Studies in Fuzziness and Soft Computing

Lotfi A. Zadeh
Ali M. Abbasov
Ronald R. Yager
Shahnaz N. Shahbazova
Marek Z. Reformat *Editors*

Recent Developments and New Directions in Soft Computing

Springer

Studies in Fuzziness and Soft Computing

Volume 317

Series editor

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

For further volumes:
http://www.springer.com/series/2941

About this Series

The series "Studies in Fuzziness and Soft Computing" contains publications on various topics in the area of soft computing, which include fuzzy sets, rough sets, neural networks, evolutionary computation, probabilistic and evidential reasoning, multi-valued logic, and related fields. The publications within "Studies in Fuzziness and Soft Computing" are primarily monographs and edited volumes. They cover significant recent developments in the field, both of a foundational and applicable character. An important feature of the series is its short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Lotfi A. Zadeh · Ali M. Abbasov
Ronald R. Yager · Shahnaz N. Shahbazova
Marek Z. Reformat
Editors

Recent Developments and New Directions in Soft Computing

Springer

Editors
Lotfi A. Zadeh
Department of Electrical Engineering
and Computer Sciences
University of California
Berkeley, CA
USA

Shahnaz N. Shahbazova
Department of Information Technology
and Programming
Azerbaijan Technical University
Baku
Azerbaijan

Ali M. Abbasov
Minister of Communications
and Information Technologies
of the Republic of Azerbaijan
Baku
Azerbaijan

Marek Z. Reformat
Electrical and Computer Engineering
University of Alberta
Edmonton, AB
Canada

Ronald R. Yager
Machine Intelligence Institute
Iona College
New Rochelle, NY
USA

ISSN 1434-9922 ISSN 1860-0808 (electronic)
ISBN 978-3-319-06322-5 ISBN 978-3-319-06323-2 (eBook)
DOI 10.1007/978-3-319-06323-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940923

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Part I
Decision-Making

Model of the Applicability of Expert System Based on Neural Networks Technology and Hybrid Systems for Decision Making

Ali M. Abbasov and Shahnaz N. Shahbazova

Abstract In this chapter it was justified mathematical model of construction of expert systems and methods of its application in educational process. It is given the mathematical formulation of a number of tasks of application of neural networks and hybrid expert systems in the subsystem of decision-making and evaluation of knowledge. By using of fuzzy logic an original method for controlling the student's knowledge was developed which as maximal closely simulating the behavior of the teacher in the student survey, which combines the power and laconism that was not previously available for automated systems. The proposed method of mathematical processing and designing of educational materials on the basis of linguistic variables allows the designer to simulate any configuration of educational materials is an important step in achieving individual learning.

1 Introduction

Nowadays, there are computer-based training systems, which in one or other form include the individual components of expert systems [1, 2]. In most cases these technologies are used as search methods that would give close correlation with maximum rating issued by teacher and accordingly associated with the improvement of block processing and determination of a grade by the test results, and that they did not disclose the full potential of expert systems.

A. M. Abbasov (✉)
Minister of Information and Communication Technologies of the Republic of Azerbaijan, 33 Z. Aliyeva str, 1000 Baku, Azerbaijan
e-mail: abbasov@mincom.gov.az

S. N. Shahbazova
Department of Information Technology and Programming, Azerbaijan Technical University, 25 H. Cavid Ave, 1073 Baku, Azerbaijan
e-mail: shahbazova@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_1,

The modern software and technical solutions have reached a level of development that can solve the problem of high levels of complexity and create a system, using as a base element of designs of low and medium complexity. The principle of modeling in application to automation of the learning process can be described as a set of simple elements, each of them by performing certain terms of responsibilities in the final system and combined with other elements creates a functionally much more complex environment, which can be called intellectual [3].

One of the developed model of expert systems should be ability to simulate the behavior of the teacher in the learning and knowledge evaluation based on fuzzy + neural + hybrid = expert systems [4].

In addition, the student is also a highly intelligent element of the system, and his analysis and reaction to his behavior should also be carefully considered.

In this chapter is suggested a methodology of designing of the components of an expert system as a set of technologies which are most applicable in tasks of complex nonlinear processes, that can develop on the basis of its intellectual environment for educational purposes.

2 Designing of the Learning Environment Which Provides Control of the Educational Process

One of the main given in this chapter problems are a designing of the learning environment which is absolutely able to function with minimal teacher participation and still provide adequate control of the educational process.

In our case, the overall structure of the expert system includes four main elements: the block of decision-making, knowledge base, database and interface to the external environment.

Decision-making block (DMB) is the core of giving to intelligence of the system. The principles of the functioning of DMB provide applicability of the system as a whole in the given task. All other elements are considered as supporting role.

The *knowledge base* contains the functions controlling by reaction of DMB in the emerging development process of events and represent itself an algorithmic model of decision-making by experts for this area of knowledge. The functions controlling by reaction algorithmizated in the form of fuzzy rules, which are relatively easy to be edited to match the specifics of application. The logistics of information flows DMB is largely depends on the functions reactions to events [5].

The database is a repository of information in any format related to this topic at the moment.

User's interface—this is the method of interaction between the user and the system. In high-power systems, the user is provided an opportunity to ask questions and get answers in ordinary language, but the development of such a system is quite a complicated task, it is easier to create a system in which the questions and answers selected within a given list.

The knowledge base contains famous facts expressed in the form of objects and conditions. In addition to the descriptive representations of reality, it involves the expression of uncertainty—limitations on the accuracy of the facts. In this respect it differs from the traditional database because of his character, not numeric and alphanumeric content. Pre-defined logical rules are used while processing of information of the database [6]. Accordingly, the knowledge base, which represents a higher level of abstraction, which deals with classes of objects are not the objects themselves.

The central component of an expert system is the mechanism that searches in the knowledge base on the rules of rational logic for obtaining solutions [7]. This component is activated by receiving a user request and performs the following tasks:

- compares the information contained in the user's request with the information the knowledge base;
- searches for certain goal or causal connection;
- evaluates the relative certainty of factors based on the relevant factors of trust associated with each factor.

The next component of an expert system is the level of trust. Facts are coming to the knowledge base. The connection between the facts presented by heuristic rules—expressions of declarative knowledge about the connection between objects. Each such rule has a component of the "IF" (background) and a component of "THEN" (the conclusion) that defines the forward and reverse causal relationship [8]. Let us consider the example.

Fact: "If a student has solved all the problems, he will get excellent."
Preconditions: "If a student was excellent, he solved all problems."

The actual approval only probable that means that the degree of certainty is not always absolute. Such statements relative certainty is often based on statistics, probability, or simply subjective preconditions. The common scale and the scale used in Intelligent Information System of Learning and Control Knowledge (IISLCK) range from 0 to 100—the higher the degree of the level of trust [9].

Let us divide the generalized problem of a number of subtasks.

The presentation of the facts in the knowledge base. The knowledge base consists of facts and rules. Facts describe what is known about the domain at the moment.

Rules establish situational, conceptual, causal, or precedent relationship between these facts. We represent the facts, identifying objects by describing their attributes and giving those values or equivalents. The word "object" is meant as a physical object (such as "evaluation", or "number of visits"), and general perceptions ("good" or "high achiever").

The attributes associated with objects that activate objects in the system (for example, "failing grade", "absenteeism"). To organize the expression of facts, we combine them into a pair of "object—the value" by combining object name with the name of the attribute.

The chain lists are using to represent objects in the knowledge base. Each item of this list is called a node and contains fields where information about the object is added. One of the fields serves as a pointer to inform the system where to find the next node in the list. Moreover, each node in the object list has a pointer to a list of values associated with the name of the object.

The coefficient of determination. Thus not all knowledge is strictly defined, the expert system must have methods of processing different degrees of confidence in expression of the given facts, moreover, if such subjective allegations tend to absoluteness.

Rules of expert system. The expert system processes the symbolic representation of reality using heuristic rules and the method of inverse chain. In this method the consultation begins with certain purpose or final result [10].

This approach is the opposite of the method of a direct of the chain, where the argument begins with the definition of the problem. A rule consists of two parts: the preconditions and the conclusion. Both precondition and conclusion are the facts of the knowledge base, expressed in pairs "object - the value". In our system, the rules have the following format [11]:

IF PRECONDITION, THEN CONCLUSION

The meaning of this format is that if the precondition is true, then the conclusion is true. These simple connection of "IF-THEN" represents the nodes on which the inference engine is moving toward the goal. The rule may include a Boolean operator "AND" for the formation for more complex expressions.

Searching solutions—the solver. The knowledge base is a description of the subject area of expert system. Solver is an interpreter of the rules, which uses the facts of this knowledge base for solving the problems. It is accomplished by formulating testing hypotheses, and test them for compliance with the stated purpose. The operator sets the goal of counseling in the form of the object name. The solver uses a set of rules, trying to get the value of the specified target object. The system continues to search for as long as one of the suggested solutions would not be correct.

Statement of the ultimate goals. First of all, the distance needed to be determined, the ultimate goal—what is the result to be achieved, when the expert system will solve the problem. The goal should express an action or event that shows the impact of expert systems on the general course of events. At this level of description the degree of uncertainty can be quite large [12].

Definition of interim targets. For the ultimate goal of each system can have a number of sub-goals—actions or specific problems. There are goals that result in achievement of the ultimate goals. At this level, the purpose can also be expressed as an action or event, but the uncertainty is going to decline. When interim goals are defined, the problem is divided into the sub problems. Each goal is the intended result and proposed program for solving individual problem. In addition to setting the desired results solve the problem, the interim goal serves another important function, because of their specific problem—it may require for their individual solutions or isolated systems of knowledge [13].

Identification of problems. Once the professionals (teachers) identified goals of the expert system by knowledge, types of problems to be solved and the method of approaches of system to their solution become obvious.

Extraction of knowledge. After determining the goals and tasks of the system, team of developers is faced with the problem of efficient extraction of expert knowledge. The most obvious method is a simple survey. The interviewed experts and gaining knowledge from handbooks—this is a direct way, but would take too much time. The use of analogies or models can significantly reduce this work. Here is one of option: to collect file expert solutions related to the problems in the subject area, and then analyze the main rules, which are based on these expert solutions. The third source of information is the direct observation or experimentation. The classic way to extract expert knowledge is going through all the stages of the expertizing process by watching expert's (teacher) behavior in a particular case.

The quality of teaching system depends on the exact definition of the characteristics of the student by several key factors: the results of the current material, mastering the previous material, current moral and psychological state of the student.

The problem of selecting further action is solved by the expert system based on these indicators. This may be a continuation of the teaching of new material, the question of the previous passed questions or completion of training.

Electronic catalog of tested student stores all data according to his ability, passing the test schedule, etc. In addition, the folder also stores personal data of the student.

In the process of working with the program the student can only not to test his knowledge, but also learn. This is achieved not only by the way the question is asked, but also by the presence of all of the questions and comments explanations given by the teacher. Access to the Internet will allow the student looking for information anywhere in the world, in the best libraries, archives, search engines, etc.

By achieving sustainable results in matters of a certain group, student can go to the next level of difficulty of the questions. This transition will allow student not to stop at the results achieved, but also develop further.

The stepping in training gives the necessary time to full mastering and strengthening of the material, and then the transition to a new, more difficult material. Each transition is accompanied by a small test on the previous material with an analysis of its mastering.

To this date, the work has been done on learning and testing services of a group of students at the same time. The shared folders are created on a particular subject or subjects leaded by teacher. With these shared folders a teacher can give tasks and exercises to a group of students, as well as check their solutions and results. The structure of shared folders to simplify the work with groups of students and teacher's access to the working directory of students allows the teacher to deal individually with each student. The same thing happens when recommending re-passing the course, and in a more detailed analysis of errors than it was done in shared directories, etc. The complex of IISLCK allows the teacher to add and edit her material, make corrections according to the latest achievements of science and culture [14].

3 The Functional Design of the Systems of Control Knowledge

One of the perspective ways to improve the functioning of the control systems of technical control of knowledge is the application of complex intelligent computer technologies, in particular, systems based on the diverse of knowledge of hybrid expert systems (ES). In hybrid ES represent different kinds of knowledge as conceptual, expert, factual and relevant different methods of its processing.

The main task in the development of hybrid systems is in the best way to combine different forms of representation and processing methods of knowledge in the processing in decision-making of ES, it means, that actual problem is to investigate the possibilities of optimal connection of different mechanisms of knowledge processing to improve the quality, mobility and efficiency of ES in solving problems and knowledge control in conditions of uncertainty.

The mobility of ES is due to the mobility of knowledge base (KB) and its ability to replenish from different information components (database, bases of expert knowledge (BEK), the base of conceptual knowledge (BCK), dynamic files, etc.), as well as various procedures of conclusion. The concretization of knowledge processing in solving problems decomposes them into accurate and inaccurate, complete and incomplete, static and dynamic, single-valued and multi-valued, etc. In addition, the expert knowledge is inaccurate due to their subjective character. The approximation and multiple meanings of knowledge processing leads to the fact that the ES has a deal not with one, but with several alternative areas. Therefore, the incompleteness of knowledge processing can be used not one, but several sources of knowledge.

The application of fuzzy logic hybrid ES of control knowledge may have at least three implementations:

(1) processing of fuzzy vagueness utterances experts, i.e. when the precondition is fuzzy variables, inference machine—mechanism of extract data thereof;
(2) using matrix of fuzzy connections, determining a number of factors and a lot of preconditions. The matrix contains the fuzzy variables relations, a measure which is represented as a real number [0, 1], and determines the cause of the condition, the transformation is a matrix and factors to a form equations of fuzzy relations, and then the resulting system was solved by the composition of the minimum-maximum;
(3) using of fuzzy conclusions. This approach is most often used in the construction of fuzzy knowledge bases [15].

The application of fuzzy hybrid ES to solve problems and control parameters of knowledge processing extends the capabilities this class of intelligent systems, increasing their flexibility and mobility, allows under equal computational resources of computers to conduct expert evaluation more number of variants, increasing the credibility and accuracy of the evaluation of the results.

In this chapter main principles of construction of neuro-fuzzy hybrid ES is considered with diverse knowledge and analysis of its functioning in conditions of uncertainty of parameters control object (of knowledge processing) with the application as a dynamic knowledge base combined models of neural networks (NN) [16].

In a hybrid neuro-fuzzy ES etalon model (EM) is stored in the knowledge base of knowledge processing form refined in the process of acquiring new knowledge. The real model is formed in a database environment, and communication with the EM via the user's requests. Solving the problem of the designing an intelligent system control quality knowledge processing-based on hybrid ES was produced with taking into account the characteristics of the environment of ES.

Hybrid ES consists of the following parts: a database that stores etalon and factual evidence about the process, the results of their comparison, conceptual, physical and info logical models, knowledge base (KB): static (knowledge is stored in the form of expert knowledge (of products)) as well as formulas, facts, dependency, tables, concepts specific subject area); dynamic (the knowledge is stored combined models of NN in the form of etalon of dynamic processes taking into account the partial or complete uncertainly parameter of control), mechanism of logic inference is based on an algorithm for generating cause and effect network of events functional—structural model; adaptation mechanism to coordinate the work of the database (DB) and KB in the process of logical inference depending on the situation, explaining the mechanism, which is an interpretation of the process of logical inference, planner, coordinating the process of solving the problem; solver for finding effective solutions to positive, negative and mixed statements of problems.

The content, form and algorithms of presentation information of hybrid ES have the possibility of varying depending on the complexity of being modeled situation, the specific and individual characteristics of the user.

This user-expert presents the expert in the knowledge in the form of sets of examples. The internal form of presentation of expert knowledge is a derivation tree. A set of examples is described by attributes and provides examples of the same structure, as defined by its attributes, which can be linked by logical transitions. In this case, the relevant trees of inference are combined in such a way that at the terminal vertex of one tree adds another tree.

Computational Model of ES and DB in solving problems under uncertainty is given in summary form:

$$W = \langle A, D, B, F, H \rangle, \tag{1}$$

where A—the set of attributes DB and KB; D—domains (attribute values of DB and KB); B—a set of functional dependencies defined over the attributes;

F—many descriptions of all types used in the B functional dependencies; H—set of fuzzy relations over a set of attributes A.

The following should considered: in each hybrid ES its defined requirements for the form of knowledge representation, and since they are different (frames, semantic networks, databases, the concepts presents in KB of ES, neural networks, fuzzy logic, genetic algorithms), that, even within a single information space in the hybrid ES, combine different knowledge is difficult. For example, in the hybrid ES diverse knowledge is stored static ES and dynamic knowledge on the state of knowledge by neural networks [17].

The modern information and computer technology (based on the approach OLE-technology) can easily share diverse knowledge within a single information space of a hybrid neuro-fuzzy ES.

It should be noted that the considered approach in the work to the creation of the intelligent systems of control and knowledge in a base of hybrid ES functioning in conditions of uncertainty allows:

- apply actively the diverse knowledge (conceptual, structural, procedural, factual, the rule base with the functions of accessories, rules and fuzzy rules DB, KB, BEK, procedures) with a combination of inference mechanisms for effective solution the problem of determining the student's knowledge;
- summarize and improve the conceptual model of representation of diverse knowledge among relational DB and the managed DBMS and interacting with the core of hybrid ES;
- solve effectively the problem of optimizing and the distribution information streams by individual subsystems ES with diverse knowledge conditions of uncertainty.

The methodology of constructing of diverse knowledge of hybrid neuro-fuzzy ES for the control the student's knowledge presents in conditions of uncertainty includes the following stages:

- The formalization of the domain (the development of a conceptual model).
- The description of knowledge model as individual concepts (knowledge) in KB.
- The formation of KB with the rule base as a managing components of intellectual core
- The description of diverse information to control the student's knowledge in the individual sub-systems of hybrid ES (DB, KB, EKB, a graphical DB the calculated files).
- Selecting the model NN and learning rules.
- Development of a program of fuzzy logic.
- Distribution of information streams between the ES and its individual subsystems.
- Testing of individual subsystems of ES with diverse knowledge
- Testing of hybrid neuro-fuzzy ES.

4 Developing the Mathematical Model of the Educational Process

The mechanism of knowledge and ways of its improvement, have a special meaning, and it is one of the most important modules of the training system, for example, on its base model developed for the full cycle of learning with minimal, or no participation of the teacher. In this case the academic year loses its special importance, since education is absolutely individual, i.e., depending on the student's abilities and efforts aimed at obtaining and the assimilation of knowledge, academic year may continue for 2–3 months and for 2–3 years.

To construct a qualitatively new system of organizing the educational process necessary to solve a number problems to improve the intelligence of especially important modules, such as educational material, construction decision-making block, creation knowledge base, implementation of control systems knowledge [18].

Before developing the mathematical model of the educational process was to select a mathematical tool, based on which will algorithmization all future system. The main requirements were the power, flexibility and simplicity for solving problems containing a large number of uncertain variables. Another component of the problem is the need for constant readiness to modify and improve the kernel of system [19].

All these requirements are answered the mathematical apparatus of fuzzy sets. Due to this, as key elements of knowledge base were used fuzzy rules in the form of easy to algorithmizated logical structures [20]:

$$\begin{array}{c} IF\,\mu_A(x_1)THEN\,\mu_B(y_1) \\ IF\,\mu_A(x_2)THEN\,\mu_B(y_2) \\ \cdots \\ IF\,\mu_A(x_n)THEN\,\mu_B(y_m) \end{array} \tag{2}$$

where $\mu_A()$—selection function of the fuzzy rules according to the incoming signal fuzzy x, $\mu_B()$—the function of selection of the reaction system according fuzzy conclusion y. The sets of A and B are the fuzzy space of possible input signals and output reaction of the system: $A = \{a_1, a_2, a_3,\ldots, a_n\}$, $B = \{b_1, b_2, b_3,\ldots, b_m\}$. Figure 1.

These expressions (2), in essence, a template functions for the logical selection corresponding reaction of the decision-making block to the incoming signal generated by a fuzzy learning system to changing external environment [21].

Applicability of the mathematical apparatus of fuzzy logic for use in an uncertain environment of the educational process, supported by opportunities to demonstrate in practice as developments in the field of intellectual challenges, from the set expert systems to models of artificial intelligence.

Creator (Developer) of Fuzzy Logic Professor Lotfi A. Zadeh in his works, recommends the use of linguistic variables for the development of systems that

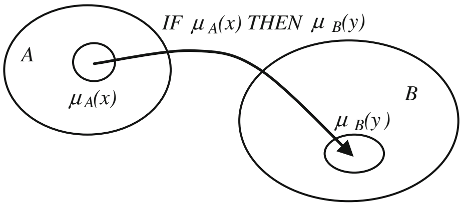


Fig. 1 Method of selection the system reaction $\mu_B(y)$ to receiving an incoming fuzzy signal $\mu_A(x)$

operate with such concepts as human judgments and which are strongly influenced by human factors [22].

Values of linguistic variables are words or phrases of natural language, characterizing the decision taken by a man and feedback reaction to decisions taken by the system.

In the case of the mathematical formalization of the educational process and in particular the object of learning "student", the linguistic variable in the database of student can simply be given as the following linguistic expressions:

$$R_i = \langle M, K, A_Q, R, T \rangle \tag{3}$$

R_i—linguistic suggestion, which characterizes the properties of an object "student" for the special case i as an indicator of knowledge in some a learning course K, educational material M, at response A to a question Q, in which evaluated R, with commentary T.

The advantage of this model, formalizing the characteristics of indicators of students' knowledge, is that every word in your the linguistic variables easy enough to be described almost any situation not provided for in advance, on the occurrence of which can be given by the corresponding reaction decisions making block.

For example, the variable M—describes a set of educational materials, can contain the name of a specific educational material, and the whole field of knowledge.

Thus, when a new discipline (e.g. ekoinformatics) or the merger of several learning courses (e.g. biochemistry, bioinformatics) modification of the fundamental principles of construction system is not required and appropriate educational material to the establishment of new training courses K_{new}, or transferring the existing subsets of the educational materials into a new set of M_{new} (4).

$$M_{new} = \left\{ \begin{array}{r} K_{new},\, K_{i1} \in M_k,\, K_{i1} \in M_k, \\ K_{i2} \in M_i,\, K_{i3} \in M_j) \end{array} \right\} \tag{4}$$

As seen from Fig. 2, the application of linguistic variables in a semi-structured environments, which is the educational process and the educational system as a whole, a mandatory requirement, otherwise the newly created system of education immediately obsolete, or will have limited applicability [23].

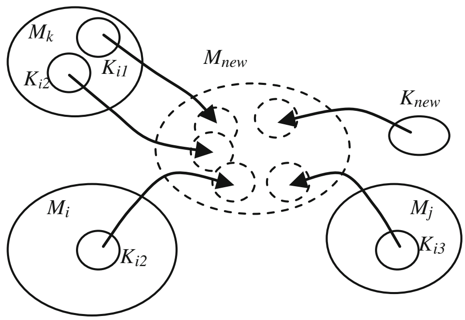


Fig. 2 Model for creating a new value of linguistic variable M_{new}—a new educational material

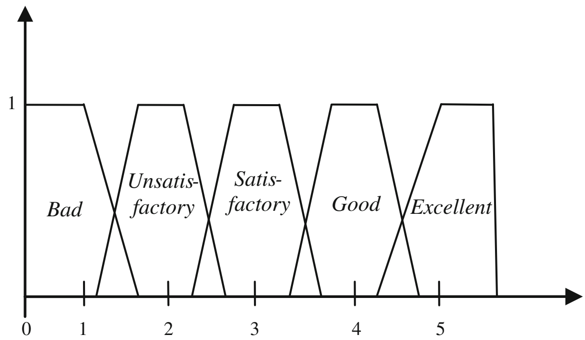


Fig. 3 Membership function of models of the evaluation of the knowledge of student on the basis of linguistic variable with values: it is bad, it is unsatisfactory, it is satisfactory, it is good, excellently

The next problem, which is chosen to match the mathematical apparatus needs improvement is the method evaluation student knowledge. If in the traditional system of evaluations, the student receives a natural number or letter as a characteristic of his knowledge in some area of educational materials, in this chapter, evaluation system is given to a fuzzy form. Figure 3 [24].

On the membership function evaluation "bad" characterized by lack of knowledge, as "excellent"—that the student possesses, the maximum volume of knowledge and skills on the studied materials [25].

Thus, evaluation of knowledge is characterized as a form of uncertainty, since it is not amenable to analysis and measurement with arbitrarily high accuracy.

As in any expert system, the core of the educational system is a knowledge base. Analyzed the existing methods for constructing knowledge bases, the choice was made in favor of fuzzy knowledge base, which provided a strong uncertainty shows a high noise immunity for a fairly simple procedure, modifications and additional study by increasing the base of fuzzy rules.

Classically, in the base of the fuzzy rules stored knowledge and experience of experts, but in an environment of educational purpose is impossible to describe the process of study and knowledge control.

Ways to determine student knowledge, to identify areas with insufficient understanding, defining moment of students' readiness to transition to the new curriculum is a multifactorial intellectual task. To solve it, a model was created, which handles two types of the systems of states. Each state is served by having the classical form, fuzzy rules [26]:

- Reaction to the situation—is used, for example, during the procedure of control of students' knowledge.

$$\begin{aligned} &IF\ input\ (current\ situation) \\ &\quad Then\ output\ (system\ reaction) \end{aligned} \tag{5}$$

  Has tactical character, for example, decision-making on sample of level of complexity of a following set question, by results of answers to the previous questions.
- Update on the results of learning procedures—used, for example, when updating information on knowledge and misconceptions students, and refinement of the curriculum—individual for each student

$$\begin{aligned} &IF\ input\ (current\ result) \\ &\quad Then\ output\ (database\ update) \end{aligned} \tag{6}$$

Has a strategic character, because, decides on the level of understanding of the relevant section of the course, and should be able to answer the question of whether the transition to the next section learning course [27].

The effectiveness of this model and the quality of processing the input information based on the use of fuzzy rules, and for example, the following fuzzy rules are applied in a block survey of the student.

The criterion for selection next question is the result of the answer to a previous question. The basic formula of the algorithm is described below:

- ……..

1. IF the Answer is correct THEN
2. GeneralWeight +=WeightQuestionI; KategoryQuestion[i] ++;
3. ELSE
4. GeneralWeight1 +=WeightQuestionI; KategoryQuestion1[i] ++;
5. END IF
6. PITCH ++;
7. I = FUZZY (IndexStudent, Subject, Pitch, GeneralWeight, GeneralWeight1, KategoryQuestion, KategoryQuestion1);
8. Transition in a Data Base to I Question
9. Reading, Formatting as a kind of HTML by means of Jscript
10. Expectation of variant of the student's answer
11. Opening the data base of the answers, in order to choose the correct answer
12. Comparison with the student's variant and preservation of an outcome in a variable named Answer
13. Jump to line 1

- ……..

Below are the fuzzy rules applied in the block of the survey of a student—there are two fuzzy rules [9]:

(1) IF the previous question the student answered correctly, THEN in this case, the student is given a more difficult question;
(2) If the previous question the student answered incorrectly, THEN the student is given a simple question.

Mathematically, they can be described by the following formula [28]:

$$V_i = \begin{cases} A_{i-1} \in A_{true} => \mu_{A-}(A_{i-1}, A_{i-2}, \ldots, A_0) \\ A_{i-1} \notin A_{true} => \mu_{A+}(A_{i-1}, A_{i-2}, \ldots, A_0) \end{cases} \tag{7}$$

Where V_i—question, which to be asked the following, A_{true}—set of correct answers, μ_{A-}—function of selection question with the incorrect answer (8), μ_{A+}—function of selection question in correct answer (9), ($A_{i\text{-}1}$, $A_{i\text{-}2}$, …, A_0)—set of answers received for this session of the control of knowledge.

$$\mu_{A-}\left(A_{temp}\right) = V\left(\frac{\begin{cases} \max(A_{temp} \notin A_{true}) \\ A_{\max} \end{cases} + \begin{cases} \min(A_{temp} \in A_{true}) \\ 0 \end{cases}}{2} \pm 2\%\right) \tag{8}$$

$$\mu_{A+}\left(A_{temp}\right) = V\left(\frac{\begin{cases} \max(A_{temp} \in A_{true}) \\ A_{\max} \end{cases} + \begin{cases} \max(A_{temp} \notin A_{true}) \\ A_{\max} \end{cases}}{2} \pm 2\%\right) \tag{9}$$

where A_{temp}—temporary set containing a set of values answers received currently by procedures of control of knowledge $A_{temp} = \{A_{i\text{-}1}, A_{i\text{-}2}, \ldots, A_0\}$, $\begin{cases} \max(A_{temp} \notin A_{true}) \\ A_{\max} \end{cases}$—selection answer with the maximum value from the set A_{temp} does not belong to the set of correct answers, or if no item, value is taken $A_{\max}$—that responds to maximum by a difficulty of the question of knowledge area that perform the procedure control of knowledge, $\begin{cases} \min(A_{temp} \in A_{true}) \\ 0 \end{cases}$—selection answer with a minimum value of the set A_{temp} belonging to the set of correct answers, or if no item is taken value 0—as the minimum possible answer to the easiest question, $\begin{cases} \max(A_{temp} \in A_{true}) \\ A_{\max} \end{cases}$—selection answer with the maximum value from the set A_{temp} belonging to the set of correct answers, or if no item, value is taken $A_{\max}$—that responds to maximum by a difficulty of the question of knowledge area that perform the procedure control of knowledge, ±2 %—is the range within which the selected number of which make an element of chance in selection of by the difficulty of questions, an additional consequence of which is the uniqueness of the lists of questions given to students, even if the order is right and wrong answers to these questions have the same, function $V()$—a procedure for selection of question from the database to by the difficulty as maximum close to the input value [29].

Difference of formulas $\mu_{A-}(A_{temp})$ (8) и $\mu_{A+}(A_{temp})$ (9) is their asymmetry, which is manifested in the fact that, following the correct answer is chosen much more difficult question, and after wrong answer choose question just a little easier. Thus, on one hand the student is motivated by the fact that the show its best knowledge, on the other hand the number of questions to determine the level of student knowledge need to ask a lot less questions than traditional testing.

Applying only these two rules allows qualitatively improve characteristics of control systems knowledge, as it allows to evaluate not only counting the number of correct and incorrect answers, but also provides, by the analysis given in the course of testing questions, the general picture of students' knowledge and the level of understanding, given the educational material.

5 Conclusion

In this chapter proposed the mathematical formulations of some problems of application of neural networks and hybrid expert systems in subsystem of decision-making and evaluation of knowledge.

Using apparatus of fuzzy logic development of an original method for controlling the student's knowledge as closely simulates the behavior of a teacher during the survey a student who combines the power and laconism, not previously accessible to automated systems. Based on improvements in several key aspects of the learning environment, obtained results showing qualitative improvement of quality fixing study materials achieved through the application of algorithm of knowledge control developed on the basis of knowledge mathematical using apparatus of fuzzy logic.

Proposed a fundamental principle of construction of such systems, which can be applied at practically any model of learning. The proposed method of mathematical design and projecting of learning materials based on linguistic variables, allows the developer to simulate any configuration of the learning materials is an important step in achieving individual learning, because now the development of individual student's plan is a technical problem selection of learning courses specific profession, for which the student showed insufficient level knowledge.

References

1. Abbasov, A.M.: Information boom: new trends and expectations, Springer series title: studies in fuzziness and soft computing. Soft Comput.: State Art Theory Novel Appl. **291**, 1–12 (2012)
2. Kandel, A., Langholz, G.: Fuzzy Control Systems, pp. 159–187. CRC Press, Boca Raton (1993)

3. Laurene, V.F.: Fundamentals of Neural Networks: Architectures, Algorithms And Applications, pp. 103–121. Prentice Hall, Englewood Cliffs (1993)
4. Zadeh, L.A., Klir, G.J., Yuan, B.: Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh, pp. 60–69 (1996)
5. Bernshteyn, L.S., Bojenyuk, A.V.: Fuzzy Models of Decision Making: Deduction, Induction, Analogy, pp. 78–99. Univ Tsure, Taganrog (2001)
6. Cordon, O., Herrera, F.: Linguistic Modeling by Hierarchical Systems of Linguistic Rules/ Technical Report # DECSAI—990 114, Department of Computer Science and A. I., University of Granada, July, 1999, pp. 187–215 (1999)
7. Barsky, A.B.: Neural Networks: Recognition, Management, Decision-Making, pp. 30–63. Finance and Statistics, Moscow (2004)
8. Shahbazova, Sh., Freisleben, B.: A network-based intellectual information system for learning and testing. In: Fourth International Conference on Application of Fuzzy Systems and Soft Computing, Siegen, Germany, pp. 308–313 (2000)
9. Shahbazova, Sh.: Applied research in the field of automation of Learning and knowledge control, SPRİNGER Series Title: Studies in Fuzziness and Soft Computing. Soft Comput.: State Art Theory Novel Appl. **14**, 223–240 (2012)
10. Ledeneva, T.M.: Fuzzy Information Processing: A Tutorial/TM Ledeneva, pp. 212–233. Voronezh State University, Voronezh (2006)
11. Bellman, R., Zadeh, L.A.: Decision-Making in Ambiguous Circumstances, Issues Analysis And Decision-Making, pp. 180–199. Springer, New York (1976)
12. Novak, V., Perfilieva, I.: Mochkorzh Mathematical Principles of Fuzzy Logic, Trans. from English, Averkina, M. (ed.). FIZMATLIT, 209–252 (2006)
13. Borisov, VV, Kruglov, VV, Fedulov, AS.: Fuzzy Models and Networks, pp. 224–284. Hot line—Telecom, Moscow (2007)
14. Gorbunova, L.G.: On the realization of the rating system in pedagogical high schools. In: Proceedings of 2nd International Technical Conference “University Education”, Penza, 1998, Part 1, pp. 105–106 (1998)
15. Nikravesh, M., Zadeh, L.A., Kacprzyk, J.: Soft Computing for Information Processing and Analysis, pp. 93–99. Springer, New York (2005)
16. Heaton, J.: Introduction to Neural Networks for C #, 2nd edn, pp. 224–231. Heaton Research, USA (2008)
17. Hanss, M.: Applied Fuzzy Arithmetic: An Introduction with Engineering Applications, 1st edn, pp. 100–116, 139–147. Springer, Berlin (2004)
18. Zadeh, L.A.: The Concept of Linguistic Variable and Its Application to the Adoption of Approximate Solutions, pp. 140–164. Springer, New York (1976)
19. Shahbazova, Sh.: Application of fuzzy sets for control of student knowledge. Appl. Comput. Math. Int. J. **10**(1), 195–208 (2008). ISSN 1683-3511 (Special Issue on Fuzzy Set Theory and Applications)
20. Zadeh, L.A., Kacprzyk, J.: Fuzzy Logic for the Management of Uncertainty, 1st edn, pp. 75–84. Wiley-Interscience, New York (1992)
21. Shahbazova, Sh.: Development of the knowledge base learning system for distance education. Int. J. Intell. Syst. **27**(4), 343–354 (2012)
22. Zadeh, L.A.: The new approach to the analysis of difficulty systems and decision processes, Mathematics Today, Knowledge, pp. 23–37 (1974)
23. Shahbazova, Sh: Investigation of the basic problems and trends of traditional education systems. Int. J. Technol. Manage. Inform. Probl. Ukraine **3**, 110–117 (2013)
24. Shahbazova, Sh: Simulating the behavior of the teacher, the use of expert systems in the field of educational systems, control systems and machines. J. Inst. Cybern. Glushkov Nat. Acad. Sci. Ukraine **3**, 68–75 (2012)
25. Yager, R.: Fuzzy Sets and Theory of Possibilities: Recent advances, pp. 391–405. Radio and Communications, Moscow (1986)

26. Shahbazova, Sh.: Functional design of the control of knowledge on base of fuzzy logic. In: International Conference on Application of Information and Communication Technology and Statistics in Economy and Education, Sofia, Bulgaria, pp. 24–31 (2012)
27. Bouchon-Meunier, B., Yager, R.R.: Fuzzy Logic and Soft Computing (Advances in Fuzzy Systems: Application and Theory), pp. 84–93, 103–119 World Scientific, Singapore (1995)
28. Shahbazova, Sh.: Application of fuzzy sets for control of student knowledge. Appl. Comput. Math. Int. J. **10**(1),195–208 (2011). ISSN 1683-3511 (Special Issue on Fuzzy Set Theory and Applications)
29. Shahbazova, Sh.: Decision-making in determining the level of knowledge of students in the learning process under uncertainty. Informatica Int. J. Comput. Inform. **37**(3), 339–344 (2013). Print edition ISSN: 0350-5596

Group Multiple Criteria Decision Making: Multiset Approach

Alexey B. Petrovsky

Abstract The chapter describes a new approach to solve problems of group multiple criteria decision making. New methods for group sorting and ordering objects, presented with many quantitative and qualitative attributes, are based on the theory of multiset metric spaces. The suggested techniques were applied to the expertise of R&D projects in the Russian Foundation for Basic Research. For selection of competitive applications and multiple criteria evaluation of project efficiency, several experts evaluated projects upon many verbal criteria.

1 Introduction

Sorting objects into several classes and ordering objects by their properties are the typical problems of multiple criteria decision making (MCDM), pattern recognition, data mining, and other areas. These problems are formulated as follows. Let $A_1, \ldots, A_n$ be a collection of objects, which are described by m attributes $Q_1, \ldots, Q_m$. Every attribute has its own scale $X_s = \{x_s^1, \ldots, x_s^{h_s}\}$, $s = 1, \ldots, m$, grades of which may be numerical, symbolic or verbal, discreet or continuos, nominal or ordinal. Ordinal grades are supposed to be ordered from the best to the worst. Attributes may have different relative importance (weights). The attribute list depends on the aim of decision analysis. It is required to range all multi-attribute

This work is partially supported by the Russian Academy of Sciences, Research Programs "Information, Control and Intelligent Technologies and Systems", "Intelligent Information Technologies, Systems Analysis and Automation"; the Russian Foundation for Basic Research, Projects No 11-07-00398, 14-07-00916; the Russian Foundation for Humanities, Projects No 12-02-00525, 12-33-10608, 14-32-13004.

A. B. Petrovsky (✉)
Institute for Systems Analysis, Russian Academy of Sciences, Moscow, Russia
e-mail: pab@isa.ru

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_2,

objects or assign every object to one of the given classes (categories) C_1, …, C_g, describe and interpret the properties of these classes of objects. The number g of object classes can be arbitrary or predefined, and the classes can be ordered or unordered.

In the case of group decision making, one and the same multi-attribute object is represented in k versions or copies, which are usually distinguished by attribute values. For example, object characteristics have been measured in different conditions or in different ways, either several experts independently evaluated objects upon many criteria.

One of basis points in MCDM area [4–6, 8–11, 14–21] is preferences of decision maker (DM) and/or expert. The person expresses his/her preferences when he/she describes properties and characteristics of the analyzed problem, compares decision alternatives, estimates the choice quality. Preferences may be represented as decision rules of mathematical, logical and/or verbal nature and explained with any language. While solving the problem, a person may behave inconsistently, make errors and contradictions. In the case of individual choice, the consistency of subjective preferences is postulated. In order to discover and correct possible inconsistent and contradictory judgments of a single DM, special procedures have to be included in MCDM methods [10].

A collective choice of several independent actors is more complicate and principally different due to a variety and inconsistency of many subjective preferences. Every DM may have his/her own personal goals, interests, valuations and information sources. As a result, individual subjective judgements of actors may be similar, concordant or discordant. Usually, in MCDM techniques, one tries to avoid possible inconsistencies and contradictions between judgements of several persons. Often many diverse points of view are replaced with a single common preference that is aggregated mostly all individual opinions. But individual preferences may be coordinated not always. Nevertheless, most of the decision methods do not pay a consideration to contradictions and inconsistencies in DMs' preferences.

In this chapter, we consider methods for group ordering and classifying objects, which are presented with many numerical and/or verbal attributes and may exist in several copies. These methods are based on the methodology of group verbal decision analysis and the theory of multiset metric spaces [9, 10, 13–18]. The suggested techniques were applied to real-life case studies in various practical areas, where several experts estimated objects upon many qualitative criteria.

2 Representation of Multi-Attribute Objects

In MCDM problems, a multi-attribute object A_i is represented as a vector or tuple (cortege) $\boldsymbol{x}_i = (x_{i1}^{e_1}, \ldots, x_{im}^{e_m})$ in the Cartesian m-space $X_1 \times \cdots \times X_m$ of attributes scales. Often qualitative variables are transformed in the numerical ones by one or another way, for example, using the lexicographic scale or fuzzy membership functions [6, 20, 21]. Unfortunately, the admissibility and validity of similar transformations of qualitative data into quantitative ones are not always justified. In methods of verbal decision analysis [9, 10], objects are described by qualitative variables without a transformation into numerical attributes.

The situation becomes more complicated when one and the same object exist in multiple versions or copies. Then, not one vector/cortege but a group of vectors/corteges corresponds to each object. So, an object A_i is represented now as a collection of k vectors/corteges $\left\{\boldsymbol{x}_i^{(1)},\ldots,\boldsymbol{x}_i^{(k)}\right\}$ where $\boldsymbol{x}_i^{(j)}=(x_{i1}^{e_1(j)},\ldots,x_{im}^{e_m(j)})$, $j=1,\ldots,k$. And this group should be considered and treated as a whole in spite of a possible incomparability of separate vectors/corteges $\boldsymbol{x}_i^{(j)}$. A collection of multi-attribute objects can have an overcomplicated structure that is very difficult for analysis.

In many group decision methods, a collection of k vectors $\left\{\boldsymbol{x}_i^{(1)},\ldots,\boldsymbol{x}_i^{(k)}\right\}$ is replaced usually by a single vector $\boldsymbol{y}_i$. Typically, this vector $\boldsymbol{y}_i$ has the components derived by averaging or weighting the values of attributes of all members of the group, or this vector is to be the mostly closed to all vectors within a group or to be the center of group. Note, however, that features of all initial vectors $\boldsymbol{x}_i^{(1)},\ldots,\boldsymbol{x}_i^{(k)}$ could be lost after such replacement. The operations of averaging, weighing, mixing and similar data transformations are mathematically incorrect and unacceptable for qualitative variables. Thus, a group of objects, represented by several tuples, can not be replaced by a single tuple. So, we need new tools of aggregation and work with such objects.

Let us consider another way for representing multi-attribute objects. Define the combined attribute scale or the hyperscale $X=X_1\cup\cdots\cup X_m$ that is a set consisted of m attribute (criteria) scales $X_s=\{x_s^{e_s}\}$. Now represent an object A_i as the following set of repeating attributes

$$\begin{aligned}\boldsymbol{A}_i=\{&k_{\boldsymbol{A}i}\left(x_1^1\right)\circ x_1^1,\ldots,k_{\boldsymbol{A}i}\left(x_1^{h_1}\right)\circ x_1^{h_1},\ldots,\\&k_{\boldsymbol{A}i}\left(x_m^1\right)\circ x_m^1,\ldots,k_{\boldsymbol{A}i}\left(x_m^{h_m}\right)\circ x_m^{h_m}\}.\end{aligned}\tag{1}$$

Here $k_{\boldsymbol{A}i}(x_s^{e_s})$ is a number of attribute $x_s^{e_s}$, which is equal to a number of experts evaluated the object A_i with the criterion estimate $x_s^{e_s}$, or a number of different conditions or instruments used to measure an attribute value $x_s^{e_s}$; the sign $\circ$ denotes that there are $k_{\boldsymbol{A}i}(x_s^{e_s})$ copies of attribute $x_s^{e_s}\in X_s$ within the description of object A_i.

Thus, the object A_i is represented now as a set of many repeating elements x (for instance, attribute values $x_s^{e_s}$) or as a multiset $\boldsymbol{A}_i=\{k_{\boldsymbol{A}i}(x_1)\circ x_1,\,k_{\boldsymbol{A}i}(x_2)\circ x_2,\ldots\}$ over the ordinal (crisp) set $X=\{x_1,\ x_2,\ \ldots\}$ that is defined by a multiplicity function $k_{\boldsymbol{A}}:X\to\mathbf{Z}_+=\{0,1,2,3,\ldots\}$ [2, 7, 13]. A multiset $\boldsymbol{A}_i$ is said to be finite when all numbers $k_{\boldsymbol{A}i}(x)$ are finite. Multisets $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be equal ($\boldsymbol{A}=\boldsymbol{B}$), if $k_{\boldsymbol{A}}(x)=k_{\boldsymbol{B}}(x)$. A multiset $\boldsymbol{B}$ is said to be included in a multiset $\boldsymbol{A}$ ($\boldsymbol{B}\subseteq\boldsymbol{A}$), if $k_{\boldsymbol{B}}(x)\leq k_{\boldsymbol{A}}(x)$, $\forall x\in X$.

There are defined the following operations with multisets:

- union $\boldsymbol{A} \cup \boldsymbol{B}$, $k_{\boldsymbol{A} \cup \boldsymbol{B}}(x) = \max[k_{\boldsymbol{A}}(x), k_{\boldsymbol{B}}(x)]$;
- intersection $\boldsymbol{A} \cap \boldsymbol{B}$, $k_{\boldsymbol{A} \cap \boldsymbol{B}}(x) = \min[k_{\boldsymbol{A}}(x), k_{\boldsymbol{B}}(x)]$;
- arithmetic addition $\boldsymbol{A} + \boldsymbol{B}$, $k_{\boldsymbol{A}+\boldsymbol{B}}(x) = k_{\boldsymbol{A}}(x) + k_{\boldsymbol{B}}(x)$;
- arithmetic subtraction $\boldsymbol{A} - \boldsymbol{B}$, $k_{\boldsymbol{A}-\boldsymbol{B}}(x) = k_{\boldsymbol{A}}(x) - k_{\boldsymbol{A} \cap \boldsymbol{B}}(x)$;
- symmetric difference $\boldsymbol{A} \Delta \boldsymbol{B}$, $k_{\boldsymbol{A} \Delta \boldsymbol{B}}(x) = |k_{\boldsymbol{A}}(x) - k_{\boldsymbol{B}}(x)|$;
- multiplication by a scalar $b \cdot \boldsymbol{A}$, $k_{b \cdot \boldsymbol{A}}(x) = b \cdot k_{\boldsymbol{A}}(x)$, $b > 0$;
- arithmetic multiplication $\boldsymbol{A} \cdot \boldsymbol{B}$, $k_{\boldsymbol{A} \cdot \boldsymbol{B}}(x) = k_{\boldsymbol{A}}(x) \cdot k_{\boldsymbol{B}}(x)$;
- direct product $\boldsymbol{A} \times \boldsymbol{B}$, $k_{\boldsymbol{A} \times \boldsymbol{B}}(x_i, x_j) = k_{\boldsymbol{A}}(x_i) \cdot k_{\boldsymbol{B}}(x_j)$, $x_i \in \boldsymbol{A}$, $x_j \in \boldsymbol{B}$.

A collection $A_1, \ldots, A_n$ of multi-attribute objects may be considered as points in the multiset metric space $(L(\boldsymbol{Z}), d)$ with the following types of distances

$$\begin{aligned} d_{1p}(\boldsymbol{A},\boldsymbol{B}) &= [m(\boldsymbol{A}\Delta\boldsymbol{B})]^{1/p}; d_{2p}(\boldsymbol{A},\boldsymbol{B}) = [m(\boldsymbol{A}\Delta\boldsymbol{B})/m(\boldsymbol{Z})]^{1/p}; \\ d_{3p}(\boldsymbol{A},\boldsymbol{B}) &= [m(\boldsymbol{A}\Delta\boldsymbol{B})/m(\boldsymbol{A}\cup\boldsymbol{B})]^{1/p}, \end{aligned} \tag{2}$$

where $p \geq 0$ is an integer, the multiset $\boldsymbol{Z}$ is the so-called maximal multiset with $k_{\boldsymbol{Z}}(x) = \max_{\boldsymbol{A}} k_{\boldsymbol{A}}(x)$, and $m(\boldsymbol{A})$ is a measure of multiset $\boldsymbol{A}$.

Multiset measure m is a real-valued non-negative function defined on the algebra of multisets $L(\boldsymbol{Z})$. The maximal multiset $\boldsymbol{Z}$ is the unit and the empty multiset $\emptyset$ is the zero of the algebra. A multiset measure m has the following properties: $m(\boldsymbol{A}) \geq 0$, $m(\emptyset) = 0$; strong additivity $m(\sum_i \boldsymbol{A}_i) = \sum_i m(\boldsymbol{A}_i)$; weak additivity $m(\cup_i \boldsymbol{A}_i) = \sum_i m(\boldsymbol{A}_i)$ for $\boldsymbol{A}_i \cap \boldsymbol{A}_j = \emptyset$; weak monotony $m(\boldsymbol{A}) \leq m(\boldsymbol{B}) \Longleftrightarrow \boldsymbol{A} \subseteq \boldsymbol{B}$; symmetry $m(\boldsymbol{A}) + m(\overline{\boldsymbol{A}}) = m(\boldsymbol{Z})$; continuity $\lim_{i \to \infty} m(\boldsymbol{A}_i) = m(\lim_{i \to \infty} \boldsymbol{A}_i)$; elasticity $m(b \cdot \boldsymbol{A}) = bm(\boldsymbol{A})$.

The distances $d_{2p}(\boldsymbol{A}, \boldsymbol{B})$ and $d_{3p}(\boldsymbol{A}, \boldsymbol{B})$ satisfy the normalization condition $0 \leq d(\boldsymbol{A}, \boldsymbol{B}) \leq 1$. $d_{3p}(\emptyset, \emptyset) = 0$ by the definition, while the distance $d_{3p}(\boldsymbol{A}, \boldsymbol{B})$ is undefined for $\boldsymbol{A} = \boldsymbol{B} = \emptyset$. For any fixed p, the metrics d_{1p} and d_{2p} are the continuous and uniformly continuous functions, the metric d_{3p} is the piecewise continuous function almost everywhere on the metric space for any fixed p.

The proposed metric spaces are the new types of spaces that differ from the well-known ones [3]. The general distance $d_{1p}(\boldsymbol{A}, \boldsymbol{B})$ is analogues of the Hamming-type distance between objects, which is traditional for many applications. The completely averaged distance $d_{2p}(\boldsymbol{A}, \boldsymbol{B})$ characterizes a difference between two objects related to common properties of all objects as a whole. And the locally averaged distance $d_{3p}(\boldsymbol{A}, \boldsymbol{B})$ reflects a difference related to properties of only both objects. In the case of sets for $p = 1$, $d_{11}(A, B) = m(A \Delta B)$ is called the Fréchet distance, $d_{31}(A, B) = m(A \Delta B)/m(A \cup B)$ is called the Steinhaus distance [3].

The measure $m(\boldsymbol{A})$ of multiset $\boldsymbol{A}$ may be determined in the various ways, for instance, as a linear combination of multiplicity functions: $m(\boldsymbol{A}) = \sum_s w_s k_{\boldsymbol{A}}(x_s^{e_s})$, $w_s > 0$. In this case, for example, the Hamming-type distance for $p = 1$ has the form

$$d_{11}(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A}\Delta\boldsymbol{B}) = \sum_{s=1}^{m} w_s \sum_{e_s=1}^{h_s} |k_{\boldsymbol{A}}(x_s^{e_s}) - k_{\boldsymbol{B}}(x_s^{e_s})|,$$

where $w_s > 0$ is a relative importance of the attribute Q_s. Various properties of multisets and multiset metric spaces are considered and discussed in [13].

3 Method of Group Ordering Multi-Attribute Objects

The method ARAMIS (Aggregation and Ranking Alternatives nearby the Multi-attribute Ideal Situations) is developed for group ordering multi-attribute objects that is based on preference aggregation [14, 17]. This method does not require pre-construction of individual rankings objects. Let us represent an object A_i that is described by many repeated quantitative and/or qualitative attributes as a multiset (1). Consider multi-attribute objects A_1, …, A_n as points of multiset metric space ($L(\boldsymbol{Z})$, d).

There are two (may be hypothetical) objects with the highest and lowest estimates by all attributes/criteria Q_1, …, Q_m. These are the best object A^+ and the worst object A^-, which can be represented as the following multisets in a metric space

$$\begin{aligned}\boldsymbol{A}^+ &= \{k \circ x_1^1, 0, \ldots, 0, k \circ x_2^1, 0, \ldots, 0, \ldots, k \circ x_m^1, 0, \ldots, 0\},\\ \boldsymbol{A}^- &= \{0, \ldots, 0, k \circ x_1^{h_1},\ 0, \ldots, 0, k \circ x_2^{h_2}, \ldots,\ 0, \ldots, 0, k \circ x_m^{h_m}\},\end{aligned} \tag{3}$$

where k is a number of experts or instrument techniques. These objects are called also as the ideal and anti-ideal situations or referent points. So, all objects may be arranged with respect to closeness to the best object A^+ or the worst object A^- in the multiset metric space ($L(\boldsymbol{Z})$, d).

The descending arrangement of multi-attribute objects with respect to closeness to the best object A^+ is constructed in the following way. An object A_i is said to be more preferable than an object A_j ($A_i \succ A_j$), if a multiset $\boldsymbol{A}_i$ is closer to the multiset $\boldsymbol{A}^+$ than a multiset $\boldsymbol{A}_j$, that is $d(\boldsymbol{A}^+, \boldsymbol{A}_i) < d(\boldsymbol{A}^+, \boldsymbol{A}_j)$ in the multiset metric space ($L(\boldsymbol{Z})$, d). The ascending arrangement of multi-attribute objects with respect to farness to the worst object A^- is constructed analogously. The final ranking multi-attribute objects is constructed as a combination of the descending and ascending arrangements. All objects can be also ordered in accordance with the index $l^+(A_i) = d(\boldsymbol{A}^+, \boldsymbol{A}_i)/[d(\boldsymbol{A}^+, \boldsymbol{A}_i) + d(\boldsymbol{A}^-, \boldsymbol{A}_i)]$ of relative closeness to the best object.

4 Method of Group Clustering Multi-Attribute Objects

Cluster analysis is a widely used approach to study the natural grouping large collections of objects and relationships between them. In clustering or classifying multi-attribute objects without a teacher, the association of objects into groups is based on their differences or similarities, which are estimated by a proximity of objects considered as points of attribute space. The principal features of cluster analysis are as follows: the choice of distance between objects in the attribute space; the choice of algorithm for grouping objects; a reasonable interpretation of the formed groups. A selection of the attribute space and the metric type depends on the properties of the analyzed objects. For the objects with manifold attributes, the most adequate is a representation as multisets and use of the metric space $(L(\boldsymbol{Z}), d)$ of measurable multisets with the basic, completely or locally averaged metric.

Traditionally, a cluster is formed as a set-theoretic union of the closest objects [1]. New operations under multisets open new possibilities for aggregation of multi-attribute objects. For example, a group (class) C_t, $t = 1, \ldots, g$ of objects can be obtained as the sum $\boldsymbol{Y}_t = \sum_i \boldsymbol{A}_i$, $k_{\boldsymbol{Y}t}(x_j) = \sum_i k_{\boldsymbol{A}i}(x_j)$, union $\boldsymbol{Y}_t = \cup_i \boldsymbol{A}_i$, $k_{\boldsymbol{Y}t}(x_j) = \max_i k_{\boldsymbol{A}i}(x_j)$ or intersection $\boldsymbol{Y}_t = \cap_i \boldsymbol{A}_i$, $k_{\boldsymbol{Y}t}(x_j) = \min_i k_{\boldsymbol{A}i}(x_j)$ of multisets $\boldsymbol{A}_i$ describing the objects, either as a linear combination of multisets $\boldsymbol{Y}_t = \sum_i b_i \cdot \boldsymbol{A}_i$, $\boldsymbol{Y}_t = \cup_i b_i \cdot \boldsymbol{A}_i$ or $\boldsymbol{Y}_t = \cap_i b_i \cdot \boldsymbol{A}_i$. When a group C_t of objects is formed by addition of multisets, all of the properties of all objects within a group are aggregated. While forming group C_t of objects by union or intersection of multisets, the best properties (maximal values of attributes) or, respectively, the worst properties (minimal values of attributes) of individual members of a group are strengthened.

In order to generate groups of objects, the following typical approaches are used in clustering techniques: (1) minimize the difference (maximize the similarity) between objects within a group (2) maximize the difference (minimize the similarity) between groups of objects. We assume, for simplicity, that distinctions between objects within the group, between some object and the group of objects, between groups of objects are determined in the same manner and given by one of the above distances d (2).

Consider basic ideas of cluster analysis of multi-attribute objects $A_1,\ldots,A_n$ represented as multisets $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$. Hierarchical clustering, when a number of the generated clusters is not defined in advance, consists of the following major stages.

Step 1 Put $g = n$, g is the number of clusters, n is the number of objects A_i. Then each cluster C_i consists of a single object A_i, and multisets $\boldsymbol{Y}_i = \boldsymbol{A}_i$ for all $i = 1,\ldots,g$.

Step 2 Calculate the distances $d(\boldsymbol{Y}_p, \boldsymbol{Y}_q)$ between all possible pairs of multiset represented clusters C_p and C_q for all $1 \leq p, q \leq g$, $p \neq q$ using one of the metrics d (2).

Step 3 Find the closest pair of clusters C_u and C_v such that $d(\boldsymbol{Y}_u, \boldsymbol{Y}_v) = \min_{p,q} d(\boldsymbol{Y}_p, \boldsymbol{Y}_q)$, and form a new cluster C_r, which represented as a sum $\boldsymbol{Y}_r = \boldsymbol{Y}_u + \boldsymbol{Y}_v$, union $\boldsymbol{Y}_r = \boldsymbol{Y}_u \cup \boldsymbol{Y}_v$, intersection $\boldsymbol{Y}_r = \boldsymbol{Y}_u \cap \boldsymbol{Y}_v$ of correspondent multisets or a linear combination of one of these operations.

Step 4 Reduce the number of clusters per unit: $g = n - 1$. If $g = 1$, then output the result and stop. If $g > 1$, then go to Step 5.

Step 5 Calculate the distances $d(\boldsymbol{Y}_p,\boldsymbol{Y}_r)$ between pairs of new multisets represented clusters C_p and C_r for all $1 \leq p,\ r \leq g,\ p \neq r$. Go to Step 3.

The algorithm builds a hierarchical tree or dendrogram by a successive aggregation of objects into groups. New objects/clusters C_p and C_q appear, while moving from the root of the tree by its branches, going at each step in one of the closest clusters C_r. The process of hierarchical clustering ends when all the objects are grouped into several classes or a single class. The procedure can also be interrupted at some stage in accordance with any rule, for instance, when the difference index exceeds the given threshold level [11].

The nature of cluster formation and results are largely depended on the type of used metric. The basic metric d_{11} and completely averaged metric d_{21} give almost identical results. In the process of clustering, 'small' objects (with small numbers of attributes) are merged firstly, and more 'large' objects are aggregated later. The resulted groups are comparable to the number of included objects, but very differ from each other by sets of characterizing attributes. The clustering with locally averaged metric d_{31} starts from combining similar objects of 'medium' and 'large' sizes with significant 'common' sets of attributes. Different 'small' objects are joined later. The final grouping objects obtained in the first and second cases can be strongly varied.

In the methods of non-hierarchical cluster analysis, the number of clusters is considered as fixed and specified in advance. For multi-attribute objects described by multisets, a generalized framework of nonhierarchical clustering includes the following stages.

Step 1 Select an initial partition of collection $A_1, \ldots, A_n$ of n objects in g clusters $C_1, \ldots, C_g$.

Step 2 Distribute all of the objects $A_1, \ldots, A_n$ by clusters $C_1, \ldots, C_g$ according to some rule. For example, calculate the distances $d(\boldsymbol{A}_i, \boldsymbol{Y}_t)$ between multisets $\boldsymbol{A}_i$ $(i = 1, \ldots, n)$ represented objects A_i and multisets $\boldsymbol{Y}_t$ represented clusters C_t $(t = 1, \ldots, g)$. Place the object $\boldsymbol{A}_i$ in the nearest cluster C_h with the distance $d(\boldsymbol{A}_i, \boldsymbol{Y}_h) = \min_t d(\boldsymbol{A}_i, \boldsymbol{Y}_t)$. Or determine the center A_t° of each cluster C_t, for instance, by solving the optimization task $J(\boldsymbol{A}_t^\circ, \boldsymbol{Y}_t) = \min_p \sum_i d\left(\boldsymbol{A}_i, \boldsymbol{A}_p\right)$. Place each object A_i in the cluster C_r with the nearest center A_r° given by the condition $d(\boldsymbol{A}_i, \boldsymbol{A}_r^\circ) = \min_t d(\boldsymbol{A}_i, \boldsymbol{A}_t^\circ)$. The center $\boldsymbol{A}_t^\circ$ of cluster C_t can coincide with one of the really existing objects A_i or be a so-called 'phantom' object that is absent in the original collection of objects but constructed as multiset.

Step 3 If all objects $A_1, \ldots, A_n$ do not change their cluster membership that was given by the initial partition in clusters $C_1, \ldots, C_g$, then output the result and stop. Otherwise return to Step 2.

The results of object classification can be estimated by a quality of the partition. The best partition can be found, in particular, as a solution of the following optimization problem: $\sum_t J(\boldsymbol{A}_t^\circ, \boldsymbol{Y}_t) \to$min, where the functional $J(\boldsymbol{A}_t^\circ, \boldsymbol{Y}_t)$ is defined above. In general, the solution of optimization problem is ambiguous since the functional $H(\boldsymbol{Y}_{\text{opt}})$ is a function with many local extrema. The final result depends also on the initial (near or far from optimal) allocation of objects into classes.

