarity measure). However, it is not clear whether other problems such as a time series indexing, information retrieval, and clustering can perform efficiently. They use a general similarity approach involving a transformation rules language [1], and hundreds of algorithms from DM to classify, cluster, segment, and index time series.

3.6 HS Filters

Certain problems have too much information throughput, generating an extensive dataset, and making it very hard to perform efficient analysis. In these cases the model may be extended to go through one or more data and/or pattern filters. The effect of this preliminary step is biasing information to focus on specific harmonics. There are three types of filters [24]:

- High-pass filters: They leave the data that are beyond a certain distance (δ) that ($t_{i+\delta}$ < tactual|pattern). Since the pattern's property $p_1..p_i$ is met, t_i exceeds the model value.
- Low-pass filters: They leave the data that are closer than a certain distance ($t_{i-\delta}$ > tactual|pattern). Since the pattern's property $p_1..p_i$ is met, t_i is lower than the model value.
- Band-pass filters: They leave the data that are within a certain distance range ($t_{i+\delta}$ > tactual|pattern > $t_{i-\delta}$). Since the pattern's property $p_1..p_i$ is met, t_i is within the model value with a certain distance.

4 The KRONOS Prototype

This section presents a summary of KRONOS presented as a proposal in [26–28].

4.1 Architecture

Kronos is a prototype that implements a model for time predictions. After collecting data from many sources, an Expert System interacts with a Knowledge Base and an intelligent HS subsystem. Its goal is to evaluate any traffic and pedestrian risks. The global design is able to interact with diverse and mobile devices, other information systems, user data, and Internet (Fig. 3).

Main components are

- **Web**: It is a source of information and requests. A web server, web service, a local server, or other host may connect with the prototype using a proper interface represented in the picture with this module.
- **Host**: The prototype has a rule-based Expert System for data prediction [24], to evaluate risks based on expert knowledge, inputs and historical statistics. It interacts with the HS subsystem to dynamically build a more precise model.
- **Input Device**: Information regarding current position, status, and requests may be provided to the prototype by one or more mobile devices, sensors, etc. Each one requires a specific interface.
- **Output Device**: As it may be used by pedestrians and drivers, it is expected to output information upon requests though mobile devices' interfaces.
- **External System**: Already existent systems may interact with the prototype using the interface represented here as this module.

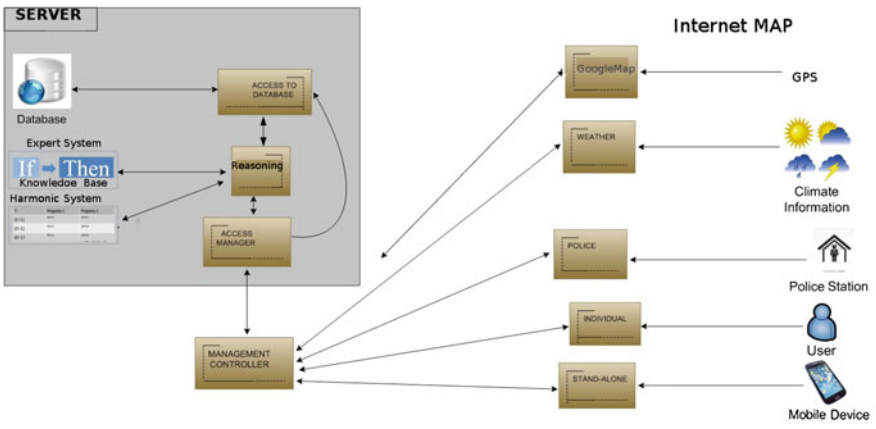


Fig. 3 KRONOS architecture

- **DBMS**: it is a Postgres database with a set of store procedures and triggers that automatically check and transform data.

4.2 Characteristics of Time Mining

Data may be collected from many sources and converted to be compatible with internal and DBM's requirements. They have the following characteristics:

- Belong to an external entity.
- Have a time stamp with a least date, time, minutes, and seconds.
- May be associated with a cyclic and complex event producing a measurable set of features.
- The source of the data is one or more identifiable and distinguishable sources.
- Events are well defined and limited in time.
- Their duration is variable or constant but they start and end at a defined point in time.
- They may undergo duration and frequency changes, but not changes in identifying characteristics.

5 Test and Evaluation

As mentioned previously, the prototype partially implements the model. The ES core is functional, it has many rules according to expert recommendations. The interfaces to user, DBMS, Maps, DBMS, are working and also part of the Harmonics system. The tests in this section use ES knowledge but not the Harmonics system, since the goal of this paper is to provide the improvement acquired by considering GPS in the risk inference.

5.1 Database

The test set are risk situations evaluated post-Morten to be able to find the accuracy of the results.

All the dataset belongs to real situations of traffic in Concepción del Uruguay city (Entre Ríos, Argentina). Figure 4 shows the map of the city.

As can be seen the map has a grid. Each cell represents a zone, described in the Database as follows: (ID-Zone, Description, Latitude, Longitude, Risk). ID-Zone is an integer that identifies the cell. Description outlines the zone. Table 3 describes the reference values and the labels in Fig. 4.

Fig. 4 City map

Table 3 Zone labeling

Risk	Risk Type	Label
0	No	N
1	High	A
2	High	A
3	Medium	M
4	Low	None
5	Low	None

The testing resulting in the results summarized in Table 4.

In Table 4, column Desc. (Description field), has the values: (c)enter, (r)oundabout, c(o)untry road, country (z)one. Column S (Subject) has the values: (d)river, (p)edestrian.

To assess the impact of considering GMT (Zone) information, the dataset was reevaluated avoiding that variable from the rules. Then, the predictions were compared to expert predictions. The number of hits and errors are in Table 5.

The legend (~G) means without GMT information, and (G) means with GMT information. Results indicate higher accuracy (94% vs. 42%). Also the system trends to underestimate risk. Analyzing these three test cases, the deviation occurs for drivers that do not use helmet/belt and are in a safe zone of the city, during the daylight hours but when weather has reduced visibility. Figure 5 shows results test by test.

The (E)xpert prediction is the darkest curve. Without GMT information (~ G) the ES prediction results are more erratic.

Table 4 Testing results

Risk P/G	Dd/mm/yy	Time	Weather	Desc.	S	km/h	Belt/helmet	Alc. (g/l)
0	01/05/15	13:55	Sunny	C	D	40	Si	0.9
0	22/05/15	12:10	Sunny	R	D	90	No	0
0	12/06/15	22:29	Sunny	O	D	40	No	0.3
0	24/11/15	08:32	Foggy	O	D	30	No	0
0	01/05/15	13:56	Sunny	C	P			0.3
0	08/05/15	11:30	Cloudy	Z	P			1
1	12/06/15	22:29	Cloudy	Z	D	122	No	0
1	28/06/15	17:43	Cloudy	C	D	140	No	0.4
1	12/06/15	23:44	Cloudy	C	P			0.2
1	12/06/15	23:45	Cloudy	C	P			0.5
1	23/11/15	23:32	Cloudy	C	D	190	No	0
1	23/11/15	23:50	Cloudy	C	D	230	No	0.9
1	12/06/15	22:20	Cloudy	Z	D	130	No	0.5
3	12/06/15	23:03	Cloudy	C	D	100	No	0
3	12/06/15	23:13	Cloudy	Z	D	100	No	0.2
3	08/09/15	21:45	Cloudy	C	D	150	No	0.1
3	02/10/15	19:21	Sunny	C	D	190	No	0.9
3	02/10/15	19:48	Sunny	C	D	190	No	0.9
3	12/06/15	23:44	Cloudy	C	P			0.51
3	12/06/15	23:44	Cloudy	C	P			1.25
0	25/11/15	11:41	cloudy	C	P			0
0	25/11/15	12:36	Fog	Z	P			2
0	25/11/15	19:36	Sunny	C	P			1.9
0	25/11/15	02:36	Sunny	C	P			1.9
0	25/11/15	02:36	Sunny	Z	P			1.5
1	25/11/13	13:53	Sunny	C	D	120		1
1	12/08/15	19:30	Cloudy	Z	D	140		0.19
1	17/08/10	19:30	Cloudy	Z	D	140		0.16
1	17/08/10	19:30	Cloudy	Z	D	125		0.18
1	17/08/10	20:30	Cloudy	Z	D	125		0.25
2	25/11/15	20:14	Cloudy	C	D	90		1
2	25/11/15	14:21	Cloudy	C	D	40		2
2	25/11/15	14:21	Cloudy	C	D	50		2
2	25/11/15	14:21	Cloudy	C	D	50		2
2	25/11/15	14:21	Cloudy	C	D	60		2
3	25/11/15	14:36	Sunny	Z	D	40		0.1
3	25/11/15	20:36	Sunny	Z	D	40		0.1
3	25/11/15	20:36	Sunny	Z	D	60		0.1
3	25/11/15	20:36	Sunny	Z	D	90		0.1

(continued)

Table 4 (continued)

Risk P/G	Dd/mm/yy	Time	Weather	Desc.	S	km/h	Belt/helmet	Alc. (g/l)
3	25/11/15	14:36	Sunny	Z	D	35		1
5	31/05/11	23:22	Cloudy	C	P			0.13
5	04/05/15	23:22	Cloudy	C	P			0.1
5	25/11/15	15:24	Sunny	C	P			0.1
5	25/11/15	20:00	Sunny	C	P			0.14
5	25/11/15	20:00	Sunny	C	P			0.1
5	09/07/14	23:48	Sunny	C	P			0.11
5	09/07/14	23:48	Sunny	C	P			0.1
5	09/07/14	23:48	Sunny	C	P			0.05
5	09/07/14	23:48	Sunny	C	P			0.03
5	09/07/14	23:48	Sunny	C	P			0.09

Table 5 Results with and without GMT

	Error (G)	%	Error (G)	%
Overestimated	11	22.00	0	00.00
Underestimated	18	36.00	3	06.00
OK	21	42.00	47	94.00

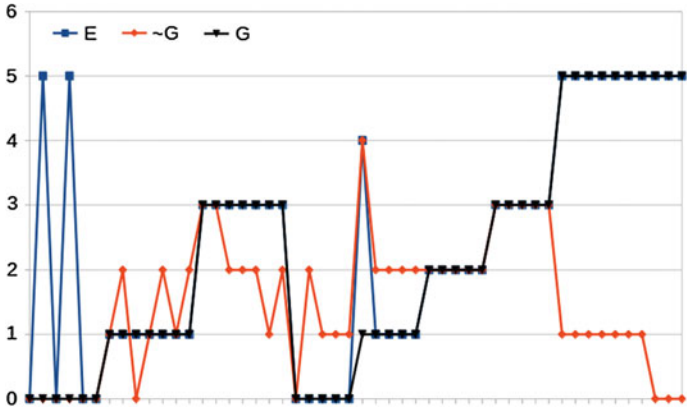


Fig. 5 Risk according Expert (E), ES without GMT (~G) and with GMT (G)

6 Conclusion and Future Work

This paper presents an outline of the KRONOS project, a prototype of a model with a dual risk evaluation mainly using statistical and heuristic approaches respectively. The key features of each one were presented as well as basic statistical information regarding how the statistical inferences can be improved using GPS information.

The comparison between the accuracy acquired in previous work has increased to 52%, and it is possible to say that GPS information has a good impact in the results. This trend must be verified with a larger number of test cases.

As a future work it remains to test HS as it was performed with ES, and find out how both perspectives may be combined to provide better results. Also a FL treatment for patterns is pending for implementation and statistical evaluation.

References

1. López De Luise, D.: MLW and bilingualism. Advances Research and Trends in New Technologies, Software, Human-Computer Interaction, and Communicability. IGI Global, USA (2013)
2. Wash, O.: Assessing pedestrian risk locations: a case study of WSDOT efforts. Department of Transportation. Washington State Library. Electronic State Publications (1998)
3. Organization for Economic Co-operation and Development (OECD): http://www.oecd.org (2014)
4. Alex Quistberg, D., Jaime Miranda, J.: Beth Ebel. Reducing pedestrian deaths and injuries due to road traffic injuries in Peru: interventions that can work. Revista Peruana de Medicina Experimental y Salud Pública. Rev Peru Med Exp Salud Publica vol. 27 n. 2 Lima Apr/Jun 2010. ISSN 1726-4634
5. Oxley, J.: Improving Pedestrian Safety. Curtin—Monash Accident Research Centre. Fact Sheet No. 6. (2004)
6. Cathy, T., D. Packman, Risk and safety on the roads: the older pedestrian. Foundation for Safety Research. New Castle University (1995)
7. Generic Risk Assessment 4.1. TSO Publisher: www.tsoshop.com.uk (2009)
8. Antov, D., Rõivas, T., Antso, I., Sürje, P.: A method for pedestrian crossing risk assessment. Transaction of Wessex Institute. doi:10.2495/UT110501 (2011)
9. Traffic Management Hazard Identification & Risk Assessment Control Form: Swinburne University of Technology. http://www.docstoc.com/docs/24507714/Traffic-Management-Health–Safety-Checklist (2009)
10. Alavi, H., Charlton, J., Newstead, S., Archer, J.: A Pedestrian Data System for Safety Analyses. Monash University Accident Research Center (MUARC), Melbourne, Victoria, Australia (2014)
11. Jiao, J., Moudon, A.: Using a Case-control approach and GIS Methods to Assess the Risk of Pedestrian Collision In Seattle, USA
12. Kim, K., Sul, J.: Development of Intersection Traffic Accident Risk Assessment Model Transportation & Environment Research Institute Ltd. kijoonkim@hotmail.com, 82-(0) 2-10-8752-1851 (2001)
13. Hautzinger, H.: Analysis Methods for Accident and Injury Risk Studies. Project No. 027763 TRACE. Deliverable 7.3.2007
14. Journey Risk Management® (JRM®): http://www.irte.com/journey-risk-management.html
15. Rodríguez-Hernández, M., Campuzano-Rincón, J.: Primary prevention measures for controlling pedestrian injuries and deaths and improving road safety. Revista Salud Pública. Rev. salud pública vol. 12 no. 3 (2010)
16. Thomas, L., Hamlett, C., Hunter, W., Gelinne, D.: Final Report to North Carolina Department of Transportation. North Carolina Department of Transportation, Traffic Engineering and Safety Systems Branch (2009)
17. Hart, J.: Measuring Pedestrian Risk and Identifying Methods to Prevent Pedestrian Accidents in Langley Park. National Fire Academy (2004)

18. Health and Safety Comission: Reducing at-work road traffic incidents. Report to Government and the Health and Safety Commission. DTLR (2001)
19. Han, J., Kamber, M.: Mining Stream. Time-Series, and Sequence Data. In Data Mining, 2nd edn. Concepts and Techniques, 2nd edn (2011)
20. Shieh, J., Keogh, E.: iSAX: indexing and mining terabyte sized time series. In: Proceedings KDD'08, pp. 623–631. ACM (2008)
21. Tak-chung, F.: Engineering applications of artificial intelligence. Eng. Appl. Artif. Intell. **24**, 164–181 (2011)
22. Bollogás, B., Das, G., Gunopulos, D., Mannila, H.: Time series similarity problems and well-separed geometric sets. Nordic J. Comput., 409–423 (2003)
23. Jagadish, H., Mendelzon, A.: Similarity based queries. In: Proceedings of the 14th PODS, vol. 95, pp. 36–45 (1995)
24. López De Luise, D.: Harmonics systems for time mining. Int. J. Modern Eng. Res. (IJMER) 3 (6, 3), 2719–2727 (2013). ISSN:2249-6645
25. Instituto Nacional de Estadísticas y sensos: México. http://www3.inegi.org.mx
26. Acuña, I., García, E., López De Luise, D., Paredes, C., Celayeta, A., Sandillú, M., Bel, W.: Traffic & Pedestrian risk inference using Harmonic Systems. SOFA, Romania (2014)
27. Celayeta, A., Paredes, C., López De Luise, D., Bel, W.: Traffic and Pedestrian risk evaluation with Harmonic Systems (Cálculo de Riesgo en tráfico y Peatón usando Sistemas Armónicos). ARGENCON (2014)
28. Mansilla, D., Sandillú, M., López De Luise, D., Bel, W.: Un Modelo con Conocimiento Experto y Sistemas Armónicos para Evaluación de Riesgos de Tráfico. EnIDI Argentina (2015)
29. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Syst. Man Cybern. **3**(1), 28–44 (1973)
30. Wolper, D.H.: The supervised learning no-free-lunch theorems. In: Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications (2001)
31. Zadeh, L.A.: A note on prototype set theory and fuzzy sets. Cognition **12**, 291–297 (1982)
32. Zadeh, L.A.: Fuzzy Sets and Applications (Selected Papers, ed. by R.R. Yager, S. Ovchinnikov, R.M. Tong, H.T. Nguyen. Wiley, Nueva York (1987)
33. Ratanamahatana, C.A., Lin, J., Gunopulos, D., Keogh, E.: Mining time series data. In: Data Mining and Knowledge Discovery Handbook, pp. 1049–1077 (2010)
34. De Nicolao, G., Ferrara, A., Giacomini, L.: On-board sensor-based collision risk assessment to improve pedestrians' safety. IEEE Trans. Veh. Technol. **56**(5), 2405–2413 (2007). doi:10.1109/TVT.2007.899209

Performance Aspects of Alamouti STBC for MIMO Channels Affected by Impulsive Noise

Mihaela Andrei and Viorel Nicolau

Abstract In general, wireless communications are affected by noise and by time-varying characteristics of propagation environment. Space-time block codes are an effective method to combat fading and also to provide spatial diversity. Besides fading and additive white Gaussian noise (AWGN), we considered two types of impulsive noise described by Middleton Class-A (AWCN) and symmetric α-stable (SαS) distributions. The AWCN model was used to highlight the Alamouti code diversity, compared with the situation when the channel is only affected by AWGN. Even in the presence of non-Gaussian noise, the diversity at both the reception and transmission has decreased the number of errors. Alamouti 2×2 performances for all three types of aforementioned noise are presented; the evaluation was performed considering the Bit Error Rate (BER) curves depending on signal-to-noise ratio. We have also used this code in image transmission in the presence of SαS noise. For all simulations, data was binary phase-shift keying (BPSK) modulated and the fading was Rayleigh type. Different values of the parameters that describe the noise models were considered.

Keywords Alamouti code · Alpha stable distribution · Image transmission · Impulsive noise · Middleton Class-A noise

M. Andrei (✉) · V. Nicolau
Department of Electronics and Telecommunications, "Dunarea de Jos"
University of Galati, Galati, Romania
e-mail: mihaela.andrei@ugal.ro

V. Nicolau
e-mail: viorel.nicolau@ugal.ro

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_8

1 Introduction

Wireless communication systems are currently in the spotlight, due to their high usage in human activities. The safety of data transmitted by such systems, on channels affected by fading, is considerably improved by using space-time block code (STBC).

STBCs ensure protection, especially at high speeds [1], and furthermore, they accomplish transmission diversity [2]. The simplest scheme is that proposed by Alamouti, with two emitting antennas [3]. This scheme is an important accomplishment in the field of communications, because it leads to good performances, in spite of having a simple decoder.

In general, any process is affected by various additive or multiplicative disturbances. In communication systems, the perturbations are additive. These are generated by various sources at different points of time and space and then are propagated through communication channels to the receivers, where they arrive as a combination of noise signals, independent or correlated. The noise that could affect the data transmission on multiple-input multiple-output (MIMO) channels can be additive white Gaussian noise or non-Gaussian (impulsive noise). Impulsive noise is an additive disturbance, independent of background noise, active at different moments of time, as very short pulses. In addition, it is a non-stationary process, whose statistical parameters may vary in time.

The main characteristic of this type of noise is high value of instantaneous power and average power ratio. As a result, impulsive noise is a significant source of errors if the pulses occur frequently and their amplitudes are much higher than background noise [4]. There are various sources that can produce non-Gaussian noise such as: automotive ignition, refrigerators, printers, microwave ovens [5], or network interference [6].

Most of the space-time block code receptors were designed for the AWGN case. That is why, in the presence of impulsive noise, their performance drops significantly, compared to the AWGN case, especially for high values of signal-to-noise ratio (SNR) [7].

Because this type of noise is often present, communications research required the development of statistical models enable to characterize it. So, in [8] Hall proposed an exogenous model, where impulsive noise is generated as a product of two independent random processes. Shao and Nikias established a symmetric stable distribution, characterized by an exponent term α and known as symmetric α-stable (SαS) distribution [9]. In many applications, for impulsive noise a canonical model is used, proposed by Middleton [4]. In this paper, Middleton Class-A and SαS distributions were considered.

In the case of MIMO communication channels with multiple receivers, the Middleton Class-A impulsive noise models are multivariable extensions of the monovariable distributions (valid for channels with a single receiver). Based on the sources of interference spatial distribution relative to the receivers, there are three types of multivariable models for Middleton Class-A impulsive noise [10]:

(a) the impulsive noises from the receivers are considered random variables, independent in space and time and evenly distributed. In this case, the noise sources are independent and each antenna receives an independent impulsive noise, generated by a monovariable probability density function (pdf). Often, the same parameters are used for all the probability densities of the random variables;
(b) the impulsive noises from the receivers are considered temporal-independent random variables, but spatially dependent and correlated between the receiving antennas. In this case, all the receiving antennas are under the influence of the same set of noise sources. Furthermore, the spatial dependency implies that the distance between the interference sources and the antennas is much greater than the distances between the antennas. As such, there is no difference between the distances from a source of interference to each antenna, and practically, the antennas receive impulsive noises that are more or less the same. The multivariable model uses a monovariable pdf, extended by the noise covariance matrix. This is the case considered in our paper.
(c) the impulsive noises from the receivers are considered temporal-independent random variables, but spatially dependent and uncorrelated.

When analyzing the behavior of communication systems in various situations and conditions, the Middleton Class-A noise model is used very often. Some of the results target wireless communication systems, like: IEEE 802.11a and IEEE 802.11b [11], other the power line communication [12]. In both cases, the system performances on a channel affected by non-Gaussian noise are significantly lower against AWGN for high signal-noise ratio (SNR) values. In a MIMO power line communication system, if the noise gets more impulsive, the Bit Error Rate (BER) increases [12]. For a MIMO system with orthogonal space-time coding (OSTBC), QPSK and 16QAM modulations, a coding gain of about 6 dB was obtained in the case of AWCN channel compared to AWGN, for low SNR [13]. If the SNR increases, the situation reverses.

In the case of the SαS distribution the situation is similar to Middleton Class-A, i.e., the non-Gaussian noise effects on MIMO systems worsen their performances. However, until now, there is no closed-form expression for the error probability [14], except for the optimal linear receivers in a single-input single-output system [15]. For space-time codes over a channel affected by fading and impulsive noise modeled SαS, [16] used Monte–Carlo simulations to compare the performance of the different decoders. The maximum-likelihood (ML) receiver leads the best performances.

This paper analyzes the Alamouti code spatial diversity on a channel affected by Middleton Class-A impulsive noise compared with an AWGN channel, for varying degrees of impulsivity: from almost Gaussian to strongly impulsive noise, given by the model parameters. It investigates the performances of Alamouti STBC with two transmitting and two receiving antennas over a channel affected impulsive noise

with SαS distribution, for a ML receiver and different values of the exponent parameter α. This type of noise was used in image transmission with the aforementioned code. In all cases, the fading was considered to be of Rayleigh type and data was BPSK modulated.

The paper is organized as follows. In Sect. 2, the impulsive noise models for the two types of noise are described and Sect. 3 presents the system model. Simulation results are presented in Sect. 4 and Sect. 5 concludes the paper.

2 Impulsive Noise Models

2.1 Symmetric-Alpha Stable (SαS) Distribution

First, we assume the SαS model to represent the impulsive noise. Some sources of impulsive noise are: underwater acoustics, low-frequency atmospheric noises and many more man-made noises [17]. Its characteristic function is [18]:

$$\varphi(t) = \exp\{j\delta\gamma - |\sigma t|^{\alpha}(1 - j\beta sign(t) \cdot w(t, \alpha))\}, \tag{1}$$

where

$$w(t, \alpha) = \begin{cases} \tan(\pi\alpha/2), \alpha \neq 1 \\ -\frac{2}{\pi}\log|t|, \alpha = 1 \end{cases} \tag{2}$$

The significance of variables in (1) is as follows [17]:

- $\alpha \in (0, 2]$—is the characteristic exponent. This parameter is the one who influences the thickness of the distribution tail. When $\alpha = 2$, the process becomes Gaussian.
- γ—represents the dispersion parameter. It is analogous to the variance from the Gaussian case.
- μ—is the location parameter and its corresponding parameter in the normal distribution is the mean.
- $\sigma \in (0, \infty)$ is the scale;
- $\beta \in [-1; 1]$—determines the distribution symmetry. Thus, if $\beta = 0$, the distribution is symmetrical about μ, if $\beta < 0$ the distribution is skewed to the left, and if $\beta > 0$—to the right.

So far, no expression has been set for the probability density function that describes the SαS distribution. However, there are two exceptions: Gaussian (if $\alpha = 2$) and Cauchy (if $\alpha = 1$) [19].

2.2 Middleton Class-A Model

The Middleton Class-A distribution statistically models the sum of electromagnetic interferences from multiple noise sources, spatially spread in an annular area around the receiver, following a Poisson distribution pattern for magnitude and uniform distribution, within the [0, 2π] interval for phase [20]. Unlike the SαS distribution, the Middleton Class-A can also include the white noise from the receiver, without changing the nature of distribution [10].

A sample of Middelton Class-A impulsive noise is given by: $n = n_g + n_i$, where n_g represents the Gaussian component and n_i is the impulsive component [21], with their variances: σ_g^2 and σ_i^2, respectively. Non-Gaussian type impulsive noise that follows the aforementioned distribution has the probability density function [22]:

$$p(n) = \sum_{m=0}^{\infty} \frac{A^m e^{-A}}{\sqrt{2\pi} m! \sigma_m} \exp\left(-\frac{n^2}{2\sigma_m^2}\right) \tag{3}$$

In the above pdf expression, m is the number of impulsive noise sources and A is the impulse index [22]. The term $m = 0$ is assigned to the Gaussian background noise component and the remaining summed components, indexed with $m > 0$, represent the impulsive noise, as a results of a sum of interferences from noise sources, spatial spread following a Poisson distribution.

Two important parameters describe the Middleton Class-A distribution: A and T. Parameter A is called impulsive index or overlapping and is the product of the average number of impulses that reach at the receiver in one second (v) from the noise sources and their average duration (T_m): $A = v \cdot T_m$ [23]. This parameter shows how impulsive is the noise at the receiver, as a result of the interference from noise sources. Depending on A value, for each moment of time, from total number of noise sources considered, only some of them will have a significant contribution in the noise of the receiver. Thus, if A has high values, it results a high density of waveform overlapping at a time and the noise is less impulsive, looking almost Gaussian (according to the Central Limit Theorem). Conversely, low values for A indicate a small overlap, so only a few sources interfere and the noise gets more impulsive. σ_m^2 is given by

$$\sigma_m^2 = \sigma^2 \cdot \frac{\frac{m}{A} + T}{1 + T}, \tag{4}$$

where $\sigma^2 = \sigma_g^2 + \sigma_i^2$ is the total noise power and

$$T = \frac{\sigma_g^2}{\sigma_i^2} \tag{5}$$

is the Gaussian factor. Analogous to parameter A, if T gets lower, the noise gets more impulsive, and if T has high values, the distribution will approach the Gaussian one.

3 STBC Model

A general scheme for a MIMO communication system, with N_T emitting and N_R receiving antennas is presented in Fig. 1. The use of this type of system significantly improves communication by providing diversity at the reception and/or emission. In this paper, we consider $N_T = 2$, $N_R = 2$, a BPSK modulator and the transmitted data are STBC encoded, which means that each antenna transmits a different version of the same input. This is an advantage because at reception it will get more signal copies, which will be affected differently by noise, interference or fading. A space-time block decoder uses all these copies to remake the transmitted data and thus, the number of errors will be lower.

The relation that describes a MIMO channel is [24]:

$$\mathbf{r} = \mathrm{H} \cdot \mathrm{x} + \mathrm{n} \tag{6}$$

where: **r** is the array of received signals, **H**—the channel matrix, **x** include the transmitted signals and **n**—the noise samples.

The channel matrix elements are the channel fading coefficients between the emitting and the receiving antennas. These can vary in time; so, at moment t, the matrix form is:

$$\mathrm{H}_t = \begin{bmatrix} h^t_{1,\,1} & h^t_{1,\,2} & \cdots & h^t_{1,\,N_T} \\ h^t_{2,\,1} & h^t_{2,\,2} & \cdots & h^t_{2,\,N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h^t_{N_R,\,1} & h^t_{N_R,\,2} & \cdots & h^t_{N_R,\,N_T} \end{bmatrix} \tag{7}$$

If the channel matrix **H** varies slowly in time, being constant during the transmission of an entire frame with L symbols, but changing from frame to frame, then the channel is quasistatic or slow fading. In this case, the channel parameters vary more slowly than those of the base-band signal, and the channel coherence time,

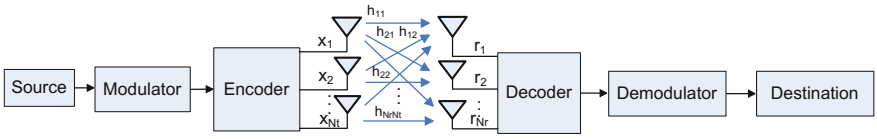


Fig. 1 The general scheme for a MIMO system with STBC encoder

denoted t_c, is bigger than the time of frame transmission, T_F: $T_F = L \cdot T < t_c$, where L is the number of symbols in the frame, and T is the transmission time of a symbol.

If the channel matrix **H** remains constant during the symbol transmission, but varies from symbol to symbol during the frame transmission, then the channel is fast fading. In this case, the coherence time is: $T < t_c < L \cdot T$.

In this study, the fading is considered to be flat and of Rayleigh type, and the coefficients to be random complex Gaussian variables, with identical distribution with zero mean and unit variance [24].

For every time moment *t*, the signal received by antenna *j* will be a linear combination of all signals, with fading and noise, and it will be given by [24]:

$$r_j^t = \sum_{i=1}^{N_T} h_{ji}^t \cdot x_i^t + n_j^t, \tag{8}$$

The scheme proposed by Alamouti has two emitting antennas and N_R receiving ones. The signals are BPSK modulated and they are transmitted as Alamouti technique: at moment *t*, the first antenna emits x_1, the second one x_2 and at the moment $t + 1$, $-x_2^*$ and x_1^*, respectively, where x^* is a complex conjugate of x [3].

According to relation (8), the signals received by antenna *j* are [24]:

$$\begin{cases} r_{j,1} = h_{j,1} \cdot x_1 + h_{j,2} \cdot x_2 + n_{j,1} \\ r_{j,2} = -h_{j,1} \cdot x_2^* + h_{j,2} \cdot x_1^* + n_{j,2} \end{cases} \tag{9}$$

The matrix form is

$$r_j = [\, r_{j,1} \quad r_{j,2} \,] = [\, h_{j,1} \quad h_{j,2} \,] \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix} + [\, n_{j,1} \quad n_{j,2} \,] \tag{10}$$

The estimated symbols $\hat{x}_1$ and $\hat{x}_2$ are given using the square Euclidean distance between the received sequence and the alleged received one. Therefore, the decoder uses the maximum-likelihood algorithm. The complexity of this type of decoder depends on the number of antennas and the modulation that was used (BPSK modulated symbols are easiest to decode). As this increases, the decoding becomes more difficult.

4 Simulation Results

This part of the paper contains three subparagraphs, which present the results as follows: Alamouti code spatial diversity analysis on an AWCN channel (Sec. 4.1), the aforementioned code performances on channel affected by SαS noise (Sec. 4.2), and visual and quantitative results on image transmission using Alamouti STBC (Sec. 4.3). For all simulations, we considered a channel affected by noise (Gaussian or impulsive), Rayleigh slow fading and BPSK modulation. The ML receiver was

used. The impulsive noises, Middleton Class-A and SαS, were generated by the Interference Modeling and Mitigation Toolbox [25].

4.1 Alamouti Code Spatial Diversity Analysis on AWCN Channel

Alamouti created a space-time code with superior performances, which combats fading and ensures diversity on an AWGN channel with Rayleigh-type fading [3]. In order to highlight the benefit of this diversity on an AWCN channel also, we performed simulations for $N_R = 2$ and 4 receiving antennas and a random sequence of input data, of dimension $N = 100$ and $N = 1000$, respectively. The following situations will be considered: the transmission is affected only by the channel fading (flat fading of type Rayleigh), in which case the H matrix is considered to be perfectly known or known with uncertainty—for this case we chose o small input sequence, of size 100; fading and background noise (Gaussian) and with Middleton Class-A additional impulsive noise, respectively. The impulsive noise model parameters were varied like this: *(A, T)* = (0.1; 0.1), (0.01; 0.01). For the last scenarios, the input sequence was considered to be of size $N = 10000$.

(a) The first case considered is that when the transmission is affected only by the fading, in the absence of background noise or other sources of non-Gaussian noise. The simulations were done for a small set of input data ($N = 100$), two emitting antennas, two receiving antennas and the H matrix known at reception. In Fig. 2, we represented the symbols received by each antenna and also the estimated symbols. It can be observed that the received symbols estimation is not in the C constellation theoretical values, but within values corrected with the energy transmitted per symbol and number of transmitters. For $N_T = 2$, the estimated values of the symbols are in the $\pm 1/\sqrt{2}$ points. Even though the transmitted symbols are received with errors (caused by fading), the decoder can correct these errors, if the channel's H matrix is perfectly known.
If the H matrix is known with some uncertainty, then the estimated values will have the same uncertainty also, but the decoder will be able to correct this estimation error. We considered two situations: when the H matrix is known at reception with 20% uncertainty and 50%, respectively. The symbols' estimated values for each situation are represented in Fig. 3a and b. The symbols are no longer in the constellation points, but around them, and it can be observed that there are no decoding errors (the errors will be red).

(b) When the transmission is affected by Rayleigh fading and AWGN noise, we consider $N_R = 2$. For simulations we used a larger data set ($N = 10000$), assumed the H matrix is known and that SNR = 5, 7, 10 dB.
For SNR = 10 dB the decoding is done with a number of $N_{err} = 11$ errors. The estimated symbols' "distribution," along with the received values, is represented in Fig. 4. The red dots represent the wrongly classified points. They are

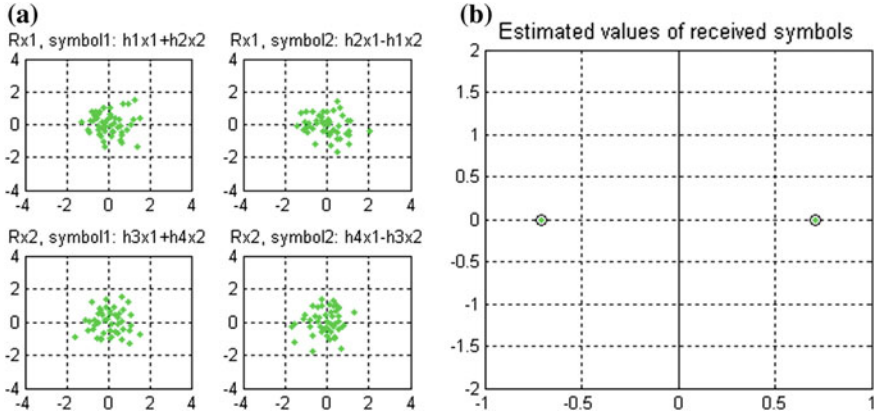


Fig. 2 Channel affected by fading and the H matrix known at reception. **a** Symbols received by each antenna; **b** estimated symbols

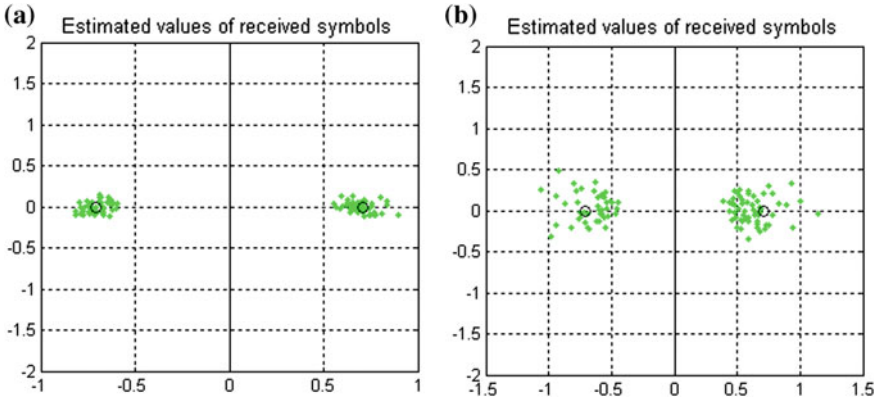


Fig. 3 Channel affected by fading and the H matrix known at reception with an uncertainty of **a** 20%; **b** 50%

placed in one of the semi-planes, left or right of the ordinate (having the real part negative or positive), but they should be in the other semi-plane.
For SNR = 7 and 5 dB, the number of errors rises N_{err} = 53 errors—at 7 dB, and 162 errors—at 5 dB, respectively, and the dots will be "distributed" like in the Figs. 5 and 6. It can be observed that most of the erroneous symbols are very close to zero.

(c) When transmission is affected by fading and AWGN noise, for N_R = 4 and H matrix known at reception, we considered the same large data set (N = 10000) and SNR = 7 dB and 5 dB, respectively.
For SNR = 7 dB, decoding is done with a number of N_{err} = 30 errors, and for SNR = 5 dB, N_{err} = 103 errors. The estimated values are represented in Fig. 7.

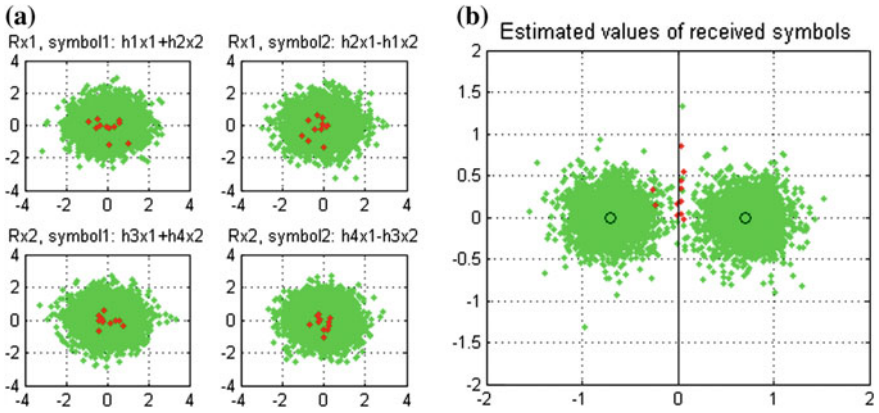


Fig. 4 Channel affected by fading and AWGN noise, $N_R = 2$. **a** Symbols received by every antenna; **b** symbols estimated at SNR = 10 dB

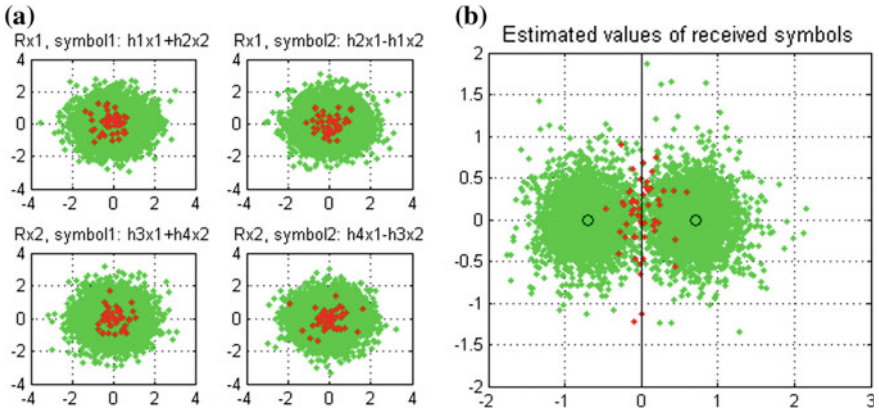


Fig. 5 Channel affected by fading and AWGN noise, $N_R = 2$. **a** Symbols received by each antenna; **b** estimated symbols, SNR = 7 dB

(d) When transmission is affected by fading, by background (Gaussian) noise and by impulsive noise, we considered $N_R = 2$, matrix H known at reception, SNR = 10 dB and *(A; T)* = (0.1; 0.1), (0.01; 0.01). The symbols distribution is given in Fig. 8. For $A = T = 0.1$, $N_{err} = 102$ errors, and for $A = 0.01$ and $T = 0.01$, the decoding is done with a number of $N_{err} = 65$ errors. A much greater spread of the wrongly classified points can be observed (compared to the AWGN case). These points, being affected by impulsive noise, have considerably passed in the opposed semi-plane. The number of errors is smaller in the case of strong impulsive noise; this has an explanation easily deductible from Fig. 8: in the case of strong impulsive noise, the impulses have greater amplitude, but they are more rare, which places the impulses far away from the

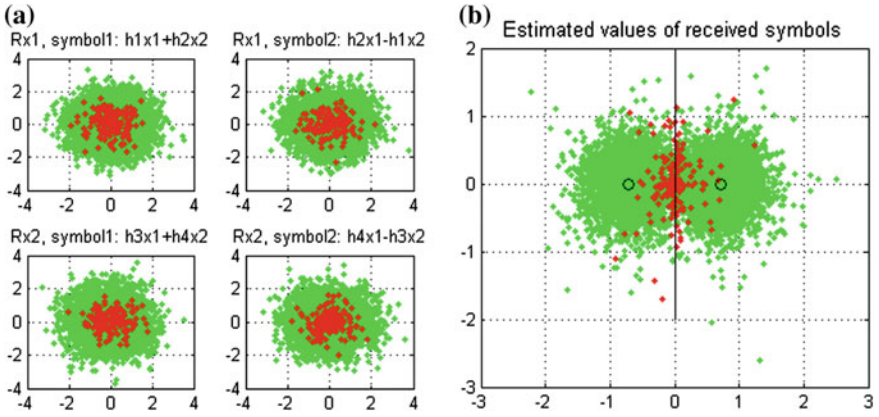


Fig. 6 Channel affected by fading and AWGN noise, $N_R = 2$. **a** Symbols received by each antenna; **b** estimated symbols, SNR = 5 dB

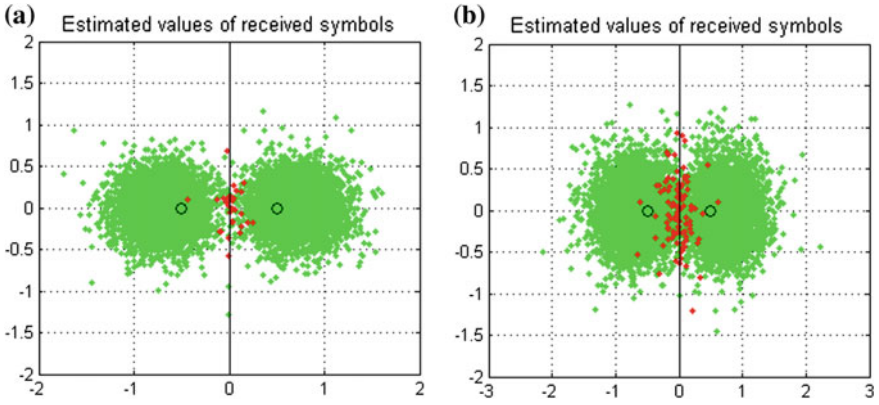


Fig. 7 Channel affected by fading and AWGN noise, $N_R = 4$. **a** SNR = 7 dB; **b** SNR = 5 dB

constellation's points. For parameters $A = T = 0.1$, the points are much closer to the constellation's ones, the erroneous ones being concentrated near 0, while for $A = T = 0.0.1$, the wrongly classified symbols are "all over the place," being less numerous, but with larger values.

(e) The last case we considered is the one when the transmission is affected by channel fading, background (Gaussian) noise and impulsive noise, but for $N_R = 4$. Assuming the H matrix is known, for SNR = 10 dB and $A = 0.1$ and $T = 0.1$, the decoding is done with a number of $N_{err} = 87$ errors, and for $A = 0.01$ and $T = 0.01$, the decoding is done with a number of $N_{err} = 52$ errors. The symbols "distribution" is represented in Fig. 9. In this case, the numbers of errors for the two parameter sets of the Middleton Class-A noise model are comparable, which means that indeed it's lucrative to have diversity

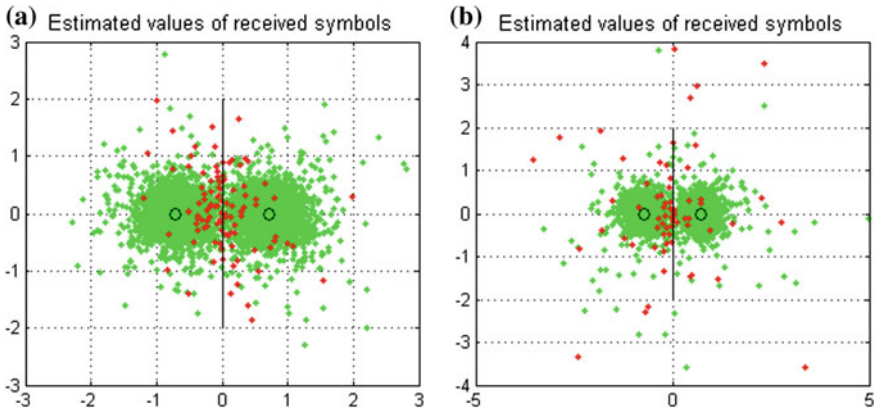


Fig. 8 AWCN channel affected by fading, $N_R = 2$, SNR = 10 dB. **a** $A = T = 0.1$; **b** $A = T = 0.01$

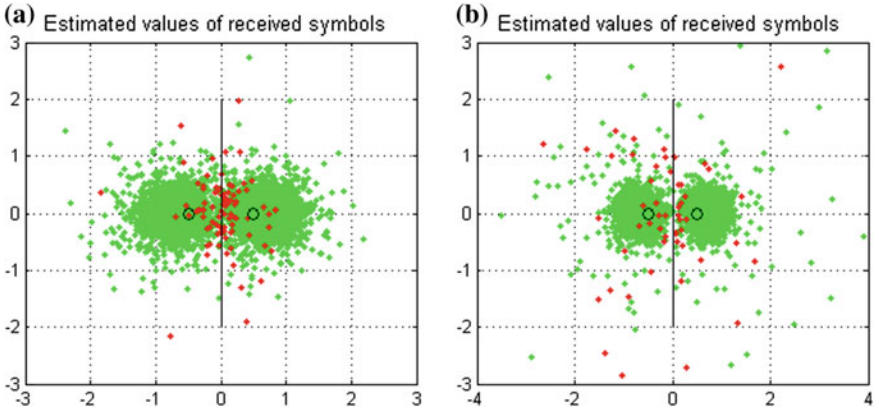


Fig. 9 AWCN channel affected by fading, $N_R = 4$, SNR = 10 dB. **a** $A = T = 0.1$; **b** $A = T = 0.01$

at reception. This situation is similar to that at point d), the number of errors being smaller for strong impulsive noise, but the impulses' high amplitude places the symbols very far from the constellation's points. Using 4 receiving antennas, at $A = T = 0.01$, we can see that the values of the estimated symbols are smaller than the ones at point d).

The results from above can be summarized in Table 1, for the AWGN channel, and Table 2, for the AWCN channel, respectively. In both cases, we considered the H matrix to be known at the receiving end, the fading to be of Rayleigh type, BPSK modulation, two emitting antennas and a data set of size N = 10000.

Table 1 Number of errors (N_{err}) for AWGN channel

SNR (dB)	N_{err}	
	Number of receiving antennas	
	2	4
5	162	103
7	53	30

Table 2 Number of errors (N_{err}) for AWCN channel

(A; T)	N_{err}	
	Number of receiving antennas	
	2	4
(0.1; 0.1)	102	87
(0.01; 0.01)	65	52

4.2 Bit Error Rate Analysis

To analyze the influence of impulsive noise on Alamouti code performances, a MIMO channel is considered with $N_T = 2$ and $N_R = 2$. The values for the model parameters are: $\alpha \in [1; 1.5; 2]$, $\beta = 0$, $\gamma = 1$, $\mu = 0$ and $\sigma = 1$. The results were compared with the cases of an AWGN channel and a Middleton Class-A (AWCN) channel with parameter $A = T = 0.01$ (highly impulsive noise) [24].

The simulation results are illustrated in Fig. 10. The impulsive noise degrades the system performances compared with AWGN. For example, for BER = 10^{-3}, the system brings a coding gain of about 3 dB when the channel is affected by Gaussian noise against SαS impulsive noise.

The poorest results are obtained in the presence of SαS noise and as α decreases, BER increases. For $\alpha = 2$, starting from SNR = 8 dB, the BER increases in case of AWCN channel. In the presence of Middleton Class-A noise, the system leads better performances for SNR values up to 3.8 dB. Beyond this point, the BER has the smallest values for AWGN channel.

4.3 Image Transmission Using Alamouti STBC

In this section, Alamouti 2 × 2 code is used for image transmission through MIMO channel. The original image is shown in Fig. 11a. It has 512 × 512 pixels with 8 bit grayscale. Figure 11b and c present the received image on AWGN and SαS ($\alpha = 1$) channels at SNR = 5 dB. The last one has the largest number of damaged pixels.

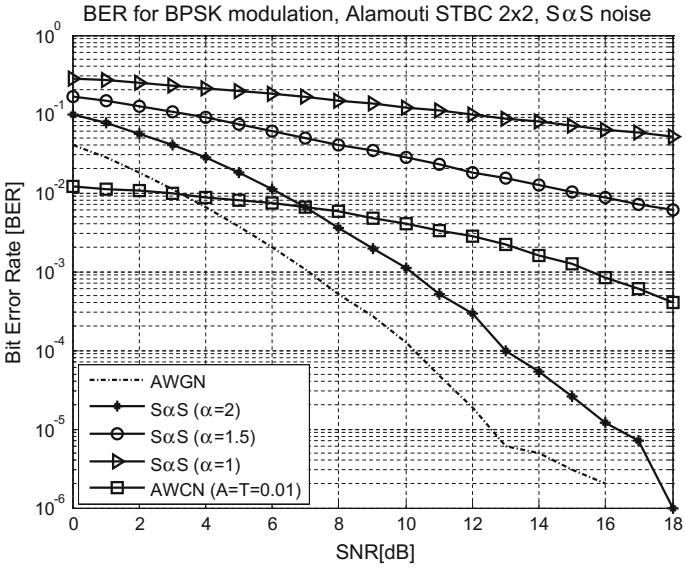


Fig. 10 BER curves for Alamouti code 2 × 2, under different type of noise

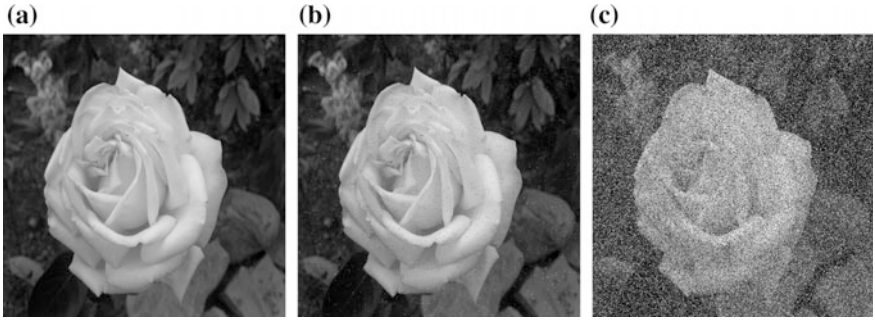


Fig. 11 **a** Original image; **b** AWGN; **c** SαS ($\alpha = 1$)

The transmitted data was BPSK modulated. The simulations were done for two SNR values: 5 and 10 dB, respectively. The image quality is assessed in terms of mean squared error (MSE) defined as:

$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [I(i,j) - \widehat{I}(i,j)]^2 \tag{11}$$

where M, N—represent the image's horizontal and vertical number of pixels, respectively; I is the original image and $\widehat{I}$ is the received image.

Table 3 Image quality metric

	SNR (dB)	Type of noise				
		AWGN	AWCN (A = T = 0.01)	SαS		
				α = 2	α = 1.5	α = 1
MSE	5	83.66	235.52	387.04	1634	4244
	10	2.37	145.37	21.78	605.43	2647

The values of MSE are collected in Table 3. The MSE values calculated for AWGN case are significantly lower than in the case of impulsive noise. It can be observed that MSE increased with the impulsive component (α). If SNR increases, the noise affects the image less, MSE having significantly lower values.

5 Conclusions

Analyzing the distribution of the estimated values of received symbols, for 2 and 4 receiving antennas, in the case of a channel affected by Gaussian and impulsive noise described by the Middleton Class-A model, it was highlighted that the spatial diversity at the receiver brings performance enhancements to the Alamouti code, on both AWGN and AWCN channels. But in the case of strongly impulsive noise ($A = T = 0.01$), the differences between the number of errors obtained for 2 and 4 receiving antennas, respectively, is not very high, while for AWGN this is cut nearly in half.

The behavior of a MIMO system, with an Alamouti 2×2 code, was investigated on a channel affected by impulsive noise and Rayleigh fading. The non-Gaussian noise was modeled with SαS type and data was BPSK modulated. The simulations were performed for different values of the exponent parameter α. The BER curves increase considerably against the AWGN channel, as α gets lower. The results were also compared with the case of AWCN channel ($A = T = 0.01$—highly impulsive). For low SNR, Middleton Class-A noise yields the best performances. The worst results, for all SNR values, are obtained for SαS noise, with $\alpha < 2$.

The image quality is strongly affected by the impulsive noise, compared to AWGN, when transmitting it on a MIMO channel. In this case, we have considered Alamouti 2×2 code, Rayleigh-type fading and SαS impulsive noise. The simulations have shown that as the noise gets more impulsive (the exponent parameter is lower), the images get more distorted (the case of $\alpha = 1$).

References

1. Tarokh, V., Seshadri, N., Calderbank, A.R.: Space–time codes for high data rate wireless communication: Performance analysis and code construction. IEEE Trans. Inf. Theory **44**(2), 744–765 (1998)
2. Vucetic, B., Yuan, J.: Space-Time Coding. Wiley, England (2003)
3. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. IEEE J. Sel. Areas Comm. **16**(8), 1451–1458 (1998)
4. Middleton, D.: Statistical-physical models of electromagnetic interference. IEEE Trans. Electromagn. Compat **EMC-19**(3), 106–127 (1977)
5. Al-Dharrab, S., Uysal, M.: Cooperative diversity in the presence of impulsive noise. IEEE Trans. Wireless Commun. **8**(9), 4730–4739 (2009)
6. Win, M., Pinto, P., Shepp, L.: A mathematical theory of network interference and its applications. Proc. IEEE **97**(2), 205–230 (2009)
7. Madi, G., Sacuto, F., Vrigenau, B., Agba, B.L., Vauzelle, R., Gagnon, F.: Impact of impulsive noise from partial discharges on wireless systems performance: applications to MIMO precoders. EURASIP J. Wirel. Commun. Network. **2011**, 186 (2011)
8. Hall, H.M.: A new model for impulsive phenomena: application to atmospheric-noise communication channels. Stanford University, Technical Report, August 1966
9. Shao, M., Nikias, C.L.: On symmetric stable models for impulsive noise. University Southern California, Los Angeles, Technical Report USC-SIPI-231 (1993)
10. Chopra, A., Evans, B.L.: Join statistics of radio frequency interference in multi antenna receivers. IEEE Trans. Signal Process. **60**(7), 3588–3603 (2012)
11. Bhatti, S.A., Shan, Q., Glover, I.A., Atkinson, R., Portugues, I.E., Moore, P.J., Rutherford, R.: Impulsive noise modeling and prediction of its impact on the performance of WLAN receiver. In: 17th European Signal Processing Conference (EUSIPCO 2009), pp. 1680–1684 (2009)
12. Yoo, J., Choe, S.: Performance of space–time–frequency coding over indoor power line channels. IEEE Trans. Commun. **62**(9), 3326–3335 (2014)
13. Gong, Y., Wang, X., He, R., Pang, F.: Performance of space-time block coding under impulsive noise environment. In: Proceedings IEEE of 2nd International Conference on Advanced Computer Control, vol. 4, pp. 445–448 (2010)
14. Rajan, Tepedelenlioglu, C.: Diversity combining over Rayleigh fading channels with symmetric alpha stable noise. IEEE Trans. Wirel. Commun. **9**(9), 2968–2976 (2010)
15. Niranjayan, S., Beaulieu, N.: The BER optimal linear rake receiver for signal detection in symmetric alpha-stable noise. IEEE Trans. Commun. **57**(12), 3585–4588 (2009)
16. Lee, J., Tepedelenlioglu, C.: Space-time coding over fading channels with stable noise. IEEE Trans. Veh. Technol. **60**(7), 396–400 (2011)
17. Zha, D., Qiu, T.: Direction finding in non-Gaussian impulsive noise environments. Digit. Signal Proc. **17**, 451–465 (2007)
18. Nikias, C.L., Shao, M.: Signal Processing with Alpha-Stable Distributions and Applications. Wiley, New York (1995)
19. Gu, W., Peters, G., Clavier, L., Septier, F., Nevat, I.: Receiver study for cooperative communications in convolved additive α-stable interference plus Gaussian thermal noise. In: Ninth International Symposium on Wireless Communication Systems, pp. 451–455 (2012)
20. Middleton, D.: Non-Gaussian noise models in signal processing for telecommunications: new methods and results for class A and class B noise models. IEEE Trans. Inf. Theory **45**(4), 1129–1149 (1999)
21. Saaifan, K.A., Henkel, W.: A spatial diversity reception of binary signal transmission over rayleigh fading channels with correlated impulse noise. In: 19th International Conference on Telecommunications (ICT 2012), pp. 1–5, Jounieh, Lebanon (2012)
22. Umehara, D., Yamaguchi, H., Morihiro, Y.: Turbo decoding over impulse noise channel. In: Proceedings of IEEE ISPLC (2004)

23. Spaulding, A., Middleton, D.: Optimum reception in an impulsive interference environment-part I: coherent detection. IEEE Trans. Commun. **25**(9), 910–923 (1977)
24. Andrei, M., Trifina, L., Tarniceriu, D.: Influence of impulse noise on Alamouti code performances. Proc. Wirel. Mobile Appl. ECUMICT **2014**, 11–21 (2014)
25. Gulati, K., Nassar, M., Chopra, A., Ben Okafor, N., DeYoung, M., Aghasadeghi, N., Sujeeth, A., Evans, B.L.: Interference modeling and mitigation toolbox 1.6, for matlab. ESP Laboratory, ECE Department, University of Texas at Austin (2011)

Evaluation of Gradient Norms on a Consistent Quadtree Grid in 2D

With Application to Curvature Filters

Zuzana Krivá and Angela Handlovičová

Abstract This paper solves the problem of how to evaluate gradients and their norms on a quadtree grid, which is deformed in such a way that the connections of centers of its adjacent elements are perpendicular to their common boundaries. On the grid, we solve the parabolic PDEs representing the curvature-driven filters based on the mean curvature flow and geodetic mean curvature flow equations in a level set formulation. The numerical solution of these equations is based on the finite volume method, where the finite volumes correspond to elements of the deformed quadtree. The described method utilizes representative points not only for the finite volumes but also for the edges forming the boundaries of grid elements. Using these points we evaluate the gradients locally. Solution values in the edge representative points are updated by balancing the fluxes in such a way that we always need only neighbors of a finite volume sharing a common edge with it, not only a vertex. This fact is important for the efficiency of the algorithm. The edge representative points have been chosen in such a way that they lie on a connection of volume representative points enabling to derive a special formula for the gradient norm. We discuss the ways of approximating the norms of the gradients, and on selected examples we show their properties.

Keywords Quadtree adaptive grid · Consistent adaptive grid · Finite volume method · Level set method · Mean curvature flow · Geodetic mean curvature flow

Z. Krivá (✉) · A. Handlovičová
Faculty of Civil Engineering, Department of Mathematics and Descriptive Geometry, Slovak Technical University Bratislava, Bratislava, Slovakia
e-mail: kriva@math.sk

A. Handlovičová
e-mail: angela@math.sk

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*, Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_9

1 Introduction

Numerical solutions of nonlinear PDEs used in image processing often need to approximate gradients and their norms on their computational grids, e.g., to perform edge detection or to evaluate the curvature. In this paper, we evaluate these entities for the finite volume grid based on a quadtree which is adjusted in such a way that the connection of representative points of two adjacent finite volumes is perpendicular to their common boundary (i.e., it fulfills the classical orthogonality property). In image processing, the quadtree is used to decrease a number of elements in the computational grid, typically in regions with homogeneous intensity where larger elements can be used. This basic quadtree grid is deformed to fulfill the property mentioned above. In such an adjusted grid, the intersection of a line segment connecting representative points (set to centers of original squares) and corresponding finite volume boundary is not generally a middle point of their common edge—the middle points are usually used to evaluate the gradients. In [3], a novel method to evaluate gradients for such a situation has been presented. It leads to a solution of a short linear system to get a gradient for a grid element, however using the properties of the grid can simplify this evaluation significantly. This method is shortly described in Sect. 3.

In this paper, we explore the use of the gradients obtained such as in [3] for approximation of norms, which are used to solve the parabolic PDEs representing the curvature filters working on the deformed quadtree grid like in Fig. 1. The edge representative points are chosen to be the intersections of connection of representative points and edges. We want them to be the points **with help of which we evaluate the gradients**. In [3], where the method has been introduced, we dealt with a linear heat equation and the regularized Perona-Malik model, where the edge detection is performed with a help of a Gaussian gradient. In the following, we first

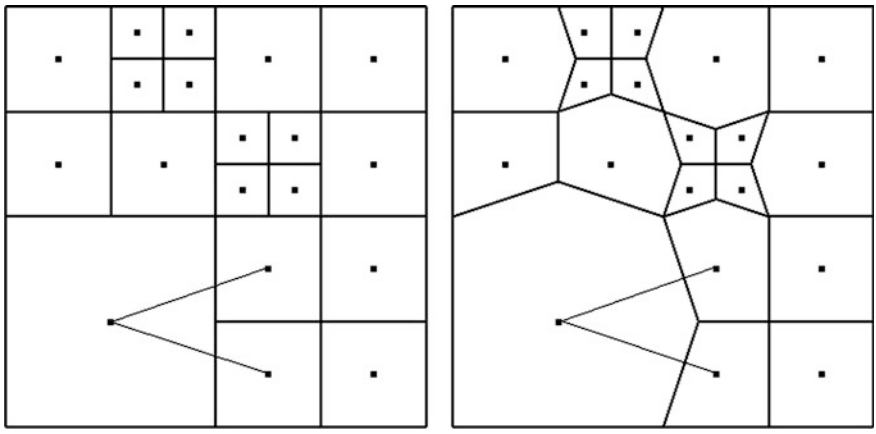


Fig. 1 *Left* the basic quadtree grid. *Right* deformation to a consistent grid

discuss the possibilities of applying such an evaluation of gradients to the model (1) for the mean curvature flow (MCF) in a level set formulation, called the curvature filter in image processing. In the last section, we explore its modification (24) known as the geodetic mean curvature flow (GMCF). In both cases, we evaluate the norm of the gradients in a different way like in [3].

The MCF model is described by the following evolutionary PDE

$$\frac{\partial u}{\partial t} = |\nabla u| \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \cdot \tag{1}$$

The effect is similar to a median filter—unsmooth boundaries are smoothed (see Fig. 6 for an example)—and the noise represented by small objects of high curvature is removed (Figs. 8 and 11). However, the boundary of the object itself can shrink: the geodetic mean curvature flow mode stops the movement of important edges, detected again by the Gaussian gradient.

2 Computational Grid

If the character of the data is such that we have large areas of the homogenous intensity we can use an irregular grid. A good example of such a grid is a quadtree composed of square elements (see Fig. 1 on the left). In the finite volume method, the piecewise constant solution on grid elements—finite volumes—is considered. These values are assigned to representative points in centers of the squares. Because the normal derivative is needed, the orthogonality property, i.e., the orthogonality of a connection of representative points of two adjacent finite volumes and their common boundary, is a desirable property. The quadtree does not fulfill this property. To satisfy it, we deform the grid as in Fig. 2. In this paper, the quadtree grid obtained by this deformation is called the **consistent**.

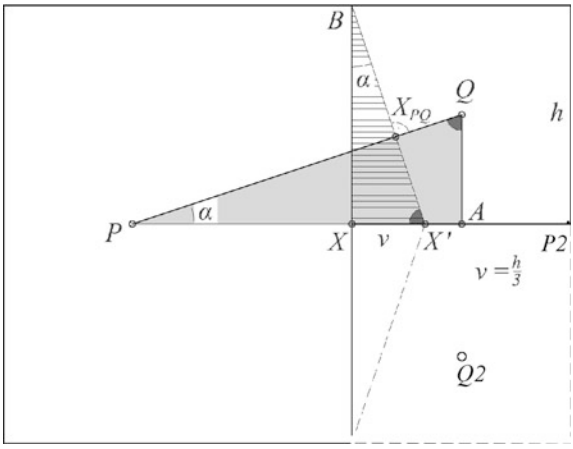


Fig. 2 Deformation of the basic quadtree grid to a consistent one

2.1 Basic Quadtree Grid

The details of encoding, building, and traversing the quadtree can be found in [4]. We mention here only the main features.

Definition. A quadtree is a recursive partition of some data of the size $2^N \times 2^N$ into four quadrants data sets of the size $2^{N-1} \times 2^{N-1}$—through its center. The subdivision is ruled by some criterion.

Encoding. The quadtree corresponding to the image of the dimensions $M \times M$ is encoded in the binary field $(M + 1) \times (M + 1)$. Each of its elements of the size $2^i \times 2^i, i = 1, \ldots, log_2 M$, has a corresponding element $(2^i + 1) \times (2^i + 1)$ in the binary field, for which we speak about central and corner positions and also about the middle positions of their sides. These positions are used to create the quadtree, moreover the quadtree with a prescribed ratio of sides. In this binary field 1 is used for division and 0 for homogeneity indication.

Traversing. To traverse the quadtree (i.e., to go through all its elements), we inspect the central position of elements in the binary field obtained by a recursive procedure, where *0* means *stop* and *1 continue in subdivision*. The central positions are also used to test the configurations of neighbors. To find a neighbor sharing a common edge, in a graded quadtree with a possible ratio of elements 1:1, 1:2 or 2:1 we need at most two tests. To find neighbors sharing a vertex only, more tests are needed—**the proposed methods avoid this testing because it utilizes only neighbors having a common edge of nonzero measure**.

Building. Let us have data defined on a regular square grid (e.g., an image with its pixel structure). To construct the quadtree, we start with merging similar valued elements forming blocks $2^i \times 2^i$, for increasing i, i.e., from leaves to the root. The original values are either preserved, if merging fails, or set to mean values of the processed elements, if merging is successful. During this process, the information about successful or unsuccessful merging is stored in the binary field mentioned above, in such a way that it allows for creating the quadtree with a prescribed ratio of elements. We require the ratio of sides of two adjacent squares to be 1:1, 1:2 or 2—it simplifies building the linear system matrix, where access to neighbors is needed, and also, the consistent grid requires this condition. In [4], the criterion steering the merging-the **coarsening criterion** is the following: cells are merged if a difference in their intensities is below a prescribed threshold ε. This criterion can be modified, i.e., to keep small elements in the vicinity of edges and will be discussed later.

2.2 Deformation into the Consistent Grid

The quadtree grid (Fig. 1 on the left) is *inconsistent* in the sense that it does not fulfill the classical orthogonality property. The grid keeping this property is an admissible mesh in the sense of the basic finite volume theory [1]—this is one of the reasons why we deform. The basic quadtree grid can be adjusted to a consistent

one procedurally (i.e., the setting of the binary field is not changed): we must adjust the shape, if two adjacent finite volumes p and q are different sized in the original quadtree (this situation will be called non conformal). If we denote the length of a common edge in the original quadtree by h and we shift the *hanging node* by $v = h/3$ (e.g., in Fig. 2 we shift X to X′), then the connection of P and Q is perpendicular to the shifted common boundary. This fact (and also the fact that BX′/PQ = 2/3) follows from the similarity of triangles ΔAQP and Δ XX′B with the ratio of their adjacent sides being 1:3. The area of p is also evaluated procedurally—it depends on a configuration of its neighbors.

2.3 Properties of the Consistent Grid

In this paragraph, we summarize the geometrical properties necessary to derive the numerical scheme (details in [3]):

$$\frac{BX'}{PQ} = \frac{2}{3}, \tag{2}$$

$$\frac{QX_{PQ}}{PX_{PQ}} = \frac{1}{4}, \tag{3}$$

$$\frac{X'P2}{QQ2} = \frac{2}{3}, \tag{4}$$

$$v = \frac{XP2}{3} = \frac{h}{3} \tag{5}$$

3 Evaluation of the Gradient on the Regular Grid

Notations. Every finite volume p has a representative point denoted by X_p. It lies in its center or in the center of the original square for a deformed element of the consistent grid. A measure of p is denoted by $|p|$. The common part of boundaries of p and q—an edge σ_{pq} must have a nonzero measure in R, which is denoted by $|\sigma_{pq}|$. Let $d_{pq} = |X_q - X_p|$ be the distance of representative points. Let us denote by $\boldsymbol{X_\sigma}$ a point on σ_{pq}, which represents the **intersection** of the line segment X_pX_q and σ_{pq}. In our consistent grid, X_pX_q is perpendicular to σ, but the intersection X_σ is not the midpoint of σ in a general case. Let us denote by X^*_σ the **midpoint** of the edge σ. By ε_p we denote the set of all edges σ of p. When we speak about a unit outer normal vector to $\sigma \in \varepsilon_p$, we denote it by $\mathbf{n}_{pq}$.

Motivation. As we have mentioned below, we want to avoid using values in the corners of the finite volumes because they can lead to too many tests. Our method has been inspired by the method derived in [2] for a regular square grid with the following set of steps:

1. On edges σ of a finite volume p, representative points X_σ are defined in the same way like we have described above: in the case of the regular square grid these intersections are midpoints at the same time: it holds $X_\sigma^* = X_\sigma$.
2. Using these points, the norm of the **gradient on *p*** is evaluated *locally*, using (6).
3. A discrete equation for (1) on the finite volume p is derived locally.
4. Solution values in X_σ are updated using a conservation principle. Only neighbors sharing an edge σ are needed.

In this method, the following formula to evaluate the norm of the gradient on p has been used:

$$\left|\nabla u_p\right| \approx \sqrt{\frac{|\sigma|}{h^2} \sum_{\sigma \in \varepsilon_p} (u_\sigma - u_p)^2}, \tag{6}$$

where h is the length of a side of squares forming the uniform grid. In the method, the fact that the connection of edge and volume representative points X_p X_σ is perpendicular to σ is used to derive (6): the midpoint is needed to perform precise integration in deriving the gradient approximation like in (10). In the consistent grid, the representative point is X_σ which can be different from the midpoint X_σ^* in general. In our paper, the formulas for the gradient and the norm use only u_σ and u_p. Due to geometrical properties of the consistent grid, it can be eventually easily eliminated from computations.

4 Evaluation of the Gradient on the Consistent Grid

Let us consider the linear approximation of the solution u over the finite volume p. At every X of the finite volume p, any linear function can be written as

$$u(X) = u(X_p) + \nabla u \cdot (X - X_p) = u_p + \nabla u \cdot (X - X_p),$$

If $X = X_\sigma$, it holds

$$u_\sigma - u_p = \nabla u \cdot (X_\sigma - X_p) \tag{7}$$

with u_σ and u_p representing the solution values in X_σ and X_p. Because the gradient of any linear function is a constant vector in R^2, the gradient over a control volume p can be expressed as follows:

$$\nabla u = \frac{1}{|p|}\int_{p} \nabla u \, dX = \frac{1}{|p|}\int_{\partial p} u\, \mathbf{n}_p \, dS$$
$$= \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} \int_{\sigma} \left(u_p + \nabla u \cdot \left(X - X_p\right)\right) \mathbf{n}_{p\sigma} \, dS$$
$$= \frac{1}{|p|} u_p \sum_{\sigma \in \varepsilon_p} |\sigma| \mathbf{n}_{p\sigma} + \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| \, \nabla u \cdot \left(X_\sigma^* - X_p\right) \mathbf{n}_{p\sigma}.$$

The first term can be shown to be equal to zero vector and the second expression, due to using X_σ^*, represents the precise integration of a linear function over the edge σ. We can write

$$\nabla u = \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| \nabla u \cdot \left(X_\sigma^* - X_p\right) \mathbf{n}_{p\sigma}. \tag{8}$$

On edges with $X_\sigma^* \neq X_\sigma$ it holds

$$X_\sigma^* - X_p = (X_\sigma - X_p) + (X_\sigma^* - X_p) \quad \text{and} \tag{9}$$

$$\nabla u = \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| \nabla_u \cdot (X_\sigma - X_p) \mathbf{n}_{p\sigma} + \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| \nabla_u \cdot (X_\sigma^* - X_\sigma) \mathbf{n}_{p\sigma} \tag{10}$$

The first vector of (10) will be denoted by $(\nabla_u)^A$. Using (7) it can be expressed as

$$(\nabla u)^A = \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| (u_\sigma - u_p) \mathbf{n}_{p\sigma} \tag{11}$$

The second term of (10) is a *correction* **C** of $(\nabla u)^A = ((u_x)^A, (u_y)^A)$. The right-hand side of (10) depends on the unknown gradient u. Further, let us denote by

$$\mathbf{c}_\sigma = (X_\sigma^* - X_\sigma) = \left((c_\sigma)_1, (c_\sigma)_2\right), \tag{12}$$
$$\nabla u = (u_x, u_y)$$

and

$$\mathbf{n}_{p\sigma} = \left(\left(n_{p\sigma}\right)_1, \left(n_{p\sigma}\right)_2\right).$$

Now (11) can be rewritten by coordinates and using (12) we get

$$(u_x, u_y) = \left((u_x)^A, (u_y)^A\right) + \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma|\left((c_\sigma)_1 u_x + (c_\sigma)_2 u_y\right)\mathbf{n}_{p\sigma}. \tag{13}$$

The Eq. (13) represents a linear system of two equations with two unknowns u_x and u_y. Let us denote:

$$\mathbf{N}_{p\sigma} = \frac{|\sigma|\mathbf{n}_{p\sigma}}{h}, \quad \mathbf{C}_\sigma = \frac{\mathrm{c}_\sigma}{h}.$$

Then (13) can be adjusted to the following form:

$$\begin{aligned} u_x(1 - \frac{h^2}{|p|}\underbrace{\sum_{\sigma \in \varepsilon_p}(C_\sigma)_1(N_{p\sigma})_1)}_{b_{11}} + u_y(1 - \frac{h^2}{|p|}\underbrace{\sum_{\sigma \in \varepsilon_p}(C_\sigma)_2(N_{p\sigma})_1)}_{b_{12}} &= (u_x)^A, \\ u_x(-\frac{h^2}{|p|}\underbrace{\sum_{\sigma \in \varepsilon_p}(C_\sigma)_1(N_{p\sigma})_2)}_{b_{21}} + u_y(1 - \frac{h^2}{|p|}\underbrace{\sum_{\sigma \in \varepsilon_p}(C_\sigma)_2(N_{p\sigma})_2)}_{b_{22}} &= (u_y)^A. \end{aligned} \tag{14}$$

This way of evaluating the gradient was explored in [3]. The coefficient matrix B depends only on the shape of the element, not on its size. As we can see in this way, we can approximate the gradient of any numerical solution u_h for which the values u_p and u_σ are known.

Example 1 Let us take the consistent quadtree grid built over a uniform grid with *32* × *32* elements (Fig. 3). We inspect the function $u(x, y) = 0.5 * (x^2 + y^2)$ defined on the interval $[-1.25; 1.25] \times [-1.25; 1.25]$. First, we consider the gradient evaluated analytically, in the representative points of the finite volumes. The result is displayed in Fig. 3. Then we evaluate the gradients using (14). In this case, the gradient of $u(x, y)$ obtained by (14) is equal to the analytical gradient (x, y) evaluated in the representative point. In Fig. 4 we can see the gradients obtained by (14) decomposed into $(\nabla_u)^A$ and **C**. In Fig. 5 the elements are colored by the magnitude of **C**. We can see that also for some non-square elements, the correction is equal to zero or it is very small. This is caused by the fact that in sums of (14), for elements having smaller neighbors, some terms cancel each other (two edges sharing a common vertex) and the matrices are either identity matrices or very close to them (the diagonal matrix with 0.985 and 1.015 on the diagonal). The significant corrections appear in cases when elements have a greater neighbor (neighbors). For an element having two greater neighbors the main diagonal of B is (1; 1), but the minor one is (±0.3; ±0.3). Also in Figs. 4 and 5, we can see that for equally shaped elements, the correction depends on the size of ∇u. More about the system matrices B and their practical evaluations can be found in [3].