Often in clustering procedures, a maximization of various indicators of objects' similarity is used instead of a minimization of distance between objects that characterizes their differences. The following indexes of objects' similarity can be introduced

$$s_1(\boldsymbol{A},\boldsymbol{B}) = 1 - m(\boldsymbol{A}\Delta\boldsymbol{B})/m(\boldsymbol{Z}), s_2(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A} \cap \boldsymbol{B})/m(\boldsymbol{Z}),$$
$$s_3(\boldsymbol{A},\boldsymbol{B}) = m(\boldsymbol{A} \cap \boldsymbol{B})/m(\boldsymbol{A} \cup \boldsymbol{B}).$$

In the case of multisets, the functions s_1, s_2, s_3 generalize the known indexes of similarity such as, respectively, the simple matching coefficient, Russell-Rao measure of similarity, Jaccard coefficient or Rogers-Tanimoto measure [1]. The simple matching coefficient s_1 and Russell-Rao measure s_2 of similarity are connected with the expression $s_1(\boldsymbol{A}, \boldsymbol{B}) = s_2(\boldsymbol{A}, \boldsymbol{B}) + s_2(\boldsymbol{A}, \boldsymbol{B})$, which is one of the possible binary decompositions of maximal multiset $\boldsymbol{Z}$ on blocks of coverings and overlapping multisets [12].

5 Method of Group Sorting Multi-Attribute Objects

Consider a problem of group classification of multi-attribute objects with teachers as follows. Several experts evaluate each object from the collection A_1, …, A_n upon all criteria Q_1, …, Q_m and make a recommendation r_t for sorting the object into one of the classes C_t, $t = 1,\ldots,g$. Need to find a simple general group rule, which aggregates a large family of inconsistent individual expert-sorting rules and assigns objects to the given classes taking into account inconsistent opinions.

The method MASKA (abbreviation of the Russian words Multi-Attribute Consistent Classification of Alternatives) is used for group sorting multi-attribute objects [14–16]. An object A_i with a multiple criteria estimates $X_s = \{x_s^{e_s}\}$, $s = 1, \ldots, m$ may be represented as the following multiset of the type (1)

$$\begin{aligned}\boldsymbol{A}_i = \{&k_{\boldsymbol{A}i}(x_1^1) \circ x_1^1, \ldots, k_{\boldsymbol{A}i}(x_1^{h_1}) \circ x_1^{h_1}, \ldots,\\ &k_{\boldsymbol{A}i}(x_m^1) \circ x_m^1, \ldots, k_{\boldsymbol{A}i}(x_m^{h_m}) \circ x_m^{h_m}, k_{\boldsymbol{A}i}(r_1) \circ r_1, \ldots, k_{\boldsymbol{A}i}(r_f) \circ r_g\},\end{aligned} \tag{4}$$

which is drawn from the domain $P = X_1 \cup \ldots \cup X_m \cup R = X \cup R$. The part of sorting attributes $R = \{r_1,\ldots,r_g\}$ is the set of expert recommendations. Here $k_{\boldsymbol{A}i}(x_s^{e_s})$ and $k_{\boldsymbol{A}i}(r_t)$ are equal to numbers of experts who gives the estimate $x_s^{e_s}$ and the recommendation r_t to the object A_i. Obviously, judgments of many experts may be similar, diverse, or contradictory. These inconsistencies express subjective preferences of individual experts and cannot be considered as accidental errors.

The representation (4) of object A_i can be written as a collective sorting rule

$$\text{IF}\langle\text{conditions}\rangle,\ \text{THEN}\,\langle\text{decision}\rangle,$$

which is associated with arguments in the formula (4) as follows. The antecedent term $\langle$conditions$\rangle$ includes the various combinations of criteria estimates $x_s^{e_s}$, which describes the object features. The consequent term $\langle$decision$\rangle$ denotes that the object A_i belongs to the class C_t, if some conditions are fulfilled. The object A_i is assigned to the class C_t in accordance with the rule of voices majority that is, for instance, the relative majority if $k_{\mathbf{A}i}(r_t) > k_{\mathbf{A}i}(r_p)$ for all $p \neq t$, or the absolute majority if $k_{\mathbf{A}i}(r_t) > \sum_{p \neq t} k_{\mathbf{A}i}(r_p)$.

In order to simplify the problem, let us assume that the collection of objects $A_1, \ldots, A_n$ is to be sorted only into two classes C_a (say, more preferable) and C_b (less preferable) that is $g = 2$. This demand is not the principle restriction. Whenever objects are to be sorted into more than two classes, it is possible to divide the object collection into two classes, then into subclasses, and so on. For instance, competitive projects may be classified as projects approved and not approved, then the not approved projects may be sorted as projects rejected and considered later, and so on.

Let us correspond to each class C_a and C_b multisets $\boldsymbol{Y}_a$ and $\boldsymbol{Y}_b$, which are formed as sums of multisets represented multi-attribute objects. In this case,

$$\begin{aligned}\boldsymbol{Y}_t = \{&k_{\mathbf{Y}t}\left(x_1^1\right) \circ x_1^1, \ldots, k_{\mathbf{Y}t}\left(x_1^{h_1}\right) \circ x_1^{h_1}, \ldots, \\ &k_{\mathbf{Y}t}\left(x_m^1\right) \circ x_m^1, \ldots, k_{\mathbf{Y}t}\left(x_m^{h_m}\right) \circ x_m^{h_m}, k_{\mathbf{Y}t}(r_a) \circ r_a, k_{\mathbf{Y}t}(r_b) \circ r_b\},\end{aligned}$$

where $k_{\mathbf{Y}t}(x_s^{e_s}) = \sum_{i \in It} k_{\mathbf{A}i}(x_s^{e_s}), k_{\mathbf{Y}t}(r_t) = \sum_{i \in It} k_{\mathbf{A}i}(r_t)$, $t = a, b$, the index subsets $I_a \cup I_b = \{1, \ldots, n\}$, $I_a \cap I_b = \varnothing$. The above expression represents the collective decision rule of all experts for sorting multi-attribute objects to the class C_t.

The problem of object classification may be considered as the problem of sorting multisets in a metric space $(L(\boldsymbol{Z}), d)$. The main idea of aggregating a large family of discordant individual expert-sorting rules in a generalized group decision rule is formulated as follows. Let us introduce a set of new attributes $Y = \{y_a, y_b\}$, which elements related to the classes C_a and C_b, and construct the following new multisets

$$\begin{aligned}&R_a = \{k_{\boldsymbol{R}a}(y_a) \circ y_a, k_{\boldsymbol{R}a}(y_b) \circ y_b\}, R_b = \{k_{\boldsymbol{R}b}(y_a) \circ y_a, k_{\boldsymbol{R}b}(y_b) \circ y_b\}, \\ &Q_j = \{k_{\boldsymbol{Q}j}(y_a) \circ y_a, k_{\boldsymbol{Q}j}(y_b) \circ y_b\}\end{aligned}$$

drawn from the set Y. Here $k_{\boldsymbol{R}a}(y_t) = k_{\mathbf{Y}t}(r_a)$, $k_{\boldsymbol{R}b}(y_t) = k_{\mathbf{Y}t}(r_b)$, $k_{\boldsymbol{Q}j}(y_t) = k_{\mathbf{Y}t}(x_s^j)$, $j = 1, \ldots, h_s$. We shall call the multisets $\boldsymbol{R}_a$, $\boldsymbol{R}_b$ as 'categorical' and the multisets $\boldsymbol{Q}_j$ as 'substantial' multisets.

Note that the distance $d(\boldsymbol{R}_a, \boldsymbol{R}_b)$ between multisets $\boldsymbol{R}_a$ and $\boldsymbol{R}_b$ is the maximal distance between objects belonging to the different classes C_a and C_b. So, the categorical multisets $\boldsymbol{R}_a$ and $\boldsymbol{R}_b$ correspond to the best binary decomposition of the objects collection into the given classes C_a and C_b according to primary sorting rules of experts

$$\begin{aligned} &\text{IF}\langle(k_{\boldsymbol{A}i}(r_a) > k_{\boldsymbol{A}i}(r_b))\rangle,\ \text{THEN}\langle\text{Object}\, A_i \in C_a\rangle,\\ &\text{IF}\langle(k_{\boldsymbol{A}i}(r_a) < k_{\boldsymbol{A}i}(r_b))\rangle,\ \text{THEN}\langle\text{Object}\, A_i \in C_b\rangle. \end{aligned} \tag{5}$$

Thus, it is necessary to construct a pair of new substantial multisets $\boldsymbol{Q}_{sa}{}^*$ and $\boldsymbol{Q}_{sb}{}^*$ for every attribute group Q_s, $s = 1, \ldots, m$ such that these multisets as points of multiset metric space are to be placed at the maximal distance. The multisets $\boldsymbol{Q}_{sa}$ and $\boldsymbol{Q}_{sb}$ aggregate groups of multisets $\boldsymbol{Q}_j$ as the sums: $\boldsymbol{Q}_{sa} = \sum_{j\in Jsa}\boldsymbol{Q}_j$, $\boldsymbol{Q}_{sb} = \sum_{j\in Jsb}\boldsymbol{Q}_j$, where the index subsets $J_{sa}\cup J_{sb} = \{1, \ldots, h_s\}$, $J_{sa}\cap J_{sb} = \varnothing$. The substantial multisets $\boldsymbol{Q}_{sa}{}^*$ and $\boldsymbol{Q}_{sb}{}^*$, which correspond to the best binary decomposition of objects for the s-th attribute Q_s and are the mostly coincident with primary expert-sorting objects into the given classes C_a and C_b, are a solution of the following optimization problem:

$$d(\boldsymbol{Q}_{sa}, \boldsymbol{Q}_{sb}) \to \max d(\boldsymbol{Q}_{sa}, \boldsymbol{Q}_{sb}) = d(\boldsymbol{Q}_{sa}{}^*, \boldsymbol{Q}_{sb}{}^*).$$

The set of attributes $Q_1, \ldots, Q_m$ can be ranged by the value of distance $d(\boldsymbol{Q}_{sa}{}^*, \boldsymbol{Q}_{sb}{}^*)$ or the level of approximation rate $V_s = d(\boldsymbol{Q}_{sa}{}^*, \boldsymbol{Q}_{sb}{}^*)/d(\boldsymbol{R}_a, \boldsymbol{R}_b)$. We shall call an attribute value $x_s^j \in \boldsymbol{Q}_{st}{}^*$, $j\in J_{st}$, $t = a, b$ that characterizes the class C_t as a classifying attribute for the correspondent class. The classifying attribute that provides the acceptable level of approximation rate $V_s \geq V_0$ is to be included in the generalized decision rule for group multicriteria sorting objects. The level of approximation rate V_s shows a relative significance of the s-th property Q_s within the generalized decision rule.

Various combinations of the classifying attributes produce the generalized decision rules for group sorting objects into the classes C_a and C_b as follows

$$\begin{aligned} &\text{IF}\left\langle(x_u^j \in \boldsymbol{Q}_{ua}{}^*)\ \text{AND}\ (x_v^j \in \boldsymbol{Q}_{va}{}^*)\ \text{AND}\ldots\text{AND}\ (x_w^j \in \boldsymbol{Q}_{wa}{}^*)\right\rangle,\\ &\qquad\qquad\qquad\qquad \text{THEN}\langle\text{Object}\, A_i \in C_a\rangle, \end{aligned} \tag{6}$$

$$\begin{aligned} &\text{IF}\left\langle(x_u^j \in \boldsymbol{Q}_{ub}{}^*)\ \text{AND}\ (x_v^j \in \boldsymbol{Q}_{vb}{}^*)\ \text{AND}\ldots\text{AND}\ (x_w^j \in \boldsymbol{Q}_{wb}{}^*),\right\rangle\\ &\qquad\qquad\qquad\qquad \text{THEN}\langle\text{Object}\, A_i \in C_b\rangle. \end{aligned} \tag{7}$$

Remark, generally, that these generalized group decision rules are quite different.

Among the objects, which have been assigned to the given class C_a or C_b in accordance with the generalized decision rule (6) or (7), there are the correctly and not correctly classified objects. So, a construction of collective decision rules for sorting multi-attribute objects, which aggregate a large number of inconsistent individual expert-sorting rules, includes not only a selection of the classifying attributes $x_s^j \in Q_{sa}{}^*, x_s^j \in Q_{sb}{}^*$, but also a determination of the correctly and contradictory classified objects.

Let us find such attribute values that maximize numbers N_a and N_b of the correctly classified objects, and minimize numbers N_{ac} and N_{bc} of the not correctly classified objects. We can find, step by step, a single criterion $Q_{ua}{}^*$, then a couple of criteria $Q_{ua}{}^*$ and $Q_{va}{}^*$, three criteria $Q_{ua}{}^*, Q_{va}{}^*, Q_{wa}{}^*$, four criteria and so on, which are included in the generalized decision rules (6) or (7), and provide the minimal difference $N_a - N_{ac}$ or $N_b - N_{bc}$. Finally, we obtain the aggregated decision rules for consistent sorting the objects

$$\begin{aligned} &\text{IF}\Big\langle\Big(\sum\nolimits_{x \in Qua^*} k_{Ai}(x) > \sum\nolimits_{x \in Qub^*} k_{Ai}(x)\Big) \text{ AND} \\ &\Big(\sum\nolimits_{x \in Qva^*} k_{Ai}(x) > \sum\nolimits_{x \in Qvb^*} k_{Ai}(x)\Big) \text{ AND}\ldots \\ &\text{AND } (k_{Ai}(r_a) > k_{Ai}(r_b))\rangle,\ \text{THEN}\langle \text{Object}\, A_i \in C_a \backslash C_{ac}\rangle, \end{aligned} \tag{8}$$

$$\begin{aligned} &\text{IF}\Big\langle\Big(\sum\nolimits_{x \in Qua^*} k_{Ai}(x) < \sum\nolimits_{x \in Qub^*} k_{Ai}(x)\Big) \text{ AND} \\ &\Big(\sum\nolimits_{x \in Qva^*} k_{Ai}(x) < \sum\nolimits_{x \in Qvb^*} k_{Ai}(x)\Big) \text{ AND}\ldots \\ &\text{AND } (k_{Ai}(r_a) < k_{Ai}(r_b))\rangle,\ \text{THEN}\langle \text{Object}\, A_i \in C_b \backslash C_{bc}\rangle. \end{aligned} \tag{9}$$

These aggregated decision rules define the specified classes $C_a \backslash C_{ac}$ (say, completely preferable) and $C_b \backslash C_{bc}$ (completely not preferable) of the correctly classified objects. These aggregated rules for consistent sorting approximate the family of initial sorting rules of many individual experts.

Simultaneously the specified class $C_c = C_{ac} \cup C_{bc}$ of the contradictory classified objects is built. Such objects satisfy the aggregated decision rule for inconsistent sorting

$$\begin{aligned} &\text{IF}\Big\langle\Big[\Big(\sum\nolimits_{x \in Qua^*} k_{Ai}(x) > \sum\nolimits_{x \in Qub^*} k_{Ai}(x)\Big) \text{ AND} \\ &\Big(\sum\nolimits_{x \in Qva^*} k_{Ai}(x) > \sum\nolimits_{x \in Qvb^*} k_{Ai}(x)\Big) \text{AND}\ldots \\ &\text{AND } (k_{Ai}(r_a) < k_{Ai}(r_b))\Big] \text{ OR } \Big[\Big(\sum\nolimits_{x \in Qua^*} k_{Ai}(x) < \sum\nolimits_{x \in Qub^*} k_{Ai}(x)\Big) \text{ AND} \\ &\Big(\sum\nolimits_{x \in Qva^*} k_{Ai}(x) < \sum\nolimits_{x \in Qvb^*} k_{Ai}(x)\Big) \text{ AND}\ldots \\ &\text{AND } (k_{Ai}(r_a) > k_{Ai}(r_b))\Big],\ \text{THEN}\langle \text{Object}\, A_i \in C_c\rangle. \end{aligned} \tag{10}$$

This aggregated rule helps a DM to discover possible inconsistencies of individual expert rules and analyze additionally the contradictory classified objects.

6 Case Studies: Multiple-Criteria Expertise of R&D Projects

The developed techniques were applied to real-life expertise of R&D projects in the Russian Foundation for Basic Research (RFBR). RFBR is the Federal agency that organizes and funds basic research, and exams their practical applications. In RFBR, there is the special peer review system for a selection of the applications and assessment of the completed projects—the original multi-expert and multi-criteria expertise, similar to that found nowhere else in the world.

Several independent experts estimated each project using special questionnaires, which include specific qualitative criteria with detailed verbal rating scales. Additionally, experts give the recommendations on whether to support the application (at the competition stage) or to continue the project (at the intermediate stage). Experts estimate the scientific and practical values of the obtained results (at the final stage when the project is ended). On the basis of expert judgments, the Expert Board of RFBR decides to approve or reject the new project, to continue the project implementation, and evaluates the efficiency of the completed project. Finally, the Expert Board of RFBR determines the size of financing the supported project.

The most of methodologies, which are applied for expert estimation in different areas, uses quantitative approaches that are based on a numerical measurement of object characteristics. However, such approaches are not suitable for the expertise in RFBR, where projects are evaluated by several experts on many qualitative criteria with verbal scales.

To select the best competitive applications, the Expert Board of RFBR is need in a simple collective decision rules, which aggregate many contradictory decision rules of individual experts described with non-numerical data. These aggregated decision rules for sorting applications have been constructed by the MASKA method, and could not been found with other known MCDM techniques.

During the RFBR expertise of the goal-oriented R&D projects, several experts (usually, three) evaluate the applications upon 11 qualitative criteria presented in the expert questionnaire. These criteria are combined in two groups such as ‘Scientific characteristics of the project’ and ‘Evaluation of possibilities for the practical implementation of the project’. The first group includes 9 criteria. These are as follows: Q_1. Fundamental level of the project; Q_2. Directions of the project results; Q_3. Goals of research; Q_4. Methods of achievement of the project goals; Q_5. Character of research; Q_6. Scientific value of the project; Q_7. Novelty of the proposed solutions; Q_8. Potential of the project team; Q_9. Technical equipment for the project realization. The second group consists of 2 criteria: Q_{10}. Completion stage of basic research suggested in the project, and Q_{11}. Applicability scope of the research results.

Each criterion has nominal or ordered scale with verbal grades. For instance, the scale X_7 of the criterion Q_7. ‘Novelty of the proposed solutions’ looks as follows; x_7^1—the solutions were formulated originally and are undoubtedly superior to the other existing solutions; x_7^2—the solutions are on the same level as other existing solutions; x_7^3—the solutions are inferior to some other existing solutions.

Additionally, every expert gives a recommendation on the feasibility of the project support using the following scale: r_1—unconditional support (grade ‘5’), r_2—recommended support (grade ‘4’), r_3—possible support (grade ‘3’), r_4—should not be supported (grade ‘2’).

The proposed approach to a competitive selection of the goal-oriented R&D projects has been tested on the real database. This base included the expert evaluations of the supported and rejected applications in the following fields: ‘Physics and astronomy’ (totally 127 projects, including 39 supported and 88 rejected applications); ‘Biology and medical science’ (totally 252 projects including 68 supported and 184 rejected applications).

Expert data was processed with the MASKA method. As a result in the fields mentioned above, it was sufficient to use combinations of only several criteria, namely Q_6, Q_{10}, and Q_{11}, in order to construct the aggregated collective decision rule for the unconditional support of project. So, this decision rule had the following form:

$$\begin{gathered}\text{IF}\left\langle \text{Object}\, A_i \text{ is estimated with the criteria grades } \left(x_6^1 \text{ or } x_6^2\right),\right.\\ \left.\text{AND } \left(x_{10}^1 \text{ or } x_{10}^2\right), \text{AND } \left(x_{11}^1 \text{ or } x_{11}^2\right)\right\rangle,\\ \text{THEN}\langle \text{Object}\, A_i \in C_a \backslash C_{ac}\rangle.\end{gathered}$$

The aggregated rule for the project support can be rewritten with a natural language as follows: “The project is unconditionally supported if the project has the exceptional or very high value of scientific significance; basic research suggested in the project are completed in the form of a laboratory prototype or key elements of development; and the project has a large or interdisciplinary applicability scope of the research results”.

To evaluate efficiency of the goal-oriented R&D projects, we used the methodology of group verbal decision analysis in the reduced attribute space. At the first stage, the complex criterion of project efficiency is constructed with the original interactive procedure HISCRA (HIerarchical Structuring CRiteria and Attributes) for reducing the dimension of attribute space [18]. A construction of complex criterion scale is considered as the ordinal classification problem, where the classified alternatives are combinations of verbal grades of criteria scales. The decision classes are verbal grades of the complex criteria. At the second stage, grades of the complex criteria are composed, step by step, by using various verbal decision methods [10]. Thus, each project is assigned into some class correspondent to the grade of complex criterion, which are obtained with different methods. At the third stage, all projects are ordered by the ARAMIS method [14, 17]. The hierarchical aggregation of initial attributes allows to generate manifold collections of complex criteria, find the most preferable solution, and diminish essentially time that a DM spends for solving a problem.

During the RFBR expertise of the completed goal-oriented R&D projects, several experts (usually, two, three or four) evaluate the obtained results upon 8 qualitative criteria presented in the expert questionnaire. These criteria are as follows: Q_1. Degree of the problem solution; Q_2. Scientific level of results; Q_3. Appropriateness of patenting results; Q_4. Prospective application of results; Q_5. Result correspondence to the project goal; Q_6. Achievement of the project goal; Q_7. Difficulties of the project performance; Q_8. Interaction with potential users of results.

Each criterion has two or three-point scale of ordered verbal grades. For example, the scale X_1 of the criterion Q_1. 'Degree of the problem solution' looks as follows: x_1^1—the problem is solved completely, x_1^2—the problem is solved partially, x_1^3—the problem is not solved. The criterion Q_6. 'Achievement of the project goal' is rated as x_6^1—really, x_6^2—non-really.

The rates of project efficiency correspond to the ordered grades on a scale of the top level complex criterion D. 'Project efficiency' as d^1—superior, d^2—high, d^3—average, d^4—low, d^5—unsatisfactory. These grades, which were considered as the new attributes that characterize the projects, was formed with four different combinations of verbal decision methods.

The real database included expert assessments of results of goal-oriented R&D projects, which had been completed in the following fields: 'Mathematics, Mechanics and Computer Science' (totally 48 projects), 'Chemistry' (totally 54 projects), 'Information and telecommunication resources' (totally 21 projects). For instance, the obtained final ranking projects on Mathematics, Mechanics and Computer Science in accordance with the index $l^+(A_i)$ of relative closeness to the best object is as follows: 23 projects have the superior level of efficiency ($l^+(A_i) = 0{,}333$), 1 project has the level of efficiency between superior and high ($l^+(A_i) = 0{,}429$), 24 projects have the high level of complex efficiency ($l^+(A_i) = 0{,}500$).

7 Conclusion

In this chapter, we considered the new tools for group ordering and sorting objects described with many numerical, symbolic and/or verbal attributes, when several copies of object may exist. These techniques are based on the theory of multiset metric spaces. Underline that verbal attributes in these methods are not transformed in or replaced by any numerical ones as, for instance, in MAUT and TOPSIS methods [6], and in fuzzy set theory [21].

The multiset approach allows us to solve traditional MCDM problems in more simple and constructive manner, and discover new types of problems never being sold earlier, while taking into account inconsistencies of objects' features and preference contradictions of many actors. The ARAMIS technique is simpler and easier than the other well-known approaches to ranking multiple criteria alternatives. The MASKA technique is the unique method for group classification of multi-attribute objects and has no analogues.

References

1. Anderberg, M.R.: Cluster Analysis for Applications. Academic Press, New York (1973)
2. Blizard, W.: Multiset theory. Notre Dame J. Formal Logic **30**, 36–66 (1989)
3. Deza, M.M., Laurent, M.: Geometry of Cuts and Metrics. Springer, Berlin (1997)
4. Doumpos, M., Zopounidis, C.: Multicriteria Decision Aid Classification Methods. Kluwer Academic Publishers, Dordrecht (2002)
5. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. EJOR **138**, 247–259 (2002)
6. Hwang, C.L., Lin, M.J.: Group Decision Making under Multiple Criteria. Springer, Berlin (1987)
7. Knuth, D.E.: The Art of Computer Programming. Semi-numerical Algorithms, vol. 2. Addison-Wesley, Reading (1998)
8. Köksalan, M., Ulu, C.: An interactive approach for placing alternatives in preference classes. EJOR **144**, 429–439 (2003)
9. Larichev, O.I., Olson, D.L.: Multiple Criteria Analysis in Strategic Siting Problems. Kluwer Academic Publishers, Boston (2001)
10. Larichev, O.I.: Verbal Decision Analysis. Nauka, Moscow (2006). (in Russian)
11. Miyamoto, S.: Cluster analysis as a tool of interpretation of complex systems. Working paper WP-87-41. Laxenburg, Austria: IIASA (1987)
12. Petrovsky, A.B.: Combinatorics of Multisets. Doklady Acad. Sci. **370**(6), 750–753 (2000). (in Russian)
13. Petrovsky, A.B.: Spaces of Sets and Multisets. Editorial URSS, Moscow (2003). (in Russian)
14. Petrovsky, A.: Group verbal decision analysis. In: Adam, F., Humphreys, P. (eds.) Encyclopedia of Decision Making and Decision Support Technologies, pp. 418–425. IGI Global, Hershey (2008)
15. Petrovsky, A.B.: Method 'Maska' for group expert classification of multi-attribute objects. Doklady Math. **81**(2), 317–321 (2010)
16. Petrovsky, A.B.: Methods for the group classification of multi-attribute objects (Part 1, Part 2). Sci. Tech. Inf. Process. **37**(5), 346–368 (2010)
17. Petrovsky, A.B.: Method for group ordering multi-attribute objects. In: Engemann, K.J., Lasker, G.E. (eds.) Advances in Decision Technology and Intelligent Information Systems, vol. XI, pp. 27–31. The International Institute for Advanced Studies in Systems Research and Cybernetics, Tecumseh, Canada (2010)
18. Petrovsky, A.B., Royzenson, G.V.: Sorting multi-attribute objects with a reduction of space dimension. In: Engemann, K.J., Lasker, G.E. (eds.) Advances in Decision Technology and Intelligent Information Systems, vol. IX, pp. 46–50. The International Institute for Advanced Studies in Systems Research and Cybernetics, Tecumseh, Canada (2008)
19. Roy, B.: Multicriteria Methodology for Decision Aiding. Kluwer Academic Publishers, Dordrecht (1996)
20. Saaty, T.: Multicriteria Decision Making: The Analytic Hierarchy Process. RWS Publications, Pittsburgh (1990)
21. Zadeh, L.A.: From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions. IEEE Trans. Circuits Syst. **45**(1), 105–119 (1999)

Knowledge Fusion in Context-Aware Decision Support: Ontology-Based Modeling and Patterns

A. V. Smirnov, T. V. Levashova, N. G. Shilov and A. A. Krizhanovsky

Abstract The purpose of this chapter is twofold: (1) introducing of a semantic modeling mechanism, which is applied to achieve context-based knowledge fusion in a decision support system and (2) discovery of context-based knowledge fusion patterns. An approach to ontology-based resource modeling is proposed. The set of resources comprises sources of data/information/knowledge, problem solving resources and various actors. The knowledge fusion patterns are generalized with regard to two aspects: (1) preserving internal structures of multiple sources from which information/knowledge is fused within the ontological structure of context and preserving internal structure of the context itself, and (2) preserving autonomies of the multiple sources and the context. Six knowledge fusion patterns have been discovered. They are simple fusion, inferred fusion, instantiated fusion, adapted fusion, flat fusion, and historical fusion.

A. V. Smirnov (✉) · T. V. Levashova · N. G. Shilov · A. A. Krizhanovsky
Laboratory of Computer Aided Integrated Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia
e-mail: smir@iias.spb.su

T. V. Levashova
e-mail: tatiana.levashova@iias.spb.su

N. G. Shilov
e-mail: nick@iias.spb.su

A. A. Krizhanovsky
e-mail: aka@iias.spb.su

A. V. Smirnov
University ITMO, St. Petersburg, Russia

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_3,

1 Introduction

Decision support systems (DSSs) heavily rely upon large volumes of data, information, and knowledge available in different sources. Whereas several years ago the main technology used to integrate data and information from multiple sources within DSS was data fusion, today the focus of data fusion has naturally changed to knowledge fusion. The goal of knowledge fusion (KF) is to integrate information and knowledge from multiple sources into a common piece of knowledge that may be used for decision making and problem solving or may provide a better insight and understanding of the situation under consideration [1–4].

The present research is a continuation of the research on knowledge logistics. The KF technology was an important constituent of the knowledge logistics approach. As main results of that approach, a conceptual framework for context-aware operational decision support [5] was developed and *generic* KF patterns were discovered [6]. A context-aware DSS, designed according to the framework, was intended to function in heterogeneous environments like pervasive or ubiquitous ones. Semantic modeling of the environmental resources supported KF processes.

This chapter has for two objectives. The first one is a description of the semantic modeling mechanism applied to achieve context-based KF in DSS. The second objective is a discovery of *context-based* knowledge fusion patterns. While the *generic* patterns generalized KF processes taking place at different stages of building and exploitation of DSS, the *context-based* patterns are intended to generalize processes in the operational stage of the DSS functioning, i.e. the stage where the context aware functions of the system come into operation.

DSS described in this chapter is intended to support decisions on involvement of autonomous resources of a DSS environment in common activities and on planning these activities. Sources of data/information/knowledge and various actors existing in such an environment are considered as resources. Semantics is the key to ensure that several information and knowledge sources come to the same meaning of the situation and information/knowledge being communicated. This explains the fact that ontologies support most efforts on KF (e.g., [7–11]). In this chapter a mechanism to ontology-based resource modeling is proposed to overcome resource heterogeneity. This mechanism enables KF processes and allows the resources to achieve interoperability.

In the DSS, the KF technology is resorted to the following objectives: (a) integration of knowledge to create a new type of knowledge that is an abstract non-instantiated model of the current situation, (b) fusion of data/information/knowledge to produce an actual picture of the current situation that serves as the basis for problem solving and decision making, (c) problem solving, (d) inference of new relationships between knowledge objects, (e) discovery of alternative problem solving methods, (f) extension of capabilities/competencies of a knowledge object with new ones. The present research proposes KF patterns discovered at the stages of DSS functioning where the listed objectives are achieved.

The rest of the chapter is organized as follows. Section 2 introduces the conceptual framework of the developed DSS and the mechanism to semantic resource modeling. Section 3 gives an insight into knowledge fusion and summarizes the effects of it. Context-based knowledge fusion patterns are discovered and systematized in Sect. 4. Main research findings and a brief discussion conclude the chapter. Throughout the chapter, the task of planning fire response actions is considered to illustrate main ideas behind the research.

2 Context Aware Decision Support System

The objectives, this chapter aims at, are treated within a context-aware DSS intended to support decisions on involvement of autonomous resources of the DSS environment in common activities and on scheduling these activities. Sources of data/information/knowledge, problem solving resources and acting resources are distinguished in the environment. All they are considered as resources. The DSS was built following the conceptual framework presented below.

2.1 Conceptual Framework

The conceptual framework adopts the idea of ontology-based context representation. The central constituent of the framework is the application ontology (AO). This ontology represents non-instantiated knowledge of the application domain. Domain and problem solving knowledge fused from different knowledge sources make up AO. In this regard, AO can be considered as a knowledge source representing two different types of knowledge fused. AO supports an object-oriented ontology representation—it is specified by sets of classes, class attributes, attribute domains, and relationships [12].

Ontological knowledge is instantiated in the context by resources. Context represents a decision situation (the setting in which the decision occurs). A situation is represented at two levels. At the *first level* it is represented by *abstract context* that specifies non-instantiated ontology knowledge relevant to the current situation. Such knowledge is extracted from the application ontology. As two components make up the application ontology, the abstract context specifies domain knowledge describing the decision situation and problems to be solved in this situation. At the *second level* the situation is represented by *operational context* that is an instantiation of the abstract context with the actual information.

A set of contextual resources is organized to instantiate the abstract context. This set is a subset of all the environmental resources. The set of contextual resources comprises data/information/knowledge sources that can provide data values to create instances of the classes represented in the abstract context or solve problems specified in it. Then, an execution sequence of the contextual resources is determined. In this way a resource network is organized. Nodes of this network are resources providing data values and/or solving problems; network arcs signify an ordering on the resource execution.

The result of problem solving is a set of alternative decisions that can be made in the current situation. These alternatives are plans for the common activities of available acting resources. The decision maker chooses one plan from the set of alternative ones and delivers it to acting resources that are in this plan. The system facilitates the process of decision making by providing efficiency criteria that the decision maker can apply to the proposed set of alternatives.

The made decision (plan), the abstract context, and the operational context along with the resource network are saved in a context archive. The operational context and the resource network are saved in their states at the instant of alternatives generation.

In order to achieve context-based knowledge instantiation and enable KF processes the mechanism of ontology-based resource modeling is used.

2.2 Ontology-Based Resource Modeling

In the mechanism described in this subsection, the resources are represented by their ontologies. The representations of the resource ontologies are compatible with the representation of AO. The mechanism purposes the aim of an alignment of the resource ontologies and AO.

The alignment is based on discovery of semantically similar names for classes and attributes from AO and the resource ontologies. Semantic distance serves as a measure of semantic similarity. The semantic distances are calculated based on a machine-readable dictionary. This dictionary was automatically extracted from Wiktionary [13].

First, names of classes and attributes occurred in AO are searched for among concepts defined in the machine-readable dictionary. For the found concepts their synonyms and associated concepts are sought. Associated concepts are words hyperlinked in the concept definitions (Table 1).

Then, an initial semantic network is built. The found concepts (i.e., names of classes and attributes from AO), their synonyms and associated concepts are nodes of this network. The nodes corresponding to the found concepts and the nodes representing their synonyms and associated concepts are linked. The links are labeled by weights. It is assumed that weight w of a link specified between two concepts t_i and t_j is assigned as:

$$w = \begin{cases} 0,5 & -t_i,\ t_j \text{ are synonyms} \\ 0,3 & -t_i,\ t_j \text{ are associated concepts}\,. \\ \infty & -t_i,\ t_j \text{ are the same concept} \end{cases} \qquad (1)$$

Table 1 Extraction from machine-readable dictionary:example for class "Fire"

Concept	Wiktionary definition	Synonyms	Associated concepts
Fire	The often *accidental* occurrence of fire in a certain place leading to its full or partial *destruction*	–	Accidental, destruction

Names of classes and attributes from AO, for which there are no corresponding concepts in the machine-readable dictionary, are included in the network in the form as they are introduced in AO.

At the next step the semantic network is augmented with nodes representing the names of classes and attributes from the resource ontologies. This is done as described in the case of AO but with a slight difference: synonyms and associative concepts are included in the network only for the names from the resource ontologies that are different from the names the network already represents.

The links between nodes representing the same names from AO and from the resource ontologies are weighted as ∞.

Next, semantic distance *Dist* (2) is calculated between nodes representing concepts from AO and nodes representing concepts from the resource ontologies:

$$Dist(t_i, t_j) = \frac{1}{\sum_S \prod_{k=s_i}^{s_j} w_k}, \tag{2}$$

where t_i—concept from a resource ontology, t_j—AO-concept; w—weight of the link between t_i and t_j; S—a set of paths from t_i to t_j, where a path $s \in S$ is formed by any number of links between t_i and t_j passing through any number of nodes.

After the semantic distances have been calculated, experts are provided with a ranked list of semantically similar concepts. This list represents concepts from the resource ontologies and semantically similar to them concepts from AO with corresponding semantic distance values. Based on the list the experts map resource ontologies and AO.

The described approach is illustrated by an example of matching an ontology-based representation of the sensor that registers the location of an emergency event against AO for the emergency management domain. Figure 1 represents the sensor ontology and a piece of AO that is relevant to the illustrative example. In this example, matches for the attribute `'Point'` from the sensor ontology are searched for in AO. In the sensor ontology the attribute `'Point'` represents a point on the map, where the emergency is taking place. The purpose is to find in AO names of classes or attributes which could be considered as similar to the attribute `'Point'`.

Part of the machine-readable dictionary relevant for the example in question is presented in Table 2. The given part is limited to the attributes `'Location'` and `'Type'` of the AO-class *Fire* to simplify the discussion. A piece of the semantic network built based on Table 2 and formula (1) is presented in Fig. 2.

Fig. 1 Specifications of emergency in sensor ontology and fire in application ontology

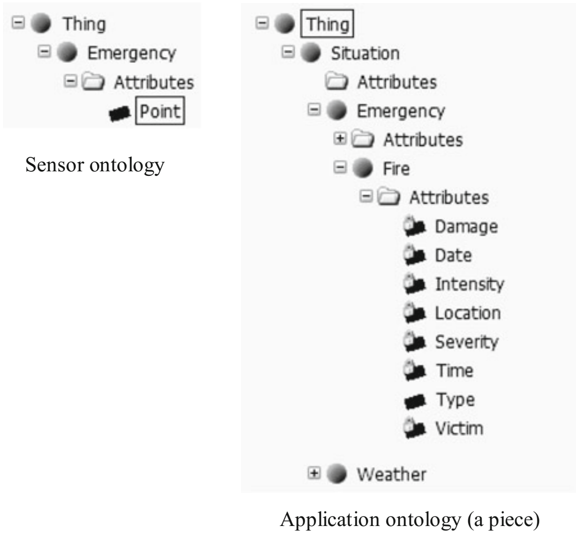
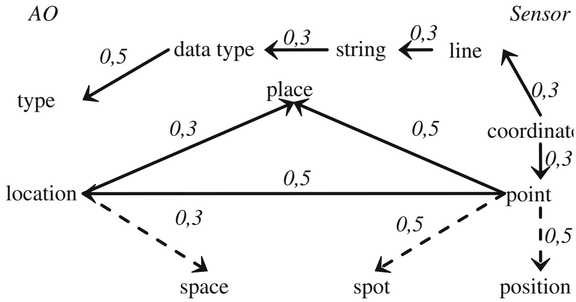


Table 2 Extraction from machine-readable dictionary: example for attribute "point"

Point	A *location* or place. A point is defined by its *coordinates*	Location, place, position, spot	Location, coordinates
Location	A particular *place* in physical *space*	–	Place, space
Type	A *tag* attached to variables and values used in determining what values may be assigned to what variables	Data type	Tag

Fig. 2 A piece of semantic network relevant to attribute "Point"



The set of paths from the concept `'Point'` to the AO-concept `'Location'` comprises two paths: (1) *location* → *place* (weight 0.3), *place* → *point* (weight 0.5); and (2) *point* → *location* (weight 0.5). Thus, semantic distance between the two concepts is calculated as:

$$Dist(point, location) = \frac{1}{0.3 \cdot 0.5 + 0.5} = 1.54.$$

The set of paths from the concept `'Point'` to the AO-concept `'Type'` comprises one path: *point* ← *coordinate* (weight 0.3), *coordinate* → *line* (weight 0.3), *line* → *string* (weight 0.3), *string* → *data type* (weight 0.3), *data type* → *type* (weight 0.5). Semantic distance between `'Point'` and `'Type'` is

$$Dist(point, type) = \frac{1}{0.3^4 \cdot 0.5} = 246.9.$$

It can be seen that the semantic distance between the concepts `'Point'` and `'Location'` is much shorter than between the concepts `'Point'` and `'Type'`. So, the experts can align the attribute `'Location'` of the class *Fire* specified in AO and the attribute `'Point'` specified in the sensor ontology. This alignment means that the sensor responsible for registration of emergency locations instantiates the attribute `'Location'` of the AO-class *Fire*.

3 Knowledge Fusion

The main result of KF is synergetic effect from integration of a wide variety of information and knowledge sources. Based on an analysis of publications on KF, several types of KF can be distinguished:

- Intelligent fusion of massive amounts of heterogeneous data/information from a wide range of distributed sources into a form which may be used by systems and humans as the foundation for problem solving and decision making [1, 2]. Intelligent fusion assumes taking into account the semantic content of the sources being fused.
- Integration of knowledge from various knowledge sources resulting in a completely different type of knowledge or new idea how to solve the problem [14, 15]. Integration of different types of knowledge (domain, procedural, derived, presentation, etc.) resulting in a new knowledge type [3] and integration of multiple knowledge sources into a new knowledge object [16, 17] belong to this type of KF.
- Combining knowledge from different autonomous knowledge sources in different ways in different scenarios, which results in discovery of new relations between the knowledge from different sources or/and between the entities this knowledge represents [18, 19].
- Re-configuration of knowledge sources to achieve a new configuration with new capabilities or competencies [20].
- Knowledge exchange to improve capabilities or competencies through learning, interactions, discussions, and practices [20]. This type of KF is supported in multi-agent systems, which are composed of multiple intelligent software entities.

- Involving knowledge from various sources in problem solving, which results in a new knowledge product [6].

The analysis above enabled to reveal the possible effects of KF:

- new knowledge object created from data/information;
- new knowledge type or knowledge product (service, process, technology, etc.);
- new relations between knowledge objects;
- new capabilities/competencies of a knowledge object;
- new problem solving method;
- solution for a problem.

In the next section the context-aware DSS is investigated for the effects above and KF patterns behind the found effects are discovered.

4 Context-Based Knowledge Fusion in DSS

Processes of KF are considered with references to abstract and operational contexts. At first, processes in DSS resulting in effects described in the precedent section are described. Demonstrations with examples from the fire response scenario accompany these descriptions. At the end of each description, a statement is formulated. Such a statement generalizes the KF result in terms of maintenance of the internal structure of the context and resource ontologies, and preservation of their autonomies.

4.1 Knowledge Fusion: Abstract Context

Abstract context is created from the single knowledge resource—AO. The procedure of the abstract context creation consists in selection in AO knowledge relevant to the decision situation, its extraction, and integration into a new knowledge object. This object corresponds to the abstract context, which can be considered as a new knowledge product fusing two types of knowledge.

Referring to the illustrative scenario, the abstract context is created for a fire situation. Figure 3 shows pieces of AO knowledge and abstract context knowledge. The abstract context significantly reduces the amount of knowledge represented in AO. The created context, among other things, specifies that in a fire situation the services provided by emergency teams and fire brigades are required. These teams and brigades can use ambulances, fire engines and special-purpose helicopters for transportation. In the figure, the problem-solving knowledge specified in the abstract context is collapsed in the class “Emergency response”. Partly, this class is shown expanded in Fig. 3 on the right.

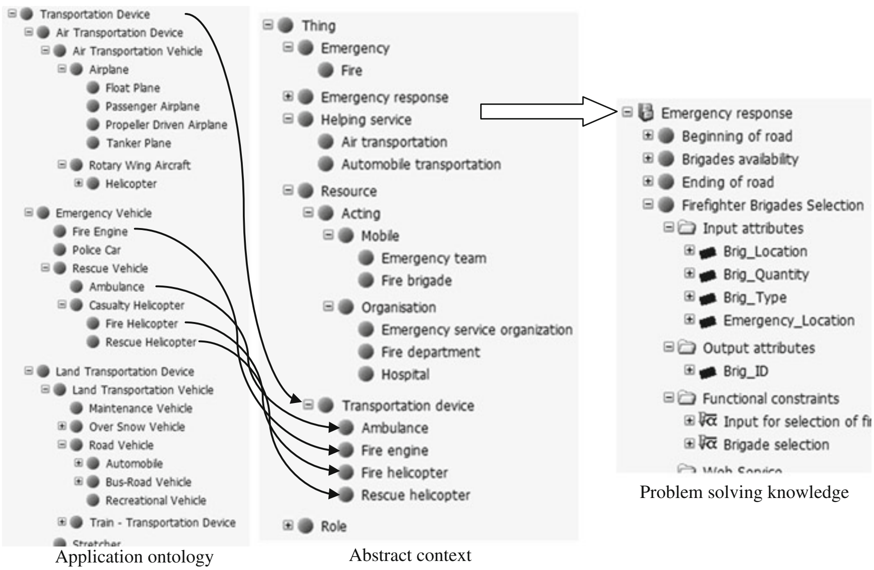


Fig. 3 Fire situation: abstract context (a fragment)

Statement 1. The procedure of the abstract context creation neither affects the internal structure of AO nor its autonomy. The abstract context becomes an autonomous object with a proper structure.

The knowledge integration may result in discovery of new relationships between the knowledge unrelated in AO. Figure 4 illustrates the case when AO specifies that a value for an input parameter of the routing method serves as a value for the attribute representing the current location of a transportation device (1). In this ontology the class "mobile" representing a mobile acting resource and the class "transportation device" are related by a functional relationship (2) assigning that the location of a mobile acting resource is the same as the location of the transportation device this resource goes by. In the abstract context a new functional relationship (3) has been inferred. This relationship means that a value for the attribute representing the current location of a mobile acting resource serves as an input parameter of the routing method. In other words, values for the both attributes representing the current location of a transportation device or the current location of a mobile resource can be used by the routing method as one of its input parameter.

Statement 2. The inferred relationships preserve the abstract context autonomy, but they change the context structure.

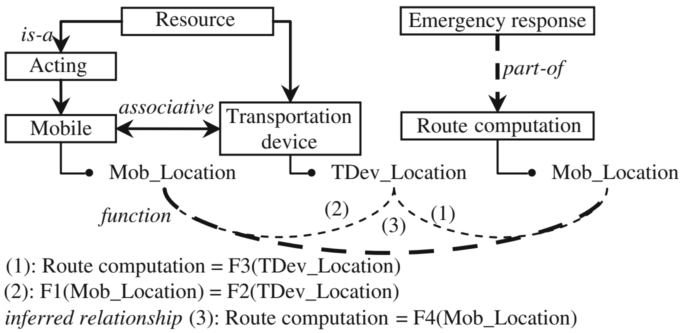


Fig. 4 Inferred relationship

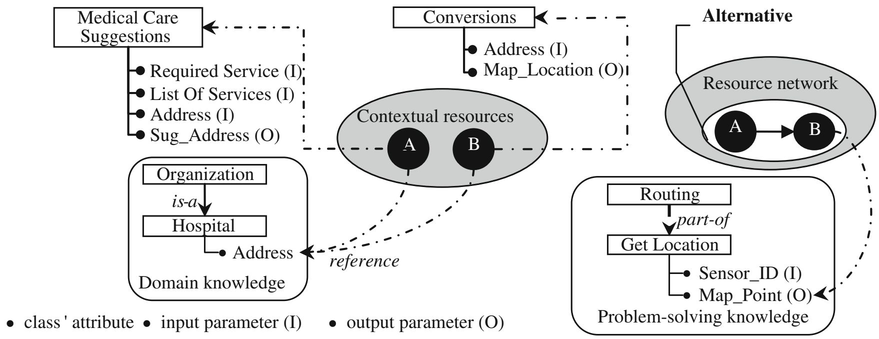


Fig. 5 Discovery of alternative problem-solving methods

The abstract contexts are reusable components of DSS. Reuse of an abstract context in settings when available resources are not intended to solve problems that are specified in this context enables to find an alternative problem-solving method. A basic condition for finding alternatives is availability of resources that provide methods that can be reused to solve the specified problems.

Figure 5 illustrates the case when the abstract context specifies routing problem as a hierarchy of methods one of which (*GetLocation*) returns the current locations of objects in the format of coordinates of a point on the map. In the example under consideration it is required to determine the locations of hospitals. The method *GetLocation* uses data from sensors. The set of contextual resources comprises no sensors dealing with static objects like hospitals. But this set comprises some other resources. One of them (*A*) implements a method (*MedicalCareSuggestions*) intended to make recommendations on which medical care organizations can be used to access a specific medical service. This resource contains a private database with information about hospitals. The other resource (*B*) implements a method (*Conversions*) that converts the address format into the format of coordinates. The execution of the methods *MedicalCareSuggestions* and *Conversions* one after another is an alternative way to calculate the hospital locations in the format of coordinates. Optionally, the methods *MedicalCareSuggestions* and *Conversions* can be included in the abstract context.

Statement 3. If alternative methods are included in the abstract context then the structure of this context is changed. Otherwise, its structure is preserved. In any case, the autonomies of the abstract context, and the sources implementing the alternative methods are preserved.

Summing up the statements for the abstract context, KF can produce a new knowledge product, identify new relationships between knowledge objects, and discover alternative problem solving methods.

4.2 Knowledge Fusion: Operational Context

An operational context is produced through the semantic fusion of data/information from multiple data/information resources within the ontology structure of the abstract context. The result of this kind of fusion is a new knowledge object (operational context) created from data/information.

In case of fire, the operational context is a GIS-based representation of the fire situation. This representation is supplemented with characteristics of the fire situation specified in the abstract context. Examples of such characteristics can be seen in Fig. 1—the attributes of class "Fire" specify them. As well, the operational context represents information about traffic situation, available routes, weather conditions, and found acting resources (their locations, availabilities, capacities, transportation devices being used, etc.). According to the abstract context (Fig. 4) the main acting resources are emergency teams, fire brigades, and hospitals.

Statement 4. As soon as the resources start instantiating the abstract context, they lose their autonomies. Sources of information and data preserve their structures; problem-solving resources dissolve within the operational context. The operational context is a new knowledge object without autonomy.

Referring to the main purpose of decision support, which is solving the problems of organization of an association or community of autonomous entities joined for a common purpose and planning their joint activities, the problem of planning fire response actions is solved for the fire situation. For the emergency teams, fire brigades, and hospitals a plan for their joint actions is produced. An example of such a plan is shown in Fig. 6. The dotted lines indicate the routes to be used for the transportations. This result relates earlier independent objects.

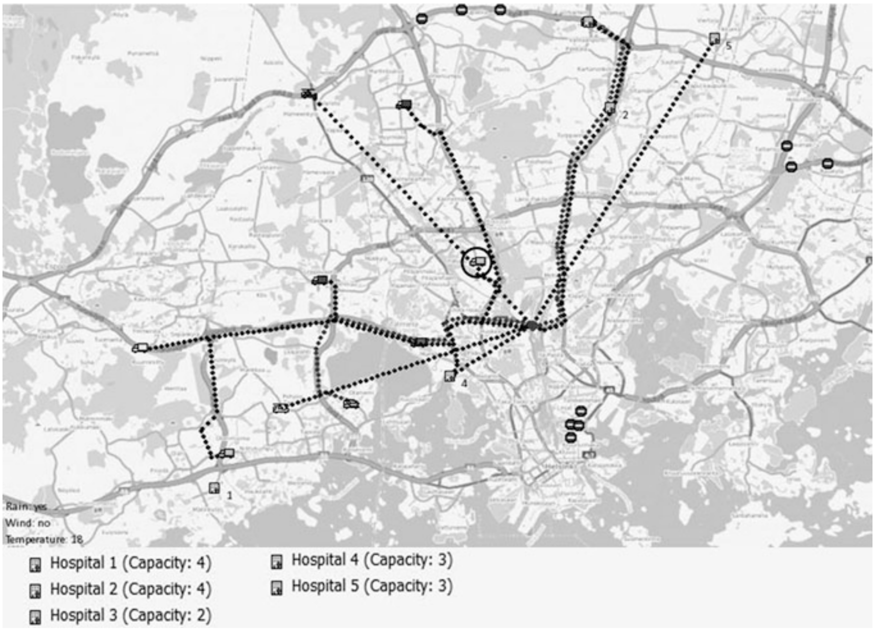


Fig. 6 Plan for actions: new relations and new knowledge product

To proceed from the above, problem solving results in new relations between resource entities represented in the operational context. Moreover, the result of problem solving is a new knowledge product (plan for actions) of a new type.

Statement 5. Resources involved in problem solving lose their autonomies and dissolve within the operational context. New relations between the resource entities affect the internal structure of the operational context.

New relations between the knowledge represented in the operational contexts can be discovered based on a comparative analysis of the operational contexts accumulated in the context archive. Finding the same entity in different operational contexts may lead to revealing new relations for this entity.

For example, the emergency team encircled in Fig. 6 participated in different emergency response actions. Some operational contexts in which this team appeared and then participated in corresponding actions do not represent any instances of the class *Emergency response organization* specified in the abstract context (Fig. 7). This suggests that the emergency team in question is a part of one of the hospitals represented in the operational contexts together with this team. Based on the operational context (Fig. 7) it can be judged that most probably the team is a part of hospital 5 represented in this context since the context does not

Fig. 7 History for an emergency team

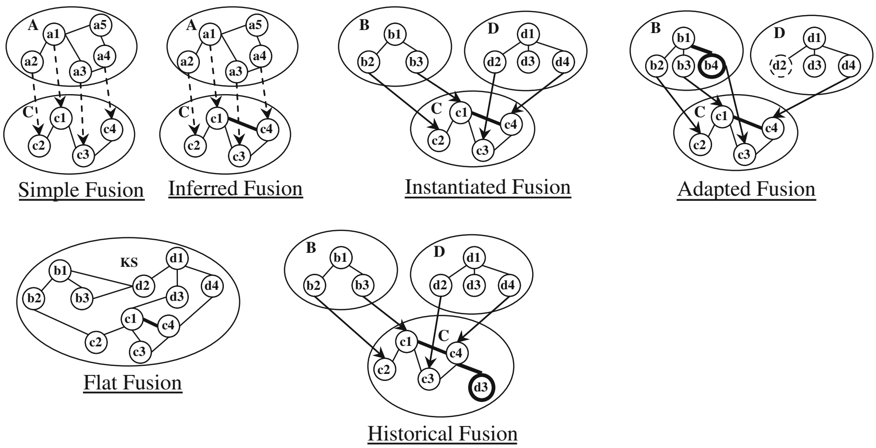
represent any other hospitals from Fig. 6 except this one. *Part-of* relation between the hospital 5 and the encircled emergency team is the new revealed relation.

Statement 6. The operational context and resources are not autonomous in the context archive. The new relation changes the structure of the operational context. The resources preserve their structures.

The resource network responsible for the instantiation has to adapt to the changing situation. In some cases, the network can be adjusted to the changed situation. An example of such a case is when a knowledge source failed, but its actions can be delegated to another resource or redistributed between other resources. This can lead to an extension of the capabilities/competencies of these resources with new ones.

For instance, an emergency team has failed in the course of actions because of road destruction, ambulance blockage, etc. This team was trained to rescue operations. If such operations are required in the current situation, then in some cases these operations can be delegated to available emergency teams. In such cases the profiles of teams agreed to take part in the rescue operations are extended with this new capability.

Statement 7. The operational context and the resources are not autonomous in the course of actions. Introducing a new capability to knowledge objects changes their internal structures.



A, B, D – nodes of the knowledge source network;
C – context;
KS –new knowledge source of new type

Fig. 8 Context-based knowledge fusion patterns in DSS

Summarizing the results of KF for the operational context, KF can produce a new knowledge object, new relations between knowledge objects, a new knowledge product of a new type, and extend capabilities/competencies of a knowledge object with new ones.

4.3 Context-Based Knowledge Fusion Patterns

The discussion above enables to distinguish patterns of context-based KF (Fig. 8). These patterns use maintenance of internal structures of contexts and knowledge sources and preserving their autonomies as the criteria. It is suggested, that essentially, all kinds of resources considered above can play a role of knowledge sources.

- *Simple fusion*: selection of pieces of knowledge from one or more knowledge sources and their integration into a new knowledge object. Initial knowledge sources preserve their internal structure and autonomies; the new knowledge object (context) becomes autonomous object with a proper structure.
- *Inferred fusion*: inference of new relations between the unrelated pieces of knowledge extracted from the same knowledge source and fused into existing context. The context preserves its autonomy, but its structure is changed.
- *Instantiated fusion*: semantic fusion of data/information from multiple sources within the ontology structure of context resulting in a new knowledge object created from data/information. The context and the sources of data & information preserve their internal structures but lose their autonomies.

Table 3 Correspondences between Statements and KF patterns

Statement	Phase of DSS functioning	Meaning	Knowledge fusion pattern
1.	Abstract context creation	Integration of multiple knowledge pieces from the same knowledge source into a new piece of knowledge	Simple fusion
2.	Abstract context creation	Inference of new relationships between originally unrelated concepts	Inferred fusion
3.	Abstract context reuse	Discovery of alternative problem-solving methods	*Not defined*
4.	Operational context producing	Instantiation of the abstract context with values from multiple data/information/knowledge sources	Instantiated fusion
5.	Generation of a set of alternatives	Problem solving	Flat fusion
6.	Contexts archiving	Discovery of new relationships between entities	Historical fusion
7.	Decision implementation	Gaining new capabilities/competencies by elements of knowledge source network	Adapted fusion

- *Adapted fusion*: adaptation of the existing knowledge source network to the context that results in gaining new capabilities/competencies by elements of the knowledge source network. The context and the initial knowledge sources do not preserve their autonomies; the structures of the knowledge sources may be changed.
- *Flat fusion*: fusion of knowledge from multiple knowledge sources to solve a problem specified in context. The solution represents a new knowledge product of a new type. The initial knowledge sources dissolve within the new knowledge object. The context and the knowledge sources do not preserve their internal structure and autonomies.
- *Historical fusion*: revealing new relations for the same knowledge object based on a comparative analysis of different contexts this object has been involved in. The context loses its autonomy, its structure is changed. Knowledge sources representing the object in question preserve their internal structures, but lose their autonomies.

The correspondences between the formulated statements and the patterns are given in Table 3. It must be pointed out, that any patterns for the Statement 3 have not been discovered. This can be explained by optional change of the abstract context structure when a new alternative problem-solving method is found. An assumption here is that the considered aspects of generalization are not sufficient to discover such a pattern.

5 Conclusion

In this chapter a conceptual framework for a context-aware DSS was introduced. An approach to ontology-based resource modeling was proposed within this framework. Usage of this approach within the framework allows the autonomous resources to make for achievement of a common purpose and to participate in joint actions. Besides this, the semantic resource models enable context-based KF processes taking place in DSS. Such processes were investigated and context-based KF patterns were discovered.

Comparatively to the existing approaches dealing with ontology based modeling of heterogeneous sources [21–23], the proposed approach introduces a measure for semantic similarity that takes into account, besides typical lexical relations, associations existing between the compared concepts. This is believed to give more accurate results, since the concepts are compared in their broad semantic environment.

While semantic modeling is a subject of many research studies, discovery of KF patterns is not a hot research topic. Up to now, some general patterns like unstructured fusion [24], convergence [14], fractal fusion [14], knowledge recombination (includes two patterns: KF and knowledge reconfiguration) [20] have been mentioned in a few research works. These patterns were discovered as a generalization of processes of knowledge interchange and combination (integration) in different distributed organizations and as a specialization of technology fusion patterns.

The present research was limited with two views on context-based KF. It was one of the reasons for a pattern that generalizes discovery of alternative problem-solving methods was not defined. Some future research is required to consider more aspects of interactions between the knowledge being fused and the contexts. Probably, the future research will enable to build formal models for KF patterns and provide some ideas which pattern is appropriate to use in the current state of the situation.

Acknowledgment The present research was supported partly by projects funded by grants 12-07-00298, 12-01-00481, 13-07-12095, 14-07-00345, 14-07-00427 of the Russian Foundation for Basic Research, the project 213 of the research program "Information, control, and intelligent technologies & systems" of the Russian Academy of Sciences (RAS), the project 2.2 of the Nano- and Information Technologies Branch of RAS, and grant 074-U01 of the Government of the Russian Federation.