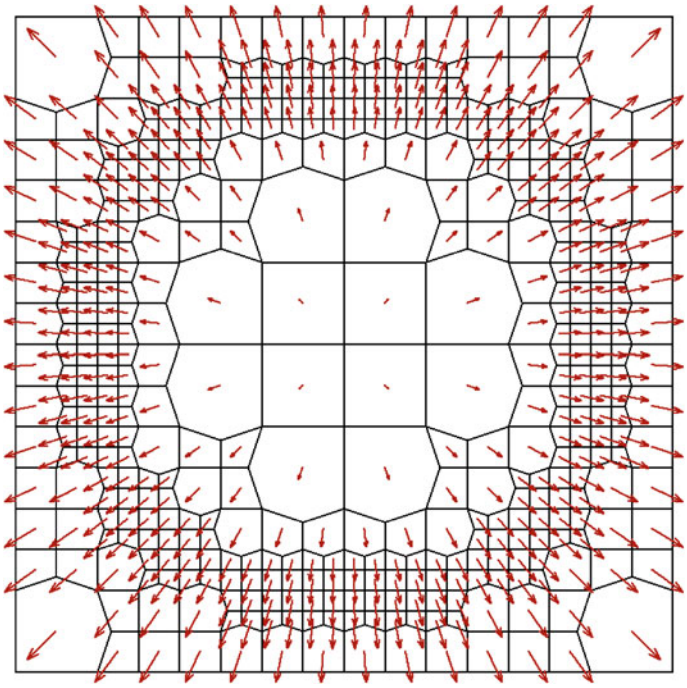


Fig. 3 Example 1. The consistent grid built over the data given on a regular grid $2^5 \times 2^5$ and the analytical gradient for the paraboloid

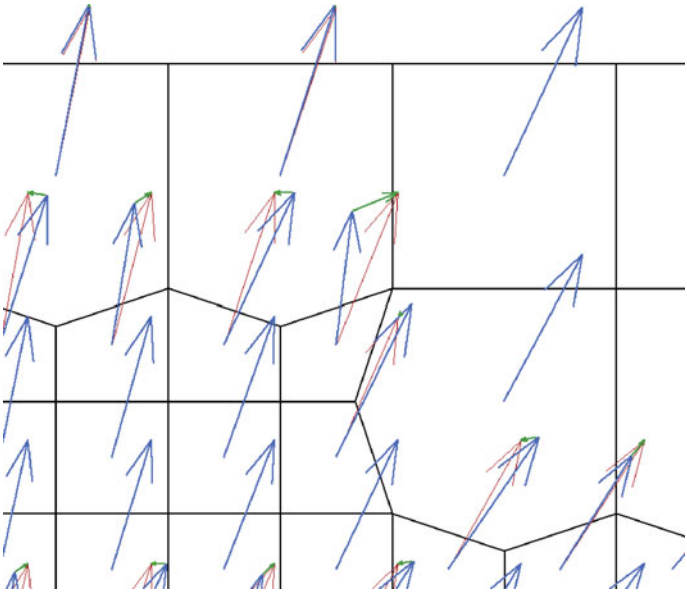


Fig. 4 Example 1. The decomposition of ∇u into the gradient $(\nabla u)^A$ (obtained using the edge representative points) and the correction vector C. The ∇u equal to analytical gradients (*thin lines* on *red*), the gradient $(\nabla u)^A$ in *blue* and corrections C in *green*

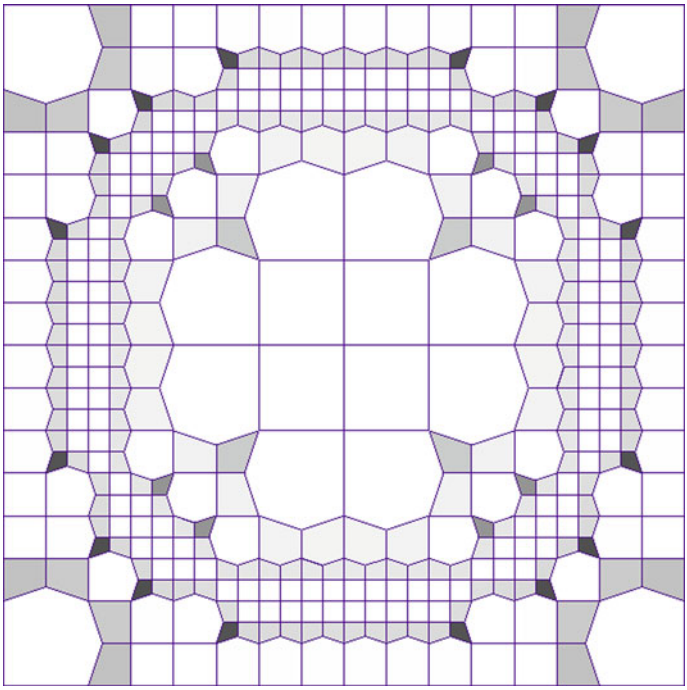


Fig. 5 Example 1. The elements of the consistent grid are colored by the magnitudes of their correction vectors **C**

5 Solving the MCF Equation

The mean curvature flow (MCF) in a level set formulation is the following initial value boundary problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= |\nabla u| \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \quad \text{in } \Omega \times [0, T], \\ \frac{\partial u(x,t)}{\partial n} &= 0 \quad \text{in} \quad \partial\Omega \times [0, T], \\ u(x, 0) &= u^0(x) \quad \text{for} \quad \text{x} \in \Omega. \end{aligned} \tag{15}$$

The computational domain Ω corresponds to the image, the abstract parameter T to the time for which the task is solved and $u^0(x)$ is the initial image.

The effect of this curvature-driven motion is as follows: imagine that we have a double-valued image representing an object on a background (the filter is contrast invariant). Then boundaries of the object are moved in the normal direction with a speed proportional to the curvature. The circles always shrinks: small very fast and large very slow. The curvature of a circle is *1/radius:* the time to vanish is given by a formula $T = R^2/2$, where R is the initial radius of the circle. A general image can

Fig. 6 *Left* the *circle* object with a noise with high curvature and an unsmooth boundary. *Middle* the object smoothed by (15). *Right* isolines representing the boundaries of the original (*red*) and of the smoothed objects. The boundary slightly shrinks. We used method [2] working on a regular grid

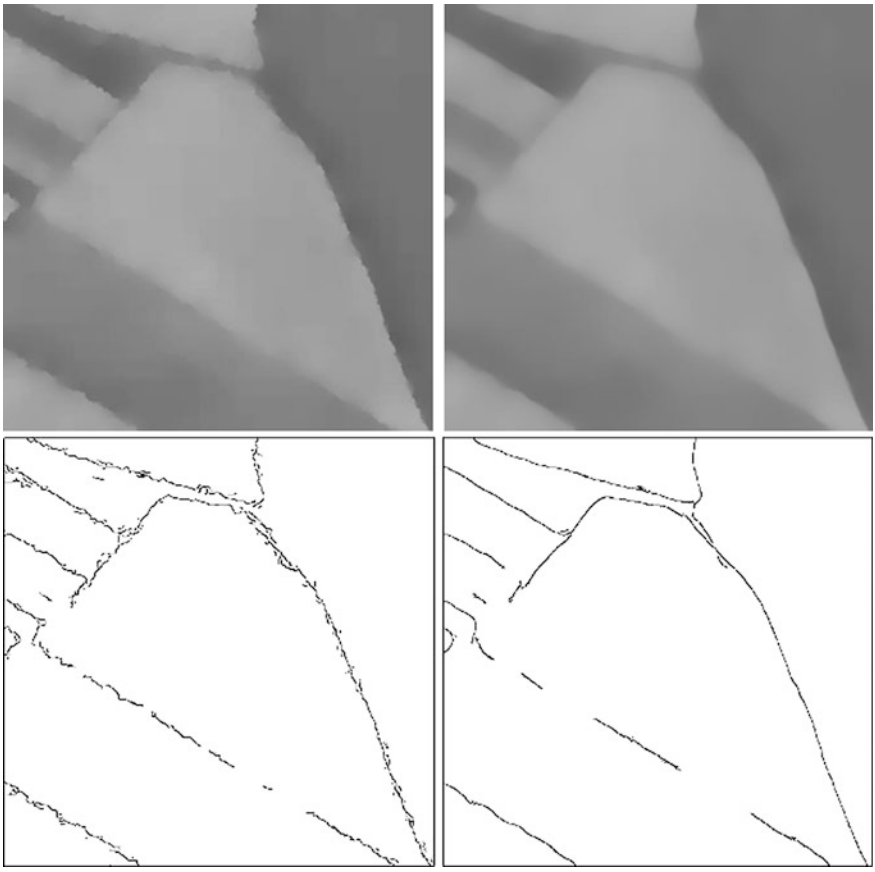


Fig. 7 Example 2. *Top* the effect of smoothing on an image with unsmooth boundaries. The boundaries are smoothed by (20). *Bottom* edges obtained by Canny edge detector to see the differences between boundaries better

be understood as a set of isolines of level sets forming a topological map of the image, all moving simultaneously by (15). The unsmooth boundaries are smoothed (see Figs. 6 and 7) as well as the noise of high curvature (see Fig. 10). The advantage of the level set method [6] is that we work with a curve on a grid and we need not solve the topological problems.

5.1 Deriving the Numerical Scheme

Now we derive the finite volume scheme numerical scheme with evaluation of gradients given by (14). The solution, constant on the finite volume p, is represented by u_p. The zero gradients in the denominator are treated with a help of the Evans–Spruck regularization

$$|\nabla u|_\varepsilon = \sqrt{u_x^2 + u_y^2 + \varepsilon^2}.$$

First, we regularize the Eq. (15) (because of possible zero gradient in a denominator) and then we divide it by $|u|_\varepsilon$:

$$\frac{\partial_t u}{|\nabla u|_\varepsilon} - \nabla \cdot \left(\frac{\nabla u}{|\nabla u|_\varepsilon} \right) = 0. \tag{16}$$

Then, as it is usual in the finite volume method, we integrate over the finite volume p and apply the divergence theorem. We get

$$\begin{aligned}
&\int_p \frac{\partial_t u}{|\nabla u|_\varepsilon} dx - \int_p \nabla \cdot \left(\frac{\nabla u}{|\nabla u|_\varepsilon} \right) dx = 0, \\
&\int_p \frac{\partial_t u}{|\nabla u|_\varepsilon} dx - \int_{\partial p} \left(\frac{\nabla u}{|\nabla u|_\varepsilon} \right) \cdot \mathbf{n}_p ds = 0. \\
&\int_p \frac{\partial_t u}{|\nabla u|_\varepsilon} dx - \sum_{\sigma \in \varepsilon_p} \int_\sigma \frac{\nabla u}{|\nabla u|_\varepsilon} \cdot \mathbf{n}_{p\sigma} ds = 0.
\end{aligned}$$

The discretization in time is done with the backward Euler time difference. For this purpose, we subdivide the whole time interval T into subintervals of the size τ which define time steps denoted by $t^i, t^n = t^{n-1} + \tau, t^0 = 0$ and $t^N = T$. We denote by u_p^n (u_σ^n) the value of solution in X_p (X_σ) at time t^n. The gradient is constant on the finite volume p, so we can write

$$\int_p \frac{\partial_t u}{|\nabla u|_\varepsilon} \approx \frac{(u_p^n - u_p^{n-1})|p|}{\tau \left| \nabla u_p \right|_\varepsilon^{n-1}},$$

where $|u_p|$ denote the evaluated L2 norm of the gradient obtained by (14) at time t^{n-1}. Let us denote by $d_{p\sigma}$ the distance of X_p from X_σ. Because the line segment X_pX_σ is perpendicular to σ, for all $\sigma \in \varepsilon_p$, we can approximate the derivative in the direction $\mathbf{n}_{p\sigma}$ by

$$\nabla u^n \cdot \mathbf{n}_{p\sigma} \approx \frac{u_\sigma^n - u_p^n}{d_{p\sigma}}. \tag{17}$$

To derive the semi-implicit scheme, we take linear terms from the current time step and nonlinear terms from a previous one. The time derivative is approximated by a backward Euler difference. Let us denote by $f_{p\sigma}$ the semi-implicit flux through boundary σ of p. We approximate

$$f_{p\sigma}^n = \int_\sigma \frac{\nabla u^n}{\left|\nabla u_p^{n-1}\right|_\varepsilon} \cdot \mathbf{n}_{p\sigma}\, ds \approx F_{p\sigma}^n = \frac{|\sigma|}{d_{p\sigma}} \frac{1}{\left|\nabla u_p^{n-1}\right|_\varepsilon} \left(u_\sigma^n - u_p^n\right), \tag{18}$$

where the evaluation of the norm of the gradient will be specified later. Because ∇u_p is constant on p, (16) gives us the linear system consisting of equations for each finite volume p. The equation is as follows:

$$\frac{\left(u_p^n - u_p^{n-1}\right)}{\left|\nabla u_p\right|_\varepsilon^{n-1}} |p| = \tau \sum_{\sigma \in \varepsilon_p} F_{p\sigma}^n. \tag{19}$$

In more details, for every p we have:

$$\frac{u_p^n - u_p^{n-1}}{\left|\nabla u_p\right|_\varepsilon^{n-1}} |p| - \tau \sum_{\sigma \in \varepsilon_p} \frac{|\sigma|}{d_{p\sigma}} \frac{1}{\left|\nabla u_p\right|_\varepsilon^{n-1}} \left(u_\sigma^n - u_p^n\right) = 0. \tag{20}$$

All possible values of $|\sigma|/d_{p\sigma}$ follow from Fig. 2 and geometrical properties of the grid (2)–(5). Let us remind that

$$\frac{d_{p\sigma}}{d_{pq}} \text{ is } 1{:}2,\ 1{:}5 \text{ or } 4{:}5.$$

Let T_{pq} denote the ratio $|\sigma|/d_{pq}$. It holds [3]:

$$T_{pq} = 1 \text{ for neighbors of the same size in the original quadtree grid}$$
$$T_{pq} = \frac{2}{3} \text{ otherwise and}$$
$$\frac{|\sigma|}{d_{p\sigma}} = T_{pq} \frac{d_{pq}}{d_{p\sigma}}.$$

5.2 Balancing the Fluxes

The value u_σ^n in the edge representative point X_σ can be obtained using the conservation principle $F_{p\sigma}^n = -F_{q\sigma}^n$. Let us denote the regularized gradient on p by f_p. We get

$$u_\sigma^n = \frac{d_{q\sigma} f_q^{n-1} u_p^n + d_{p\sigma} f_p^{n-1} u_q^n}{d_{q\sigma} f_q^{n-1} + d_{p\sigma} f_p^{n-1}}. \tag{21}$$

Usually we evaluate

$$u_\sigma^n - u_p^n = \frac{d_{q\sigma} f_p^{n-1} \left(u_q^n - u_p^n\right)}{d_{q\sigma} f_q^{n-1} + d_{p\sigma} f_p^{n-1}}. \tag{22}$$

The steps of the algorithm are as follows:

1. We build a balanced quadtree grid (in linear time). The values u_σ^0 are obtained by a linear interpolation.
2. During traversing the quadtree (procedurally treated as a consistent grid), we evaluate the linear system coefficients and store them into a linear list.
3. The linear system is solved using SOR method.
4. We rebuild the grid and evaluate new values of gradients with u_σ^n obtained by (21).
5. We go to Step 3.

Example 2 The efficiency of the adaptive algorithm depends on the decrease of elements in the grid. As an example of an application, we have chosen a cut of a filtered SAR image. SAR images are large-scale images of the Earth obtained by airborn or spaceborn **r**adars with a **s**ynthetic **a**perture. These images are degraded by a speckled noise with a strong granular pattern. We take an image filtered by the Perona-Malik algorithm working on an adaptive grid of this kind with evaluation of the gradients like in (14). We can see the uneven boundaries caused by the speckle (Fig. 7 on the left). In Fig. 7 on the right, we see the result of smoothing by the algorithm based on (20) described above. To see the effect of smoothing better, we add images displaying edges obtained by the Canny edge detector.

The algorithm soon starts to work on about 5% of the original number of elements. We performed 10 time steps with $\tau = 2$. The final adaptive grid is displayed in Fig. 8.

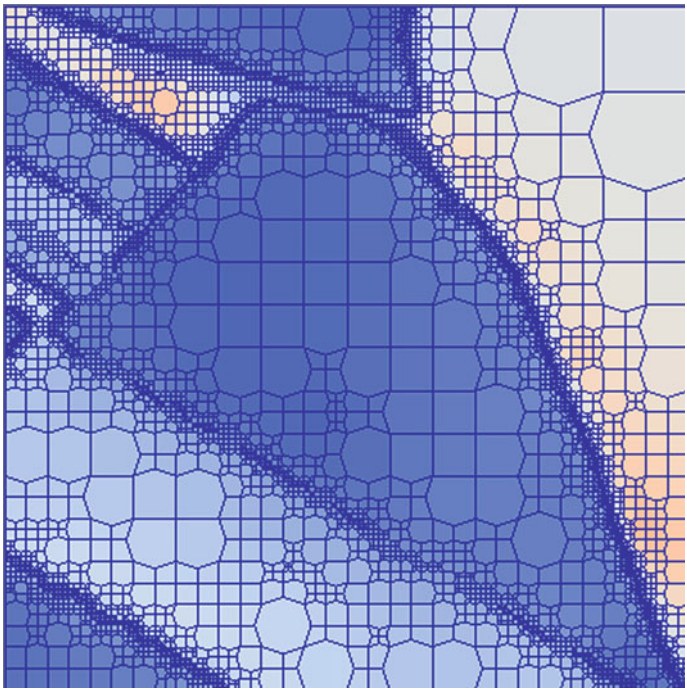


Fig. 8 Example 2. The adaptive grid for the final step of the adaptive MCF algorithm based on (20). The size of the original image for the algorithm (reduced) was 512 × 512. At the end, the number of grid elements was about 5% of the original

5.3 Removing Salt and Pepper Noise

Now let us take an image containing the salt and pepper noise (see Fig. 9 as an example). This noise corresponds to a configuration as in Fig. 9 on the left. In such a situation we obtain a zero gradient, the curvature is equal to zero and the noise is kept.

We are going to introduce a different approximation of the gradient corresponding to those in [2].

Let us consider a linear function again. Let us perform

$$|\nabla u|^2 = \nabla u \cdot \nabla u = \left((\nabla u)^A + \mathbf{C}\right) \cdot \nabla u$$
$$= \left(\frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| \nabla u \cdot (X_\sigma - X_p)\mathbf{n}_{p\sigma} + \frac{1}{|p|}\sum_{\sigma \in \varepsilon_p} |\sigma| \nabla u \cdot (X_\sigma^* - X_\sigma)\mathbf{n}_{p\sigma}\right).\nabla u.$$

Now, we can use the fact that the connection of the edge and the volume representative points is perpendicular to the boundary. We can write

$$X_\sigma - X_p = d_{p\sigma}\mathbf{n}_{p\sigma}$$

Let us substitute for $X_\sigma - X_p$ in the previous equation:

$$\begin{aligned}
&\left(\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}|\sigma|\nabla u\cdot(d_{p\sigma}\mathbf{n}_{p\sigma})\frac{d_{p\sigma}}{d_{p\sigma}}\mathbf{n}_{p\sigma}\right)\cdot\nabla u+\mathbf{C}\cdot\nabla u\\
&=\left(\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}|\sigma|\left(\nabla u\cdot(d_{p\sigma}\mathbf{n}_{p\sigma})\right)\left(\frac{d_{p\sigma}}{d_{p\sigma}}\mathbf{n}_{p\sigma}\cdot\nabla u\right)\right)+\mathbf{C}\cdot\nabla u\\
&=\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}|\sigma|\frac{1}{d_{p\sigma}}\left(\nabla u\cdot(d_{p\sigma}\mathbf{n}_{p\sigma})\right)^2+\mathbf{C}\cdot\nabla u\\
&=\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}\frac{|\sigma|}{d_{p\sigma}}\left(u_\sigma-u_p\right)^2+\mathbf{C}\cdot\nabla u=\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}T_{pq}\frac{d_{p\sigma}}{d_{pq}}\left(u_\sigma-u_p\right)^2+\mathbf{C}\cdot\nabla u.
\end{aligned}$$

In the case of a linear function, we see (Fig. 4) that on a sharp element the second part—the projection of ∇u on $\mathbf{C}$—is significant and cannot be omitted. In [2], where (16) was solved on a regular square grid, the norm corresponding to the first part has been used (see (6)) and the mathematical and numerical properties of this scheme have been studied. For examples with a known analytical solution, the experimental order of convergence for the semi-implicit scheme has been reported as 2.

In the case of the arbitrary numerical function u, the identity (7) does not hold and $u_\sigma - u_p \neq \nabla u\cdot(X_\sigma - X_p)$. Our approximation of the norm of u, we denote it by grad u, on every finite volume p will be defined as

$$\begin{aligned}
&|\nabla u|:=\sqrt{\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}\frac{|\sigma|}{d_{p\sigma}}\left(u_\sigma-u_p\right)^2+C.\nabla u_l},\\
&C=\nabla u_l-(\nabla u)^A,
\end{aligned}\tag{23}$$

where, ∇u_l is evaluated by (14). (Now the subscript l helps to distinguish between the approximation of a gradient by a linear function and arbitrary one.) Let us denote the regularized approximation of the norm by (23) as f_p^2, and the regularized approximation of the norm of the gradient approximated by (14) as f_p^1. In the case of f_p^1, we first use the left-hand side of (7) and then square it. In the case of f_p^2, we first use the right-hand side, square it, and then substitute by the right-hand side of (7). Thus in the case of the configuration like in Fig. 9, the norm of the gradient is different from zero and the salt and pepper noise is removed (see Fig. 12).

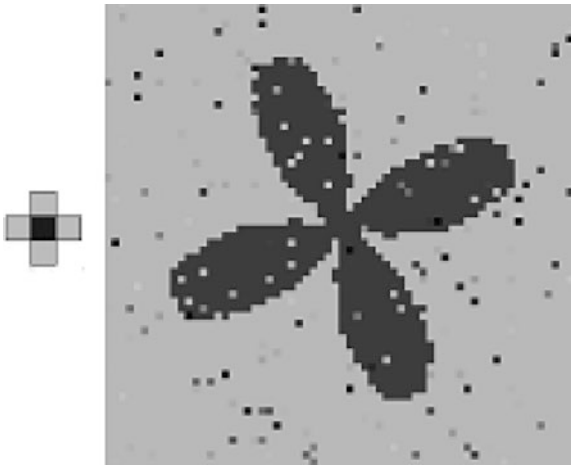


Fig. 9 The salt and pepper noise on a double-valued image. In the situation depicted on the left, the gradient is equal to the zero vector

5.4 Correctness of the Definition

We show that the definition (23) is correct and (23) cannot be negative.

Let us remind that

$$\nabla u_l^A = \frac{1}{|p|} \sum_{\sigma \in \varepsilon_p} |\sigma|(u_\sigma - u_p)\mathbf{n}_{p\sigma}.$$

We want to incorporate this term into (23).

Let us introduce several identities which will be put together at the end. From $(a/\sqrt{\varepsilon} - \sqrt{b}\sqrt{\varepsilon})^2 \geq 0$ if follows

$$-ab \geq -\frac{a^2}{2\varepsilon} - \frac{b^2\varepsilon}{2}$$

and

$$C \cdot \nabla u_l = \nabla u_l^2 - \nabla u_l^A \cdot \nabla u_l^A \geq \nabla u_l^2 - \frac{(\nabla u_l^A)^2}{2\varepsilon} - \frac{\nabla u_l^2 \varepsilon}{2}$$

Also it holds that

$$\begin{aligned}
\left(\nabla u_l^A\right)^2 &= \frac{1}{|p|^2}\left(\sum_{\sigma \in \varepsilon_p} |\sigma|(u_\sigma - u_p)\mathbf{n}_{p\sigma}\right)^2 \\
&\leq \frac{2}{|p|^2} \sum_{\sigma \in \varepsilon_p} |\sigma|^2 (u_\sigma - u_p)^2 \\
&\leq \underbrace{\frac{2max|\sigma|d_{p\sigma})}{|p|}}_{\alpha} \frac{1}{|p|} \sum_{\sigma \in \varepsilon_p} \frac{|\sigma|}{d_{p\sigma}} (u_\sigma - u_p)^2,
\end{aligned}$$

where *max* is taken for all σ of element p. After rearrangement

$$\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}\frac{|\sigma|}{d_{p\sigma}}(u_\sigma-u_p)^2\geq\frac{|p|}{2\max|\sigma|\left|d_{p\sigma}\right|}\left(\nabla u^A\right)^2=\frac{1}{\alpha}\left(\nabla u_l^A\right)^2$$

we have that

$$\begin{aligned}&\sqrt{\frac{1}{|p|}\sum_{\sigma\in\varepsilon_p}\frac{|\sigma|}{d_{p\sigma}}(u_\sigma-u_p)^2+C.\nabla u_l}\\&\geq\sqrt{\frac{1}{\alpha}\left(\nabla u_l^A\right)^2+\nabla u_l^2-\frac{\left(\nabla u_l^A\right)^2}{2\varepsilon}-\frac{\nabla u_l^2\varepsilon}{2}}\\&=\sqrt{\underbrace{\left(\frac{1}{\alpha}-\frac{1}{2\varepsilon}\right)}_{I}\left(\nabla u_l^A\right)^2+\underbrace{\left(1-\frac{\varepsilon}{2}\right)}_{II}\nabla u_l^2}.\end{aligned}$$

We only need to show that *I* and *II* are positive. The term *II* is positive if $\varepsilon<2$. Now we must find ε such that the following inequality holds

$$\frac{1}{\alpha}-\frac{1}{2\varepsilon}>0$$

and this is true if

$$\varepsilon>\frac{\alpha}{2}.$$

So our ε must fulfill the following condition:

$$\frac{\alpha}{2}<\varepsilon<2.$$

If our geometry fulfills

$$max\left(|\sigma|\left|d_{p\sigma}\right|\right)<2|p|$$

then it is possible to find such ε that the approximation of the gradient norm given by (23) is positive.

In our case, the above inequality holds because of the geometrical properties of the grid mentioned in (2)–(5), at the end of Sect. 5 and the limited number of element shapes.

Example 3 We take again the consistent quadtree grid from Example 1 (see Fig. 3). On the paraboloid function $u(x, y) = 0.5 * (x^2 + y^2)$ which is defined on the same interval $[-1.25;1.25] \times [-1.25;1.25]$, we approximate the norm of the

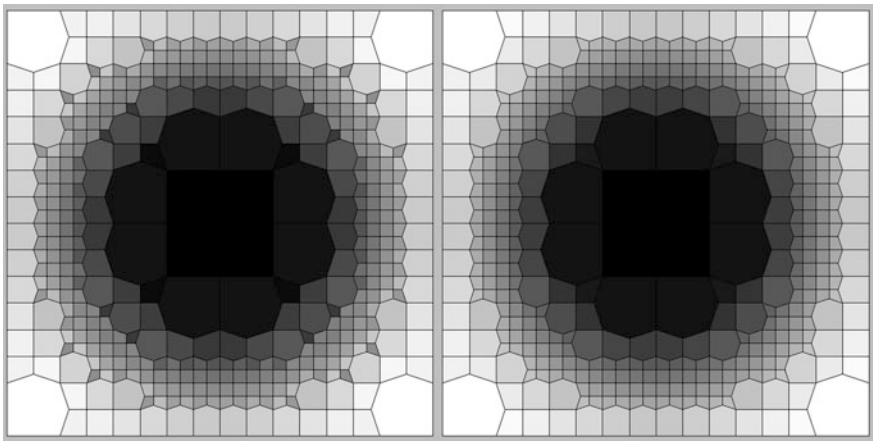


Fig. 10 Example 3. On the *left:* we evaluate (23) without a correction. On the *right*: we evaluate (23)

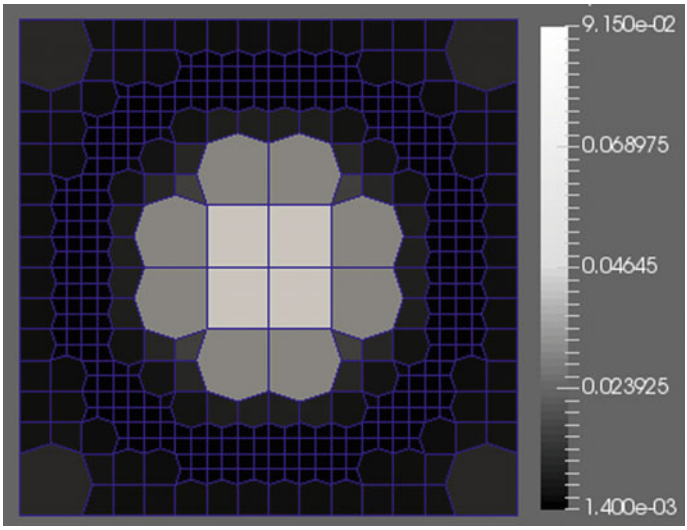


Fig. 11 Example 3. We display $\left|f_p^2 - f_p^1\right|$

gradient by (23). First, we omit the second term with the correction, the results are displayed in Fig. 10 on the left, then we use entire (23), the results are displayed in Fig. 10 on the right. In Fig. 11, we show the comparison of the first and the second approaches displaying $\left|f_p^2 - f_p^1\right|$. The results differ on larger elements corresponding to nonlinear situations.

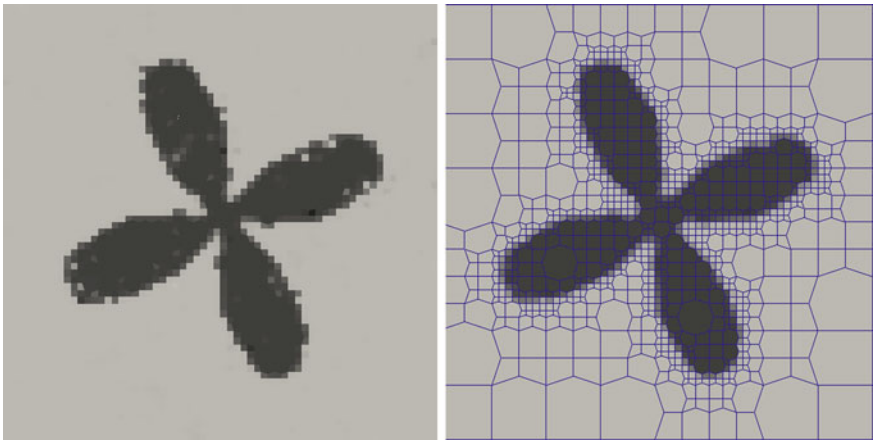


Fig. 12 Example 4. Filtration of image form Fig. 9 using (25). The second and the tenth time steps

6 Modification of MCF to Geodetic MCF

To prevent objects from shrinking, the edges in the image are detected by a gradient edge detector and using a decreasing function g with the norms of the gradients as the input, the movement of boundaries is slowed down. Outputs of g are numbers from (0;1]. This model is represented by the equation

$$\frac{\partial u}{\partial t} = |\nabla u| \nabla \cdot \left(g(|\nabla G_\sigma * u|) \frac{\nabla u}{|\nabla u|} \right). \tag{24}$$

The Gaussian gradient as an argument of g cannot be omitted—otherwise also salt and pepper configuration could lead to a small value in the output of g and slowed down the motion.

In our final experiment we use the scheme

$$\frac{u_p^n - u_p^{n-1}}{\left| f_p^2 \right|_\varepsilon^{n-1} \tau} |p| - \sum_{\sigma \in \varepsilon_p} \frac{|\sigma| g(f_p^1)}{\left| d_{p\sigma} f_p^2 \right|_\varepsilon} \left(u_\sigma^n - u_p^n \right) = 0. \tag{25}$$

The norm f^1 is suitable for detection of object boundaries and does not respond to the salt and pepper noise. On the other hand, f^2 is used for a curvature. (In this case of the special artificial image used in the experiment, we could omit smoothing of the gradient by the Gaussian).

Example 4 Using (25) we remove the salt and pepper noise and enhance the boundaries of the object from Fig. 9. The size of the original image is 64×64. In Fig. 11 we see the result after two time steps ($\tau = 1$) where most of the salt and

pepper noise is removed [5]. Then we show the solution after 10 time steps, when the boundary damaged by the noise is repaired, together with a computational grid. After two steps, the initial number of elements 4096 decreased to 1537 (37.5%), after 10 time steps to 1009 (24.6%). To detect the edges we used the function $g(s) = 1/(1 + Ks^2)$ in this experiment where K was set to 200.

Conclusion. The necessity of evaluating the gradient (14) to get (23), instead of to evaluate the norm directly, like on a regular grid, seems to be a drawback of the method. The time to get it depends on an implementation (e.g., the gradient need not to be corrected in the case of a zero correction, i.e., for some shapes [3], and also, if the first part of (23) is equal to zero.) If we solve GMCF, on the other hand, the gradient can be used to detect the boundaries because its norm f_p^1 does not respond to salt and pepper like noise. In the case of evaluating the Gaussian gradient, instead of $|\nabla(G * u)|$ we can perform $|G * \nabla u|$ and solve the linear heat equation on the adaptive grid in a fast way [5].

Acknowledgements This work was supported by VEGA1/0608/15, VEGA1/0714/15 and VEGA 1/0728/15. I would also like to thank my colleague Karol Mikula for advice and for oral communication.

References

1. Eymard, R., Gallouet, T., Herbin, R.: Finite volume method. In: Handbook for Numerical Analysis, vol. 7, pp. 713–1020. Elsevier, Amsterdam (2000)
2. Eymard, R., Handlovičová, A., Mikula, K.: Study of a finite volume scheme for the regularised mean curvature flow level set equation. IMA J. Numer. Anal. **31**, 813–846 (2011)
3. Krivá, Z., Handlovičová, A., Mikula, K.: Adaptive cell-centered finite volume method for diffusion equations on a consistent quadtree grid. Adv. Comput. Math. (accepted, 2015). doi:10.1007/s10444-015-9423-2
4. Krivá, Z., Mikula, K.: An adaptive finite volume scheme for solving nonlinear diffusion equations in image processing. J. Vis. Commun. Image Represent. **13**(1/2), 22–35 (2002)a
5. Petrovič, P.: Počítanie nelineárnych difúznych rovníc na konzistentných adaptívnych mriežkach. Diploma thesis. STU, Bratislava (2015)
6. Sethian, J.A.: Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Material Science. Cambridge University Press, New York (1999)

Human–Robot Interaction Through Robust Gaze Following

Sorin M. Grigorescu and Gigel Macesanu

Abstract In this paper, a probabilistic solution for gaze following in the context of joint attention will be presented. Gaze following, in the sense of continuously measuring (with a greater or a lesser degree of anticipation) the head pose and gaze direction of an interlocutor so as to determine his/her focus of attention, is important in several important areas of computer vision applications, such as the development of nonintrusive gaze-tracking equipment for psychophysical experiments in Neuroscience, specialized telecommunication devices, *Human–Computer Interfaces* (HCI) and artificial cognitive systems for *Human–Robot Interaction* (HRI). We have developed a probabilistic solution that inherently deals with sensor models uncertainties and incomplete data. This solution comprises a hierarchical formulation of a set of detection classifiers that loosely follows how geometrical cues provided by facial features are used by the human perceptual system for gaze estimation. A quantitative analysis of the proposed architectures performance was undertaken through a set of experimental sessions. In these sessions, temporal sequences of moving human agents fixating a well-known point in space were grabbed by the stereovision setup of a robotic perception system, and then processed by the framework.

1 Introduction

Head movements are commonly interpreted as a vehicle of interpersonal communication. For example, in daily life, human beings observe head movements as an expression of agreement or disagreement in a conversation, or even as a sign of confusion. On the other hand, gaze shifts are usually an indication of intent, as they commonly precede action by redirecting the sensorimotor resources to be used. As a

S.M. Grigorescu (✉) · G. Macesanu
Department of Automation, Transilvania University of Brasov, Mihai Viteazu 5,
500174 Braşov, Romania
e-mail: s.grigorescu@unitbv.ro

G. Macesanu
e-mail: gigel.macesanu@unitbv.ro

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_10

Fig. 1 Gaze following in the context of joint attention for HRI, using the ROVIS system on a Neobotix® MP 500 mobile platform

consequence, sudden changes in gaze direction can express alarm or surprise. Gaze direction can also be used for directing a person to observe a specific location. To this end, during their infancy, humans develop the social skill of *joint attention*, which is the means by which an agent looks at where its interlocutor is looking at by producing an eye-head movement that attempts to yield the same focus of attention. Over nine months of age, infants are known to begin to engage with their parents/caregivers in an activity in which both look at the same target through joint attention.

As artificial cognitive systems with social capabilities become more and more important due to the recent evolution of robotics towards applications where complex and human-like interactions are needed, basic social behaviors such as joint attention have increasingly become important research topics in this field. Figure 1 illustrates the ROVIS[1] (*Robust Vision and Control Laboratory*) gaze following system at work, under the context of joint attention for *Human Robotic Interaction* (HRI). Gaze following thus represents an important part of building a social bridge between humans and computers. Researchers in robotics and artificial intelligence have been attempting to accurately reproduce this type of interaction in the last couple of decades, and, although much progress has been made [1], dealing with perceptual uncertainty still renders it difficult for these solutions to work adaptively.

Gaze following is an example for which the performance of artificial systems is still far from human adaptivity. In fact, the gaze following adaptivity problem can be stated as follows: how can gaze following be implemented under nonideal circumstances (perceptual uncertainty, incomplete data, dynamic scenes, etc.)? Figure 2 demonstrates how incomplete data, arguably the issue where the lack of adaptivity and underperformance of artificial systems are most apparent, might influence the outcome of gaze following.

In the following text, we propose a robust solution to facial feature detection for human–robot interaction based on (i) a feedback control system implemented at the image processing level for the automatic adaptation of the system's parameters, (ii) a

[1] http://rovis.unitbv.ro.

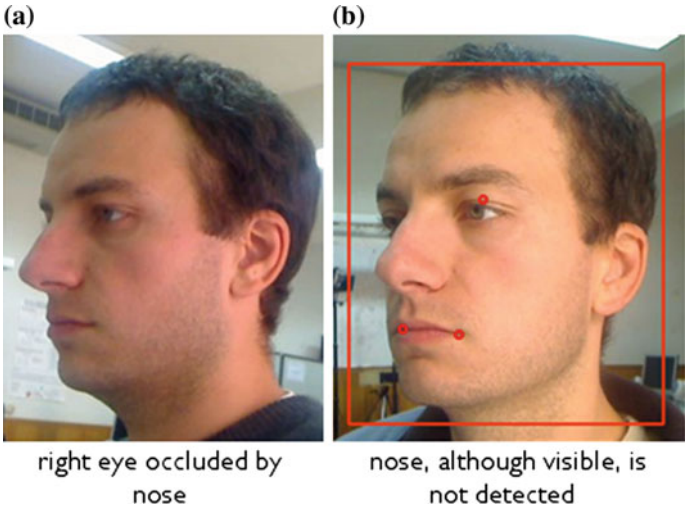


Fig. 2 Examples of probable gaze following failure scenarios due to incomplete data: facial features occluded in profile views (**a**), or failure of feature detection algorithms (**b**)

cascade of facial features classifiers, and (iii) a *Gaussian Mixture Model* (GMM) for facial points segmentation. The goal is to obtain a real-time gaze following estimator which can cope with uncertainties and incomplete data. The proposed system aims at the robust computation of the human gaze direction in the context of joint attention for HRI.

2 Related Work

2.1 Gaze Following

In recent years, the problem of gaze following has been extensively studied. Physiological investigations have demonstrated that the brain estimates the gaze as a mixture of eye direction and head position and orientation (pose) [2]. By itself, head pose provides an estimate that represents a coarse approximation of gaze direction that can be used in situations in which the eyes are invisible (e.g., when observing a distant person, or when sunglasses occlude the eyes) [3]. When the eyes are not occluded, the head pose is an extra marker that can be used to estimate the direction of the gaze. The gaze direction estimation problem, as it is solved by the human brain, can therefore be subdivided into two fundamental and *sequential* subproblems: *head pose estimation* and *eye gaze estimation*.

The consequences of such a solution are twofold: partial information can be used to already arrive to an estimate; however, this happens at the expense of biasing. As

Fig. 3 Wollaston illusion: although the eyes are the same in both images, the perceived gaze direction is dictated by the orientation of the head. (Adapted from [2, 3])

an illustration of this drawback, in Fig. 3 is shown [2] that the interpretation of the gaze for an observer is deviated in the direction of the head. In any case, the error propagated by erroneously estimating one of the features is greatly compensated by the fact that the human brain is able to yield an estimate *even when only presented with partial or incomplete information*. Moreover, visual features used to detect a face or an eye do not need to be the same for both cases, so they can be detected independently, which makes the problem more tractable.

Consequently, the following paragraphs will present a summarized survey of solutions for each subproblem.

In the survey by [3], solutions for head pose estimation are divided into eight categories: seven represent pure methods, while the remaining are hybrid methods, i.e., combinations of the other methods. The article ends by presenting a quantitative comparison of the performance of these methods.

As mentioned in this survey, most of the computer vision based head pose calculation algorithms have diverged greatly from the results of psychophysical experiments as to how the brain tackles this problem. In fact, the former are concentrated on *appearance-based* methods, while the latter takes into account how the human perceives the pose of the head based on *geometrical cues* [3].

Geometrical approaches, as shown in Fig. 4, attempt to detect head features as accurately as possible in order to compute the pose of the head. An example of a geometrical approach for head pose estimation is presented in [4], where monocular images are used as input information. The proposed algorithm makes minimal assumptions, compared with other methods, about the facial features structure. Knowing the positions of the nose, eyes, and mouth, the facial normal direction can be obtained from one of the next two methods [4], also used in our work:

1. Using two relations: the nose tip and the line between the far corners of the mouth ($R_1 = \frac{l_m}{l_f}$); the line between one eye with the correspondent far corners of the mouth and the distance given by the nose tip; and the line connecting one eye with the far corners of the mouth ($R_2 = \frac{l_n}{l_f}$);
2. Using the line between the eye extremities and the far mouth corners.

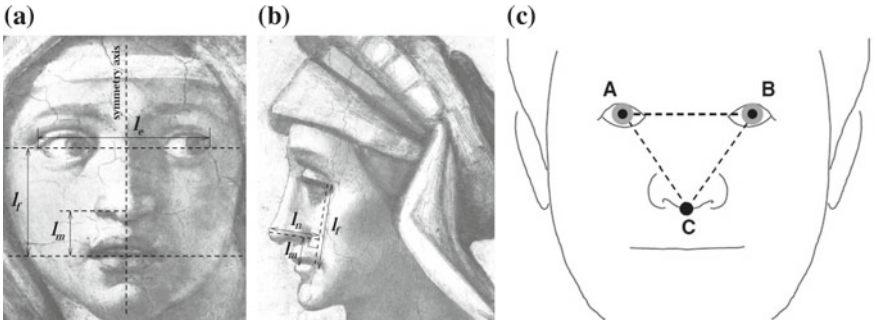


Fig. 4 Geometrical relations between facial features (Adapted from [4]). 3D gaze orientations can be computed using the distances between detected facial features, such as the eyes, nose and mouth

The derivation of the roll, pitch, and yaw for a human head is presented in [5]. The assumption from this article is that the four points that describe the eye are collinear. The position is obtained using the line through the four eye points and the nose tip. The main difficulties with this method are related to the pitch direction estimation, which uses an anthropometric face analysis [5]. The yaw and the pitch are obtained from eye corners and the intrinsic camera parameters (focal length).

The method proposed by [6] uses the model of the face and the eye, deduced from anthropometric features in order to determine the head orientation. This method uses only three points (e.g., eye centers and the middle point between the nostrils) to perform the desired task. Their model uses the following assumption: $d(A, C) = d(B, C)$; $d(A, B) = kd \cdot d(A, C)$; $d(A, B) = 6,5$ cm, where A and B are the central points of each eye and C is the middle point between the nostrils.

Another solution for head pose estimation is introduced in [7]. The main idea here is to consider an isosceles triangle, with corners in both eyes and in the center of the mouth. The direction of the head is computed if we assume that one side of the triangle lies on the image plane, such that applying a trigonometric function we can estimate the angle between the triangle plane and the image plane [7].

Finally, an alternative method for head estimation is supposed to use multiple cameras [8] with accurate calibration information available. Skin color segmentation is performed on each camera, and then data fusion is performed, resulting in a 3D model of the head. The orientation of the head is estimated based on a particle filter.

2.2 Facial Features Extraction

Feature detection represents a subtopic within the head pose estimation problem. An accurate estimate for the eye, nose, or the mouth represents an intermediate stage, in which essential information used by the geometrical approach for head pose estimation is computed. Methods for gaze estimation, presented in the following section,

include eye feature detection. Detection of other important facial features, such as the mouth and the nose, is discussed next.

Mouth recognition is dealt with methods such as the ones suggested in [9, 10]. A common approach for detecting the mouth is by pre-segmenting the color red on a specific patch of the image. Both methods use a ROI (Region of Interest) extracted after head segmentation, in which the mouth is approximately segmented, after a color space conversion is performed (such as RGB to HSI (*Hue, Saturation, Intensity*) [9], or RGB to *Lab* [10]). On the other hand, nose detection algorithms use Boosting classifiers, commonly trained with Haar-like features [11], or the 3D information of the face, as in [12].

As suggested in [13], most of the methods used for eyes detection and segmentation can be divided into shape-based, appearance-based and hybrid methods. The shape-based technique uses the detection of the iris, the pupil, or the eyelids to locate the eye. Particular features, such as the pupil (dark/bright pupil region) or cornea reflections are used in appearance-based approaches, while the hybrid method tries to combine the advantages of both methods.

The shape-based algorithm proposed in [14], built on the isophote curvature concept, i.e., the curve that connects points of the same intensity, is able to deliver accurate eye localization from a web camera. The main advantage of using this concept is that the shape of the isophotes is invariant to rotation or to linear illumination changes. The eye location can be determined using a combination of Haar features, dual orientation Gabor filters and eye templates, as described in [15].

Unsupervised learning algorithms, such as the *Independent Component Analysis* (ICA), are used in [16] for eyes extraction, based on the fact that the eye is a stable facial feature. The two stages technique determines first a rough eye ROI using ICA and the gray-level image intensity variance, and second, the eye center point is computed from image intensity data.

Finally, an alternative method which uses two visual sensors is proposed in [17]: a wide-angle camera for face detection and rough eyes estimation and an active pan–tilt–zoom camera to focus on the rough detected ROIs. The method considers the face as a 3D terrain surface and the eye areas as ”pits” and ”hillsides” regions. The eyes 2D positions are chosen using a (GMM). A similar dual stereo camera system is also proposed in [18], where a wide-angle camera detects the face and an active narrow *Field of View* (FoV) system tracks the eyes at high resolution.

As mentioned above, most methods tackle the problem of gaze direction estimation using either head pose or eyes direction estimation. However, papers such as [14, 19, 20] present hybrid approaches that combine head pose and eye direction estimation for obtaining the subject's gaze direction.

In [14], a hybrid solution for eye detection and tracking, combining the detection results with a *Cylindrical Head Model* (CHM) for head direction estimation, is presented. In [19], the gaze's direction is computed in two stages, after a camera calibration process: first the eyes orientation vector is determined with respect to the head's coordinate system and, second, the final gaze direction estimate is given by a fusion between the determined eyes and head's poses. Both approaches have lim-

itations in estimating the gaze's orientation when either the eyes or the poses of the head are imprecise.

The technique from [20] describes a human gaze direction algorithm from a combination of *Active Appearance Models* (AAM) and a CHM. Although the approach seems to perform well in off-line experiments, real-time scenarios are not presented. One other notable facial features extractor is the Flandmark system [21], which, despite its real-time capabilities and ability to detect and track facial features from frontal faces, fails to recognize features when the pose of the head has a slight offset from the frontal view.

3 Controlling a Machine Vision System

In a robotics application, the purpose of the machine vision system is to perceive the environment through a camera module.

An image processing chain is usually composed of low (e.g., image enhancement, segmentation) and high (e.g., object recognition) level image processing methods. In order for the high level operations to perform properly, the low level ones have to deliver reliable information. In other words, object recognition methods require reliable input coming from previous operations [22].

In order to improve the image processing chain, we propose to control the low level vision operation through a feedback loop derived from the higher level components. In [23, 24], the inclusion of feedback structures within vision algorithms for improving the overall robustness of the chain is suggested.

The core idea of the feedback control system for adapting the low level vision operations is presented in Fig. 5, where the control signal u, or *actuator variable*, is a parameter which controls the processing method, whereas the *controlled variable* y is a measure of image processing quality.

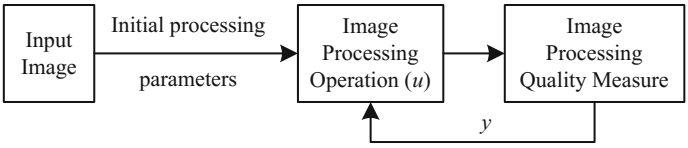


Fig. 5 Feedback adaptation of a computer vision algorithm. The image processing quality measure y is used as a feedback control variable for adapting the parameters of the vision algorithms using the actuator u

4 Image Processing Chain

The gaze following image processing chain, depicted in Fig. 6, contains four main steps. We assume that the input is an 8-bit gray-scale image $I = J^{V \times W}$, of width V and height W, containing a face viewed either from a frontal or profile direction, where $J = \{0, \ldots, 255\}$. (v, w) represents the 2D coordinates of a specific pixel. The face region is obtained from a face detector.

First, a set of facial features ROI hypotheses $\mathbf{H} \in \{h_{le}, h_{re}, h_n, h_m\}$, consisting of possible instances of the left h_{le} and right h_{re} eyes, nose h_n and mouth h_m, are extracted using a local features estimator which determines the probability measure $p(\mathbf{H}|I)$ of finding one of the searched local facial region. The number of computed ROI hypotheses is governed by a probability threshold T_h, which rejects hypotheses with a low $p(\mathbf{H}|I)$ confidence measure. The choice of the T_h threshold is not a trivial task when considering time critical systems, such as the gaze estimator, which, for a successful HRI, has to deliver in real-time the 3D gaze orientation of the human subject. The lower T_h is, the higher the computation time. On the other hand, an increased value for T_h would reject possible "true positive" facial regions, thus leading to a failure in gaze estimation. As explained in the following, in order to obtain a robust value for the hypotheses selection threshold, we have chosen to adapt T_h with respect to the confidences provided by the subsequent estimators from Fig. 6, which take as input the facial regions hypotheses. The output probabilities coming from these estimation techniques, that is, the spatial estimator and the GMM for point-wise feature extraction, are used in a feedback manner within the extremum seeking control paradigm.

Once the hypotheses vector $\mathbf{H}$ has been built, the facial features are combined into the spatial hypotheses $\mathbf{g} = g_0, g_1, \ldots, g_n$, thus forming different facial region combinations. Since one of the main objectives of the presented algorithm is to identify facial points of frontal, as well as profile faces, a spatial vector s_i is composed either from four, or three, facial ROIs:

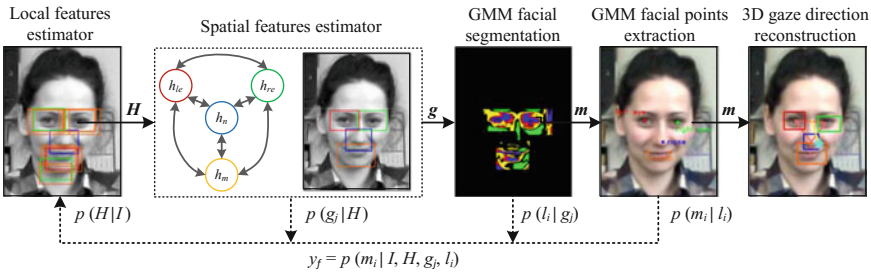


Fig. 6 Block diagram of the proposed gaze following system for facial feature extraction and 3D gaze orientation reconstruction. Each processing block within the cascade provides a measure of feature extraction quality, fused within the controlled variable y_f (see Eq. 2)

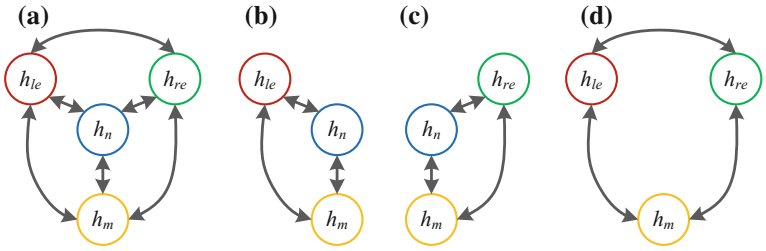


Fig. 7 Different spatial combinations of features used for training the four classifiers. **a** All four facial features. **b**, **c**, **d** Cases where only three features are visible in the sample image

$$g_i = \{h_0, h_1, h_2, h_3\} \cap \{h_0, h_1, h_2\}, \tag{1}$$

where $h_i \in \{h_{le}, h_{re}, h_n, h_m\}$.

The extraction of the best spatial features combination can be seen as a graph search problem $g_j = f : G(\mathbf{g}, \mathbf{E}) \rightarrow \Re$, where $\mathbf{E}$ are the edges of the graph connecting the hypotheses in $\mathbf{g}$. The considered features combinations are illustrated in Fig. 7. Each combination has a specific spatial probability value $p(g_j|\mathbf{H})$ given by a spatial estimator trained using the spatial distances between the facial features from a training database.

Once the spatial distributions of the probable locations of the facial features ROIs are available, their pointwise location m_i is determined using a GMM segmentation method. Its goal is to extract the most probable facial pointwise locations m_i given the GMM pixel likelihood values $p(l_i|g_j)$. The most relevant point features for computing the 3D gaze of a person are the centers of the eyes, tip of the nose, and corners of the mouth.

The described data analysis methods are used to evaluate a feature space composed of the local and spatial features.

Having in mind the facial feature points extraction algorithm described above, it can be stated that the confidence value y_f of the processing chain in Fig. 6 is a probability confidence measure obtained from the estimators cascade:

$$y_f = p(m_i|I, \mathbf{H}, g_j, l_i). \tag{2}$$

Since the whole described processing chain is governed by a set of parameters, such as the threshold T_h for selecting the vector $\mathbf{s}$, we have chosen to adapt it using an extremum seeking control mechanism and the feedback variable y_f, derived from the output of the gaze following structure illustrated in Fig. 6. The final 3D gaze orientation vector $\vec{\varphi}(m_i)$, representing the roll, pitch, and yaw of the human subject, is determined using the algorithm proposed in the work of Gee and Cipolla [4].

5 Performance Evaluation

5.1 Experimental Setup

In order to test the performance of the proposed gaze following system, the following experimental setup has been prepared.

The system has been evaluated on the *Labeled Faces in the Wild* (LFW) database [25]. LFW consists of 13,233 images, each having a size of $250 \times 250px$. In addition to the LFW database, the system has been valuated on an Adept Pioneer® 3-DX mobile robot equipped with an RGB-D sensor delivering $640px \times 480px$ size color and depth images. The goal of the scenarios is to track the facial features of the human subject in the HRI context. The error between the real and estimated facial feature's locations was computed offline.

For evaluation purposes, two metrics have been used:

- the mean normalized deviation between the ground truth and the estimated positions of the facial features:

$$d(\mathbf{m}, \hat{\mathbf{m}}) = \tau(\mathbf{m}) \frac{1}{k} \sum_{i=0}^{k-1} \|m_i - \hat{m}_i\|, \tag{3}$$

  where k is the number of facial features, $\mathbf{m}$ and $\hat{\mathbf{m}}$ are the manually and estimated annotated positions of the eyes, nose and mouth, respectively, and $\tau(\mathbf{m})$ is a normalization constant:

$$\tau(\mathbf{m}) = \frac{1}{\|(m_{le} + m_{re}) - m_m\|}. \tag{4}$$

- the maximal normalized deviation:

$$d^{\max}(\mathbf{m}, \hat{\mathbf{m}}) = \tau(\mathbf{m}) \max_{j=0,\ldots,k-1} \|m_i - \hat{m}_i\|. \tag{5}$$

5.2 Competing Detectors

The proposed gaze following system has been tested against three open source detectors.

(1) *Independent facial feature extraction*: The detector is based on the Viola–Jones boosting cascades and returns the best detected facial features, independent of their spatial relation. The point features have been considered to be the centers of the computed ROIs.
The boosting cascades, one for each facial feature, have been trained using a

few hundred samples for each eye, nose, and mouth. The searching has been performed several times at different scales, with Haar-like features used as inputs to the basic classifiers within the cascade. From the available ROI hypotheses, the one having the maximum confidence value has been selected as the final facial feature.

(2) *Active Shape Models*: An *Active Shape Model* (ASM) calculates a set of feature points along the facial features contours of the eyes, nose, mouth, eyebrows, or chin. An ASM is initially trained using a set of manually marked contour points. The open source AsmLib, based on OpenCV, has been used as candidate detector. The ASM is trained using manually marked face contours. The trained ASM model determines variations in the training dataset using *Principal Component Analysis* (PCA), which enables the algorithm to estimate if the contour is a face.

(3) *Flandmark*: *Flandmark* [21] is a deformable part model detector of facial features, where the detection of the point features is treated as an instance of struc-

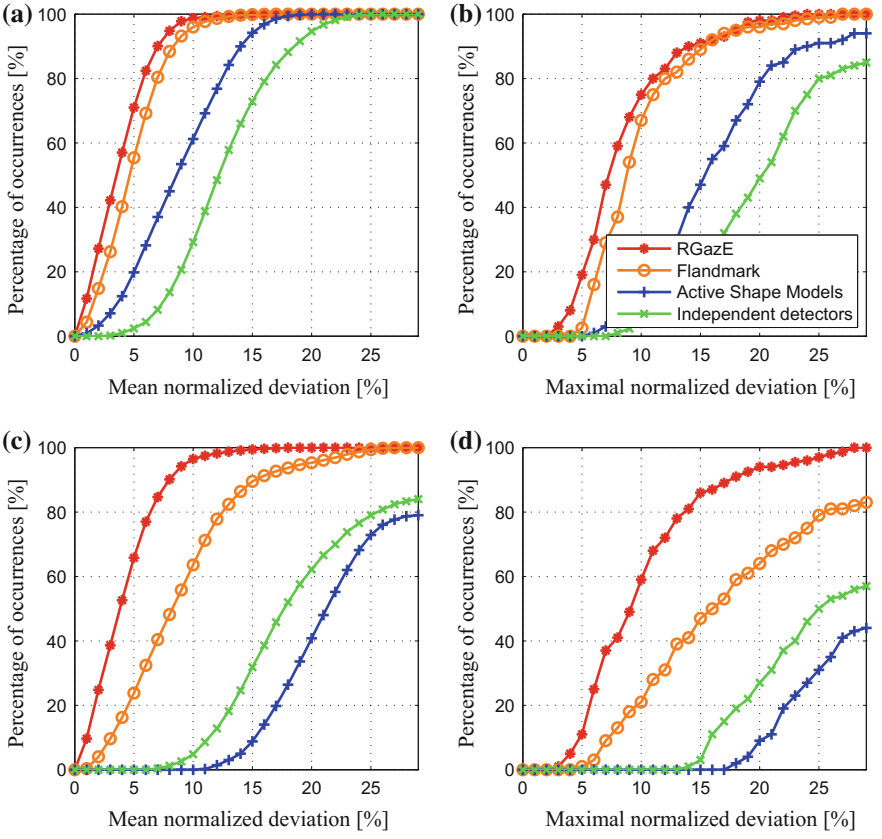


Fig. 8 Cumulative histograms for the mean and the maximal normalized deviation shown for all competing detectors applied on video sequences with frontal (**a**, **b**) and profile (**c**, **d**) faces

tured output classification. The algorithm is based on a *Structured Output Support Vector Machine* (SO-SVM) classifier for the supervised learning of the parameters for facial points detection from examples.

In comparison to our gaze following system, which uses a segmentation step for determining the pointwise location of the facial features, flandmark considers the centers of the detected ROIs as the point location of the eyes, nose, and mouth.

The mean and maximal deviation metrics were used to compare the accuracy of the four tested detectors with respect to the ground truth values available from the benchmark databases. Especially for the evaluation of the computation time, the algorithm has also been tested on a mobile robotic platform.

The cumulative histograms of the mean and maximal normalized deviation are shown in Fig. 8 for frontal and profile faces. In all cases, the proposed estimator delivered an accuracy value superior to the ones given by the competing detectors. If the accuracy difference between our algorithm and Flandmark is relatively low for the case of frontal faces, it actually increases when the person's face is imaged from a profile view.

An interesting observation can be made when comparing the independent detectors with the ASM one. Although the ASM outperforms independent facial feature extraction on frontal faces, it does not perform well when the human subjects are viewed from the lateral. This is due to the training nature of the ASM, where the input training data is made of points spread on the whole frontal area (e.g., eyes, eyebrows, nose, chin, cheeks, etc.).

6 Conclusion

In this paper, a robust facial features detector for 3D gaze orientation estimation has been proposed. The solution is able to return a reliable gaze estimate, even if only a partial set of facial features is visible. The paper brings together algorithms for facial feature detection, machine learning, and control theory. During the experiments, we investigated the system's response and compare the results to ground truth values. As shown in the experimental results section, the method performed well with respect to various testing scenarios. As future work, the authors consider the possibility of extending the framework for the simultaneous gaze estimation of multiple interlocutors and the adaptation of algorithm with respect to the robot's egomotion.

Acknowledgements We hereby acknowledge the structural founds project PRO-DD (POS-CCE, O.2.2.1., ID 123, SMIS 2637, ctr. No 11/2009) for providing the infrastructure used in this work.

References

1. Scassellati, B.: Theory of mind for a humanoid robot. Auton. Robots **12**(1999), 13–24 (2002)
2. Langton, S.R.H., Honeyman, H., Tessler, E.: The influence of head contour and nose angle on the perception of eye-gaze direction. Atten. Percept. Psychophys. **66**(5), 752–771 (2004)
3. Chutorian, E., Trivedi, M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–629 (2009)
4. Gee, A., Cipolla, R.: Determining the gaze of faces in images. Image Vis. Comput. **12**(10), 639–647 (1994)
5. Horprasert, T., Yacoob, Y., Davis, L.: Computing 3-d head orientation from a monocular image sequence. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 242–247, Oct 1996
6. Kaminski, J., Knaan, D., Shavit, A.: Single image face orientation and gaze detection. Mach. Vis. Appl. **21**(3), 85–98 (2009)
7. Nikolaidis, A., Pitas, I.: Facial feature extraction and pose determination. Pattern Recogn. **33**(11), 1783–1791 (2000)
8. Canton-Ferrer, C., Casas, J., Pardas, M.: Head orientation estimation using particle filtering in multiview scenarios. In: Multimodal Technologies for Perception of Humans, vol. 4625, pp. 317–327. Springer, Berlin (2008)
9. Pantic, M., Tomc, M., Rothkrantz, L.: A hybrid approach to mouth features detection. In: 2001 IEEE International Conference on Systems, Man, and Cybernetics, vol. 2, pp. 1188–1193 (2001)
10. Skodras, E., Fakotakis, N.: An unconstrained method for lip detection in color images. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1013–1016 (2011)
11. Gonzalez-Ortega, D., Diaz-Pernas, F., Martinez-Zarzuela, M., Anton-Rodriguez, M., Diez-Higuera, J., Boto-Giralda, D.: Real-time nose detection and tracking based on adaboost and optical flow algorithms. In: Intelligent Data Engineering and Automated Learning, vol. 5788, pp. 142–150. Springer, Berlin (2009)
12. Werghi, N., Boukadia, H., Meguebli, Y., Bhaskar, H.: Nose detection and face extraction from 3d raw facial surface based on mesh quality assessment. In: 36th Annual Conference on IEEE Industrial Electronics Society, pp. 1161–1166 (2010)
13. Hansen, D., Ji, Q.: In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 78–500 (2010)
14. Valenti, R., Sebe, N., Gevers, T.: Combining head pose and eye location information for gaze estimation. IEEE Trans. Image Process. (2011)
15. Ke, L., Kang, J.: Eye location method based on haar features. In: 2010 3rd International Congress on Image and Signal Processing, vol. 2, pp. 925–929 (2010)
16. Hassaballah, M., Kanazawa, T., Ido, S.: Efficient eye detection method based on grey intensity variance and independent components analysis. Comput. Vis. IET **4**(4), 261–271 (2010)
17. Reale, M., Canavan, S., Yin, L., Hu, K., Hung, T.: A multi-gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing. IEEE Trans. Multimedia **13**(3), 474–486 (2011)
18. Beymer, D., Flickner, M.: Eye gaze tracking using an active stereo head. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 451–458 (2003)
19. Ronsse, R., White, O., Lefevre, P.: Computation of gaze orientation under unrestrained head movements. J. Neurosci. Methods **159**, 158–169 (2007)
20. Sung, J., Kanade, T., Kim, D.: Pose robust face tracking by combining active appearance models and cylinder head models. Int. J. Comput. Vis. **80**, 260–274 (2008)
21. Uřičář, M., Franc, V., Hlaváč, V.: Detector of facial landmarks learned by the structured output SVM. In: Csurka, G., Braz, J. (eds.) VISAPP '12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications, vol. 1, pp. 547–556. SciTePress—Science and Technology Publications, Portugal, Feb 2012

22. Hotz, L., Neumann, B., Terzic, K.: High-level expectations for low-level image processing. In: KI 2008: Advances in Artificial Intelligence. Springer, Berlin (2008)
23. Ristic, D.: Feedback structures in image processing. Ph.D. dissertation, Bremen University, Institute of Automation, Bremen, Germany, Apr 2007
24. Grigorescu, S.M.: Robust machine vision for service robotics. Ph.D. dissertation, Bremen University, Institute of Automation, Bremen, Germany, June 2010
25. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments, University of Massachusetts, Amherst. Technical Report 07–49, Oct 2007

The Detection of the Retina's Lesions in Optical Coherence Tomography (OCT)

Joanna Swiebocka-Wiek

Abstract The Optical Coherence Tomography (OCT) is a very modern, noninvasive, and noncontact optical imaging technique. It is dedicated to different types of ocular tissues such as for example: the retina, optic disk, or cornea. During the examination, OCT is used for the early diagnosis of diseases such as: glaucoma, macular degeneration (AMD), diabetic changes in the retina, macular hole, macular edema, and eye cancer. These may inevitably lead to blindness, hence it is so important to ensure the patient of early and accurate diagnosis. The main goal of this paper is to propose a preliminary, automated method of detecting the occurrence of pathological changes in the retina (cysts and inflammation).

1 Introduction

The retina is a blood membrane with multiple vessels, located at the back of the eye. Its primary function is to receive visual signals. Nerve cells belonging to the retina are arranged in layers and connected to the brain by the optic nerve. There are many diseases which change the system of blood vessels in the retina. That is why the retina's examination is a source of much valuable information about the patient's state of health and the subject of numerous studies on the diagnostic's methods development. One of the most important parts of the retina is the macula (or macula lutea from latin *macula* meaning spot and *lutea*, meaning yellow) which is an oval-shaped pigmented area close to the retina's center. It has a diameter of around 5.5 mm and it is responsible for high-acuity vision. Within the macula two other important structures are placed a fovea and foveola that both contain a high density of cones (photoreceptors). The great importance of this structure may be illustrated by the fact that even small damage or inflammation of the macula may lead to loss of central vision and, in extreme cases, even blindness. Diseases leading to impaired work macula is

J. Swiebocka-Wiek (✉)
Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, 30–059 Kraków, Krowodrza, Poland
e-mail: jsw@agh.edu.pl

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_11

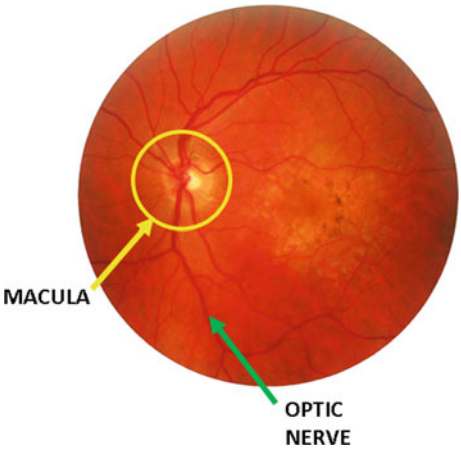


Fig. 1 The appearance of the retina. The location of the optic nerve and macula was marked

detailed later in this paper. An example of the retina with optic nerve and macula location was demonstrated in Fig. 1.

The application of interferometry for the purpose of in vivo imaging human eye structures has fascinated scientists for a long time. Initial work, using white, visible light were presented at the ICO-15 SAT Conference in 1990 [1]. In the same year two works of Naohiro Tanno were published, proposing a new technique which became a leadership approach in the field of tissue imaging, in the micrometer scale resolution [2, 3].

This method called Optical Coherence Tomography (OCT) is based on receiving and processing the optical signal which mainly uses light. It captures three-dimensional (3D) images with very good (even with micrometers) spatial resolution, obtained from scattering centers. It is highly useful in medical applications (biological tissues and their cross-section imaging). The huge and still growing popularity of OCT as an imaging technique is the reason for developing image processing algorithms. The subject of this work is to present a method for automatically distinguishing images of healthy and diseased retina. The OCT as the diagnostic technique is limited to imaging in the range of 1–2 mm below the tissue's surface. It is caused by the fact that in case of greater depths the proportion of undistracted light is too small to be properly detected and registered. It is also worth to mention that in case of OCT, no earlier preparation of a biological tissue or structure is needed. The images can be obtained without any contact with an object (what in the case of the retina's examination is a huge convenience) or through a transparent membrane. It is also important to highlight that all procedure take place using near-infra-red light which is safe for the eye and therefore significantly minimizes the likelihood of a sample's damage.

1.1 Theoretical Basics

Optical Coherence Tomography allows to obtain high-resolution images because, as it was already mentioned in previous section, it is based on applying near to the infrared (IR) wavelength light waves, instead of much lower radio frequency waves (like in case MRI application) or sound frequency waves (used in USG diagnostic procedure) [4].

An optical beam is directed onto the tissue and a small part which is reflected from the elements below the surface is recorded [5]. However, most of the light is not reflected, but is scattered at high angles. In conventional imaging methods widely scattered light obscures, however OCT uses interferometry to register the length of the path traveled by the photons and reject most of those that have been scattered several times before reaching the receiver. Therefore, OCT allows to create detailed three-dimensional (3D) images even for very thick samples, simply by the rejection of background signal, which disturb the perception of light reflected directly from the surface of the sample [6].

Among the many noninvasive methods dedicated for three-dimensional imaging, OCT shows similarity to ultrasound imaging, which also uses the echoes occurrence [7]. As a diagnostic technique OCT is limited to imaging only 1–2 mm deep under the surface of the biological tissue. Deeply scattered light is greater and the amount of light reflected without dissipation is too small to be registered. It is worth emphasizing that using the OCT method does not require prior preparation of the sample, and images can be obtained without touching it, even by a layer of transparent film or window [7]. The OCT method is based on white light or low coherence interferometry. It uses light from the broadband light source (super-luminescent diode), which is divided in a beam splitter into two beams: reference and a sample one. It allows to receive the retina's profile where reflectivity is versus depth. Some part of the light which is backscattered from the retina interfere with the reference beam. It allows to obtain an interference pattern which is used for measuring the light with respect to the tissue's depth profile. Owing to the fact that axial resolution is between 5 and 7 μm it allows to obtain the results comparable with in vivo retina's biopsy [8]. The optical setup for the OCT procedure was shown in Fig. 2.

1.2 Clinical Application

One of the medical areas where over the past decade OCT has become one of the most important and widespread examinations is ophthalmology [9, 10]. As a method which seems to have revolutionized the clinical practice, it is commonly used for noninvasive eye tests and to obtain detailed, cross-sectional images of the inside of the retina, retina nerve fiber layer and optic nerve [6]. The first images showing retina's structure were published in 1993 [11]. Retinal OCT imaging provides high-resolution imagery of subsurface retinal features that were inaccessible for previous

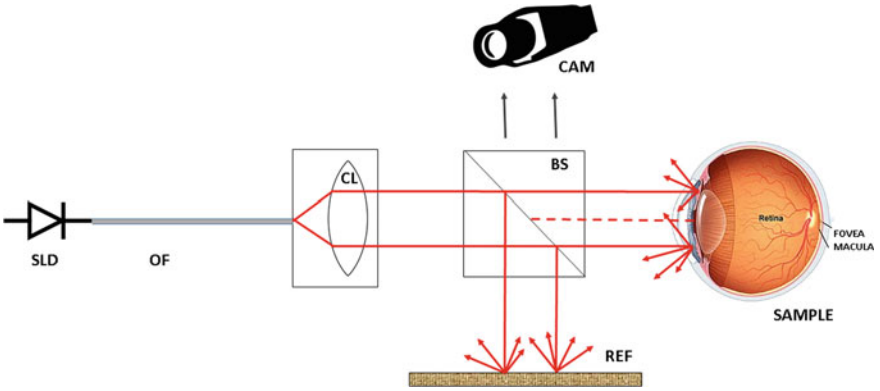


Fig. 2 OCT optical setup (*SLD*—super-luminescent diode, *OF*—optical fiber, *CL*—convex lens, *BS*—beam splitter, *CAM*—CMOS camera, *REF*—reference, *SAMPLE*—retina)

techniques (fluorescein angiography or ophthalmic ultrasound). Due to the possibility of imaging choroidal thickness in the retina, the most common indications for OCT examination are:

- age-related macular degeneration (AMD),
- diabetic maculopathy,
- glaucoma,
- macular edema of different origin,
- macular hole or fibrosis,
- central serous retinopathy.

Macular edema, which is the main subject of interest in this work, occurs in many ophthalmological diseases, when fluid and proteins gather on (*intraretinal*) or under (*subretinal*) the macula. As a consequence the macula thickens and swells. The swelling is the most important reason for a patient's central vision distortion (as it was mentioned, the macula consists of tightly framed photoreceptors (cones), responsible for clear and sharp vision of objects that are placed directly in the center of a person's field of view.

In Figs. 3 [12] and 4 [13] the comparison of the OCT scan for health and pathologically changed retina was presented.

What is more, OCT methods (spectroscopic OCT, PS-OCT) also started to be used in interventional cardiology as a method of ischemic heart disease diagnosis (structural examination of the vasculature in the coronary artery) and molecular analysis. OCT technique is also applied in the field of oncology for evaluation of the surgical margins or the detection of small lesions which are unable to be diagnosed in overall examination. Relatively recent development is OCT application in case of imaging dental structures or musculoskeletal tissues [10, 14]. Generally, changes that occur in the retina, capable of being diagnosed using OCT technique can be divided into two categories:

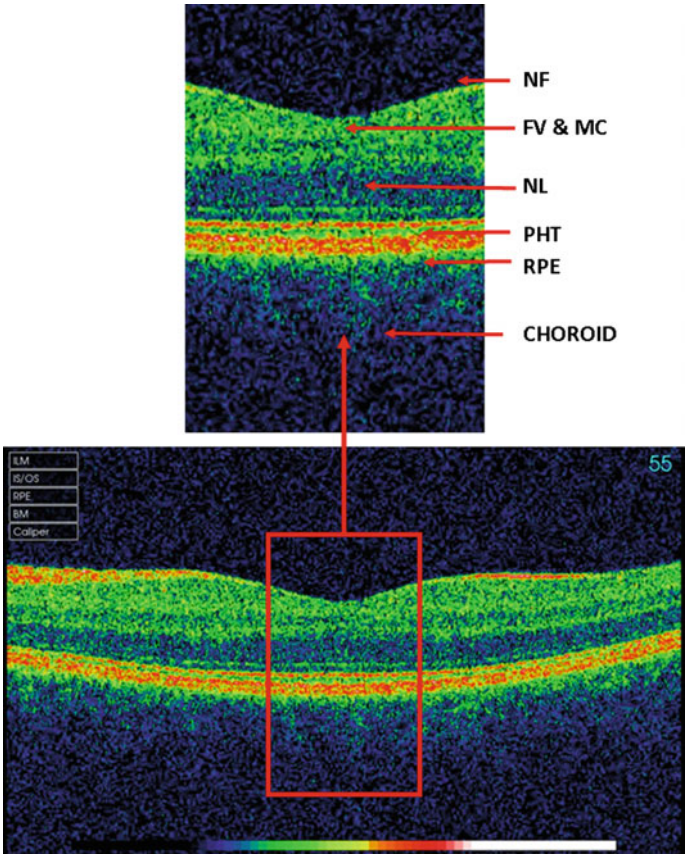


Fig. 3 The cross-section through the central part of the retina of healthy eyes in the OCT examination. *FV*—fovea, *MC*—macula, *NF*—nerve fiber, *NL*—nucleus layer (inner and outer layer), *PHT*—photoreceptors (inner and outer complex) *RPE*—retina pigment epithelium

(1) vascular changes usually associated with the presence of fluid within the retina (subretinal fluid) or gathering the fluid above the photoreceptor layer (intraretinal fluid),
(2) changes related to the break in the retina's structure (macular holes, pseudomacular holes).