References

1. Scherl, R., Ulery, D.L.: Technologies for army KF. Final report, Monmouth: Monmouth University, Computer Science Department, West Long Branch, Report no. ARL-TR-3279 (2004)
2. Alun, P., Hui, K., Gray, A., Marti, P., Bench-Capon, T., et al.: Kraft: an agent architecture for KF. Int. J. Coop. Inf. Syst. **10**(1–2), 171–195 (2001)
3. Holsapple, C.W., Whinston, A.B.: Building blocks for decision support systems. In: Ariav, G., Clifford, J. (eds.) New Directions for Database Systems, pp. 66–86. Ablex Publishing Corp, Norwood (1986)

4. Phan-Luong, V.: A framework for integrating information sources under lattice structure. Inform. Fusion **9**(2), 278–292 (2008)
5. Smirnov, A., Pashkin, M., Chilov, N., Levashova, T.: Constraint-driven methodology for context-based decision support. J. Decis. Syst. **14**(3), 279–301 (2005)
6. Smirnov, A., Pashkin, M., Chilov, N., Levashova, T., Haritatos, F.: Knowledge source network configuration approach to knowledge logistics. Int. J. Gen. Syst. **32**(3), 251–269 (2003)
7. Bossé, É., Valin, P., Boury-Brisset, A.-C., Grenier, D.: Exploitation of a priori knowledge for information fusion. Inform. Fusion **7**(2), 161–175 (2006)
8. Gu, J., Xu, B., Chen, X.: An XML query rewriting mechanism with multiple ontologies integration based on complex semantic mapping. Inform. Fusion **9**(4), 512–522 (2008)
9. Yao, J., Raghavan, V.V., Wu, Z.: Web information fusion: a review of the state of the art. Inform. Fusion **9**(4), 446–449 (2008)
10. Little, E.G., Rogova, G.L.: Designing ontologies for higher level fusion. Inform. Fusion **10**(1), 70–82 (2009)
11. Dapoigny, R., Barlatier, P.: Formal foundations for situation awareness based on dependent type theory. Inform. Fusion **14**(1), 87–107 (2013)
12. Smirnov, A., Pashkin, M., Chilov, N., Levashova, T.: Knowledge logistics in information grid environment. Future Gener. Comp. Syst. **20**(1), 61–79 (2004)
13. Wiktionary, the free dictionary. Internet: http://en.wiktionary.org/. Accessed 10 Oct 2012
14. Lee, K.-R.: Patterns and processes of contemporary technology fusion: the case of intelligent robots. Asian J. Technol. Innov. **15**(2), 45–65 (2007)
15. Grebla, H.A., Cenan, C.O., Stanca, L.: Knowledge fusion in academic networks, BRAIN: Broad Res. Artif. Intell. Neurosci. **1**, 2 (2010). http://www.edusoft.ro/brain/-index.php/brain/article/download/60/145. Accessed 10 Oct 2012
16. Kuo, T.-T., Tseng, S.-S., Lin, Y.-T.: Ontology-based KF framework using graph partitioning. In: Chung, P.W.H., Hinde, C.J., Ali, M. (eds.) Proceedings of IEA/AIE 2003. LNCS, vol. 2718, pp. 11–20 (2003)
17. Gou, J., Yang, J., Chen, Q.: Evolution and evaluation in KF system. In: Mira, J. Alvarez, J.R. (eds.) Proceedings of IWINAC 2005. LNCS, vols. 2718, 3562, pp. 192–201 (2005)
18. Laskey, K.B., Costa, P., Janssen, T.: Probabilistic ontologies for KF. In: Proceedings of the 11th International Conference Information Fusion, IEEE (2008). http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=-4632375. Accessed 10 Oct 2012
19. Jonquet, C., LePendu, P., Falconer, S., Coulet, A., Noy, N.F., et al.: NCBO Resource Index: ontology-based search and mining of biomedical resources. J. Web Semant. **9**(3), 316–324 (2011)
20. Lin, L.Y., Lo, Y.J.: Knowledge creation and cooperation between cross-nation R&D institutes. Int. J. Electron. Bus. Manag. **8**(1), 9–19 (2010)
21. Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Huebner, S.: Ontology-based integration of information—a survey of existing approaches. In: Proceedings of the Workshop on Ontologies and Information Sharing at the International Joint Conference Artificial Intelligence (IJCAI), pp. 108–117 (2001)
22. Kalfoglou, Y., Schorlemmer, M.: Ontology mapping: the state of the art. Knowl. Eng. Rev. **18**(1), 1–31 (2003)
23. Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. J. Data Semant. **4**, 146–171 (2005)
24. Chen, J., McQueen, R.J.: Knowledge transfer processes for different experience levels of knowledge recipients at an offshore technical support center. Inf. Technol. People **23**(1), 54–79 (2010)

Preference Function Reconstruction for Multiple Criteria Decision Making Based on Machine Learning Approach

L. M. Lyubchyk and G. L. Grinberg

Abstract The problem of expert preference function reconstruction in decision making process of multicriterion comparative assessment of set of object is considered. The problem is reduced to integral indicator identification using available data of object's performance indexes measurements as well as expert estimation of integral indicators values for each object and feature weights. Based on machine learning approach and expert estimations concordance technique, the solution of preference function recovering problem is obtained in the form of optimal nonlinear object feature convolution.

1 Introduction

One of the most important problems of decision making theory is multiple criteria comparative assessments and ordering of objects or alternative decisions based on expert judgments [1–3]. The well-known and widely practiced approach to such a problem is the reduction of a set of partial performance criteria (indexes), which characterized some object's features, to the generalized one, known as an integral indicator [4], which, in fact, is an expert measure of object's performance. Integral indicator as an aggregate performance index of objects (alternatives) over for all criteria should be constructed on the basis of expert preferences and reflects the

L. M. Lyubchyk (✉)
Department of Computer Mathematics and Mathematical Modeling, National Technical University "Kharkiv Polytechnic Institute", Kharkiv 61002, Ukraine
e-mail: lyubchyk@kpi.kharkiv.edu

G. L. Grinberg
Department of Economical Cybernetics and Marketing Management, National Technical University "Kharkiv Polytechnic Institute", Kharkiv 61002, Ukraine
e-mail: glngrinberg@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_4,

priorities of the decision maker. The choice of integral indicator model is an intricate problem, because expert preference function structure is usually unknown.

Different approaches to expert preference function construction are considered in [5–7]. In practice, the commonly used preference function model is a linear convolution of partial performance indexes defines the object's features, where feature weights are given by experts [2, 8]. An alternative is direct expert estimation of integral indicator using observed (measured) values of partial performance indexes for each object to be compared with subsequent approximation of obtained experimental dependence.

A simple linear feature convolution is not always adequately representing the actual expert preferences, which may be very complicated and appears in a complex nonlinear dependence of integral indicator from partial performance indexes. For such a case, for instance, an appropriate approximation of preference function may be used [5, 6], but it leaves open the question about the choice of approximation model, which provides a sufficiently high accuracy of approximation under small amount of data.

The possible contradiction between expert estimations of object's integral indicators and feature weights may be overcome by expert estimation concordance approach, proposed by Shakin and Strijov [9, 10]. An appropriate concordance procedure performs the simultaneous correction of object's integral indicators and feature weights expert estimations in order to obtain the concordant decision.

In this chapter the problem of preference function reconstruction is considered with the purpose of expert preferences eliciting using measured performance indexes ("object–feature" data) as well as expert estimations of features weights and object's integral indicators. Therefore the preference function reconstruction may be considered as specific identification problem [11], which may be solved by means of machine learning approach [12].

The use of support vector machine (SVM) approach [13] combined with kernel-based methods [14] provides a significant reduction in the number of estimated parameters of integral indicator model and allow to reconstruct the complex nonlinear structure of expert preference function.

Herewith expert estimations of feature weights in linear feature convolution, which may be considered as a first approximation for nonlinear feature convolution, and can be used as *a priory* information for optimal expert estimation concordance according to the proposed technique.

2 Problem Statement

Consider decision making comparison process for similar objects or decision alternatives, which characterized by a set of performance criteria.

Introduce a set of comparable objects or alternatives $\Im = \{\upsilon_1, \ldots, \upsilon_n\}$ and a set of partial performance indexes $\Re = \{\rho_1, \ldots, \rho_m\}$, where each index defines some object feature. It is assumed that a set of features describes the objects performance or quality according to some criteria.

Each object υ_i characterized by the vector of partial performance indexes, namely, measured features $x_i^{\mathrm{T}} = \left(x_i^1, \ldots, x_i^m\right)$, where x_i^j is a j-th actual feature measured value for i-th object (alternative).

Thus a set of observation may be represented in the form of data matrix ("object–feature" table) $X_n = \left\{x_i^j\right\}_{i,j=1}^{n,m}$.

In order to compare the objects or decision alternatives based on expert estimation, it is first necessary to evaluate a generalized performance index or integral indicator [4], which characterized the quality or performance of each object.

The integral indicator is a scalar real-valued function $J(x)$, which is *a priory* unknown and defines by expert's preferences. At that the integral indicator value $J(x_i)$ for each object $\upsilon_i,\ i = \overline{1,n}$ associates with corresponding vector of measured features x_i.

Consider the situation, when integral indicator is constructed using the suitable expert estimations. It is assumed that for investigated set of objects an expert or expert group based on available data and their own preferences produce the vector $q_n^{\mathrm{T}} = (q^1, \ldots, q^n),\ q^i = J(x_i),\ i = \overline{1,n}$ of linear-scaled integral indicator estimates for each object and a vector of feature weights $w^{\mathrm{T}} = (w_1, \ldots, w_m)$, which may be treated as an expert estimates of features relative significance.

It is necessary using available data of measured partial performance indexes (features) X_n and expert estimates q_n, w to reconstruct integral indicator function $J(x)$.

Therefore such an approach may be considered as an expert preference eliciting based on expert indicator and feature weights estimations and "object–feature" data.

The integral indicator model is taken in the quasilinear form $\hat{J}(x) = \phi^{\mathrm{T}}(x)c$, where $c^{\mathrm{T}} = (c_1, \ldots, c_M)$—vector of unknown model parameters, $\varphi^{\mathrm{T}}(x) = (\varphi_1(x), \ldots, \varphi_M(x))$—vector of predetermined coordinate functions, M is a model dimension.

Let us evaluate the parameters estimates for integral indicator model $\hat{J}(x)$ as a solution of optimization problem, which formalized a goal of matching the estimated and the measured integral indicators:

$$R(c) = \sum_{i=1}^{n} \left(q^i - \varphi^{\mathrm{T}}(x_i)c\right)^2 + \Omega(c, w) \rightarrow \min_c \tag{1}$$

The first item in (1) describes the approximation quality of objects performance expert estimates, characterized by integral indicator values for each object, whereas the second one allows expert estimation of features weights and may be considered as an *a priory* information for unknown model parameters c. The specific form of $\Omega(c, w)$ may be evaluated using concordance technique considered below.

Thus the solution of (1) defines the data-based expert preference function model. In fact, an integral indicator approximation $\hat{J}(x)$ may be treated as an optimal nonlinear convolution of partial performance indexes.

The procedure of expert preference function reconstruction based on "object–feature" measured data and concordance of objects performance indexes and features weights expert estimations is illustrated in Fig. 1.

3 Optimal Concordance for Linear Preference Function

3.1 Expert Estimations Concordance

Consider at first a linear model of expert preference function, at that the integral indicator is the linear combination of features and their weights (linear features convolution model of integral indicator).

Using available "object–feature" measured data X_n under linear feature convolution model assumption, a mutual transformation of expert estimates of objects indices q_n vector and features weights vector w can be performed:

$$(q_n, w) \xrightarrow{X_n} \left(q = X_n w, \quad w_n = X_n^+ q_n\right), \tag{2}$$

where linear operator X_n maps the vector of expert estimates of features weights w to the vector q and pseudo-inverse operator X_n^+ maps the vector of expert estimates of objects indexes q_n to the vector w_n.

In general case, the estimated and mapped vectors are usually different, namely $q_n \neq q, w \neq w_n$, so it is necessary to extricate the contradictions between measured data and expert estimates via expert concordance approach.

Vectors of expert estimates of object's integral indicators and features weights will be concordant, if they satisfied the concordance condition $\bar{q} = X_n \bar{w}$.

In accordance to the *Concordance Theorem* [9], there exist such scalar values $\alpha, \beta \in [0, 1]$, $\alpha + \beta = 1$, so vectors $w_\alpha = \alpha w + (1 - \alpha) w_n$, $q_\beta = \beta q + (1 - \beta) q$ meet concordance condition $q_\beta = X_n w_\alpha$.

In such a way, concordant expert estimates of integral indicators and feature weights can be defined as following:

$$\begin{aligned} w_\alpha &= \alpha w + (1 - \alpha) X_n^+ q_n, \\ q_\alpha &= (1 - \alpha) q + \alpha X_n w, \end{aligned} \tag{3}$$

where $\alpha = 1 - \beta \in [0, 1]$ is the object's integral indicators versus feature weights importance parameter.

It is evident, that there exist a set of vectors which satisfy the concordance conditions (3), so the problem of optimal expert estimations concordance arises.

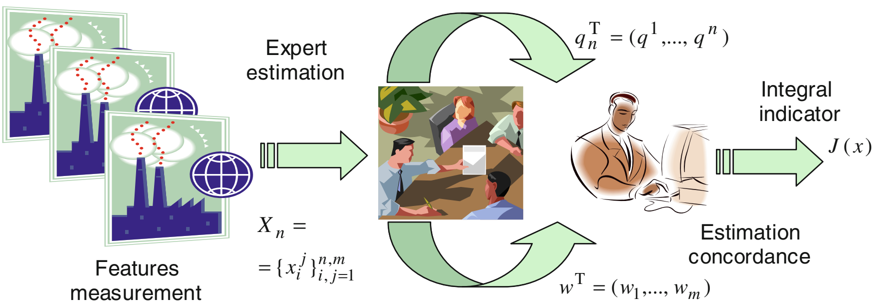


Fig. 1 Preference function reconstruction based on expert estimations concordance and “object–feature” measured data

3.2 Optimal Expert Estimations Concordance

Expert estimates of object’s indicators and features weights will be referred to as *optimally concordant* for linear feature convolution model regarding measured data X_n if they minimize the concordance functional:

$$
\begin{aligned}
&M(q,w) = \frac{1}{2}\|q_n - q\|^2 + \frac{1}{2}\gamma\|(w - w_0)\|^2,\\
&q = X_n w, \quad e^T w = 1,
\end{aligned} \tag{4}
$$

where w_0—vector of *a priory* values of feature weights in linear convolution, γ—weight coefficient, which defines degree of belief for expert estimates of object performance indexes versus feature weights, $e^T = (1\ 1\ldots1)$.

Constrains in (4) match the relationship between object’s indicators and feature weights expert estimation for linear preference function and weight coefficients normalization requirements.

Note that the second item in (4) acts as a regularizing component, which in addition ensures computational stability of optimal concordant expert estimation.

The solution of optimization problem (4) may be obtained using Lagrange function

$$L(q,w,\lambda,\mu) = \mathrm{M}(q,w) + \lambda^T(q_n - X_n w) + \mu\big(1 - e^T w\big). \tag{5}$$

Using Kunch-Takker condition of optimality for (5), the relationship between optimal values of Lagrange multipliers can be obtained:

$$\mu = m^{-1} e^T X_n^T \lambda, \quad \lambda = P_n^{-1}(q_n - X_n w_0), \tag{6}$$

where $P_n = \gamma I_n + X_n \Pi_m X_n^T$, $\Pi_m = I_m - m^{-1} e e^T, I_n, I_m$—identity matrices of appropriate dimension.

Taking into account obtained dependence (6), the solution of optimal expert estimation concordance problem may be obtained in the following form:

$$\begin{aligned}\bar{w} &= w_0 + \Pi_m X_n^{\mathrm{T}} P_n^{-1}(q_n - X_n w_0),\\ \bar{q} &= X_n w_0 + \Psi_n P_n^{-1}(q_n - X_n w_0),\end{aligned} \tag{7}$$

where $\Psi_n = X_n \Pi_m X_n^{\mathrm{T}},\ P_n = \gamma I_n + \Psi_n$.

It is evident, that optimal concordant expert estimations (7) have the following limit properties: $\gamma \to 0:\ \bar{w} = X_n^+ q_n,\ \ \gamma \to \infty:\ \bar{w} = w_0,$ which describe the extreme cases of linear feature convolution weights forming: under full *a priory* information absence $(\gamma = 0)$ and under full confidence for expert estimation of feature weights $(\gamma = \infty)$.

4 Nonlinear Preference Function Reconstruction

4.1 Integral Indicator Identification

Consider a generalization of proposed approach for nonlinear expert preference function model which defines the appropriate nonlinear integral indicator (nonlinear feature convolution).

Consider the nonlinear model of preference function as $\hat{J}(x) = \varphi^{\mathrm{T}}(x)c$, where $c^{\mathrm{T}} = (c_1, \ldots, c_M)$—vector of preference function model parameters, subjected to estimation, $\varphi^{\mathrm{T}}(x) = (\varphi_1(x), \ldots, \varphi_M(x))$—vector of coordinate functions.

Using the kernel-based approach, the coordinate function vectors $\varphi(x_i)$ is taken hereby that its scalar products will be positive define kernel function $K(x, x_i) = \varphi^{\mathrm{T}}(x)\varphi(x_i)$, as a radial-basis function $K(x, x_i) = \exp\left(-\mu\|x - x_i\|^2\right)$, where $\mu > 0$—function parameter.

In accordance with accepted preference function model, vector of integral indicators expert estimates $q_n^{\mathrm{T}} = (q_1, \ldots, q_n)$ for each object may be represented as

$$q_i = \hat{J}(x_i) + \xi_i, \quad i = \overline{1, n}, \tag{8}$$

where ξ_i—"expert measurement" errors, which allow inaccuracy of expert estimates.

From (8) following, that measurement equation for preference function model may be represented in the form:

$$\begin{aligned}q_n &= \Phi_n c + \xi,\\ \Phi_n^T &= (\varphi(x_1), \ldots, \varphi(x_n)),\ \ \xi = (\xi_1, \ldots, \xi_n).\end{aligned} \tag{9}$$

In accordance with support vector approach (SVM) [5], the unknown model parameters estimates may be obtained as a solution of regularized functional minimization problem

$$I(c) = \frac{1}{2}\|\xi\|^2 + \frac{1}{2}\gamma\|(c - c_0)\|^2, \\ \xi = q_n - \Phi_n c, \tag{10}$$

where c_0—vector of *a priory* values of integral indicator model parameters, $\gamma > 0$—regularization parameter, which provides computational stability of estimation procedure.

An equivalent conjugate optimization problem taking into account (9), (10) may be stated using Lagrange function

$$L(c, \xi, \lambda) = I(c) \; + \; \lambda^{\mathrm{T}}(q - \Phi_n c - \xi), \tag{11}$$

where $\lambda^{\mathrm{T}} = (\lambda_1, \ldots, \lambda_n)$—vector of Lagrange multipliers.

Using the conditions for optimality for conjugate optimization problem for Lagrange function (11), the expression for model parameters optimal estimates and conjugate variables may be obtained in the form:

$$c^* = c_0 + \gamma^{-1}\Phi_n^{\mathrm{T}}\lambda^*, \;\; \lambda^* = \gamma \mathrm{A}_\gamma^{-1}(q_n - \Phi_n c_0), \\ \mathrm{A}_\gamma = \gamma I_n + \mathrm{K}_n, \; \mathrm{K}_n = \Phi_n \Phi_n^T = \left\{K(x_i, x_j)\right\}_{i,j=1}^{n,n}. \tag{12}$$

4.2 Optimal Concordance for Nonlinear Preference Function Model

The resulting estimate of preference function model parameters depends of *a priory* value of estimated parameters c_0. To find them it is naturally to use the available information of feature weights expert estimates w, defining the relative importance of particular criteria.

Interpreting the linear convolution of criteria as a first step of preference function nonlinear model approximation, take an optimal *a priory* value of preference function model parameters c_0 from the condition of best approximation of objects integral indexes $q_0 = \Phi_n c_0$ by the vector of its expert estimates obtained for measured data X_n using linear feature convolution $\tilde{q}_n = X_n w$..

Consequently, in order to find *a priory* value of preference function parameters, which are optimally concordant with expert estimations of feature weights w, consider the corresponding optimization problem for auxiliary regularized functional:

$$I_0(c_0) = \frac{1}{2}\|\zeta\|^2 + \frac{1}{2}\omega\|c_0\|^2,$$
$$\zeta = \tilde{q}_n - q_0 = X_n w - \Phi_n c_0. \tag{13}$$

where $\omega > 0$—regularization parameter.

Using the appropriate Lagrange function for optimization problem (13)

$$L(c_0, \zeta, v) = I_0(c_0) + v^{\mathrm{T}}(X_n w - \Phi_n c_0 - \zeta), \tag{14}$$

where $v^{\mathrm{T}} = (v_1, \ldots, v_n)$—vector of Lagrange multipliers, the solution of constrained optimization problem (13) takes the following form:

$$c_0^* = \omega^{-1}\Phi_n^{\mathrm{T}} v^*, \; v^* = \omega \mathrm{A}_\omega^{-1} X_n w,$$
$$\mathrm{A}_\omega = \omega I_n + \mathrm{K}_n, \tag{15}$$

where in accordance with (12) $\mathrm{K}_n = \Phi_n \Phi_n^{\mathrm{T}}$.

From the (12), (15) following, that the resultant expression for nonlinear preference function model, based on feature measured data and optimally concordant expert estimates, takes the following form:

$$\hat{J}_n(x) = \varphi^{\mathrm{T}}(x)\Phi_n d,$$
$$d = [\mathrm{A}_\gamma^{-1} q_n + (I_n - \mathrm{A}_\gamma^{-1}\mathrm{K}_n)\mathrm{A}_\omega^{-1} X_n w]. \tag{16}$$

Finally, taking into account the obvious relation

$$\varphi^{\mathrm{T}}(x)\Phi_n = (K(x, x_1), \ldots, K(x, x_n)), \tag{17}$$

the obtained nonlinear preference function model may be represented as kernel function linear combination determined in the points of corresponding feature measurements:

$$\hat{J}_n(x) = \sum_{i=1}^{n} d_i K(x, x_i), \; d = (d_1, \ldots, d_n), \tag{18}$$

where coefficients d_i of linear kernel function linear combination (18), in accordance with (16), depends from measured data matrix X_n and expert estimates $\{q_n, w\}$.

Thus, formula (18), gives, in fact, the desired preference function model in the form of nonlinear convolution of feature measurements optimally concordant with expert estimates of integral indicators and feature weights.

It should be noted, that obtained estimates depends only from symmetric positive-defined $(n \times n)$ matrix K_n, formed from kernel functions, determined in measurement points; thereof the estimates calculation requires the inversion of well-posed matrix.

Moreover, the final expression for integral indicator estimate (18) doesn't directly include the coordinate functions $\{\varphi_i(x)\}$ in explicit form, which eliminates the need of their prior choice. As a result only kernel function $K(x, x_i)$ is used for indicator model construction.

This enable the possibility of quite complex preference function approximation, at that the number of estimated model parameters doesn't exceed the number of feature measurement.

Later on, obtained preference function model may be used for new objects integral indicator assessment using available feature measurements without direct expert participation.

5 Numerical Example

Consider as an example the problem of integral indicator construction for expert assessment of Thermal Power Plants, which is necessary to compare the plants according to their impact on the environment, so the criterion of ecological footprint is used.

At that the integral indicator of Power Plant should characterize its pollution emission. Such a problem is closely related with Kyoto Index assessment [15].

In order to illustrate the proposed approach, the numerical example from [15] is used, where the compared objects are some of US Power Plants. The main features, according to the criterion of ecological footprint, includes total net generation, emission of anthropogenic greenhouse gas CO_2 and other air pollutants NO_X and SO_2.

The waste measurements are used as statistical information which is necessary for "feature–object" data specification.

Power Plants "feature–object" measured data as well as object's integral indicators expert estimations are taken from http://strijov.com/papers/iiesky.pdf and presented in Table 1.

The expert estimates of feature weights, which used as *a priory* information under expert estimates concordance procedure for nonlinear preference function model, are presented in Table 2.

The Integral Indicator of ecological footprint for considered Power Plants was constructed using the proposed method of preference function kernel-based identification along with expert estimates concordance. In contrast to [15], where integral indicator is taken as a linear convolution of the object's features, the nonlinear integral indicator is designed using presented methodology.

The obtained integral indicator of Power Plants regarding the feature variables $x_1 = CO_2,\ x_2 = NO_X + SO_2$ is presented at Fig. 2.

The obtained nonlinear model of integral indicator more accurate represents the complex structure of expert preferences in comparison with simple linear feature convolution.

Table 1 "Object–feature" measurement data and expert estimation of integral indicators

#	Plant name	Total net generation	Emission			Indicator's expert estimations
			CO_2	NO_X	SO_2	
		10^6 kWh	Short tons per month			0–100
1	Beckford	458,505	191	10	41	76
2	East Bend	356,124	147	16	13	89
3	Miami Fort	484,590	204	28	33	62
4	Zimmer	818,435	329	5	64	24

Table 2 Expert estimation of feature weight

#	Feature	Weight
1	Total net generation	0.2
2	CO_2 emission	0.5
3	NO_X emission	0.2
4	SO_2 emission	0.1

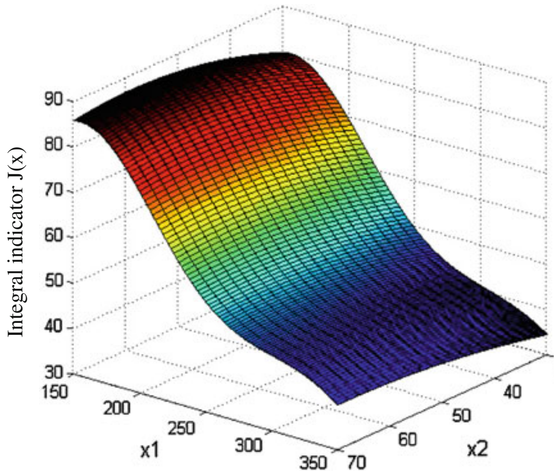


Fig. 2 The constructed integral indicator or power plants

6 Conclusion

The developed approach allows considering the problem of multiple criteria nonlinear convolution as a problem of preference function identification based on both feature measurement data and expert estimates of integral indicators and feature weights. The proposed generalization of expert estimates concordance idea for the case of nonlinear preference function guaranties on optimal concordance of measurement and expert data, whereas machine learning approach coupled with kernel-based technique ensure the possibility of more accurate approximation of expert preference function with complex structure.

The considered identification scheme may be compared with neural network structure with radial-basis activation function and weights, corresponding to estimated preference model parameters. At that under the sufficiently large training data set the suitable learning algorithm may be used for weights tuning. The further development of proposed approach is related with kernel function parameters well as regularization parameters choice.

References

1. Yu, P.: Multiple Criteria Decision Making: Concepts, Techniques, and Extension. Plenum Press, NY, London (1885)
2. Figueira, J., Greco, S., Ehrgott, M.: Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, Berlin (2005)
3. Hora, S.: Expert judgment in risk analysis. In: Everitt, B., Melnick, E. (eds.) Encyclopedia of Quantitative Risk Assessment. Wiley, Chichester (2008)
4. Aivazian, S., Mkhitariyan, V.: Applied Statistics and Essential Econometrics. UNITI, Moscow (1998)
5. Barzilai, J.: Measurement and preference function modeling. Intern. Trans. Oper. Res. **12**, 173–183 (2005)
6. Podvezko, V., Podviezko, A.: Dependence of multicriteria evaluation result on choice of preference functions and their parameters. Technol. Econ. Dev. Econ Baltic J. Sustain. **16**(1), 143–158 (2010)
7. Amponsah, S., Darkwah, K., Inusah, A.: Logistic preference function for preference ranking organization method for enrichment evaluation (PROMETHEE) decision analysis. Afr. J. Math. Compu. Sci. Res. **5**(6), 112–119 (2012)
8. Noghin, V.: Relative importance of criteria: a quantitative approach. J. Multi-Criteria Decis. Anal. **6**, 355–363 (1997)
9. Strijov, V., Shakin, V.: Index construction: the expert-statistical method. Environ. Res. Eng. Manage. **4**(26), 51–55 (2003)
10. Strijov, V., Letmathe, P.: “Integral indicators based on data and rank-scale expert estimations”, Intellectual Information Processing. In: Conference Proceedings, Cyprus, pp. 107–110, October 2010
11. Ren, Y., Papalambros, P.: Design preference elicitation: exploration and learning. Tn: Proceedings of 18th International Conference on Engineering Design, vol. 10, pp. 149–158 (2011)
12. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)
13. Suykens, A., Gestel, T., Brabanter, J., Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. World Scientific, Singapore (2002)
14. Smola, A.: Learning with kernel. MIT Press, Cambridge, MA (2002)
15. Molak, V., Shakin, V., Strijov, V.: Kyoto index for the power plants in the USA. In: The third Moscow International Conference on Operations Research (ORM2001), pp. 80–81. Moscow, Russia (2001)

A Multi Criteria Group Decision Making Process Based on the Soft Fusion of Coherent Evaluations of Spatial alternatives

Gloria Bordogna and Simone Sterlacchini

Abstract In this chapter, a soft fusion approach of a set of aligned registered images is described, in which each image represents a distinct theme of the same territory. The objective of the fusion is to evaluate the spatial units of the territory, pixels or cells, by a group of experts through associating to the spatial units scores representing their suitability as best locations with respect to given criteria expressed by soft constraints. The decision process first determines automatically the coherence among a majority of experts' evaluations with respect to each given criterion and then computes the coherent evaluation of the criterion representing the majority opinion. Finally, it computes the overall coherent evaluation with respect to a majority of the criteria that is used to rank the spatial units. The fuzzy majority is defined by linguistic quantifiers; the coherence among a fuzzy majority of experts is defined based on the Minkowski OWA operators while the fuzzy majority of coherent evaluations is modeled by an Induced OWA operator.

1 Introduction

Many applications require the evaluation of spatial information in order to take decisions: examples are environmental, social, and economic resources location and allocation, environmental status indicators definition, impact factor evaluation.

G. Bordogna (✉)
Institute for Electromagnetic Sensing of the Environment CNR—National Council of Research, Via Bassini 15, Milano, Italy
e-mail: bordogna.g@irea.cnr.it

G. Bordogna · S. Sterlacchini
Institute for the Dynamics of Environmental Processes CNR—National Council of Research University of Milano Bicocca, Milano, MI, Italy
e-mail: sterlacchini@idpa.cnr.it

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_5,

Often, a group of experts is asked to select or rate a set of alternatives which are distinct spatial units (locations or areas) of a territory for the target purpose [1]. This is done by evaluating some of the territory's properties, represented in distinct registered aligned thematic images of the territory, based on distinct criteria. Generally, the criteria are expressed by linguistic terms such as *low, medium, high*, whose semantics can be represented within the fuzzy framework by soft constraints on the distinct properties of the territory [2]. For example, the disposition of an area to the occurrence of fires is greater if the area has a *high* aridity while it is *low* if it is a wetland. Each expert can associate a distinct semantics to the linguistic terms so that the experts' opinions can be coherent among each other on some property but may conflict on others.

In this chapter, a soft fusion approach of coherent evaluations of spatial alternatives is described. In this approach the coherent evaluations of each property are aggregated to generate the group evaluation of the spatial alternatives by a soft fusion process, consisting of two steps: in the former, each expert evaluates each alternative, i.e., pixel of an image, by specifying soft constraints; in the latter, the coherent evaluations of the experts on each criterion are aggregated so as to weight differently their impact in determining the global judgment of each alternative by considering their distinct trust determined, based on the coherence of their evaluations.

Linguistic quantifier-guided Ordered Weighted Averaging (*OWA*) operators [3, 4] have been indicated as a powerful means to define soft fusion strategies of spatial information based on the concept of fuzzy majority 155,319. In [5] a fuzzy majority approach has been proposed to model a land suitability problem based on multi criteria group decision making, where the group solution map synthesizes the majority of the decision-makers' preferences. In their approach the quantifier-guided fusion function generating the group map is defined by an Induced Ordered Weighted Averaging (*IOWA*) operator as proposed in [6] so that the opinions of the decision makers supporting each other in the fuzzy majority are taken into account in generating the group map. However, in this approach, the evaluations on each criterion are expressed by linguistic terms that are internally converted into fuzzy numbers and aggregated by each expert independently from the other experts. So the subjective interpretation of the semantics of linguistic terms by each expert is not modeled but is assumed as homogeneous. Further, the distinct trust of the experts is not modeled.

In our approach, the opinions of the experts on each criterion are compared to assess their coherence so that the final evaluation of each pixel is generated by a distinct fuzzy majority of experts on each criterion. In fact, the coherent evaluation is obtained by a quantifier-guided fusion, defined by an *IOWA*, where both the weights of the *IOWA* and the order of the arguments are based on the coherence and the distinct trust of the experts.

In Sect. 2, the classic approach to model group decision making is analyzed and compared with our proposal. In Sect. 3, the evaluation of alternatives on single criteria by each expert is described. In Sect. 4, we propose the computation of a soft coherence measure. In Sect. 5, the proposal to compute the coherent evaluation of each criterion expressing the opinion of a fuzzy majority of experts is described. In Sect. 6, the approaches to determine the overall evaluation of alternatives are outlined. The conclusions summarize the main contents.

2 Rational of the Approach with Respect to the Classic Group Decision Making Procedure

The objective of a group decision activity is the identification of the alternative(s) which is (are) judged the best by the majority of the experts.

In a spatial context, the alternatives can be distinct locations, i.e., pixels or areas of a territory that is described by several registered aligned thematic maps.

Experts give their the options on the basis of some evaluation scheme (see Fig. 1); this can be either implicitly assumed by the experts, or explicitly specified in the form of a set of predefined qualitative criteria, as in the case we are considering in this chapter.

Examples of qualitative criteria are *high* slope, *mostly* sand lithology, etc. Each expert can specify his/her subjective meaning of a qualitative criterion by defining its semantics through a soft constraint that the pixels of an image representing a theme must satisfy, such as a minimum, maximum, acceptance level or range of values of slope, types of lithology, etc.

For each alternative, a pixel, a performance judgment is obtained by each expert by evaluating each soft constraint defined on the domain of the pixels' values in an image representing a property of the territory.

At this point, the first goal of most of the group decision making approaches defined in the literature is the computation of the expert's performance judgment of each alternative by aggregating the scores obtained by applying all the soft constraints defined on the available images. To this end a desired trade-off among the satisfaction degrees of soft constraints can be adopted by each expert, so that the expert's performance judgment reflects a desired concurrency or compensation among the criteria.

Other techniques to evaluate the alternatives are based, for example, on the specification of preference relations or subjective probabilities or linguistic terms defined on an ordinal scale [7–10].

Once the alternatives have been evaluated by all the experts, the main problem is to compare experts' judgments to verify the consensus among them [11]. In case of unanimous consensus, the evaluation process ends with the selection of the best alternative(s) that corresponds to the alternative(s) with the greatest consensual performance judgment.

However, in real situations humans rarely come to unanimous agreement: this has led to evaluate not only crisp degrees of consensus (degree 1 for full and unanimous agreement) but also intermediate degrees between 0 and 1, corresponding to partial agreement among all experts. Furthermore, full consensus

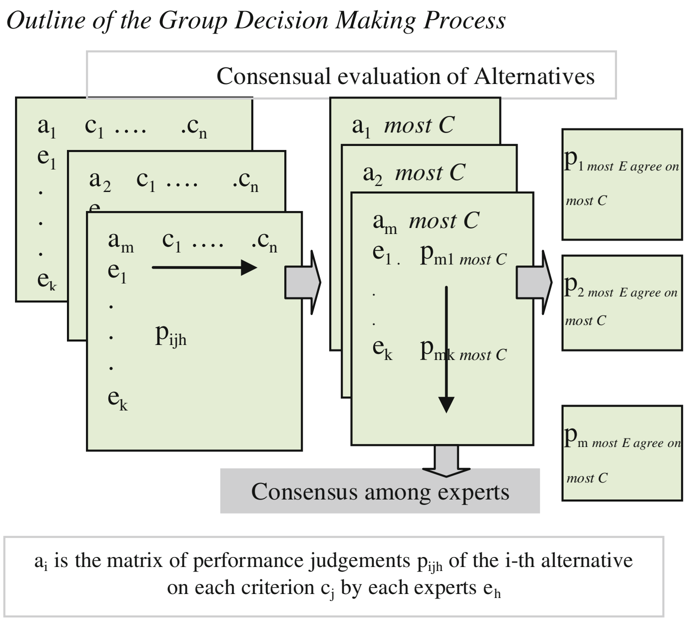


Fig. 1 Classic schema of a multi criteria group decision process

(degree = 1) can be considered not necessarily as a result of unanimous agreement among *all* the experts, but it can be obtained even in case of agreement among a fuzzy majority of the experts, for example *most* experts 1,411. An outline of this two-step decision process is depicted in Fig. 1 where each matrix contains the performance judgments of an alternative by each expert on each criterion: the first step computes the global evaluation of each alternative by each expert on all or most criteria. Consensus in computed based on a similarity measure among the experts' performance judgments of the alternatives.

In our approach, we focus on the fact that with the classic multi criteria group decision making schema two experts may obtain the same performance judgement for an alternative even if their single performance judgments of the criteria are completely different.

In fact, what matters in this approach is the overall performance of each alternative by each expert, and their comparison is irrespective of the conflicts or incoherence, between the performance judgments of the single criterion.

This scheme is suited to compute a global evaluation of the alternatives by a group of experts in two distinct situations:

- in case of implicit criteria, when each expert uses a different set of criteria to evaluate the alternatives;
- in the case of explicit criteria, when their aggregation can be performed freely by each expert based on their distinct decision attitudes. For example, an expert

can consider all criteria completely not compensative one another for taking the decision, i.e., they must be all satisfied and, on the contrary, another expert considers the lack of satisfaction of a criterion completely replaceable by the satisfaction of any other criterion as proposed in [5].

Nevertheless this decision schema is not adequate in two other cases:

- in the cases in which we need to introduce **uniformity** in considering the compensation/concurrency of the performance judgments of the given criteria. This can be necessary **for providing distinct scenarios modelling distinct attitudes to risk of the decision maker**: from example one taking into account just the best satisfied criterion, and the opposed one considering all the criteria;
- in the case in which one wants a robust decision, determined by taking into account the performance judgments on each criterion of the experts who expressed coherent evaluations of the criterion;
- in order to take into account the distinct subjective semantics of the same linguistic terms used by experts to judge the alternatives based on each criterion.

These three last cases are those that most often occur in the case of decisions involving the evaluation of spatial alternatives, generally performed based on a set of explicit criteria. In such case the decision requires the uniformity of the decision attitude in determining the disposition/hazard of regions of a territory to natural events. For example, in the case of evaluation of susceptibility maps of environmental phenomena, the experts are indeed models, implemented by distinct software tools, and specialized to compute the performance judgment of a criterion, i.e., a factor score contributing to the occurrence of the phenomenon, which is due to a property of the territory. A more robust global performance judgement of the criterion can be assessed by taking into account the coherent evaluations of the single tools or experts.

2.1 Outline of the Group Decision Making Process

Figure 1 recaps the classic group decision making process and Fig. 2 our proposed approach. It can be seen that in the classic schema first the fusion of the criteria evaluations for each expert is performed and then the fusion of the experts judgements on all criteria, while in our proposal first the coherent experts' performance judgements on each criterion are computed and successively they are fused to compute the overall coherent performance of each alternative.

In this last case, first we compute for each criterion the coherence among the performance judgments among all (or most) experts. Notice that, for each criterion and each alternative a distinct subset of experts can express coherent evaluations.

In doing this, central to our approach is the definition of coherence among the experts' evaluations over a single criterion.

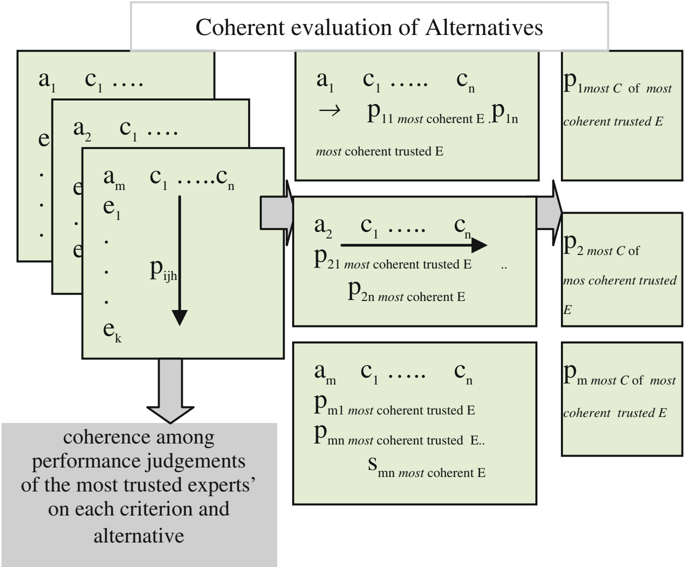


Fig. 2 Proposed schema of decision process computing the overall performance of alternatives by the coherent evaluation of criteria by a fuzzy majority of experts

Then, the global satisfaction of each criterion for each alternative is computed by aggregating the coherent performance judgments of the experts. Once we have these global satisfactions of all the criteria for each alternative, we can aggregate them to derive a final ranking of the alternatives: this last aggregation can have distinct semantics, depending on the fact that we want to consider the criteria completely compensative one another or not, or partially compensative.

In the next section, we are going to describe this multi criteria group decision process which is soft at distinct levels:

- in modelling the distinct reliability of spatial data and trusts of the experts;
- in modelling the distinct interpretations of coherence among the experts performance judgments on each criterion;
- in computing the coherent performance judgement on each criterion by the fuzzy majority of coherent experts,
- finally, in modelling the decision attitude that is most appropriate.

3 Evaluating Alternatives on Single Criteria

The objective of the decision process is ranking the locations (spatial units) of a given territory that have characteristics suitable for the allocation or location or resources, or for the identification of spatial objects or even areas susceptible to the occurrence of a phenomenon according to a group of experts or models. This is a multi criteria decision making process where the criteria are the factors that must be assessed independently and then combined to assess the suitability of the areas of the territory.

The decision process is performed on each spatial unit independently from the others.

For sake of simplicity, as sources of information we assume to have spatial data consisting of M registered aligned images, each one describing a physical or administrative property of the territory under study.

The domains of pixels values in each image can be different: they can be numeric continuous or discrete, ordinal or nominal.

We consider a group of K experts that are called to evaluate the spatial units based on a set of N common criteria.

Each criterion c_i is expressed by a linguistic predicate whose semantics is defined by a soft constraint on the domain of values of an image. For example "*high slope*" is defined with a non decreasing membership function on the domain [0, 90] of the slope map.

Each expert can specify his/her subjective definition of the membership functions of the soft constraints on the domain of the property. This way, the experts' evaluations of the distinct criteria are made consistent one another and can be aggregated.

Further, to each expert we can associate a trust value $t_1,\ldots,t_k \in [0, 1]$ to represent either their skill and ability in performing the evaluations, or the validity of the evaluation due to several reasons: either because the source of the data cannot be completely trusted or because one knows that the means of acquisition are not enough sophisticated and generate systematic errors; not least, because data are a result of a subjective analysis, such as surveyed data.

4 Evaluating the Coherence of Performance Judgments of a Fuzzy Majority of Experts

4.1 Definition of Coherence

Let us define a coherence measure between two vectors $A = (a_1,\ldots, a_K)$ and $B = (b_1,\ldots, b_K)$ with $a_i\ b_i \in [0, 1]$ based on the weighted Minkowski distance, by considering the trust $(t_1,\ldots, t_K)$ (such that they sum to 1) associated with each dimension of the two vectors, as follows:

$$coherence(A,B) = \left(\sum_{k=1}^{K} t_k(1 - |a_k - b_k|)^{\alpha}\right)^{\frac{1}{\alpha}} \tag{1}$$

with $-\infty < \alpha < +\infty$.

We can obtain a range of special similarity measures by setting distinct values for the norm parameter α. For example, when $\alpha = 1$ and all trust scores are equal, we obtain the inner product. If $\alpha = 2$, the complement of the weighted Euclidean distance. If $\alpha = 0$, the complement of the weighted geometric distance. If $\alpha = -1$ the complement of the weighted Harmonic distance.

Our first purpose is to compute the coherence of an expert with respect to *all* the other experts of the group in evaluating each single alternative *a* with respect to each single criterion *c* by taking into account the distinct trust scores of the experts $(t_1,\ldots,t_k)$. To this end we assume that a *K* dimensional space is defined in which all the experts' judgments on alternative *a* with respect to criterion *c* are represented by a vector:

$$P = (p_{\mathrm{ac1}}, \ldots, p_{\mathrm{acK}}).$$

The coherence of the i-th expert evaluation p_{aci} with respect to all the other experts is defined as *coherence* (P, R_i), according to definition (1), in which $R_i = (p_{\mathrm{aci}},\ldots, p_{\mathrm{aci}})$ is the *K* dimensional reference vector representing the opinion of the i-th expert.

Coherence(P, R_i) considers all the *K* experts (all single components of the two vectors *P* and R_i).

A more flexible definition of the coherence measure taking into account a fuzzy majority *Q* of the experts and their distinct trust can be obtained by computing it through a Minkowski similarity OWA operator (*MOWA*) of dimension *K* with importance weights as defined in the following subsections [12].

4.2 Definition of the MOWA Operator

MOWA: $R^K \times R^K \rightarrow R$ has an associated weighting vector *W* of dimension *K* such that $\sum_{i=1}^{K} w_i = \mathbf{1}$ and $w_\mathrm{i} \in [0, 1]$ and

$$MOWA(s_1, \ldots, s_K) = \left(\sum_{i=1}^{K} w_i S_i^{\alpha} \right)^{\frac{1}{\alpha}} \quad \text{with } -\infty < \alpha < +\infty \tag{2}$$

where S_i is the *i*-th largest of the $s_\mathrm{j} = 1 - |a_\mathrm{i} - b_\mathrm{i}|$; it is the individual similarity between the *i*-th components of the two vectors *A* and *B*.

The main aspect of the *MOWA* operator is the reordering of the arguments based upon their values. This means that the weights of *W*, instead of being associated with a specific argument are associated with a particular position in the ordering. This reordering makes the *MOWA* operator a non linear operator. Thus, while the weights $(t_1,\ldots, t_n,)$ in formula (1) are indications of the reliability of the vector components, the semantics of *W* in formula (2) is the importance of an ordered position of the vector components, thus its choice determines the semantics of the aggregation.

In particular, if the weighting vector W is such that $w_1 = 1$ then we obtain as a result of $MOWA(s_1,\ldots,s_K) = \max(s_1,\ldots,s_K)$ instead when $w_K = 1$ we obtain $MOWA(s_1,\ldots,s_K) = \min(s_1,\ldots,s_K)$.

To compute an "or-like" similarity between two vectors, which means weighting more heavily the contributions of the greatest individual similarities, we can set $w_i > w_j$ for $i < j$, while, on the contrary, if we want an "and-like" semantics in order to minimize the overall similarity between two vectors we can weight more heavily the smallest individual similarities by setting $w_i < w_j$ for $i < j$.

When all the $w_i = 1/K \ \forall i = 1,\ldots,K$, the *MOWA* operator reduces to definition (1) with equal trust scores.

In [13] it has been proposed that a monotone not decreasing linguistic quantifier [14] $Q:[0, 1] \rightarrow [0, 1]$ can be used to specify the concept of a fuzzy majority for modeling group decision making process. By a linguistic quantifier it is possible to defined flexible notions of majority. The crisp notion of majority (linguistically expressed by *greater than* 50 %) corresponds to $Q(x) = 1$ f-or $x = round(K/2)$, $Q(x) = 0$ otherwise. If $Q_1(x) \le Q_2(x)$ $\forall x$ in [0, 1] the linguistic quantifier Q_1 defines a stricter fuzzy majority than Q_2, as in the case of the quantifiers *almost all* with respect to *most*.

Here we use this notion to generate a quantifier-guided *MOWA* operator computing a value reflecting the truth of a proposition:

Q most trusted dimensions of the two vectors A and B are similar.

This is achieved by computing the weights w_i for $i = 1,\ldots,K$ of of *MOWA* as it has been done for the *OWA* operator with distinct importance in [15, 16]:

$$w_i = Q\left(\sum_{j=1}^{i} e_j\right) - Q\left(\sum_{j=0}^{i-1} e_j\right) \text{ with } \sum_{i=1}^{K} e_i = \sum_{i=1}^{K} t_i = 1 \quad (3)$$

where $e_j = t_h$, the trust score of S_h in definition (2) which is the h-th greatest among the arguments $s_1,\ldots,s_K$.

It can be shown that the w_i obtained by applying (3) are defined in [0, 1] and their sum is 1.

This way w_i increases with the trust of the argument to which it is associated. The arguments with null trust play no role and have a zero weight.

4.3 Definition of Qcoherence

Now we can define a quantifier notion of coherence measure between two vectors A and B, *Qcoherence,* based on the *MOWA* operator with weighting vector obtained by applying definition (3) as follows:

$$Qcoherence(A,B) = MOWA(1 - |a_1 - b_1|, \ldots, 1 - |a_K - b_K|) \tag{4}$$

Formula (4) reduces to the coherence measure in (1) when Q = *averagely all* is the identity function $Q(x) = x$, $\forall x \in [0, 1]$.

By defining two fuzzy majorities Q_1 and Q_2 such that $Q_1(x) \leq x \leq Q_2(x)$, $\forall x$ in [0, 1] we have:

$$Q_1coherence(A,B) \leq coherence(A,B) \leq Q_2coherence(A,B) \tag{5}$$

which means that the coherence measure defined in (4) between two same vectors A and B with same trust scores produces a lower equal value and an upper equal value with respect to the *coherence* measure in (1).

The coherence of the *i*-th expert evaluation p_{aci} of alternative a with respect to criterion c and by considering a fuzzy majority Q of experts with distinct trust can be computed as $Qcoherence(P, R_i)$ according to definition (3) and (2) in which $R_i = (p_{aci}, \ldots, p_{aci})$ is the K dimensional reference vector representing the opinion of the *i*-th expert.

5 Soft Fusion of Coherent Performance Judgements

5.1 Fusing the Coherent Evaluations on Each Criterion

Let us indicate by $coherence_i = Qcoherence(P, R_i)$ the coherence of an expert i with Q other experts. If $coherence_i$ is low, it means that the i-th expert does not belong to the fuzzy majority Q of the group of experts expressing coherent performance p_{ac} of the criterion c for the alternative a.

The values $coherence_i$ can be used to indicate the weight of the *i*-th expert's evaluation p_{aci} in determining the group coherent evaluation of the alternative a with respect to the criterion c by a fuzzy majority Q of coherent trusted experts. This corresponds to compute $p_{ac\ of\ Q\ cohrent\ trusted\ experts}$, i.e., the opinion of the Q majority of most trusted experts that expressed coherent evaluations of alternative a with respect to criterion c.

This can be done as proposed in [6] by applying an *IOWA* operator defined in the following subsection.

5.2 Definition of the IOWA Operator

The *IOWA* operator of dimension K is a non linear aggregation operator *IOWA*: $[0, 1]^K \rightarrow [0, 1]$ with a weighting vector $W = (w_1, w_2, \ldots, w_K)$, with $w_j \in [0, 1]$ and $\sum_{j=1}^{K} w_j = 1$, defined as:

$$IOWA(<x_1, u_1, >, \ldots, <x_K, u_K>) = \sum_{i=1}^{K} w_i x_{u-index(i)} \tag{6}$$

in which $X = (x_1, \ldots, x_K)$ is the argument vector to be aggregated and $U = (u_1, \ldots, u_K)$ is the inducing order vector such that it determines the order in which the elements of X have to be taken into account in the aggregation. Specifically, $x_{u\text{-}index(i)}$ is the element of vector X associated with the i-th smallest inducing order value u among the values $(u_1, \ldots, u_K)$.

Example For example, given $W = (0, 0, 0.5, 0.5)$, $X = (0.1, 0.3, 0.8, 1)$, $U = (3, 8, 6, 2)$ we obtain:

IOWA(<0.7,3>, <0.2,8>, <0.8,6>, <1,2>) = = 0*1 + 0*0.7 + 0.5*0.8 + 0.5*0.2 = 0.5 while considering a distinct inducing order vector $U = (8, 1, 2, 4)$ we obtain: *IOWA*(<0.7, 8>, <0.2, 1>, <0.8, 2>, <1, 4>) = 0*0.2 + 0*0.8 + 0.5*1 + 0.5*0.7 = 0.85.

In order to compute the opinion of a fuzzy majority Q in [6] it was proposed to derive the inducing order vector U and the weighting vector W based on the support *Supp* of each argument to aggregate, which is a measure of the proximity of the argument to the other arguments defined as follows:

$$Supp(x_i, X) = \sum_{j=1\ldots,K} (1 \text{ if } |x_i - x_j| < \beta \text{ else } 0). \tag{7}$$

Then, the inducing order vector is defined as:
$U = (u_1, \ldots, u_K) = (Supp(x_1, X), \ldots, Supp(x_K, X))$

Further, given the quantifier Q and U the weighting vector $W = (w1, \ldots, wK)$ of the *IOWA* operator is derived as follows:

$$w_i = \frac{Q(m_i/K)}{\sum_{j=1}^{K} Q(m_j/K)} \tag{8}$$

in which $m_i = argmin_i(u_1, \ldots, u_K)$ is the i-th smallest element among $(Supp(x_1, X), \ldots, Supp(x_K, X))$.

With this definition the *IOWA* operator do not take into account the distinct trust associated with the arguments to aggregate. Specifically, U is defined independently from the distinct trust scores $(t_1, \ldots, t_K)$ of the elements in X.

In order to compute the performance of alternative a with respect to criterion c by Q most trusted experts so as to take into account the trust score t_i of each expert i and his/her coherence $Qcoherence(P, R_i)$ with Q other trusted experts, we propose to set the induced ordering vector U as follows:

$$U = (Qcoherence(P, R_1), \ldots, Qcoherence(P, R_K)) \tag{9}$$

and then we compute $p_{ac\ of\ Qcohrent\ trusted\ experts}$ by applying an *IOWA* operator:

$$p_{ac\ of\ Qcohrent\ trusted\ experts} \quad = IOWA(<p_{ac1}, u_1>, \ldots, <p_{acK}, u_K>) \tag{10}$$

in which the weighing vector *W* is obtained by following definition (8) in which we have computed:

$$m_i = \frac{\text{argmin}_i(u_1 * t_1, \ldots, u_K * t_K)}{\sum_{i=1}^{K} \text{argmin}_i(u_1 * t_1, \ldots, u_K * t_K)} \tag{11}$$

this way m_i is the *i*-th smallest element among the *Qcoherence* of the experts multiplied by their trust degrees:

$(Qcoherence(P, R_1)^* t_1, \ldots, Qcoherence(P, R_K)^* t_K)$.

5.3 Example of Computation of the Opinion of a Fuzzy Majority of Experts Without Considering Their Trust

Consider the following quantifier *most*

$$most(y) = \begin{cases} 1 & y \geq 0.8 \\ 2y - 0.6 & 0.3 < y < 0.8 \\ 0 & y \leq 0.3 \end{cases} \tag{12}$$

and the performance judgments of the experts.

$P = (0.1, 0.6, 0.8, 0.2, 0.9)$ with trust $T = (0.4, 0.2, 0, 0.4, 0)$.

Notice that experts [17] and [5] are not trusted at all $t_3 = t_5 = 0$. Thus, their evaluations 0.8 and 0.9 should be disregarded in computing the group evaluation, while the evaluation of the second expert, having trust $t_2 = 0.2$, should not influence too much the result.

Let us set $\beta = 0.3$ for computing $U = Supp$ in formula (7).

By applying (6) we obtain:

$W = (0.143, 0.143, 0.143, 0.143, 0.428)$, and, finally, $IOWA(P, U) = 0.599$ that is a bit closer to the performance judgements of the not trusted experts that to most trusted ones.

5.4 Example of Computation of the Opinion of a Fuzzy Majority of Experts by Considering Their Distinct Trust

Now, let us make the same example discussed above with our method having *most* as defined in formula (12),

$P = (0.1, 0.6, 0.8, 0.2, 0.9)$ and trust $T = (0.4, 0.2, 0, 0.4, 0)$.

We first compute the *Qcoherence* of each expert by applying definitions (4) based on the application of formula (3) for computing W and on definition (2) with parameter $\alpha = 1$:

Qcoherence = (0.86, 0.64, 0.44, 0.88, 0.34).

Then we set the inducing order vector U = *Qcoherence* as defined in (9).

By applying definition (11) taking into account *Qcoherence* and T = (0.4, 0.2, 0, 0.4, 0) we obtain: M = (0, 0, 0.155, 0.417, 0.427).

By applying definition (8) we obtain the weighting vector W = (0, 0, 0, 0.48, 0.52).

Finally, we can apply the fusion defined by the *IOWA* in definition (6):

IOWA(<0.1,0.86>, <0.6,0.64>, <0.8,0.44>, <0.2,0.88>, <0.9,0.34>) = 0.1*0.48 + 0.2*0.52 = 0.152.

This result 0.152 is much lower than 0.599, obtained by the basic method proposed in [6], and better synthesises the coherent evaluations of the trusted majority of experts.

6 Computing the Coherent Evaluation of Each Alternative

Once we have the group coherent evaluations on all criteria for each alternative, computed as proposed in the previous section, we can aggregate them in order to compute the overall coherent evaluation of each alternative. This last aggregation can be defined as outlined in Sect. 1 by considering the objective of the decision process and the decision attitude.

Either total or partial compensation among the criteria can be modeled by aggregations defined for example by *OWA* operators [4, 16] and Generalized Conjunctive Disjunctive *GCD* operators [18]. When requiring that all criteria are fully, simultaneously, satisfied to state the disposition of a spatial unit to the occurrence of a critical natural event, one models a cautious decision, in the sense that he/she does not want to overestimate the risk that might happen in the given area. On the contrary, by requiring full compensation among the criteria one models an alarming decision in the sense that the satisfaction of one criterion is enough to set an alarm on the spatial unit.

Another aspect of the aggregation of the criteria evaluations is modeling their partial optionality, so that the overall evaluation of the alternative that satisfies the mandatory criteria is reworded or penalized depending on the degree of satisfaction of the optional criteria. This kind of aggregations can be modeled, for example, by discounting operators [19] and non monotonic operators [20].

A further aspect is modeling the hierarchical structure of the criteria: the criteria can be organized in a tree hierarchy that represents the criteria and sub-criteria desired aggregation flow, reflecting the decision preferences regarding the combination of the different criteria evaluations. This can be modeled by Logic Scoring of Preference (LSP) operators [21] as in [1, 22].

Finally, the overall evaluation of an alternative could synthesize the coherent evaluations of a fuzzy majority of the criteria. This can be done by specifying a linguistic quantifier Q and applying the procedure based on the *IOWA* operator proposed in the previous section.

7 Conclusions

The chapter proposes a multi criteria group decision making process for ranking spatial alternatives based on a soft fusion of coherent evaluations. The approach computes for each spatial unit of a territory an overall evaluation that reflects the coherent performance judgments of distinct groups of experts. This allows modeling the fact that the experts may have inhomogeneous knowledge of the area they are called to evaluate, i.e. they know better the conditions characterizing a given place with respect to the others.

This process could be applied also to rank any type of alternative in the cases in which it is important to take into account the coherence among the experts' evaluations on each criterion, or when the aggregation of the criteria evaluations must be homogeneous for all experts, reflecting a given decision attitude or need.

The novelty of the proposal is the definition of the coherence of a fuzzy majority of the experts performance judgments based on *MOWA* operators, and the definition of a new approach based on *IOWA* operators for computing the representative opinion of a fuzzy majority of a group of experts with distinct trust.

References

1. Dujmović J.J., De Tré G., Van de Weghe N.: LSP suitability maps. Soft Comput. **14**, 421–434 (2010)
2. Robinson, P.B.: A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. Trans. GIS **7**(um1), 3–30 (2003)
3. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. Syst. Man Cybern. **18**, 183–190 (1988)
4. Yager, R.R.: A framework for multi-source data fusion. Inf. Sci. **163**, 175–200 (2004)
5. Boroushaki, S., Maczewski, J.: Using the fuzzy majority approach for GIS-based multicriteria group decision-making. Comput. Geosci. **36**, 302–312 (2010)
6. Pasi, G., Yager, R.: Modeling the concept of majority opinion in group decision making. Inf. Sci. **176**, 390–414 (2008)
7. Bezdek, J., Spillman, B., Spillman, R.: A fuzzy relation space for group decision theory. Fuzzy Sets Syst. **1**, 255–268 (1978)
8. Herrera, F., Herrera-Viedma E.: Linguistic decision analysis: Steps for solving decision problems under linguistic information. Fuzzy Sets Syst. **115**(1), 67–82 (2000)
9. Herrera, F., Herrera-Viedma, E., Martnez, L.: A fusion approach for managing multi-granularity linguistic terms sets in decision making. Fuzzy Sets Syst. **114**(1), 43–58 (2000)
10. Kacprzyk, J., Roubens, M. (eds.): Non conventional Preference Relations in Decision Making, Springer, Berlin (1988)

11. Kacprzyk, J., Fedrizzi, M., Nurmi, H.: Group decision making and consensus under fuzzy preferences and fuzzy majority. Fuzzy Sets Syst. **49**, 21–31 (1992)
12. Merigo, J.M., Gil-Lafuente, A.M.: Using OWA operator in the Minkowski distance. Int. J. Soc. Hum. Sci. **2**, 564–572 (2008)
13. Kacprzyk, J.: Group decision making with a fuzzy linguistic majority. Fuzzy Sets Syst. **18**, 105–118 (1986)
14. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Comput. Math. Appl. **9**, 149–184 (1983)
15. Yager, R.R.: Interpreting linguistically quantified propositions. Int. J. Intell. Syst. **9**, 541–569 (1994)
16. Yager, R.R.: Quantifier guided aggregation using OWA operators. Int. J. Intell. Syst. **11**, 49–73 (1996)
17. Bordogna, G., Boschetti M., Brivio A., Carrara P., Pagani M., Stroppiana D.: Fusion strategies based on the OWA operator in environmental applications, in recent developments in the ordered weighted averaging operators: theory and practice. In: Yager, R. and Kacprzyk, J. Beliakov, G. (eds.), Studies in Fuzziness and Soft Computing, vol. 265/2011, pp. 189–207. Springer, Berlin (2011)
18. Dujmović J.J., Larsen H.L.: Generalized conjunction/disjunction. Int. J. Approximate Reasoning **46**(3), 423–446 (2007)
19. Dubois, D., Prade, H.: Combination of Fuzzy Information in the Framework of Possibility Theory. In: M.A. Abidi, R.C. Gonzalez (eds.), pp. 481–505. Academic Press, New York (1992)
20. Yager, R.R.: Nonmonotonic set theoretic operations. Fuzzy Sets Syst. **42**, 173–190 (1991)
21. Dujmović J.J.: Preference logic for system evaluation. IEEE Trans. Fuzzy Syst. **15**(6), 1082–1099 (2007)
22. Dujmović J.J., Scheer D.: Logic aggregation of suitability maps. In: Proceedings of the IEEE World Congress on Computational Intelligence, Barcelona, Spain, pp. 2222–2229 (2010)
23. Bone, C., Dragicevic, S., Roberts, A.: Integrating high resolution remote sensing, GIS and fuzzy set theory for identifying susceptibility areas of forest insect infestations. Int. J. Remote Sens. **26**(21), 4809–4828 (2005)
24. Bordogna, G., Fedrizzi M., Pasi G.: A linguistic modeling of consensus in group decision making based on OWA operators. IEEE Trans. Syst. Man Cybern. A **27**(1), 126–133 (1997)
25. Fedrizzi, M., Kacprzyk, J., Zandrozny, S.: An interactive multi user decision support system for consensus reaching processes using fuzzy logic with linguistic quantifiers. Decis. Support Syst. **4**, 313–327 (1988)
26. Tran, L.T., Knight, C.G., O'Neill, R.V., Smith, E.R., Riitters, K.H., Wickham, J.: Environmental assessment, fuzzy decision analysis of integrated environmental vulnerability assessment of the Mid-Atlantic region. Environ. Monit. **29**(6), 845–859 (2002)

Distributed Manufacturing Scheduling Based on a Dynamic Multi-criteria Decision Model

M. L. R. Varela and R. A. Ribeiro

Abstract Distributed manufacturing scheduling is increasingly necessary in nowadays global manufacturing environments and assumes primal importance to ensure enhanced solutions for such globally distributed manufacturing scheduling problems. In this chapter an approach based on a dynamic multi-criteria decision model is proposed, which enables (re)scheduling strategies and trade-offs between different performance measures. In this dynamically changing environment, real-time changes may occur in production and there is a need for a global view and manufacturing (re)scheduling to improve the globally distributed manufacturing scenario. The approach main aim is to support scheduling decision making, namely through reliable and timely deliveries, as well as improved manufacturing management of available resources. An illustrative example, integrating a set of manufacturing cells is provided to clarify the approach.