Due to the fact that the proposed algorithm is dedicated to the detection of vascular lesions, the second type of changes will not be the subject of interest in this paper. OCT applications for some of the most common edema's diseases are discussed. The ability to confirm the presence of fluid in the retina, choroidal malformations, and distortions in retina's thickness are very helpful in clinical decisions and planning the treatment.

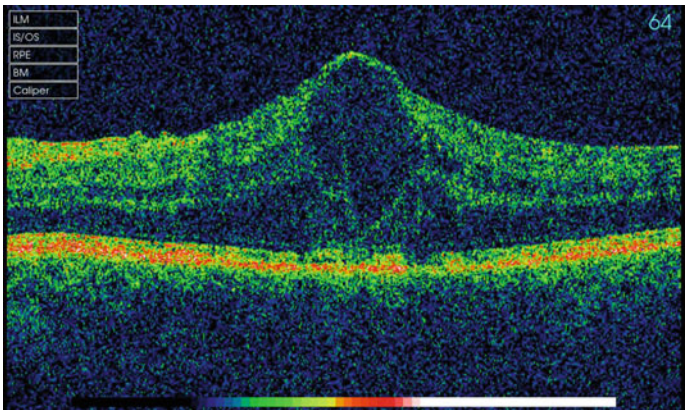


Fig. 4 The cross-section through the central part of the retina of the eye. Macula in the center part of the image is pathologically covered by cysts

2 Results

All the procedures were implemented in the MATLAB environment. The main steps of the algorithm are

(1) choosing the green channel from the image as most representative according to image diagnostic value,
(2) image thresholding,
(3) noise reduction (removing single isolated white pixels),
(4) improvement of the boundary line between the retina's structures using morphological operators,
(5) edge detection,
(6) calculation of the shape coefficients for the biggest structures,
(7) identification of a cyst's presence by finding objects with specified shape coefficient,
(8) comparison of potential cyst and macula localisation.

Verification of the algorithms efficiency was made for three specific diseases, characterized by vascular edema in the retina

(1) cystoid macular edema (CME) in diabetes,
(2) branch retinal vein occlusion (BRVO)—2 cases,
(3) idiopathic polypoidal choroidal vasculopathy (IPCV).

The results were discussed in the following subsections. All steps of the algorithm were described in details in the section related to analysis of OCT image in diabetes. Because of the fact that in the other two cases, the algorithm has an identical process and steps, its description was considered as unnecessary.

2.1 Diabetic and Cystoid Macular Edema (DME and CME)

Chronic or uncontrolled diabetes (type II) can influence retina's blood vessels including and cause leaks of fluid, blood, and occasionally fats into the retina causing its swelling [8]. Based on Fig. 6 of the retina affected diabetes all the steps of the proposed algorithm are discussed in details. Analysis of Fig. 3 (presenting healthy retina) and Fig. 4 (presenting macular edema) shows that in the RGB color scheme, the greater contribution comes from the green channel. Moreover, the human eye has the highest sensitivity to green color (the rods are most sensitive at a wavelength of 500 nm and suppositories at a wavelength of 550 nm). That is why it is considered as a most important in further analysis. In the next step, after choosing only the green channel (reducing image in color to grayscaling image), the image was filtered with a global threshold to divide the retina into separate structures. Unfortunately binarisation reveals the existence of unwanted noise in the form of numerous individual white pixels. The most probable cause of the noise is the resolution OCT method. To remove this effect, morphological operators (mainly erosion) were implemented.

Morphological operations are based on the use of the movable transformation core, called a structural element. It may have different shapes and sizes and contain any combination of 0 and 1 values. If a pixel's value is not significant it could be marked in the structural element as z. In this paper, the following morphological operations were tested and applied [?]:

- **Dilatation** which is an operation that thickens objects in a binary image (or growth). Dilation is one of the most fundamental morphological operations. The manner and extent of this process is controlled by the structural element, which is compared with each pixel of the image. In other words if at least one pixel in the neighborhood has a value equal to "1" the focal point also receives it (in another case the value "0" is assigned). Types of structural element strongly affects the output image.
- **Erosion** which is also one of the most fundamental morphological operations that thins (or shrinks) objects in a binary image. This operation applies a rotated structural element for each pixel in the image. If even one pixel in the neighborhood has a value equal to 0, the focal point also receives this value. Otherwise, its value does not change. This is an operation which is the inverse of dilatation. Erosion is significantly influenced by the choice of the structural element.
- **Opening** Assembling of dilatation and erosion processed on the original image. It causes image smoothing (removal of details, the greater the structural element is used, the stronger image smoothing can be observed).
- **Closing** Assembling of erosion and dilatation processed on the original image. It removes all the holes in the image and the concave lower than the structural element (the greater structural element, the more elements are filled in).
- **Hit or miss** transformation. The main reason of this kind of transformation is to identify some specified pixels configurations (isolated foreground pixels belonging to the line segments or endpoints).

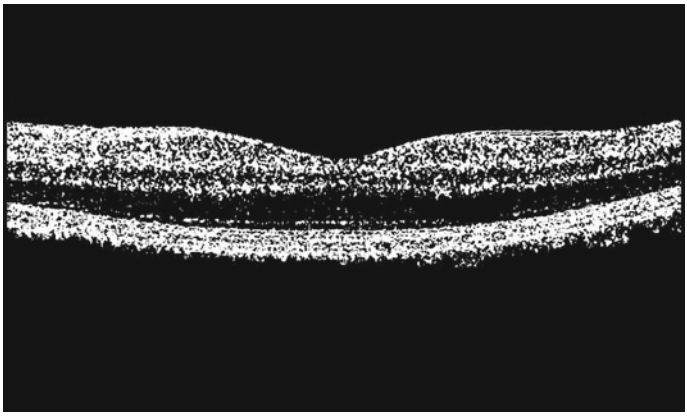


Fig. 5 The healthy retina's cross-section OCT image after thresholding and noise removing

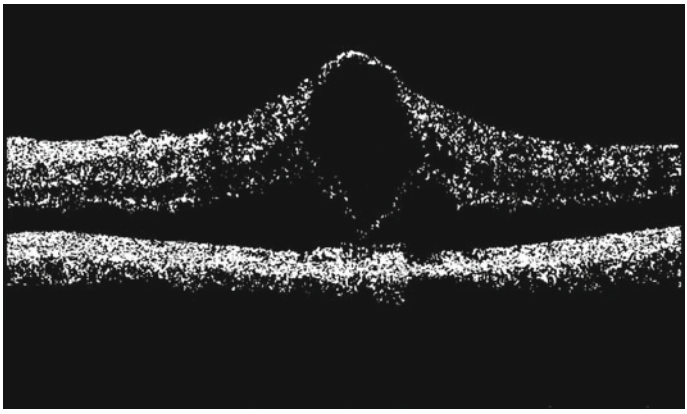


Fig. 6 The pathologically changed retina's cross-section OCT image after thresholding and noise removing

The results of image thresholding and noise removing by erosion were shown in Figs. 5 and 6.

To ensure the best possible result achieved at this stage of the algorithm, it is necessary to precisely analyze the influence of the various structural elements (in terms of their size and shape), which is used during morphological operation sequence applications. The MATLAB environment allows to test nine types of structural elements: diamond (with defined radius), disk (with defined radius), line (with defined length and angle), octagon (with defined distance from the structuring element origin to the octagon's size, along to the vertical and horizontal axis), pair (the structure with two elements), periodic line, rectangle (with specified two-piece vector defining its size), square (with specified width) and even arbitrary structural element (shape defined by user). Initially, while choosing the optimal structural element, the *line* and

periodic line elements were rejected according to their shape: both are vectors with the spatial orientation angle as an input parameter: applying a horizontal or vertical vector would lead to a fast connecting of the retina's structures and thickening of the vascular layers, which in case of edema's diseases has also a diagnostic value. On the other hand, applying diagonal vectors would lead to obtaining sharpened objects borders and, therefore, would make it impossible in further analysis to use shape coefficients to find and identify ellipses. Due to keeping the right edge course, the element's size should not be too high. *Disk* with radius equals to 3 has 25 elements (5×5 matrix) and looks exactly the same as *square* element with width equal to 5 and gives same unsatisfying results, that is why they were excluded from the further analysis. During *square* and *rectangular* elements examination it was observed that increasing their size by even one pixel can cause unwanted blurring of the edges. Although, after decreasing its size a *rectangular* element was used as very effective for removing a single pixel (or small group of pixels) within the image. Morphological operators are also useful in case of image segmentation; specially using dilation which causes the growth of the white area defining the boundaries between the retina's structures and in a consequence their better separation. To ensure that the components of the image are properly distributed, structural elements being matrices of 5×5 pixels were applied. Result of these operations were shown in the Fig. 7.

The last, but also the most important step is to calculate a shape coefficient W for each structure. It describes the relationship between object circuit L to its surface S and helps to distinguish the objects. The main advantage of using these parameters to describe objects is their lack of sensitivity for rotation and scaling. Higher values are noted for structures with elongated shape, Coefficients for basic shapes and dependence between L and S were shown on the Fig. 8.

Higher values are noted for structures with elongated shape. According to the fact that we can predict the elliptical shape of cysts, we can also predict their shape

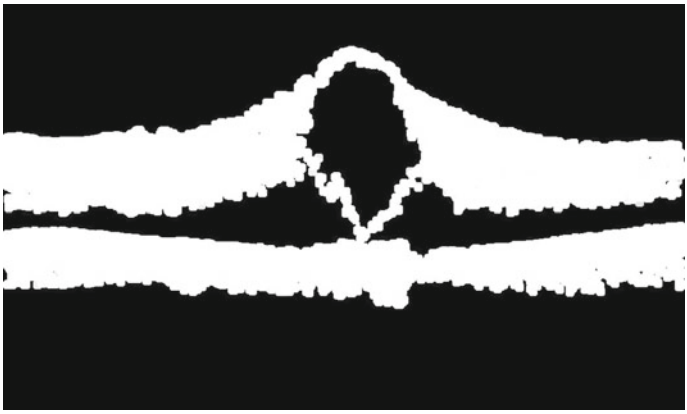


Fig. 7 The pathologically changed retina (cross-section) after morphological filtration for image segmentation

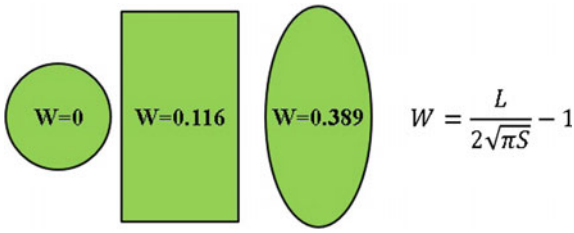


Fig. 8 Shape coefficients for basic geometric figures

coefficients value as approximately equal to 0.39. As a consequence, it is possible to find in the image suspicious objects by calculating their W value and compare them with the theoretical value.

The simplest method, also used in this paper, is to calculate the L value as the total number of contour pixels (white pixels having at least one black pixel in their neighborhood), and the surface S as the sum of all white pixels being a part of the examined object.

Three objects which might be seen in Fig. 7 were labeled and their W values were enumerate. The object which W value were the nearest to 0.39 should be chosen as a potential pathological tissue. In this paper, the nearest W value, equal to 0.31 was received for object with label 2.

The next step is to check the object localisation in the image. According to Fig. 4, the cyst should cover the macula, which is why it must be localized in the central part of the image. An effective method to verify the nature of an object suspected of being a change is to verify the macula position in the image of a healthy eye (determining its center), and then designate the center of the object in the image of the diseased retina. If the obtained value coincides within the limit's established uncertainty (in this work abroad these adopted 20 pixels), then we can assume that the test object is indeed a cyst tissue covering the macula. In the OCT images, the macula is localized in the same place as the fovea (a small depression in the retina of the eye where visual acuity is highest. The center of the field of vision is focused in this region, where retinal cones are particularly concentrated). Finding the location of the macula is based on finding a cavity in the upper layer of the retina (the lowest point belonging to the retina's top border). Its position along the X axis designates the macula center as it is shown in Fig. 9.

After thresholding the input image is a monochrome bitmap (matrix where white pixels are ones, and black-zeroes). The implemented algorithm works on a T vector of equal length to the horizontal size of an image (width) in pixels. At the beginning T vector is filled with zeroes. Then, gradually vector T is filled with ones by iterative operation of binary OR with consecutive rows of an image: T = T and Vi, where i is the number of iteration, starting from 0. Algorithm is terminated when there is only one 0 value in the T vector on an j position. In the effect (j, i) coordinates are found that point the position of the macula on the image.

The resulting value of the position of the macula (487) is compared within the accepted uncertainty (40 pixels) with a calculated means of elliptical structures in the

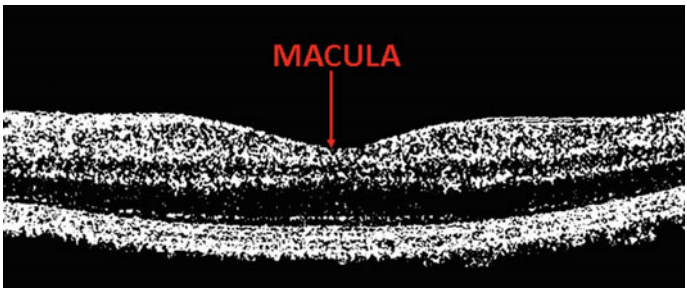


Fig. 9 Macula localisation in healthy retina's cross-section

image of the diseased retina (519). Comparing these two values taking into account the accepted uncertainty, the central location of the elliptical tissue (and its pathological nature) were confirmed.

2.2 Branch Retinal Vein Occlusion (BRVO)

Branch retinal vein occlusion (BRVO) is a non-curable, common (16 millions cases in the world) retinal vascular disease appearing in the elderly (60–70 years old patients). Mainly it is caused by the occlusion of one of the branches of central retinal vein branches. It is characterized by a strong distortion of the retina (the course of the optic nerve) without distorting the photoreceptor layer (Fig. 10 [13]). It is possible to observe the presence of both intraretinal cysts and several edemas filled with subretinal fluid.

All algorithm steps were applied in a similar order like in the cystoid macular edema case. One elliptical change located in the center of the image was found (the shape coefficient equals to 0.326). On the other hand, its location is not consistent with the location of the macula in the assumed margin of error (40 pixels). The edema

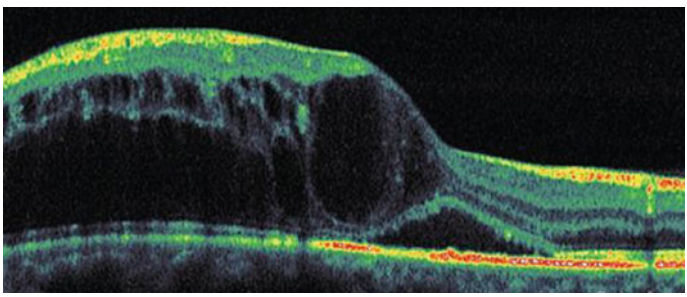


Fig. 10 The pathologically changed retina (cross-section) in case of branch retinal vein occlusion disease

Fig. 11 The pathologically changed retina (cross-section) in branch retinal vein occlusion disease after morphological filtration

covers the macula, but it is shifted slightly to the left part of the image. Probably it is caused by the individual course of BRVO disease. It is strongly needed to examine the algorithm's efficiency in case of other patients with this kind of retina's distortion. Furthermore, in the case of lesser cysts the shape analysis did not give satisfactory results. The main cause is the selection of structural element during the morphological operations (retina's segmentation). Both difficulties will be the subject of further examination. The final results of the algorithm's application were shown in Fig. 11.

Figures 12 [15] and 13 show the analysis of another BRVO case. During visuals it was easy to observe a single subretinal change with no intraretinal distortions. The course of the optic nerve was only slightly disturbed in the central part of the image; the retina's shape is not distorted. Already at the stage image thresholding, a high degree of the tissue's segmentation was obtained. As a consequence, after noise removing, morphological operations in order to improve the degree of separation of objects were not applied. After assigning labels to the separated objects, many small

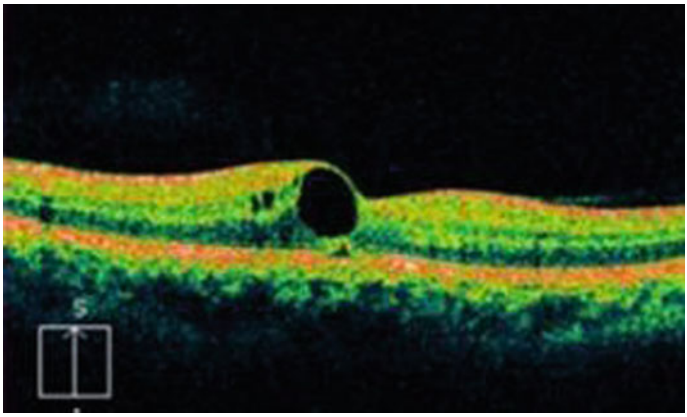


Fig. 12 The pathologically changed retina (cross-section) in BRVO disease

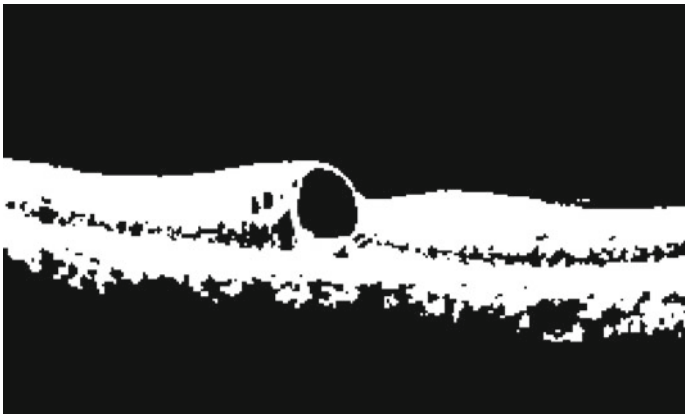


Fig. 13 The pathologically changed retina (cross-section) in BRVO (case 2) disease after morphological filtration

and isolated groups of pixels were detected although only one of them (the biggest one) has proper shape coefficient (0.341) and the location (507) to fulfill criteria to be recognized as an edema.

2.3 Idiopathic Polypoidal Choroidal Vasculopathy (IPCV)

Idiopathic polypoidal choroidal vasculopathy (IPCV) is a disease entity characterized by vascular changes within the retina of intense edema in its central part. An example of the retina of a person suffering from this condition was shown in Fig. 14 [13]. It is possible to distinguish three significant cysts and a large one filled with intraretinal fluid. The biggest one is localized in the central part of an image, covering the macula. In addition below the photoreceptors layer (red line along the image)

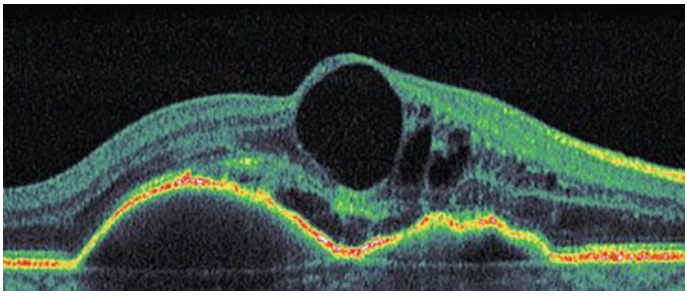


Fig. 14 The pathologically changed retina (cross-section) in case of IPCV disease

Fig. 15 The pathologically changed retina (cross-section) in IPCV disease after morphological filtration

there is a big choroidal vessel, strongly distorted by subretinal fluid accommodation. Further analysis shows that the whole retina's shape is distorted in comparison with the healthy retina (Fig. 3 [12]): both nerve fiber layer (upper edge of the retina) and a line of photoreceptors. However, the shape of the retina should not affect the algorithm's results.

The application of the algorithm gave the expected positive results: confirmation of the elliptical shape changes in the image as its location characteristic of angioedema in the course of the IPCV disease (for accepted margin of uncertainty of 40 pixels). The only aspect that requires improving may be the fact that during the image processing two minor changes were combined into one (in the thresholding stage), which distorted the analysis of their shape coefficients. The final image processing results were shown in Fig. 15.

3 Summary and Conclusions

In biomedical applications, specially in ophthalmology, OCT is a very attractive diagnostic method with unique properties: it allows imaging of the tissue's morphology with a much higher resolution than other imaging techniques, such as MRI or ultrasound (OCT resolution might be even higher than 0.01 mm). The main advantages of the OCT method are

- subsurface images with microscopic resolution,
- fast, accurate imaging of tissue building,
- no need prior sample preparation,
- lack of dangerous ionizing radiation (the procedure can be repeated many times and performed in patients of all ages and pregnant women as well)

In all analyzed cases developed methodology gives satisfactory results. The intended goal was achieved. Preliminary retina image processing, segmentation, and shape analysis of the individual structures helped to find potentially pathological change. Using information about the expected position of cysts in the OCT image

Table 1 The results of the algorithm for selected edematous retinal disorders

Retina's disorder	Shape coefficient	Edema's location
CME (diabetes)	0.319	519
BRVO	0.326	613
BRVO (case 2)	0.341	507
IPCV	0.334	524
Health eye	–	487

(overlaying cysts on the macula) made it possible to check whether it is consistent with the expected location of the edema. The confirmation of this thesis has allowed to evaluate this method, detecting pathological changes in the retina as effective. The achieved results are summarized in Table 1.

With no doubt, the algorithm requires further work and improvement in order to identify other, less common lesions of the retina and a more precise location of the macula in the retina(which will reduce the expected margin of error).

Additionally, the presence and the appearance of cysts and edema of the retina is individual and depends on the course of disease in a particular patient, consequently, in some cases difficult to predict, and automatically extract the unique pattern of the disease. Figures 16 [13] and 17 [13] show two diabetes cases with significant and abnormal retina deformation. In the first case, a single change has non-elliptical shape (and thus impossible for identification by searching ellipse in the image). In the second one, there are many various small, elliptical changes, with mutually comparable size (lack of parent change). These difficulties mean that the proposed algorithm

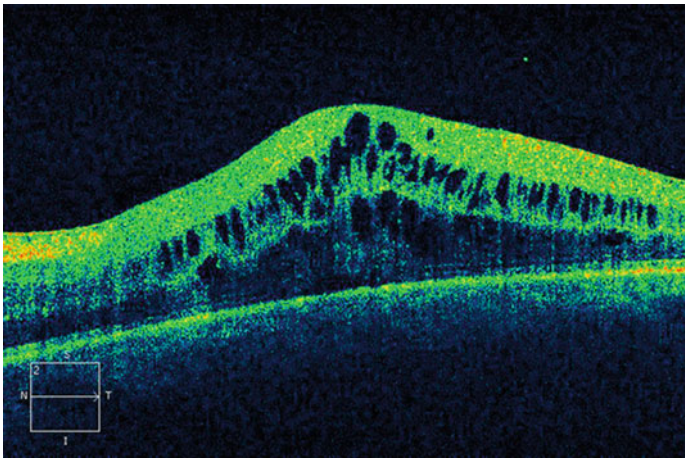


Fig. 16 Retina of the person with diabetes diagnosis (case 2). Numerous, small, subretinal changes with *elongated shape* without a dominant change in the *center* of the picture where macula should be localized

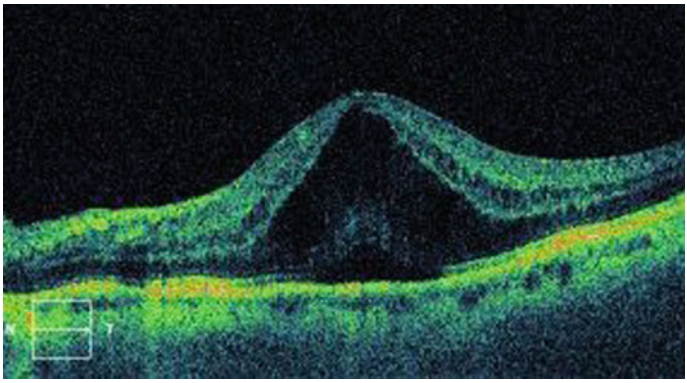


Fig. 17 Retina of the person with diabetes diagnosis (case 3). Vascular single change in the position of potentially coinciding with the location of the macula (the stage of visual evaluation) of an unusual *triangular shape*. This shape causes difficulties in the operation of the algorithm seeking change of *elliptical shape*

in its current form requires additional visual assessment and adaptation based on the larger number of clinical cases.

It should also be noted that in its present form, the algorithm does not distinguish disease entities but only confirms the presence of edema. To distinguish between disease entities it should be discussed to expand the proposed method of analysis module number of changes, shift changes in the parent company, and change the shape of the retina itself, taking into account the course of the optic nerve and the photoreceptor layer. These challenges are related to, but inseparable from the assessment of the effectiveness of the algorithm for a large database of images diseases ophthalmological, which will be the subject of further work. To sum up, the most important challenges and difficulties associated with the development of the method are

- the correlation between the appearance of the retina and the individual course of disease in the case of each patient,
- the inability to distinguish discussed diseases (although detecting edemas is very effective),
- expanding the possibilities of the algorithm for detection of retinal diseases other than vascular abnormalities like holes, the retina's geometric distortions or AMD.

These improvements will be the subject of further research.

Acknowledgements The work was financed (co-financed) by the Polish Ministry of Science and Higher Education (MNiSW).

References

1. Fercher, A.F.: Ophthalmic interferometry. In: von Bally, G., Khanna, S. (eds.) Proceedings of the International Conference on Optics in Life Sciences, pp. 221–228, Aug 1990
2. Tanno, N., Ichikawa, T., Saeki, A.: Lightwave reflection measurement. Japanese Patent No. 2010042 (1990)
3. Chiba, S., Tanno, N.: Backscattering optical heterodyne tomography. In: Prepared for the 14th Laser Sensing Symposium (1991)
4. Bouma, B.E., Tearney, G.J.: Handbook of Optical Coherence Tomography. Marcel Dekker Inc., New York, Basel (2002)
5. Bourquin, S., Seitz, P., Salathe R.P.: Optical coherence topography based on a two-dimensional smart detector array. Opt. Lett. **26**(8), 512–514 (2001)
6. Yeow, J.T.W., Yang, V.X.D., Chahwan, A., Gordon.: Micromachined 2-D Scanner for 3-D optical coherence tomography. Sensors Actuators **117**(2) (2005)
7. Drexler, W., Morgner, U., Ghanta, R.K., Kartner, F.X.: Ultrahigh-resolution ophthalmic optical coherence tomography. Nat. Med. **7**(4), 502–507 (2001)
8. Adhi, M., Duker, J.S.: Optical coherence tomography—current and future applications. Curr. Opin. Ophtamol. **24**(3), 213–221 (2013)
9. Bezerra, H.G., Costa, M.A., Guagliumi, G., Rollins, A.M.: Intracoronary optical coherence tomography: a comprehensive review clinical and research applications. JACC Cardiovasc. Interv. **2**(11), 1035–1046 (2009)
10. Zysk, A.M., guyen, F.T., Oldenburg, A.L., Marks, D.L.: Optical coherence tomography: a review of clinical development from bench to bedside. J. Biomed. Opt. **5**(12) (2007)
11. Fercher, A.F., Hitzenberger, C.K., Drexler, W., Kamp, G., Sattmann, H.: In Vivo optical coherence tomography. Am. J. Ophthalmol. **116**, 113–114 (1993)
12. http://www.eyesitetexas.com. Accessed Feb 2016
13. http://www.optos.com. Accessed Feb 2016
14. Podoleanu, A.: Optical Coherence Tomography. http://www.birpublications.org/doi/ref/10.1259/bjr/55735832. Accessed Feb 2016
15. Kharousi, N.A., Upender, K.W., Sitara, A.: Current applications of optical coherence tomography in ophthalmology, Chapter 1. In: INTECH (2013). doi:10.5772/53961. ISBN 978-953-51-1032-3

Part III
Computational Physics and Applied Mathematics

Study of *R*-Factors Used in Structure Determination by Use of Genetic Algorithms from Powder Diffraction Data Consisting of a Small Number of Very Broad Peaks

T. Kozik and W. Łużny

Abstract Genetic algorithms (GAs) are an alternative to local optimization procedures in structure refinement. As a nondeterministic method, they may possibly yield better results. Some general principles of the Rietveld refinement procedure have been adapted to use with GAs, including quantities which enable the algorithm to determine how well a model diffraction pattern fits the experimental curve, named fit factors or *R*-factors. The search for the structure of the crystalline regions of the PANI/CSA conducting polymer system is specific in that the crystalline component used for structure determination is very different from typical diffraction patterns. It belongs to a class of diffraction curves exhibiting a small number of very broad peaks. As this introduces ambiguity not present in the formulation of the problem for the original Rietveld method, it should be investigated whether typical *R*-factors may still be used. It is determined that a simple sum of squares based factor and one of the original Rietveld factors R_f may still be useful, but using R_{wp} may lead to qualitatively poor results. Some new, custom *R*-factors are introduced and investigated for performance, along with the original ones. Advanced formulas combining different aspects of curve fitting appear to be the best choices for fit factors while still remaining computationally efficient.

Keywords Genetic algorithms · Computer modeling · Crystalline structure determination · Fit factors · Rietveld factors

T. Kozik (✉) · W. Łużny
Faculty of Physics and Applied Computer Science, AGH University of Science and Technology Krakow, Kraków, Poland
e-mail: Tomasz.Kozik@fis.agh.edu.pl

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*, Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_12

1 Introduction

The Rietveld refinement procedure [1] is commonly used in crystallography for structure determination from X-ray powder diffraction data. For a model structure described using a crystallographic unit cell including a collection of atoms with defined coordinates within the cell and with a given symmetry, the positions of all crystallographic reflexes can be calculated. An intermediate step is taking into account different form factors for each atom within the cell and calculating the structure factor. The method describes how to transform discrete reflexes into diffraction peaks of a nonzero width using profile functions, obtaining in the end a calculated powder diffraction pattern.

To modify the initial model and continue obtaining better and better models, which in the end should become a resemblance of the actual crystal structure, the procedure takes advantage of typical methods used for local optimization like the steepest descent method. Before this is possible, a means of grading a model diffractogram versus the target diffraction curve must be introduced. What is needed is a quantity which describes how similar the pattern obtained for the model is to the experimental one. While a simple sum of squared differences between the two patterns calculated for each value of the scattering angle would be such a quantity, the Rietveld method defines other such quantities, which may be better. These quantities, named fit factors or simply *R*-factors, have the common property of minimizing their values for two identical curves. The local optimization method used in the algorithm is given the goal of minimizing the value of the chosen fit factor, which means that ideally the two diffraction patterns match and the model structure is exactly the same as the actual one.

The final model obtained using the procedure is actually a local minimum of the *R*-factor used with the given target pattern. It cannot be guaranteed that the method will converge to the global minimum, which means that the obtained model structure may not match the physical one as well as possible, but may only be its approximation. However, such a structure may be good enough to be accepted as a model of reality.

It is important to note that Rietveld refinement requires a starting model to refine. This means that if this initial structure is already reasonably good, the method will most probably successfully find the best possible one. However, if the initial model is not good enough, the procedure may converge to a local minimum of the fit factor which is far from the global one and is a model not acceptable. With a poor starting structure it may be impossible to find a reasonably good solution at all. In some cases, construction of the initial model is very difficult. This is why before Rietveld refinement can be used, extensive research of the studied structure must usually be conducted first. Despite the above potential difficulties, the method has a history of successful applications.

However, there exist global optimization procedures. A genetic algorithm (GA) is a type of nondeterministic optimization procedure with many applications

in various fields. One of them is precisely structure determination from diffraction data [2].

The idea behind a GA is to imitate the natural phenomenon of evolution in the biological world. It requires defining a collection of parameters describing a single model. Such a collection is named a phenotype, and each model is named an individual. The phenotypes also have encoded forms, more convenient to process by the algorithm, named genotypes. The procedure begins with not one initial model, but several of them, named a population as a whole. It can be even entirely random, so there is no need to define a starting structure at all (though the scope of the model is a limitation). In each iteration, also called a generation, the GA processes the population, gradually improving the quality of the individuals. Just like in the biological world, where individuals which are better adapted to the environment have a higher probability of surviving and reproducing, individuals which are better solutions to the problem being solved by the GA have a higher probability of taking part in the creation of individuals forming the next generation. Once a (somewhat artificial) stop condition is met, the best individual of the last population is the solution found by the GA.

The evolution in the GA imitates only chosen properties of the natural phenomena it is based on. However, this appears to be sufficient, and genetic algorithms are widely used for various optimization tasks. To apply a GA in structure determination, some of the principles taken from the Rietveld method need not to be altered at all, like the structure factor calculations and profile functions. Along with these methods, the original *R*-factors were adapted for use with GAs as well.

While Rietveld refinement is usually successful in the case of typical crystal structures and there is no need to abandon it in this scope, the structure of molecular crystals may be complicated enough to pose a problem for this method. A computer program named CrystalFinder is described in [3], showing that a properly designed genetic algorithm is capable of finding the structure of several known molecular crystals basing only on their X-ray diffraction pattern. There are also reported cases of genetic algorithms finding models of molecular structures which were previously unknown [4].

What is not stressed enough is that the above applies in full to problems with target diffraction curves consisting of a fairly large number of narrow peaks. There is little ambiguity in such diffraction patterns—the peaks of the curves are usually well defined, and as long as care was taken when subtracting the background, the same applies to the valleys between peaks. The number of peaks is large enough to provide optimization procedures with enough data to fit the values of many model parameters.

This is very different in the case of polymer science. X-ray diffraction experiments may be performed on polymer samples, resulting in a diffraction curve which is difficult to analyze. Such a pattern usually consists of three components—a background of high intensity, an amorphous component consisting of a few extremely broad peaks, and a crystalline component. If one were to attempt to

investigate the structure of the crystalline phase using methods typical for the analysis of usual crystals, separating the crystalline component from the other two is required. This may be a difficult and ambiguous task itself. The resulting crystalline component may be treated like a powder diffraction pattern in further investigation, because as long as there is no texture, crystalline areas of the polymer are randomly oriented within the sample, nested in the amorphous phase (most polymer samples have such a structure). However, before methods described earlier in this introduction may be applied, it is important to note that the prepared diffraction curve is very special.

The crystalline component is different from a typical powder diffraction pattern in that it consists of only a few peaks. What is more, these peaks are very broad, potentially spanning across positions of several crystalline reflexes. Thus it is not even clear which peaks correspond to individual reflexes and which are composed of more than one. It must be stressed that this broadening cannot be eliminated by improving scientific apparatus, because it is primarily caused by the physical nature of the sample. Namely, the crystalline areas within the sample, although often accounting for a major fraction of the sample volume, are individually very small.

The above-described ambiguity is the source of problems when attempting to determine the structure of the crystalline areas of polymers. It is not surprising that Rietveld methods usually fail in finding a model of these areas, often leading to local minimums of the fit factors. Such models correspond to diffraction curves which are not even qualitatively good. It is reasonable to expect that genetic algorithms may be able to overcome the obstacles and find a good model.

However, it should be pointed out that at this point one of the principles adapted from Rietveld refinement for genetic algorithms are still the original *R*-factors or the very simple sum of squared residuals. It is reasonable to question if they are still useful given how much the task, namely the target diffraction pattern, has changed from the original idea. Investigating this is the aim of this paper.

The PANI/CSA conducting polymer system is an example of a case in which the structure of the crystalline areas remains unknown. Emeraldine, which is a form of polyaniline, may transit into a conducting state (PANI) by protonation using acids. Camphorsulphonic acid (CSA) is the acid used in this case. In 1997 [5], the first models of the structure of the crystalline areas of this system were published. However, until this day no model has been widely accepted.

While this system is being investigated using molecular dynamics simulations [6], a specialized program capable of searching for this unknown structure, basing only on chemical knowledge and the crystalline component of the diffraction pattern, was also created. It was thoroughly described and tested for fictional target input in [7]. The software described there is applied to the actual experimental crystalline component with the purpose of investigating the performance of various *R*-factors. A description of this investigation is provided in this paper.

2 Basic *R*-Factors

Fit factors are used as quantities describing how similar a model diffraction pattern is to a target one. Such a factor must minimize its value for a perfect fit of the two curves. Any formula with this property may, in principle, be used as an *R*-factor. However, not all such formulas may be equally good for the given task.

A diffraction curve is in practice a collection of points. Each of the k points is given by two values—the intensity I_k and scattering angle value $2\theta_k$. The most basic and intuitive formula for a fit factor that can be given is a simple sum over k of the squared differences between the observed and calculated intensities, I_k^{obs} and I_k^{calc} respectively, at each matching $2\theta_k$ for the two collections of points. In the study, a square root of such a sum was considered as one of the initially used *R*-factors and named *RSSR*:

$$RSSR = \sqrt{\sum_k \left(I_k^{obs} - I_k^{calc}\right)^2} \tag{1}$$

Two modifications of such an obvious formula exist and are introduced in the original Rietveld method:

$$R_f = \frac{\sum_k \left|\sqrt{I_k^{obs}} - \sqrt{I_k^{obs}}\right|}{\sum_k \sqrt{I_k^{obs}}} \tag{2}$$

$$R_{wp} = \sqrt{\frac{\sum_k w_k \left(I_k^{obs} - I_k^{calc}\right)^2}{\sum_k w_k \left(I_k^{obs}\right)^2}} \tag{3}$$

The most commonly used one appears to be R_{wp}, with the introduced weights w_k. Interestingly, exact formulas for the used weights are very seldom disclosed in papers describing the usage of this factor. The values for the weights are either assigned by taking into account statistical uncertainties for the intensities (in one way or another), or are given by the formula:

$$w_k = \frac{1}{I_k^{obs}} \tag{4}$$

This latter convention was chosen for investigation, as the former may be difficult to apply for the studied case of polymer crystalline components, which are superimposed on a background and an amorphous component of the overall diffraction pattern, and means of separation are often qualitative. What is more, this

convention makes it possible to study a fit factor much different in behavior than the other two introduced so far.

3 New *R*-Factors Designed

Preliminary results obtained using *R*-factors described so far were not as good as one would wish for them to be. Details of this are given in Sect. 7. What is more, the software described in [7] was prepared with the possibility of launching several genetic algorithms one after another with some different parameters, among them the type of *R*-factor used. Designing at that point more fit factors for the sake of supplying the software with more tasks to complete and for the purpose of attempting to find formulas which would lead to better results was a natural course of action.

The Pearson product-moment correlation coefficient is a very general quantity for measuring the linear dependency or correlation between samples of two variables x and y, each sample of the same size. It is given by the following formula:

$$\rho = \frac{\sum_k (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_k (x_k - \bar{x})^2}\sqrt{\sum_k (y_k - \bar{y})^2}} \tag{5}$$

This quantity has the property of taking its maximal value, 1, for total positive correlation, and minimal value, −1, for total negative correlation. This formula can be applied for curve fitting if the intensities of the curves are introduced as samples to apply the coefficient to. While the original formula maximizes its value for a perfect fit, a simple conversion makes it ready to use as an *R*-factor, further named R_1:

$$R_1 = 1 - \rho \tag{6}$$

The above formula takes the value of 0 for a perfect fit that is a total positive correlation. Total negative correlation is not an interesting case at all from this point of view (and also unreachable in this problem), but takes a value of 2.

The second initial idea was applying the most simple *RSSR* factor for calculations again, this time not directly to the two curves, but to their first derivatives. Since the diffraction patterns are given as sets of points, an approximation of the value of the derivative at each point must be made. The decision to simply approximate it with a central difference was made, which means that for each point:

$$I_k' = \left.\frac{dI}{d(2\theta)}\right|_k \approx \frac{I_{k+1} - I_{k-1}}{2\theta_{k+1} - 2\theta_{k-1}} \tag{7}$$

As a result, the new *R*-factor named R_4 or $RSSR'$ was introduced:

$$R_4 \equiv RSSR' = \sqrt{\sum_k \left(\left(I_k^{obs}\right)' - \left(I_k^{calc}\right)' \right)^2} \tag{8}$$

4 Elaborating on the *R*-Factors

Preliminary calculations using these new *R*-factors showed that they are not quantities obviously outperforming the original fit factors. Some details concerning this issue are given in Sect. 7.

However, they appeared to be partially working in that the final model diffraction patterns obtained using them exhibit peaks properly localized more or less in the place of the experimental peaks. The flaws of the model patterns were that the intensities were usually too low. The interpretation of this regularity was that the designed formulas are lacking a term which would allow the algorithm to distinguish between better and worse individuals among those with diffraction patterns correlated to some extent with the target one. At the same time, a term which would allow individuals with very intense but slightly misplaced peaks to reproduce would be useful. Misplacement of peaks contributes to an increase in the fit factor value, while potentially tied to valuable genetic information.

A fairly simple idea is used—implementing two new *R*-factors which are products of two terms. In both cases the first term is one of the previously designed factors, and the other is the *RSSR* factor:

$$R_2 = R_1 \cdot RSSR \tag{9}$$

$$R_5 = R_4 \cdot RSSR \tag{10}$$

While the first term enables the algorithm to fit basing on correlation or derivatives, the second is a pure point-to-point fit. Hopefully such a combination will perform well—this will be covered in Sect. 7.

A variety of other fit factors were used. Many of them were based on the *R*-factors described so far, but with additional weights for different scattering angle values. These were intended to help fit diffraction peaks which were difficult to obtain by the algorithm. They were even sometimes not present in the solutions at all. Practice showed that none of them improved the results. One final fit factor which usually enables the algorithm to reach interesting models, good enough to still be in use, will be covered in this paper:

$$R_3 = \sum_k \exp\left(m \frac{\left(I_k^{obs} - I_k^{obs}\right)^2}{I_{\max}^2} \right) \tag{11}$$

In the above formula I_{max} is the highest possible intensity value (in the convention introduced in [7] this is 100 arbitrary units—a.u.), and m is a parameter. The factor is used with $m = 5$ so far.

5 Scope of the Modeling

The crystallographic unit cell which may possibly be used to model the structure of the PANI/CSA crystalline areas is proposed as a triclinic one with inversion symmetry. The triclinic unit cell angles α, β, and γ are enough to describe the parallelepiped of the unit cell, because its edge lengths a, b, and c may be calculated from those three angles by taking into account the assumption of inter planar distance values d100 = 19.362514 Å, d020 = 3.5 Å, and d001 = 9.373439.

Within the unit cell, two polyaniline (PANI) dimers and two CSA counter ions of opposite chirality are contained. It is enough to describe only the layout of half of those contents, since the other will be calculated by applying inversion symmetry.

What must be given is the offset along the unit cell axes a and b at which the polymer chain enters the cell a_{off} and b_{off} (by convention, the chain is modeled from one of the nitrogen atoms), the angle of rotation of the polymer chain by its axis φ_{PANI} and the torsion angle of the aromatic rings τ_{ring}. The CSA ion is described by its absolute coordinates within the cell and by angles describing its orientation. It is assumed (though this property can be changed in the software) that the CSA ions are rigid, which is reasonable. More parameters are available in the software, but only these are used in the genetic algorithm described in this paper.

It must be also mentioned that the length of the polymer chain dimer is tied to the cell constant c. This means that as the cell constant varies, the C–N–C angle within the dimer varies accordingly, ensuring that the polymer chain continues through consecutive cells along the c direction (the polymer chain axis is parallel to the c edge). This means that as the c constant decreases, the polymer chain is being compressed to fit into the cell, although without changing the covalent bond lengths. This also means that there is an upper limit to the scope of the model, in which the polymer chain has C–N–C angles of 180°. It cannot expand any further without stretching the covalent bonds. This means that the model breaks for larger values of c and this must be avoided in the genetic algorithm. Of course, it is not possible for such a situation to occur in reality.

6 Genetic Algorithm Used

To make a summary of the properties of different factors observed throughout hundreds of launches of the genetic algorithm, we prepared a series of launches using the prepared software [7], described below.

A total of eight launches was prepared, each different from one another in the type of *R*-factor used. Each of the factors described in detail in Sects. 2, 3 and 4 were used.

The common parameters of the algorithms were as follows. Each algorithm processed 200 individuals for 500 generations. The type of encoding used was real number encoding. The initial genotype values for the 200 individuals in the first generation were (pseudo)randomly generated, within the ranges of variation summarized in Table 1.

The profile width for each reflex was given by a constant formula:

$$FWHM = \sqrt{0.02 \cdot \tan^2\theta - 0.01 \cdot \tan\theta + 2.0}, \tag{12}$$

where θ is half of the scattering angle value for the reflex. Reflexes were profiled using a Gaussian function.

For calculating the fitness of each individual based on *R*-factor value, the procedure of scaling was used, in which for each generation the maximal and minimal value of the fit factor in the population was first found, after which the quantity:

$$r_k = \frac{R_k - R_{\min}}{R_{\max} - R_{\min}} \tag{13}$$

was calculated for each individual. Based on this, fitness was calculated using the exponential fitness function (one of the three implemented in the used software)

$$F_k = \exp(-r_k) \tag{14}$$

Using such a formula means that in each generation the best fit individual has a fitness equal to 1, and the worst equal to $1/e$. For selection, the common roulette method was used, using the fitness values to scale the roulette wheel.

Table 1 Ranges of variation for the GA parameters

Parameter	Range of variation
α	72°–108°
β	72°–108°
γ	72°–108°
a_{off}	0–5 Å
b_{off}	0–3.5 Å
φ_{PANI}	0°–360°
τ_{ring}	−30°–30°
a_{CSA}	0–10 Å
b_{CSA}	0–7 Å
c_{CSA}	0–7 Å
φ_{CSA}	0°–360°
θ_{CSA}	0°–360°
τ_{CSA}	0°–360°

Regarding genetic operations, averaging crossover between two individuals with the probability of occurring 70% for each value encoded in the genotype was performed. This means that for each corresponding values in two genotypes assigned to crossover a (pseudo)random number in the range from 0 to 1 was generated, multiplied by the difference between the two values and used to alter each of the values by addition or subtraction accordingly. For mutations, for each value encoded in the genotype, a variable from the Cauchy distribution with the width parameter gamma equal to 0.05 was generated, multiplied by half of the variation range of the genotype value and added to the value, as long as this operation did not produce a value outside of the variation range. This is called a perturbation of the genotype values. The Cauchy distribution is much better than a Gaussian distribution in this case because of the non-negligible probability of obtaining large numbers from the distribution, which enables the sporadic occurrence of long range mutations, crucial for the ability of the algorithm to explore the space of possible solutions to the problem.

For succession, that is forming a population for the next generation, the elite succession method was used with elite size equal to one. This means that the best individual from each generation is guaranteed to proceed to the next generation (although it loses this privilege if it is no longer the best one in the next generation). This prevents recession in the course of the algorithm, but makes the algorithm prone to getting stuck in local minimums of the *R*-factor used. However, this may be overcome by the long-range mutations enabled using the Cauchy distribution described earlier.

7 Investigation of Performance

The program keeps a log of the best *R*-factor values for each generation. Such a convergence log is useful for analysis—for example, if significant improvements were still systematically occurring shortly before the algorithm ended, it may suggest that the number of generations should be increased. On the other hand, if the algorithm spent several final generations not improving the solution, perhaps a smaller number of generations would be enough.

The convergence logs will be used to calculate the performance of each *R*-factor by calculating the relative decrease in the value between the best of the initial (pseudo)random individuals and the final one, expressed in percents:

$$r_{gain} = \frac{R_{initial}^{best} - R_{output}}{R_{initial}^{best}} \tag{15}$$

Another parameter which may be used to grade how good the *R*-factors are, is the time *t* taken by the algorithm to complete its task. An individual launch completes in under an hour in all of the cases. Although it is understandable that

differences in speed of the order of minutes are of no concern for calculations, one should formally wish for an *R*-factor as good as possible and as inexpensive (in terms of computing time) as possible.

What is more, it would be informative to note which launches yielded the best results of the series, regardless of how big of an improvement was made from the initial random value. This would require calculating the value of one chosen *R*-factor for all of the final results and comparing the results. The most obvious choice is *RSSR*, since it is closest to the pure sum of squares. However, it should be noted that a result quantitatively better in the sense of *RSSR* is not necessarily qualitatively better (what is fitted is not just any curve, it is a very special diffraction curve).

Finally, marking in which generation g_{found} was the solution generated (in other words, for how many generations the best obtained solution was not improved) is also relevant information.

A summary of the values of all the above parameters for each of the launches is provided in Table 2.

The first immediately noticeable conclusion is that the time taken by each launch of the GA is roughly the same. This means that performing the steps of the genetic algorithm takes enough computational load for the differences in computing time caused using different fit factors to be unobservable.

The worst two factors, judging by r_{gain}, appear to be R_4 and R_{wp}. It should be pointed out that in the case of the R_{wp} factor, the algorithm has converged very early and returned the worst solution of all launches. The diffraction pattern of the model found by the GA should be investigated, and is shown in Fig. 1 as an intensity (in arbitrary units scaled to 100 at the highest recorded intensity for each curve) versus scattering angle (in degrees) plot, along with the experimental curve.

One can observe that using the R_{wp} fit factor leads to qualitatively bad solutions. The model diffraction pattern in Fig. 1 is representative, meaning using R_{wp} usually leads to such curves. Except for the well fit first peak, the solution diffraction pattern hardly exhibits any peaks at all. This demands explanation.

What is at fault for this qualitative disagreement is the form for the weighting factor used. Its property is attributing a very high weight to the fit of the model curve to the experimental points located in the valleys between the peaks of the target pattern. This means that as good as this *R*-factor is in practice for fitting

Table 2 Performance parameters for the eight launches

R used	r_{gain}	*t* [min:s]	*RSSR*	g_{found}
R_f	29.23%	37:22	404.80	386
R_{wp}	18.66%	37:02	749.23	180
RSSR	32.17%	37:04	414.47	392
R_1	52.70%	37:29	463.86	391
R_2	60.10%	37:00	442.32	281
R_3	39.47%	37:07	436.13	260
R_4	7.82%	37:14	475.08	479
R_5	40.53%	37:09	400.78	488

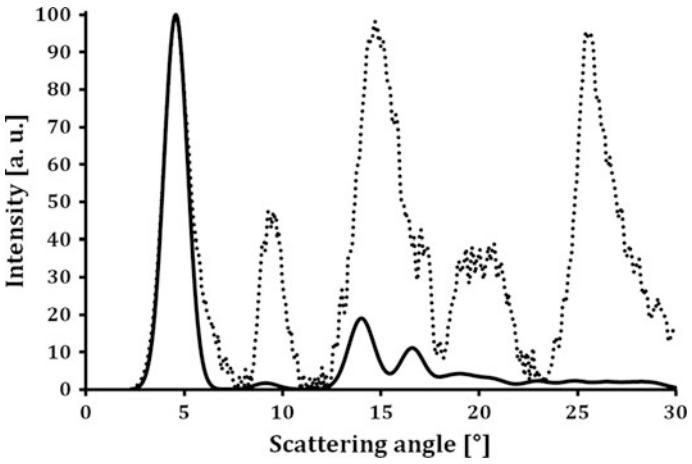


Fig. 1 Model pattern (*solid*) obtained using R_{wp} and experimental (*dotted*)

diffraction patterns exhibiting a large number of very narrow peaks, it is prone to exploitation. The models taking advantage of this feature are ones with almost entirely flat diffraction curves, which may appear in a genetic algorithm in the case of fitting diffraction patterns consisting of a small number of very broad peaks. Such models are attributed a low value of the factor due to being well fit in the valleys and despite being poorly fit to the peaks, because the peak areas are attributed a relatively low weight. This means that unless a model has perfectly positioned peaks from the start, if it has strong peaks at all that at least slightly cover a valley, it is assigned a very high fit factor value.

It should be stressed that this property is very bad for the genetic algorithm itself. It actually prevents the algorithm from searching for better solutions, since any individual with stronger peaks than the ones taking advantage of the property of the weight factor is immediately assigned a high value of the fit factor. This is counterproductive, since it is this type of individuals that introduce possibly good genetic information into the population, possibly enabling the creation of individuals acceptable as solutions later in the algorithm.

Judging by drop in the fit factor value, the best parameters at first appear to be R_1, R_2, R_3, and R_5. The highest *RSSR* value of the obtained models is calculated for the one obtained using the R_1 factor.

Initially mentioned in Sect. 4, the property of the fit factors R_1 and R_4, at this point considered worse than the rest, may be shown by presenting one of the obtained model diffraction curves. An intensity versus scattering angle plot of the diffraction pattern obtained using R_4 is shown in Fig. 2. It exhibits, though not as severely as many models obtained using this fit factor, the general property of these two factors of enabling a proper positional fit of most of the peaks, but with too low intensity. It is worth pointing out that the considered case was the only one (among the eight launches) which yielded a poorly matched first peak, which is usually very

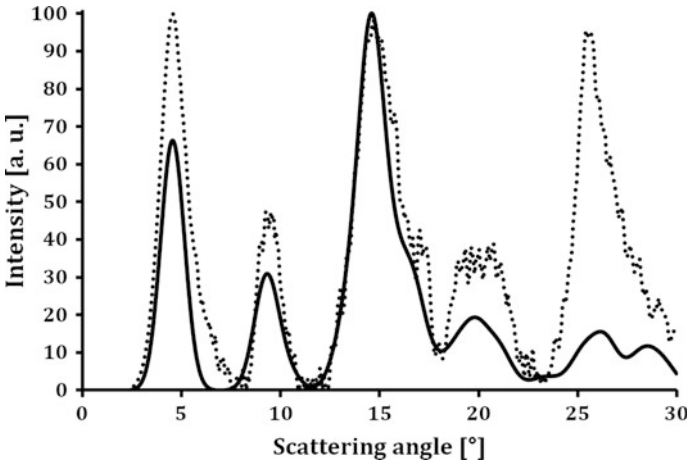


Fig. 2 Model pattern (*solid*) obtained using R_4 and experimental (*dotted*)

easy to find by the GA, as it corresponds to the (100) reflex. Interestingly, the third peak appears to be a good fit.

Viewing the diffraction patterns of the three best models obtained in the eight launches enables qualitative verification. This is a very important step—the genetic algorithm does not guarantee finding the global optimum. A local minimum found by the GA, while quantitatively good, may be qualitatively poor. None of the diffraction patterns of the three models is qualitatively poor in this case. The qualitatively best obtained pattern is the one for the model obtained using R_5. It is shown in Fig. 3 as an intensity versus scattering angle plot, on the background of the experimental curve.

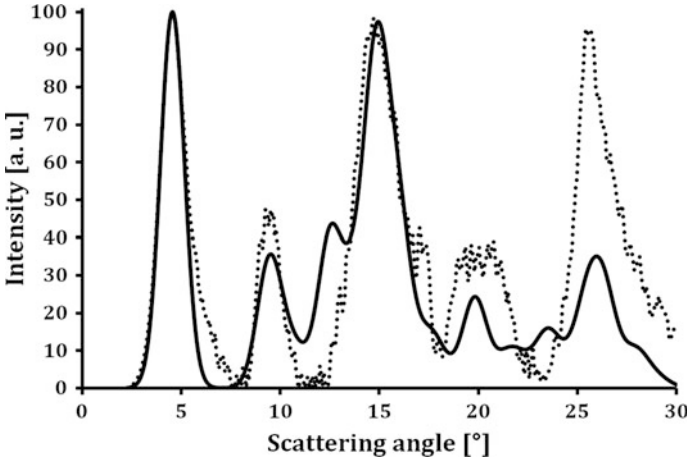


Fig. 3 Model pattern (*solid*) obtained using R_5 and experimental (*dotted*)

While the diffraction pattern is definitely not qualitatively bad, some remarks concerning it should be made. Peaks occur at the positions of all of the experimental peaks. The first peak is very well fitted. This should not be very surprising, since the properties of the model make it easy for the GA to find it here. The second peak, although seemingly perfectly positioned, lacks some intensity. The third peak appears to be well fitted. However, the asymmetry of its right slope is not reproduced by the model curve. Several peaks obviously overlapping to form the fourth peak of the experimental diffraction pattern were not properly reproduced by the GA. The fifth peak is properly positioned, and the asymmetry of its right slope is visible, but its intensity is far too low. This is a recurring problem in the modeling of the crystalline areas of the PANI/CSA system.

Interestingly, some extra peaks occur on the model curve. A peak is present on the left slope of the experimental third peak. Also a small peak is present in the valley between the fourth and fifth experimental peak.

Due to all of the above remarks, although the model is a proof that the genetic algorithm itself is working well and has the potential of finding the actual structure, this specific solution cannot be proposed as an actual model of the PANI/CSA crystalline structure.

It should be noted that the GA has no means of verifying the found structure in terms of atomic configuration, that is whether or not a definite collision between atoms, not bonded, occurs. This requires, e.g., investigating a visualization of the structure found by the algorithm. Two chosen visualizations with orthogonal projection used are shown in Figs. 4 and 5. In the first, the structure is viewed along the *c* direction of the (triclinic) unit cell, with some extra cells to show the structure better. In the second the same applies, but the structure is viewed along the *b* axis.

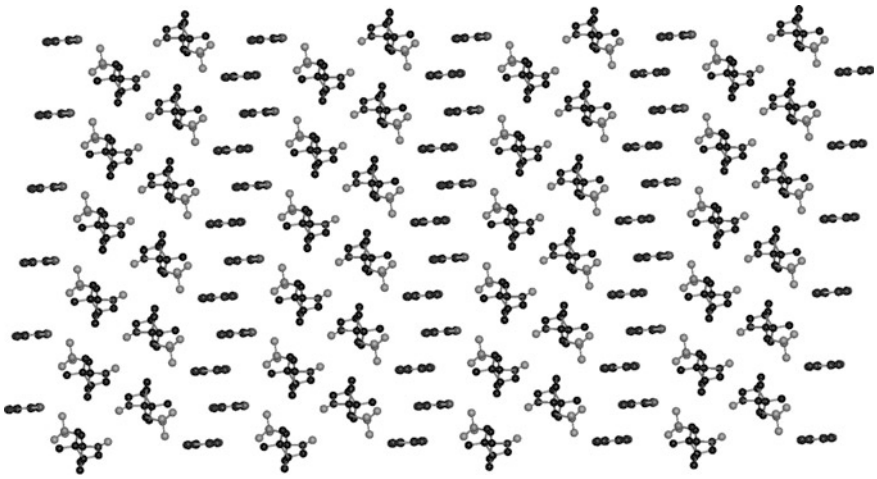


Fig. 4 The best obtained (R_5) structure in the investigation, viewed along the *c* cell axis; 4 unit cells along the *a* axis (in *horizontal plane*) and 6 unit cells along the *b* axis (in *vertical plane*) visible

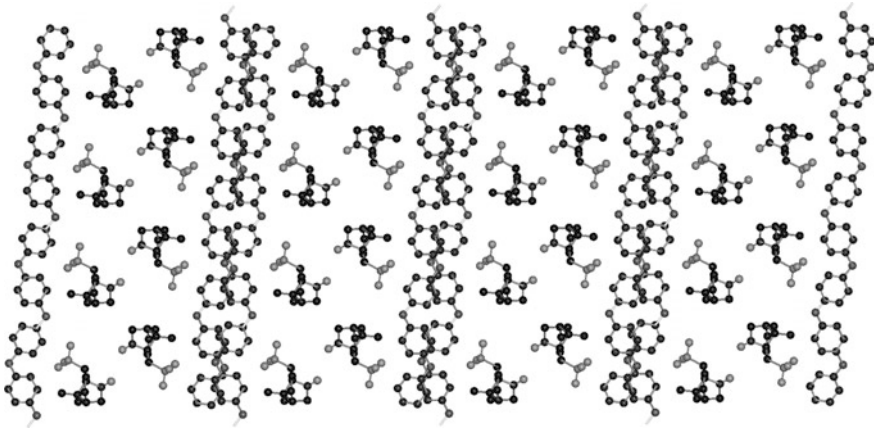


Fig. 5 The best obtained (R_5) structure in the investigation, viewed along the *b* cell axis; 4 unit cells along the *a* axis and *c* axis are visible

CSA ions and PANI chains are drawn in a distinguishable way. The view along the *a* axis is not shown as CSA counter ions and the protonated PANI chains overlap in the projection, making the image unclear and not allowing to make any observations. Visualizations were made using the software described in [7].

A very good feature of the model is that no visible collisions between atoms can be observed. The CSA ions are far away from each other and from the polymer chains. The chains also do not collide with each other. The model is designed to reproduce partially the bilayer structure postulated in [6]. One of the differences is the ordering of CSA ions along the *c* direction—here it is the same enantiomer, with different enantiomers nearest to different chains, whereas there the ordering is alternating along chain.

8 Discussion

It could be argued that performing one series of eight runs to measure the relative decrease in *R*-factor value obtained by the algorithm is a poor estimation of the performance of a factor and a much better approach would be to perform several launches and average the relative decrease. While this is a good suggestion, several hundreds of algorithms were launched throughout the search for the actual structure of the crystalline areas of the PANI/CSA polymer system by the authors. This is enough to observe enough cases of one *R*-factor outperforming others or being not useful enough to keep on including it in the search. Performance results presented in this paper could therefore be verified by the authors for agreement with observations made so far.

Another remark that could be made is that for curve fitting, one would expect much lower values of the *R*-factors used. In fact, it could even be argued that both the initial and final values of the factors used by the GAs are in all cases extremely high, and that performance analysis is due to this questionable. A counterargument would be drawing attention to the form of the experimental curve—although the main experimental peaks are clearly visible, it is very noisy. This means that even if a perfect structure was to be ever obtained within the scope of the current model, with a diffraction pattern which could be qualitatively called perfect, it would still have a high value of, e.g., the sum of squares. This does not pose a problem for the genetic algorithm as long as the perfect fit is still a local minimum of the *R*-factor, although with a value higher than the lowest possible to obtain with an exact match of two sets of points. What is more, it is impossible for the algorithm to attempt to fit individual peaks in the positions of noise spikes due to the profile width which is simply too large to enable such a behavior.

9 Conclusions

To conclude, diffraction patterns composed of a small number of broad peaks are a special case when used for structure determination. They can be obtained by extracting the crystalline component from experimental results of X-ray diffraction performed on polymer samples. The Rietveld refinement method, usually successful for structure refinement from curves consisting of a large number of narrow peaks, often fails in the case of these special patterns. A method alternative to local optimization used in the original algorithm is a genetic algorithm, free of the need of an initial model and capable of finding a global optimum.

Out of several techniques adapted to use with genetic algorithms from Rietveld refinement, the formulas used to compare the model curve with the experimental one, called *R*-factors, require some attention. While a simple sum of squares of residuals or its square root, or other formulas proposed in the original method work well for experimental patterns consisting of a large number of narrow peaks, it may be questioned if they are still applicable in the special case of a small number of very broad peaks.

Investigating the performance of various *R*-factors using a software designed for searching for the structure of the crystalline areas of the PANI/CSA polymer system using genetic algorithms shows that R_{wp}, due to a property enabling nearly flat diffraction patterns to be graded well when using it, is not particularly useful in this case. It was also found that the usage of fairly complicated factors, combining correlation or fit of the first derivative with a pure least sum of squares type of fit may yield very good results. Not only does the usage of these factors significantly improve the quality of the solutions found by the genetic algorithm, but also the increase in computing time caused by the use of such factors is still acceptable, or even unnoticeable. They seem to outperform original *R*-factors in this specific task.

Acknowledgements This work is supported in the form of a scholarship from KNOW—The Marian Smoluchowski Krakow Research Consortium "Matter–Energy–Future".

References

1. Rietveld, H.W.: A profile refinement method for nuclear and magnetic structures. J. Appl. Cryst. **2**, 65–71 (1969)
2. Harris, K.D.M., Johnston, R.L., Habershon, S.: Applications of evolutionary computation in structure determination from diffraction data. Struct. Bond. **110**, 55–94 (2004)
3. Łużny, W., Czarnecki, W.: Application of genetic algorithms to model the structure of molecular crystals. Polimery **59**(7–8), 542–548 (2014)
4. Kariuki, B.M., Serrano-González, H., Johnston, R.L., Harris, K.D.M.: The application of a genetic algorithm for solving crystal structures from powder diffraction data. Chem. Phys. Lett. **280**, 189–195 (1997)
5. Łużny, W., Samuelsen, E.J., Djurado, D., Nicolau, Y.F.: Polyaniline protonated with camphorsulphonic acid: modelling of its crystalline structure. Synth. Met. **90**, 19–23 (1997)
6. Śniechowski, M., Borek, R., Piwowarczyk, K., Łużny, W.: New structural model of PANI/CSA conducting polymer system obtained by molecular dynamics simulations. Macromol. Theor. Simul. **24**, 284–290 (2015)
7. Kozik, T., Łużny, W.: Modeling of the crystalline structure of the complex system containing doped polyaniline by use of genetic algorithms. J. Edu. Technol. Sci. **2**, 9–14 (2015)

Influence of Inlet Positions on the Flow Behavior Inside a Photoreactor

M. Poliński and Z. Stęgowski

Abstract Efficiency of a photoreactor depends on the irradiation dose. Fluid residence time distribution (RTD) reflects hydrodynamic behavior of the flow. A computational model was built on a base and fitting a previous radiotracer experiment. Results of three simulations for three different configurations and height flow rate are presented and discussed below. This paper shows usefulness of CFD modeling as an imaging tool, which can be used to retrieve detailed, local information about the flow.

Keywords Photoreactor · Residence time distribution · Tracers · CFD

1 Introduction

Effluent disinfection is often done in photoreactors, which consist of a set of UV lamps that irradiate the wastewater. Productivity of such apparatus depends on disinfection kinetics. Radiation intensity decreases with distance from the lamps with respect to exponential behavior according to Beer's law. In a perfect photoreactor, radiation dose would be equal in every fluid element. For this reason, flow shall be characterized with high mixing rate in tangential direction to UV lamps. Favored trajectory and stagnant zones should be reduced as much as possible, because their result downgrades the efficiency and raises operational costs. A variety of photoreactors designs was previously studied. One of the recent solutions was proposed by Moreira [1], who proposed a device with four vertical lamps, an upward flow and configurable inlets. They used it to conduct the radiotracers experiment in order to obtain Residence Time Distributions (RTD).

M. Poliński (✉) · Z. Stęgowski
Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: michal.polinski@fis.agh.edu.pl

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*, Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_13

Data gathered during their investigation was used as a starting point and validation reference for numerical modeling. Computational Fluid Dynamics (CFD) methods was deployed. Workflow of reconstructing RTD using virtual prototyping methods was described by Sugiharto et al. [2] Furman and Stęgowski, respectively, [3] have described RTD–CFD the junction in case of axial mixing in a pipe flow. The RTD curve contains comprehensive information about flow pattern, but values like local velocity and local effective dissipation rate are hidden because of its integral nature. The objective of this work is to retrieve and to show detailed information about local values of the flow inside the photoreactor, which cannot be directly quantified during radiotracer measurement. This paper describes the used workflow and present hydrodynamic behavior in a photoreactor with various inlets set up.

2 Experimental Part

This work has its origin in the measurements performed by Moreira [1] at the Center for the Development of Nuclear technology in Belo Horizonte in Brazil. As mentioned in the introduction they designed and built the photoreactor as a vertical tube 875 mm tall and with a diameter of 200 mm. PVC tubes were used, because of their common availability. Authors claim that this design should be very simple to build and does not require sophisticated tools and expensive materials. A set of four UV lamps was mounted inside—also vertically. Outflow was placed near the top. Inflows were situated close to the bottom and they had a possibility to choose from three varied configurations:

- Configuration 1: One central bottom inlet,
- Configuration 2: One lateral inlet,
- Configuration 3: Three lateral inlets equally spaced around the bottom of the reactor with flowrate divided into three equal parts.

During their experiments, different configurations were tested with flow rates starting from 0.112 dm^3/s to 0.881 dm^3/s.

For this paper, the upper flowrate level cases were chosen, because it emphasizes differences between the inlet configurations. To be precise, the following cases have been selected to in this paper:

- Simulation 1: 0.869 dm^3/s flow rate in one central bottom inlet setup (not done in experiment),
- Simulation 2: 0.869 dm^3/s flow rate in one lateral inlet set up,
- Simulation 3: 0.881 dm^3/s flow rate.

3 Residence Time Distribution

Residence time distribution (RTD) is a commonly used technique for describing flow patterns in a large class of applications, and particularly for investigating reactors. This term was first introduced by MacMullin and Weber in 1935 and the first application was proposed by Dankwerts [4] in 1953. Later it has been described in a multitude of scientific papers like Dudukovic [5] and Levenspiel [6].

Residence time distributions are usually obtained by injecting tracer at the examined flow inlet and then measuring its concentration at least at the outlet of the flow system. Normalized form of RTD could be treated as a probability distribution function for time space that describes the quantity of time a fluid element spends inside the flow system.

Injection of tracer shall not change the flow characteristics, so it shall be inert to the flow—have the same physical properties like density, viscosity, as the main fluid in the examined flow. Besides that, the substance chosen for the marker shall be clearly distinguishable from the current. There are multiple different approaches for the tracer choice. One of them is the use of a small amount of radioactive isotopes like proposed in Hector Constant-Machado et al. [7] Because emitting γ-radiation radiotracers are easy distinguishable and it is not needed to use high concentrations, they are also neutral for the flow. Emission of γ-radiation is proportional to the concentration of tracer in the flow. There are also other non-radioactive techniques for obtaining RTD, like the method that uses phosphorescent marker particles, which could be detected via light sensitive photomultiplier. An example of such measurements is shown in Harris et al. [8].

With an assumption of a constant flow rate RTD or the impulse response E(t) is usually defined as ratio of the outflow concentration $C_{out}(t)$ to the total amount of dissolved tracer.

$$E(t) = \frac{C_{out}(t)}{\int_0^\infty C_{out}(t)dt} \tag{1}$$

This is the simplest case, when inflow concentration $C_{in}(t)$ could be assumed as a Dirac delta function. For other cases, when injection is non-instantaneous $C_{out}(t)$ function is described as a product of the convolution of inflow $C_{in}(t)$ impulse by the impulse response E(t).

$$C_{out}(t) = \int_0^t C_{in}(t)E(t-x)dx \tag{2}$$

RTD experiments are a crucial concept to characterize flow and mixing inside the reactor. It helps to distinguish if the flow is close to one of opposed ideal reactor: plug reactor or perfect mixing reactor. The basic approach of examination RTD is a method of moments. The first moments of the measured residence time distribution yield the mean residence time in flow system. The second moment help

estimate the dispersion of tracer in the flow. Higher moments are also in use. In most cases, residence time distribution characterizes with positive skewness. The presence of stagnant zones in flow results with higher third moment.

Ideal plug flow reactor is characterized with no mixing in an axial direction. The fluid elements do not change their order during the flow, so residence time is equal for each single droplet of fluid. There is no dispersion at all, so the variance of an ideal plug flow reactor is equal to zero. Therefore, the residence time distribution is a Dirac delta function shifted by mean residence time T.

$$E(t) = \delta(t - T) \tag{3}$$

Another boundary case is a perfect mixing reactor also called a continuous stirred-tank reactor. This model is based on the assumption that inlet flow is continuously mixed with bulk volume of a reactor. All the time the outflow has the same uniform structure as the reactor volume. The residence time distribution for the perfect mixing reactor is described by exponential decay that is:

$$E(t) = \frac{1}{T} e^{-\frac{t}{T}} \tag{4}$$

Any residence time distribution could be mathematically described as a superposition of ideal plug flow reactors and perfect mixing reactors. Many compartment models were proposed to explain flow behavior. The agreement of such modeling with experimentally obtained RTD grows with number of blocks, compartments used to build a model—starting from one plug flow reactor with junction of one perfect mixing cell like proposed in Moreira [1] paper up to twenty shown by Hocine et al. [9]. Another examples of this quasi analytical approach for flow modeling could be found in Hocine [10], Haris et al. [8, 11] and also in Blet et al. [12].

As it is shown in the Fig. 1 the RTD describes the flow in black box manner, so in case to deeply analyze the flow it must be combined with another method like previously described method of moments or compartment models. Anyway it could be also used as a rich comparison data for direct simulation methods, because RTD data store information about whole flow in an integrate manner. For that reason, we use experimental residence time distributions as a reference point for computational fluid dynamics calculations to find what details have the highest influence for the stream.

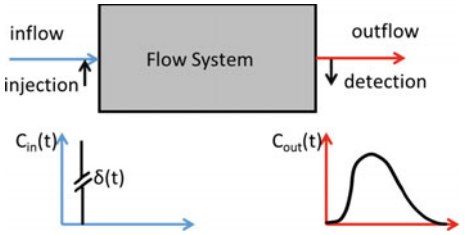


Fig. 1 Residence time distribution

4 Numerical Methods and Simulation Setup

Computational fluid dynamics (CFD) methods were used to model the flow in the photoreactor and then to simulate residence time distribution for each of chosen case. The commercial Fluent package [13] was used to drive the calculations. This section shortly presents governing equations and two-step workflow, which was used. First, the flow was calculated using a finite volume method. Therefore, particle tracking on previously calculated stationary flow was used to retrieve RTD.

4.1 Navier-Stokes Equations

The majority of fluid dynamics problems are solved using Navier–Stokes equations (N–S). This description of motion of viscous fluid substances is named after Claude–Louis Navier and George Gabriel Stokes and it was introduced in 1822. Until today, this set of equation has no analytical solution and smooth solutions to the Navier–Stokes equations are listed as one of the Millennium Prize Problems in mathematics, which are proposed by the Clay Mathematics Institute as the seven most important open problems in mathematics, with a reward of million US dollars for solving. They express the conservation of momentum principle. N–S equations are formed by applying Newton's second law of fluid motion with assumption that the stress in the fluid is the sum of the effects of viscosity and the effects associated with the pressure. Using Einstein's summation convention as follows:

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \rho\vec{g} + \overrightarrow{F_c} + \upsilon\frac{\partial^2 u_i}{\partial x_j \partial x_j} \tag{5}$$

where u_i denotes the velocity in Cartesian coordinates, t stands for time, p is pressure, density is denoted by ρ, kinematic viscosity is represented with υ. Gravitational acceleration is marked by $\vec{g}$ and $\overrightarrow{F_c}$ stands for Coriolis force.

Applying Eq. (5) together with conservation of mass, with assumption of incompressibility, i.e.,

$$\frac{\partial(u_i)}{\partial x_i} = 0 \tag{6}$$

results with complete psychical description for fluid in motion in Cartesian coordinates.

As it was mentioned above, there is as yet no smooth, general solution for the Navier–Stokes equation, but there are many specific solutions for particular problems, which are usually solved using dimensional analysis. Order of magnitude of each separate part of equations is calculated. Parts smaller than orders are neglected. For example, in laboratory scale, which shall be taken account in this paper the

gravitational acceleration $\vec{g}$ and Coriolis force $\overrightarrow{F_c}$ are negligible in comparison with viscous and pressure term. These assumptions yield

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \upsilon \frac{\partial^2 u_i}{\partial x_j \partial x_j} \tag{7}$$

Depending on the solving problem, different assumptions are also applied.

4.2 Reynolds Averaged Navier–Stokes Equation

The stationary Reynolds-Averaged Navier–Stokes (RANS) approach was used to drive the calculations of flow motion. It is the oldest and the best known way to model turbulent flow. The estimated Reynolds number for the investigated system is greater than 3,000. Therefore, mathematical model used to describe the flow needs to take into account its turbulent behavior. The method behind the equations is called Reynolds decomposition and it was firstly proposed by Osborne Reynolds [14] in 1895. Making an assumption that physical value averaged in time:

$$\underline{\varphi} = \frac{1}{\Delta t}\int_0^{\Delta t} \varphi_{(t)} dt \tag{8}$$

could be described as a sum of time average and fluctuations around it:

$$\varphi = \underline{\varphi} + \varphi' \tag{9}$$

Fluctuating part averaged in time in the same manner as shown above is equal to zero. These assumptions put inside the momentum conservation law yields

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \underline{p}}{\partial x_i} + \upsilon \frac{\partial \underline{u_i}}{\partial x_j \partial x_j} - \rho \frac{1}{\partial x_j}\underline{u_j' u_i'} \tag{10}$$

The left-hand side remains in the same form. Also the pressure derivative source term is not changing during Reynolds averaging. Fluctuations remain only in diffusion term. The component containing fluctuations is called the Reynolds stress tensor:

$$\tau_{ij}' = \underline{u_i' u_j'} \tag{11}$$

The term described by the equation above contains six unknown correlations, which is too many to solve the described system of equations and leads to closure problem. As a result another postulations need to be added. Due to this issue,

turbulence modeling has to be implemented to change the above equation system to the form of a closed set.