1 Introduction

In dynamic and stochastic manufacturing systems, production planners and manufacturing engineers can benefit from better understanding how (re)scheduling strategies affect system's performance. This knowledge will support them to design and operate better manufacturing scheduling systems.

M. L. R. Varela
Department of Production and Systems, School of Engineering, University of Minho, 4800-058 Guimarães, Portugal
e-mail: leonilde@dps.uminho.pt

R. A. Ribeiro (✉)
UNINOVA - CA3, Campus FCT-UNL, 2829-516 Caparica, Portugal
e-mail: rar@uninova.pt

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_6,

Dynamic scheduling has been solved using different approaches and techniques, as for instance [1–4]: heuristics, meta-heuristics, knowledge-based systems, fuzzy logic, neural networks, Petri nets, hybrid techniques, and multi-agent systems.

Most manufacturing systems operate in dynamic environments where usually inevitable unpredictable real-time events may cause a change in the scheduled plans, and a previously feasible schedule may turn infeasible when it is released to the shop floor. Examples of such real-time events include machine failures, arrival of urgent jobs, due date changes, etc.

In dynamic, stochastic manufacturing systems (see for example [5–9], unpredictable events like breakdowns, expedite orders, quality problems, and material shortages occur during processing. Although careful scheduling coordinates activities to maintain productivity, these disruptions can render the desired schedule infeasible. Re-scheduling attempts to diminish the loss by creating a new schedule that more accurately reflects the current state of the production system.

An approach to solve the kind of problems referred above is proposed in this chapter. This contribution relies on the use of a dynamic multi-criteria decision model (spatial-temporal) for better supporting globally distributed manufacturing scheduling.

For accomplishing a clear and structured idea about the main subjects underlying this work, this chapter is organised as follows. Section 1 defines the problem of dynamic scheduling and the categories of real-time events. Next, Sect. 2 presents some important related work regarding approaches, techniques and systems used to solve the problem of dynamic (re)scheduling. Section 3 briefly describes our proposed approach for distributed dynamic manufacturing scheduling. Section 4 introduces an illustrative example of our proposal. Finally, Sect. 5 presents a conclusion and planed future work.

2 Related Work on Manufacturing Scheduling

In most real-world environments, scheduling is an ongoing reactive process where the presence of a variety of unexpected disruptions is usually inevitable, as well as continually forces reconsideration and revision of pre-established schedules.

Many approaches to solve the problem of static scheduling are often impractical in real-world environments, and the near-optimal schedules with respect to the estimated data may become obsolete when they are released to the shop floor. Vieira et al. [8] outline the limitations of the static approaches to scheduling in the presence of real-time information and presents a number of issues that have come up in recent years on dynamic scheduling.

2.1 The Dynamic Scheduling Problem

Literature on dynamic scheduling usually addresses a significant number of real-time events and their effects in various manufacturing systems, such as: single machine systems, parallel machine systems, flow shops, job shops, and flexible manufacturing systems.

Real-time events have been classified into two categories [1, 3, 5, 6]:

- Resource-related: machine breakdown, operator illness, unavailability or tool failures, loading limits, delay in the arrival or shortage of materials, defective material (material with wrong specification), etc.
- Job-related: rush jobs, job cancellation, due date changes, early or late arrival of jobs, change in job priority, changes in job processing time, etc.

Dynamic scheduling has been defined under three categories [6–10]: completely reactive scheduling, predictive–reactive scheduling, and robust pro-active scheduling.

2.2 Completely Reactive Scheduling

In completely reactive scheduling no firm schedule is generated in advance and decisions are made locally in real-time.

Priority dispatching rules are frequently used. A dispatching rule is used to select the next job with highest priority to be processed, from a set of jobs awaiting service, at a machine that becomes free. The priority of a job is determined based on job and machine attributes. Dispatching rules are quick, usually intuitive, and easy to implement. In contrast, global scheduling has the potential to significantly improve shop performance, when compared to myopic dispatching rules, where it is hard to predict system performance, as decisions are made locally in real-time.

2.3 Predictive-Reactive Scheduling

Predictive-reactive scheduling is the most common dynamic scheduling approach used in manufacturing systems [11]. Predictive-reactive scheduling/rescheduling is a process in which schedules are revised in response to real-time events.

Most predictive-reactive scheduling strategies are based on simple schedule adjustments that consider only shop efficiency. The new schedule may deviate significantly from the original schedule, which can seriously affect other planning activities based on the original schedule and may lead to poor schedule performance [11]. It is therefore desirable to generate predictive-reactive schedules that are robust. Robust predictive-reactive scheduling focuses on building predictive-reactive schedules to minimize the effects of disruptions on the performance measure value of the realised schedule [12–14].

2.4 Robust Pro-active Scheduling

Robust pro-active scheduling approaches focus on building predictive schedules, which satisfy performance requirements predictably in a dynamic environment [6, 7]. The main difficulty of this approach is the determination of the predictability measures.

2.5 Rescheduling in the Presence of Real-Time Events

Rescheduling in the presence of real-time events needs to address two issues: how and when to react to real-time events. The first issue concerns the definition of rescheduling strategies to react to real-time events, and the second issue addresses the problem of when to reschedule.

Regarding the first issue, what strategies to use to reschedule, the literature focus on two main strategies [5, 6, 15]: schedule repair, and complete rescheduling. Schedule repair refers to some local adjustment of the current schedule and may be preferable because of the potential savings in CPU times and the stability of the system is preserved. Complete rescheduling regenerates a new schedule from scratch. Complete rescheduling might, in principle, be better in maintaining optimal solutions, but these solutions are rarely achievable in practice and require prohibitive computation time.

Regarding the second issue, when to reschedule, three policies have been proposed in the literature [6, 15]: periodic, event driven, and hybrid. The periodic and hybrid policies have received special attention under the name rolling time horizon [8, 9, 16–18].

In the periodic policy, schedules are generated at regular intervals, which gather all available information from the shop floor. The dynamic scheduling problem is decomposed into a series of static problems that can be solved by using classical scheduling algorithms. The periodic policy yields more schedule stability and less schedule nervousness. Unfortunately, following an established schedule in the face of significant changes in the shop floor status, may compromise performance since unwanted products or intermediates may be produced. Determining the rescheduling period is also a difficult task.

In event driven policy, rescheduling is triggered in response to an unexpected event that alters the current system status. Most approaches for dynamic scheduling use this policy.

A hybrid policy reschedules the system periodically and also when an exception occurs. Events usually considered are machine breakdowns, arrival of urgent jobs, cancellation of jobs, or job priority changes.

3 A Distributed Dynamic Manufacturing Scheduling Model

As exposed before, important work has already been put forward by different authors, for dealing with distributed dynamic manufacturing scheduling problems, but there is still much room for improving these complex problems. Therefore the main purpose of this work consists on providing a novel approach for better solving this kind of problems. Our contribution is a scheduling decision support system, based on a fuzzy dynamic multi-criteria decision model (DMCDM) [19–21], as well as on web services, XML modeling and related technologies [22–24]. Here we will focus on the dynamic approach, but we have in mind that the XML and other documents will be stored on a distributed data repository, spread through a set of dynamically updated collaborating businesses within a globally distributed manufacturing environment [24–26].

It is rather important to consider well-organized manufacturing systems, which include appropriate requirements for enabling collaborative management, in terms of intra and inter factories manufacturing scheduling. Figure 1 illustrates this kind of scenario, which integrates four manufacturing cells, each one including two similar machine-centers, for jobs processing. The manufacturing cells perform their work integrated in a network, in the scope of a distributed manufacturing system (DMS), where a brokering service plays an important role in several distinct aspects, namely assignment of orders received from clients to the distributed working cells.

In the DMS context, integration or collaboration of humans and technology - in a versatile environment - play fundamental roles to obtain suitable and efficient decisions about manufacturing scheduling solutions, for the global integrated manufacturing system. Another important aspect is to ensure user-friendly interfaces to facilitate sharing information among the manufacturing cells, occurring inside the DMS, to support enlightened decisions in a widened and integrated manufacturing management context.

In this context, reconfigurability dynamics and business alignment are important requirements for considering using a dynamic multi-criteria decision support model [19–21] for dealing with responsiveness of market demands that require increasingly shorter product life cycles and shorter time to market. Furthermore, these problems are constantly forcing product life cycles to suffer frequent redesigns, which imply requirements for increased dynamics to ensure accurate and timely responses to problems arising dynamically, in a real-time basis. These dynamic reconfiguration requisites are the motivation for the multi-criteria decision-making model, within the DMS dynamic reconfiguration support, which we are going to describe next.

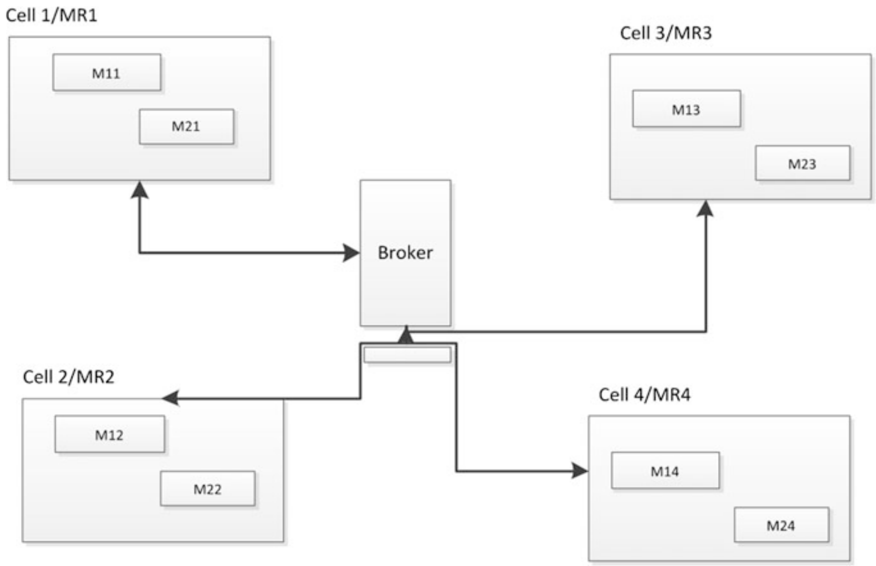


Fig. 1 Inter and intra manufacturing scheduling scenario within a DMS

In general, the aim of multiple-criteria decision-making is to find the best compromise solution from a set of feasible alternatives assessed with respect to a predefined set of criteria [21, 27, 28]. This type of decision problems is widespread in real-life situations, and many approaches have been proposed in the literature to deal with the static decision process, from utility methods to scoring and ranking ones [21, 27, 28]. However, when facing dynamic decision-making problems, where feedback from step to step is essential (e.g. any periodical evaluation of collaborating businesses), very few contributions can be found in the literature (see, for example [20, 25, 29]). Usually, dynamic multi-criteria decision-making (DMCM) belongs to spatial-temporal contexts, in that exploration of the problem might result in new alternatives being considered, others being discarded, and a set of criteria to be similarly altered [19].

In distributed manufacturing contexts, the problem of multiple resources manufacturing scheduling can be easily understood as a temporal multi-criteria decision-making problem: periodically and dynamically, businesses express their preferences with respect to manufacturing resources, for instance, manufacturing cells, which can then be ranked and selected to improve the complete manufacturing scheduling.

Therefore, each candidate manufacturing resource, at each decision moment (t) is assessed by collaborating businesses according to some set of decision criteria (such as lack of reliability, or speed, time and costs) that may also change over time. These assessments are then distilled down (aggregated) to a single (static) rating using some aggregation operator. After, in dynamic contexts, this rating

Fig. 2 Dynamic decision making model

value has to be joined (aggregated) with historical information to produce dynamic ratings that take in consideration past behaviors (historic rating). Finally a linear programming model can be used to cover processes that include collaboration of many-to-many business. This process goes on (feedback process) and in any new iteration both the alternatives and criteria may change, new ones can be added and/or others can be deleted, according to the real-time events. Figure 2 (based on [20, 21]) displays the general dynamic process.

With this dynamic model in mind we can now generalize the scheduling scenario, depicted in Fig. 1. Let us consider a time instant t and m collaborating businesses (CB_j), which are planning their orders on a set of n manufacturing resources (MR_i). Each collaborating business CB_j is assumed to be evaluated by the behavior regarding breakdowns on orders satisfaction, i.e. the value represents a penalty, where the lower the better ("good behavior"). At each time instant t an evaluation, regarding penalties for breakdowns, is produced for each CB_j (e.g. monthly periodicity) and this value is combined with the corresponding historical information from the previous period, t-1, to obtain *Pi* (details about this aggregation process are presented in [19, 21, 29]). Furthermore, each collaborating business has a certain production demand D_j and each manufacturing resource has a maximum production capacity, C_i. Moreover, the variables of the network of collaborating businesses and manufacturing resources includes the quantities x_{ij} that collaborating business, CB_j order from manufacturing resources, MR_i.

Another important aspect of the proposed model is related to the requisite of a satisfaction level imposed for each candidate manufacturing resource; in our case it consists on pre-defined values for penalties to express that if the penalty is higher than a certain threshold, M_i, the MR_i is eliminated as an alternative manufacturing resource candidate on that iteration.

Hence, the total lack or un-satisfaction levels of all collaborating businesses (L), in terms of the breakdowns related to deficiency or nonexistence of orders satisfactions is minimized.

4 Illustrative Example

Let us consider a manufacturing context that includes four different manufacturing resources, for short (MR_1–MR_4), distributed worldwide, each one being able to process a set of four distinct jobs (J_1–J_4), with a same processing time of 1 time unit. Moreover, we will consider four distinct scenarios (R_1–R_4) where each MR_i will produce either 1 or 2 or 3 or 4 jobs (i.e. all possible combinations can be distributed by the businesses MR_i).

We will also consider that each job requires a setup cost/time for being processed on a manufacturing resource, which varies according to the number of jobs to be processed on that manufacturing resource, as follows: 0.8 time units if the whole set of four jobs is processed on one of the four manufacturing resources available; 1.5 time units if three of them are processed on a given manufacturing resource; 2 time units if 2 of the jobs are processed on the same manufacturing resource; and 3 time units if only one job is processed on a manufacturing resource. Moreover, the jobs also have to be delivered to the final clients and this final transportation cost/time will be 0 time units if the job is processed next to the corresponding client location (i), assuming that MR_1 is located in the same location as Client 1 (C_1) and so forth (e.g. MR_2 is located next to Client 2, C_2). Otherwise, if a job has to be delivered to some other location, the corresponding transportation time of job j to location i follows the rule of time = |i-j| time units, for example, the time for transportation of job 2 to manufacturing resource 3 is equal to 1 (|2–3|) time units, and so one.

Under a collaborative context let us now consider the set of all alternative scenarios for jobs allocation for being processed and delivered to the corresponding four clients—placed next to each of the four manufacturing resource available—as follows:

- Scenarios considering only 1 job per manufacturing resource include 24 situations (R_1^1 to R_1^{24});
- Scenarios considering 2 jobs per manufacturing resource include 36 situations (R_2^1 to R_2^{36}). These occur when two of the 4 jobs will be processed on one of the four manufacturing resources available and the remaining two jobs on another available manufacturing resource;
- Scenarios considering 3 jobs include 48 situations (R_3^1 to R_3^{48}) that arise from the context of processing three of the set of the four jobs on one of the four manufacturing resources available and the remaining job being processed on another manufacturing resource of the four available;
- Scenarios considering 4 jobs include 4 different situations (R_4^1 to R_4^4) correspond to processing each set of four jobs on one of the four manufacturing resources available.

After establishing the possible scenarios we can calculate the best interscheduling, expressed on time units, for the example, which totals nine alternative solutions.

From the obtained results, we select the best solutions (minimums) for each manufacturing resource:

(1) Scenario $R_4^2 = \{(MR_2, J_1, J_2, J_3, J_4)\} = 9$
(2) Scenario $R_4^3 = \{(MR_3, J_1, J_2, J_3, J_4)\} = 9$
(3) Scenario $R_2^2 = \{(MR_1, J_1, J_2); (MR_3, J_3, J_4)\} = 10$

(4) Scenario $R_2^3 = \{(MR_1, J_{1,} J_2); (MR_4, J_{3,} J_4)\} = 10$
(5) Scenario $R_2^{19} = \{(MR_2, J_{1,} J_2); (MR_3, J_{3,} J_4)\} = 10$
(6) Scenario $R_2^{20} = \{(MR_2, J_{1,} J_2); (MR_4, J_{3,} J_4)\} = 10$
(7) Scenario $R_3^{15} = \{(MR_2, J_{1,} J_{2,} J_3); (MR_4, J_4)\} = 11$
(8) Scenario $R_3^{28} = \{(MR_1, J_1); (MR_3, J_{2,} J_{3,} J_4)\} = 11$
(9) Scenario $R_1^1 = \{(MR_1, J_1); (MR_2, J_2); (MR_3, J_3); (MR_4, J_4)\} = 16$

It is obvious that the best solutions are the ones from scenario R_4^2 and R_4^3 with either MR_2 or MR_3 (total time = 9) producing all the jobs. The worst scenario is R_1^1 (total time 16) where each job is divided per manufacturing resource (mainly due to the set-up times involved).

Now, if we are in a collaborative environment we could expect some negotiation to take place to ensure selection of the best option in terms of members of the "social network" established by the four manufacturing resources. Notice that this negotiation will correspond to a second iteration in our dynamic model. Since the 2 best scenarios for R_3 (total costs 11) R_1^1 present the worst results for the second round of negotiation we will only consider the 6 best candidate alternatives, independent of the manufacturing resources used. Furthermore, let us now consider that we have information about the historic timely deliveries (mean lateness) from each manufacturing resource, hence, this allow us to run a second round of negotiations/re-scheduling to decide which MR_i will win the contract.

Table 1 depicts the results for the candidate scenarios of using this criterion to decide between the six best candidates. Table 1 also depicts the total times to produce the combination of jobs in each resource. Observing this table the results are inconclusive since R_4^2 and R_2^{20} obtained the same global result of 12 time units. Using this simple re-scheduling does not bring much added value to decision makers and a third round of negotiation would have to take place to select between the best two scenarios.

Now, we will consider the influence of historical information (dynamical aspect of the MCDM model) to close the loop of information, based on previous behaviour/performance and un-satisfaction levels of the manufacturing resources being analysed. For the second phase of the dynamic MCDM, where we want to include an additional performance measure, regarding past experience related with lateness of orders, more specifically, the mean lateness of the manufacturing resources, regarding the execution of a similar set of jobs. Furthermore we will also discard the worst 3 scenarios from the first phase of our DMCDM, and only keep the 6 best candidate solutions for the second iteration.

Now, we will compare these results with the ones obtained in a second iteration with our dynamic MCDM model. In this second iteration the same new criteria is used, but the main difference is that the approach takes in consideration past information from the previous iteration. The additional criterion (mean lateness) is included in the second phase of the DMCDM by aggregating it with the historic results (obtained through the application of the first phase). Here we selected the simple weighted sum, as aggregation operator, with a relative importance of 40 %

Table. 1 New performance measure with historical information about mean lateness of resources for each scenario

Scenario	R_4^2	R_4^3	R_2^2	R_2^3	R_2^{19}	R_2^{20}
New criteria: mean lateness	3	6	3	4	3	2
Total times	12	15	13	14	13	12

for the historic information and 60 % for the current new criteria. In the general dynamic MCDM any other suitable aggregation operator could have been used both for combining criteria as well as we for combining historic information with current one [19]. The updated final results (second iteration) obtained for this example are the following:

(1) Scenario R_4^2 = 9*0.4 + 3*0.6 = 5.4
(2) Scenario R_4^3 = 9*0.4 + 6*0.6 = 7.2
(3) Scenario R_2^2 = 10*0.4 + 3*0.6 = 5.8
(4) Scenario R_2^3 = 10*0.4 + 4*0.6 = 6.4
(5) Scenario R_2^{19} = 10*0.4 + 3*0.6 = 5.8
(6) Scenario R_2^{20}= 10*0.4 + 2*0.6 = 5.2

Observing the results, we see that the best solution obtained is now R_2^{20} and this scenario consists on producing jobs 1 and 2 on manufacturing resource 2 and jobs 3 and 4 on manufacturing resource 4, with a global time unit of 12 and an evaluation value of 5.2.

Notice that all solutions vary not only according the additional performance measure of mean lateness, but also consider the total time (historic information), regarding jobs processing, setup times and transportation times. Furthermore, this model spatial-temporal characteristic allows (along re-scheduling steps) including new criteria, changing input values and so forth, without forgetting past behaviours.

This model allows that sometimes we may make a decision based on a trade-off situation in terms of lowest total time, if we are considering time based performance measures for our decision making or other kind of performance measures. For instance, cost or resources utilization or combined situations, according to each preferred situation occurring on each decision scenario, and even considering adding other criteria related to promoting collaboration between businesses or any other constraints.

Moreover, when execution times approximate—more or less equally—preferable situations tend to occur, as we enter on a collaborative environment. Therefore, it may be wise to highlight that, in a globally distributed market of resources it is relevant to pay attention not only to local production scheduling approaches but also to global ones, considering the diverse situations related to inter manufacturing environments planning and scheduling alternative scenarios, besides the intra manufacturing scenario. Concluding, our dynamic re-scheduling approach allows historical information to play an important role in supporting decision makers, by providing enriched information about previous behaviour of alternative solutions.

5 Conclusion

In this chapter we showed the importance, as a competitive strategy, to explore and use a dynamic multi-criteria decision model to better support collaborative manufacturing scheduling systems, particularly in today's Internet and Intranets, for solving distributed dynamic manufacturing scheduling problems. A simple illustrative example was presented, which highlighted how the decision making process could be supported within globally distributed manufacturing scenarios, for instance, based on a set of distributed manufacturing cells, each one integrating a set of manufacturing resources available for producing a set of jobs. Although the main goal was to highlight decision support in manufacturing scheduling resolution, the collaborative decision support system could also play other important roles. For instance, for enabling an easy and user friendly interface for problem data introduction and processing as well as easy access to solving methods and its implementation(s), for further intra manufacturing system (e.g. cell, or manufacturing resource scheduling, occurring in the context of the whole distributed manufacturing system scenario).

As future work, we plan to develop a platform for solving distributed manufacturing scheduling problems occurring in real-time distributed manufacturing environments, for instance, either for intra or inter cellular manufacturing scheduling scenarios. Concepts, related to problems and alternative solving methods will be modelled through XML and put available through a globally distributed network, where a set of collaborating business are dynamically integrated.

Acknowledgment The authors wish to acknowledge the support of: (1) The Foundation for Science and Technology—FCT, under the scope of the financed Project on "Ubiquitous oriented embedded systems for globally distributed factories of manufacturing enterprises"—PTDC/EME-GIN/102143/2008, and (2) EUREKA, under the Project E!4177-Pro-Factory UES; (3) NSF Grant #2003168 and CNSF Grant #9972988.

References

1. Suresh, V., Chaudhuri, D.: Dynamic scheduling a survey of research. Int. J. Prod. Econ. **32**(1), 53–63 (1993)
2. Shukla, C.S., Chen, F.F.: The state of the art in intelligent real-time FMS control: a comprehensive survey. J. Intell. Manuf. **7**, 441–455 (1996)
3. Stoop, P.P.M., Weirs, V.C.S.: The complexity of scheduling in practice. Int. J. Oper. Prod. Manag. **16**(10), 37–53 (1996)

4. Brandimarte, P., Villa, A.: Modelling manufacturing systems: from aggregate planning to real-time control. Springer, Berlin (1999)
5. Cowling, P.I., Johansson, M.: Using real-time information for effective dynamic scheduling. Eur. J. Oper. Res. **139**(2), 230–244 (2002)
6. Vieira, G.E., Hermann, J.W., Lin, E.: Rescheduling manufacturing systems: a framework of strategies, policies and methods. J. Sched. **6**(1), 36–92 (2003)
7. Mehta, S.V., Uzsoy, R.: Predictable scheduling of a single machine subject to breakdowns. Int. J. Comput. Integr. Manuf. **12**(1), 15–38 (1999)
8. Vieira, G.E., Herrmann, J.W., Lin, E.: Analytical models to predict the performance of a single machine system under periodic and event-driven rescheduling strategies. Int. J. Prod. Res. **38**(8), 1899–1915 (2000)
9. Aytug, H., Lawley, M.A., McKay, K., Mohan, S., Uzsoy, R.: Executing production schedules in the face of uncertainties: a review and some future directions. Eur. J. Oper. Res. **161**(1), 86–110 (2005)
10. Herroelen, W., Leus, R.: Project scheduling under uncertainty: survey and research potentials. Eur. J. Oper. Res. **165**(2), 289–306 (2005)
11. Ouelhadj, D., Petrovic, S.: A survey of dynamic scheduling in manufacturing systems. J. Sched. **12**, 417–431, (2009) (Springer)
12. Wu, S.D., Storer, R.H., Chang, P.C.: A rescheduling procedure for manufacturing systems under random disruptions. In: Proceedings Joint USA/German Conference on New Directions for Operations Research in Manufacturing, pp. 292–306 (1991)
13. Wu, S.D., Storer, R.H., Chang, P.C.: One machine rescheduling heuristics with efficiency and stability as criteria. Comput. Oper. Res. **20**(1), 1–14 (1993)
14. Leon, V.J., Wu, S.D., Storer, R.H.: Robustness measures and robust scheduling for job shops. IIE Trans. **26**(5), 32–41 (1994)
15. Sabuncuoglu, I., Bayiz, M.: Analysis of reactive scheduling problems in a job shop environment. Eur. J. Oper. Res. **126**(3), 567–586 (2000)
16. Church, L.K., Uzsoy, R.: Analysis of periodic and event-driven rescheduling policies in dynamic shops. Int. J. Comput. Integr. Manuf. **5**(3), 153–163 (1992)
17. Ovacik, I.M., Uzsoy, R.: Rolling horizon algorithms for a single-machine dynamic scheduling problem with sequence-dependent set-up times. Int. J. Prod. Res. **32**(6), 1243–1263 (1994)
18. Sabuncuoglu, I., Karabuk, S.: Rescheduling frequency in an FMS with uncertain processing times and unreliable machines. J. Manuf. Syst. **18**(4), 268–283 (1999)
19. Campanella, G., Ribeiro, R.A.: A framework for dynamic multiple criteria decision making. Decis. Support Syst. **52**(1), 52–60 (2011). doi: http://dx.doi.org/10.1016/j.dss.2011.05.003
20. Campanella, G., Pereira, A., Ribeiro, R. A., Varela, L. R.: Collaborative dynamic decision making: a case study from B2B supplier selection. In: J. Hernández, E., Zarate, P., Dargam, F., Delibašic, B., Liu, S., Ribeiro R. (eds.) Decision Support Systems—Collaborative Models and Approaches in Real Environments: EWG-DSS, LNBIP 121, 2011. Springer, berlin (2012). doi: 10.1007/978-3-642-32191-7
21. Campanella, G., Varela, L. R., Ribeiro, R. A.: A model for B2B selection. In: De Baets, B., Fodor, J., Serodio, C., Couto, P., Melo-Pinto P. (eds.) Advances in Intelligent and Soft Computing, vol. 107, pp. 221–228. Springer, Berlin. doi: 10.1007/978-3-642-24001-0_21
22. Ceponkus, A., Hoodbhoy, F.: Applied XML. Wiley Computer Publishing, USA (1999)
23. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, New York (1989)
24. Arrais-Castro, A., Varela, M. L. R., Putnik, G. D., Ribeiro, R. A.: Collaborative network platform for multi-site production. In: Hernández, J. E., Zarate, P., Dargam, F., Delibašic, B., Liu, S., Ribeiro, R. (eds.) Decision Support Systems—Collaborative Models and Approaches in Real Environments: EWG-DSS 2011, LNBIP 121. Springer, Berlin. doi: 10.1007/978-3-642-32191-7
25. Varela, M. L. R., Putnik, G. D, Ribeiro, R. A.: A web-based platform for collaborative manufacturing scheduling in a virtual enterprise. Information and Communication Technologies for the Advanced Enterprise an International Journal (ict), vol. 2

26. Magalhães, R., Varela, M.L.R., Carmo-Silva, S.: Web-based decision support system for industrial operations management. Rom Rev Precis Mech. Opt. Mechatron. **37**, 159–165 (2010)
27. Chen, S.-J., Hwang, C. L., Hwang, F. P.: Fuzzy multiple attribute decision making: methods and applications, vol. 375. Lecture Notes in Economics and Mathematical Systems. Springer, Berlin (1992)
28. Triantaphyllou, E.: Multi-criteria decision making methods: a comparative study, vol. 44. Applied Optimization. Springer, Berlin (2000)
29. Chuu, S.-J.: Group decision-making model using fuzzy multiple attributes analysis for the evaluation of advanced manufacturing technology. Fuzzy Sets Syst. **160**, 586–602, Elsevier (2009)

Part II
Fuzziness and Probability

Probabilistic Reasoning in a Fuzzy Context

Giulianella Coletti and Barbara Vantaggi

Abstract By adopting the interpretation of the membership of a fuzzy set in terms of coherent conditional probability, we study probabilistic reasoning under coherence when the information on the statements is vague and the probabilistic information is imprecise or incomplete.

1 Introduction

Models and tools for handling jointly probability and vagueness are need to be performed in order to deal with problems involving both these aspects. For this aim we refer to the interpretation, given in [6] (see also [3, 7]), of the fuzzy membership as a coherent conditional probability. This allows to have a general framework in which it is possible to handle uncertainty due to different causes and information coming from different sources, maintaining consistency with the model of reference.

In the literature there are other papers (see e.g. [17, 25, 30]) proposing to reread fuzzy concepts as conditional probabilities, but only coherence guaranties an effective tool for controlling global consistency and ruling the inferential procedures.

We recall that a general inferential problem can be seen simply as an extension of an assessment to other events, maintaining consistency with the framework of reference. In the probabilistic framework this problem reduces in the simplest case to Bayes rule (where likelihood and prior distribution are completely assessed on the

G. Coletti (✉)
Department of Mathematics and Computer Science, University of Perugia, via Vanvitelli, 1, 06123 Perugia, Italy
e-mail: coletti@dipmat.unipg.it

B. Vantaggi
Department of S.B.A.I., University La Sapienza of Rome, via Scarpa 16, 00161 Roma, Italy
e-mail: barbara.vantaggi@sbai.uniroma1.it

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_7,

same space). However, the hypotheses of Bayes Theorem usually are not satisfied and so we need to handle generalized Bayesian inferential procedures whose result is in general not unique, but consists on an interval of coherent values (see e.g. [5]).

We analyze the above problem in some general cases, when the "prior probability distribution" can be not completely available or it is related to sets of events different from those where the membership function (given in terms of likelihood) is defined. In these cases first of all it is necessary to check whether probabilistic and fuzzy information are globally coherent: in fact they can be separately coherent but coherence can fail when considered together. These problems, related to a "partial prior" or to the fact that the prior distribution is given on a partition less fine than that where membership is defined, require to manage also lower and upper probabilities and to make inference also in this setting.

Furthermore we show how to manage some probabilistic IF-THEN rules, in particular we show how to compute the probability associated to these rules among fuzzy sets. For this aim we recall a notion of inclusion [10, 29] that allows us to handle the rules IF-THEN among "fuzzy events" and show how to propagate coherent interval related to the IF-THEN rules.

2 Coherent Lower Conditional Probability

The approach to probability adopted here is based on *coherence* (a notion that goes back to de Finetti [14]). The starting point is conditional probability as a function of two variables ruled by a set of axioms.

Definition 1 Let $\mathcal{E} = \mathcal{B} \times \mathcal{H}$, with $\mathcal{B}$ a Boolean algebra and $\mathcal{H}$ an additive set (i.e., closed with respect to finite logical sums), not containing $\emptyset$. A function $P : \mathcal{E} \to [0, 1]$ is a conditional probability if the following conditions hold:

(C1) $P(H|H) = 1$, for every $H \in \mathcal{H}$,
(C2) for any $H \in \mathcal{H}$ the function $P(\cdot|H)$ is a (finitely additive) probability on $\mathcal{B}$,
(C3) for every $A \in \mathcal{B}$, $E \wedge H \in \mathcal{H}$,

$$P(E \wedge A|H) = P(E|H)P(A|E \wedge H).$$

We recall also an easy consequence of the above axioms, i.e. the *disintegration formula* for the probability of an event $E|H$ with respect to a partition of an event H

$$P(E|H) = \sum_{k=1}^{N} P(H_k|H)P(E|H_k) \quad (1)$$

We notice that it is not possible to construct a conditional probability by taking as starting point *just one* probability, since a conditional probability is essentially a *class* of probabilities, linked by *(C3)* (see [13, 16, 26]).

Actually, when $P_0(\cdot) = P(\cdot|\Omega)$ is strictly positive on $\mathcal{H}$, any conditional probability can be derived as a ratio by means of this unique "unconditional" probability P_0; while otherwise (when P_0 is not strictly positive on $\mathcal{H}$) to get a similar representation we need to resort to a class of unconditional probabilities (see [2, 4, 5]), each of them defined where the previous one is equal to zero.

The above definition of (conditional) probability is strictly based on the Boolean structure of the domains. Actually, in reasoning under uncertainty the knowledge base usually consists on a bunch of (conditional) events and one is interested into extending it to new events.

The concept of coherence, introduced by de Finetti [13] in probability theory, has the fundamental role to manage partial assessments of an uncertainty measure and its enlargements. In other words coherence is a tool to check whether a function defined on an arbitrary set of (conditional) events is consistent with a specific uncertainty measure and to make inference in the general sense, that is to extend this function to new conditional events, by maintaining coherence.

Definition 2 Given an arbitrary set $\mathcal{F} = \{E_i|H_i\}$ of conditional events, a real function P on $\mathcal{F}$ is a coherent conditional probability assessment if there exists a conditional probability $P'(\cdot|\cdot)$ extending P on $\mathcal{E} = \mathcal{B} \times \mathcal{H}$, where $\mathcal{B}$ is the Boolean algebra spanned by the events $\{E_i, H_i\}$ and $\mathcal{H}$ the additive set spanned by the events $\{H_i\}$.

Concerning coherence, we recall the following fundamental result (essentially due to de Finetti [13]), which consider the equivalent concept of coherence, expressed in terms of coherent betting scheme:

Theorem 1 Let $\mathcal{C} = \{E_i|H_i\}$ be any family of conditional events, and take an *arbitrary* family $\mathcal{C}^* \supseteq \mathcal{C}$. Let P be an assessment on $\mathcal{C}$; then there exists a (possibly not unique) coherent extension of P to $\mathcal{C}^*$ *if and only if* P is a coherent lower conditional probability on $\mathcal{C}$.

In particular, if $\mathcal{C}^* = \mathcal{C} \cup \{E|H\}$, then the possible coherent values $P(E|H)$ are all the values of a suitable closed interval $[p', p''] \subseteq [0, 1]$, with $p' \leq p''$ (see e.g. [8, 28, 34, 36]). If moreover $\mathcal{C}$ is finite, then it is possible to compute the bounds of the interval $[p', p'']$ by solving a linear programming problem (see e.g. [5]).

This may give rise to non-uniqueness of the extended probability: in order to continue to work with coherent conditional probabilities, it would be necessary to make successive choices of the possible extensions inside the coherent intervals. This leads not only to a loss of information, but also to possible different conclusions, depending both on the order in which the "new" events are considered and on the relevant chosen values.

On the other hand, avoiding any intervention during the inferential process, we have a class of coherent probabilities consistent with the given assessment. Therefore, we need to handle coherent lower and upper probabilities [5], or in other words with the natural extensions [34].

Definition 3 A pair of functions $\underline{P}$ and $\overline{P}$ from $\mathcal{F}$ to $[0,1]$ are conjugate coherent lower and upper conditional probabilities if there exists a family of coherent conditional probabilities $P_i : \mathcal{F} \to [0,1]$ whose lower and upper envelope coincide with $\underline{P}$ and $\overline{P}$, respectively.

Obviously, if $E|H, E^c|H \in \mathcal{F}$ for any pair of conjugate coherent lower and upper probabilities it follows

$$\underline{P}(E|H) = 1 - \overline{P}(E^c|H).$$

Now let $\mathcal{C} = \{E_1|H_1, \ldots, E_n|H_n\}$ be an arbitrary finite family of conditional events, denote by $\mathcal{C}' = \mathcal{C} \bigcup \{E_1^c|H_1, \ldots, E_n^c|H_n\}$ $(i = 1, \ldots, n)$, a pair of real functions $(\underline{P}, \overline{P})$ on $\mathcal{C}$ are coherent conjugate lower and upper probabilities if and only if the function $\underline{P'}$ on $\mathcal{C}'$, defined as $\underline{P'}(E_i|H_i) = \underline{P}(E_i|H_i)$ and $\underline{P'}(E_i^c|H_i) = 1 - \overline{P'}(E_i|H_i)$, is a coherent lower probability (see e.g. [34]).

In the literature different characterizations of coherent lower conditional probability assessments are present, see for instance [34, 36]. All of them use the "compactness property" of coherence, assuring that an assessment is coherent with a probability if and only if it is coherent in any finite subset.

We recall here a characterization (for simplicity) given for a finite family of conditional events (see [5, 8]), which is the most useful for the check of coherence.

This condition is based on the notion of atoms related to a finite set of conditional events $E_1|H_1, \ldots, E_n|H_n$ which are all the possible events obtained by the following conjunctions $E_1^* \wedge H_1^* \wedge \cdots \wedge E_n^* \wedge H_n^*$ where E_i^* (similarly for H_i^*) stands for either E_i or E_i^c.

Given a partition $\mathcal{H} = \{H_i\}_{i=1,\ldots,m}$, the events E_i $(i = 1, \ldots, n)$, are said *logically independent with respect to* $\mathcal{H}$ if, for any $H_j \in \mathcal{H}$,

$$\wedge_{i=1}^n E_i^* \wedge H_j = \emptyset, \text{ implies } E_i^* \wedge H_j = \emptyset$$

for some $i = 1, \ldots, n$.

In particular, the events E_i, with $i = 1, \ldots, n$, are *logically independent* (with respect to Ω) if the cardinality of the set of relevant atoms is 2^n.

The following Theorem gives an operative tool to check coherence by solving a sequence of linear systems where the unknowns are probabilities of atoms.

Theorem 2 Let $\mathcal{C} = \{E_1|H_1, \ldots, E_n|H_n\}$ be an arbitrary finite family of conditional events and denote by $\mathcal{A}_o$ the set of the relevant atoms A_r. For a pair of real functions $(\underline{P}, \overline{P})$ on $\mathcal{C}$ the following statements are equivalent:

(a) $\underline{P}, \overline{P}$ are conjugate coherent conditional lower and upper probabilities on $\mathcal{C}$ (i.e. $\underline{P}$ is a coherent lower probability on $\mathcal{C}'$);

(b) for every event $F_i|H_i \in \mathcal{C}'$ there exists a sequence of compatible systems S_β $(\beta = 0, \ldots, k \leq 2n)$, with non negative unknowns $x_r^\beta = P_\alpha(A_r)$, $A_r \in \mathcal{A}_\beta$ $(\mathcal{A}_0 = \mathcal{A}_o^{\mathcal{F}}, \mathcal{A}_\beta = \{E \in \mathcal{A}_{\beta-1} : \sum_{A_r \subseteq E} x^{\beta-1} = 0\})$,

$$\begin{cases} \sum\limits_{A_r \subseteq F_i \wedge H_i} x_r^\beta = \underline{P}'(F_i|H_i) \sum_{A_r \subseteq H_i} x_r^\beta \\ \sum\limits_{A_r \subseteq F_j \wedge H_j} x_r^\beta \geq \underline{P}'(F_j|H_j) \sum_{A_r \subseteq H_j} x_r^\beta \quad (j \neq i), [\text{for } E_k|H_k \in \mathcal{C} \text{ such that} \sum_{A_r \subseteq H_k} \mathrm{x}_r^{\beta-1} = 0] \\ \sum\limits_{A_r \subseteq H_o^\beta} x^{\beta_r} = 1 \end{cases}$$

where $H_o^o = H_o = H_1 \vee \cdots \vee H_n$, $x_r^{\beta-1}$ denotes a solution of $(S_{\beta-1})$ and, for $\beta \geq 1$, H_o^β is the logical sum of the H_{j_i}'s such that $\sum_{A_r \subseteq H_{j_i}} x_r^{\beta-1} = 0$. Finally put, for all H_{j_i}'s, $\sum_{A_r \subseteq H_{j_i}} x_r^{-1} = 0$

The proof (see for instance [5]) is based on the characterization theorem for coherent conditional probability and in the fact that, for a finite set of conditional event, if there exists a family of coherent conditional probabilities $\prod$ whose lower envelope (inf) coincides with $\underline{P}$ then there exists a finite subclass $\prod'$ of $\prod$ such that the minimum over it coincides again with $\underline{P}$.

We recall the following result [6], which can be deduced by Theorem 2 with $\underline{P} = \overline{P} = P$:

Corollary 1 Let $\mathcal{C} = \{E|H_i\}_{i \in I}$, where $card(I)$ is arbitrary and the events H_i's are a *partition* of Ω. Then any function $P : \mathcal{C} \rightarrow [0, 1]$ such that

$$P(E_i|H_i) = 0 \quad \text{if} \quad E_i \wedge H_i = \emptyset \tag{2}$$

$$P(E_i|H_i) = 1 \quad \text{if} \quad H_i \subseteq E_i \tag{3}$$

(and taking otherwise any value in the interval $[0, 1]$) is a coherent conditional probability.

Then, we call likelihood function any function $L : \mathcal{C} \rightarrow [0, 1]$ satisfying conditions (2) and (3).

Remark 1 It is easy to see that the lower envelope of a family of likelihood functions is still a likelihood function.

Corollary 2 Let $\mathcal{C} = \{E_j|H_i\}$ $(i = 1, \ldots, n; j = 1, \ldots, m)$ be a set of conditional events such that the events E_j are logically independent with respect to a *partition* $\mathcal{H} = \{H_i\}_{j=1,\ldots,m}$ of Ω and let $\underline{P}(H_i)$ and $\overline{P}(H_i)$ be a coherent lower and upper probability on $H_1, \ldots, H_m$ respectively.

Then for every pair of likelihood functions (L_1, L_2) on $\mathcal{C}$ such that, for any $E_j|H_i \in \mathcal{C}$

$$L_1(E_j|H_i) \leq L_2(E_j|H_i)$$

the two global assessments

$$\prod\nolimits_1 = \{L_1(E_j|H_i), \underline{P}(H_i)\}$$

and

$$\prod\nolimits_2 = \{L_2(E_j|H_i), \overline{P}(H_i)\}$$

are conjugate coherent conditional lower and upper probabilities on $\mathcal{C} \bigcup \mathcal{H}$.

Proof Coherence follows from Theorem 2. Since the events E_j's are logically independent with respect to $\mathcal{H}$. It follows that any atom of the form $\wedge_j E_j^* \wedge H_i$ is possible whenever there is no E_j with $E_j \wedge H_i = \emptyset$ or $H_i \subseteq E_j$. By considering the relevant system, the subsystem related to each H_i (without the equations involving $\underline{P}(H_i)$ and $\overline{P}(H_i)$ and the last one) admits solution.

Moreover, since the set of events H_i is a partition, the relevant system contains independent equations each related to a conditioning event H_i. So it admits a solution, which is given by a suitable vector obtained by concatenation of the solutions related to H_i. In fact, these equations of the system are satisfied, since $\underline{P}(H_i)$ and $\overline{P}(H_i)$ on $H_1, \ldots, H_m$ are coherent lower and upper probabilities. □

2.1 Extension

In the following we give a procedure for computing the coherent interval of the extension of a coherent lower conditional probability to a new conditional event $E|H$, that is a simplification of that given in [5].

This procedure consists into a linear programming whose constraints are expressed by just one sequence of linear systems (instead of a set of sequences of linear systems). It exploits the possibility to give probability 0 to the conditioning event H so to have less constraints as possible in order to compute the maximum and minimum value for $E|H$.

Consider for that the systems in Theorem 2 condition (b): since $\underline{P}$ is a lower probability, the solution of this quoted system are probabilities and so are solutions also of the following modified system

$$\begin{cases} \sum\limits_{A_r \subseteq F_i \wedge H_i} x_r^\alpha \geq \underline{P}'(F_i|H_i) \sum\limits_{A_r \subseteq H_i} x_r^\alpha \\ \sum\limits_{A_r \subseteq F_j \wedge H_j} x_r^\alpha \geq \underline{P}'(F_j|H_j) \sum\limits_{A_r \subseteq H_j} x_r^\alpha \; (j \neq i), \quad [\text{for } E_k|H_k \in \mathcal{C} \text{ s.t.} \sum\limits_{A_r \subseteq H_k} \mathrm{x}_r^{\alpha-1} = 0\,] \\ \sum\limits_{A_r \subseteq H^o} x^{\alpha_r} = 1 \end{cases} \tag{4}$$

In order to exploit the zero probability values on H, we look for the minimum index β such that the following systems are not consistent:

$$\begin{cases} \sum\limits_{A_r \subseteq F_i \wedge H_i} x_r^\beta \geq \underline{P}'(F_i|H_i) \sum\limits_{A_r \subseteq H_i} x_r^\beta \\ \sum\limits_{A_r \subseteq F_j \wedge H_j} x_r^\beta \geq \underline{P}'(F_j|H_j) \sum\limits_{A_r \subseteq H_i} x_r^\beta (j \neq i), \; [\text{for } E_k|H_k \in \mathcal{C} \text{ s.t.} \sum\limits_{A_r \subseteq H_k} x_r^{\beta-1} = 0] \\ \sum\limits_{A_r \subseteq H^o} x^{\beta_r} = 1 \\ \sum\limits_{A_r \subseteq H} x^{\beta_r} = 0 \end{cases} \tag{5}$$

Now since the equation $\sum\limits_{A_r \subseteq H} x^{\beta_r} = 0$ is not compatible with the other equations that are consistent (since system (4) has solution for every α), we consider the following system

$$\begin{cases} \sum\limits_{A_r \subseteq F_i \wedge H_i} x_r^\beta \geq \underline{P}'(F_i|H_i) \sum\limits_{A_r \subseteq H_i} x_r^\beta \\ \sum\limits_{A_r \subseteq F_j \wedge H_j} x_r^\beta \geq \underline{P}'(F_j|H_j) \sum\limits_{A_r \subseteq H_j} x_r^\beta \,(j \neq i), \, [\text{for } E_k|H_k \in \mathcal{C} \text{ such that} \sum\limits_{A_r \subseteq H_k} \mathrm{x}_r^{\beta-1} = 0] \\ \sum\limits_{A_r \subseteq H} x^{\beta_r} = 1 \end{cases} \tag{6}$$

that has solution. Actually, all solutions of system (4) are obtained by normalizing the solution of system (6).

Then, the values $\underline{p}$ and $\overline{p}$ for $E|H$ can be computed by finding the minimum and maximum value, respectively, of

$$\sum_{A_r \subseteq E \wedge H} x^{\beta_r},$$

under the above system (6).

This shows that even if the checking of coherence of lower conditional probabilities is more difficult than g-coherence [1, 20], for the extension we could refer to the analogous system arising from g-coherence or equivalently for natural extension [28, 34].

3 Fuzzy Sets as Coherent Conditional Probabilities

We adopt the interpretation of fuzzy sets in terms of coherent conditional probabilities, introduced in [5–7]: the idea behind such interpretation is related to that given in the seminal work [22]. We briefly recall here the main concepts.

3.1 Main Definition

Let X be a (not necessarily numerical) variable, with range $\mathcal{C}_X$, and, for any $x \in \mathcal{C}_X$, let us indicate by x the event $\{X = x\}$.

Let φ be any *property* related to the variable X and let us refer to the state of information of a real (or fictitious) person that will be denoted by "You". A coherent conditional probability assessment $P(E_\varphi|x)$, $x \in \mathcal{C}_X$ measures the degree of belief of You in E_φ, when X assumes the different values of its range. Corollary 1 assures in fact that any assessment $\{P(E|x)\}_{x\in X}$ is a coherent conditional probability assessment.

Then $P(E_\varphi|\cdot)$ comes out to be a natural interpretation of the membership function $\mu_\varphi(\cdot)$, according to [6] (see also [5, 7]).

Definition 4 For any variable X with range $\mathcal{C}_X$ and a related property φ, the *fuzzy subset* E^*_φ of $\mathcal{C}_X$ is the pair

$$E^*_\varphi = \{E_\varphi\,,\,\mu_{E_\varphi}\},$$

with $\mu_{E_\varphi}(x) = P(E_\varphi|x)$ for every $x \in \mathcal{C}_X$.

By Remark 1 it follows that the lower envelope of a set of membership functions related to the same property φ and the same X is still a membership function.

3.2 Operations

By referring to [6] we recall the operations between fuzzy subsets: the binary operations of union and intersection and that of complementation can be obtained directly by using the rules of coherent conditional probability and the logical independence between E_φ and E_ψ.

For this aim let us denote by $\varphi \vee \psi$, $\varphi \wedge \psi$, respectively, the properties "φ or ψ", "φ and ψ".

Define

$$E_{\varphi\vee\psi} = E_\varphi \vee E_\psi \,,$$

$$E_{\varphi\wedge\psi} = E_\varphi \wedge E_\psi \,.$$

Let us consider two fuzzy subsets $E^*_\varphi,\ E^*_\psi$, corresponding to the same variable X, with the events $E_\varphi,\ E_\psi$ logically independent. As proved in [6], for any given x in the range of X, the assessment $P(E_\varphi \wedge E_\psi|x) = v$ is coherent if and only if takes values in the interval

$$\max\{P(E_\varphi|x) + P(E_\psi|x) - 1, 0\} \leq v$$
$$\leq \min\{P(E_\varphi|x), P(E_\psi|x)\}.$$

From probability rules, given a value to $P(E_\varphi \wedge E_\psi|x)$, we get also the value of $P(E_\varphi \vee E_\psi|x)$.

Then we put

$$E^*_\varphi \cup E^*_\psi = \{E_{\varphi\vee\psi}\,,\,\mu_{\varphi\vee\psi}\}\ ,$$

$$E^*_\varphi \cap E^*_\psi = \{E_{\varphi\wedge\psi}\,,\,\mu_{\varphi\wedge\psi}\}\ ,$$

with

$$\mu_{\varphi\vee\psi}(x) = P(E_\varphi \vee E_\psi|x)\ ,$$

$$\mu_{\varphi\wedge\psi}(x) = P(E_\varphi \wedge E_\psi|x)\,.$$

Moreover, denoting by $E^*_{\neg\varphi}$ the complementary fuzzy set of E^*_φ, the relation $E_{\neg\varphi} \neq (E_\varphi)^c$ holds, since the propositions "You *claim* $\neg\varphi$" and "You *do not claim* φ" are logically independent. In fact, we can claim both "X has the property φ" and "X has the property $\neg\varphi$", or only one of them or finally neither of them; similarly are logical independent E_φ and E_ψ, where ψ is the superlative of φ.

Then, while

$$E_\varphi \vee (E_\varphi)^c = C_X\,,$$

we have instead

$$E_\varphi \vee E_{\neg\varphi} \subset C_X\,,$$

and, if we consider the union of a fuzzy subset and its complement

$$E^*_\varphi \cup (E^*_\varphi)' = \{E_{\varphi \vee \neg\varphi}\,,\, \mu_{\varphi \vee \neg\varphi}\}$$

we obtain in general a *fuzzy subset* of (the universe) C_X.

For two fuzzy subsets E^*_φ, E^*_ψ, corresponding to the random quantities X_1 and X_2, respectively, the following choice for the membership of conjunction and disjunction is coherent [7]:

$$\mu_{\varphi\vee\psi}(x,x') = P(E_\varphi \vee E_\psi | A_x \wedge A_{x'}), \tag{7}$$

$$\mu_{\varphi\wedge\psi}(x,x') = P(E_\varphi \wedge E_\psi | A_x \wedge A_{x'})\,. \tag{8}$$

with the only constraints

$$\begin{aligned} \max\{\mu_\varphi(x) + \mu_\psi(x') - 1\,,\,0\} \\ \leq \mu_{\varphi\wedge\psi}(x,x') \\ \leq \min\{\mu_\varphi(x) + \mu_\psi(x')\}\,. \end{aligned} \tag{9}$$

and

$$\mu_{\varphi\vee\psi}(x,x') = \mu_\varphi(x) + \mu_\psi(x') - \mu_{\varphi\wedge\psi}(x,x')\,. \tag{10}$$

4 Extension on "Fuzzy Events"

For simplicity we refer to variables X with a finite codomain, for an infinite version see [12]. First of all we notice that, in this context, the concept of fuzzy event, as introduced by Zadeh, is an ordinary event of the kind

$$E_\varphi = \text{``You claim that } X \text{ is } \varphi\text{''}.$$

We recall that for any probability distribution on the events x the global assessment

$$\{\mu_\varphi(x), P(x)\}$$

is coherent (see Corollary 2) and so coherently extendible to E_φ. According to Eq. (1), it is easily to see that the only coherent value for the probability of E_φ is

$$P(E_\varphi) = \sum_{x \in \mathcal{C}_X} \mu_{\varphi_i}(x) P(x),$$

which coincides with Zadeh's definition of the probability of a "fuzzy event" [32].

Now let $\mathcal{F}(\mathcal{C}_X)$ be a finite family of fuzzy subsets $E^*_{\varphi_i} = (E_{\varphi_i}, \mu_{\varphi_i})$ related to the (possibly coincident) components X_i of a vector X (with the events $\{E_{\varphi_i}\}$ logically independent with respect to the partition generated by X).

For every joint probability distribution on the events $\{X = \mathbf{x}\} = (X_1 = x_1, \ldots, X_n = x_n)$ the global assessment $\{\mu_{\varphi_i}, P(\mathbf{x})\}$ is coherent (see Corollary 2).

Moreover it is easy to prove (see also [6]) that for every t-norm $\odot$ and its dual t-conorm $\oplus$ of the class of Frank [19], the probability assessment

$$\{P_\odot(E_{\varphi_i}), P_\odot(E_{\varphi_i} \wedge E_{\varphi_j})\}$$

is still coherent, with

$$P_\odot(E_{\varphi_i}) = \sum_{x_i \in \mathcal{C}_{X_i}} \mu_{\varphi_i}(x_i) P_i(x_i)$$

$$P_\odot(E_{\varphi_i} \wedge E_{\varphi_j}) = \sum_{\mathbf{x}_{ij} \in \mathcal{C}_{(X_i, X_j)}} (\mu_{\varphi_i}(x_i)) \odot (\mu_{\varphi_j}(x_j)) P_{ij}(\mathbf{x}_{ij})$$

(where P_i $[P_{ij}]$ is the marginal on X_i $[(X_i, X_j)]$).

So we can extend $P_\odot$ to events $E_{\varphi_i} \vee E_{\varphi_j}$: this extension is univocally determined by coherence as:

$$\begin{aligned} &P_\odot(E_{\varphi_i} \vee E\varphi_j) \\ &\quad = \sum_{\mathbf{x}_{ij} \in \mathcal{C}_{(X_i, X_j)}} (\mu_{\varphi_i}(x_i)) \oplus (\mu_{\varphi_j}(x_j)) P_{ij}(\mathbf{x}_{ij}) \\ &\quad = P_\odot(E_{\varphi_i}) + P_\odot(E_{\varphi_j}) - P_\odot(E_{\varphi_i} \wedge E_{\varphi_i}). \end{aligned}$$

Remark 2 The condition of logical independence of events E_φ, E_ψ is crucial for proving all the above assertions. Nevertheless this condition is not so strong, in fact generally events E_φ representing fuzzy sets are logically independent, even those seemingly linked.

The above assessment $P_\odot$ is a coherent conditional probability, so by using Theorem 1 it can be furthermore extended to any conditional event $A|B$ where A, B are events of the algebra $\mathcal{B}$, with $B \neq \emptyset$. This extension is not unique in general, but for the events $A = E_{\varphi_i}$ and $B = E_{\varphi_j}$ $(i \neq j)$, with $P_\odot(E_{\varphi_j}) > 0$ the only coherent extension is:

$$P_{\odot}(E_{\varphi_i}|E_{\varphi_j}) = \frac{\sum\limits_{\mathbf{x}_{ij} \in \mathcal{C}_{(X_i,X_j)}} (\mu_{\varphi_i}(x_i)) \odot (\mu_{\varphi_j}(x_j)) P_{ij}(\mathbf{x}_{ij})}{\sum\limits_{x_j \in \mathcal{C}_{X_j}} \mu_{\varphi_j}(x_j) P_j(x_j)}.$$

We call the above extension of P a coherent $\odot$-*extension.*

In the case in which we have $P_{\odot}(E_{\varphi_j}) = 0$ for some event E_{φ_j}, we obtain in general a not unique extension to the events $E_{\varphi_i}|E_{\varphi_j}$.

We note that one has $P_{\odot}(E_{\varphi_j}) = 0$ if and only if $P_j(x_j) = 0$ for every $x_j \in \mathcal{C}_{X_j}$ such that $\mu_{\varphi_j}(x_j) > 0$. In this case to obtain a unique extension we need to have also the conditional probability $P(\cdot|B)$, where B is the logical sum of the events $x_j \in \mathcal{C}_{X_j}$ such that $P(E_{\varphi_j}|x_j) = 0$ (the proof of its coherence goes along the same line of that of Corollary 2).

In this case in fact, since $P(B|B) = \sum\limits_{\mathbf{x} \subseteq B} P(x|B) = 1$, at least one event $\mathbf{x} \subseteq B$ is such that $P(\mathbf{x}|B) > 0$ and so

$$P_{\odot}(E_{\varphi_i}|E_{\varphi_j}) = \frac{\sum\limits_{\mathbf{x}_{ij}} (\mu_{\varphi_i}(x_i) \odot \mu_{\varphi_j})(x_j) P(\mathbf{x}_{ij}|B)}{\sum\limits_{x_j} \mu_{\varphi_j}(x_j) P(x_j|B)}.$$

This follows by Theorem 2 (with $\underline{P} = \overline{P} = P$), which states that there exists an index α such that

$$P_{\odot}(E_{\varphi_i}|B) = \frac{P_{\alpha}(E_{\varphi_i} \wedge B)}{P_{\alpha}(B)} ,$$

and in our context P_{α} arises from $P(\mathbf{x}|B)$.

Remark 3 We remark that the values $P_{\odot}(E_{\varphi_i}|E_{\varphi_j})$ computed by the above formulas are coherent only when the events E_{φ} and E_{ψ} are logically independent, so, for instance the same formula cannot be used for obtaining the coherent extension of $P_{\odot}$ to $E_{\varphi}|E_{\varphi}$ which is necessarily 1, independently of the Frank t-norm used for computing union and intersection between the fuzzy sets related to the family of logically independent events E_{φ_i}.

4.1 Lower Probability of Fuzzy Events

In this section till now we consider the case where the probability distribution on the variable X, whose fuzzy set refers, is available. However in many situations this condition does not hold (see e.g. [3]): consider, for instance, the elements of a data base and a statistical information elicited on a set of classes related to some characteristic and a membership assessed taking into account the classes of another attribute.

We analyze different situations of this kind. First of all consider the case where we have a lower probability on the algebra generated by X.