4.3 The Parameters Characteristic for Turbulence

Root mean squared average, which is also known as quadratic mean, could be calculated using the following definition:

$$\varphi'_{rms} = \sqrt{\frac{1}{\Delta t}\int_0^{\Delta t} (\varphi'_{(t)})^2 dt} \tag{12}$$

Applying the above formula to the term responsible for the turbulence kinetic energy per unit mass could be extracted, so the turbulence kinetic energy takes the form:

$$k = \frac{1}{2}\underline{u'_i u'_i} \tag{13}$$

Therefore, the amount of energy dissipated per unit mass, per unit time follows the rule:

$$\varepsilon = \frac{\upsilon}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{14}$$

4.4 Eddy Viscosity-Based Turbulence Models

Eddy viscosity-based turbulence models assumes the undefined Reynolds stress tensor using two hypotheses. The first one is the Boussinesq hypothesis, which involves the proportionality of Reynolds stress to the mean velocity gradients with proportional factor of turbulent viscosity υ_t. In our words, the Reynolds stress term behaves in the same way as viscosity stress.

$$\underline{u'_i u'_j} = -\upsilon_t\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) - \frac{2}{3}\delta_{ij}k \tag{15}$$

where δ_{ij} is a Kroneker delta function and υ_t denotes the turbulent viscosity. It is a scalar value, which means that it is assumed to be isotopic. The minus sign before the right-hand side means that turbulence tends to stop the flow via turbulence viscosity, which is taking energy from main stream and transport it to another direction.

The first turbulence model was proposed by Ludwig Prandtl in 1925 [15]. The mixing length model supposes eddy viscosity to be proportional to a velocity scale with a proportionality constant described by turbulent length scale L, that is,

$$\upsilon_t \propto LU \tag{16}$$

4.5 K-Epsilon RNG Equations

Applying dimensional analysis to the Eq. 11 velocity scale U could be described as a square root of turbulence kinetic energy $\sqrt{k}$ and then turbulent length scale L could be qualifying as by largest existing scale of turbulence in examined problem. There are many different approaches to choice the length scale L and there are depending on specified turbulent quantity. The k-ε RNG was employed in this paper. It relates the turbulent length scale to a kinetic energy dissipation rate ε, so the turbulent viscosity is expressed as

$$\upsilon_t = C_\mu \frac{k^2}{\varepsilon} \tag{17}$$

Proportionality constant C_μ is an empirical value. It is usually taken as 0.09.

The transport equations for the turbulence kinetic energy k and its dissipation rate ε are shown in the Eqs. 14 and 15.

$$\rho u_i \frac{\partial k}{\partial x_i} = \mu_t \left(\frac{1}{2} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) \right)^2 + \frac{\partial}{\partial x_i} \left(\alpha_k \mu_{eff} \frac{\partial k}{\partial x_i} \right) - \rho\varepsilon \tag{18}$$

$$\begin{aligned} \rho U_i \frac{\partial \varepsilon}{\partial x_i} = {} & C_1 \left(\frac{\varepsilon}{k} \right) \mu_t \left(\frac{1}{2} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) \right)^2 \\ & + \frac{\partial}{\partial x_i} \left(\alpha_\varepsilon \mu_{eff} \frac{\partial \varepsilon}{\partial x_i} \right) - C_2 \rho \left(\frac{\varepsilon^2}{k} \right) - R \end{aligned} \tag{19}$$

R is an additional scalar related to mean strain and turbulence quantities. This model was chosen because of an opportunity to account for the effects of smaller scales of motion. It was designed to improve accuracy in simulation of rotating flows, therefore mixing results also. A detailed description of k-epsilon RNG model and constant values can be found in Yakhot et al. [16] and Tennekes and Lumley [17].

4.6 Finite Volume Method and Boundary Conditions

Solving the equations for fluid main velocity (10) and its turbulence parameters (18), (19) requires the use of numerical methods. These equations were discretized on the numerical mesh using the finite volume method (FEM). This numerical approach represents analytic form of partial differential equations in shape of set of algebraic equations. Values are solved on a discrete computational mesh made up from set of small finite volumes. Using Gauss-Ostrogradsky's theorem divergence terms are converted into surface integral. The flow is discretized to the form of fluxes at the surfaces between adjacent finite volumes.

The zero step, when applying FVM, is to create system geometry and mesh, which follows corresponding dimensions of experimental setup. Tetrahedral mesh was used for creating the computational grid. There were approximately 3,000,000 cells used. It is shown in Fig. 2.

The following boundary conditions were used:

- No-slip wall boundary condition was set on the photoreactor sides—zero velocity and standard wall functions for flow calculations and reflection for discrete phase motion were assumed.

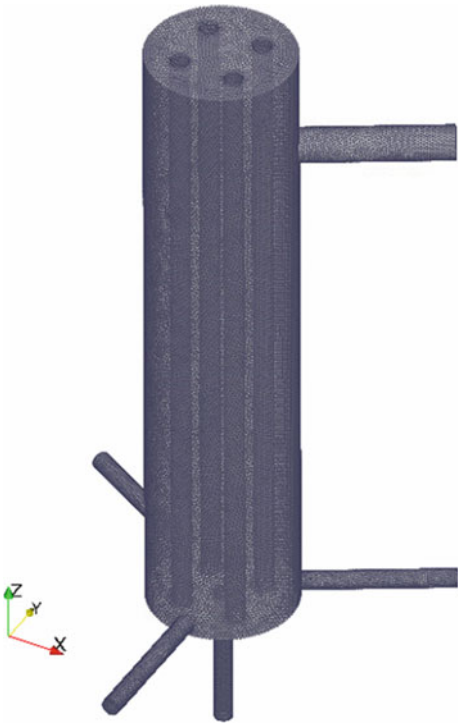


Fig. 2 Computational mesh

- Velocity inlets for inlet surfaces with constant velocity profile. In order to obtain physically correct axial velocity profile at the entrance to photoreactor inflow tubes was additionally extended.
- Outflow boundary condition was set for the outlet.

4.7 Particle Tracking

Simulating Residence time distribution requires additional calculations to the previously calculated stationary flow. Using prior simulated velocity field, a particle tracking method was deployed to retrieve RTD. Similar approach as described by Zhang and Chen [18], Cantu-Perez et al. [19, 20] was used. Numerous trajectories were simulated using the Lagrangian particle tracking method in case to collect statistics and determine probability distribution function.

For each simulation, more than 10,000 individual particle tracks were calculated. Particles were described with the same physical values as fluid volume. The droplet size d was set at 10^{-6} m, which is a value far below the size of the involved computational grid. Particle motion is driven by the Stokes drag law and it is described by the equation:

$$\frac{du_{i,p}}{dx} = \left(\frac{\upsilon 18}{d^2 C_c}\right)\left(\underline{u_i} - u_{i,p}\right) \tag{20}$$

Discrete random walk (DRM) model was used to simulate stochastic velocity fluctuations in the flow. The fluctuating velocity component follows Gaussian probability distribution and is proportional to the turbulent kinetic energy k calculated in the previous step that is:

$$u_i^{'} = \zeta\sqrt{\frac{2}{3}k} \tag{21}$$

5 Results and Discussion

This section presents the results of the conducted simulations. On the experimental field, it was not possible to look inside the flow. One of the reasons was that photoreactor sides have to be painted black, because of safety reasons resulting from the use of UV and γ radiation. Various measurement techniques like particle image velocimetry [21], different kinds of anemometry are costly or flow disturbing. Applying CFD techniques flow could be analyzed without restrictions in any place inside the system, but in need of a fit to experimental data. RTD measurements are taken at the inlet and outlet of the system and it covers information about

the whole flow, but any particular local information must be recalculated and resituated from it. Many simulation loops were done to obtain the best fit with experimental data.

The most intuitive way to describe the flow is visualization of velocity field. It is presented in two ways. Mean velocity trajectories are shown in Fig. 3.

Detailed local velocity directions are presented in Fig. 4 as vectors applied to horizontal cross-sections.

It can be observed that in the central inlet configuration a big part of volume is a stagnant zone, which is omitted by a mean flow. Introduction of lateral inlet configuration produces a swirling flow with significantly decreased stagnant zone. One lateral inlet is characterized by large spin. At the bottom part, it results with height rates velocities close to the sides. It produces local pressure drop in the center, which is the origin of the backflow. Dividing input flow into free lateral inlets, a swirling part of the flow was decreased as well. Opposite inlet position results with the most balanced axial velocity profile of all simulated cases.

Another interesting value that should be shown is effective viscosity, which is proportional to diffusion coefficient. It is described by equation X and it contains both physical values describing the turbulence: turbulent kinetic energy k and dissipation rate ε. The photoreactor should be characterized by a high mixing rate in tangential direction to the UV lamps. Effective viscosity rate inside it is presented on vertical cross-sections shown in Fig. 5.

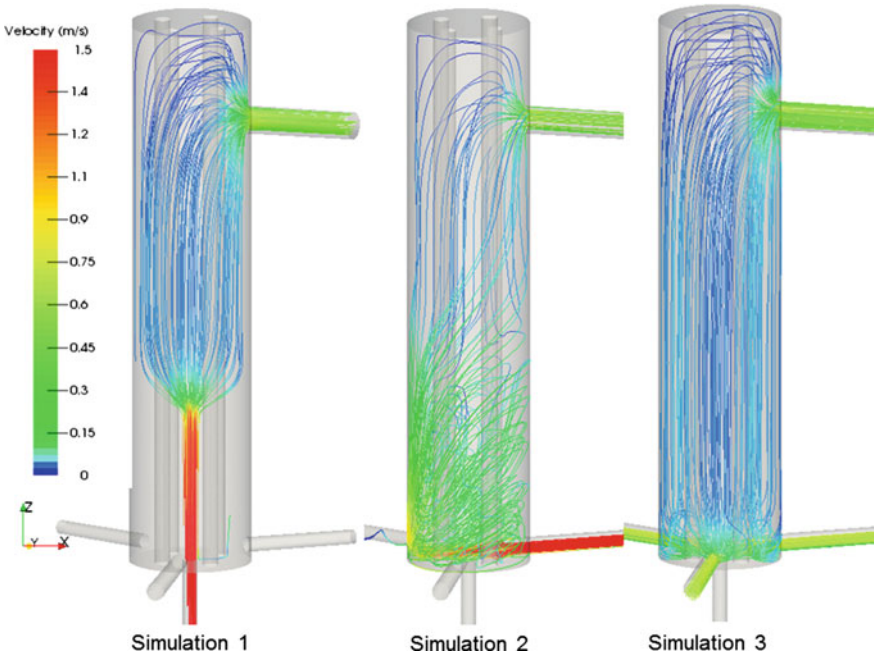


Fig. 3 Mean velocity streamlines

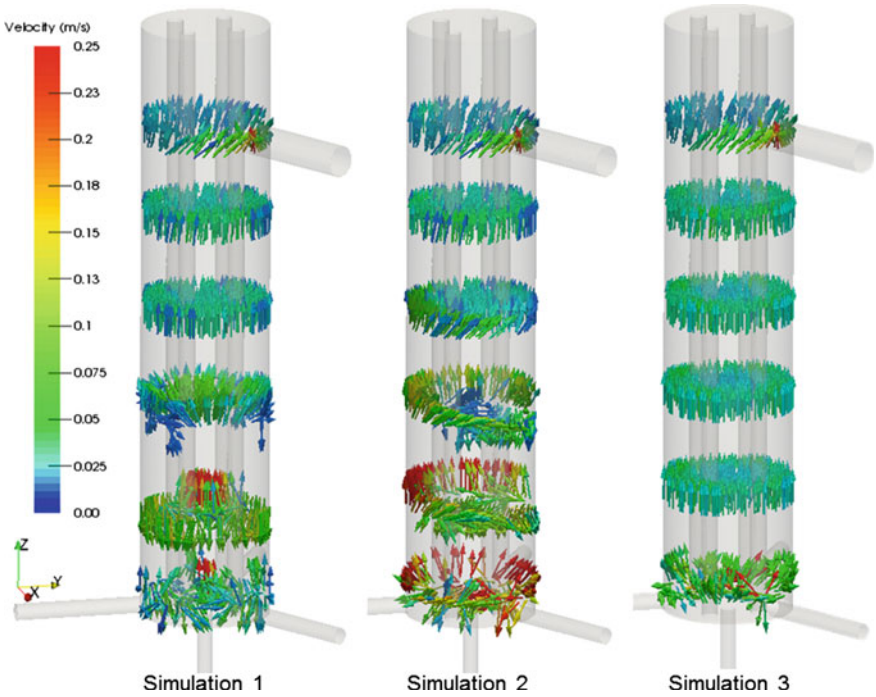


Fig. 4 Velocity Vectors

Water viscosity in a standard state is approximately equal to 0.001 kg/ms so the greater part of effective viscosity is the turbulence viscosity. As is describes the rate of subgrid motion, it is related to dispersion in small scales. There are significant differences between tested configurations, but in all cases the highest effective viscosity is in the center. Spin inside photoreactor is the highest in the simulation 2 case, which is also characterized by the lowest small scales motion. The highest viscosity rate is in the simulation 3 case, which is desired. It is likely to have high exchange rate close to the UV lamps in case of getting the dose per particle as equal as possible. Turbulent viscosity rate affects the specific particle path. Figure 6 presents 5 specific particle tracks.

In the illustration for simulation 1 it could be seen that in the bottom side parts, where the velocities are negligible in comparison to the main jet in the center, there is still a little movement. It might be called a dead zone, but it is not one, because there is still fluid exchange between the bottom and upper part of the system. Particles that are crossing through this stagnant zone are characterized by very long residence time in the system. It is desired to avoid such situations, because it has a tendency of an unnecessary overdose. As it was described earlier, simulation 2 has the lowest turbulence part and it produces the sharpest particle trajectories. Anyway, the backflow in the center results with trajectories with long overall residence time. Trajectories in simulation 3 case are fuzzy.

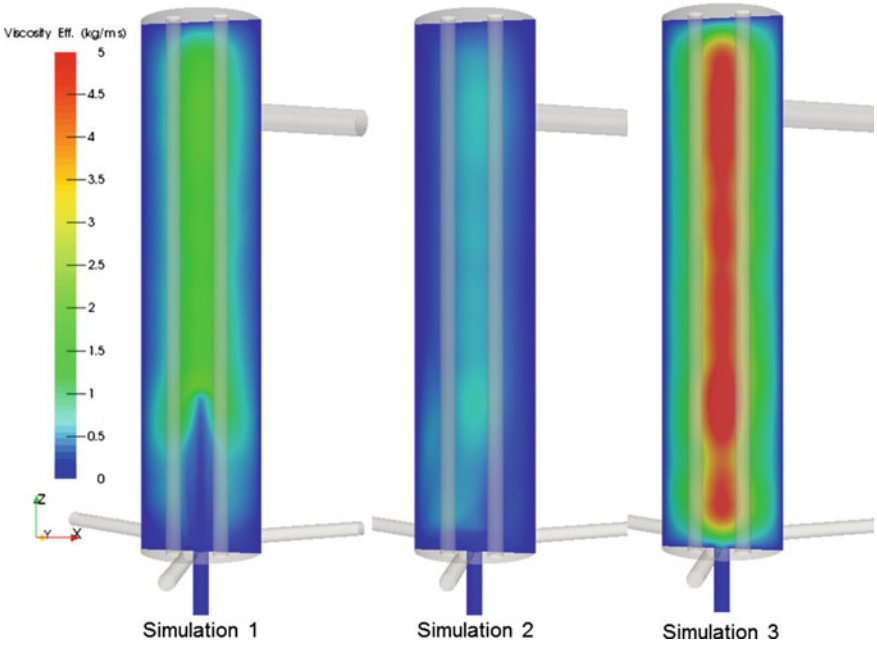


Fig. 5 Effective velocity field

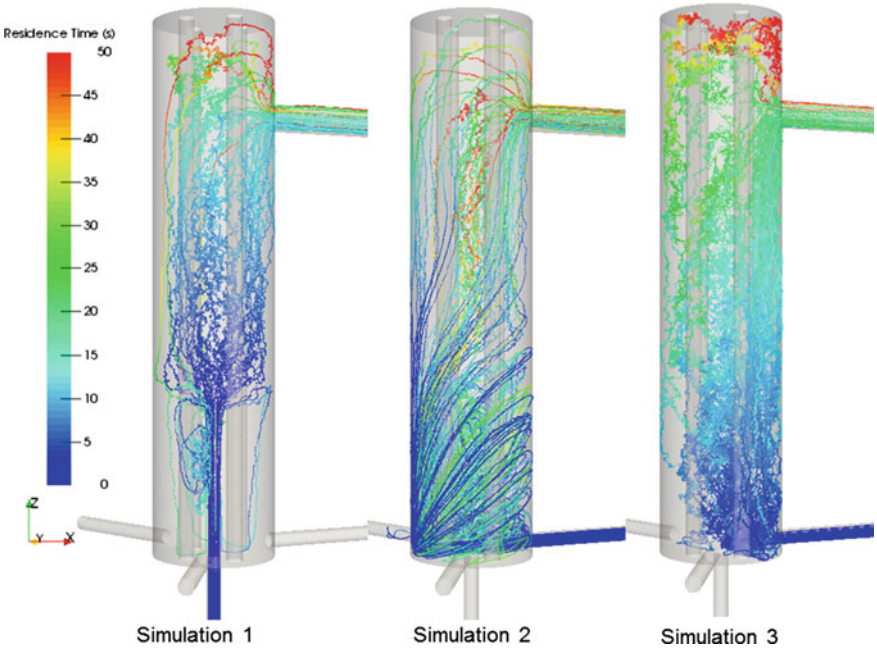


Fig. 6 Specific particle trajectories

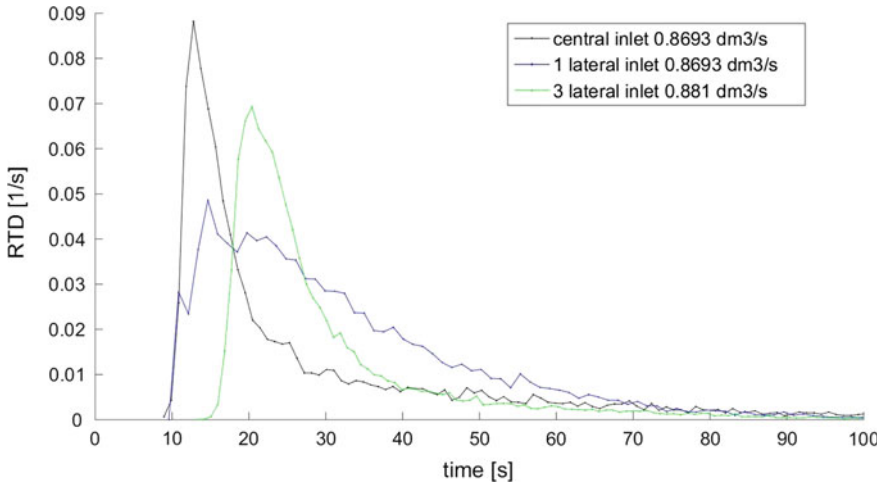


Fig. 7 Residence time distribution

The most equal axial velocity profile in that case results with that particle residence time tends to be proportional to the distance from the inlet. All flow characteristics described below lead to different shape of RTD. It was calculated as an overall particle residence time in the system for many simulated particles. For each curve shown in Fig. 6, more than 10,000 individual particle tracks were calculated.

As could be expected there are major differences between obtained RTDs, which is presented on Fig. 7.

6 Conclusion

This paper shows the use of computational fluid dynamics methods in case to illustrate and comment on flow behavior in the photoreactor previously investigated by a radiotracer experiment. Computer simulation could be used to retrieve detailed, local information about the flow. It is also convenient for retrieving residence time distributions. The height flow rate was chosen to emphasize the differences between three different inlet configuration. Major contrast between three investigated setups was shown. Simulation 2 is the worst case. It characterizes with big spin, which results with presence of undesired backflow. It has widely spread RDT, so radiation dose would differ much over the fluid elements. Simulation 1 case has the narrowest RDT over all simulated cases, which is expected regarding to fact that equal radiation dose distribution is one of the goals in photoreactors design. From the other hand, it produces big stagnant zone at the inlet side, which results with wasting the active volume of an apparatus. Approximately one-third of volume is out of the main stream and stays inside a photoreactor for a much longer time. That

fluid part might collect too high a radiation dose and its reduce dose collected by main stream. Simulation 3 has RTD somewhere in between the other two, so it might not be found as the best solution, when looking only at the RTD curve. Using CFD simulation, it could be seen that it also has the most balanced axial velocity profile and characterizes with the highest small-scale mixing rate, so it tends to be a most efficient set up in all test cases. In all three simulations there is still room for improvement near the outlet, because the upper part of volume above the outlet results with prolongation of residence time. This work describes hydrodynamic behavior of the flow inside the photoreactor, but its efficiency is also associated with disinfection kinetics. It shall be taken account in the succeeding investigations as well as the outflow design.

Acknowledgements The authors acknowledge the support of originators [1] for providing experimental data needed to lead simulations made for this research. This research was supported in part by PL-Grid Infrastructure. M.P. acknowledges his benefit from Ph.D. scholarships founded by Marian Smoluchowski Cracow Scientific Consortium—KNOW.

References

1. Moreira, R.M., Pinto, A.M.F., Mesnier, R., Leclerc, J-P.: Influence of inlet positions on the flow behaviour inside a photoreactor using radiotracers and colored tracer investigation. Appl. Radiat. Isot. **65**, 419–427 (2004)
2. Sugiharto, S., Stegowski, Z., Furman, L., Su'Ud, Z., Kurniadi, R., Waris, A., Abidin, Z.: Dispersion determination in a turbulent pipeflow using radiotracer data and CFD analysis. Comput. Fluids **79**, 77–81 (2013)
3. Furman, L., Stęgowski, Z.: CFD models of jet mixing and their validation by tracer experiments. Chem. Eng. Process. **50**(3), 300–304 (2011). ISSN 0255-2701
4. Danckwerts, P.V.: Continuous flow systems, distribution of residence times. Chem. Eng. Sci. **2**(1), 1–13 (1953)
5. Dudukovic, M.P.: Tracer methods in chemical reactors. Techniques and applications. Chem. React. Des. Technol., NATO ASI Series (1986)
6. Levenspiel, O.: Chemical Reaction Engineering, 3rd edn. Wiley, New York (1999)
7. Constant-Machado, H., Lecrerc, J.P., Avilan, E., Landaeta, G., Anorga, N., Capote, O.: Flow modeling of a battery of industrial crude oil/gas separators using ^{113m}Im tracer experiments. Chem. Eng. Process. **44** (2005)
8. Haris, A.T., Davidson, J.F., Thorpe, R.B.: A novel method for measuring the residence time distribution in short time scale particulate systems. Chem. Eng. J. **89** (2002)
9. Hocine, S.: Identification de modeles de procedes par programmation mixte deterministe. Ph. D. thesis, INP Toulouse (2006)
10. Hocine, S., Pibouleau, L., Azzaro-Pantel, C., Domenech, S.: Modelling systems defined by RTD curves. Comput. Chem. Eng. **32** (2008)
11. Haris, A.T., Davidson, J.F., Thorpe, R.B.: Particle residence time distributions in circulating fluidised beds. Chem. Eng. Sci. **58** (2003)
12. Blet, V., Berne, Ph., Chaussy, C., Perrin, S., Schweich, D.: Characterization of a packed column using radioactive tracers. Chem. Eng. Sci. (1999)
13. FLUENT Inc., FLUENT 6 User's Manual, 2006
14. Osborne, R.: On the dynamical theory of incompressible viscous fluids and the determination of the criterion. Philos. Trans. R. Soc. London A **186**, 123–164 (1895)

15. Prandtl, L.: Z. angew. Meth. Mech **5**(1), 136–139 (1925)
16. Yakhot, V., Orszag, S.A., Thangam, S., Gatski, T.B., Speziale, C.G.: Development of turbulence models for shear flows by a double expansion technique. Phys. Fluids A **4**(7), 1510–1520 (1992)
17. Tennekes, H., Lumley, J.L.: A First Course in Turbulence.
18. Zhang, Z., Chen, Q.: Experimental measurements and numerical simulation of particle transport and distribution in ventilated rooms. Atmos. Environ. **40**, 419–427 (2006)
19. Cantu-Perez, A., Barrass, S., Gavriilidis, A.: Residence time distributions in microchannels: comparison between channels with herringbone structures and rectangular channel. Chem. Eng. J. **160** (2010)
20. Cantu-Perez, A., Barrass, S., Gavriilidis, A.: Hydrodynamics and reaction studies in a layered herringbone channel. Chem. Eng. J. **167** (2011)
21. Pruvost, J., Legrand, J., Legentilhomme, P., Doubliez, L.: Particle image velocimetry investigation of the flow-field of a 3D turbulent annular swirling decaying flow induced by means of a tangential inlet. Exp. Fluids **29** (2000)

Simulations of Transport Characteristics of Core-Shell Nanowire Transistors with Electrostatic All-Around Gate

Tomasz Palutkiewicz, Maciej Wołoszyn and Bartłomiej J. Spisak

Abstract A mathematical model of the investigated semiconductor core-shell nanowire transistor with all-around gate, computation methods and calculated transport characteristics of the device are presented. The influence of applied gate voltage and drain-source voltage on the potential energy profile of the system is determined, electric current flowing through it is calculated and dependence of operation regime of the device on these voltages is discussed.

1 Introduction

Vertical semiconductor core-shell nanowires can be manufactured with surrounding gate electrodes (all-around gates) by various top-down, bottom-up or mixed methods [1–4]. Such devices can be used as efficient transistor components because transmission through it can be controlled by applied gate voltage.

In top-down methods, the nanodevice is carved out of a larger piece of bulk material which is partially covered by some protective layers to distinguish between a produced device and the rest of material. Methods of this kind include electrophoresis, optical lithography, or electron-beam lithography [5, 6].

On the other hand, bottom-up methods consist of growing the nanodevice on the surface of the substrate by adding atoms of subsequent elements, which allows fabrication of a nanowires build of many layers of different materials. In this category methods like electrochemical deposition, vapor–liquid–solid growth or epitaxy can be found [2, 5].

These technologies allow production of efficient three-dimensional transistors (surrounding gate transistors or SGT) which can be placed with very high density on a single chip and can be very useful in advanced nanoelectronic devices. Experimental realization of such transistors is presented in [4, 7], and gated nanowires that are

T. Palutkiewicz (✉) · M. Wołoszyn · B.J. Spisak
Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, Al. A. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: palutkiewicz@fis.agh.edu.pl

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*, Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_14

shown in them have typical static current–voltage characteristics of three terminal devices.

We present computational study of the nanostructure of this kind—a core-shell nanowire transistor with surrounding gate placed asymmetrically in the vicinity of the drain electrode, and discuss influences of applied gate and drain-source voltages its on transport properties. Potential energy profiles of the device in different states, transmission coefficients and static current–voltage characteristics are calculated and assessed. Algorithm and implementation is also briefly described.

The paper is organized as follows: in Sect. 2, a three-dimensional model of the semiconductor core-shell nanowire with the asymmetrically placed all-around gate is presented; equations, theoretical methods used for their solving and investigation of the transport properties of the considered nanowire are described in Sect. 3; Sect. 4 contains presentation and discussion of the obtained results; a brief summary and conclusions are given in Sect. 5.

2 Model of Nanowire Transistor

A semiconductor core-shell nanowire transistor with all-around gate electrode and circular cross-section is modeled as a rod with large aspect ratio of length and diameter. It consists of few cylindrical layers of semiconductors, metals and oxides with different material constants (starting from the axis): a core, a shell, an insulation, and a gate electrode. The shape of the nanowire is fully characterized by radius and thicknesses of every layer (r_c, t_s, t_o, t_g), length of the whole nanowire and the gate (l_w, l_g) and distance of the gate from drain electrode (l_d)—see Table 1. To obtain a

Table 1 Parameters of calculations

Wire length	$l_w = 1200$ nm	
Gate length	$l_g = 200$ nm	
Gate-drain distance	$l_d = 50$ nm	
Electrode length	$l_e = 100$ nm	
Wire core radius	$r_c = 62$ nm	
Wire shell thickness	$r_s = 18$ nm	
Oxide thickness	$t_o = 20$ nm	
Core relative permittivity	$\epsilon_c = 13$	$In_{0.7}Ga_{0.3}As$
Shell relative permittivity	$\epsilon_s = 10$	$In_{0.5}Ga_{0.5}As$
Oxide relative permittivity	$\epsilon_o = 120$	TiO_2
Effective mass	$m^* = 0.041m_0$	$In_{0.7}Ga_{0.3}As$
Temperature	$T = 4$ K	
Fermi energy	$\mu_S = 0.01$ eV	

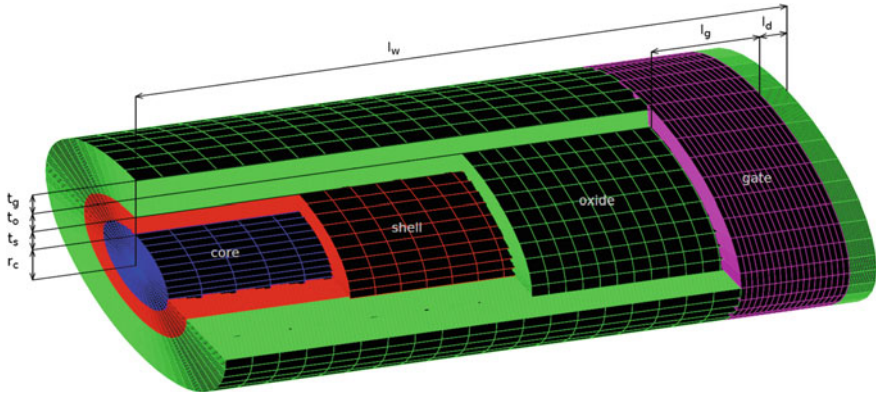


Fig. 1 Schematic (not in scale) of the modeled core-shell nanowire transistor with all-around gate in the vicinity of the drain electrode. Source and drain electrodes are not included in the picture. Dimensions, materials, and material constants are described in Table 1

proper calculation box for numerical methods, certain lengths of electrodes are also specified.

It is assumed that both ends of the wire are attached to reflectionless reservoirs of electrons through the perfect contacts but the gate is isolated from the wire—there is no electric transport to or from it. These reservoirs are termed source (s) and drain (d). A schematic of the single nanodevice is presented in Fig. 1. A typical nanowire transistor consists of an array of tens or hundreds of such nanodevices connected to common source and drain electrodes. In this work we consider them separately.

3 Calculations

To determine the potential profile $V(\mathbf{r})$ of the nanowire, Poisson's equation with varying electric permittivity is solved in 3D

$$-\nabla \cdot [\epsilon(\mathbf{r})\nabla V(\mathbf{r})] = \rho(\mathbf{r}), \tag{1}$$

where $\epsilon(\mathbf{r})$ is the position-dependent electric permittivity set as piecewise constant function for each layer and $\rho(\mathbf{r})$ is the density of electric charge which currently is uniformly set to 0. This equation is numerically solved with Dirichlet boundary conditions where $V(\mathbf{r}) = 0$ is set at limits of computation box and $V(\mathbf{r}) = V_g$ is set at gate electrode. Since electric potentials are additive, at this point drain-source voltage (V_{ds}) is set to 0 and external linear potential caused by it will be added in later steps of calculations. Parallel explicit multigrid relaxation is used to solve the equation.

Although the implicit method has better convergence and requires less memory as it is performed in-place, it destroys the symmetry of the obtained solution which can create slight differences between degenerated eigenstates of Schrödinger's equation calculated in the next step, and indispose optimizations of following steps calculations—each of these states would have to be considered separately. Usage of an explicit method ensures that initial symmetry of the system is preserved.

Density of the grid is increased 4 times for every dimension simultaneously in order to obtain final resolution. Each time all sizes are doubled and intermediate points are calculated as arithmetic means of surrounding values: first in every dimension, then on diagonals. Perpendicular cross-sections of the final grid are used in the following calculations. A one-dimensional longitudinal profile calculated for the axis of the wire is also used in later steps.

Parallelization of this step is performed at intermediate level—inside the Poisson's equation solving method, inside each of its iterations, but at most external loop—for every plane of the grid. Every line of every plane and every cell of every line is calculated sequentially, but multiple planes are calculated simultaneously. Parallelization is most efficient when the count of calculations to be performed simultaneously is the same or a small multiple of the count of available computation cores (processors) (but the count of parallel threads should not exceed the count of cores, surplus calculations need to be performed later in sequence to avoid switching of threads); so there is no reason to parallelize multiple levels of calculations (for example for multiple dimensions) when a single level has a greater size than count of available cores. If only hundreds of cores are available, parallelization of one dimension is enough, for two dimensions it would need tens of thousands and for three dimensions millions of cores to work at full efficiency (there are hundreds of points in each dimension).

In the next step one-particle spinless Schrödinger's equation for conduction band electrons is solved within adiabatic [8–11] and effective-mass approximation:

$$\left[-\frac{\hbar^2}{2m^*}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right) + U_\perp\left(x,y|z;V_g\right)\right]\chi_n\left(x,y|z\right) = E_n^\perp\left(z;V_g\right)\chi_n\left(x,y|z\right), \tag{2}$$

where $U_\perp\left(x,y|z;V_g\right)$ is a perpendicular energy profile of cross-section at distance of z from source electrode under V_g, $\chi_n\left(x,y|z\right)$ is a perpendicular wave function and $E_n^\perp\left(z;V_g\right)$ is the energy of n-th eigenstate at that cross-section;

$$\left[-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2} + E_n^\perp\left(z;V_g\right) + U_\parallel\left(V_{ds},z\right)\right]\phi_n\left(z\right) = E\phi_n\left(z\right). \tag{3}$$

where $\phi_n\left(z\right)$ is a parallel wave function and $U_\parallel\left(V_{ds},z\right) = e\left(V_{ds}/l_w\right)z$ is a parallel energy profile of electric field caused by applied drain-source voltage. The effect of modes mixing is neglected in this approach.

Equation (2) is solved by the finite differences method with Dirichlet boundary conditions $\lim_{(x,y)\to\infty}\chi_n(x,y;z)=0$ while (3) is solved by quantum transmit-

ting boundary method [12] with open boundary conditions and plain waves of equal amplitude entering the system from both end electrodes (wave functions which leave the system can be calculated from the solution). From Eq. 2 we obtain only eigenenergies and eigenstates of transport channels while Eq. 3 is solved for every channel and every discrete energy value in considered range. Wave function of n-th channel is given by the formula

$$\psi_n(\mathbf{r}) = \chi_n(x, y|z)\,\phi_n(z)\,. \tag{4}$$

and electric current at the drain (I_d) is calculated by the formula [13, 14]:

$$I(V_{ds}, V_g, T) = \frac{e}{\pi\hbar}\sum_n \int_0^\infty dE$$
$$T_n^{\parallel}(E; V_{ds}, V_g) \times [f_{FD}(E; \mu_S, T) - f_{FD}(E; \mu_D, T)], \tag{5}$$

where f_{FD} is Fermi–Dirac distribution function of the electrons in the source (drain) contact with electrochemical potential $\mu_{S(D)}$ where $\mu_D = \mu_S - eV_{ds}$, T is temperature and $T_n^{\parallel}(E; V_{ds}, V_g)$ is transmission coefficient of n-th channel given by the formula

$$T_n^{\parallel}(E; V_{ds}, V_g) = \frac{J_n\left(l_w, E\right)}{J_n(0, E)}, \tag{6}$$

where $J_n(z, E)$ is probability current in the ingoing and outgoing wave given by the formula

$$J_n(z, E) = \frac{\hbar}{m^*}\mathfrak{Im}\left\{\phi_n^*(z)\,\frac{d}{dz}\phi_n(z)\right\}. \tag{7}$$

Parallelization of the second step is performed at high level—every cross-section is calculated sequentially, but multiple cross-sections are calculated at the same time. The count of cross-sections is similar to the count of planes in previous step, but in this particular problem, a bit smaller: cross-sections in vicinity of the source electrode are far from the gate, so it has no significant influence on them and they have very similar, almost the same, plain potential profile as it can be seen at Fig. 2. As some of them can be omitted from calculations of this step to save time and result from adjacent one will be used instead. In this case, the count of parallel calculations is also in the order of hundreds, which justifies limitation of previous parallelization to just one dimension—it is the same for both cases.

In third step results for various drain-source voltages are calculated simultaneously. Count of these values depends on range and resolution of V_{ds} calculations and is fully independent from sizes of previous parallel sections of algorithm; but this step of the problem is relatively small and is calculated in short time; so in case of such necessity, a greater number of iterations in a sequence is acceptable.

Our software can dynamically lock and free resources such as computation cores or operating memory, but their availability can also be limited by operating system

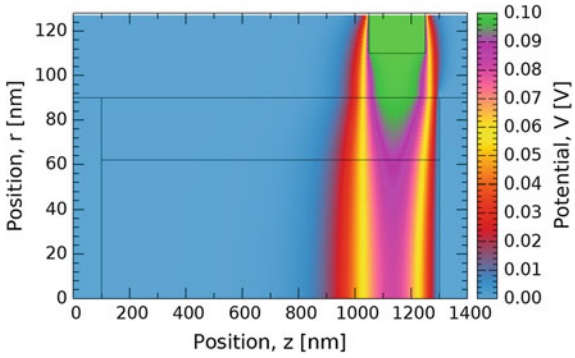


Fig. 2 Map of potential dependency on cylindrical coordinates in the nanowire for $V_g = 0.1$ V applied to the gate. *Lines* on the graph indicate borders of layers of the nanodevice

and other processes, usually it is somehow managed, so we have to assume that it is assigned as requested and constant for the whole duration of calculations and therefore the algorithm is optimized for about one hundred cores.

4 Results

In this section results of calculations for a single cylindrical core-shell nanowire transistor with all-around gate with parameters having values given in Table 1 are presented (Figs. 3 and 4).