A particular but interesting case is when the lower probability is 2-monotone, i.e. for any A, B in the algebra, $\underline{P}(A \vee B) \geq \underline{P}(A) + \underline{P}(B) - \underline{P}(A \wedge B)$. In fact in this case the Choquet integral can be used [15] for computing the lower bound of all possible coherent values of probability of a fuzzy event E_φ, by using directly the lower probability $\underline{P}$. More precisely, being $\mu(\cdot) = p(E|\cdot)$ a function with values in [0, 1]

$$\int \mu d\underline{P}_H = \int_0^1 \underline{P}_H(s : \mu(z) \geq x)dx$$

where the first integral is a choquet integral with respect to $\underline{P}_H$. This happens for instance when the lower probability on a partition (or algebra) is obtained as lower envelope of the coherent extensions of a probability defined on a different partition (or algebra): in this case in fact, independently of the logical constraints among the events of the two partition we obtain a lower probability completely monotone (belief function) (see e.g. [9, 18]) and so 2-monotone. Therefore

$$\{\underline{P}_\odot(E_{\varphi_i}), \underline{P}_\odot(E_{\varphi_i} \wedge E_{\varphi_j})\}$$

can be computed by means of the Choquet integral.

A different situation can be obtained when we extend on a superalgebra more than one probability assessed on different algebras. For instance when we have marginal probabilities and we are interested on obtain the joint probability (on the conjunctions). In this case in fact we could have an extension not 2-monotone, as the following example shows:

Example 1 Consider two logically independent partitions $\mathcal{E}_1 = \{E_1^1, \ldots, E_3^1\}$ and $\mathcal{E}_2 = \{E_1^2, \ldots, E_3^2\}$, that means $E_i^1 \wedge E_j^2 \neq \emptyset$ (for $i, j = 1, 2, 3$). Consider two probabilities P_1, P_2 on $\mathcal{E}_1, \mathcal{E}_2$, respectively: $P_j(E_i^j) = 1/3$ for $i = 1, 2, 3$ and $j = 1, 2$.

Taking the algebra $\mathcal{A}$ generated by the two partitions and so having atoms $E_{rs} = E_r^1 \wedge E_s^2$ for $r, s = 1, 2, 3$ and the extension $\underline{P}$ on $\mathcal{A}$ of P_1, P_2 we have for the events $A = \vee_{s=1}^3 E_{1s}$ and $B = \vee_{r=1}^3 E_{r1}$

$$\underline{P}(A \vee B) = 1/3, \underline{P}(A) = 1/3 = \underline{P}(B) = 1/3, \underline{P}(A \wedge B) = 0,$$

and so $\underline{P}$ is not 2-monotone.

In the cases where 2-monotonicity does not hold the values obtained by Choquet integral do not coincide with $\{\underline{P}_{\odot}(E_{\varphi_i}), \underline{P}_{\odot}(E_{\varphi_i} \wedge E_{\varphi_j})\}$ and so we need to compute them by the general procedure, using programming optimization systems.

Analogous consideration hold for the computations of

$$\underline{P}_{\odot}(E_{\varphi_i} \vee E\varphi_j).$$

In order to extend the assessment to any conditional event $A|B$ where A, B are events of the algebra $\mathcal{B}$, with $B \neq \emptyset$, we can make the following considerations: the extension is not unique in general, however when we are in the conditions of Example 1, that is the marginal distributions of X_i and X_j are available and $P_{\odot}(E_{\varphi_j}) > 0$, then the coherent lower extension satisfies the equation

$$\underline{P}_{\odot}(E_{\varphi_i}|E_{\varphi_j}) = \frac{\underline{P}_{\odot}(E_{\varphi_i} \wedge E_{\varphi_j})}{P(E_{\varphi_j})}. \tag{11}$$

Then we can conclude that in this case the lack of unicity $\underline{P}_{\odot}(E_{\varphi_i}|E_{\varphi_j})$ is due to the lack of unicity of $\underline{P}_{\odot}(E_{\varphi_i} \wedge E_{\varphi_j})$.

If $P_{\odot}(E_{\varphi_j}) = 0$ for some event E_{φ_j}, then the lack of unicity can be induced by both the causes described before: in this case the lower extension need to be determined through the sequence of systems, described in Sect. 2.

Finally we consider the situation where different experts give different memberships μ_i for the same property φ. We recall that (see Remark 1) the lower envelope of these memberships is still a likelihood $\underline{\mu}$ but we cannot use it for computing the lower probability of the event "You claim the property φ," that is

$$\underline{P}(E_{\varphi}) = \inf \sum \mu_i(x)P(x),$$

since in general we have only

$$\inf \sum \mu_i(x)P(x) \geq \sum \underline{\mu}(x)P(x).$$

We conclude with an example stressing the need of dealing with imprecise probabilities, in presence of partial and vague information:

Example 2 Let H_i (i $= 1, \ldots, 5$) be five urns, contains n balls of various sizes $s_1, \ldots, s_m$ with $m \leq n$, but having different composition with respect to the size. Let us consider also three mutually exclusive and exhaustive events A, B, C, having respectively probability (p_1, p_2, p_3), and such that if A is verified, then one of the urns H_1 and H_2 is chosen, if B is verified, then one among H_2, H_3, H_4 is chosen and finally one between H_4 and H_5 is chosen when C is verified.

We have the following problem: if we do not know which urn has been chosen, what is the probability that a ball drawn at random is small, under the hypothesis that in the urn "most balls are large"?

Given an urn H_i, let us indicate by $n_1, \ldots, n_m$ the number of balls of different sizes $s_1, \ldots, s_m$.

Then the fraction of large balls is for us equal to

$$v_1 P(E_S|s_1) + \cdots + v_m P(E_S|s_m)$$

where $v_i = n_i/n$ and E_S is the event "You claim that the size (of the drown ball) is small".

Note that the fraction of large balls coincides with the probability of $E_S|H_i$ where H_i stands also for the event "the urn H_i is chosen".

First of all we need to compute the probability of E_S should be obtained by disintegration rule, when the probability distribution on H_i is available, but in this case we have a family of probability distributions $\mathcal{P}$ compatible (or coherent) with (p_1, p_2, p_3). Then to compute all the coherent values for $P(E_S)$ we must consider all the probability distributions $p \in \mathcal{P}$ and then to compute the lower and upper value of

$$P(E_S) = \sum_1^r p(H_i) P(E_L|H_i).$$

Now we consider the same problem under the hypothesis that "most balls are large". Introduce for that a new (random variable): $\Pi =$ "percentage of balls with a size such that you claim large".

The range of Π is

$$\{v^j = v_1^j P(E_L|s_1) + \cdots + v_m^j P(E_L|s_m)\},$$

where E_L is the event "You claim that the size (of the drown ball) is large" and v_i^j is the percentage of s_i in the (possible) urn H_j. So the values of Π are exactly the probabilities $P(E_L|H_j)$.

Now we consider the event E_H= "You claim that Π_L is high" and the fuzzy set (E_{LH}, μ_{LH}), with $\mu_{LH}(v^r) = P(E_{LH}|\Pi = v^r)$.

In order to compute the values of $P(E_S|E_{LH})$, by assuming the independence of E_S and E_{LH} conditionally to any possible composition of an urn H_i, one has as lower probability

$$\underline{P}(E_S|E_{LH}) = \inf_p \sum_k P(E_S|H_k) p(H_k|E_{LH})$$

(and similarly for the upper) with

$$p(H_k|E_{LH}) = \frac{p(H_k) P(E_{LH}|H_k)}{\sum_k p(H_k) P(E_{LH}|H_k)}.$$

5 Probabilistic Fuzzy Reasoning

A relevant problem in literature is the managing of fuzzy rules based systems: they are essentially composed by some rules of the form

"IF A THEN B, with a given probability",

where either premise A and consequence B of the rule can be fuzzy sets [24, 27, 33, 35].

For example, we consider the task of determining the class $y \in \{C_1, \ldots, C_c\}$ to which a data point $\mathbf{x} = (x_1, \ldots, x_n)$ belongs.

For this aim a probabilistic fuzzy classifier needs to be performed, by giving some rules such as:

if $\mathbf{x}$ is E_{φ_j} then $y = C_k$ with probability $p_{k|j}$.

The main problem related to the performance of a probabilistic fuzzy system consists into the determination or elicitation of the probabilities (or lower probabilities) $p_{k|j}$,:

$$p_{k|j} = P(C_k|E_{\varphi_j}) = \frac{P(E_{\varphi_j}|C_k)P(C_k)}{\sum_{r=1}^{c} P(E_{\varphi_j}|C_r)P(C_r)};$$

When the unicity is not guarantied, the lower and upper envelope of a class of such probabilities must be computed.

The values $p_{k|j}$ could be seen as values of degree of inclusion [29].

In the following we discuss different situations by using an example.

Example 3 Let us consider an urn with balls of different sizes and colors and let X_S be the variable "size of the balls" with $\mathcal{C}_{X_S} = \{s_1, \ldots, s_n\}$, and X_C be the variable "color of the balls" with $\mathcal{C}_{X_C} = \{c_1, \ldots, c_m\}$.

Moreover we take in consideration the following properties φ_j $(j = 1, 2, 3)$ related to X_S and $\psi_i = (i = 1, 2)$, related to X_C

$\varphi_1 =$ "small",
$\varphi_2 =$ "medium",
$\varphi_3 =$ "large";
$\psi_1 =$ "dark",
$\psi_2 =$ "light".

Given the fuzzy sets $E^*_{\varphi_j} = (E_{\varphi_j}, \mu_{\varphi_j})$ $(j = 1, 2, 3)$, with $\mu_{\varphi_j}(s_r) = P(E_{\varphi_j}|s_r)$ for any $s_r \in \mathcal{C}_{X_S}$, let us consider the problem of determining

$$p_{k|j} = P(c_k|E_{\varphi_j})$$

with $c_k \in \mathcal{C}_{X_C}$.

The simplest case is when a joint probability on the vector (X_S, X_C) is available, in fact in this case coherence together conditional independence of E_{φ_j} and X_C given X_S implies

$$P(c_k|E_{\varphi_j}) = \sum_{d_r} P(c_k|d_r \wedge E_{\varphi_j})\, P(d_r|E_{\varphi_j}) = \sum_{d_r} P(c_k|d_r) \frac{P(d_r)\mu_{\varphi_j}(d_r)}{\sum_{d_s} P(d_s)\mu_{\varphi_j}(d_s)}.$$

If the joint probability is obtained by an extension procedure which start from an other partition (for instance finer then that of interest) then we can compute upper and lower bound of $P(c_k|E_{\varphi_j})$ by means of Choquet integral using the upper and lower bound of the coherent extensions, that in this case is 2-monotone [15].

When indeed only the marginal probability distribution of X_S and X_C (or a conditional probability of X_C given X_S)) are given and the two variables are not stochastically independent, we could compute the lower and upper probabilities on $\mathcal{C}_{(X_S,X_C)}$ and then to compute the extension in a similar way than in Eq. (11).

Suppose now we are interested into compute the probability that "the ball is dark if the size is small", which in our context is the coherent values of $P(E_{\psi_1}|E_{\varphi_1})$. The value could be not unique and in the case that the value 1 is coherent we could consider a weak implication [11] (or maximal degree of inclusion [10]).

In this last case the probabilistic fuzzy rules become of this kind:

IF X_S is small THEN $X_C$is dark with probability 1

and the entailment follows the rules of default logic.

More precisely, we start from a list of IF-THEN rules (weak implications) denoted as a set $\mathcal{D}$ of (ordered) pairs (E_φ, E_ψ) of fuzzy subsets, with degree

$$I(E_\psi, E_\varphi) = P(E_\varphi|E_\psi)$$

of fuzzy inclusion (of E_ψ in E_φ) equal to 1, and denote any such set a MDFI–set ("Maximum Degree of Fuzzy Inclusion" set).

We wonder now whether, given $\mathcal{D}$, it is possible to find further pairs of fuzzy subsets in $\mathcal{F}_\mathcal{C}$ with maximum degree of fuzzy inclusion (MDFI–pairs).

Even if for any coherent assessment on $\mathcal{C}$ its enlargement to a family $\mathcal{K} \supseteq \mathcal{C}$ is not (in general) unique, nevertheless for some events we can have a unique coherent extension, so giving rise to the important concept of *entailment*.

More generally MDFI–set $\mathcal{D}$ entails the pair (E_φ, E_ψ) of fuzzy sets with degree belonging to an interval $[p', p'']$ if the coherent value for $P(E_\psi|E_\varphi)$ are all the values in $[p', p'']$.

In particular the MDFI–set $\mathcal{D}$ strictly entails the pair (E_φ, E_ψ) if the only coherent value for $P(E_\psi|E_\varphi)$ is $p', p'' = 1$.

As proved in [10] the MDFI–set satisfy all the inferential rules of default logic [23].

Concerning the rules IF-THEN with probability belonging to an interval *I*, where the extremes are lower and upper conditional probabilities, we could denote by MDFIP-set the set of these rules, and the entailment satisfies the System P rules, see [21].

References

1. Biazzo, V., Gilio, A., Lukasiewicz, T., Sanfilippo, G.: Probabilistic logic under coherence: complexity and algorithms. Ann. Math. Artif. Intell. **45**, 35–81 (2005)
2. Coletti, G.: Coherent numerical and ordinal probabilistic assessments. IEEE Trans. Syst Man Cybern **24**, 1747–1754 (1994)
3. Coletti, G., Gervasi, O., Tasso, S., Vantaggi, B.: Generalized Bayesian inference in a fuzzy context: From theory to a virtual reality application. Comput. Stat. Data Anal. **56**(4), 967–980 (2012)
4. Coletti, G., Scozzafava, R.: Characterization of coherent conditional probabilities as a tool for their assessment and extension. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **4**, 103–127 (1996)
5. Coletti, G., Scozzafava, R.: Probabilistic Logic in a Coherent Setting. Trends in logic, n.15. Kluwer, Dordrecht (2002)
6. Coletti, G., Scozzafava, R.: Conditional probability, fuzzy sets, and possibility: a unifying view. Fuzzy Sets Syst. **144**, 227–249 (2004)
7. Coletti, G., Scozzafava, R.: Conditional probability and fuzzy information. Comput. Stat. Data Anal. **51**, 115–132 (2006)
8. Coletti, G., Scozzafava, R.: The role of coherence in eliciting and handling imprecise probabilities and its application to medical diagnosis. Inf. Sci. **130**, 41–65 (2000)
9. Coletti, G., Scozzafava, R.: Toward a general theory of conditional beliefs. Int. J. Intell. Syst. **21**, 229–259 (2006)
10. Coletti, G., Scozzafava, R., Vantaggi, B.: Coherent conditional probability, fuzzy inclusion and default rules. In: Proceedings of First World Conference on Soft Computing. San Francisco (2011)
11. Coletti, G., Scozzafava, R., Vantaggi, B.: Weak implication in terms of conditional uncertainty measures. In: Lecture Notes in Computer Science LNAI 4724, pp. 139–150 (2007)
12. Coletti, G., Vantaggi, B.: Hybrid models: probabilistic and fuzzy information. In: Kruse, R.; Berthold, M.R.; Moewes, C.; Gil, M.Ã.; Grzegorzewski, P.; Hryniewicz, O. (eds.) Synergies of Soft Computing and Statistics for Intelligent Data Analysis, Advances in Intelligent Systems and Computing, vol. 190, pp. 389–397 (2013)
13. de Finetti, B.: Sull'impostazione assiomatica del calcolo delle probabilità. Annali Univ. Trieste **19**, 3–55 (1949) (Engl. transl. in: Chap. 5 of Probability, Induction, Statistics, Wiley, London, 1972)
14. de Finetti, B.: Teoria della probabilità 1974. Einaudi, Torino (1970) (Engl. transl.: "Theory of Probability", Vol. 1 and 2, Wiley, Chichester)
15. Denneberg, D.: Non-Additive Measure and Integral. Kluwer, Dordrecht (1994)
16. Dubins, L.E.: Finitely additive conditional probabilities, conglomerability and disintegration. Ann. Probab. **3**, 89–99 (1975)
17. Dubois, D., Moral, S., Prade, H.: A semantics for possibility theory based on likelihoods. J. Math. Anal. Appl. **205**, 359Â–380 (1997)

18. Fagin, R., Halpern, J.: Uncertainty, belief and probability. Comput. Intell. **7**, 160–173 (1991)
19. Frank, M.J.: On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes Math. **19**, 194–226 (1979)
20. Gilio, A.: Algorithms for precise and imprecise conditional probability assessments. In: Coletti, G., Dubois, D., Scozzafava, R. (eds.) Mathematical Models for Handling Partial Knowledge in Artificial Intelligence. Plenum Press, New York, pp. 231–254 (1995)
21. Gilio, A.: Probabilistic reasoning under coherence in system P. Ann. Math. Artif. Intell. **34**, 5–34 (2002)
22. Hisdal, E.: Are grades of membership probabilities. Fuzzy Sets Syst. **25**, 325–348 (1988)
23. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artif. Intell. **55**, 1–60 (1992)
24. Liu, Z., Li, H.-X.: A probabilistic fuzzy logic system for modeling and control. IEEE Trans. Fuzzy Syst. **13**(6), 848-Â–859 (2005)
25. Loginov, V.I.: Probability treatment of Zadeh membership functions and theris use in patter recognition. Eng. Cybern. **2**, 68–69 (1966)
26. Krauss, P.H.: Representation of conditional probability measures on boolean algebras. Acta Math. Acad. Scient. Hungar. **19**, 229–241 (1968)
27. Meghdadi, A.H., Akbarzadeh-T, M.-R.: Probabilistic fuzzy logic and probabilistic fuzzy systems. In: Proceedings of IEEE International Fuzzy Systems Conference, pp. 1127–1130 (2001)
28. Miranda, E., de Cooman, G., Couso, I.: Lower previsions induced by multi-valued mappings. J. Stat. Plan. Infer. **133**(1), 173–197 (2005)
29. Scozzafava, R., Vantaggi, B.: Fuzzy inclusion and similarity through coherent conditional probability. Fuzzy Sets Syst. **160**, 292–305 (2009)
30. Singpurwalla, N.D., Booker, J.M.: Membership functions and probability measures of fuzzy sets (with discussion). J. Am. Statist. Assoc. **99**, 867–889 (2004)
31. Zadeh, L.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
32. Zadeh, L.A.: Probability measures of fuzzy events. J. Math. Anal. Appl. **23**, 421427 (1968)
33. Zadeh, L.A.: Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. J. Stat. Plan. Infer. **105**, 233–264 (2002)
34. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
35. Waltman, L., Kaymak, U., van den Berg, J.: Maximum likelihood parameter estimation in probabilistic fuzzy classifiers. In: Proceedings of IEEE International Conference on Fuzzy Systems, pp. 1098–1103 (2005)
36. Williams, P.M.: Notes on conditional previsions. Technical Report, School of Mathematical and Physical Sciences, University of Sussex 1975

Approximate Z-number Evaluation Based on Categorical Sets of Probability Distributions

Saied Tadayon and Bijan Tadayon

Abstract In this chapter, we present a method for approximate evaluation of Zadeh's Z-numbers using category sets of probability distributions corresponding to similar certainty measures.

1 Introduction

Lotfi A. Zadeh introduced the concept of a Z-Number denoted as an ordered pair (A, B), where A and B are fuzzy numbers (typically perception-based and described in natural language), in order to describe the level of certainty or reliability of a fuzzy restriction of a real-valued uncertain variable X in Z-valuation (X, A, B) [1]. For example, the proposition "the price of ticket is usually high", may be expressed as a Z-valuation (price or ticket, high, usually). In Z-valuation, the certainty component B describes the reliability of the possibilistic restriction, R, for the random variable X, where

$$R(X) : X \text{ is } A, \tag{1}$$

with the reliability restriction given by

$$\text{Prob}(X \text{ is } A) \text{ is } B. \tag{2}$$

S. Tadayon · B. Tadayon (✉)
Z Advanced Computing, 11204 Albermyrtle Road, Potomac, MD 20854, USA
e-mail: bijantadayon@ZAdvancedComputing.com

S. Tadayon
e-mail: saiedtadayon@ZAdvancedComputing.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_8,

In another words, the certainty component B, restricts the probability measure of A, denoted by v,

$$v = \text{Prob}(X \text{ is } A) = \int_X \mu_A(x) \cdot p_x(x) \cdot dx, \quad (3)$$

where $\mu_A(x)$ is the membership function of x in fuzzy set A on X domain, and p_X is the probability distribution of X. Therefore, the certainty component B indirectly restricts the possibilities of various (candidate) hidden probability distributions of X by

$$\mu_B(v) = \mu_B\left(\int_X \mu_A(x) \cdot p_x(x) \cdot dx\right), \quad (4)$$

where $\mu_B(v)$ is the membership function of the probability measure v in fuzzy set B.

In this chapter, we present a method to approximate Z-valuation, based on categories (sets) of p_X's with similar probability measures (or resulting in similar certainty measure), as an approach to reuse predetermined calculations of probability measures. First, we demonstrate an example of Z-valuation without such approach, and then, we present an approximate approach to Z-valuation via categorical sets of probability distributions.

1.1 Z-valuation: Basics

The Z-valuation uses the mapping of the test scores given by (4) to each of hidden probability distribution candidates of X (see [1] and [2]), collectively referred to as

$$\text{Prob. Distrib. Candidates} = \{p_i\}, \quad (5)$$

where i numerates different candidates. Figure 1 conceptually illustrates the mapping, where each p_i is first mapped to a probability measure of A, v_i, and then mapped to a test score determined by B, where

$$v_i = \mu_A \cdot p_i = \int_X \mu_A(x) \cdot p_i(x) \cdot dx, \quad (6)$$

and

$$ts_i = \mu_B(v_i). \quad (7)$$

Note that the dot symbol in $(\mu_A \cdot p_i)$ in (6) is used as shorthand for the probability measure [1].

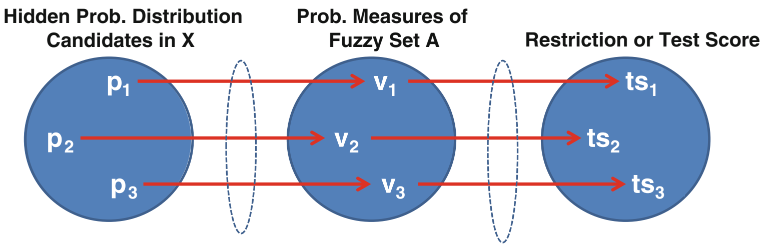


Fig. 1 Test score mapping to hidden probability distribution candidates p_i in X, for Z-valuation (X,A,B)

Via the extension principle, Zadeh illustrated the application of the restriction (test scores) on $p_{x,i}(x)$ (i.e., probability distribution candidates in X domain) to other entities [1]. For example, the restriction on $p_{x,i}(x)$ can be extended to the possibilistic restriction on the corresponding probability distributions, $p_{y,i}(y)$, in Y domain, where

$$Y = f(X).$$

In such a case, the restrictions can further be extended to the probability measures, w_i, of a fuzzy set A_y in Y domain, based on $p_{y,i}(y)$. The aggregation of the best test scores for w_i would determine the certainty component B_Y in Z-valuation (Y, A_Y, B_Y), based on the original Z-valuation (X, A_X, B_X), as indicated in Fig. 2 which illustrates the extension of test scores to Y domain.

For simplicity, as shown in Fig. 2, three probability distribution candidates in X domain, $p_{x,1}$, $p_{x,2}$, and $p_{x,3}$, are assigned test scores ts_1 and ts_2, via certainty restriction on probability measures v_1 and v_2 (with $p_{x,2}$, and $p_{x,3}$, having the same probability measure v_2 for A_X). By applying f(X) to each probability distribution candidate in X domain, we can obtain a corresponding probability distribution in Y domain, denoted as $p_{y,i}$, which can be used to compute the corresponding probability measure of A_Y (assume given), denoted as w_i. In this example, $p_{y,1}$, and $p_{y,2}$ (mapped from $p_{x,1}$, and $p_{x,2}$) result in the same probability measure w_2 (or aggregated w bin), while $p_{y,3}$ (mapped from $p_{x,3}$) maps into w_1. In this simple example, the aggregation of the best test scores for $p_{y,i}$, denoted as $ts(p_{y,i})$, in w domain (e.g., in each w bin) would result in the following membership function for B_Y:

$$\mu_{B_Y}(w_1) = ts_2$$
$$\mu_{B_Y}(w_2) = \max(ts_1, ts_2).$$

In other words, in this scenario,

$$\mu_{B_Y}(w) = \sup_{\forall p_{y,i}} ts(p_{y,i}) \qquad (8)$$

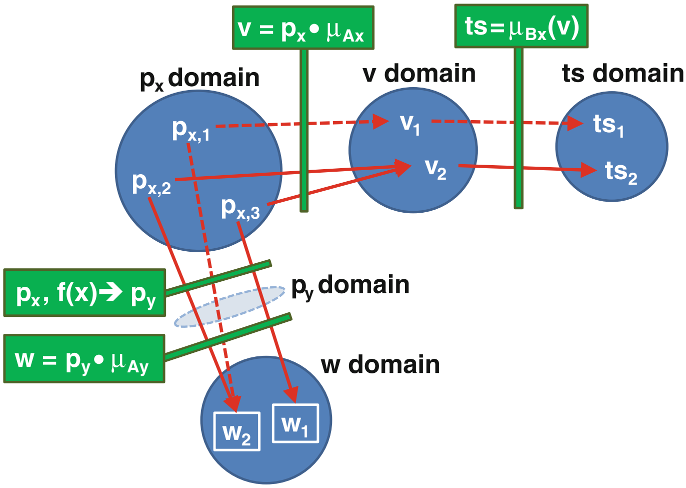


Fig. 2 Test score mapping from X domain to Y domain and aggregation of test scores on probability measures, w, for Z-valuation (Y, A_Y, B_Y)

subject to

$$w = \mu_{A_Y} \cdot p_{y,i}.$$

In case of single variable dependency Y = f(X), the probability measure w can be evaluated by unpacking the probability distribution in Y as illustrated by (9) and transforming the integration over X domain as shown in (10), without explicitly evaluating $p_{y,i}$:

$$w_i = \mu_{A_Y} \cdot p_{y,i} = \int_Y \mu_{A_Y}(y) \cdot p_{y,i}(y) \cdot dy$$

$$= \int_Y \mu_{A_Y}(y) \cdot \sum_j \frac{p_{x,i}(x_j)}{|f'(x_j)|} \cdot dy \tag{9}$$

where j denotes the consecutive monotonic ranges of f(X) in X domain, and x_j is the solution for $f^{-1}(y)$, if any, within the monotonic range j, for a given y. This takes into account that the probability ($p_{y,i}$. dy) for an event within the infinitesimal interval of [y, y + dy] in Y domain, is the summation of the infinitesimal probabilities from various infinitesimal intervals [x_j + dx_j] (if applicable) in X domain, where for each j:

$$dy = f'(x_j) \cdot dx_j.$$

Therefore, with repacking the integration (9) in X domain over the consecutive monotonic ranges of f(X), we obtain:

$$w_{\mathrm{i}} = \int_X \mu_{A_Y}(f(x)) \cdot p_{x,i}(x) \cdot dx \tag{10}$$

Furthermore, if f(X) is monotonic (i.e., $f^{-1}(y)$ has only one solution in X, if any) *AND* μ_{A_Y} is obtained from μ_{A_X} via the extension principle by applying f(X) to A_X, then w_i is guaranteed to be equal to v_i for all candidate probability distributions $p_{x,i}$, because $\mu_{A_Y}(y) = \mu_{A_X}(x)\, for\, \forall y = f(x)$ in such a case. This also means that in such a case, B_Y becomes equal to B_X, and no additional computation would be necessary.

1.2 Z-valuation: Example

To illustrate an example of Z-valuation, assume the followings are given:

$$X = (A_X, B_X),$$
$$Y = f(X) = (X+2)^2, \text{ and}$$
$$A_Y.$$

The goal is to determine the certainty value B_Y for the proposition that (Y is A_Y), i.e., the Z-valuation (Y, A_Y, B_Y).

For purpose of this example, assume Figs. 3, 4, and 5 depict the membership functions for A_X, B_X, and A_Y, respectively. The function f(X) is also depicted in Fig. 6.

In this example, the set of candidate probability distribution for X was constructed using Normal distributions with mean (m_x) ranging from −2 to 2 and standard deviation (σ_x) ranging from 0^+ (close to Dirac delta function) to 1.2. Figures 7 and 8 depict the probability measure of A_X, denoted as v, based on (3) and each of these probability distribution candidates represented by a point on (m_x, σ_x) plane. These also illustrate the contour maps of constant probability measures. Figures 9 and 10 depict the test scores (denoted as ts) for each probability distribution candidate, based on the application of certainty component B_X to each probability measure, v, via (4). Given that B_X imposes a test score on each v, the probability distribution candidates that form a contour (on (m_x, σ_x) plane) for constant v, also form a contour for the corresponding test score. However, given that a range of v values may result in the same test score (e.g., for v less than 0.5 or above 0.75 in this example), some test score contours on (m_x, σ_x) plane collapse to flat ranges (e.g., for test scores 0 and 1, in this example) as depicted on Figs. 9 and 10.

By applying (10), we can then determine the probability measure of A_Y (in Y domain), denoted as w, based on the probability distribution candidates in X domain

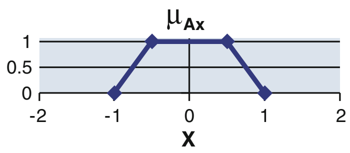


Fig. 3 The membership function of A_X, e.g., "X is around zero"

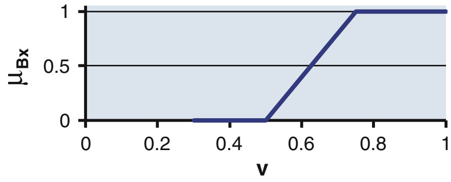


Fig. 4 The membership function of B_X, e.g., "Likely"

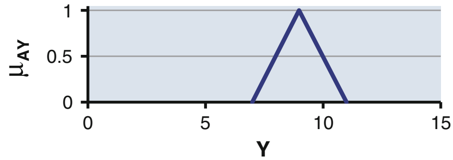


Fig. 5 The membership function of A_Y, e.g., "Y is about nine"

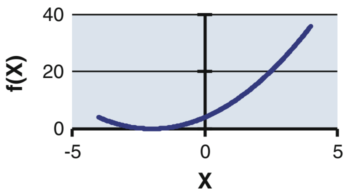


Fig. 6 Diagram depicting f(X)

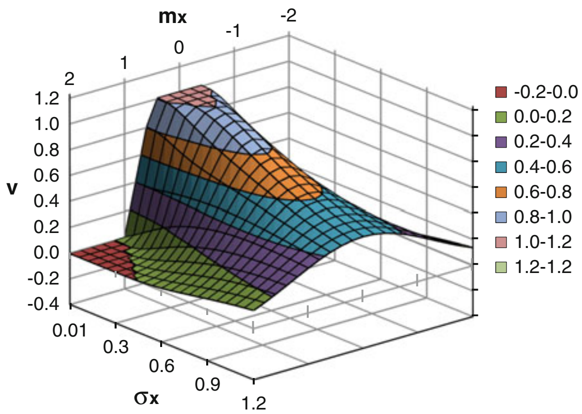


Fig. 7 The probability measure of A_X, v, per each (Normal) probability distribution candidate represented by (m_X, σ_X)

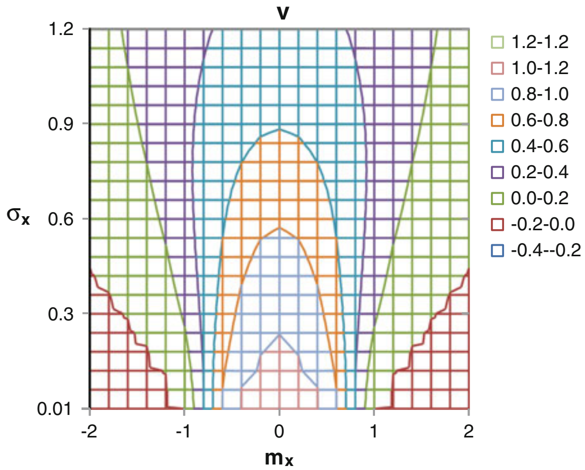


Fig. 8 The contours of the probability measure of A_X,v, per each (Normal) probability distribution candidate represented by (m_X, σ_X)

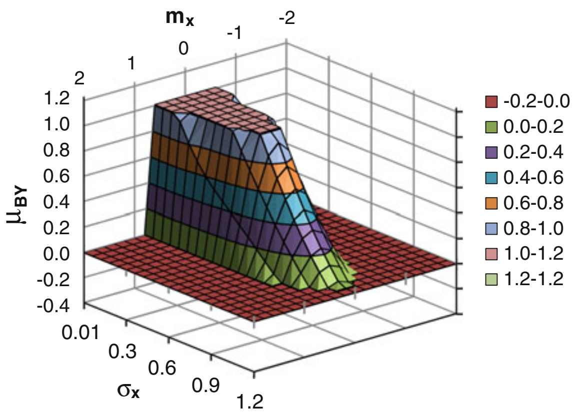


Fig. 9 The test score based on certainty measure B_X for each (Normal) probability distribution candidate represented by (m_X, σ_X)

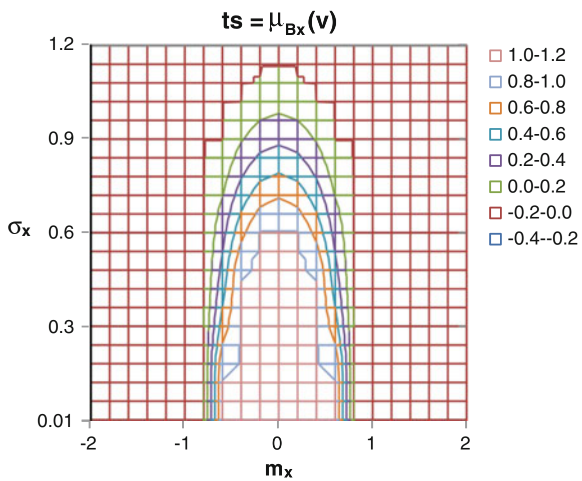


Fig. 10 The test score based on certainty measure B_X for each (Normal) probability distribution candidate represented by (m_X, σ_X)

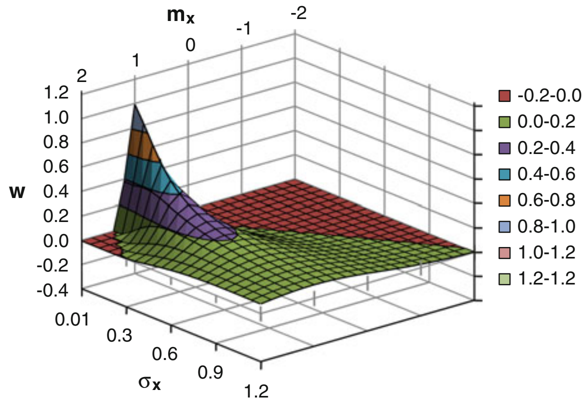


Fig. 11 The probability measure of A_Y,w, per each probability distribution (Normal) candidate represented by (m_X, σ_X)

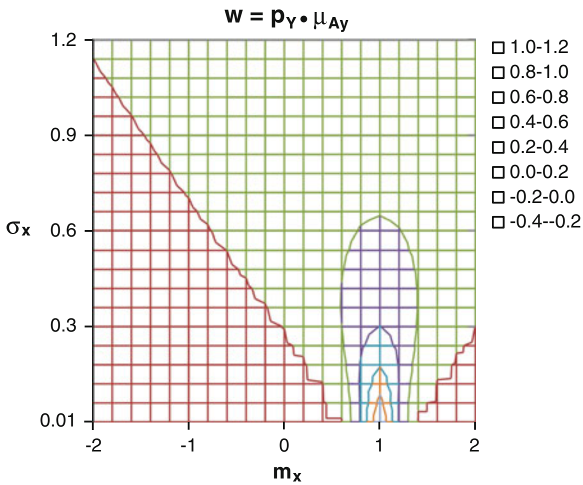


Fig. 12 The contours of the probability measure of A_Y,w, per each probability distribution (Normal) candidate represented by (m_X, σ_X)

(i.e., bypassing the direct calculation of the corresponding probability distributions in Y domain). The probability measure w is depicted in Figs. 11 and 12 for each probability distribution candidate in (m_x, σ_x) plane.

Given that each probability distribution candidate is associated with a possibility restriction test score (as shown for example in Fig. 10), such test score can be applied and correlated with the probability measure w (shown for example in Fig. 12). A given w (or a w bin) may be associated with multiple test scores as indicated by contours of constant w or regions of very close or similar w in Fig. 12.

Therefore, to assign a final test score to a given w (or w bin) based on (8), we can determine the maximum test score for all w's associated with the given w bin.

The result of an intermediate step for determining the maximum test score for correlated w's (i.e., falling in the same w bin) is illustrated in Fig. 13, on the (m_x, σ_x) plane (for illustrative comparison with Fig. 11).

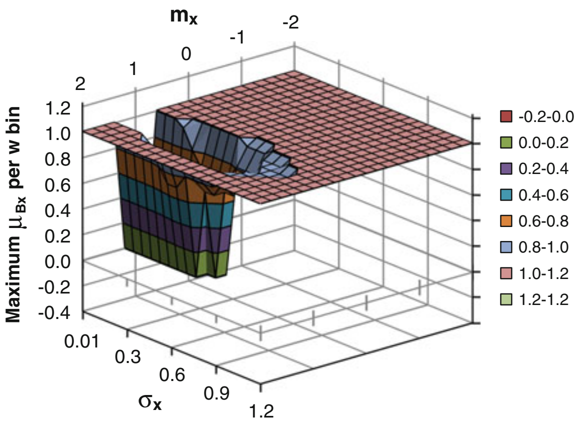


Fig. 13 The maximum test score for a w-bin associated with each probability distribution (Normal) candidate represented by (m_X, σ_X)

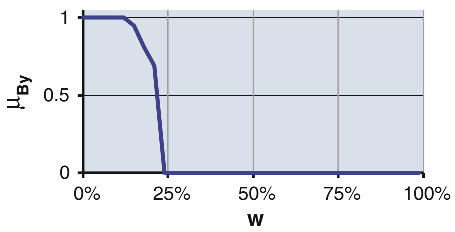


Fig. 14 The maximum test scores for w-bins defining the membership function of w in fuzzy set B_Y, e.g., "significantly less than 25 %."

The resulting maximum test score associated with a given w bin defines the membership function of w (or a value of w representing the w bin) in B_Y, as depicted for this example in Fig. 14. As shown in Figs. 11 and 13, where w is high, the maximum associated test score is low, resulting in B_Y which represents "significantly less than 25 %" for this example.

2 Z-Valuation Using Granular Category Sets

2.1 Predetermined Category Sets: Test Scores, Probability Measures, and Probability Distributions

The probability measure of A_X, denoted as v, may be predetermined and reused, given that the integration in (3) may be normalized based on the general shape of the membership function of A_X and the class/parameters of probability distribution candidates. In normalized form, for example, a category of normalized membership function may be defined as symmetric trapezoid with its support at interval $[-1,1]$ with a single parameter, β, indicating the ratio of its core to its support (as shown in Fig. 15). Examples of classes of probability distribution are Normal distribution and

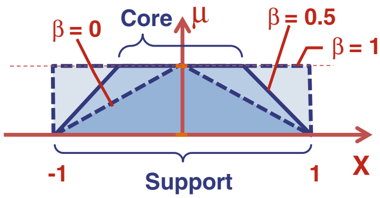


Fig. 15 Membership function parameter β (ratio of core to support) adjusts the symmetric trapezoid shape from triangular with ($\beta = 0$) to crisp with ($\beta = 1$)

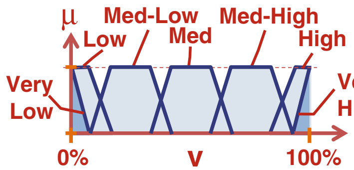


Fig. 16 Examples of various granular (fuzzy) sets of probability measures

Poisson distribution, with their corresponding parameters normalized with respect to normalized A_X. For example, for Normal distribution, the parameters (m_x, σ_x) may be normalized with respect to half width of the support having the origin of the normalized coordinate translated to cross zero at the center of the support.

Furthermore, we may reduce the level and complexity of computation in approximating the Z-valuation by using a granular approach. For example, for a category of normalized A_X (e.g., symmetric trapezoid with β of about 0.5 as shown in Fig. 15), we may predetermine relations/mapping (or a set of inference rules) between (fuzzy or crisp) subset of probability distribution candidates (of a given class such as Normal or Poisson distribution) and (fuzzy or crisp) subsets of probability measures, v's (as for example shown in Fig. 16).

Let V_j denote a category/set of probability measures of A_X (e.g., probability measure "High"), where j numerates such categories in v domain. Each V_j corresponds to a range or (fuzzy or crisp) subset of probability distribution candidates, denoted by C_j whose p_i members are defined via the following membership function:

$$\mu_{C_j}(p_i) = \mu_{V_j}(\mu_A \cdot p_i) = \mu_{V_j}\left(\int_X \mu_A(x) \cdot p_i(x) \cdot dx\right), \tag{11}$$

Therefore according to (11), we may predetermine C_j via a similar method of applying test scores to the probability distribution candidates, p_i (as for example shown in Fig. 9), by replacing B_x with V_j. For example, the categories of probability measure V_{Low} and V_{High} (shown in Figs. 17 and 18, respectively), correspond to the (category) fuzzy sets of probability distribution candidates, denotes as C_{Low} and C_{High} (with labels used in place of j), with a membership function depicted in Figs. 19 and 20, respectively.

Furthermore, the certainty levels (test scores) may also be made into granular (fuzzy or crisp) sets TS_k, e.g., in order to reduce the complexity of calculation

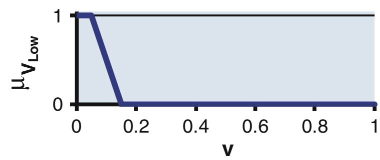


Fig. 17 Membership function of v in V_{Low}

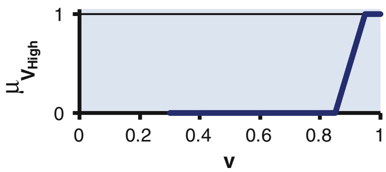


Fig. 18 Membership function of v in V_{High}

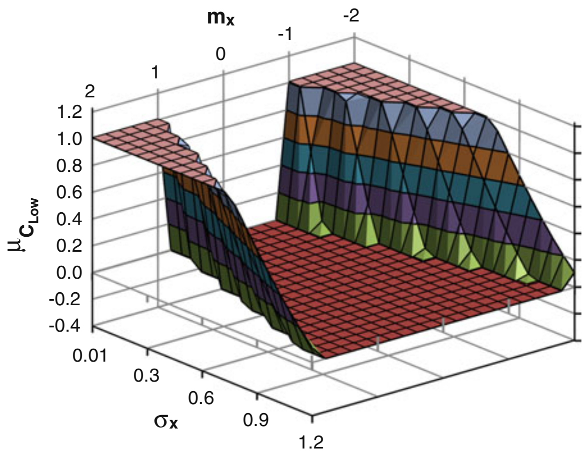


Fig. 19 Membership function of p_i in C_{Low} (with p_i represented by its parameters (m_X, σ_X)

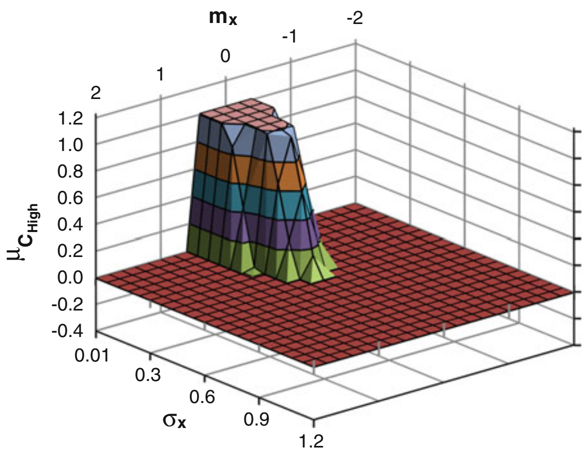


Fig. 20 Membership function of p_i in C_{High} (with p_i represented by its parameters (m_X, σ_X)

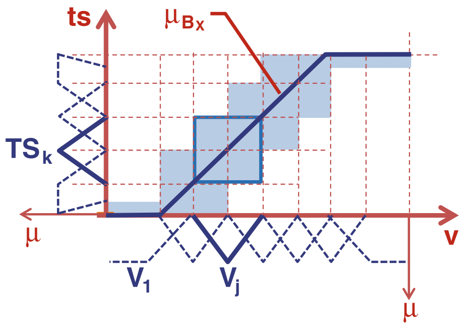


Fig. 21 An example of granularizing/mapping of B_X, via (V_j,TS_k) pairs

during the aggregation process of Z-valuation. Index k numerates these test score category sets. Figure 16 may also serve as an example of such categorization (with test score replacing v).

In one approach, the certainty component B_X is granularly decomposed or mapped (or approximately expressed) via pairs of probability measure and test score category sets, i.e. (V_j, TS_k)'s, as for example demonstrated in Fig. 21. In one approach, each relation pair may be further associated with a $weight_{j,k}$ that indicates the degree of mapping of B_X among the pairs (e.g., when TS_k is a predefined set). For example:

$$weight_{j,k} = \sup_{v \in [0,1]} \left(\mu_{V_j}(v) \wedge \mu_{TS_k}\left(\mu_{B_X}(v)\right) \right).$$

In one scenario, the decomposition of B_X may be expressed as series of tuples in the form (V_i, TS_k, $weight_{j,k}$) or simply as a matrix with $weight_{j,k}$ as its elements. Given the correspondence between C_j and V_j, the granular test score sets TS_k's are also associated with granular probability distribution candidate sets, C_j's (with the same $weight_{j,k}$).

In another approach, a non-categorical test score (e.g., a fuzzy or crisp set) TS_j is determined for each V_j (and C_j), e.g., by using extension principle, based on mapping via B_X:

$$\mu_{TS_j}(ts) = \sup_{v' \in [0,1]} \left(\mu_{V_j}\left(v'\right) \right), \tag{12}$$

subject to $ts = \mu_{B_X}\left(v'\right)$.

2.2 Computation and Aggregation via Normalized Categories

One advantage of reusing the predetermined normalized categories is the reduction in number of calculations such as the integration or summation in determining probability measures per individual probability distribution candidates in X domain or their corresponding probability distributions in Y domain, per (4) and (8). In addition, instead of propagating the test scores via an individual probability distribution candidate, the extension of the test scores may be done at a more granular level of the probability distribution candidate subsets, C_j, which are typically far fewer in number than the individual probability distribution candidates. However, the aggregation of test scores for Z-valuation, e.g., for (Y, A_Y, B_Y), will involve additional overlap determination involving various normalized category sets, as described below.

The normalization of symmetrical trapezoid membership function A_Y, e.g., "Y is about nine," as shown in Fig. 5, involves shifting the origin by −9 and scaling the width by 0.5 (in Y domain) in order to match the position and width of the support to the normalized template depicted in Fig. 15 (with $\beta = 0$ determined as the ratio of the core to support). Note that such normalization (translation and scaling) also impacts the location and scaling of associated p_y's (e.g., mean and standard deviation) in order to preserve the probability measure of A_Y per (8).

Note that the predetermined categorical subset of probability distributions in Y domain, denoted as $C_{Y,j}$, that is associated with V_j, may be distinct from the corresponding one in X domain, denoted as $C_{X,j}$, e.g., due to parameters such as β (or the class of the membership, such as trapezoid or ramp). For example, Fig. 22 illustrates the membership function of $C_{Y,High}$, for normalized A_Y ($\beta = 0$), for comparison with $C_{X,High}$, depicted in Fig. 20, for the same values of normalized probability distribution parameters.

2.2.1 Mapping in X Domain

In one approach to estimate (10), we may determine (or approximate) $\mu_{A_Y}(f(x))$ in X domain as for example depicted in Fig. 23, labeled $\mu_{A_{Y\to X}}(x)$. Then, we may proceed with mapping and normalization of the membership function to one or more normalized categories of membership functions (e.g., a symmetric trapezoid shape with ($\beta = 0$)).

In such an approach, the normalization effects on Ax and $A_{Y\to X}$ are combined into a transformation operation, T (e.g., translation and scaling) used to also transform the normalized probability distribution parameters (e.g., mean and standard deviation). Thus, T also transforms the predetermined subsets of probability distribution candidates, $C_{X,j}$, to $C^T_{X,j}$, e.g., via the extension principle, as follows:

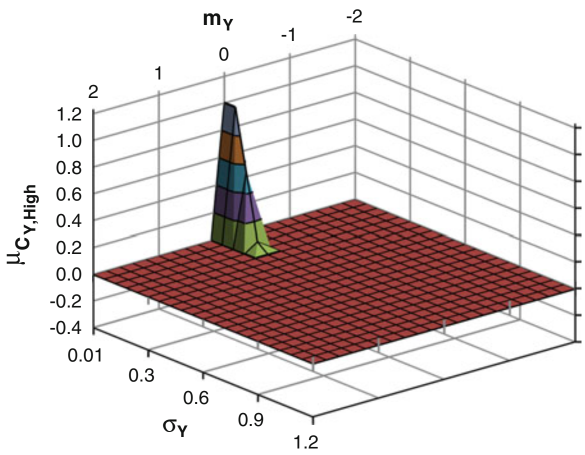


Fig. 22 Membership function of p_y in $C_{Y,High}$ (with p_y represented by its parameters (m_Y, σ_Y)

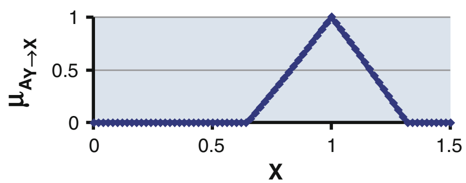


Fig. 23 Membership function $\mu_{A_{Y \to X}}(x)$

$$\mu_{C_{X,j}^T}\left(p_{X,i}^T\right) = \sup_{\forall p_{X,i}} \mu_{C_{X,j}^T}\left(p_{X,i}\right),$$

subject to

$$p_{X,i}^T = \mathrm{T}\left(p_{X,i}\right), \tag{13}$$

where $p_{X,i}^T$ represents the transformed probability distribution candidate (in X domain) from $p_{X,i}$.

Since in our example, μ_{Ax} (depicted in Fig. 3) is already in a normalized form, we focus on the transformation due normalization of $\mu_{A_{Y \to X}}(x)$. Note that in Fig. 11, the outline of probability measure w for ($\sigma_X = 0+$) is the same as the membership function $\mu_{A_{Y \to X}}(x)$ prior to the normalization, as depicted in Fig. 23. To normalize $\mu_{A_{Y \to X}}(x)$, the membership function must be scaled by factor of about 3, denoted by s, and translated by the amount of -3 (or -1 before scaling), denoted by t. The ordered translation and scaling operations, denoted by T_t and T_s respectively, define the transformation operation which also transforms a probability distribution (13) by scaling and translating its parameters, for example:

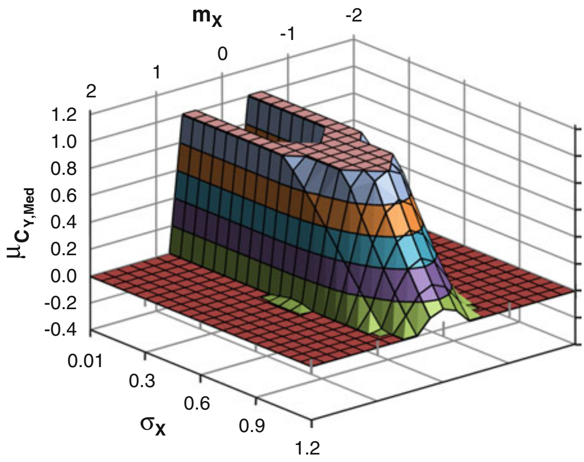


Fig. 24 Membership function $C_{Y,Med}$

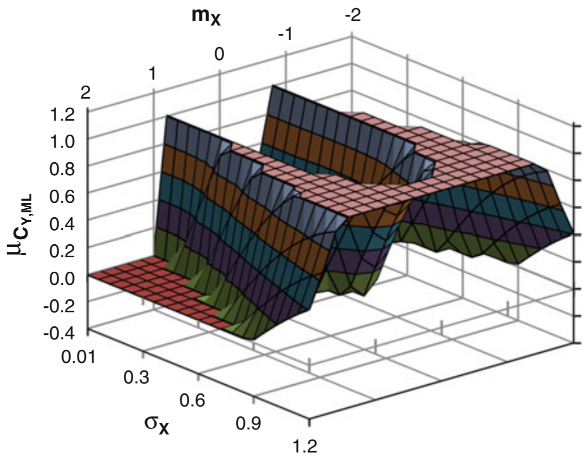


Fig. 25 Membership function $C_{Y,ML}$

$$p_{X,i}^{T} = \mathrm{T}\left(p_{X,i}\right) = \mathrm{T_t} \cdot \mathrm{T_s} \cdot p_{X,i}, \tag{14}$$

with

$$\begin{aligned} \mathrm{T_s} \cdot p_{X,i} &= \mathrm{T_s}\left(m_{X,i}, \sigma_{X,i}\right) = \left(s \cdot m_{X,i}, s \cdot \sigma_{X,i}\right), \\ \mathrm{T_t} \cdot p_{X,i} &= \mathrm{T_t}\left(m_{X,i}, \sigma_{X,i}\right) = \left(m_{X,i} + t, \sigma_{X,i}\right). \end{aligned}$$

Once normalized, $\mu_{A_{Y \to X}}(x)$ is associated with a predetermined subset(s) of normalized probability distributions, $C_{Y,j}$'s (e.g., as shown in Figs. 22, 24 and 25 for j as "High," "Med," and "Med-Low" (or "ML"), respectively). To associate $C_{Y,j}$ with the test score value(s) (e.g., $TS_{X,n}$) assigned to $C_{X,n}$ (shown for example

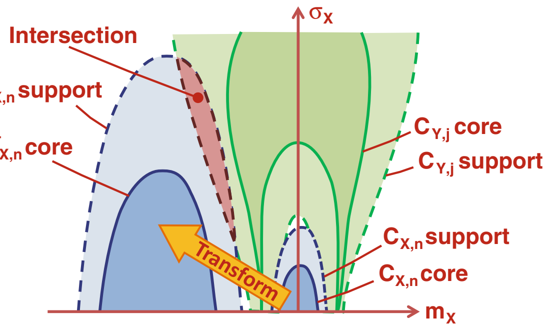


Fig. 26 Illustrating the fuzzy intersection of $C_{Y,j}$ and $C^T_{X,n}$, where $C^T_{X,n}$ is transformed from $C_{X,n}$ via scaling and translation. For the predetermined category sets $C_{Y,j}$ and $C_{X,n}$, $C_{Y,ML}$ and $C_{X,High}$ are used from Figs. 25 and 20

in Fig. 20 with n as "High"), the relative position and scaling of $C_{Y,j}$ and $C_{X,n}$ are adjusted by transforming $C_{X,n}$ to $C^T_{X,n}$ per (13), to determine the intersection between $C^T_{X,n}$ and $C_{Y,j}$, for example by:

$$I_{j,n} = \sup_{\forall p^T_{X,i}} \left(\mu_{C^T_{X,n}}\left(p^T_{X,i}\right) \wedge \mu_{C_{Y,j}}\left(p^T_{X,i}\right)\right), \tag{15}$$

where $I_{j,n}$ describes a grade for overlap between $C^T_{X,n}$ and $C_{Y,j}$. Figure 26 schematically illustrates the (fuzzy) intersection of $C^T_{X,n}$ and $C_{Y,j}$, with n being "High" and j being "ML", based on the predetermined category sets $C_{X,High}$ and $C_{Y,ML}$ from Figs. 20 and 25, respectively.

For example, as shown in Fig. 26, $C^T_{X,High}$ overlaps $C_{Y,ML}$ (to a degree), while it may not intersect $C_{Y,Med}$ (which is depicted in Fig. 24).

If $I_{j,n}$ exceeds an (optional) overlap threshold value, then we may apply the category test score TS_k associated with $C_{X,n}$, to $C_{Y,j}$. Note that the association with TS_k was determined based on B_X, e.g., through mapping of μ_{B_X} to the relation pairs ($V_{X,n}$, $TS_{X,k}$). This means that the category set of probability measures $V_{Y,j}$ associated with $C_{Y,j}$ may get associated with category test score $TS_{X,k}$, as well. In general, $V_{X,n}$ and $V_{Y,j}$ may be sets of probability measures belonging to the same family of sets (i.e., without X or Y dependencies). The steps from B_X to approximating B_Y is conceptually summarized as:

$$\left.\begin{array}{l} B_X \stackrel{map}{\longrightarrow} \left(V_{X,n}, TS_{X,k}\right) \\ \left.\begin{array}{l} B_X \rightarrow C_{X,n} \stackrel{T}{\rightarrow} C^T_{X,n} \\ A_Y \stackrel{f}{\rightarrow} A_{Y\rightarrow X} \rightarrow C_{Y,j} \end{array}\right\} \rightarrow I_{j,n} \end{array}\right\} \rightarrow \left(V_{Y,j}, TS_{X,k}\right) \stackrel{approx.}{\longrightarrow} B_Y.$$

The determination of the test scores for $V_{Y,j}$ may be implemented via a set of fuzzy rules linking $C_{X,n}$ and $C_{Y,j}$. For example, the antecedent of each rule is triggered if the corresponding $I_{j,n}$ is above an overlap threshold, and the consequent of the rule assigns $TS_{X,k}$'s (or an aggregate of $TS_{X,k}$'s based on $weight_{n,k}$ for a given n) to a variable $SC_{Y,j}$. A simpler test score assignment rule may use a non-categorical test score $TS_{X,n}$ which is determined for each $V_{X,n}$, e.g., via (12), based on the mapping through B_X:

$$Rule_{j,n} : if \left(I_{j,n}\right) then \left(\mathrm{SC}_{Y,j} \ is \ TS_{X,n}\right) \tag{16}$$

However, in correlation/aggregation of assigned (fuzzy) test scores to variable $SC_{Y,j}$, we must consider the maximization of test score required by (8). For example, in aggregating the rules for $SC_{Y,j}$, we may use α-cuts to determine an aggregated (fuzzy) result, denoted as $AGSC_{Y,j}$, as follows:

$$AGSC_{Y,j} = \underset{n}{\mathrm{MAX}}\left(Correl\left(I_{j,n}, TS_{X,n}\right)\right) \tag{17}$$

where $Correl\left(I_{j,n}, TS_n\right)$ modifies the membership function of $TS_{X,n}$ by correlating it with the factor $I_{j,n}$, e.g., via scaling or truncation. Membership function of B_Y is then approximated by a series of fuzzy relations ($V_{Y,j}$, $AGSC_{Y,j}$).

For a given w (probability measure of A_Y), $\mu_{BY}(w)$ may be approximated as a fuzzy number (or a defuzzified value), by further aggregation using fuzzy relations ($V_{Y,j}$, $AGSC_{Y,j}$), e.g.:

$$\mu_{B_Y}(w, ts) = \sup_j\left(\mu_{V_{Y,j}}(w) \wedge \mu_{AGSC_{Y,j}}(ts)\right). \tag{18}$$

2.2.2 Overlap Approximation

An approach to approximate or render the overlap (15) between the category sets, such as $C_{X,n}$, may use α-cuts to present each crisp α-cuts of predetermined category set as a set of points in (m, σ) space. These sets of points may be modeled efficiently, e.g., based on graphical models, optimized for fast transformation and intersection operations. For example, the models that use peripheral description for the α-cuts allow robust and efficient determination of intersection and avoid the need to transform all the points within the set individually, in order to reduce the computation involved in (13).

2.2.3 Estimation Using Contour Approach

In addition to predetermining $C_{X,n}$, based on $V_{X,n}$, for a normalized set A_X, we can predetermine various α-cuts of probability measures (e.g., depicted as contours of constant v in Figs. 7 and 8) or various α-cuts of associated test scores (e.g., depicted as contours of constant test scores, ts, in Figs. 9 and 10) for a set of predefined (e.g., most frequently used) B_X components. These α-cuts that represent sets of probability distribution candidates in (m,σ) space (already associated with specific test scores) may be transformed per (13) and intersected with $C_{Y,j}$ in extending their test scores to $V_{Y,j}$. In essence, this is similar to the previous analysis except $V_{X,n}$ and $TS_{X,n}$ become singleton, and $C_{X,n}$ becomes a crisp set, while $C_{Y,j}$ and $V_{Y,j}$ are predetermined (crisp or fuzzy) set.

Another approach uses (e.g., piecewise) representation of B_X (not predefined) where based on inspection or description, key values of v associated with key values of test scores may readily be ascertained (e.g., based on α-cuts), resulting in a set of (v_i, ts_i) pairs. Then, the predetermine α-cuts of probability measures (e.g., depicted as contours of constant v in Figs. 7 and 8) are used to interpolate the contours of constant ts_i's in (m, σ) space, based on the corresponding v_i values. Again, these crisp contours of constant (crisp) ts_i's, may be transformed and intersected with $C_{Y,j}$ to extend the test scores to $V_{Y,j}$ for estimating B_Y.

For quick estimation of B_Y in an alternate approach, the predetermined α-cuts (i.e., w's) of probability measures for normalized A_Y may be used (similar to those shown in Figs. 7 and 8 based on A_X), in essence, turning $V_{Y,j}$ to a singleton and $C_{Y,j}$ to a crisp set (contour) for carrying out the intersect determination. The estimates for $\mu_{BY}(w)$ may be determined via interpolation between the aggregated test score results obtained those w values associated with the α-cuts.

Acknowledgment We are very grateful to Prof. Lotfi A. Zadeh for his mentorship and encouragement and to Ms. Ixel Chavez for all her help.

References

1. Zadeh, L.A.: A note on Z-numbers. Inf. Sci. **181**, 2923–2932 (2011)
2. Yager, R.: On Z-valuations using Zadeh's Z-numbers. Int. J. Intell. Syst. **27**(3), 259–278 (2012)

Part III
Fuzzy Recognition/Classification and Image Processing

Spatiotemporal Human Brain Activities on Recalling Names of Body Parts II

T. Yamanoi, Y. Tanaka, M. Otsuki, S. Ohnishi, T. Yamazaki and M. Sugeno

Abstract The authors measured electroencephalograms (EEGs) from subjects who were looking at line drawings of body parts and recalling their names silently. The equivalent current dipole source localization (ECDL) method is applied to the event related potentials (ERPs): summed EEGs. ECDs are located in the ventral pathway. The areas are related to the integrated process of visual recognition of pictures and the retrieval of words. Some of these areas are also related to image recognition and word generation. ECDs are localized to the primary visual area V1, to the ventral pathway (ITG: Inferior Temporal Gyrus), to the parahippocampus (ParaHip), the right angular gyrus (AnG), to the right supramarginal gyrus (SMG) and to the Wernike's area. Then ECDs are localized to the Broca's area, to the post central gyrus (PstCG) and to the fusiform gyrus (FuG), and again to the

T. Yamanoi (✉) · S. Ohnishi
Faculty of Engineering, Hokkai-Gakuen University, Sapporo, Hokkaido 064-0926, Japan
e-mail: yamanoi@lst.hokkai-s-u.ac.jp

S. Ohnishi
e-mail: ohnishi@eli.hokkai-s-u.ac.jp

Y. Tanaka
Graduate School of Engineering, Hokkai-Gakuen University, Sapporo
Hokkaido 064-0926, Japan
e-mail: 6512102y@hgu.jp

M. Otsuki
Faculty of Health Sciences, Hokkaido University, Sapporo, Hokkaido 060-0808, Japan
e-mail: lasteroideb612@pop.med.hokudai.ac.jp

T. Yamazaki
Department of Bioscience and Bioinformatics, Kyushu Institute of Technoogy, Iizuka, Fukuoka 680-4, Japan
e-mail: t-ymzk@bio.kyutech.ac.jp

M. Sugeno
European Center for Soft Computing, Edificio Cientifico Technologico, 33600 Mieres-Asutrias, Spain
e-mail: michio.sugeno@softcomputing.es

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_9,

Broca's area. These areas are related to the integrated process of visual recognition of pictures and the retrieval of words. Some of these areas are also related to image recognition and word generation. And process of search and preservation in the memory is done from the result of some ECDs to the paraHip.

1 Introduction

According to research on the human brain, the primary process of visual stimulus is first processed on V1 in the occipital robe. In the early stage, a stimulus from the right visual field is processed in the left hemisphere and a stimulus from the left visual field is processed in the right hemisphere. Then the process goes to the parietal associative area [1].

Higher order processes of the brain thereafter have their laterality. For instance, 99 % of right-handed and 70 % of left-handed have their language areas in the left hemisphere, the Wernicke's area and the Broca's area [2–4].

By presenting words written in *kanji* (Chinese characters) and others written in *hiragana* (Japanese alphabet) to the subjects, researchers measured electroencephalograms (EEGs). Some of the present authors have used the same methodology as the previous research [5]. Then those stimuli and both data were summed and averaged according to the type of the stimuli and the subjects. As a result, event related potentials (ERPs) were obtained. ERPs peaks were detected and analyzed by equivalent current dipole source localization (ECDL) [6] at that latency using three dipoles model. In both the recognition of the *kanji* and *hiragana*, researchers localized equivalent current dipole (ECD) nodes from early components of ERPs to the V1, V2, and the inferior temporal gyrus (ITG). After that, ECDs were localized to the Wernicke's area and the Broca's area.

On the other hand, clinical lesion studies have shown that lesions causing disabilities of naming and comprehension of objects are dissociated depending on the target categories, e.g., artificial or biological things. These symptoms are called category-specific disorders [7].

Using the same methodology as that in the above mentioned and previous researches [5, 8–12], some of the present authors elucidated spatiotemporal human brain activities during language or image recognition.

In the present study, we measured electroencephalograms (EEGs), in order to investigate the brain activity while subjects were looking at line drawings of body parts and recalling the name of presented body parts. The data were summed and averaged according to the type of stimuli in order to obtain event related potentials (ERPs). Peak ERPs were detected and analyzed using the equivalent current dipole source localization (ECDL) method [6]. The chapter is a continuation of the previous research [12].

Fig. 1 Presented images of human body part

2 EEG Measurement Experiments

One subject was a 22-year-old female (MN) that had normal visual acuity. She was left-handed, and, from the previous experiment, her dominant language area was considered to be located in the right hemisphere. The other subject was a 22-year-old male (HT) that had also normal visual acuity. He was right-handed. The subjects put on 19 active electrodes and watched a 21-inch CRT 30 cm in front of her. Their heads were fixed on a chin rest on the table.

Each image was displayed on the CRT. Stimuli were simple monochrome images (line drawings) of parts of the human body. Images were of a foot, mouth, finger, ear, and hand (Fig. 1). First, a fixation point was presented, and then a stimulus was presented. Both of those durations were 3,000 ms. EEGs were measured on the multi-purpose portable bio-amplifier recording device (Polymate AP1524; TEAC) by means of the electrodes; the frequency band was between 1.0 and 2,000 Hz. Output was transmitted to a recording PC.

We measured the subject's EEGs on each visual stimulus. So as to effectively execute the ECDL method, EEGs were summed and averaged according to the type of human part to get event-related potentials (ERPs). To each subject, we tried the experiment twice. So as to distinguish these experiments, we labelled as HT1, HT2, MN1 and MN2 to each ERP.

According to these ERPs, the following three characteristics were found: (1) A positive peak existed around the latency of 400 ms; (2) A large negative peak existed around 450 ms; and (3) A positive peak appeared around 500 ms, attenuated gradually, and converged around 700 ms (Fig. 2).

Then the ECDL method was applied to each ERP. Because the number of recording electrodes was 19, three ECDs at most were estimated by use of the PC-based ECDL analysis software "SynaCenterPro [6]" from NEC Corporation. The goodness of fit (GOF) of ECDL was more than 99 %.

3 Results of ECDL Analysis

In these figures from Figs. 3, 4, 5 and 6, the left picture shows a sagittal view, the middle an axial view and the right a coronal view. From these three views, one can understand a location of the ECD in a three dimensional space. Localized ECDs by the ECDL method are indicated by white dots in these figures.

Fig. 2 Examples of event-rerated potentials (ERPs) by the present experiment

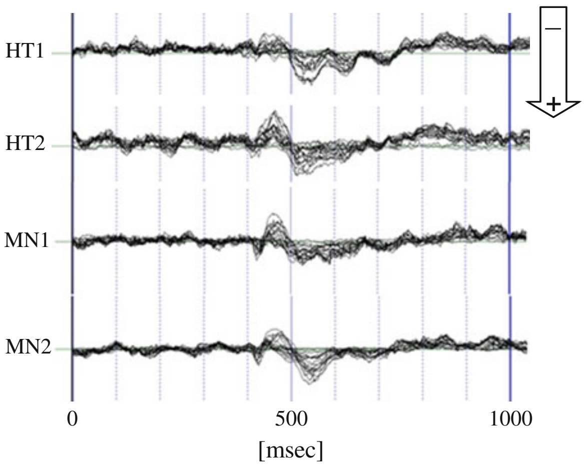
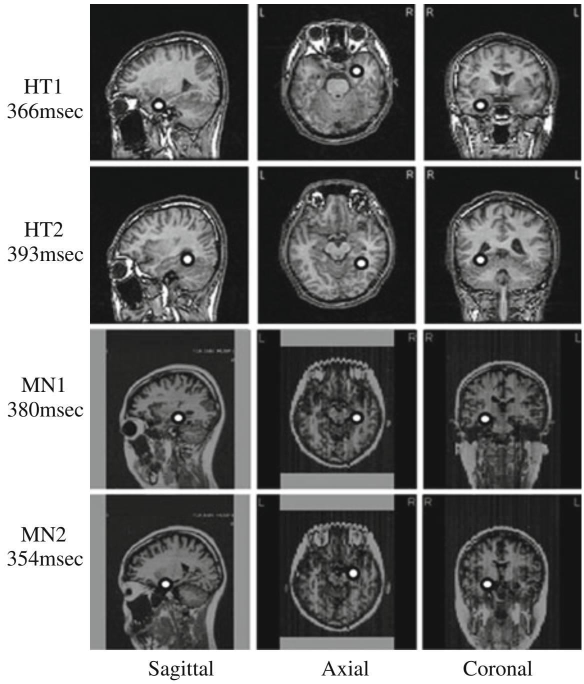


Fig. 3 ECDs localized to the *right* ParaHip



Some examples of localized ECDs are depicted in Figs. 3, 4, 5 and 6. These processes are done in series or in parallel. The relationship between ECDs and its latency is summarized in Tables 1 and 2.

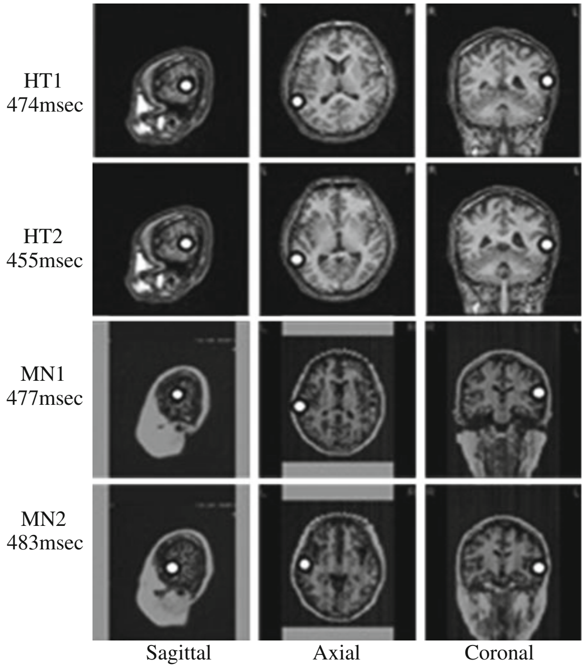


Fig. 4 ECDs localized to the Wernicke's area

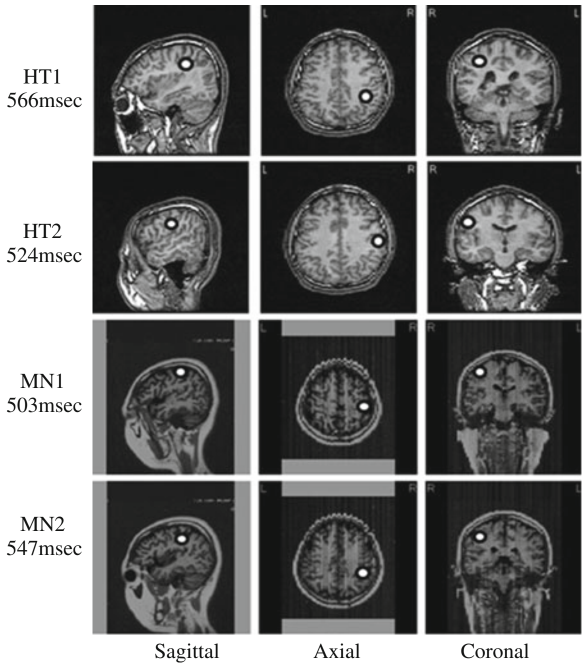


Fig. 5 ECDs localized to the post central Gyrus

Fig. 6 ECDs localized to the Broca's area

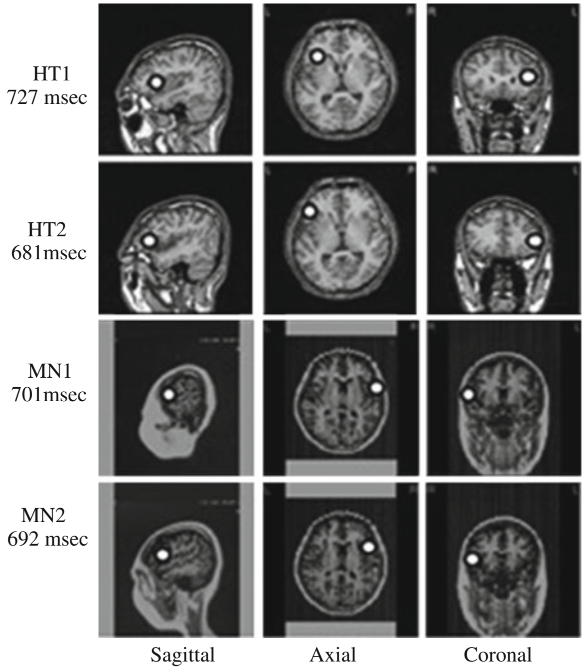


Table 1 Relationship between localized source and its latency (HT)

Subject	V1	ITG	Right ParaHip
HT1	119	326	366
HT2	131	323	393
	Right AnG	Wernicke	Right Broca
HT1	373	474	524
HT2	427	455	485
	Right ParaHip	Right PstCG	Left FuG
HT1	530	566	537
HT2	506	524	556
	Right ParaHip	Broca	
HT1	610	727	
HT2	590	681	[msec]

4 Discussion

In this study, we call the pathway among early visual recognition process and language recognition the input pathway. And we call the pathway among higher recognition and recalling the output pathway.

According to the subject HT, the input pathway was observed: V1 → right ITG → the right ParaHip → the right angular gyrus (AnG) → the Wernicke's area (Fig. 7), and the input pathway of the subject MN, was observed: V1 → left ITG →

Table 2 Relationship between localized source and its latency (MN)

Subject	V1	ITG	Right ParaHip
MN1	127	292	380
MN2	106	334	354
	Right SMG	Broca	Wernicke
MN1	443	457	477
MN2	430	455	483
	Right ParaHip	Right PstCG	Right FuG
MN1	481	503	546
MN2	494	547	548
	Right ParaHip	Right Broca	
MN1	575	701	
MN2	592	692	[msec]

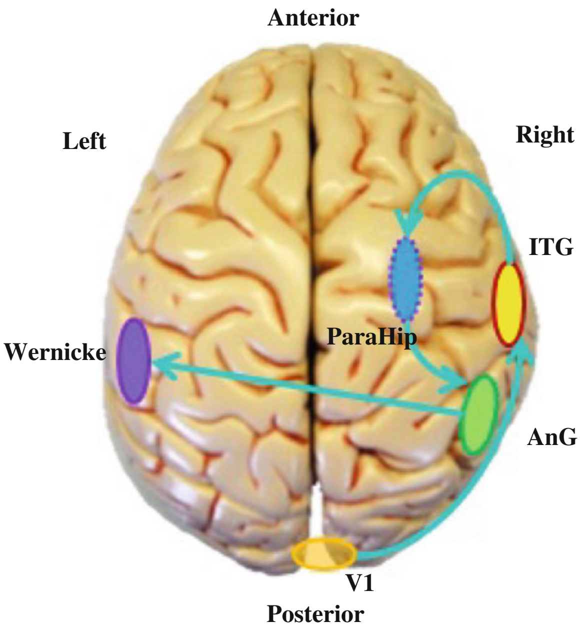


Fig. 7 Input pathway (HT): *bold line* denotes ECD on surface, *dash line* denotes ECD inside

the right ParaHip → the right supramarginal gyrus (SMG) → the Wernicke's area (Fig. 8).

The output pathway of the subject HT was observed: the right Broca's area (Broca's homologue) → the right ParaHip → the right post central gyrus (PstCG) → the left fusiform gyrus (FuG) → the right ParaHip → the Broca's area (Fig. 9). And the output pathway of the subject MN was observed: the Broca's area → the right ParaHip → the right PstCG → the right FuG → the right ParaHip → the Broca's homologue (Fig. 10). Both of output pathways included the PstCG is supposed as somatosensory area.

Fig. 8 Input pathway (MN)

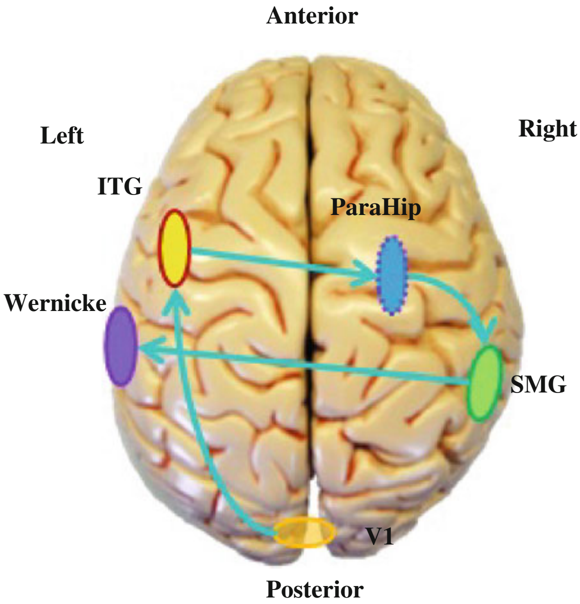
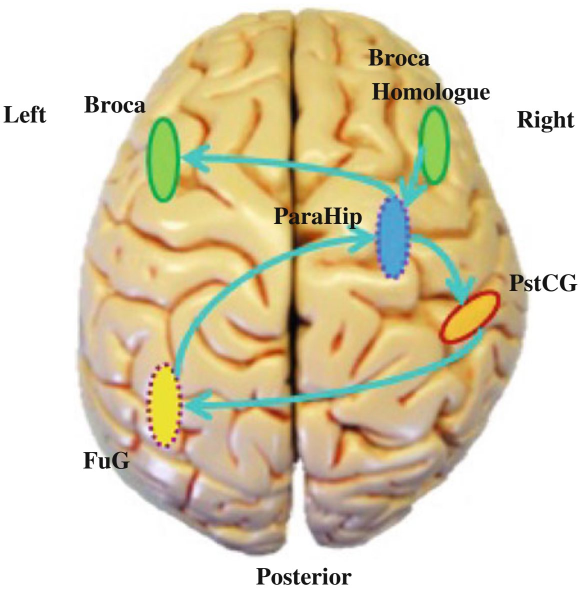


Fig. 9 Output pathway (HT)



The input pathway had been found in other studies [9, 10], and the output pathway had been found to another study [13]. These results show that the brain activities for observing static visual stimuli are related to the same pathway

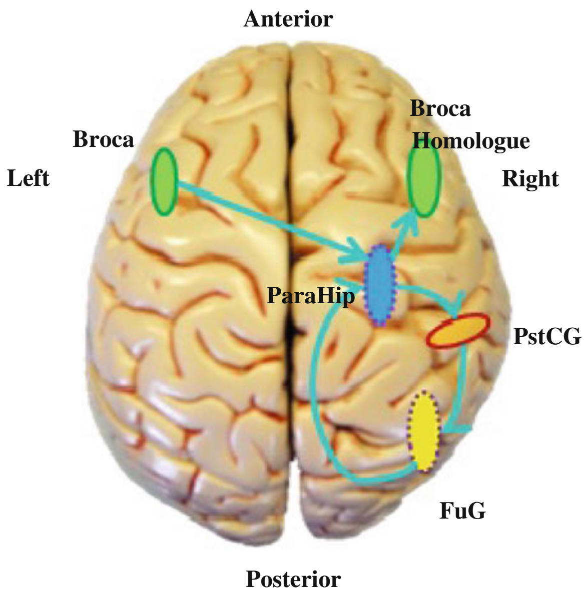


Fig. 10 Output pathway (MN)

regardless of visual stimuli (e.g. character, symbol or line drawing). And the output pathway is found to other studies [8, 13]. These results show that the brain activities on recalling process are related to the same pathway regardless of task (e.g. direction or name).

Almost the same pathways are found to the subjects HT and MN in case of recalling a name of "mouth." However, the estimated areas of the Broca and FuG are opposite between HT and MN. It is said that the dominant language area is opposite in some left-handed person, therefore, the dominant language area is supposed to be different between these two subjects.

5 Conclusion

In this study, we estimated human brain activities while human subjects who were looking at line drawings of the human body parts and recalling their names silently. ECDs were localized to the word generation area and the image recognition area.

In the previous research, we have detected a pathway regarding with the recalling of the names of body parts. By use of ECDL method, ECDs were localized to the right angular gyrus, the right fusiform gyrus and the right temporal pole. These areas are related to the integrated process of visual recognition of picture and the recalling of word. Some of these areas are also related to the image recognition and word generation. The chapter is a continuation of the previous research [12].

In case of the subject HT, estimated activities concentrate to the left hemisphere, e.g. the Broca's area and the Wernicke's area, so his language area is supposed to be the left hemisphere.

In case of the subject MN, left-handed person, although the input pathway is the same as HT, the out pathway is different from HT. It should be noted that there might be the difference of the dominant hemisphere between input and output of the language on her, or she might use both hemispheres in language process.

Because of some activities on the right ParaHip is observed, the process of search and preservation to the memory is done here. Further, we observe activities on the PstCG, which is a part of the somatosensory area, so the subjects made some somatosensory process during recalling the name of "mouth".

Acknowledgment This research was supported by a project of the High-Tech Research Center of Hokkai-Gakuen University, with the grant-in-aide from the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

References

1. McCarthy, R.A., Warrington, E.K.: Cognitive Neuropsychology: A Clinical Introduction. Academic Press, San Diego (1990)
2. Geschwind, N., Galaburda, A.M.: Cerebral Lateralization: The Genetical Theory of Natural Selection. Clarendon Press, Oxford (1987)
3. Parmer, K., Hansen, P.C., Kringelbach, M.L., Holliday, I., Barnes, G., Hillebrand, A., Singh, K.H., Cornelissen, P.L.: Visual word recognition: the first half second. NuroImage **22–4**, 1819–1825 (2004)
4. Iwata, M., Kawamura, M., Otsuki, M. et al.: Mechanisms of Writing, Neurogrammatology , pp. 179–220. IGAKU-SHOIN Ltd (2007). (in Japanese)
5. Yamanoi, T., et al.: Dominance of Recognition of Words Presented on Right or Left Eye: Comparison of Kanji and Hiragana: Modern Information Processing, from Theory to Applications, pp. 407–416. Elsevier Science B.V., Oxford (2006)
6. Yamazaki, T., et al.: PC-based multiple equivalent current dipole source localization system and its applications. Res. Adv. Biomed. Eng. **2**, 97–109 (2001)
7. Martin, A., Ungerleider, L.G., Haxby, J.V.: Category specificity and the brain: the sensory/motor model of semantic representations of objects. In: Gazzaniga, M.A. (ed.) Higher Cognitive Functions: The New Cognitive Neurosciences, pp. 1023–36. MIT Press, Cambridge (2000)
8. Sugeno, M., Yamanoi, T.: Spatiotemporal analysis of brain activity during understanding honorific expressions. J. Adv. Comput. Intell. Intell. Informatics **15**(9), 1211–1220 (2011)
9. Yamanoi, T., Toyoshima, H., Yamazaki, T., Ohnishi, S., Sugeno, M., Sanchez, E.: Micro robot control by use of electroencephalograms from right frontal area. J. Adv. Comput. Intell. Intell. Informatics **13**(2), 68–75 (2009)
10. Toyoshima, H., Yamanoi, T., Yamazaki, T., Ohnishi, S.: Spatiotemporal brain activity during hiragana word recognition task. J. Adv. Comput. Intell. Intell. Informatics **15**(3), 357–361 (2011)

11. Yamanoi, T., Toyoshima, H., Ohnishi, S., Sugeno, M., Sanchez, E.: Localization of the brain activity during Stereovision by use of dipole source localization method. In: The Forth International Symposium on Computational Intelligence and Industrial Application, pp. 108–112. (2010)
12. Yamanoi, T., Tanaka, Y., Otsuki, M., Ohnishi, S., Yamazaki, T., Sugeno, M.: Spatiotemporal human brain activities on recalling names of body parts. J. Adv. Comput. Intell. Inf. **17**(3), 386–391 (2013)
13. Hayashi, I., Toyoshima, H., Yamanoi, T: A Measure of Localization of Brain Activity for the Motion Aperture Problem Using Electroencephalogram, Developing and Applying Biologically-Inspired Vision System: Interdisciplinary Concept, Chapter9, pp. 208–223. (2012)

Novel Image Fusion Based on F-transform

Irina Perfilieva and Marek Vajgl

Abstract We propose to use the modern technique of the F-transform in order to show that it can be successfully applied to the image fusion. We remind two working algorithms (SA—the simple algorithm, and CA—the complete algorithm) which are based on the F-transform and discuss, how they can be improved. We propose a new algorithm (ESA—the enhanced simple algorithm) which is effective in time and free of frequently encountered shortcomings.

1 Introduction

In this contribution, we consider the problem of image fusion, which is one of many subjects of image processing. The image fusion aims at integration of complementary distorted multisensor, multitemporal and/or multiview scenes into one new image which contains the "best" parts of each scene. Thus, the main problem in the area of image fusion is to find the less undistorted scene for every given pixel.

This work was partially supported by the European Regional Development Fund in the IT4 Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

I. Perfilieva (✉) · M. Vajgl
Centre of Excellence IT4 Innovations, Institute for Research and Applications of Fuzzy Modeling, Division of the University of Ostrava, 30, dubna 22, 701 03 Ostrava 1, Czech Republic
e-mail: irina.perfilieva@osu.cz

M. Vajgl
e-mail: marek.vajgl@osu.cz

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_10,

A local focus measure is traditionally used for selection of an undistorted scene. The scene which maximizes the focus measure is selected. Usually, the focus measure is a measure of high frequency occurrences in the image spectrum. This measure is used when a source of distortion is connected with blurring which suppresses high frequencies in an image. In this case, it is desirable that a focus measure decreases with an increase of blurring.

There are various fusion methodologies currently in use. The methodologies differ according to different mathematical fields: statistical methods (e.g., using aggregation operators, such as the MinMax method [1]), estimation theory [2], fuzzy methods (see [3, 4]), optimization methods (e.g., neural networks, genetic algorithms [5]) and multiscale decomposition methods, which incorporate various transforms, e.g., discrete wavelet transforms (for a classification of these methods see [6]; a classification of wavelet-based image fusion methods can be found in [7], and for applications for blurred and unregistered images, refer to [8]).

In our approach we propose to use the modern technique of the F-transform and to show that it can be successfully applied to the image fusion. Our previous attempts have been reported in [9–12]. The original motivation for the F-transform (a short name for the fuzzy transform) came from fuzzy modeling [13, 14]. Similarly to traditional transforms (Fourier and wavelet), the F-transform performs a transformation of an original universe of functions into a universe of their "skeleton models" (vectors of F-transform components) in which further computation is easier. Moreover, sometimes, the F-transform can be more efficient than its counterparts. The F-transform proves to be a successful methodology with various applications: image compression and reconstruction [15, 16], edge detection [17, 18], numeric solution of differential equations [19], time-series procession [20].

The F-transform based approach to the image fusion has been proposed in [11, 12]. The main idea is a combination of (at least) two fusion operators, both are based on the F-transform. The first fusion operator is applied to F-transform components of scenes and is based on a robust partition of the scene domain. The second fusion operator is applied to the residuals of scenes with respect to inverse F-transforms with fused components and is based on a finer partition of the same domain. Although this approach is not explicitly based on focus measures, it uses the fusion operator which is able to choose an undistorted scene among available blurred. In this contribution, we analyze two methods of fusion that have been discussed in [11, 12] and propose a new method which can be characterized as a weighted combination of those two. We show that

- the new method is computationally more effective than the Complete Algorithm of fusion and has better quality than the Simple Algorithm of fusion, both have been proposed in [11, 12].

2 F-transform

Before going into the details of image fusion, we give a brief characterization of the F-transform technique applied herein (we refer to [13] for a complete description).

Generally speaking, the F-transform is a linear mapping from a set of ordinary continuous/discrete functions over domain P onto a set of discrete functions (vectors) defined on a fuzzy partition of P. We assume that the reader is familiar with the notion of *fuzzy set* and the way(s) of its representation. In this chapter, we identify fuzzy sets with their membership functions. In the below given explanation, we will speak about the F-transform of an image function u which is a discrete function $u : P \to \mathbb{R}$ of two variables, defined over the set of pixels $P = \{(i,j)\,|i = 1,\ldots,N, j = 1,\ldots,M\}$ and taking values from the set of reals $\mathbb{R}$. Throughout this text, we will always assume that M, N and u have the same meaning as above.

Let $[1,N] = \{x|\, 1 \le x \le N\}$ be an interval on the real line $\mathbb{R}$, $n \ge 2$ a number of fuzzy sets in a fuzzy partition of [1, N], and $h = \frac{N-1}{n-1}$ the distance between nodes $x_1,\ldots,x_n \in [1,N]$, where $x_1 = 1$, $x_k = x_1 + (k-1)h$, $\quad k = 1,\ldots,n$. Fuzzy sets $A_1,\ldots,A_n : [1,N] \to [0,1]$ establish a *h-uniform fuzzy partition* of [1, N] if the following requirements are fulfilled:

(i) for every $k = 1,\ldots,n$, $A_k(x) = 0$ if $x \in [1,N] \backslash [x_{k-1}, x_{k+1}]$, where $x_0 = x_1$, $x_{N+1} = x_N$;
(ii) for every $k = 1,\ldots,n$, A_k is continuous on $[x_{k-1}, x_{k+1}]$, where $x_0 = x_1$, $x_{N+1} = x_N$;
(iii) for every $i = 1,\ldots,N$, $\quad \sum_{k=1}^{n} A_k(i) = 1$;
(iv) for every $k = 1,\ldots,n$, $\quad \sum_{i=1}^{N} A_k(i) > 0$;
(v) for every $k = 2,\ldots,n-1, A_k$ is symmetrical with respect to the line $x = x_k$.

The membership functions of the respective fuzzy sets in a fuzzy partition are called *basic functions*. The example of triangular basic functions $A_1,\ldots,A_n$, $\quad n \ge 2$, on the interval [1, N] is given below.

$$A_1(x) = \begin{cases} 1 - \frac{(x-x_1)}{h}, & x \in [x_1, x_2], \\ 0, & \text{otherwise}, \end{cases}$$

$$A_k(x) = \begin{cases} \frac{|x-x_k|}{h}, & x \in [x_{k-1}, x_{k+1}], \\ 0, & \text{otherwise}, \end{cases}$$

$$A_n(x) = \begin{cases} \frac{(x-x_{n-1})}{h}, & x \in [x_{n-1}, x_n], \\ 0, & \text{otherwise}. \end{cases}$$

Let us remark that

(a) the shape (e.g., triangular or sinusoidal) of a basic function in a fuzzy partition is not predetermined and can be chosen according to additional requirements, e.g. smoothness, etc., see [13],
(b) if the shape of a basic function of a uniform fuzzy partition of [1, N] is chosen, then the basic function can be uniquely determined by the number n_p of points, which are "covered " by every "full" basic function A_k where

$$n_p = |\{i \in [1 < n] | A_k(i) > 0\}|, \quad k = 2, \ldots, n-1,$$

(In this case, we assume that $n \geq 3$).

Similarly, a uniform fuzzy partition of the interval [1, M] with $m \geq 2$ basic functions $B_1, \ldots, B_m$ can be defined. Then the fuzzy partition of $P = [1, N] \times [1, M]$ is obtained by $n \times m$ fuzzy sets $A_1 \cdot B_1, \ldots, A_n \cdot B_m$. Below, we will always assume that n, m denote quantities of fuzzy sets in fuzzy partitions of [1, N] and [1, M], respectively.

Let $u : P \to \mathbb{R}$ and fuzzy sets $A_k \cdot B_l, k = 1, \ldots, n, \ l = 1, \ldots, m$, establish a fuzzy partition of [1, N] × [1, M]. The (direct) *F-transform* of u (with respect to the chosen partition) is an image of the map $F[u] : \{A_1, \ldots, A_n\} \times \{B_1, \ldots, B_m\} \to \mathbb{R}$ defined by

$$F[u](A_k \times B_l) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} u(i,j) A_k(i) B_l(j)}{\sum_{i=1}^{N} \sum_{j=1}^{M} A_k(i) B_l(j)}, \tag{1}$$

where $k = 1, \ldots, n, \ l = 1, \ldots, m$. The value $F[u](A_k \times B_l)$ is called an *F-transform component* of u and is denoted by $F[u]_{kl}$. The components $F[u]_{kl}$ can be arranged into the matrix representation as follows:

$$\mathbf{F}_{nm}[u] = \begin{pmatrix} F[u]_{11} & \ldots & F[u]_{1m} \\ \vdots & \vdots & \vdots \\ F[u]_{n1} & \ldots & F[u]_{nm} \end{pmatrix}. \tag{2}$$

The *inverse F-transform* of u is a function on P, which is represented by the following inversion formula, where $i = 1, \ldots, N, \ j = 1, \ldots, M$:

$$u_{nm}(i,j) = \sum_{k=1}^{n} \sum_{l=1}^{m} F[u]_{kl} A_k(i) B_l(j). \tag{3}$$

It can be shown that the inverse F-transform u_{nm} approximates the original function u on the domain P. The proof can be found in [13, 14].

3 The Problem of Image Fusion

Image fusion aims at the integration of various complementary image data into a single, new image with the best possible quality. The term "quality" depends on the demands of the specific application, which is usually related to its usefulness for human visual perception, computer vision or further processing. More formally, if u is an ideal image (considered as a function of two variables) and $c_1, \ldots, c_K$ are acquired (input) images, then the relation between each c_i and u can be expressed by

$$c_i(x, y) = d_i(u(x, y)) + e_i(x, y), \quad i = 1, \ldots, K$$

where d_i is an unknown operator describing the image degradation, and e_i is an additive random noise. The problem of fusion consists in finding an image $\hat{u}$ such that it is close to u and it is better (in terms of a chosen quality) than any of $c_1, \ldots, c_K$. This problem occurs e.g., if multiple photos with focuses on different objects of the same scene are taken.

4 Image Decomposition for Image Fusion

Let us explain the mechanism of fusion with the help of the F-transform. It is based on a chosen decomposition of an image. We distinguish a one-level and a higher-level decomposition. We assume that the image u is a discrete real function $u = u(x, y)$ defined on the $N \times M$ array of pixels $P = \{(i,j) \mid i = 1, \ldots, N, \ j = 1, \ldots, M\}$ so that $u : P \to \mathbb{R}$. Moreover, let fuzzy sets $A_k \times B_l, \quad k = 1, \ldots, n, \ l = 1, \ldots, m$, where $2 \le n \le N$, $2 \le m \le M$ establish a fuzzy partition of $[1, N] \times [1, M]$.

We begin with the following representation of u on P:

$$u(x, y) = u_{nm}(x, y) + e(x, y), \tag{4}$$

$$e(x, y) = u(x, y) - u_{nm}(x, y), \tag{5}$$

where $0 < n \le N$, $0 < m \le M$, and u_{nm} is the inverse F-transform of u and e is the respective first difference. If we replace e in (4) by its inverse F-transform e_{NM} with respect to the finest partition of $[1, N] \times [1, M]$, the above representation can then be rewritten as follows:

$$u(x, y) = u_{nm}(x, y) + e_{NM}(x, y), \forall (x, y) \in P. \tag{6}$$

We call (6) a *one-level decomposition* of u on P.

If function u is smooth, then the function e_{NM} is small, and the one-level decomposition (6) is sufficient for our fusion algorithm. However, images generally contain various types of degradation that disrupt their smoothness. As a result, the function e_{NM} in (6) is not negligible, and the one-level decomposition is insufficient for our purpose. In this case, we continue with the decomposition of the first difference e in (4). We decompose e into its inverse F-transform $e_{n'm'}$ (with respect to a finer fuzzy partition of $[1,N] \times [1,M]$ with $n' : n < n' \leq N$ and $m' : m < m' \leq M$ basic functions, respectively) and the second difference e'. Thus, we obtain the *second-level decomposition* of u on P:

$$u(x,y) = u_{nm}(x,y) + e_{n'm'}(x,y) + e'(x,y),$$
$$e'(x,y) = e(x,y) - e_{n'm'}(x,y).$$

In the same manner, we can obtain a *higher-level decomposition* of u on P:

$$\begin{aligned} u(x,y) = {} & u_{n_1m_1}(x,y) + e^{(1)}_{n_2m_2}(x,y) + \cdots \\ & + e^{(k-2)}_{n_{k-1}m_{k-1}}(x,y) + e^{(k-1)}(x,y), \end{aligned} \tag{7}$$

where

$$\begin{aligned} & 0 < n_1 \leq n_2 \leq \cdots \leq n_{k-1} \leq N, \\ & 0 < m_1 \leq m_2 \leq \cdots \leq m_{k-1} \leq M, \\ & e^{(1)}(x,y) = u(x,y) - u_{n_1m_1}(x,y), \\ & e^{(i)}(x,y) = e^{(i-1)}(x,y) - e^{(i-1)}_{n_im_i}(x,y), \end{aligned}$$
$$i = 2, \ldots, k-1.$$

Below, we will be working with the two decompositions of u that are given by (6) and (7).

5 Two Algorithms for Image Fusion

In [12], we proposed two algorithms:

1. The *simple* F-transform-based fusion algorithm (SA) and
2. The *complete* F-transform-based fusion algorithm (CA).

These algorithms are based on the decompositions (6) and (7), respectively.

The principal role in fusion algorithms CA and SA is played by the *fusion operator* $\kappa : \mathbb{R}^K \to \mathbb{R}$, defined as follows:

$$\kappa(x_1, \ldots, x_K) = x_p, \quad \text{if } |x_p| = \max(|x_1|, \ldots, |x_K|). \tag{8}$$

5.1 Simple F-transform-Based Fusion Algorithm

In this section, we give a "block" description of the SA without technical details which can be found in [12] and not be repeated here. We assume that $K \geq 2$ input images $c_1, \ldots, c_K$ with various types of degradation are given. Our aim is to recognize undistorted parts in the given images and to fuse them into one image.

(1) Choose values n, m such that $2 \leq n \leq N$, $2 \leq m \leq M$ and create a fuzzy partition of $[1, N] \times [1, M]$ by fuzzy sets $A_k \cdot B_l$, $k = 1, \ldots, n$, $l = 1, \ldots, m$.
(2) Decompose input images $c_1, \ldots, c_K$ into inverse F-transforms and error functions according to the one-level decomposition (6).
(3) Apply the fusion operator (8) to the respective F-transform components of $c_1, \ldots, c_K$, and obtain the fused F-transform components of a new image.
(4) Apply the fusion operator to the to the respective F-transform components of the error functions e_i, $\quad i = 1, \ldots, K$, and obtain the fused F-transform components of a new error function.
(5) Reconstruct the fused image from the inverse F-transforms with the fused components of the new image and the fused components of the new error function.

The SA based fusion is very efficient if we can guess values n, m, that characterize a proper fuzzy partition. Usually, this is done manually according to user's skills. The dependence on fuzzy partition parameters can be considered as a main shortcoming of this otherwise effective algorithm. Two recommendations follow from our experience:

- For complex images (with many small details), higher values of n, m give better results,
- If a triangular shape of a basic function is chosen, than the generic choice of n, m is such that the corresponding values of n_p, m_p are equal to 3 (recall that n_p is a number of points, which are covered by every full basic function A_k).

In this section, the algorithm SA is illustrated on examples "Table" and "Castle", see Figs. 1 and 2. There are two inputs of the image "Table" (Fig. 3) and four ones of the image "Castle" (Fig. 4).

5.2 Complete F-transform-Based Fusion Algorithm

The CA based fusion does not depend on one choice of fuzzy partition parameters (as in the case of the SA), because it runs through a sequence of increasing values n, m. The description of the CA is similar to that of the SA except for the step 4 which is repeated in a cycle. Therefore, the quality of fusion is high, but the implementation of the CA is rather slow and memory consuming, especially for large images. For the illustration, see Figs. 5 and 6.

Fig. 1 The SA fusion of the image "Table", approx. run time: 1, 6 s

Fig. 2 The SA fusion of the image "Castle", approx. run time: 1, 9 s

5.3 Fusion Artefacts

In this section, we characterize input images, for which it is reasonable to apply the SA or CA. By doing this, we put restrictions on inputs which are acceptable by the algorithms SA and CA. First of all, input images should be taken without shifting or rotation. Secondly, blurred parts of input images should not contain many small details like leaves on trees, etc. If it is so, then the fusion made by SA or CA can leave "artefacts", like "ghosts" or "lakes", see the explanation below where we assume that there are two input images for the fusion.

- **Ghosts**—this happens when a sharp edge of a non-damaged input image is significantly blurred in the other one. As a result of the SA or CA, the edge is perfectly reconstructed, but its neighboring area is affected by the edge presence (see Fig. 7).
- **Lakes**—this may happen in both cases when the fusion is performed by the SA or CA. In the case of SA, a "lake" is a result of choosing neighboring areas with significantly different colors from different input images. In the case of SA, a "lake" is a result of rounding off numbers (see Fig. 8).

Fig. 3 Two inputs of the image "Table". The toy is *blurred* in the *left image,* and vice versa, it is the only sharp part in the *right* one

Fig. 4 Four inputs (sharp zones vary from the north-west to the south-east quatre) of the image "Castle"

Fig. 5 Table—the CA fusion of the image "Table", two input images, approx. run time: 111 s

Fig. 6 Table—the CA fusion of the image "Castle", four input images, approx. run time: 359 s

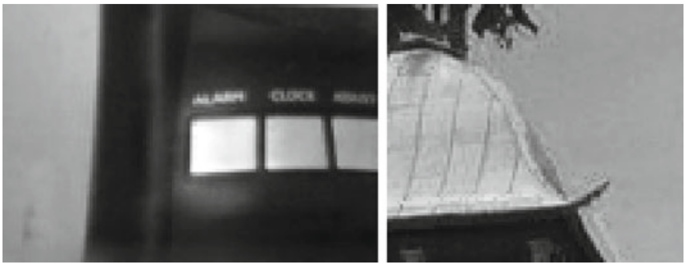


Fig. 7 Example of "ghost"—the *white area* around the *left* button (*left image*) and the doubled edge of the roof (*right image*)

Fig. 8 Example of "lake"—the *color* of this area is significantly different from the respective *colors* of input images

6 Improved F-transform Fusion

The main purpose of this contribution is to create a method which will be as fast as the SA and as efficient as the CA. The following particular goals should be achieved:

- Avoid running through a long sequence of possible partitions (as in the case of CA);
- Automatically adjust parameters of a fusion algorithm according to a level of blurring and a location of a blurred area in input images;
- Eliminate situation which can lead to "ghosts" and "lakes" in a fused image.

6.1 Proposed Solution

The main idea of the improved F-transform fusion is to enhance the SA by adding another run of the F-transform over the first difference (4). Our explanation is as follows: the first run of the F-transform is aimed at edge detection in each input image, while the second run propagates only sharp edges (and their local areas) to the fused image. The informal description of the *enhanced simple algorithm* (ESA) is given below.

for all input images **do**

Compute the *inverse F-transform*

Compute the *first absolute difference* between the original image and the inverse F-transform of it

Compute the *second absolute difference* between the first one and its inverse F-transform and set them as weights of pixels

end for

for all pixels in an image **do**

Compute the value of *sow*—the sum of weights over all input images

Fig. 9 Table—The ESA fusion of the image "Table", two input images, approx. run time: 10 s

for all input images **do**

Compute the value of *wr*—the ratio between the weight of a current pixel and *sow*

end for

Compute the fused value of a pixel in the resulting image as a weighted (by *wr*) sum of input image values

end for

Although the algorithm ESA is written for gray scale input images, there is an easy way how to extend it to color images which are represented in RGB or YUV models. Our tests were performed for both of them. In the case of RGB, the respective R, G, or B channels were processed independently and then combined. In the second case of YUV, the Y-part of the model was used to achieve weights (this part contains the most relevant information about the image intensity), while the U-part and the V-part were processed with the obtained weights.

Let us remark that the ESA-fused images are (in general) better than each of the SA or CA. It can be visually seen on the chosen above examples in Figs. 9 and 10. The main advantages of the ESA are:

- **Time**—the executing time is smaller than in the case of the CA (in the examples above it is as follows: 11 versus 111 ("Table"), 18 versus 359 ("Castle"). The quality of the ESA fusion is better than that of the SA. Examples of run times and memory consumption are presented in Table 1 (notice that the memory consumption significantly depends on memory management of implementation environment).
- **Ghosts effect is reduced**—the "ghost" effects (they are seen around the tower roof in the image "Castle", and around the buttons and the clock in the image "Table") are removed as it can be seen on 11 (Fig. 11).
- **Lakes effect is eliminated**—the "lakes" are almost eliminated as it can be seen from Figs. 8, 9 and 12.

Fig. 10 Table—The ESA fusion of the image “Castle”, four input images, approx. run time: 18 s

Table 1 Basic characteristics of the three algorithms applied to the tested images

Image set	Resolution	Time (s)			Memory (MB)		
		CA	SA	ESA	CA	SA	ESA
Castle	1120 × 840	359	1.9	19.0	160	35	102
Table	852 × 639	111	1.6	11.0	95	3	38
Balls	1600 × 1200	340	1.2	36	270	58	152

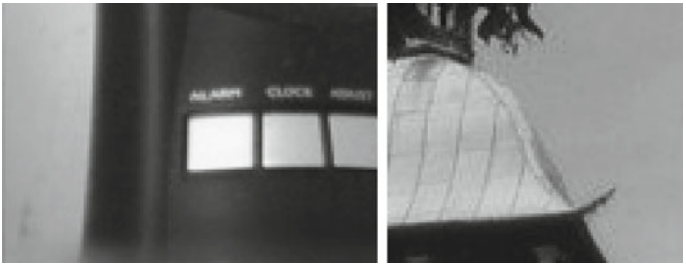


Fig. 11 Example of “Ghost” artefact reduction by ESA

6.2 Comparison Between Three Algorithms

In this section, we show that in general, the ESA fusion has better execution parameters than the SA or CA fusion. We experimented with numerous images which due to the space limitation, cannot be presented in this chapter. An exception is made for one image “Balls” with geometric figures to show how the fusion methods reconstruct edges. In Fig. 13, two inputs of the image “Balls” are given, and in Fig. 14, three fusions of the same image are demonstrated.

Fig. 12 Example of “Lakes” artefact reduction by ESA

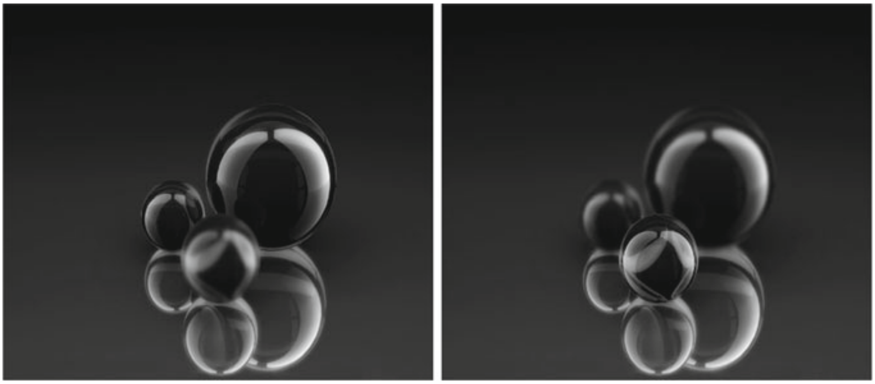


Fig. 13 Two inputs for the image “Balls”. The central ball is *blurred* in the *left image*, and vice versa, it is the only sharp ball in the *right* one

In Table 1, we demonstrate that the complexity (measured by the execution time or by the used memory) of the newly proposed ESA fusion is greater than the complexity of the SA and less than the complexity of the CA.

In Table 2, we demonstrate that the quality of fusion (measured by the values of MSE and PSNR) of the newly proposed ESA fusion is better (the MSE value is smaller) than the quality of the SA and in some cases (the image “Balls”), is better than the quality of the CA. Table 2 does not contain the values of MSE and PSNR for the image “Table”, because (as it happens in reality) there was no original (non-distorted) image at disposal.

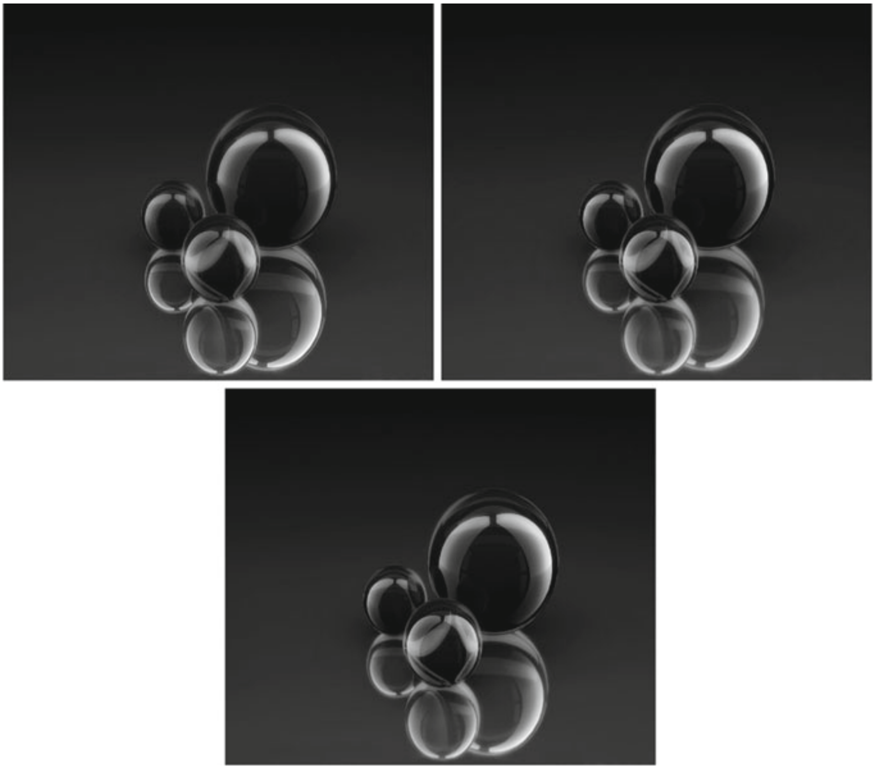


Fig. 14 The SA (*left*), CA (*right*) and ESA (*bottom*) fusions of the image "Balls". The best quality has the ESA fusion (cf. Table 2)

Table 2 MSE and PSNR characteristics of the three fusion methods applied to the tested images

Image set	MSE			PSNR		
	CA	SA	ESA	CA	SA	ESA
Castle	9.48	42.48	14.15	40.62	37.51	40.61
Balls	1.28	6.03	0.86	48.91	43.81	52.57

7 Conclusion

In this chapter, we continued our research started in [9–12] on effective fusion algorithms. We proposed the improved method of the F-transform based fusion which is free from following imperfections: long running time, dependence on initial parameters which characterize a proper fuzzy partition, presence of fusion artefacts, like "ghosts" or "lakes".

References

1. Blum, R.S.: Robust image fusion using a statistical signal processing approach. Inf. Fusion **6**(2), 119–128 (2005)
2. Loza, A., Bull, D., Canagarajah, N., Achim, A.: Non-gaussian model-based fusion of noisy images in the wavelet domain. Comput. Vis. Image Underst. **114**(1), 54–65 (2010)
3. Singh, H., Raj, J., Kaur, G., Meitzler, T.: Image fusion using fuzzy logic and applications. In: Proceedings of IEEE International Conference on Fuzzy Systems, vol. 1, pp. 337–340 (2004)
4. Ranjan, R., Singh, H., Meitzler, T., Gerhart, G.: Iterative image fusion technique using fuzzy and neuro fuzzy logic and applications. Fuzzy Information Processing Society. NAFIPS 2005. Annual Meeting of the North American, pp. 706–710 (2005)
5. Mumtaz, A., Majid, A.: Genetic algorithms and its application to image fusion. In: 4th International Conference on Emerging Technologies ICET 2008, pp. 6–10 (2008)
6. Piella, G.: A general framework for multiresolution image fusion: from pixels to regions. Inf. Fusion **4**(4), 259–280 (2003)
7. Amolins, K., Zhang, Y., Dare, P.: Wavelet based image fusion techniques an introduction, review and comparison. ISPRS J. Photogramm. Rem. Sens. **62**(4), 249–263 (2007)
8. Šroubek, F., Flusser, J.: Fusion of blurred images. In Blum, L.Z.R. (ed.), Multi-Sensor Image Fusion and Its Applications, Signal Processing and Communications Series, CRC Press, San Francisco (2005)
9. Daňková, M., Valášek, R.: Full fuzzy transform and the problem of image fusion. J. Elect. Eng. **12,** 82–84 (2006)
10. Perfilieva, I., Daňková, M.: Image fusion on the basis of fuzzy transforms. Proceedings of 8th International FLINS Conference, pp. 471–476, Madrid (2008)
11. Perfilieva, I., Daňková, M., Hodáková, P., Vajgl, M.: The use of F-transform for image fusion algorithms. In: Proceedings of International Conference of Soft Computing and Pattern Recognition (SoCPaR 2010), pp. 472–477
12. Hodáková, P. Perfilieva, I., Daňková, M., Vajgl, M.: F-transform based image fusion, In: Ukimura, O. (ed.) Image Fusion, InTech, pp. 3–22 (2011) http://www.intechopen.com/articles/show/title/f-transform-based-image-fusion
13. Perfilieva, I.: Fuzzy transforms: theory and applications. Fuzzy Sets Syst. **157**, 993–1023 (2006)
14. Perfilieva, I.: Fuzzy transforms: A challenge to conventional transforms. In P. W. Hawkes (ed.), Advances in Images and Electron Physics, vol. 147, Elsevier Academic Press, pp. 137–196, San Diego, 2007
15. Perfilieva, I., Pavliska, V., Vajgl, M., De Baets, B.: Advanced image compression on the basis of fuzzy transforms. In: Proceedings Conference on IPMU'2008, pp. 1167–1174, Torremolinos (Malaga) (2008)
16. Di Martino, F., Loia, V., Perfilieva, I., Sessa, S.: An image coding/decoding method based on direct and inverse fuzzy transforms. Int. J. Appr. Reasoning **48**, 110–131 (2008)
17. Perfiljeva, I., Daňková, M., Hodáková, P., Vajgl, M.: Edge detection using F-transform. In: Proceedings of ISDA 2011, pp. 672–677, Cordoba (2011)
18. Perfiljeva, I., Hodáková, P., Hurtík, P.: F1-transform edge detector inspired by canny's algorithm. In: Communications in Computer and Information Science. Advances on Computational Intelligence, pp. 230–239. Springer, Heidelberg, Catania (2012)
19. Štěpnička, M., Valášek, R.: Numerical solution of partial differential equations with help of fuzzy transform. In: Proceedings of the 2005 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'05), pp. 1153–1162. Reno (2005)
20. Perfilieva, I., Novák, V., Pavliska, V., Dvořák, A., Štěpnička, M.: Analysis and prediction of time series using fuzzy transform. In: Proceedings of IEEE International Conference on Neural Networks WCCI 2008, pp. 3875–3879. Hong Kong (2008)

Improving the Accuracy of a Fuzzy-Based Single-Stroke Character Recognizer by Antecedent Weighting

A. Tormási and L. T. Kóczy

Abstract In this chapter we present an improved version of the fuzzy based single-stroke character recognizer introduced in previous works. The modified recognition method is able to reach higher accuracy in the character recognition without any significant effect on the computational complexity of the algorithm. Different fuzzy rule and antecedent weighting techniques were successfully used to improve the efficiency of fuzzy systems especially in classification problems. The altered recognizer reached 99.49 % average recognition rate with 26 different single-stroke symbols (based on Palm's Graffiti alphabet) without learning user-specific parameters or modifying the rule-base. The new algorithm has the same computational complexity as the original system does.

1 Introduction

There are many problems of usability and ergonomics are posed by the current virtual keyboards. These are using most of the display reducing the maximal size of the content at the expense of the data context which again reduces the user's performance during the workflow.

This chapter was supported by the National Scientific Research Fund Grant OTKA K75711 and OTKA K105529, a Széchenyi István University Main Research Direction Grant and EU grant TÁMOP 421 B, TÁMOP 4.2.2/B-10/1-2010-0010.

A. Tormási (✉) · L. T. Kóczy
Department of Automation, Széchenyi István University, Győr 9026, Hungary
e-mail: tormasi@sze.hu

L. T. Kóczy
e-mail: koczy@sze.hu

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_11,

The best alternative for a keyboard replacement could be handwriting systems. Processing written text by computers nevertheless has a long history. In this field there are still many ongoing research and development projects aiming to achieve more accurate recognition of handwriting. In her study LaLomia determined 97 % as the general user acceptance rate for handwriting recognizers [1].

It is important for industrial users and end-users to improve the recognition rate of handwriting recognizers with the smallest possible increase of the computational complexity even if the growth of hardware performance is faster than the growth of resource requirement of the recognition method.

In this chapter we present an improved version of a fuzzy based single-stroke character recognizer (FUBAR). The modified system has reached 99.49 % average recognition rate with 26 different single-stroke symbols based on Palm's Graffiti alphabet. The accuracy of the new method is close to or even slightly beyond the results of other commercial and academic recognizers. Beside the high recognition rate the developed method has lower computational complexity than the investigated algorithms which makes FUBAR a notable competitor of other character recognition systems.

This chapter consists of five sections: After the Introduction in Sect. 2 the basic concept of our original character recognizer is presented. In Sect. 3 the steps of weight calculation and the modifications made on the original recognition method are proposed. The test results and comparisons with other known recognition systems are introduced in Sect. 4. In the last section observations are summarized, future works and other possible applications of the system approach are discussed.

2 The Concept of the Original Recognition Method

During the design of the original recognition algorithm [2] besides on acceptable recognition accuracy three main goals were targeted. The first objective was to limit the resources needed for the method as a basic requirement for the use of the system in portable devices. A solution has been worked out to eliminate geometrical transformations from the method so we could reduce the overall computational complexity. A simple and fast exchangeability of the symbol set was our second goal for a better user friendliness. The last objective was to design an easily modifiable symbol-base of writing for the adaptation step which makes our recognizer able to learn the user specific style. In the early stage of the development we decided to handle the stroke segmentation as a separate problem so we could focus on the concept of the recognition engine.

The basic concept of FUzzy BAsed Recognizer is shown in Fig. 1.

The system uses a modified Palm Graffiti alphabet (strokes for the letter "G" and "U" are changed) to make the single-strokes more similar to the Hungarian style of capital letters. The two alternation introduced can be seen in Fig. 2.

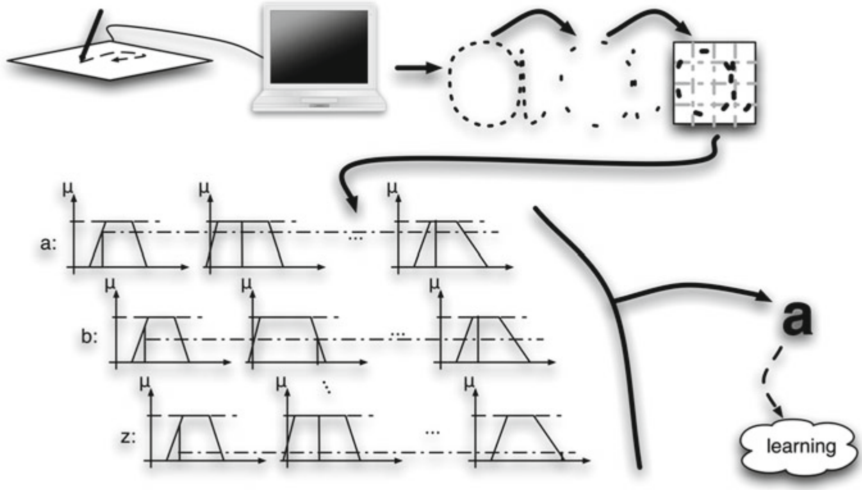
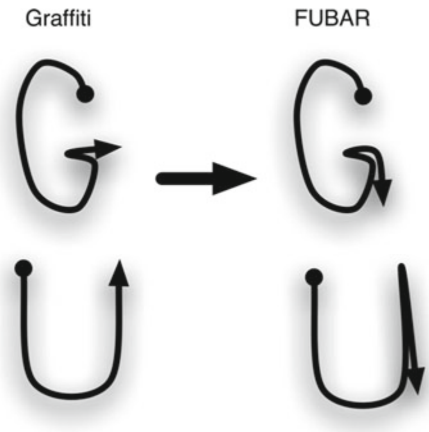


Fig. 1 Model of FUBAR

Fig. 2 Symbol "G" and "U" in Graffiti and FUBAR



2.1 The Input Signal

Each individual stroke could be represented as a three-dimensional continuous function sampled by the digitizer tablet. The system collects all the coordinates in chronological order representing the digital ink.

Due to hardware bottlenecks the input device is not capable of collecting all parts of the signal and the distance between the sampled points may differ. The range between sampling points depends on the writing speed and the available hardware resources.