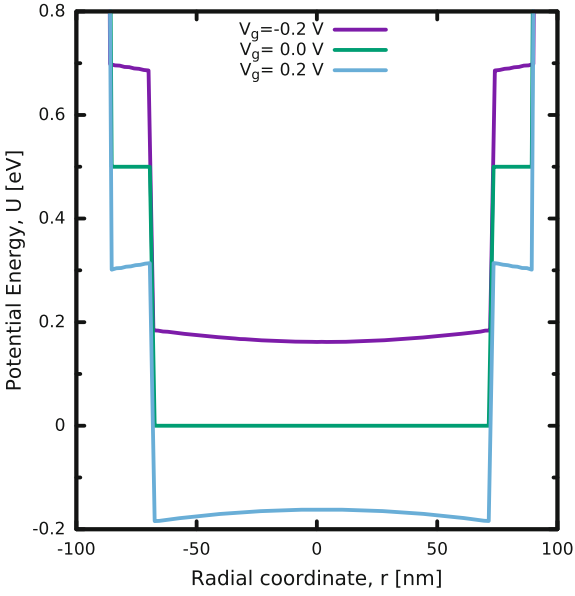


Fig. 3 Cross-section of the wire perpendicular to its axis in the middle of the gate electrode ($z = 1150$ nm) at different applied gate voltages V_g and $V_{ds} = 0$ V

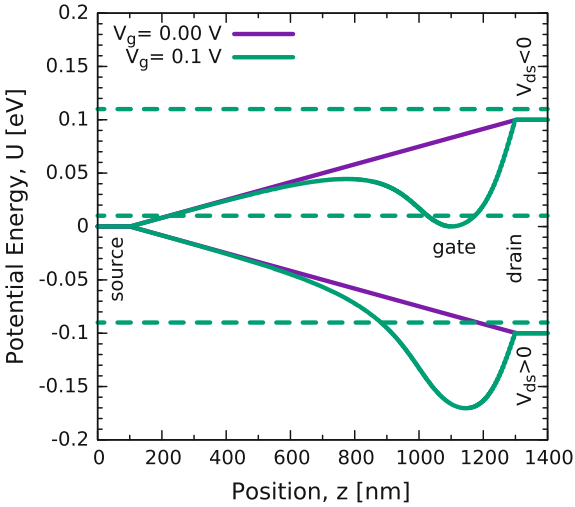


Fig. 4 Example potential energy profiles for selected applied gate voltages V_g and positive and negative drain-source voltage V_{ds} along the axis of the wire. *Dashed lines* represent Fermi levels in the source and drain electrodes

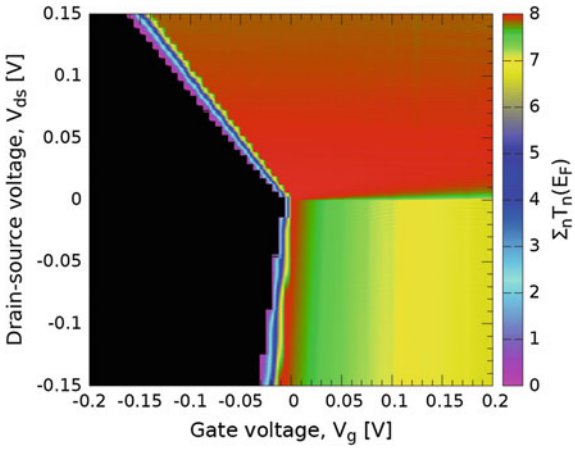


Fig. 5 Total transmission coefficient dependence on drain-source voltage V_{ds} and gate voltage V_g. Threshold voltage dependence on both input voltages is observed in second quarter of the graph and strong transmission dependence on gate voltage is observed in fourth quarter

With such parameters a coherent transport regime with eight well separated open channels in the transport window was obtained. The total transmission coefficient is visible at Fig. 5.

Current-voltage (I-V) characteristics of the investigated device are shown at Fig. 6. As can be seen the applied gate voltage (V_g) has strong influence on transmission of conduction electrons and it can be used to control the flow of the current in the analyzed nanodevice. The saturation current depends significantly on the gate voltage when $V_{ds} < 0$ and $V_g > 0$, so this can be regarded as transistor operation regime of the nanowire with all-around gate designed and placed as in our experiment. For $V_{ds} > 0$ saturation current remains almost the same and only threshold voltage changes significantly. Because of large length of the gate electrode, $V_g < 0$ causes a very wide potential barrier to appear in the system and it can completely

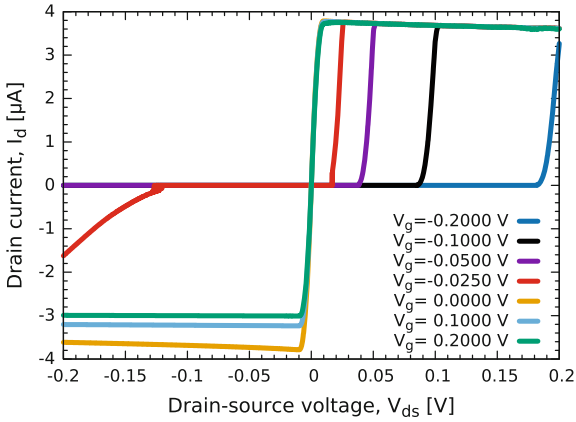


Fig. 6 The current–voltage characteristics of core-shell nanowire transistor with asymmetrically placed all-around gate for selected gate voltages V_g

prevent electronic transmission through the nanodevice. In this case the threshold value of this voltage depends on V_{ds}, which bows the barrier efficiently reducing its length. For $V_{ds} > 0$ and $V_g > 0$ total transmission coefficient is close to the number of open channels, which means that its value is close to 1 for every open channel and electrons can pass through the device coherently—these voltage values have almost no influence on the transport through the considered device.

5 Conclusion

Properties of a cylindrical three-dimensional semiconductor core-shell nanowire transistor with all-around gate were considered in the coherent regime of electronic transport. In the studied case conduction electrons are confined to the core of nanowire. Because of asymmetric location of the gate near the drain contact, the sign of drain-source voltage unequivocally determines the regime of transistor operation, where the gate voltage can be used to control electronic current, which works this way only for negative value of V_{ds} and positive values of V_g. For positive values of V_{ds} the saturation current remains almost constant for every value of V_g, while negative values of V_g totally block the transport in the wire if V_{ds} is not positive and large enough.

Acknowledgements T.P. is supported by the scholarship of Krakow Smoluchowski Scientific Consortium from the funding for National Leading Research Centre by Ministry of Science and Higher Education (Poland).
A preliminary version of this paper was presented as [15].

References

1. Hobbs, R.G., Petkov, N., Holmes, J.D.: Semiconductor nanowire fabrication by bottom-up and top-down paradigms. Chem. Mater. **24**(11), 1975 (2012)
2. Tomioka, K., Kobayashi, Y., Motohisa, J., Hara, S., Fukui, T.: Selective-area growth of vertically aligned GaAs and GaAs/AlGaAs core-shell nanowires on Si(111) substrate. Nanotechnology **20**(14), 145302 (2009)
3. Fang, M., Han, N., Wang, F., Yang, Z.-X., Yip, S.P., Dong, G., Hou, J.J., Chueh, Y., Ho, J.C.: III-V nanowires: synthesis, property manipulations, and device applications. J. Nanomater. **2014**, 1 (2014)
4. Tomioka, K., Yoshimura, M.: A III-V nanowire channel on silicon for high-performance vertical transistors. Nature **488**, 189 (2012)
5. Hoobs, R., Holmes, J.: *Semiconductor Nanowire Fabrication via Bottom-Up and Top-Down Paradigms* (2011)
6. Nassiopoulou, A., Gianneta, V., Katsogridakis, C.: Si nanowires by a single-step metal-assisted chemical etching process on lithographically defined areas: formation kinetics. Nanoscale Res. Lett. **6**, 597–605 (2011)
7. Larrieu, G., Han, X.-L.: Vertical nanowire array-based field effect transistors for ultimate scaling. Nanoscale **5**, 2437 (2013)
8. Yacoby, A., Imry, Y.: Quantization of the conductance of ballistic point contacts beyond the adiabatic approximation. Phys. Rev. B **41**, 5341 (1990)
9. Maaø, F.A., Zozulenko, I.V., Hauge, E.H.: Quantum point contacts with smooth geometries: exact versus approximate results. Phys. Rev. B **50**, 17320 (1994)
10. Brandbyge, M., Jacobsen, K.W., Nørskov, J.K.: Scattering and conductance quantization in three-dimensional metal nanocontacts. Phys. Rev. B **55**, 2637 (1997)
11. Wołoszyn, M., Spisak, B., Adamowski, J., Wójcik, P.: Magnetoresistance anomalies resulting from stark resonances in semiconductor nanowires with a constriction. J. Phys. Condens. Matter **26**(32), 325301 (2014)
12. Lent, C., Kirkner, D.: Quantum ballistic transport in a dual-gate Si transistor. J. Appl. Phys. **67**, 6353 (1990)
13. Di Ventra, M.: *Electrical Transport in Nanoscale Systems*. Cambridge University Press (2008)
14. Palutkiewicz, T., Wołoszyn, M., Adamowski, J., Wójcik, P., Spisak, B.: Infuence of geometrical parameters on the transport characteristics of gated core-multishell nanowires. Acta Physica Polonica A **129**(1A) (2016) (44th International School and Conference on the Physics of Semiconductors Jaszowiec 2015)
15. Palutkiewicz, T., Wołoszyn, M., Spisak, B.J.: Simulations of transport characteristics of core-shell nanowire transistors with electrostatic all-around gate. In: 2015, Presentation at Congress on Information Technology, Computational and Experimental Physics (CITCEP 2015)

On Some Recent Construction Methods for Bivariate Copulas

Radko Mesiar and Anna Kolesárová

Abstract Recently, we have introduced several new methods of constructing bivariate copulas. In this overview paper, we recall and exemplify some of them. We first introduce ultramodular copulas and then we present two construction methods for bivariate copulas based on ultramodular copulas. As an application, the construction of DUCS copulas is shown. The third presented construction method is based on quadratic polynomials of three variables. A quadratic construction is applied for deriving special classes of perturbed copulas. Finally, particular modular functions are considered for constructing copulas. The discussed construction methods are illustrated by several examples.

1 Introduction

Recall that a function $C : [0,1]^2 \to [0,1]$ is a (bivariate) copula whenever

(i) C is grounded, i.e., for all $x \in [0,1]$,

$$C(x,0) = C(0,x) = 0,$$

(ii) 1 is its neutral element, i.e., for all $x \in [0,1]$,

$$C(x,1) = C(1,x) = x,$$

R. Mesiar (✉)
Faculty of Civil Engineering, Department of Mathematics and Descriptive Geometry,
Slovak University of Technology in Bratislava, Radlinského 11,
810 05 Bratislava, Slovakia
e-mail: radko.mesiar@stuba.sk

A. Kolesárová
Faculty of Chemical and Food Technology, Institute of Information Engineering,
Automation and Mathematics, Slovak University of Technology in Bratislava,
Radlinského 9, 812 37 Bratislava, Slovakia
e-mail: anna.kolesarova@stuba.sk

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0_15

(iii) C is supermodular, i.e., for all $\mathbf{x}, \mathbf{y} \in [0,1]^2$,

$$C(\mathbf{x} \vee \mathbf{y}) + C(\mathbf{x} \wedge \mathbf{y}) \geq C(\mathbf{x}) + C(\mathbf{y}).$$

For more details concerning copulas we recommend the monographs [7] and [22]. The important role of copulas in statistics is stressed by the Sklar theorem, showing that for any random vector $Z = (X, Y)$, for all $x, y \in \overline{\mathbb{R}}$, $F_Z(x, y) = C\left(F_X(x), F_Y(y)\right)$ for some copula C (which is unique whenever Z is continuous). The Sklar theorem brings not only a representation of joint distribution functions, but it is also a method for constructing bivariate distribution functions. To increase the fitting potential of recent software tools for stochastic modeling, new kinds of copulas are welcomed.

In this contribution, we bring several new construction methods for bivariate copulas which we have recently proposed and studied in our papers [12–15, 19, 20]. A preliminary version of this paper was presented at CITCEP 2015 in Cracow in our lecture entitled "On some construction methods for bivariate copulas".

The paper is organized as follows. In Sect. 2, ultramodular copulas are introduced and two related construction methods are described. We also show how the DUCS copulas can be obtained by a construction method based on ultramodular copulas. Section 3 brings quadratic constructions of copulas. As an illustration, quadratic constructions are applied for obtaining two special classes of perturbed copulas. In Sect. 4 we apply modular functions to construct bivariate copulas. Finally, in concluding remarks some other construction methods introduced by our working group are briefly mentioned.

2 Ultramodular Copulas in Constructions of Bivariate Copulas

Based on the results of [12, 14], we first introduce the role of ultramodularity in copula constructions, including some examples.

A copula $C : [0,1]^2 \to [0,1]$ is *ultramodular* [13, 14] if and only if for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in [0,1]^2$ satisfying $\mathbf{x} + \mathbf{y} + \mathbf{z} \in [0,1]^2$ we have

$$C(\mathbf{x} + \mathbf{y} + \mathbf{z}) + C(\mathbf{x}) \geq C(\mathbf{x} + \mathbf{y}) + C(\mathbf{x} + \mathbf{z}). \tag{1}$$

Note that ultramodular copulas are just copulas with convex horizontal and vertical sections.

Out of the three basic copulas W, M and the product copula Π, which are given by $W(x, y) = \max\{0, x + y - 1\}$, $M(x, y) = \min\{x, y\}$ and $\Pi(x, y) = xy$, only W and Π are ultramodular. However, the upper Fréchet–Hoeffding bound M is ultramodular on the upper left triangle

$$\Delta = \{(x, y) \in [0,1]^2 \mid x \leq y\}.$$

Theorem 1 *([14], Theorem 3.1) Let $C : [0,1]^2 \to [0,1]$ be an Archimedean copula with a two times differentiable additive generator $f : [0,1] \to [0,\infty]$. Then C is ultramodular if and only if f' is constant or $\frac{1}{f'}$ is a convex function.*

As a consequence of Theorem 4.1 in [14] we have

Theorem 2 *Let $C_1, C_2, D : [0,1]^2 \to [0,1]$ be copulas and assume that D is ultramodular. Then, for all monotone non–decreasing functions $f_1, f_2, g_1, g_2 : [0,1] \to [0,1]$ with $D\left(f_1(x), f_2(x)\right) = D\left(g_1(x), g_2(x)\right) = x$ for all $x \in [0,1]$, also the function $E : [0,1]^2 \to [0,1]$ given by*

$$E(x,y) = D\left(C_1\left(f_1(x), g_1(y)\right), C_2\left(f_2(x), g_2(y)\right)\right) \tag{2}$$

is a copula.

Remark 1 Construction (2) can be generalized for ultramodular n–ary aggregation functions $A : [0,1]^n \to [0,1]$, see [8], copulas $C_1, \ldots, C_n : [0,1]^2 \to [0,1]$, and functions $f_1, \ldots, f_n, g_1, \ldots, g_n : [0,1] \to [0,1]$. So, for example, if one considers the n–ary product Π as A, the function $E : [0,1]^2 \to [0,1]$ given by

$$E(x,y) = \prod_{i=1}^{n} C_i\left(f_i(x), g_i(y)\right),$$

where $\prod_{i=1}^{n} f_i = \prod_{i=1}^{n} g_i = id_{[0,1]}$ and $f_1, \ldots, f_n, g_1, \ldots, g_n$ are non–decreasing, is a copula for all copulas $C_1, \ldots, C_n$.

Example 1 Here is an example of the construction proposed in Theorem 2:
For each copula C and all $\alpha, \beta \in [0,1]$, the function $E : [0,1]^2 \to [0,1]$ given by $E(x,y) = C\left(x^{\alpha}, y^{\beta}\right) C\left(x^{1-\alpha}, y^{1-\beta}\right)$ is a copula (this result was obtained independently in [11], see also [17]). Putting $C = W$ and $\alpha = \beta = 0.5$, we obtain the Clayton copula with parameter -0.5 (see [22]) given by $C_{-0.5}(x,y) = \left(\max\left\{\sqrt{x} + \sqrt{y} - 1, 0\right\}\right)^2$.

Recently, based on our proposal of univariate conditioning of copulas [18], Durante and Jaworski [6] have characterized generated univariate conditioning stable copulas.

Recall that if C is a bivariate copula and $\alpha \in]0,1]$, then the copula $C_{(\alpha)} : [0,1]^2 \to [0,1]$ given by

$$C_{(\alpha)}(x,y) = \frac{C\left(\varphi^{(-1)}(x), \alpha y\right)}{\alpha},$$

where $\varphi^{(-1)}(x) = \sup\{t \in [0,1] \mid C(t,\alpha) < \alpha x\}$, is called a univariate conditional copula of C. A copula C is called univariate conditional stable whenever, for any $\alpha \in]0,1]$, the corresponding univariate conditional copula $C_{(\alpha)}$ coincides with C, i.e., if $C_{(\alpha)} = C$.

Consider an additive generator $f\colon [0,1] \to [0,\infty]$ of an Archimedean copula [22], i.e., a strictly decreasing convex continuous function satisfying $f(1) = 0$. Then the functions C_f, $C^f\colon [0,1]^2 \to [0,1]$ given by

$$C_f(x,y) = \begin{cases} 0 & \text{if } x = 0, \\ xf^{(-1)}\left(\frac{f(y)}{x}\right) & \text{otherwise,} \end{cases}$$

and

$$C^f(x,y) = \begin{cases} 0 & \text{if } x = 0, \\ xf^{(-1)}\left(1 - \left(\frac{f(1-y)}{x}\right)\right) & \text{otherwise,} \end{cases}$$

where $f^{(-1)}$, given by $f^{(-1)}(x) = \min\{f(0), x\}$, is the pseudo-inverse of f, are generated univariate conditioning stable copulas. For example, taking the function $f(x) = 1 - x$, i.e. an additive generator of the lower Fréchet–Hoeffding bound W, then we have $C_f = W$ and $C^f = M$.

Now, let C be an Archimedean copula with an additive generator f, and consider any monotone functions $d, \tilde{d}\colon [0,1] \to [0,1]$ such that $d(x)\tilde{d}(x) = x$ for each $x \in [0,1]$. Putting in Theorem 2 as a ultramodular copula D the product Π, and $C_1 = \Pi$, $C_2 = C_f, f_1 = d, f_2 = \tilde{d}$, $g_1 = 1$ and $g_2 = id_{[0,1]}$, then, applying (2), we have that the function $E\colon [0,1]^2 \to [0,1]$ given, for all $x \neq 0$, by

$$E(x,y) = \tilde{d}(x)C_f(d(x),y) = xf^{(-1)}\left(\frac{f(y)}{d(x)}\right),$$

is a copula. Similarly, for $x \neq 0$, the function

$$H(x,y) = \tilde{d}(x)C^f(d(x),y) = x\left(1 - f^{(-1)}\left(\frac{1-f(y)}{d(x)}\right)\right),$$

is a copula. Such copulas were introduced in [19] under the name DUCS copulas (Distorted Univariate Conditioning Stable copulas).

Note that the introduction of DUCS copulas was inspired by a similar distortion of Archimedean copulas resulting into Archimax copulas $C_{f,F}\colon [0,1]^2 \to [0,1]$, given, for all $(x,y) \in]0,1[^2$, by

$$C_{f,F}(x,y) = f^{(-1)}\left((f(x)+f(y))\,F\left(\frac{f(x)}{f(x)+f(y)}\right)\right),$$

where $F\colon [0,1] \to [0,1]$ is a dependence function characterized by a convexity and the property $F(x) \geq \max\{x, 1-x\}$. For more details we recommend [2, 9].

Next, we will need the Schur concavity of a copula $D : [0,1]^2 \to [0,1]$ on the upper left triangle $\Delta = \{(x,y) \in [0,1]^2 | x \leq y\}$, which means that for all $(x,y) \in \Delta$ and for all $\varepsilon > 0$ with $(x+\varepsilon, y-\varepsilon) \in \Delta$ we have

$$D(x, y) \leq D(x + \varepsilon, y - \varepsilon).$$

Theorem 3 *Let C be a bivariate copula and let D be a binary copula which is ultramodular and Schur concave on the upper left triangle Δ. Then the function $D(C, C^*)$ is a copula, where C^* is the dual copula given by*

$$C^*(x, y) = x + y - C(x, y).$$

It is remarkable that $D(C, C^*)$ in Theorem 3 preserves the ultramodularity and the Schur concavity on Δ of the copulas C and D.

Proposition 1 *Let C, D be bivariate copulas which are ultramodular and Schur concave on the upper left triangle Δ. Then also the copula $D(C, C^*)$ is ultramodular and Schur concave on Δ.*

It turns out that the ultramodularity of D is a necessary condition if we expect $D(C, C^*)$ to be a copula for each copula C.

Theorem 4 *Let D be a bivariate copula such that for each binary copula C the function $D(C, C^*)$ is a copula. Then D is ultramodular on the upper left triangle Δ.*

Example 2 The product copula Π is both ultramodular and Schur concave, and thus $\Pi_C : [0, 1]^2 \to [0, 1]$ given by $\Pi_C(x, y) = \Pi(C, C^*)(x, y) = C(x, y)(x + y - C(x, y))$ is a copula for any bivariate copula C.

Remark 2 Observe that $\Pi_C \leq \Pi$, i.e., Π_C is a negative quadrant dependent copula [22], independently of C. For example, $\Pi_M = \Pi$. Moreover, putting $C^{(1)} = C$, $C^{(n+1)} = \Pi_{C^{(n)}}$ for $n = 1, 2, \ldots$, we have

$$\lim_{n \to \infty} C^{(n)} = W.$$

A similar observation holds for any ultramodular copula D. Indeed, The ultramodularity of D implies $D \leq \Pi$, and thus $D(C, C^*) \leq \Pi_C$. Putting $C_D^{(1)} = C$, $C_D^{(n+1)} = D\left(C_D^{(n)}, C_D^{(n)*}\right)$, we obtain $C_D^{(n)} \leq C^{(n)}$ and hence $\lim_{n \to \infty} C_D^{(n)} = W$.

3 Quadratic Constructions of Copulas

Next, we recall quadratic constructions of copulas [15] and their stochastic interpretation based on the results of [5]. Inspired by the fact that the copula Π_C, introduced in Example 2, can be seen as a composite function,

$$\Pi_C(x, y) = P(x, y, C(x, y)),$$

where P is a quadratic polynomial of the form $P(x, y, z) = z(x + y - z)$, in [15] we were interested in such ternary quadratic polynomials P,

$$P(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j$$

with coefficients $a, \ldots, j \in \mathbb{R}$, for which the function $C_P : [0, 1]^2 \to [0, 1]$ given by

$$C_P(x, y) = P(x, y, C(x, y)) \tag{3}$$

is a copula for each bivariate copula C.

A construction (3) of a copula C_P by means of a copula C and a quadratic polynomial P is called a quadratic construction of a copula. Polynomials P giving via (3) a copula for any copula C, are said to be universal polynomials for quadratic constructions of copulas.

Example 3 (i) If we consider the polynomial $P(x, y, z) = z^2 - xz - yz + 2z$ and the basic copulas W, M and Π, then

- for $C = W$ we have $W_P = W$;
- for $C = M$ we have $M_P(x, y) = \min\{x, y\}(2 - \max\{x, y\}) > M(x, y)$, which shows that M_P is not a copula;
- for $C = \Pi$, $\Pi_P(x, y) = xy + xy(1 - x)(1 - y)$, i.e. Π_P belongs to the Farlie–Gumbel–Morgenstern family of copulas.

(ii) For the polynomial $P(x, y, z) = z^2$ and any copula C we obtain that for all $x \in [0, 1]$, $C_P(x, 1) = P(x, 1, C(x, 1)) = (C(x, 1))^2 = x^2$, i.e., $e = 1$ is not a neutral element of C_P, thus C_P is never a copula.

We can see that the polynomial $P(x, y, z) = z(x + y - z)$ is universal for quadratic constructions of copulas, the polynomial P, considered in Example 3 (i), does not always lead to a copula, and the polynomial $P(x, y, z) = z^2$ cannot be used for a quadratic construction of copulas in any case. The following theorem gives a complete characterization of all universal polynomials for quadratic constructions of copulas.

Theorem 5 *For a copula $C : [0, 1]^2 \to [0, 1]$, let C_P be a function defined on $[0, 1]^2$ by (3), i.e.,*

$$C_P(x, y) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j,$$

where $z = C(x, y)$ and $a, \ldots, j \in \mathbb{R}$. Then the following are equivalent.

(i) For any copula C, C_P is a copula.

(ii) C_P is given by

$$C_P(x, y) = cC^2(x, y) + dxy - cxC(x, y) - cyC(x, y) + (1 + c - d)C(x, y),$$

with coefficients c, d satisfying the conditions

$$0 \le d \le 1, \ 0 \le d - c \le 1.$$

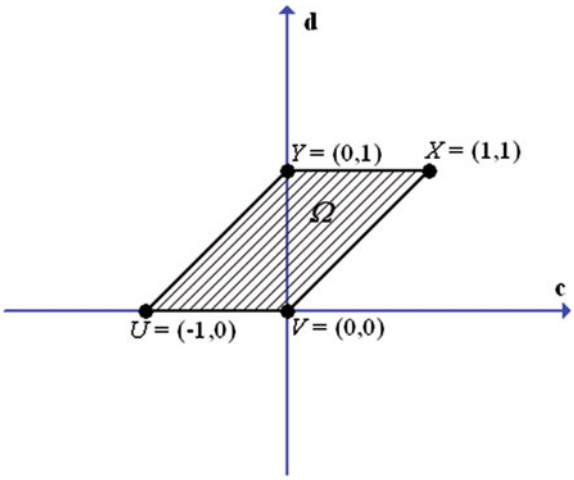


Fig. 1 The domain Ω of all possible pairs of coefficients (c, d)

Let

$$\Omega = \{(c,d) \in \mathbb{R}^2 \mid 0 \le d \le 1,\ 0 \le d - c \le 1\}, \tag{4}$$

see Fig. 1., and for $(c, d) \in \Omega$, let

$$P_{c,d}(x,y,z) = cz^2 + dxy - cxz - cyz + (1 + c - d)z. \tag{5}$$

Due to the convexity of the set Ω, each copula $C_{P_{c,d}}$ can be expressed as a convex combination of copulas $C_{P_{-1,0}}$, $C_{P_{0,0}}$, $C_{P_{1,1}}$ and $C_{P_{0,1}}$ corresponding to the vertices $U = (-1, 0)$, $V = (0, 0)$, $X = (1, 1)$ and $Y = (0, 1)$ of the set Ω, where

$$\begin{aligned}
C_{P_{-1,0}}(x,y) &= -C^2(x,y) + xC(x,y) + yC(x,y) \\
&= C(x,y)(x + y - C(x,y)) = \Pi_C, \\
C_{P_{0,0}}(x,y) &= C(x,y), \\
C_{P_{0,1}}(x,y) &= xy = \Pi(x,y), \\
C_{P_{1,1}}(x,y) &= C^2(x,y) + xy - xC(x,y) - yC(x,y) \\
&\quad + C(x,y).
\end{aligned}$$

Remark 3 Due to the previous result, to any copula C, we can assign a two-parametric class of copulas, namely the class $\left(C_{P_{c,d}}\right)_{(c,d)\in\Omega}$. For each copula C, it is a convex class of copulas, always containing copulas C and Π. Note, that the copulas of the type $C_{P_{c,0}}$ and $C_{P_{c,1}}$ already appeared in [5]. However, the results given there were obtained by probabilistic methods.

We say that a copula C is invariant under the quadratic construction (3) generated by a quadratic polynomial P, if for all $(x, y) \in [0, 1]^2$ we have $C_P(x, y) = C(x, y)$.

As was shown in [15], except the case $c = d = 0$ (when $C_{P_{0,0}} = C$ for each C), the only copula C invariant under the quadratic construction generated by a polynomial $P_{c,d}$, see (5), is the Plackett copula C_α^{Pl} [22] with parameter $\alpha = \frac{d}{d-c}$ if $d \neq c$, and $C = M = C_\infty^{Pl}$ if $d = c \neq 0$. Recall that the Plackett copulas are given by

$$C_\alpha^{Pl}(x,y) = \begin{cases} \frac{[1+(\alpha-1)(x+y)] - \sqrt{[1+(\alpha-1)(x+y)]^2 - 4\alpha(\alpha-1)xy}}{2(\alpha-1)}, & \alpha > 0, \alpha \neq 1, \\ \Pi(x,y), & \alpha = 1. \end{cases}$$

As mentioned above, for $c = -1$ and $d = 0$, the corresponding quadratic construction coincides with the construction described in Example 2, i.e., $C_{P_{-1,0}} = \Pi_C$. This construction also has the following interesting stochastic interpretation, see [5].

Consider continuous independent identically distributed random vectors (X_1, Y_1) and (X_2, Y_2) characterized by a copula C, and such that X_1, Y_1, X_2, Y_2 are uniformly distributed over $[0, 1]$. Then the random vector (Z_1, Z_2),

$$(Z_1, Z_2) = \begin{cases} \big(\min(X_1, X_2), \max(Y_1, Y_2)\big) & \text{with pp. } 0.5, \\ \big(\max(X_1, X_2), \min(Y_1, Y_2)\big) & \text{with pp. } 0.5, \end{cases}$$

is characterized by the copula Π_C.

Now, consider $c = d = 1$. The polynomial $P_{1,1}$, see (5), can be written as

$$P_{1,1}(x,y) = z + (x - z)(y - z).$$

For simplicity let us denote this polynomial by Q, i.e., $P_{1,1} = Q$. By Theorem 5, the polynomial Q is universal for the quadratic construction of copulas, i.e., the function $C_Q \colon [0,1]^2 \to [0,1]$,

$$C_Q(x,y) = C(x,y) + (x - C(x,y))(y - C(x,y))$$

is a copula for each copula C. Due to the convexity of the class of all bivariate copulas, for any $\lambda \in [0,1]$, the function $C_{\lambda,Q} \colon [0,1]^2 \to [0,1]$ given by $C_{\lambda,Q} = (1-\lambda)C + \lambda C_Q$, whose values can be written as

$$C_{\lambda,Q}(x,y) = C(x,y) + \lambda(x - C(x,y))(y - C(x,y)),$$

is a copula. The parametric class $\{C_{\lambda,Q}\}$ of copulas varies from C to C_Q, and its members $C_{\lambda,Q}$ are called perturbations of a copula C, see [20].

Note that starting from the smallest copula $C = W$, we obtain $W_Q = \Pi$, and that copulas $\Pi_{\lambda,Q}$, $\lambda \in [0,1]$, are known as the Farlie–Gumbel–Morgenstern copulas. Several interesting statistical properties of perturbed copulas $C_{\lambda,Q}$ can be found in [16].

Similarly, starting from any copula C, we can introduce by means of a quadratic construction another family $\{C_{\lambda,H}\}_{\lambda\in[-1,0]}$ of perturbed copulas. Considering

$c = -1$, $d = 0$ and denoting the polynomial $P_{-1,0}$ by H, for any $\alpha \in [0,1]$, we can construct a copula $(1-\alpha)C + \alpha C_H$, which is given by

$$(1-\alpha)C(x,y) + \alpha C(x,y)(x+y-C(x,y)).$$

Transforming $\alpha \mapsto -\lambda$, we obtain copulas $C_{\lambda,H}$, with $\lambda \in [-1,0]$, which can be written in the form

$$C_{\lambda,H}(x,y) = C(x,y) + \lambda C(x,y)(C(x,y) - x - y + 1).$$

For example, for the greatest copula $C = M$, the copula $M_{-1,H} = \Pi$. The perturbed copulas $\Pi_{\lambda,H}$, $\lambda \in [-1,0]$, are the Farlie-Gumbel-Morgenstern copulas. Moreover, the union of the above mentioned classes $\{\Pi_{\lambda,Q}\}_{\lambda\in[0,1]}$ and $\{\Pi_{\lambda,H}\}_{\lambda\in[-1,0]}$ gives the complete family of the Farlie–Gumbel–Morgenstern copulas, i.e., $\{C_\lambda^{FGM}\}_{\lambda\in[-1,1]}$, where $C_\lambda^{FGM}(x,y) = xy + \lambda xy(1-x)(1-y)$.

4 Modular Functions in Constructions of Copulas

Another kind of construction of singular copulas is related to modular functions [3]. Recall that a function $A : [0,1]^2 \to [0,1]$ is called an aggregation function if it is monotone and satisfies two boundary conditions $A(0,0) = 0$ and $A(1,1) = 1$ [8].

An aggregation function A is said to be

- modular if for all $\mathbf{x}, \mathbf{y} \in [0,1]^2$ we have

$$A(\mathbf{x}) + A(\mathbf{y}) = A(\mathbf{x} \vee \mathbf{y}) + A(\mathbf{x} \wedge \mathbf{y});$$

- 1–Lipschitz if for all $\mathbf{x} = (x_1, y_1), \mathbf{y} = (x_2, y_2) \in [0,1]^2$ we have

$$|A(\mathbf{x}) - A(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|_1 = |x_1 - x_2| + |y_1 - y_2|.$$

In [3] we have proved the following result:

Theorem 6 *Let $A : [0,1]^2 \to [0,1]$ be a modular 1–Lipschitz aggregation function. Then the function $\tilde{A} : [0,1]^2 \to [0,1]$ given by*

$$\tilde{A}(x,y) = \min\{x, y, A(x,y)\} \tag{6}$$

is a copula.

Note that for any copula $C : [0,1]^2 \to [0,1]$, the function $A_C : [0,1]^2 \to [0,1]$ given by

$$A_C(x,y) = \frac{C(x,x) + C(y,y)}{2}$$

satisfies the constraints of Theorem 6, and then $\tilde{A}_C$ given by (6) is a diagonal copula introduced in [23]. Theorem 6 can be generalized by considering a 1–Lipschitz non-decreasing modular function $A\colon [0,1]^2 \to \mathbb{R}$ such that $A(1,1) \geq 1$ and $A(0,0) \geq 0$. Also under such relaxed constraints, the construction (6) gives a copula $\tilde{A}$.

We also have the next related method based on the smallest copula W.

Theorem 7 *Let $A\colon [0,1]^2 \to \mathbb{R}$ be a non-decreasing modular 1–Lipschitz aggregation function such that $A(1,0) \leq 0$ and $A(0,1) \leq 0$. Then the function $\overline{A}\colon [0,1]^2 \to [0,1]$ given by*

$$\overline{A}(x,y) = \max\{W(x,y), A(x,y)\} \tag{7}$$

is a copula.

Remark 4 Note that the modularity of a function A in Theorems 6 and 7 can be replaced by supermodularity and still the functions $\tilde{A}$ given in (6) and $\overline{A}$ given in (7) are copulas.

5 Concluding Remarks

We have described and discussed construction methods for bivariate copulas based on ultramodular copulas, on quadratic polynomials and on modular functions.

In addition to them, we still recall some other recently studied construction methods for copulas. Conic copulas were introduced in [10]. A close relation between Archimax copulas [1] and conic copulas [10] was shown in [4]. Also interesting, especially for fitting purposes, seems to be an approach based on perturbations of particular copulas [20]. Finally, recall DUCS copulas which were introduced and discussed in [19] and partially also discussed in Sect. 2.

Acknowledgements The work on this presentation was supported by the grant APVV–14–0013.

References

1. Alsina, C., Frank, M.J., Schweizer, B.: Associative functions on intervals: a primer of triangular norms. World Scientific, Singapore (2006)
2. Capéraà, P., Fougères, A.L., Genest, C.: Bivariate distribution with given extreme value attractor. J. Multivar. Anal. **72**, 30–49 (2000)
3. De Baets, B., De Meyer, H., Kalická, J., Mesiar, R.: On the relationship between modular functions and copulas. Fuzzy Sets Syst. **268**, 110–126 (2015)
4. Dibala, M., Vavríková, L.: The relations between conic and archimax copulas. Submitted

5. Dolati, A., Úbeda-Flores, M.: Constructing copulas by means of pairs of order statistics. Kybernetika **45**, 992–1002 (2009)
6. Durante, F., Jaworski, P.: Invariant dependence structure under univariate truncation. Statistics **46**, 263–267 (2012)
7. Durante, F., Sempi, C.: Princ. Copula Theor. CRC/Chapman & Hall, Boca Raton, FL (2016)
8. Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions: encyclopedia of mathematics. Cambridge University Press (2009)
9. Jágr, V.: Generalization of archimax copulas for higher dimensions. PhD. Thesis, STU Bratislava (2012)
10. Jwaid, T., De Baets, B., Kalická, J., Mesiar, R.: Conic aggregation functions. Fuzzy Sets Syst. **167**, 3–20 (2011)
11. Khoudraji, A.: Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivariées. PhD. Thesis, Université Laval, Québec (1995)
12. Klement, E.P., Kolesárová, A., Mesiar, R., Saminger–Platz, S.: On the role of ultramodularity in the construction of binary copulas. Submitted manuscript
13. Klement, E.P., Manzi, M., Mesiar, R.: Ultramodular aggregation functions. Inf. Sci. **181**(19), 4101–4111 (2011)
14. Klement, E.P., Manzi, M., Mesiar, R.: Ultramodularity and copulas. Rocky Mt. J. Math. **44**, 189–202 (2014)
15. Kolesárová, A., Mayor, G., Mesiar, R.: Quadratic constructions of copulas. Inf. Sci. **310**, 69–76 (2015)
16. Komorník, J., Komorníková, M., Kalická, J.: Dependence measures for perturbations of copulas, Fuzzy Sets Syst. Submitted manuscript
17. Liebscher, E.: Construction of asymmetric multivariate copulas. J. Multivar. Anal. **99**, 2234–2250 (2008)
18. Mesiar, R., Jágr, V., Juráňová, M., Komorníková, M.: Univariate conditioning of copulas. Kybernetika **44**, 807–816 (2008)
19. Mesiar, R., Pekárová, M.: Ducs copulas. Kybernetika **46**(6), 1069–1077 (2010)
20. Mesiar, R., Komorník, J., Komorníková, M.: Perturbation of bivariate copulas. Fuzzy Sets Syst. **268**, 127–140 (2015)
21. Moynihan, R.: On τ_T semigroups of probability distribution functions II. Aequat. Math. **17**, 19–40 (1978)
22. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
23. Nelsen, R.B., Fredricks, G.A.: Diagonal copulas. In: Distributions with given marginals and moment problems, 121–128 (1997)

Author Index

P. Kulczycki et al. (eds.), *Information Technology and Computational Physics*,
Advances in Intelligent Systems and Computing 462,
DOI 10.1007/978-3-319-44260-0