The varying distance between points renders more difficulty in recognizing the stroke by the average density of points in the extraction grid. To solve this problem re-sampling is unavoidable and it is done in the following (Pre-Processing) subsection.

2.2 Pre-processing

This step fixes the density of points and it also works as anti-aliasing making the strokes more readable.

The first and last points of the stroke are kept for reference. After that a filtering algorithm calculates the distance between the last added (or first) points and the following ones. If the distance reaches the minimum threshold then the point will be added to the re-sampled stroke.

To calculate the distance between the points the method uses Euclidean distance. Equation (1) represents the filtering algorithm.

$$\underline{l'} = \{l_1\} \cup \left\{ l_j \middle| l \underset{\substack{j \in \{N_{n-1} - N_1\} \\ \mathrm{j} > \mathrm{k}}}{\arg\min} \left[\left|d\left(l'_{k-1}, l_j\right) - \gamma\right|\right], k = \underset{\substack{p \\ i \in \{N_{n-1} - N_1\}}}{\arg\min} \left[\left|d\left(l'_{i-1}, l_p\right) - \gamma\right|\right], i = \dim \underline{l'} \right\} \cup \{l_n\}, \quad (1)$$

where l is the list of collected points, l' is the list of re-sampled points and γ represents the minimum distance between points and *argmin* gives a minimal position a at which f is minimized.

2.3 Feature Extraction

Input signals are identified by the width/height ratio of the stroke and the average number of stroke-points in the rows and columns of the *fuzzy grid* drawn around the input stroke.

The first system used crisp grids (with sharp borders) but tests pointed out that if the angular offset of the input stroke and the etalon symbol were different then distribution of the points in the grid would also differ as shown in Fig. 3. This might cause a considerable reduction in recognition rates.

As a solution we designed a grid with blurred boundaries which will be referred to as fuzzy grid (each grid row and column defined as a fuzzy set [3] constituting Ruspinian-partitions [4]). In the case where a point is located close to a boundary it will be counted as a member of both columns and rows with different membership degrees calculated by the exact location as shown in Fig. 4.

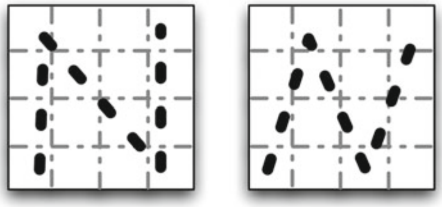


Fig. 3 Normal and oblique "N" symbol in a grid

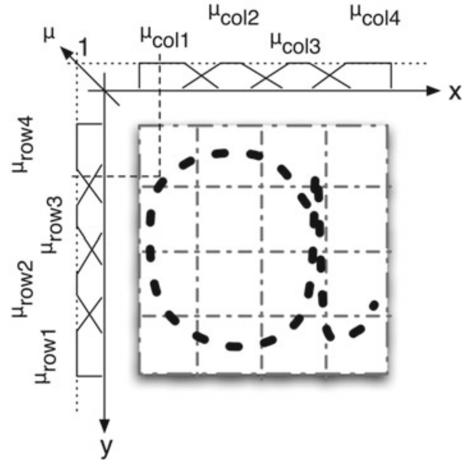


Fig. 4 A symbol in a 4-by-4 fuzzy grid

Point distributions in the fuzzy grid are calculated by the following two formulas:

$$c_i = \frac{\sum_{j=1}^{\dim l'} \left(\mu_{col_i}(x_j)\right)}{\dim l'}, \tag{2}$$

$$r_i = \frac{\sum_{j=1}^{\dim l'} \left(\mu_{row_i}(y_j)\right)}{\dim l'}, \tag{3}$$

where c_i represents the point distribution in the column i and r_i represent the same for the rows, l' is a list of the filtered stroke-points, x_j and y_j are the x- and y-coordinates of element j.

2.4 The Inference

Each symbol in the set is represented by a single fuzzy rule [5, 6] and an N_{sample} sized FIFO queue which stores symbol samples previously written by the user and used only during the adaptation step.

Previously collected stroke-features are used as input parameters for the fuzzy rules. Each rule is evaluated with the features of the current input stroke. The number of rules is equivalent to the number of the symbols in the base set.

Between the parameters of the rule we use min t-norms as AND operators. The consequent part of the rules represents the degree of matching the parameters of the input stroke and the parameters of the symbol represented by the given rules. For the inference we use the Takagi and Sugeno method [6] and the best fitting rule (with the highest rule match) will be chosen as the output of the inference.

2.5 The Adaptation Phase

After evaluating the inference the next phase is the supervised adaptation. The user has to set the target symbol which he/she wanted to enter via the recognition interface. Using the features of the input stroke and the previously stored samples the system tunes the parameters of the fuzzy sets in the rule base.

During adaptation the system must consider the features of stored samples as much as the new input stroke itself. All parameters of the new symbol must fit to the tuned fuzzy set of the target symbol as much as it is possible without decreasing the fitness of the stored samples. At the same time the method has to minimize the overlap of the target and non-target fuzzy rules without modifying the fitness of the samples stored in the non-target symbols.

To reduce the computational complexity of the system it has been decided to use an evolutionary method [7] for the adaptation process. Evolutionary algorithms are able to provide acceptable (sub)optimal solutions for many problems in a short time with smaller resource requirement than exact optimum search. In the developed algorithm classic evolutionary solutions cannot be used due to the special constraints for the different dynamic fitness functions that have to be applied for the symbols at the same process. Without these constraints the fuzzy sets would overlap in the different rules which would decrease the recognition rate.

As a solution to the overlap problem we extended the bacterial evolutionary algorithm [8] with “punish” and “reward” options. The method rewards the target symbol and punishes all other, non-target symbols by using different fitness functions containing the special constraints. The reward fitness function maximizes the recognition rate for the input symbol and the user-samples stored by the target symbol while minimizes the recognition rate for the non-target symbol samples. The punish fitness function maximizes the recognition rate for the stored user-samples in the current symbol and minimizes the recognition rate for the input symbol and for the stored user-samples from other symbols.

The algorithm consists of the following four phases:

2.5.1 Converting Symbols Into Bacteria

Each membership functions of rules are converted into four numbers per function (and nine functions per rule) and stored in an individual vector which represents the chromosomes of a given bacterium.

2.5.2 Creating Colonies and Initial Populations

At the start of this step each symbol is represented by one bacterium which represents the rule of the symbol. Each colony is created by an individual bacterium as presented in Fig. 5.

The bacterium in the colonies are replicated $N_{replicant}$ times. A colony with the original bacterium and the replicates are representing the initial population.

Previously collected symbol-samples are also stored in the colonies for further use.

2.5.3 Bacterial Mutation (Local Optimization)

Bacterial mutation is an operator used on each bacterium in a population separately. A bacterium (selected from the population) is cloned N_{clone} times. Next step of this phase is selecting randomly a chromosome and replacing it in the clones except in the original bacterium with a random value (mutation).

After modifying a chromosome we have to check the validity of the bacterium by inspecting the chromosomes. Points of a trapezoidal membership function represented by four chromosomes must be ordered by value as presented on Fig. 6.

If the chromosomes are not in a good order then the algorithm fixes it. The algorithm orders the original bacterium and all its clones by fitness value and chooses the best one which replaces the selected chromosome in the others.

This procedure (cloning—mutation—test—evaluation—replacement) is repeated until all the chromosomes have not been chosen. After selecting all the chromosomes the algorithm puts the best bacterium back into the population and the others will be deleted. As a result the new bacterium will have a better or at least the same fitness value (in case of unavailing mutation).

These steps are repeated in populations until all the bacteria have not been chosen as shown in Fig. 7.

2.5.4 Gene Transfer (Global Optimization)

During this phase the gene transfer operator is used on all the populations (with mutated bacterium) separately. The method chooses a population ordered by fitness value and divides it into two pieces, the group of the good and the group of bad bacteria.

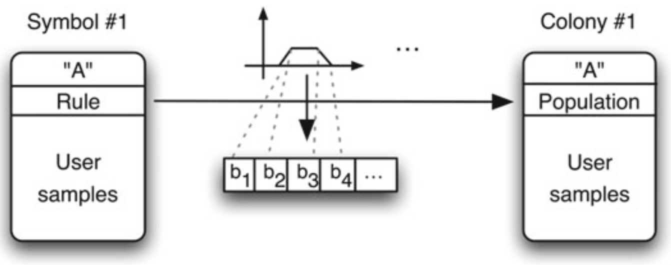


Fig. 5 Creation of a population from a symbol

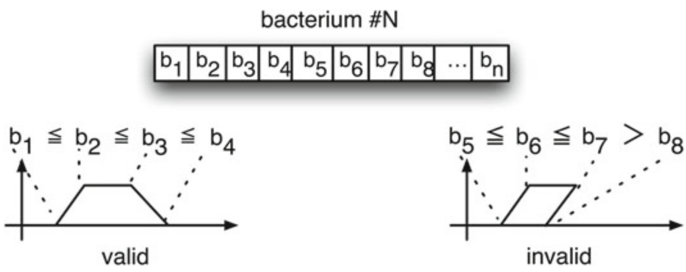


Fig. 6 Validity test of a bacterium

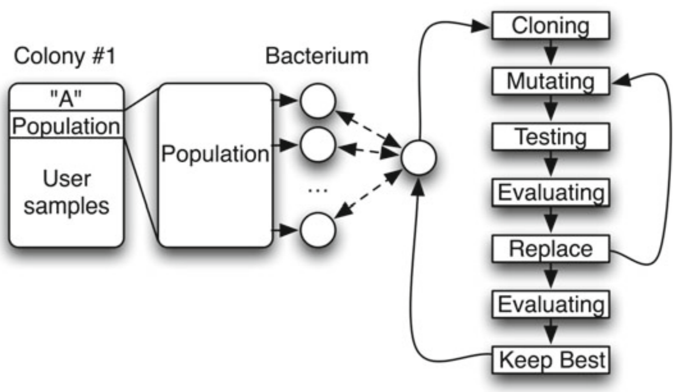


Fig. 7 Process of bacterial mutation

One bacterium is chosen from the good ones (let us call it "source bacterium") and another one from the bad ones (let us call it "target bacterium"). The source bacterium replaces a randomly chosen chromosome in the target bacterium (gene transfer) in the third step.

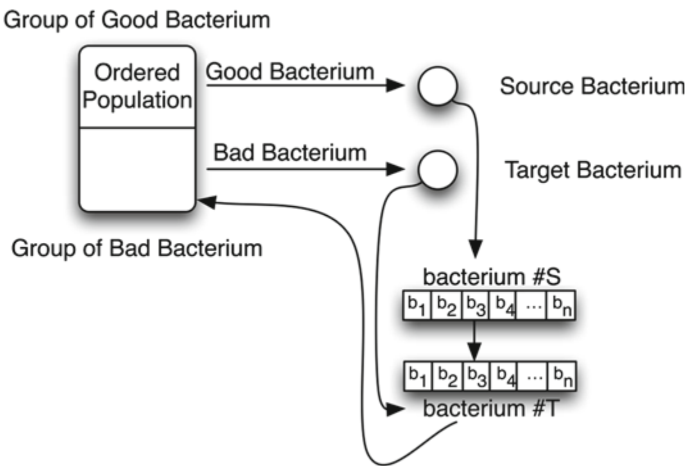


Fig. 8 Process of infection

As described in Phase 3, after the modification of a chromosome the membership functions represented by the bacterium must be validated (and adjusted if it is necessary). The algorithm repeats this phase N_{inf} times which stands for the maximal number of infections. After using gene transfer on all the populations we have to increase N_{gen} which indicates the number of the current generation. If $N_{gen} < N_{max\text{-}gen}$ (maximal number of generations) then repeat the algorithm from Phase 3 else choose the best bacterium from all the colonies, convert them back to symbols and replace the original symbols with the new ones as presented in Figs. 8 and 9.

3 Weighting of the Rule Input Parameters

There are numerous works on different fuzzy rule and antecedent weighting methods and applications with promising results especially in the classification problems [9–11]. The results of the analysis of the different rule parameter values for the test sample set showed that the different input parameters have different value of reliability during the rule-evaluation. The results introduced in the previously mentioned chapters showed that the accuracy of the designed recognition method could be increased by the weighting of the different input-parameters for each rule according to their reliability value.

The recognition method with a 6-by-6 fuzzy grid recognizes 9,942 from 10,000 input symbols (99.42 %) [12]. The same system with a 6-by-4 fuzzy grid recognizes 11,727 letters from 11,818 (99.23 %) for the same computational price [13]. Considering the previous researches, the computational complexity and the accuracy of the method, a recognizer with a 6-by-4 fuzzy grid have been chosen as the basic system for the improvement.

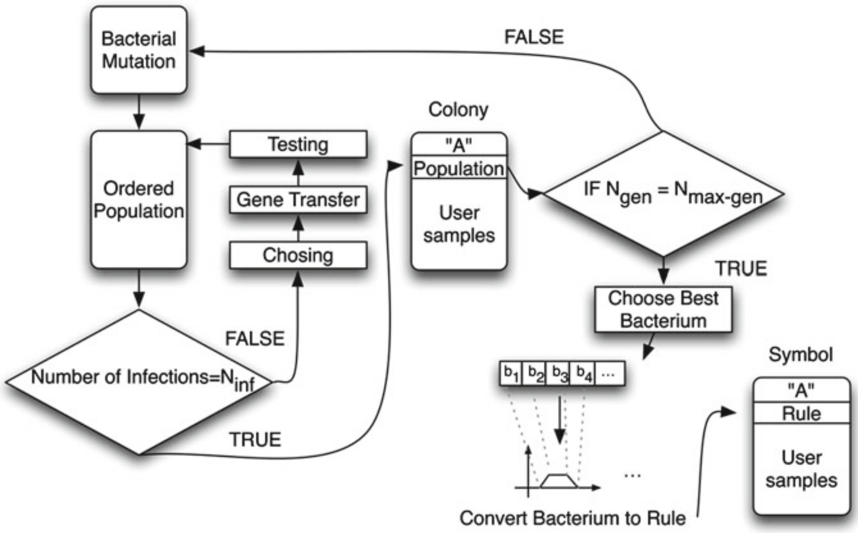


Fig. 9 Process of gene transfer

Before the calculation of the rule input parameter weights the reliability value must be calculated first. The original symbol sample set has been used to determine the reliability parameters, same sample set have been used during the preparation of the rule-base. 180 samples per symbol have been collected from 12 different students of the Széchenyi István University which has been divided into three sample groups: Set 1 to determine the rule-base and the input-parameter weights; Set 2 for the training of the rule-base by learning user-specific parameters; Set 3 for the accuracy tests. The adaptation phase was disabled during the tests of weighted rule input-parameters due to its possible influence to the recognition rate.

After the evaluation of the rule-base the differences have been calculated between the results of input-parameters of the recognized symbol and the missed target-symbol in the cases of attempted recognitions with wrong results. The analyses of the parameter-wise frequency of causing wrong results and the maximum values of the differences have been used to define the reliability rank of the rule input-parameters.

In the next step weights have been calculated for each dimension of symbol-rules. The algorithm considers the parameter values of a given dimension according to the weight associated to it. If the weight is *1* then the algorithm will use the parameter like the original method. If the weight is *0* then the parameter will be completely ignored and it will have no influence to the result.

The affects to the recognition rates of weighting the input-parameters of rules with the minimal parameter-reliability values which are lower than an α threshold have been analyzed. In some cases the weights have had influence only on the weighted symbol as presented in Fig. 10.

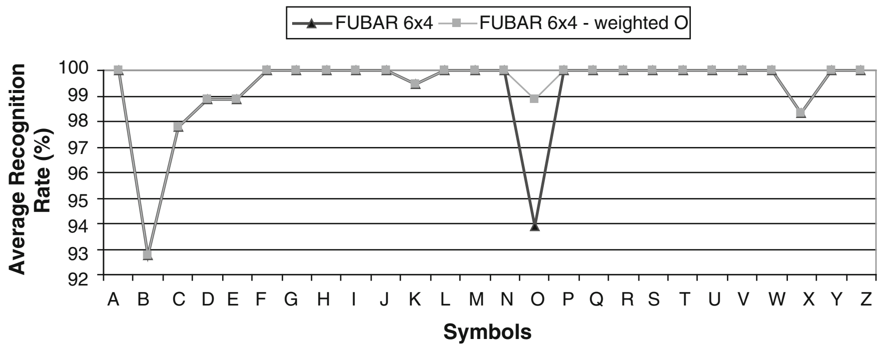


Fig. 10 Average recognition rates of FUBAR with a 6-by-4 fuzzy grid and weighted "O"

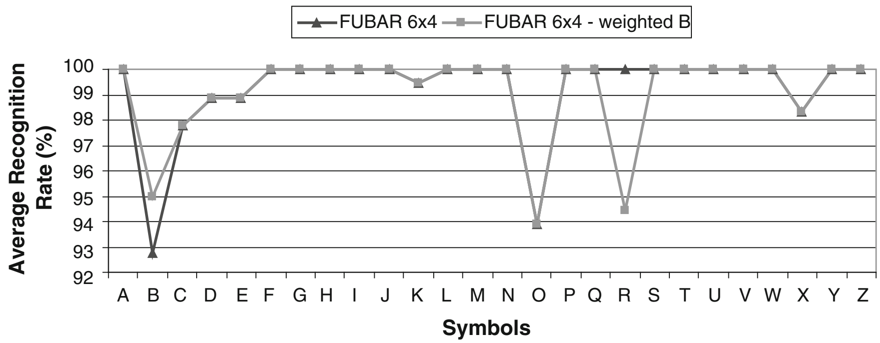


Fig. 11 Average recognition rates of FUBAR with a 6-by-4 fuzzy grid and weighted "B"

But in some cases the decrease of recognition rate for the non-weighted symbols was greater than the increase in the accuracy of the weighted symbols as shown in Fig. 11. When the reached improvement was less than the caused deterioration then the weighting of the input parameter of the given symbol was rejected.

If the increase of the recognition rate was not sufficient then the parameter with the second smallest reliability-value has been weighted for a given symbol. These steps have been applied to all affected input parameter-values of symbols.

In the next phase the joint impact on the accuracy has been analyzed for the different parameter weightings of symbols. The combinations with a lower increase than decrease in the recognition rate have been ignored from this phase.

The results showed that the best average recognition rate was 99.49 % which is better than the average accuracy of the original system and the results for the method with a 6-by-6 fuzzy grid. This result was achieved by the weighting of D_{row4}, K_{col4}, O_{wh}, O_{row3}, O_{col4}, O_{col6}, X_{row1} input parameters. The results are introduced in Fig. 12.

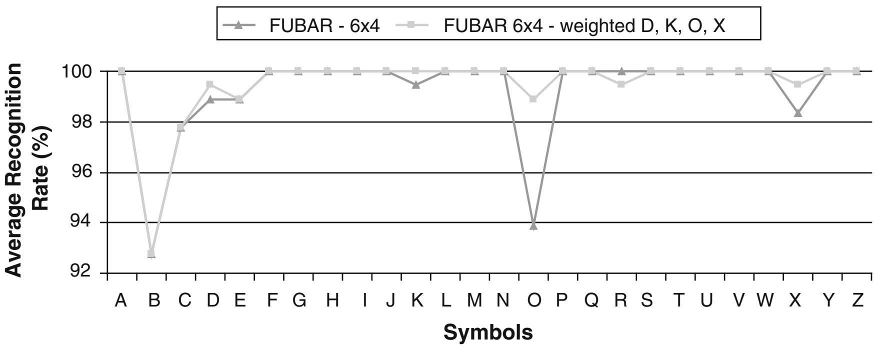


Fig. 12 Average recognition rates of FUBAR with a 6-by-4 fuzzy grid and weighted “D”, “K”, “O”, “X”

The continuous analysis and update of the reliability parameter might be necessary because it depends on the user specific writing style and the definition of the symbols in the rule-base. The adaptation phase included in the original system uses all needed parameters and it is easy to modify to recalculate the reliability parameter for a small computational cost.

4 Results

The original system reached 99.23 % average accuracy rate using a 6-by-4 fuzzy grid with 26 different symbols. The recognition rates for individual symbols are shown in Table 1.

With weighted rule input parameters the same system has a 99.49 % average recognition rate. The recognition rates for letters are shown in Table 2.

The average accuracy of the modified system is higher than the original system’s one but some letters have a lower individual recognition rate. The reduction in the results is caused by the similar look of different symbol pairs, like “R” and “B”.

The developed method has outstanding results compared to other recognizers. The average recognition rates of the different FUBAR versions and other known single-stroke based recognition engines are shown in Fig. 13.

In the study of Fleetwood et al. [14], they compared the performance of Palm virtual keyboard and Graffiti recognizer. The users reached 98 % accuracy with the keyboard and 91 % recognition rate with the Graffiti single-stroke recognizer.

In another study Költringer and Grechenig [15] analyzed the performance of the improved Graffiti (also known as Graffiti 2) which was able to recognize few multi-stroke gestures too (representing i, k, t and x characters). Test results showed 86.03 % recognition rate.

Table 1 Average symbol recognition rates of FUBAR with 6-by-4 fuzzy grid

Symbols	Average recognition rate (%)	Symbols	Average recognition rate (%)
A	100	N	100
B	92.7778	O	93.8889
C	97.7778	P	100
D	98.8889	Q	100
E	98.8889	R	100
F	100	S	100
G	100	T	100
H	100	U	100
I	100	V	100
J	100	W	100
K	99.4444	X	98.3333
L	100	Y	100
M	100	Z	100

Table 2 Average symbol recognition rates of FUBAR with 6-by-4 fuzzy grid and weighted D, K, O, X symbols

Symbols	Average recognition rate (%)	Symbols	Average recognition rate (%)
A	100	N	100
B	92.7778	O	98.8889
C	97.7778	P	100
D	99.4444	Q	100
E	98.8889	R	99.4444
F	100	S	100
G	100	T	100
H	100	U	100
I	100	V	100
J	100	W	100
K	100	X	99.4444
L	100	Y	100
M	100	Z	100

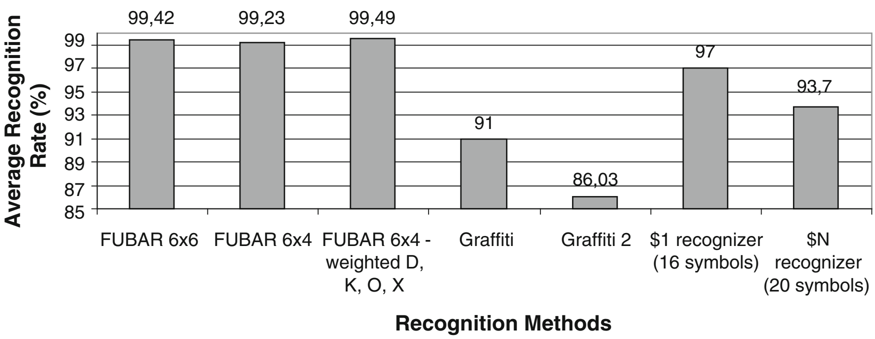


Fig. 13 Average recognition rates of different recognition methods

The $1 recognizer presented by Wobbrock et al. reached 97 % accuracy using 16 gestures with one loaded template per symbol [16]. Our system reached 100 % recognition rate with 16 symbols using only one rule per symbol.

$N (the improved $1) recognizer reached 93.7 % accuracy using 20 multi-stroke symbols with more than three loaded templates per gesture [17].

5 Conclusion and Future Work

The developed system reached a recognition rate that is well over 97 % which is the user acceptance threshold defined by LaLomia.

The used algorithm does not contain difficult geometrical transformations such as scaling and rotating to save resource and time. The computational complexity of the pre-processing in FUBAR is $O(N_{p0})$ where N_{p0} is the number of points representing the input stroke (with the order of 100). During the feature extraction phase the algorithm complexity is $O(N_p)$ where N_p stands for the number of points in the stroke after resample phase (with the order of 10). In the inference phase the complexity is $O(N_{sym} \bullet N_{dim})$ where N_{sym} is the number of symbols (26 by default) in the system and N_{dim} is the dimension of fuzzy rules. This is less than the computational complexity of other investigated recognition systems.

The symbol set of the designed system could be easily changed by adding or removing fuzzy rules which describe the value of the parameters of the given symbol. There is no need for training, restarting or any computation after the modification of the symbol-set.

The weighting of different rule input parameters is an efficient improvement to the algorithm considering the increase of the recognition rate and that the computational complexity did not change. The recognition rate has reached greater than all other published academical and commercial recognition systems and the computational complexity is lower than in all other analyzed systems.

With different improvements the developed system could have a lower computational complexity and a greater recognition rate.

Currently we are analyzing the effect of different hierarchical rule-bases on recognition rates and computational complexity.

Different methods are under development to support multi-stroke character recognition.

The capability of off-line character recognition is needed for further tests on public character bases and for possible industrial applications. This could be done by adding a “Phase 0” to the algorithm which provides usable stroke information from image input.

It is also important to analyze the accuracy of the developed method using different alphabets (e.g. Greek).

We have plans to add a dictionary(-like) support to the recognition algorithm. Currently it does not use any word or letter based support which could increase the recognition rate.

References

1. LaLomia, M.J.: User acceptance of handwritten recognition accuracy. In: Companion Proceedings of CHI ’94, New York, p. 107 (1994)
2. Tormási, A., Botzheim, J.: Single-stroke character recognition with fuzzy method. In: Balas, V.E. et al. (eds.) New Concepts and Applications in Soft Computing SCI, vol. 417, pp. 27–46. Springer, Berlin (2012)
3. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
4. Ruspini, E.H.: A new approach to clustering. Inf. Control **15**(1), 22–32 (1969)
5. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man Mach. Stud. **7**, 1–13 (1975)
6. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst. Man Cybern. **SMC-15**, 116–132 (1985)
7. Holland, J.H.: Adaption in Natural and Artificial Systems. The MIT Press, Cambridge (1992)
8. Nawa, N.E., Furuhashi, T.: Fuzzy system parameters discovery by bacterial evolutionary algorithm. IEEE Trans. Fuzzy Syst. **7**(5), 608–616 (1999)
9. Ishibuchi, H., Nakashima, T.: Effect of rule weights in fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst. **9**(4), 506–515 (2001)
10. van den Berg, J., Kaymak, U., van den Bergh, W.M.: Fuzzy classification using probability-based rule weighting. In: Proceedings of the 11th IEEE International Conference on Fuzzy Systems, Hawaii (2002)
11. Ishibuchi, H., Yamamoto, T.: Rule weight specification in fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst. **13**(4), 428–435 (2005)
12. Tormási, A., Kóczy, L.T.: Comparing the efficiency of a fuzzy single-stroke character recognizer with various parameter values. In: Greco, S. et al. (eds.) Proceedings of IPMU 2012, Part I. CCIS, vol. 297, pp. 260–269 (2012)
13. Tormasi, A., Kóczy, L.T.: Efficiency and accuracy analysis of a fuzzy single-stroke character recognizer with various rectangle fuzzy grids. In: Proceedings of CSCS ’12, Szeged, pp. 54–55 (2012)
14. Fleetwood, M.D. et al.: An evaluation of text-entry in Palm OS—Graffiti and the virtual keyboard. In: Proceedings of HFES ‘02, Santa Monica, CA, pp. 617–621 (2002)
15. Költringer, T., Grechenig, T.: Comparing the immediate usability of Graffiti 2 and virtual keyboard. In: Proceedings of CHI EA ’04, New York, pp. 1175–1178 (2004)
16. Wobbrock, J.O., Wilson, A.D., Li, Y.: Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In: Proceedings of UIST ‘07, pp. 159–168. ACM Press, New York (2007)
17. Anthony, L., Wobbrock, J.O.: The $N multi-stroke recognizer. In: Proceedings of GI’10, Ottawa, pp. 245–253 (2010)

The Development of an Algorithmic Model of Object Recognition Using Visual and Sound Information Based on Neuro-fuzzy Logic

Sabina Shahbazzade

Abstract The chapter considers the problem of recognition of visual and sound information by constructing a virtual environment, which allows a qualitatively simple system to carry out experiments and create an algorithmic model of pattern recognition comparable to human capabilities. The aim of the research is to obtain an algorithmic model that can extract from the surrounding world, "meaningful" (visual and sound) objects to link with the relevant lexical concepts which are the atomic building blocks of intelligence. The research is dedicated to the development of machine intelligence with the phased increase in the complexity of the behavioral model of artificial personality (AP), with the goal being experimental research in the problem of artificial intelligence.

1 Introduction

This chapter is devoted to the creation of an algorithmic model of a system potentially capable of eventually recognizing visual 2D and 3D objects, and sound information in a complex artificial virtual environment.

A virtual environment is a space in which objects of recognition are placed from the library of objects (3D models and corresponding textures) and sound fragments. In this environment an artificial person imitating human behavior is placed.

The time in a virtual environment is calculated independently of the actual elapsed time out of system which allows one to create an artificial personality (AP) unlimited computing power.

Numerous experiments carried out on standard visual and sound objects, showed that in the recognition of visual objects or natural speech, there is some

S. Shahbazzade (✉)
Baku, Azerbaijan
e-mail: sh.sabina86@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_12,

threshold detail comparable with the size of the object. Threshold detail is characteristic of the optimization, proper selection, which provides maximum information content with minimal interference, and, accordingly, the deviation from this value greatly reduces the reliability of recognition results.

In fact, the main thing is not the specific details of object, but the sufficient volume required for the recognition of information, contained in a given area, and the geometric dimensions of objects, which in each case is different. This geometric area which includes the necessary and sufficient volume of characteristic information is called a Unitary Square (US).

US for visual objects are measured in units of space, but sounds are measured in units of time. US for visual information represent some geometric area and for sound information it represents duration of the sound signal, including the whole or a large part of the visual object (or the sound object) is sufficient for its identification or classification.

2 Methods of Local Focus, the Visual and Sound Information

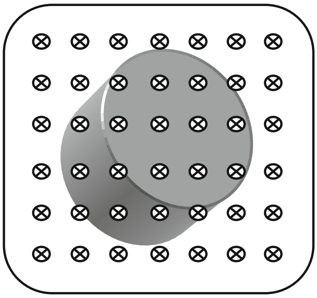
Research and experiments have shown that for that system the only available method for identification is the method of local focus, representing methods of imitation of the behavior of natural intelligence.

The parameter of the depth of the scene is very important in order to parse the scene for visual objects, due to the fact that relative to some of observers for each object corresponds to its own focal length. The essence of the method is a series of sets of measurements of focal lengths, the levels of which indicate an approximate map of surface of the visual object [4–6]. If the focal length increases sharply, then either the given area is out of the boundaries of the object or there is a hole in the object. In any case, within a finite amount of time, the system generates a map of the focal lengths and the actual image of this area (in the future image-contour).

The stage of recognition of visual information is a function of calculating the similarity of objects on which the system is already trained and information received by the image-contour of the target object. The algorithm of the similarity function is based on comparing, with a small level of allowable error, focus maps of compared objects [18, 23] (Fig. 1).

Parsing and recognition of sound objects is based on properties of sound information, which consist due to the fact that relative to some listeners of each sound, information in a certain time period corresponds to the proper frequency area [15].

Figure 2a shows the frequency density of a sound fragment with the word "rock" (circled in red oval), obtained from the appropriate unitary area (Fig. 2b) spoken by female voice with background noise. If the word was spoken by a male voice, then the fundamental frequency density would be in the area circled in the blue square. If several clusters of densities are recognized, the procedure is carried out on each of them regardless of their form [2, 7]. The longer the sound



∞	∞	∞	∞	∞	∞	∞
∞	∞	2	2	2	∞	∞
∞	1.3	2	2	2	2	∞
∞	1.3	2	2	2	2	∞
∞	1	1.3	1.6	1.8	∞	∞
∞	∞	∞	∞	∞	∞	∞

Fig. 1 Illustration of the application of a method of local focuses

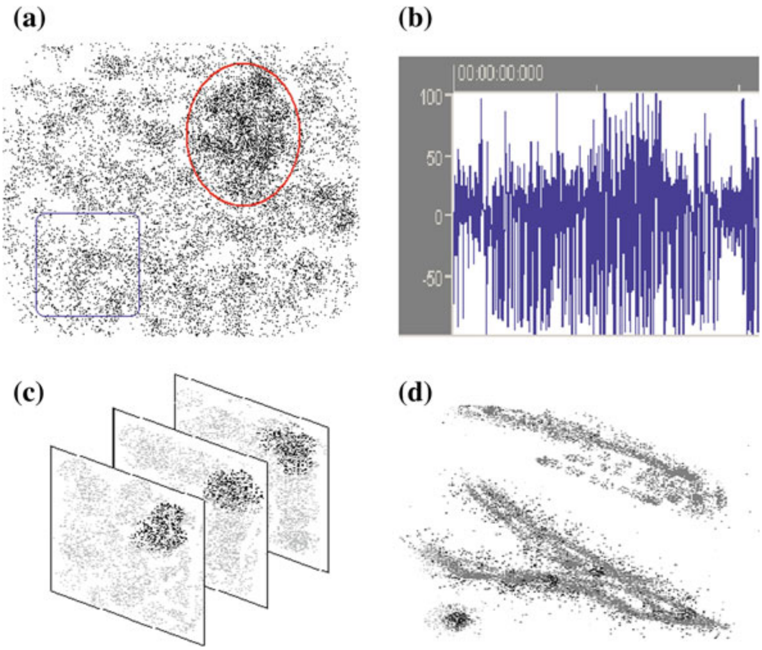


Fig. 2 **a** Frequency density of the word “rock” with a lot of background noise; **b** The unitary square (US) of sound object “rock”; **c** An example of multiple frequency densities; **d** Frequency density of non-verbal information

information, the greater the density of data that is negatively for the quality of analysis. Therefore, to achieve 25–30 % coverage is necessary to add a new bitmap (Fig. 2c). Analysis of the frequency density of artificial sounds has characteristic appearances in the form of dense points, lines or waves. Figure 2d shows analysis of symphonic musical instruments, some forms of violin strokes are represented in the upper part of the figure, brass bifurcated is in the center of the figure, and drums are represented by the point at the bottom left corner.

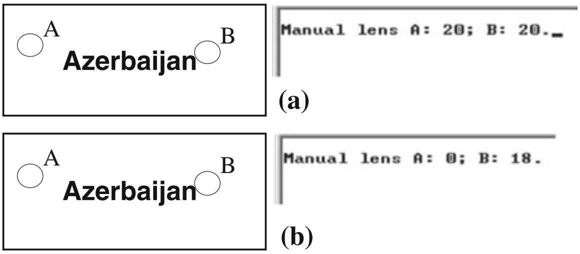


Fig. 3 Imitation of volume of flat geometric shapes: **a** usual 2d model, **b** 2d model with emulation third dimension

3 Recognition of Flat Objects (2D)

Classically presented sheet and painted on it geometric shapes are the flat objects (Fig. 3a) and, accordingly, the focal lengths of them are same, that does not allow creating a map of focal lengths in the case of 3-dimensional objects. To preserve the universality of the recognition subsystem of visual information, the problem was solved by an imitation of the third dimension for flat objects.

If one does not take into account the particular solution that provides almost 100 % results, the only method of teaching which gives satisfactory results is the method of converting the flat form of geometric shapes into 3D. The sheet on which the figure is drawn is a distant background (In Fig. 3b, focal distance of the point A is 0, i.e. as far as possible).

This approach makes it possible to produce a clear analysis of visual objects, but a steady recognition method has not yet been found. For example, about 65 % accuracy is attained in recognizing letters and about 75 % accuracy is attained in word recognition.

4 Learning and Self-Learning

An education system and quality control of procedure for pattern recognition is performed using the base previously prepared of training models. On the whole education system is based on modules:

- Decision-making process under the experimental conditions;
- The mechanism of formation of a database object recognition;
- Subprogram control achieved during the learning outcomes.

The four fundamental learning stages are illustrated below:

1. Object preparation—the creation of two-dimensional models for photography and three-dimensional models for visual objects;
2. Selection of lexical concepts—an important intellectual stage. The researcher should make a list (possibly with one element) of lexical concepts that in his opinion is consistent and is closely associated with a new object [17];

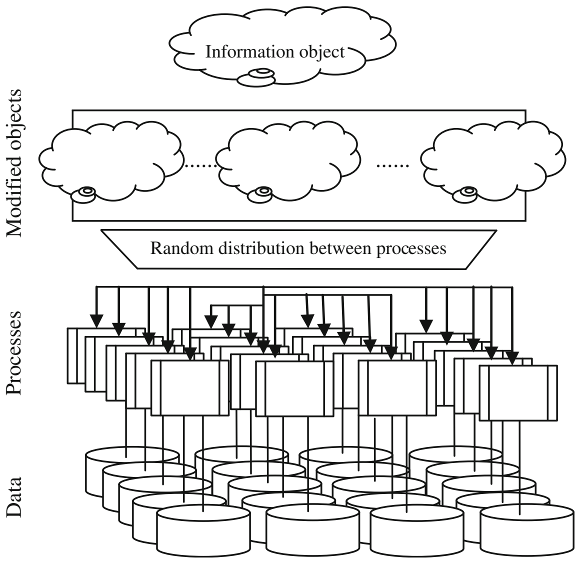


Fig. 4 Carrying out of learning—a stage of storing of object

3. Learning process—learning is a special process in which the system is proposed to recognize the new object of information. This is usually performed in a room in which only one new object is located at any given time. During this phase of the experiment, after selecting an object from the environment and at the end of stage recognition, the system receives a list of lexical concepts for corresponding new object;
4. Saving of object—In this critically important stage, quality implementation is crucial for sustainable functioning of the system as a whole. Saving of the object is a function that creates about 20 copies of the object, each of which has some minor distortion of the geometric dimensions of objects and their orientation, and color distortion, including black and white image. After receiving the modified copies and the original, the system randomly distributes between 1 % of the available system of modular processes and saves them in the relevant database [11].

The basis of self-learning is the idea of self research by AP, using available visual and sound information, to parse and save object recognition as nameless patterns (Fig. 4) [24].

Procedurally, self-learning is carried out in the following stages:

1. An unknown object (system has not been previously trained to this object) is placed in the virtual room and loading the procedure of recognition;
2. If during the recognition of a target object, recognition accuracy of less than 50 % is achieved, or an uncertain controversial situation with 70 % reliability occurs, then the object is stored under the nameless marker;
3. The same object is placed in the virtual room again and re-start the procedure of recognition;

4. If the result of recognition of the target object gives reliability of recognition similar to the corresponding nameless marker, comparable to 70 % and above, and it does not create controversy between other objects, then the object is considered as identified;
5. Otherwise, the process must revert to the second stage.

Upon the successful identification of the target object in 8–12 cases for the same nameless marker, the system is allowed to ask the question “What is it?” and operator introduces the lexical concepts, and it becomes a full marker.

The system asks the operator a question about the meaning of the lexical concept recognized objects, to obtain the largest possible number of synonyms. Few there are objects which may have more than 12 synonyms, other objects may be only 2–3 synonym. To solve this problem, was picked up by an algorithm which, when receiving a large number of identical synonyms limited to eight questions that would have on the one hand reduce the burden on operators, on the other hand, eliminate or reduce the possibility of false labeling.

If the marker already occurs in the system, then the set of visual or sound examples of the object joins an existing set [22].

With sound objects there is an additional uncertainty, which requires the intervention of a researcher in the learning process associated with the uncertainty of the main group of frequencies carrying the information. For recognizing sound information, indication of the frequency range mostly required for 5–8 first objects of information that define a library of groups of sound frequencies used in the sequel to defined of voice information.

Basically, for system important the only verbal information, which has a fairly clear system of formation and it is well identified.

As nameless lexical concepts using simple numeric codes (0001, 0002, 0003,…). It is quite feasible, since the lexical concepts play a minor role in the recognition process, which allows the researcher to monitor the correct functioning of the system [9].

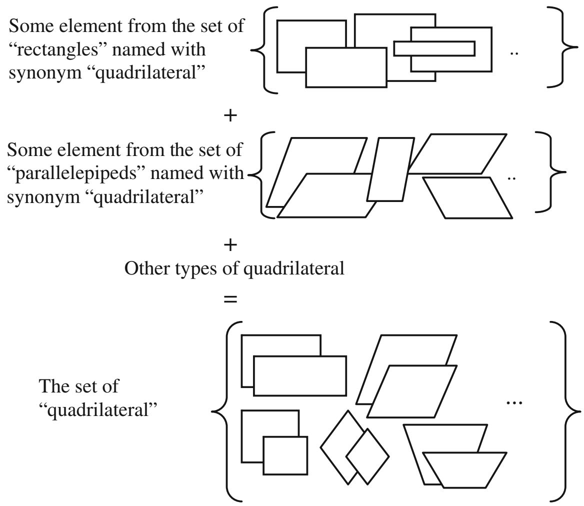
5 Method of Classification of Information Objects

As described above, visual and sound objects are stored in bundles with the lexical concepts exercising the function of the associative marker. The marker is the cornerstone of ensuring the viability of the method of classification.

In fact, the lexical concepts (a marker of information objects) are the links of a higher order of intelligence, which are based on the atoms of information—words of their native language, in which every person thinks [1].

As the number of information objects associated with the markers, it is definitely a system of classification of self-organization, with this classification as learning becomes more complex. In this case, regardless of the number of related markers, the quality of self-organization does not get worse, since the formation of bonds is a natural character. [6].

Fig. 5 Creating a virtual set of "quads" as an example of self organization of objects



All visual and sound information enters into the system as equal objects and as learning becomes associative, groups (bundles/ligament: the lexical concepts ↔ object), the growth of the number of objects of information will lead to the fact that some markers (A and B) will refer to each other like objects (A↔B), but at the set of marker (A) is more extensive than (B) (for diversity rather than on quantity, A↔C,…), therefore, the marker (B) is a subclass to the marker (A). In such a way, a hierarchy of classes and subclasses is formed (Fig. 5).

For example, it is assumed that the system is able to freely distinguish three types of quadrilaterals: squares, rectangles and parallelepipeds. Then into the further learning in the system was introduced the concept of a quadrilateral in combination with a square, a rectangle and a parallelepiped. A consequence of learning is to create a virtual (as nowhere is anything is created), parent class of quadrangles covering the child classes (markers) of squares, rectangles and parallelepipeds that is perfectly valid in terms of geometry. In addition, one can see that the class (marker) square could already be a child of the rectangle, if in the set of rectangles there is an object recognizable as the square [8].

6 Resolving a Dispute Situation: Hybrid Method of Decision-Making

One of the main tasks of the hybrid subsystem of decision-making is the task of improving the quality of recognizable images. [3] Decision made by the additional features of the image-analysis test circuits and reference objects with a fuzzy inference according to Fig. 6 [14].

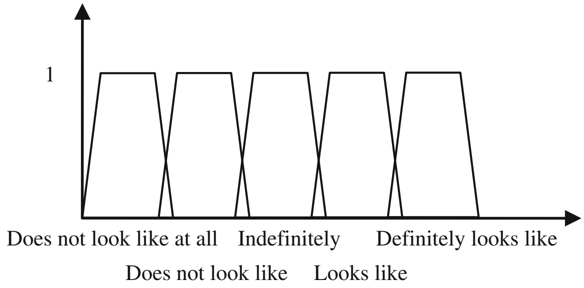


Fig. 6 Fuzzy representation of the results of pattern recognition

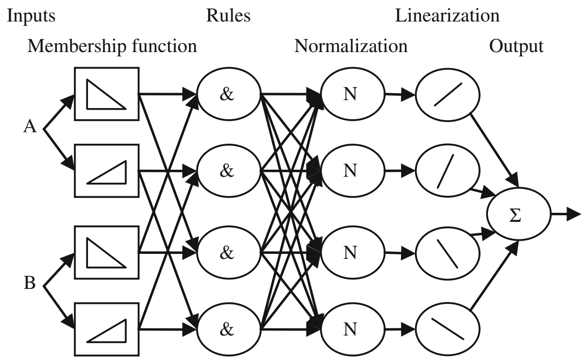


Fig. 7 The structure of applied ANFIS Network

To obtain satisfactory results of the decisions has been selected well-proven fuzzy neural network, which is built on the principle of Adaptive Neural Fuzzy Inference Systems (ANFIS), shown in Fig. 7 [13].

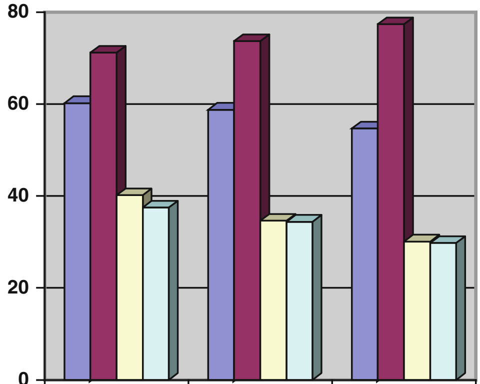
Decision-making procedure is repeated until the system is not sure what that one lexical marker characterizing the images as similar to test image, with the condition that the gap between the images of other lexical markers will be more than 10–15 % [16].

In general, a hybrid method of decision-making [26] produces the potential with equal chances, of selecting one object information from the three options, improving the quality of decisions made by approximately 10–15 %.

Thus, if the current average level of reliability in decision-making is around 70–75 %, neuro-fuzzy logic [20] will support this level at a drawdown to 55–60 % and it can be considered of quite satisfactory quality.

The necessity of application of the additional solving device is possibly due to the occurrence of disputable situations in which reliability of the most probable object of the information (A) differs from the second for probability (B) less than on 10 %.

Generally this means that the probability of needing a second solving device is relatively low and the determining factor is the presence of a 10 % area of the most likely one [25].



	1	2	3
Square	60,2	58,8	54,7
Rectangle	71,2	73,7	77,4
Parallelepiped	40,2	34,6	30,1
Trapeze	37,5	34,4	29,8

Fig. 8 Sequence from the three procedures to improve level of recognition of rectangle

Dynamic generation of the two (or more if needed) software implementations of fuzzy neural networks and their learning back propagation algorithm is the basic method of direct supervised learning of multilayer fuzzy neural directivity networks [12].

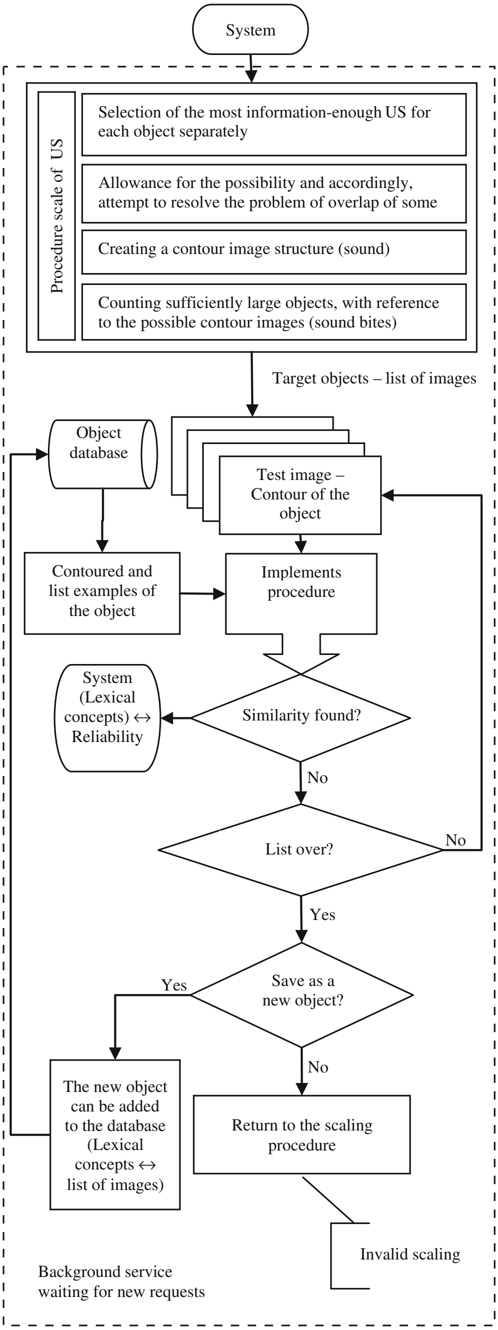
The size of a fuzzy neural network is determined by the volume of the object information. The main difference in application of fuzzy neural networks in the given work, is that each fuzzy neural network has learned only one of the patterns it is researching.

A feature of application is the absence of a priori given information in any form (usually put in logic of work of systems), and also the separation of procedure of decision-making at occurrence of disputable situations on two stages:

- One fuzzy neural network learning object (A), but the second object (B), while recognized object to the learning is not involved.
- At the end of training and the first and second fuzzy neural networks, the recognize objects send to their input, and holding for comparing of object proximity relevant of the visual or sound samples. The results of both of tests are applied to the second stage.
- The second stage is the processing of the fuzzy neural images as a of result of the reaction to the etalon object to the most probable objects of information that are transferred to the chart of implements in the form of specific fuzzy variables.

According to these values obtained by applying the algorithm of fuzzy comparisons {Definitely looks like $(+0.3, +0.5) = 10\ \%$; Looks like $(+0.1, -0.3) = 8\ \%$; Indefinitely $(-0.1, +0.1) = 4\ \%$; Does not look like $(-0.3, -0.1) = 0\ \%$; Does not look like at all $(-0.5, -0.3) = -3\ \%$}, choosing the closest indicator functions that are added to the competing object information (lexical concepts) (Fig. 8).

Fig. 9 Organizational block diagram of service modules in an infinite cycle of recognition



If after the decision making [21] the most likely object increases the gap with its competitors to a safe level (10–15 %), then the system decides in favor of the most probable object of information and accordingly, the associated lexical concepts, otherwise it is considered that the match object is not found and the algorithm passes to the interrupted work (search, record, etc.).

7 Algorithmic Model of Functioning

In Fig. 9, the operation of the system in standard conditions of the experiments. This block diagram of represents an algorithm of the main server, perform the coordination task management of the system modules. The organizational block diagram illustrates the overall cycle of functioning of the service modules that are involved in pattern recognition in an infinite cycle and in the case of successful recognition send results to the main service, and when new patterns are received automatically initiate a new task [10, 19, 20].

8 Conclusion

This article is the result of research and development of methods for recognition of object information and communication with the lexical concept without optimizing simplifications. It is dedicated to the research of creating of an automated system for the recognition of complex images and natural speech. When the learning is applied using a powerful computing system, the results will be able to recognize the visual and audio information, even at a high level of interference.

Suggested solution of the problem of allocation of objects of information from an overall picture and rather each other, is a method of local focuses for the visual information and a method of frequency density for the sound information is offered. The applied method is able to parse visual scenes and sound samples of high complexity, as well as to identify their constituent objects.

The algorithmic model of the functioning of the system is illustrated. Functional block diagrams describe the algorithm providing parse and recognition of the visual and sound information on the basis of architecture “the main service ↔ modular services”. The developed function of calculation of similarity of objects being the functional center of modular services, represents the unique algorithm which has shown satisfactory results and high stability of indicators of reliability.

The mechanism of a hybrid neuro-fuzzy method of decision-making reveals the occurrence of disputable situations. Defined the concept of the controversial situation, which is a consequence of issuing by the system of the two or more markers as a result recognition. Parameters and algorithms of functioning of hybrid system are also experimental factors which give satisfactory results on quality of accepted decisions.

The developed method of self-learning represents the mechanism of postponed learning, when the system learns to define the presence of new objects, self-learning, with the subsequent definition by the lexical information characterizing the given object.

The practical application of these scientific results is the ability to integrate visual and sound information around the lexical concepts that will be important in the further development of the system in the field of artificial intelligence.

References

1. Abakumov, V.G., Krilov, V.N., Antoshuk, S.G.: Detection and recognition of attributes of the objects using a spherical model of the visual analyzer. Electron. Commun. **8**(2), 211–212 (2000)
2. Belyavchev, V.G., Voskoboiniov, Y.E.: Local adaptive filtering algorithms for digital images. Sci. Bull. NQTU. **3**, 21–32 (1997)
3. Breiman, Friedman: Classification and Regression Trees, Olshen and Stone. Wadsworth and Brooks, Monterey (1985)
4. Dai, Y., Nakano, Y.: Recognition of facial images with low resolution using a Hopfield memory model. Pattern Recogn. **31**, 159–167 (1998)
5. Drouin, M., Abou-Kandil, H., Mariton, M., Duc, G.: Une nouvelle methode de decompozision-coordinasion 1 re partie: Principe et mise en ocuvre. APII. **3**, 205–226 (1985)
6. Duc, G., Drouin, M., Mariton, M., Abou-Kandil, H.: Une nouvelle methode de decomposition-coordination. 2 e partie: Application a la compensation des systemes multivariables. APII. **3**, 227–242 (1985)
7. Frei, W.: Image enhancement by histogram hyperbolization. Comput. Graph. Image Proc. **6**(3), 286–294 (1977)
8. Gonzalez, R.C., Fittes, B.A.: Gray–level transformation for interactive image enhancement. Mech. Mach. Theory. **12**, 111–112 (1975)
9. Gonzalez, R.C., Wintz, P.: Digital Image Processing, p. 505. Addison–Wesley, Reading (1987)
10. Gutta, S., Wechsler, H.: Face recognition using hybrid classifiers. Pattern Recogn. **30**, 539–553 (1997)
11. Hall, E.L.: Almost uniform distribution for computer image enhancement. IEEE Trans. Comput. **23**(2), 207–208 (1974)
12. Jacobsen, X., Zscherpel, U., Perner, P.: A comparison between neural networks and decision trees. Machine learning and data mining in pattern recognition. Lecture Notes in Artificial Intelligence, pp. 144–158 (1999)
13. Jang, J.S.R.: ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Trans. Syst. Man Cybern. **23**(5/6), 665–685 (1993)
14. Jang, J.S.R., Gulley, N.: The Fuzzy Logic Toolbox for Use with MATLAB. The MathWorks Inc, Natick (1995)
15. Lee, Y.: Handwritten digit recognition using K neares-neighbor, radial-basis function, and back-propagation neural networks. Neural Comput. **3**, 440–449 (1991)
16. Michie, D., Spiegelharter, D.J., Taylor, C.C (eds.): Machine Learning, Neural and Statistical Classification. Ellis Horwood, New York (1994)
17. Petrou, M.: Learning in pattern recognition. Machine learning and data mining in pattern recognition. Lecture Notes in Artificial Intelligence, pp. 1–12 (1999)
18. Ranganath, S., Arun, K.: Face recognition using transform features and neural networks. Pattern Recogn. **30**, 1615–1622 (1997)

19. Sergeev, V.V.: Application of a pattern recognition tasks in digital image processing. Autosummetry. **2**, 63–76 (1998)
20. Shahbazova, S.H.N., Suleymanov, M.R.: Stages of recognition probability images. In: Proceedings of AzTU (2004)
21. Shahbazova, S.H.N., Bonfig, K.V., Suleymanov, M.R.: Questions of construction of system of recognition of non-standard image and speech in the architecture of a hybrid vehicle based on neuro-fuzzy logic. In: 2nd international conference on application of fuzzy systems and software computing and analysis system, decision and control. Barcelona, Spain (2004)
22. Sabah, G.: Knowledge representation and natural language understanding. AI Commun. **6**(3–4), 155–186 (1993)
23. Valentin, D., Abdi, H., O'Toole, A.J., Cottrell, G.W.: Connectionist models of face processing: a survey. Pattern Recognit. **27**, 1209–1230 (1994)
24. Vetter, T., Poggio, T.: Linear object classes and image synthesis from a single example image. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 733–742 (1997)
25. Yoon, K.S., Ham, Y.K., Park, R.-H.: Hybrid approaches to frontal view face recognition using the hidden markov model and neural network. Pattern Recogn. **31**, 283–293 (1998)
26. Zadeh, L.A.: Fuzzy sets and information granularity. In: Gupta, M., Ragade, R., Yager R. (eds.) Advances in Fuzzy Set Theory and Applications, pp. 3–18. Kluwer Acadamic Press, Amsterdam (1979)

Image Thresholding by Grouping Functions: Application to MRI Images

Aranzazu Jurio, Daniel Paternain, Miguel Pagola and Humberto Bustince

Abstract In this work we present a thresholding algorithm for greyscale images. Our proposal is the use of grouping functions to find the best threshold. These functions are able to measure the belongingness of a grey intensity to the background or to the object of the image, so the best threshold is the one associated with the highest grouping value.

1 Introduction

One of the most used techniques in image segmentation is thresholding or segmentation by greylevels [1], [2], [3]. In thresholding, the different objects of the image are characterized just by the intensity of each pixel. This technique consists in finding a threshold t such that the pixels whose intensities are lower or equal to t belong to the background of the image while the intensities that are greater than t belong to the object, or vice versa [4]. The advantages of this kind of procedures

This work was supported by the Research Services of the Universidad Publica de Navarra and by the National Science Foundation of Spain, references TIN2010-15055 and TIN2011-29520.

A. Jurio (✉) · D. Paternain · M. Pagola · H. Bustince
Departamento de Automatica y Computacion, Universidad Publica de Navarra,
Pamplona Spain
e-mail: aranzazu.jurio@unavarra.es

D. Paternain
e-mail: daniel.paternain@unavarra.es

M. Pagola
e-mail: miguel.pagola@unavarra.es

H. Bustince
e-mail: bustince@unavarra.es

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_13,

with respect to other segmentation algorithms are the simplicity and low computational cost. This is why thresholding is commonly used as a first step of more complex segmentation algorithms.

In this work we present a new thresholding algorithm, generalizing previous fuzzy approaches [4], [5]. Our proposal is based on the construction, for every possible grey intensity, of two fuzzy sets (Q_{B_t} and Q_{O_t}) representing the belongingness of every greylevel to the background $\left(\mu_{Q_{B_t}}(q)\right)$ and to the object $\left(\mu_{Q_{O_t}}(q)\right)$ of the image respectively. The objective is to find the threshold for wich the belongingness of every grey intensity to the object or to the background is maximum $\left(\mu_{Q_{B_t}} = 1 \text{ or } \mu_{Q_{O_t}} = 1\right)$, so we are completely sure that those pixels belong to the background or to the object of the image.

To solve the aim of the work, we propose the use of grouping functions, as a bivariate aggregation function that gets the maximum value if and only if one of the arguments is 1. In this work we study the axiomatization of these functions, propose some construction methods and relate grouping functions to overlap functions [6].

We also show an illustrative example for a medical imaging application, where we have to segment some magnetic resonance images. The purpose is to separate the gray matter from the white matter of a brain, which is a very helpful process to evaluate some diseases like Alzheimer or schizophrenia.

The rest of the contribution is organized in the following way. We start recalling some preliminary concepts in Sect. 2. In Sect. 3 we study grouping functions, their relations with overlap functions and some construction methods. In Sect. 4 we present our image thresholding algorithm, and in Sect. 5 we show an illustrative example. We finish with some conclusions in Sect. 6.

2 Preliminaries

A strict negation [7] is a continuous and strictly decreasing function $N : [0,1]^2 \rightarrow [0,1]$ such that $N(0) = 1$ and $N(1) = 0$. A strong negation is a strict negation that is also involutive, it means, $N(N(x)) = x$ for all $x \in [0,1]$.

A triangular norm (t-norm) is a symmetric and associative bivariate aggregation function $T : [0,1]^2 \rightarrow [0,1]$ such that $T(x,1) = x$ for all $x \in [0,1]$. Some examples of t-norms are minimum function $T_M(x,y) = \min(x,y)$ or product function $T_P(x,y) = x \cdot y$. A triangular conorm (t-conorm) is a symmetric and associative bivariate aggregation function $S : [0,1]^2 \rightarrow [0,1]$ such that $S(x,0) = x$ for all $x \in [0,1]$. Some examples of t-conorms are the maximum function $S_M(x,y) = \max(x,y)$ or the probabilistic sum function $S_P(x,y) = x + y - x \cdot y$ [8–10].

In this work we use restricted equivalence functions (*REF*) to build the fuzzy sets associated with an image [11], [4].

Definition 1 A function $REF : [0,1]^2 \to [0,1]$ is called restricted equivalence function if it satisfies the following conditions:

(1) $REF(x,y) = REF(y,x)$ for all $x, y \in [0,1]$;
(2) $REF(x,y) = 1$ if and only if $x = y$;
(3) $REF(x,y) = 0$ if and only if $x = 1$ and $y = 0$ or $x = 0$ and $y = 1$;
(4) $REF(x,y) = REF(N(x), N(y))$ for all $x, y \in [0,1]$, being N a strong negation;
(5) if $x \le y \le z$ then $REF(x,y) \ge REF(x,z)$ and $REF(y,z) \ge REF(x,z)$, for all $x, y, z \in [0,1]$.

3 Grouping Functions

Definition 2 A function $G_G : [0,1]^2 \to [0,1]$ is a grouping function if it satisfies the following conditions:

$(G_G1) G_G(x,y) = G_G(y,x)$ for all $x, y \in [0,1]$;
$(G_G2) G_G(x,y) = 0$ if and only if $x = y = 0$;
$(G_G3) G_G(x,y) = 1$ if and only if $x = 1$ or $y = 1$;
$(G_G4) G_G$ is non-decreasing;
$(G_G5) G_G$ is continuous.

Observe that a grouping function is a particular case of binary aggregation.

We can find a relation between grouping functions and overlap functions, defined in [6]. We use this relation to present several construction methods of grouping functions.

3.1 Overlap Functions

Definition 3 A function $G_O : [0,1]^2 \to [0,1]$ is an overlap function if it satisfies the following properties:

(G_O1) G_O is symmetric.
$(G_O2) G_O(x,y) = 0$ if and only if $xy = 0$.
$(G_O3) G_O(x,y) = 1$ if and only if $xy = 1$.
$(G_O4) G_O$ is non-decreasing.
$(G_O5) G_O$ is continuous.

Theorem 1 *Let G_O be an overlap function and let* N *be a strict negation. Then.*

$$G_G(x,y) = N(G_O(N(x), N(y))) \tag{1}$$

is a grouping function. Reciprocally, we have that

$$G_O(x,y) = N(G_G(N(x),N(y))) \tag{2}$$

is an overlap function.

Proof (G_G1), (G_G4) and (G_G5) are direct. $(G_G2) G_G(x,y) = 0 = N(G_O(N(x), N(y)))$ if and only if $G_O(N(x),N(y)) = 1$ if and only if $N(x) = N(y) = 1$ if and only if $x = y = 0$. $(G_G3), G_G(x,y) = 1 = N(G_O(N(x);N(y)))$ if and only if $G_O(N(x), N(y)) = 0$ if and only if $N(x) = 0$ or $N(y) = 0$ if and only if $x = 1$ or $y = 1$.

Based on the relation between overlap functions and t-norms, in this work we proof that any associative grouping function is also a t-conorm. However, the reciprocal of this theorem does not hold, as it is clear if we consider any non-continuous t-conorm.

Theorem 2 *Let G_G be an associative grouping function. Then G_G is a t-conorm.*

Proof We just need to proof that 0 is the neutral element of G_G. Because of the continuity of G_G and $G_G(0,1) = 1$ and $G_G(0,0) = 0$, we can say that for any $x \in]0,1[$ there exists a $y \in]0,1[$ such that $x = G_G(y,0)$. Then $G_G(x,0) = G_G(G_G(y,0),0) = G_G(y,G_G(0,0)) = G_G(y,0) = x$ and in a similar way $G_G(0,x) = x$.

Example 1 *An associative grouping function and therefore a t-conorm is the maximum function.*

$$G_G(x,y) = \max(x,y) \tag{3}$$

Theorem 3 Let $G_G1,\ldots,G_Gm$ be m grouping functions and let $w_1,\ldots,w_m$ be m non-negative weights such that $\sum_{i=1}^m w_i = 1$. Then the convex sum $G = \sum_{i=1}^m w_i G_G i$ is a grouping function.

Proof Direct.

3.2 Construction of Grouping Functions

In the next theorem we study a construction method of grouping functions from two functions f and h that satisfy certain properties.

Theorem 4 The function $G_G : [0,1]^2 \rightarrow [0,1]$ is a grouping function if and only if

$$G_G(x,y) = \frac{f(x,y)}{f(x,y) + h(x,y)} \tag{4}$$

for $f, h : [0, 1]^2 \to [0, 1]$ such that

(1) f and h are symmetric;
(2) f is non-decreasing and h is non-increasing;
(3) $f(x, y) = 0$ if and only if $x = y = 0$;
(4) $h(x, y) = 0$ if and only if $x = 1$ or $y = 1$;
(5) f y h are continuous functions.

Proof We have to take into account that $f(x, y) + h(x, y) \neq 0$ for all $(x, y) \in [0, 1]^2$. Then the necessity is straightaway taking $f(x, y) = G_G(x, y)$ and $h(x, y) = 1 - G_G(x, y)$. (Sufficiency) (G_G1) (G_G2) (G_G3) and (G_G5) are direct. (G_G4) If $x_1 \leq x_2$ then $f(x_1, y) \leq f(x_2, y)$ and $h(x_2, y) \leq h(x_1, y)$. So we have that $f(x_1, y)h(x_2, y) \leq f(x_2, y)h(x_1, y)$. Multiplying both sides of the equality we have $f(x_1, y)f(x_2, y) + f(x_1, y)h(x_2, y) \leq f(x_1, y)f(x_2, y) + f(x_2, y)h(x_1, y)$. We can rewrite $G_G(x_1, y) = \frac{f(x_1,y)}{f(x_1,y)+h(x_1,y)} \leq \frac{f(x_2,y)}{f(x_2,y)+h(x_2,y)} = G_G(x_2, y)$.

Example 2 If we take $f(x, y) = \max(x, y)$ and $h(x, y) = \sqrt{(1 - x)(1 - y)}$ we have

$$G_G(x, y) = \frac{\max(x, y)}{\max(x, y) + \sqrt{(1 - x)(1 - y)}} \tag{5}$$

Example 3 If we take $f(x, y) = \max(x, y)$ and $h(x, y) = (1 - x)(1 - y)$ we have

$$G_G(x, y) = \frac{\max(x, y)}{\max(x, y) + (1 - x)(1 - y)} \tag{6}$$

Example 4 If we take $f(x, y) = 1 - \sqrt{(1 - x)(1 - y)}$ and $h(x, y) = \min((1 - x), (1 - y))$ we have

$$G_G(x, y) = \frac{1 - \sqrt{(1 - x)(1 - y)}}{1 - \sqrt{(1 - x)(1 - y)} + \min((1 - x), (1 - y))} \tag{7}$$

A deeper study on some properties of grouping functions can be found in [12].

4 Thresholding Algorithm Based on Grouping Functions

In this work we propose the use of grouping functions as the metric to calculate the optimal threshold of an image. To do so, we construct, for every greylevel, a fuzzy set associated with the background and a fuzzy set associated with the object of the image. Applying a convex combination of several grouping functions to these sets, we choose the suitable threshold for each image. The scheme of our proposal is shown in Algorithm 1.

4.1 Construction of Fuzzy Sets Associated with the Image

In thresholding problems with only one threshold, we suppose that the image is divided into two areas, so we can separate one object from the background. Based on the study presented in [4], in this work we construct two fuzzy sets (Q_{B_t} associated with the background and Q_{O_t} associated with the object) from restricted equivalence functions, bearing in mind the following reasoning: the more similar is a greylevel (q) to the average of the background intensities (analogously to the object intensities) the higher the membership value of that intensity to the fuzzy set associated with the background (object) is.

Algorithm 1 Thresholding algorithm

for $t = \{0, 1, \ldots, L-1\}$ (For every greylevel) **do**
  Construct a fuzzy set associated with the background of the image (Q_{B_t}).
  Construct a fuzzy set associated with the object of the image (Q_{O_t}).
  for $q = \{0, 1, \ldots, L-1\}$ (For every greylevel) **do**
    Calculate several grouping functions for $Q_{B_t}(q)$ and $Q_{O_t}(q)$.
    Calculate the convex combination of previous functions obtaining a new grouping function $G_{G_{comb}}(q)$.
  end for
  Calculate the weighted sum of previous grouping

$$\sum_{q=0}^{L-1} G_{G_{comb}}(\mu_{B_t}(q), \mu_{O_t}(q)) \cdot h(q)$$

  where $h(q)$ is the number of pixels whose intensity is q.
end for
Take as best threshold $t*$ the one associated with the maximum sum of grouping:

$$t* = \arg\max_t \sum_{q=0}^{L-1} G_{G_{comb}}(\mu_{B_t}(q), \mu_{O_t}(q)) \cdot h(q)$$

For a fixed greylevel t, we start by calculating the average value of the intensities belonging to the background ($m_B(t)$) and to the object ($m_O(t)$) using the following expressions:

$$m_B(t) = \frac{\sum_{q=0}^{t} q \cdot h(q)}{\sum_{q=0}^{t} h(q)} \quad m_O(t) = \frac{\sum_{q=t+1}^{L-1} q \cdot h(q)}{\sum_{q=t+1}^{L-1} h(q)}$$

Let REF be a restricted equivalence function, we construct the fuzzy sets Q_{B_t} and Q_{O_t} with the following membership functions, for every greylevel $q = 0, 1, \ldots, L-1$:

$$\mu_{Q_{B_t}}(q) = REF\left(\frac{q}{L-1}, \frac{m_B(t)}{L-1}\right) \tag{8}$$

$$\mu_{Q_{O_t}}(q) = REF\left(\frac{q}{L-1}, \frac{m_O(t)}{L-1}\right) \tag{9}$$

We can proof that, with this construction method and due to property (2) of Definition 1, a greylevel has maximum membership degree to the background (object) fuzzy set only if its intensity is the same as the average intensities of the background (object) of the image.

- $\mu_{Q_{B_t}}(q) = 1$ if and only if $q = m_B(t)$.
- $\mu_{Q_{O_t}}(q) = 1$ if and only if $q = m_O(t)$.

4.2 *Grouping Calculus*

To calculate the grouping value associated with each possible threshold t, we use n different grouping functions $(G_G1, G_G2, \ldots, G_Gn)$. A grouping function takes two arguments and calculates the group level between both of them. In this case, we compute the grouping, for every greylevel, between the membership degree to the fuzzy set associated with the background and to the fuzzy set associated with the object.

Using the result obtained in Theorem 3, we combine the n grouping functions previously calculated. In this way, we obtain a new grouping function that, experimentally, it outperforms the result obtained by the worst grouping expression selected. This step helps us to solve the problem of choosing a grouping expression not suitable for a specific problem, what finishes in wrong results.

Once we have one sole value for the grouping of every greylevel, we calculate the sum. This is the value for the grouping associated with the threshold t.

4.3 Selection of the Maximum Grouping

Every possible threshold $t = \{0, 1, \ldots, L-1\}$ has a grouping value associated with it, calculated as the sum of the grouping function in several points. To get the best threshold, we choose the one associated with the highest grouping value. We choose the maximum value because of grouping functions properties. It means, the sum is maximum if $G_G(\mu_{B_t}(q), \mu_{O_t}(q)) = 1$ for all $q = \{0, 1, \ldots, L-1\}$. By property ($G_G2$) of Definition 2 this is achieved in two cases:

- $\mu_{B_t}(q) = 1$, so $q = m_B(t)$. In this case we are completely sure that the pixels whose intensity is q belong to the background of the image, because this intensity is exactly the average intensity of all the pixels of the background.
- $\mu_{O_t}(q) = 1$, so $q = m_O(t)$. In this case we are completely sure that the pixels whose intensity is q belong to the object of the image, because this intensity is exactly the average intensity of all the pixels of the object.

In this sense, by choosing the highest grouping value we are selecting the threshold for which all the pixels whose intensity is lower than the threshold are very closed to the average of background (object) intensities and all the pixels whose intensity is greater than the threshold are very closed to the average intensity of the object (background).

5 Illustrative Example

In this section we show the performance of the proposed algorithm over 10 T1-weighted magnetic resonance images (see Fig. 1). These images are provided by the Center for Morphometric Analysis at Massachusetts General Hospital (available at http://www.cma.mgh.harvard.edu/ibsr/). The aim of the segmentation of this kind of images is to separate each of the pixels inside the brain into one of the following two types: grey matter and white matter. This segmentation can be viewed as part of a volumetric analysis of the brain regions, which is very useful to evaluate the evolution of diseases such as Alzheimer, epilepsy or schizophrenia [13, 14]. To measure the quality of the segmented results, we compare them with an ideal handmade segmentation provided at the same webpage (see Fig. 2). This comparison is measured by the percentage of well classified pixels.

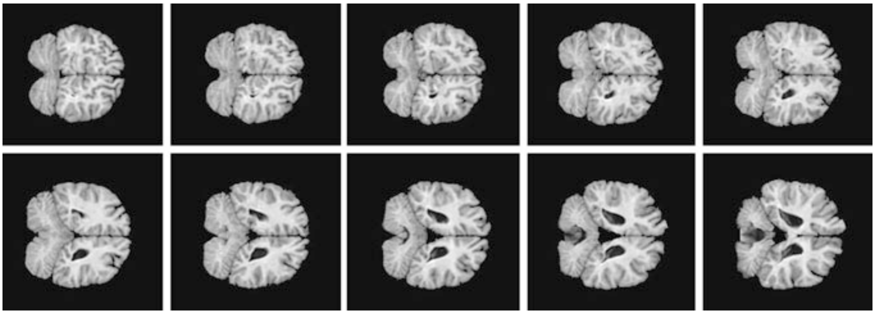


Fig. 1 Original images

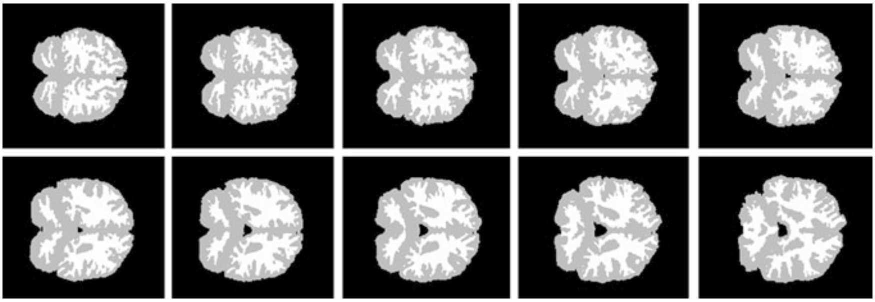


Fig. 2 Ideal handmade segmentations

In this example we use four grouping functions for the step 5 of the algorithm:

- $G_G1(x, y) = \max(x, y)$
- $G_G2(x, y) = \frac{\max(x,y)}{\max(x,y) + \sqrt{(1-x)(1-y)}}$
- $G_G3(x, y) = 1 - \sqrt{(1-x)(1-y)}$
- $G_G4(x, y) = x + y - xy$

To calculate the convex combination of grouping functions we use the same weight for each of them. In this way, each of the four weights is 0.25. If we a priori know that some grouping functions are more suitable than others for a specific image, then it is recommended to use different values for the weights, using greater values for these suitable functions (for example, using weighted means).

In Fig. 3 we see the segmentations obtained by our method for every image.

Next we study the different segmentations obtained by each one of the four proposed grouping functions in relation to the final result obtained by the consensus of all of them. To do so, in Figs. 4, 5 and 6 we show different results for three images. In the first row we see the segmentation obtained by each one of the four grouping functions. In the second row we show in white the pixels well

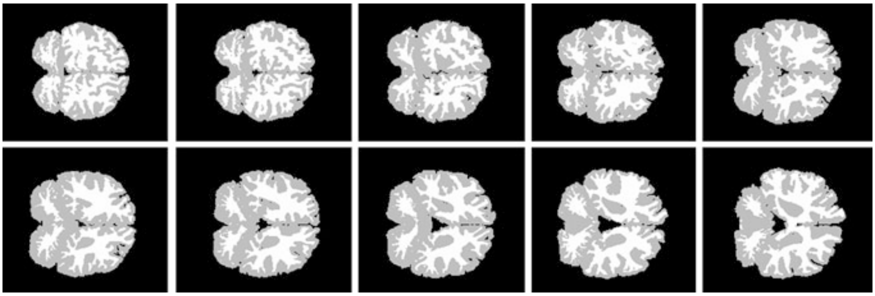
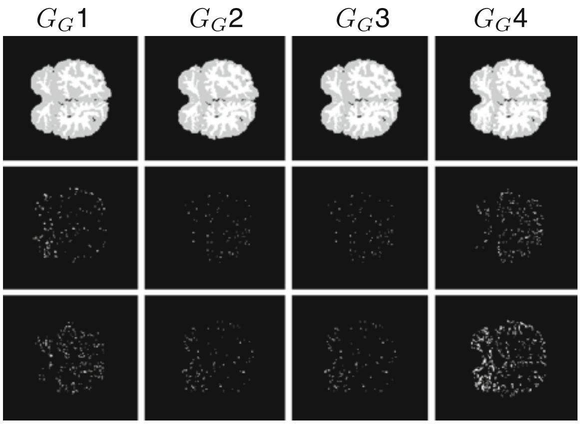


Fig. 3 Obtained segmentations by our algorithm

Fig. 4 Segmentations obtained by each one of the four grouping functions (*first row*). In the *second row* we show in *white* the pixels well classified by every individual grouping function that are wrong classified in the convex combination. In the *third row* we show the pixels wrong classified by every individual grouping function that are well classified in the convex combination



classified by every individual grouping function that are wrong classified in the convex combination. Finally, in the third row we show in white the pixels wrong classified by every individual grouping function that are well classified in the convex combination.

As we can see, depending on the image we segment, some grouping functions are more suitable for thresholding purposes than others. This fact is confirmed with the number of pixels well and wrong classified with respect to the convex combination. In Table 1 we show the thresholds obtained for every image and the percentage of well classified pixels. The first column shows the results obtained by the convex combination of grouping functions, and columns 2–5 show the results obtained by each one of the functions.

As we can experimentally see, the consensus of grouping functions always provides a middle threshold value among the ones obtained by every grouping function and so forth, in most cases the percentage of well classified pixels is intermediate too. In this way, we know that our algorithm does not always get the best possible result with grouping functions. However, the threshold got by the

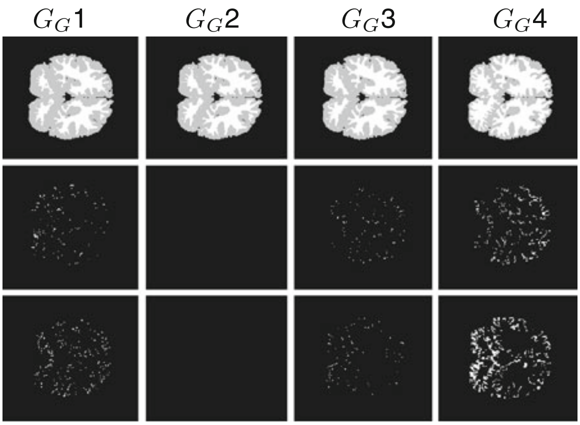


Fig. 5 Segmentations obtained by each one of the four grouping functions (*first row*). In the *second row* we show in *white* the pixels well classified by every individual grouping function that are wrong classified in the convex combination. In the *third row* we show the pixels wrong classified by every individual grouping function that are well classified in the convex combination

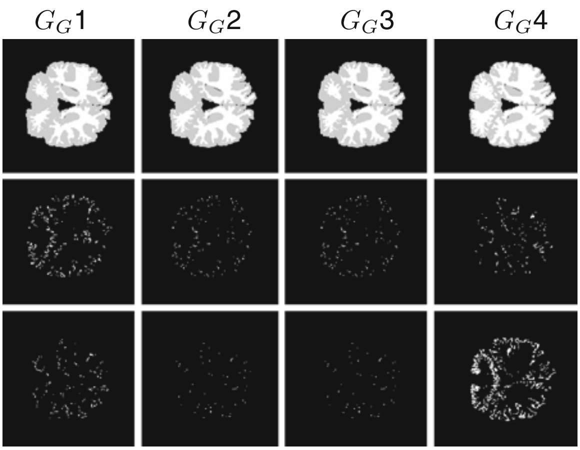


Fig. 6 Segmentations obtained by each one of the four grouping functions (*first row*). In the *second row* we show in *white* the pixels well classified by every individual grouping function that are wrong classified in the convex combination. In the *third row* we show the pixels wrong classified by every individual grouping function that are well classified in the convex combination

agreement tends to be better or equal than the worst of the thresholds got by every of the grouping functions. As we have said, this fact solves the problem of choosing a grouping function suitable for every one of the images.

Finally, we compare our thresholding method based on grouping functions with Otsu's algorithm [3], as it is one of the most used thresholding methods. In Table 2 we show the obtained threshold by both algorithms as well as the percentage of well classified pixels. We can see that our method gets the best results for all images in the experiment, so we can say it improves Otsu's thresholding method for this set of images.

Table 1 Thresholds and percentages got by every grouping function and the consensus of all of them

	Consensus		Grouping 1		Grouping 2		Grouping 3		Grouping 4	
	Threshold	%	Threshold	%	Threshold	%	Threshold	%	Threshold	%
Image 1	158	97.15	162	97.53	158	97.15	158	97.15	156	96.93
Image 2	161	97.25	166	97.75	161	97.25	161	97.25	157	96.73
Image 3	167	98.25	172	98.2	165	98.18	165	98.18	158	97.32
Image 4	166	97.4	175	97.53	166	97.4	166	97.4	159	96.64
Image 5	173	97.67	178	97.6	173	97.67	173	97.67	160	96.54
Image 6	173	97.7	177	97.67	173	97.7	173	97.7	160	96.61
Image 7	174	97.6	180	97.48	174	97.6	172	97.56	158	96.06
Image 8	173	97.94	177	97.83	173	97.94	173	97.94	156	96.15
Image 9	164	96.5	171	96.93	166	96.67	166	96.67	152	94.7
Image 10	165	96.2	165	96.2	165	96.2	165	96.2	150	94.46

Table 2 Thresholds and percentages got by our method and Otsu's one

	Consensus		Otsu	
	Threshold	%	Threshold	%
Image 1	158	97.15	155	96.82
Image 2	161	97.25	156	96.58
Image 3	167	98.25	159	97.45
Image 4	166	97.4	159	96.64
Image 5	173	97.67	162	96.8
Image 6	173	97.7	157	96.13
Image 7	174	97.6	154	95.44
Image 8	173	97.94	150	95.24
Image 9	164	96.5	137	92.35
Image 10	165	96.2	143	93.52

6 Conclusions

In this work we have presented a thresholding algorithm for greyscale images based on grouping functions. These functions, applied to our problem, measure the belongingness of a greylevel intensity to the background or to the object of the image. In this way, we choose the threshold associated with the highest grouping value to segment the image. One of the advantages of our proposal is avoiding the selection of a suitable grouping function for each image, by means of a convex combination of several of them.

References

1. Sahoo, P.K., Soltani, S., Wong, A.K.C., Chen, Y.C.: A survey of thresholding techniques. Comput. Vis. Graph. Image Process. **41**, 233–260 (1988)
2. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electron. Imaging **13**, 146–168 (2004)
3. Otsu, N.: A threshold selection method from gray level histograms. IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979)
4. Bustince, H., Barrenechea, E., Pagola, M.: Image thresholding using restricted equivalence functions and maximizing the measure of similarity. Fuzzy Sets Syst. **158**, 496–516 (2007)
5. Huang, L.K., Wang, M.J.J.: Image thresholding by minimizing the measure of fuzziness. Pattern Recogn. **28**, 41–51 (1995)
6. Bustince, H., Fernandez, J., Mesiar, R., Montero, J., Orduna, R.: Overlap functions. Nonlinear Anal. Theor. Meth. Appl. **78**, 1488–1499 (2010)
7. Trillas, E.: Sobre funciones de negación en la teoría de conjuntos difusos. Stochastica **III**(1), 47–59 (1979)
8. Alsina, C., Frank, M., Schweizer, B.: Associative Functions. Triangular Norms and Copulas. World Scientific, Singapore (2006)
9. Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: a guide for practitioners. Studies in Fuzziness and Soft Computing, vol. 221. Springer, Heidelberg (2007)

10. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms, trends in logic. Studia Logica Library, vol. 8. Kluwer Academic Publishers, Dordrecht (2000)
11. Bustince, H., Barrenechea, E., Pagola, M.: Restricted equivalence functions. Fuzzy Sets Syst. **157**, 2333–2346 (2006)
12. Jurio, A., Bustince, H., Pagola, M., Pradera, A., Yager, R.R.: Some properties of overlap and grouping functions and their application to image thresholding. Fuzzy Sets Syst. **229**, 69–90 (2013)
13. Caldairou, B., Passat, N., Habas, P., Studholme, C., Rousseau, F.: A non-local fuzzy segmentation method: application to brain MRI. Pattern Recogn. **44**(9), 1916–1927 (2011)
14. Mayer, A., Greenspan, H.: An adaptive mean-shift framework for MRI brain segmentation. IEEE Trans. Med. Imaging **28**(8), 105–118 (2009)

Part IV
Fuzzy Control, and Intelligent Monitoring

Noise Technologies for Operating the System for Monitoring of the Beginning of Violation of Seismic Stability of Construction Objects

T. A. Aliev, N. F. Musayeva and U. E. Sattarova

Abstract We analyzed the difficulties of operating the system for monitoring of seismic stability and technical condition of building structures. Technologies are proposed for solving the problem of monitoring at early stages of violation of seismic stability with application of sets of informative attributes consisting of variance, correlation, static and dynamic noise characteristics. The results of computational experiments are given.

1 Introduction

It is known that a system of monitoring of seismic stability and prediction of changes in the technical condition of high-rise buildings and building structures generally has to contain the following subsystems: primary sensors subsystem; primary data transmission and collection subsystem; subsystem for processing of obtained data and submitting the processing results to the specialized service [1].

Primary data transmission and collection subsystem is to provide data digitalization and transmission from all sensors to the local server, where all data is stored. Data digitalization provides acceptable resolution and frequency for each type of sensors. The server receives requests, processes the received requests and provides only authorized external access. The subsystem for data processing and

T. A. Aliev
Institute of Cybernetics, Baku, Azerbaijan
e-mail: telmancyber@rambler.ru

N. F. Musayeva (✉) · U. E. Sattarova
Azerbaijan University of Architecture and Construction, Baku, Azerbaijan
e-mail: musanaila@gmail.com

U. E. Sattarova
e-mail: ulker.rzaeva@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_14,

submittal contains the software providing visualized representation of processed primary data for the operator. Operator position subsystem provides information on the state of the control object by operator's request.

Only correct and adequate processing of signals that come from sensors measuring vibration, oscillation, tilts, deflation, rolls, strain in building structures, etc. and their timely submittal to the operator will ensure carrying out of operational complex of measures on preventing early wear, damage, and defects, such as cracks, bends, tilts, deformation, deflation, etc., and allow controlling physical ageing and obsolescence. On the other hand, in most cases for many high-rise buildings, there are certain troubles in prediction of changes in the object condition at early stages of fault origin with application of known methods of calculation of dispersion, correlation, spectral, static and dynamic characteristics. They allow one to detect only explicit faults at best. Analysis of emergency origin demonstrates that emergency situations are always preceded by hidden microfaults that emerge as microwears, microsagging, microvibration, microcracks, etc. in some units of the investigated building structure under. Their timely detection makes it possible to predict possible changes in the condition of a building structure, which can be used for warning and prevention of grave failures. The paper thereby considers one of alternative solutions to the problem of monitoring of seismic stability and the technical condition of high-rise buildings, building structures and strategic objects in seismically active regions with application of the technology and methods for development of robust sets of informative attributes.

2 Problem Statement

As is known, the prediction process can be generally represented as a combination of three elements: (1) pattern set Z formed from estimates of statistical characteristics, auto- and cross-correlation functions, spectral characteristics, static and dynamic characteristic corresponding to each i-state of all k of possible states of the building structure; (2) set V formed from current similar informative attributes that carry information on the current state; (3) rules of identification of F that compares each element of set Z to an element of set V, and vice versa, compares each element of set V to an element of set Z. The totality of elements of Z and V forms data support of prediction subsystem.

In this case, set Z_X corresponding to the normal state of a building structure, when signals $X_i(t)$ are received from sensors, can be represented as follows:

$$Z_X = \begin{bmatrix} D(\varepsilon_1) = 0 & D(\varepsilon_2) = 0 & \ldots & D(\varepsilon_n) = 0 \\ R_{\overset{\circ}{X}_1 \overset{\circ}{X}_1}(\mu) & R_{\overset{\circ}{X}_1 \overset{\circ}{X}_2}(\mu) & \ldots & R_{\overset{\circ}{X}_1 \overset{\circ}{X}_n}(\mu) \\ R_{\overset{\circ}{X}_2 \overset{\circ}{X}_1}(\mu) & R_{\overset{\circ}{X}_2 \overset{\circ}{X}_2}(\mu) & \ldots & R_{\overset{\circ}{X}_2 \overset{\circ}{X}_n}(\mu) \\ \ldots & \ldots & \ldots & \ldots \\ R_{\overset{\circ}{X}_n \overset{\circ}{X}_1}(\mu) & R_{\overset{\circ}{X}_n \overset{\circ}{X}_2}(\mu) & \ldots & R_{\overset{\circ}{X}_n \overset{\circ}{X}_n}(\mu) \\ a_{11} & a_{12} & \ldots & a_{1n} \\ b_{11} & b_{12} & \ldots & b_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{k1} & a_{k2} & \ldots & a_{kn} \\ b_{k1} & b_{k2} & \ldots & b_{kn} \\ c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ c_{m1} & c_{m2} & \ldots & c_{mn} \\ W_{11} & W_{12} & \ldots & W_{1n} \\ W_{21} & W_{22} & \ldots & W_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ W_{m1} & W_{m2} & \ldots & W_{mn} \\ & & \ldots & \end{bmatrix}, \tag{1}$$

where $D(\varepsilon_i) = 0$ is the value of variance of noise $\varepsilon_i(t)$, which in the normal operation of the building structure equals zero; $R_{\overset{\circ}{X}_i \overset{\circ}{X}_j}(\mu),\ i,j = \overline{1,n}$ are estimates of auto- and cross-correlation functions; $a_{ij},\ b_{ij},\ i = \overline{1,k},\ j = \overline{1,n}$ are Fourier spectral expansion coefficients; $c_{ij},\ i = \overline{1,m},\ j = \overline{1,n}$ are coefficients of results of solving static identification problem; $W_{ij}, i = \overline{1,m},\ j = \overline{1,n}$ are transfer functions; n is the quantity of input parameters, m is the quantity of output parameters, k is the quantity of coefficient in Fourier series.

Sets, which are similar to set Z_X, are composed for each deviation from the normal state of the building structure and allow predicting a failure state as a result of identification.

However, signals $g_i(t) = X_i(t) + \varepsilon_i(t)$ received from sensors measuring vibration, oscillation, tilts, deflation, rolls, etc. are in most cases distorted with noise, i.e. stationary state, normalcy of distribution law are violated in those signals, correlation between the useful signal and the noise is absent, etc. Then, instead of set Z_X, set Z_g is obtained, elements of which contain noise errors:

$$Z_g = \begin{bmatrix} D(\varepsilon_1) & D(\varepsilon_2) & \ldots & D(\varepsilon_n) \\ R_{\overset{\circ}{g}_1 \overset{\circ}{g}_1}(\mu) & R_{\overset{\circ}{g}_1 \overset{\circ}{g}_2}(\mu) & \ldots & R_{\overset{\circ}{g}_1 \overset{\circ}{g}_n}(\mu) \\ R_{\overset{\circ}{g}_2 \overset{\circ}{g}_1}(\mu) & R_{\overset{\circ}{g}_2 \overset{\circ}{g}_2}(\mu) & \ldots & R_{\overset{\circ}{g}_2 \overset{\circ}{g}_n}(\mu) \\ \ldots & \ldots & \ldots & \ldots \\ R_{\overset{\circ}{g}_n \overset{\circ}{g}_1}(\mu) & R_{\overset{\circ}{g}_n \overset{\circ}{g}_2}(\mu) & \ldots & R_{\overset{\circ}{g}_n \overset{\circ}{g}_n}(\mu) \\ a_{11}^* & a_{12}^* & \ldots & a_{1n}^* \\ b_{11}^* & b_{12}^* & \ldots & b_{1n}^* \\ a_{21}^* & a_{22}^* & \ldots & a_{2n}^* \\ b_{21}^* & b_{22}^* & \ldots & b_{2n}^* \\ \ldots & \ldots & \ldots & \ldots \\ a_{k1}^* & a_{k2}^* & \ldots & a_{kn}^* \\ b_{k1}^* & b_{k2}^* & \ldots & b_{kn}^* \\ c_{11}^* & c_{12}^* & \ldots & c_{1n}^* \\ c_{12}^* & c_{22}^* & \ldots & c_{2n}^* \\ \ldots & \ldots & \ldots & \ldots \\ c_{m1}^* & c_{m2}^* & \ldots & c_{mn}^* \\ W_{11}^* & W_{12}^* & \ldots & W_{1n}^* \\ W_{21}^* & W_{22}^* & \ldots & W_{2n}^* \\ \ldots & \ldots & \ldots & \ldots \\ W_{m1}^* & W_{m2}^* & \ldots & W_{mn}^* \\ & & \ldots & \end{bmatrix}, \quad (2)$$

where $D(\varepsilon_i) = 0$ is the value of variance of noise $\varepsilon_i(t)$, which is different from zero in the presence of noise;

$R_{\overset{\circ}{g}_i \overset{\circ}{g}_j}(\mu) = R_{\overset{\circ}{X}_i \overset{\circ}{X}_j}(\mu) + \Lambda_{\overset{\circ}{X}_i \overset{\circ}{X}_j}(\mu),\ i,j = \overline{1,n}$ are estimates of auto- and cross-correlation functions of noisy signals, $\Lambda_{\overset{\circ}{X}_i \overset{\circ}{X}_j}(\mu)$ is the value of errors of estimates $R_{\overset{\circ}{g}_i \overset{\circ}{g}_j}(\mu)$; $a_{ij}^* = a_{ij} + \lambda_{a_{ij}},\ b_{ij}^* = b_{ij} + \lambda_{b_{ij}},\ i = \overline{1,k},\ j = \overline{1,n}$ are Fourier spectral expansion coefficients for noisy signals, $\lambda_{a_{ij}},\ \lambda_{b_{ij}}$ is the value of errors of coefficients $a_{ij}, b_{ij}; c_{ij}^* = c_{ij} + \lambda_{c_{ij}},\ i = \overline{1,m},\ j = \overline{1,n}$ are coefficients of results of solving identification problem of noisy signals static, $\lambda_{c_{ij}}$ is the value of error $c_{ij}; W_{ij}^* = W_{ij} + \lambda_{W_{ij}},\ i = \overline{1,m},\ j = \overline{1,n}$ are transfer functions of noisy signals, $\lambda_{W_{ij}}$ is the value of error of estimates of are transfer functions W_{ij}; n is the quantity of input parameters, m is the quantity of output parameters, k is the quantity of coefficient in Fourier series.

Building the sets V_{g_i}, which characterize the current state of the construction object, presents similar difficulties.

Works [1–8] therefore offer algorithms for robust correlation and spectral monitoring, monitoring with application of noise characteristics, as well as algorithms for improvement of conditioning of correlation matrices in solving of static identification problems and problems of dynamics of high-rise buildings, construction and strategic objects in seismically active regions. Using those algorithms, one can build robust matrices Z_X^R, $V_{X_i}^R$, elements of which are close to elements of matrices Z_X, V_{X_i}. Considered below are algorithms for building robust matrices Z_X^R, $V_{X_i}^R$.

3 Technology and Algorithms for Monitoring of High-Rise Buildings, Construction and Strategic Objects in Seismically Active Regions by Means of Sets of Informative Attributes

Robust matrices Z_X^R, $V_{X_i}^R$ should be built in four stages. At the first stage, estimates of values of noise variances $D^*(\varepsilon_i)$ and robust estimates of auto- and cross-correlation functions $R^R_{\overset{\circ}{X}_i\overset{\circ}{X}_i}(\mu)$. Robust coefficients a_n^R, b_n^R of Fourier spectral expansion are calculated at the second stage. At the third stage, with application of robust correlation matrices $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)$, $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{Y}}(0)$ formed from robust estimates of auto- and cross-correlation functions $R^R_{\overset{\circ}{X}_i\overset{\circ}{X}_i}(0)$, $R^R_{\overset{\circ}{X}_i\overset{\circ}{Y}}(0)$, static identification problem is solved and robust coefficients c_{ij}^R are calculated. At the fourth stage, with application of robust correlation matrices $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{Y}}(\mu)$ formed from robust estimates of auto- and cross-correlation functions $R^R_{\overset{\circ}{X}_i\overset{\circ}{X}_i}(\mu)$, $R^R_{\overset{\circ}{X}_i\overset{\circ}{Y}}(\mu)$, dynamic identification problem is solved and robust W_{ij}^R transfer functions are calculated.

First stage. Calculation of values of noise variances $D^*(\varepsilon_i)$ and robust estimates of auto- and cross-correlation functions $R^R_{\overset{\circ}{X}_i\overset{\circ}{X}_i}(\mu)$.

1. Values of autocorrelation functions $R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu)$, $i=\overline{1,n}$ at time shift $\mu = 0$ are calculated from the following formula:

$$R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu) = \frac{1}{N}\sum_{k=1}^{N} \overset{\circ}{g}_i(k\Delta t)\,\overset{\circ}{g}_i(k\Delta t). \tag{3}$$

2. Values of autocorrelation functions $R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu)$, $i=\overline{1,n}$ at time shift $\mu = 1\cdot i\Delta t$ are calculated from the following formula:

$$R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu) = \frac{1}{N}\sum_{k=1}^{N} \overset{\circ}{g}_i(k\Delta t)\,\overset{\circ}{g}_i((k+1)\Delta t). \tag{4}$$

3. Values of autocorrelation functions $R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu),\ i = \overline{1,n}$ at time shift $\mu = 2 \cdot i\Delta t$ are calculated from the following formula:

$$R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu) = \frac{1}{N}\sum_{k=1}^{N} \overset{\circ}{g}_i(k\Delta t)\, \overset{\circ}{g}_i((k+2)\Delta t). \tag{5}$$

4. Values of variances $D^*(\varepsilon_i)$ of noises $\varepsilon_i(t)$ are calculated from the following formula:

$$D^*(\varepsilon_i) = R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu = 0 \cdot \Delta t) - 2R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu = 1 \cdot \Delta t)R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu = 2 \cdot \Delta t). \tag{6}$$

5. Robust estimates of autocorrelation function $R^R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu = 0)$ are calculated by means of the following formula:

$$R^R_{\overset{\circ}{X}_i \overset{\circ}{X}_i}(\mu = 0) = R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu = 0) - D^*(\varepsilon_i). \tag{7}$$

Second stage. Calculation of robust coefficients a_n^R, b_n^R of Fourier spectral expansion.

Robust estimates a_n^R are determined in the following way.

1. Noise variance $D^*(\varepsilon)$ and the mean value of relative error of readings $\overline{\lambda}_{rel}$ are determined

$$\overline{\lambda}_{rel} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\overset{\circ}{g}(i\Delta t)\,\overset{\circ}{g}(i\Delta t) + \overset{\circ}{g}(i\Delta t)\,\overset{\circ}{g}((i+2)\Delta t) - 2\,\overset{\circ}{g}(i\Delta t)\,\overset{\circ}{g}((i+1)\Delta t)\right]}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\overset{\circ}{g}{}^2(i\Delta t)}}$$
$$= \frac{D^*(\varepsilon)}{D(g)}; \tag{8}$$

2. Values $\Pi^+,\ \Pi^-,\ N^+$ and N^- are determined

$$\Pi^+ = \overline{\overset{\circ}{g}(i\Delta t)\,\cos n\omega(i\Delta t)}^{\,+}, \tag{9}$$

$$\Pi^- = \overline{\overset{\circ}{g}(i\Delta t)\,\cos n\omega(i\Delta t)}^{\,-} \tag{10}$$

as well as the number of positive N^+ and negative N^- sums of $g^\circ(i\Delta t)\cos n\omega(i\Delta t)$.

3. Conditions $N^+ = N^-$ and $\Pi^+ = \Pi^-$ are checked, during fulfilment of which application of conventional algorithms is recommended.

   3.1. When conditions $N^+ \neq N^-$ and $\Pi^+ = \Pi^-$ hold true, formula for determination of robust estimates a_n^R is represented as follows

$$a_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t) - \frac{1}{4}\left|N_{a_n}^{+} - N_{a_n}^{-}\right| A(i\Delta t)\right\}, \tag{11}$$

where $A(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t)}$

3.2. When conditions $N^+ > N^-$ and $\Pi^+ \neq \Pi^-$ hold true, estimates a_n^R are determined from the expression

$$a_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t) - \frac{1}{4}\left(N_{a_n}^{+} - N_{a_n}^{-}\right)A^{+}(i\Delta t) - \frac{1}{2}\left[N - \left(N_{a_n}^{+} - N_{a_n}^{-}\right)\right]\left[A^{+}(i\Delta t) - A^{-}(i\Delta t)\right]\right\}, \tag{12}$$

where $A(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t)}$, $A^{+}(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t)}^{+}$, $A^{-}(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t)}^{-}$.

3.3. If $N^+ < N^-$ and $\Pi^+ \neq \Pi^-$ take place, estimates a_n^R are determined from the expressions:

$$a_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t) - \frac{1}{4}\left(N_{a_n}^{-} - N_{a_n}^{+}\right)A(i\Delta t) - \frac{1}{2}\left[N - \left(N_{a_n}^{-} - N_{a_n}^{+}\right)\right]\left[A^{+}(i\Delta t) - A^{-}(i\Delta t)\right]\right\}. \tag{13}$$

3.4. If conditions $N^+ = N^-$ and $\Pi^+ \neq \Pi^-$ hold true, estimates a_n^R are determined from the formula

$$a_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t) - \frac{1}{2}N\left[A^{+}(i\Delta t) - A^{-}(i\Delta t)\right]\right\}. \tag{14}$$

Robust estimates b_n^R are determined in the following way.

1. Noise variance $D^*(\varepsilon)$ and the mean value of relative error of readings $\overline{\lambda}_{rel}$ are determined from the expression (8).
2. Values Π^+, Π^- are determined

$$\Pi^{+} = \overline{\overset{\circ}{g}(i\Delta t)\ \sin n\omega(i\Delta t)}^{+}, \tag{15}$$

$$\Pi^{-} = \overline{\overset{\circ}{g}(i\Delta t)\ \sin n\omega(i\Delta t)}^{-} \tag{16}$$

as well as the number of positive N^+ and negative N^- sums of $g^{\circ}(i\Delta t)\sin n\omega(i\Delta t)$.

3. Conditions $N^+ = N^-$ and $\Pi^+ = \Pi^-$, are checked, during fulfilment of which application of conventional algorithms is recommended.

3.1 When conditions $N^{+} = N^{-}$ and $\Pi^{+} = \Pi^{-}$ hold true, formula for determination of robust estimates b_n^R is represented as follows

$$b_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t) - \frac{1}{4}\left|N_{b_n}^{+} - N_{b_n}^{-}\right| B(i\Delta t)\right\}, \tag{17}$$

where $B(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t)}$.

3.2 When conditions $N^{+} > N^{-}$ and $\Pi^{+} \neq \Pi^{-}$ hold true, estimates b_n^R are determined from the expression

$$b_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t) - \frac{1}{4}\left(N_{b_n}^{+} - N_{b_n}^{-}\right)B^{+}(i\Delta t) - \frac{1}{2}\left[N - \left(N_{b_n}^{+} - N_{b_n}^{-}\right)\right]\left[B^{+}(i\Delta t) - B^{-}(i\Delta t)\right]\right\}, \tag{18}$$

where $B(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t)}$, $B^{+}(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t)}^{+}$, $B^{-}(i\Delta t) = \overline{\lambda}_{rel}\ \overline{\overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t)}^{-}$.

3.3 If $N^{+} < N^{-}$ and $\Pi^{+} \neq \Pi^{-}$ take place, estimates b_n^R are determined from the expressions:

$$b_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t) - \frac{1}{4}\left(N_{b_n}^{-} - N_{b_n}^{+}\right)B(i\Delta t) - \frac{1}{2}\left[N - \left(N_{b_n}^{-} - N_{b_n}^{+}\right)\right]\left[B^{+}(i\Delta t) - B^{-}(i\Delta t)\right]\right\}. \tag{19}$$

3.4 If conditions $N^{+} = N^{-}$ and $\Pi^{+} \neq \Pi^{-}$ hold true, estimates b_n^R are determined from the formula

$$b_n^R = \frac{2}{N}\left\{\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t) - \frac{1}{2}N\left[B^{+}(i\Delta t) - B^{-}(i\Delta t)\right]\right\}. \tag{20}$$

Third stage. Building robust correlation matrices $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0)$, $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{Y}}(0)$, solving static identification problem and calculation of robust coefficients c_{ij}^R.

1. Values of autocorrelation functions $R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu)$, $i = \overline{1,n}$ at time shift $\mu = 0$ are calculated from the formula (3).
2. Values of autocorrelation functions $R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu)$, $i = \overline{1,n}$ at time shift $\mu = 1 \cdot i\Delta t$ are calculated from the formula (4).
3. Values of autocorrelation functions $R_{\overset{\circ}{g}_i\overset{\circ}{g}_i}(\mu)$, $i = \overline{1,n}$ at time shift $\mu = 2 \cdot i\Delta t$ are calculated from the formula (5).
4. Values of variances $D^{*}(\varepsilon_i)$ of noises $\varepsilon_i(t)$ are calculated from the formula (6).

5. Robust estimates of autocorrelation function $R^R_{\overset{\circ}{g}_i \overset{\circ}{g}_i}(\mu = 0)$ are calculated by means of the formula (7).
6. Robust correlation matrix $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)$ is formed by means of the following expression:

$$\vec{R}^R_{\overset{\circ}{X}\circ}(0) = \begin{bmatrix} R_{\overset{\circ}{g}_1\overset{\circ}{g}_1}(0) - D^*(\varepsilon_1) & R_{\overset{\circ}{g}_1\overset{\circ}{g}_2}(0) & \dots & R_{\overset{\circ}{g}_1\overset{\circ}{g}_n}(0) \\ R_{\overset{\circ}{g}_2\overset{\circ}{g}_1}(0) & R_{\overset{\circ}{g}_2\overset{\circ}{g}_2}(0) - D^*(\varepsilon_2) & \dots & R_{\overset{\circ}{g}_2\overset{\circ}{g}_n}(0) \\ \dots & \dots & \dots & \dots \\ R_{\overset{\circ}{g}_n\overset{\circ}{g}_1}(0) & R_{\overset{\circ}{g}_n\overset{\circ}{g}_2}(0) & \dots & R_{\overset{\circ}{g}_n\overset{\circ}{g}_n}(0) - D^*(\varepsilon_i) \end{bmatrix}$$
$$\approx \begin{bmatrix} R_{\overset{\circ}{X}_1\overset{\circ}{X}_1}(0) & R_{\overset{\circ}{X}_1\overset{\circ}{X}_2}(0) & \dots & R_{\overset{\circ}{X}_1\overset{\circ}{X}_n}(0) \\ R_{\overset{\circ}{X}_2\overset{\circ}{X}_1}(0) & R_{\overset{\circ}{X}_2\overset{\circ}{X}_2}(0) & \dots & R_{\overset{\circ}{X}_2\overset{\circ}{X}_n}(0) \\ \dots & \dots & \dots & \dots \\ R_{\overset{\circ}{X}_n\overset{\circ}{X}_1}(0) & R_{\overset{\circ}{X}_n\overset{\circ}{X}_2}(0) & \dots & R_{\overset{\circ}{X}_n\overset{\circ}{X}_n}(0) \end{bmatrix} \approx \vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0). \tag{21}$$

7. Similarly, correlation matrix $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{Y}}(0)$ is formed.
8. The following matrix equation is solved

$$\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0) \cdot \vec{C}^R = \vec{R}^R_{\overset{\circ}{X}\overset{\circ}{Y}}(0) \tag{22}$$

and robust coefficients c^R_{ij} of static identification equation are calculated from the expression

$$\vec{C}^R = \left[\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)\right]^{-1} \vec{R}^R_{\overset{\circ}{X}\overset{\circ}{Y}}(0). \tag{23}$$

Fourth stage. Building robust correlation matrices $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{Y}}(\mu)$, solving dynamic identification problem and calculation of robust transfer functions W^R_{ij}.

1. Value of autocorrelation functions $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ at time shift $\mu = 0$ is calculated from the following formula:

$$R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) = \frac{1}{N}\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\, \overset{\circ}{g}(i\Delta t). \tag{24}$$

2. Value of autocorrelation functions $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ at time shift $\mu = 1 \cdot i\Delta t$ is calculated from the following formula:

$$R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) = \frac{1}{N}\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\,\overset{\circ}{g}((i+1)\Delta t). \tag{25}$$

3. Value of autocorrelation functions $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ at time shift $\mu = 2 \cdot i\Delta t$ is calculated from the following formula:

$$R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) = \frac{1}{N}\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\,\overset{\circ}{g}((i+2)\Delta t). \tag{26}$$

4. Value of variance $D^{*}(\varepsilon)$ of noise is calculated from the following formula

$$D^{*}(\varepsilon) = R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu = 0 \cdot \Delta t) - 2R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu = 1 \cdot \Delta t) + R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu = 2 \cdot \Delta t). \tag{27}$$

5. Robust estimates of autocorrelation function $R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ are calculated by means of the following formula:

$$R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu = 0) = R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu = 0) - D^{*}(\varepsilon). \tag{28}$$

6. Robust correlation matrix $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0)$ is formed by means of the following expression:

$$\vec{R}^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) = \begin{Vmatrix} R_{\overset{\circ}{g}\overset{\circ}{g}}(0) - D^{*}(\varepsilon) & R_{\overset{\circ}{g}\overset{\circ}{g}}(\Delta t) & \ldots & R_{\overset{\circ}{g}\overset{\circ}{g}}[(N-1)\Delta t] \\ R_{\overset{\circ}{g}\overset{\circ}{g}}(\Delta t) & R_{\overset{\circ}{g}\overset{\circ}{g}}(0) - D^{*}(\varepsilon) & \ldots & R_{\overset{\circ}{g}\overset{\circ}{g}}[(N-2)\Delta t] \\ \ldots & \ldots & \ldots & \ldots \\ R_{\overset{\circ}{g}\overset{\circ}{g}}[(N-1)\Delta t] & R_{\overset{\circ}{g}\overset{\circ}{g}}[(N-2)\Delta t] & \ldots & R_{\overset{\circ}{g}\overset{\circ}{g}}(0) - D^{*}(\varepsilon) \end{Vmatrix}$$
$$\approx \begin{Vmatrix} R_{\overset{\circ}{X}\overset{\circ}{X}}(0) & R_{\overset{\circ}{X}\overset{\circ}{X}}(\Delta t) & \ldots & R_{\overset{\circ}{X}\overset{\circ}{X}}[(N-1)\Delta t] \\ R_{\overset{\circ}{X}\overset{\circ}{X}}(\Delta t) & R_{\overset{\circ}{X}\overset{\circ}{X}}(0) & \ldots & R_{\overset{\circ}{X}\overset{\circ}{X}}[(N-2)\Delta t] \\ \ldots & \ldots & \ldots & \ldots \\ R_{\overset{\circ}{X}\overset{\circ}{X}}[(N-1)\Delta t] & R_{\overset{\circ}{X}\overset{\circ}{X}}[(N-2)\Delta t] & \ldots & R_{\overset{\circ}{X}\overset{\circ}{X}}(0) \end{Vmatrix} \approx \vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu). \tag{29}$$

7. Similarly, correlation matrix $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{Y}}(\mu)$ is formed.
8. The following matrix equation is solved

$$\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\vec{W}(\mu) = \vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{Y}}(\mu) \tag{30}$$

and robust transfer functions W^{R}_{ij} of matrix equation of dynamic identification are calculated from the expression

$$\vec{W}_{ij}^{R} = \left[\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right]^{-1} \vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{Y}}(\mu). \tag{31}$$

After all four stages are complete, robustness conditions are provided, i.e. the following equalities hold true:

$$D^{*}(\varepsilon_i) \approx D(\varepsilon_i), \tag{32}$$

$$R^{R}_{\overset{\circ}{g_i}\overset{\circ}{g_j}}(\mu) \approx R_{\overset{\circ}{X_i}\overset{\circ}{X_j}}(\mu), \tag{33}$$

$$a_{ij}^{R} \approx a_{ij},\ b_{ij}^{R} \approx b_{ij}, \tag{34}$$

$$c_{ij}^{R} \approx c_{ij},\ W_{ij}^{R} \approx W_{ij}. \tag{35}$$

It allows forming robust correlation matrices Z_X^R, $V_{X_i}^R$

$$Z_X^R = \begin{bmatrix} D^{*}(\varepsilon_1) & D^{*}(\varepsilon_2) & \dots & D^{*}(\varepsilon_n) \\ R^{R}_{\overset{\circ}{X_1}\overset{\circ}{X_1}}(\mu) & R^{R}_{\overset{\circ}{X_1}\overset{\circ}{X_2}}(\mu) & \dots & R^{R}_{\overset{\circ}{X_1}\overset{\circ}{X_n}}(\mu) \\ R^{R}_{\overset{\circ}{X_2}\overset{\circ}{X_1}}(\mu) & R^{R}_{\overset{\circ}{X_2}\overset{\circ}{X_2}}(\mu) & \dots & R^{R}_{\overset{\circ}{X_2}\overset{\circ}{X_n}}(\mu) \\ \dots & \dots & \dots & \dots \\ R^{R}_{\overset{\circ}{X_n}\overset{\circ}{X_1}}(\mu) & R^{R}_{\overset{\circ}{X_n}\overset{\circ}{X_2}}(\mu) & \dots & R^{R}_{\overset{\circ}{X_n}\overset{\circ}{X_n}}(\mu) \\ a_{11}^R & a_{12}^R & \dots & a_{1n}^R \\ b_{11}^R & b_{12}^R & \dots & b_{1n}^R \\ a_{21}^R & a_{22}^R & \dots & a_{2n}^R \\ b_{21}^R & b_{22}^R & \dots & b_{2n}^R \\ \dots & \dots & \dots & \dots \\ a_{k1}^R & a_{k2}^R & \dots & a_{kn}^R \\ b_{k1}^R & b_{k2}^R & \dots & b_{kn}^R \\ c_{11}^R & c_{12}^R & \dots & c_{1n}^R \\ c_{12}^R & c_{22}^R & \dots & c_{2n}^R \\ \dots & \dots & \dots & \dots \\ c_{m1}^R & c_{m2}^R & \dots & c_{mn}^R \\ W_{11}^R & W_{12}^R & \dots & W_{1n}^R \\ W_{21}^R & W_{22}^R & \dots & W_{2n}^R \\ \dots & \dots & \dots & \dots \\ W_{m1}^R & W_{m2}^R & \dots & W_{mn}^R \\ & & \dots & \end{bmatrix}, \tag{36}$$

for which the following equalities hold true:

$$Z_X^R \approx Z_X, \tag{37}$$

$$V_{X_i}^R \approx V_{X_i}. \tag{38}$$

Thus, as a result of application of robust technology for formation of state matrices Z_X^R, $V_{X_i}^R$, a possibility arises to detect faults at early stages for reliable prediction of the technical condition of high-rise buildings or building structure.

4 The Results of Computational Experiments

Computational experiments have been carried out to prove the validity of the developed algorithms. The experiments were carried out in the following way.

First, a technology was developed for modeling the signals coming from the sensors installed in the most informative spots of high-rise buildings and building structures; the results of their processing are used to estimate the seismic stability and technical condition of high-rise buildings, construction objects and strategic objects. It is demonstrated that according to their nature, the signals received from the sensors can be divided into four groups: (1) deterministic, when the signal is non-random, undistorted and is some function of time; time relationship of the signal can be defined with any mathematical expression; (2) deterministic (non-random) signal distorted by a noise, which is a random function; (3) random signal not containing distortions; (4) random signal distorted by a noise, which is a random function.

Therefore, we model a bank of useful signals $X(t)$ with different sampling intervals Δt in the form of sums of sine functions and cosine functions with specified amplitudes a_k, b_k and frequencies $k\omega_0$:

$$X(i\Delta t) = \sum_{k=1}^{k_{end}} [a_k \cos(k\omega_0(i\Delta t)) + b_k \sin(k\omega_0(i\Delta t))]. \tag{39}$$

The bank of useful signals $X(t)$ and noises $\varepsilon(t)$ is also modeled by means of the random number generator in the form of random sequences with beta-, exponential-, gamma-, normal-, rayleigh- and weibull distribution laws at different values of parameters of those distributions Noisy signals $g(i\Delta t) = X(i\Delta t) + \varepsilon(i\Delta t)$ are formed with different distribution laws and "useful signal-noise". Thus, the bank contains both the signals, for which the classical conditions hold true, and the signals, for which the classical conditions are violated. Charts and histograms have been built for all of those signals. MATLAB computing environment was used in the modeling of the signals of all four groups and in the computational experiments.

We developed a technology for experimental research of random signals by means of noise characteristics. In each computational experiment, variances were calculated for random noises, using the traditional algorithm, i.e. the formula

$$D(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N} \overset{\circ}{\varepsilon}(i\Delta t)\,\overset{\circ}{\varepsilon}(i\Delta t). \tag{40}$$

Then, the variance $D^{*}(\varepsilon)$ of the noise $\varepsilon(v\Delta t)$ was calculated for the noisy signal $g(v\Delta t)$ from the formula (6). The value of relative error was taken as the indicator of efficiency of the experiment:

$$proc = (D(\varepsilon) - D^{*}(\varepsilon))/D(\varepsilon)^{*}100\%. \tag{41}$$

Given below are the results of one experiment with the useful signal $x0 = 40$ sin (t) and the random noise complying with exponential-distribution $f(\varepsilon\|\mu) = \frac{1}{\mu}e^{-\frac{\varepsilon}{\mu}}$with parameter $\mu = 5$

Our numerous experiments confirmed the efficiency of the developed algorithm.

A technology was developed and the results of computational experiments are given for correlation monitoring of seismic stability and technical condition of high-rise buildings, construction objects and strategic objects in seismically active regions.

To conduct the computational experiments, the following were calculated: (1) estimates of the correlation functions $R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ of the useful signal $X(t)$ and the noisy signal $g(t)$; (2) the value of variance $D^{*}(\varepsilon)$ of the noise $\varepsilon(i\Delta t)$ from the expression (27); (3) robust estimates of the normalized auto-correlation function from the expression (28). The comparative analysis followed, and the values of relative errors of autocorrelation functions $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ of the noisy signals $g(t)$ and robust estimates $R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ were calculated from the following expressions:

$$\Delta R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu) = \left|\frac{R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) - R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)}{R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)}\right| \cdot 100\%; \tag{42}$$

$$\Delta R^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu) = \frac{\left|R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) - R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right|}{R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)} \cdot 100\%. \tag{43}$$

Given below are the results of the computational experiment with the input useful signal $X(i\Delta t) = 40$ sin $(i\Delta t) + 100$ and the noise $\varepsilon(t)$ complying with the normal distribution law with mathematical expectation $m_{\varepsilon} \approx 0$ and variance $D(\varepsilon) \approx 90$ (Table 2), i.e. the classical conditions are observed. In the computational experiment with the useful signal $X(i\Delta t) = 50$ sin $(i\Delta t) + 5$ cos $(0.5 \cdot i\Delta t) + 2$ sin $(i\Delta t) + 5$ cos $(13 \cdot i\Delta t) + 15$, the noise $\varepsilon(t)$ complying with the lognormal distribution with mathematical expectation $m_{\varepsilon} \approx 0$ and variance $D(\varepsilon) \approx 665$ (Table 3), the classical conditions are violated.

Table 1 Results of the calculation of the variance estimate for the noise of the noisy signal

N	Estimates	$\Delta t = \pi/50$	$\Delta t = \pi/100$	$\Delta t = \pi/200$	$\Delta t = \pi/400$
1	$D(\varepsilon)$	27.7153	23.6353	28.2664	25.9530
2	$D^{*}(\varepsilon)$	23.8590	25.0362	26.8281	25.6202
3	*proc,%*	15.5922	5.9272	5.0884	1.2820

Table 2 Results of the calculation of the robust autocorrelation function with the classical conditions for the useful signal and noise holding

	$R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	$R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	$R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	$\Delta R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	$\Delta^{R} R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$
1	2	3	4	5	6
Experiment N1	800.000	888.7157	808.0540	11.0895	1.0067
$D(\varepsilon) = 89.7642$	799.6052	801.1015	801.1015	0.1871	0.1871
	798.4214	794.1490	794.1490	0.5351	0.5351
$D_{\varepsilon} = 80.6617$	796.4496	798.2796	798.2796	0.2298	0.2298
	793.6918	790.7896	790.7896	0.3657	0.3657

Table 3 Results of the calculation of the robust autocorrelation function with the classical conditions for the useful signal and noise violated

	$R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	$R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	$R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	$\Delta R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	$\Delta^{R} R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$
1	2	3	4	5	6
Experiment N6	1377.0	1963.9	1366.1	42.6213	0.7927
$D(\varepsilon) = 665.69$	1376.6	1320.6	1320.6	4.0687	4.0687
	1375.3	1275.0	1275.0	7.2900	7.2900
$D^{*}(\varepsilon) = 597.8$	1373.2	1317.9	1317.9	4.0313	4.0313
	1370.4	1279.5	1279.5	6.6343	6.6343

The following conclusions were made based on the analysis of the obtained results.

(1) The conditions (Tables 1 and 2, Column 1)

$$D^{*}(\varepsilon) \approx D(\varepsilon) \tag{44}$$

hold true.

(2) For the estimates of the correlation functions $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ of the noisy signals, the following inequality holds true (Tables 1 and 2, Columns 2, 3)

$$R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu) \neq R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu). \tag{45}$$

(3) For the robust estimates of the correlation functions $R^{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$, the equality (11) holds true (Tables 2 and 3, Columns 3, 4).

A technology was developed and the results of computational experiments are given for spectral monitoring of seismic stability and technical condition of high-rise buildings, construction objects and strategic objects in seismically active regions. For this purpose, the following were calculated:

(1) estimates of the coefficients of Fourier series of the useful signal, using the conventional algorithms:

$$a_0 = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{X}(i\Delta t), \tag{46}$$

$$a_n = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{X}(i\Delta t)\cos n\omega(i\Delta t), \tag{47}$$

$$b_n = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{X}(i\Delta t)\sin n\omega(i\Delta t); \tag{48}$$

(2) estimates of the coefficients of Fourier series of the noisy signal, using the conventional algorithms:

$$a_n^* = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\cos n\omega(i\Delta t), \tag{49}$$

$$b_n^* = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{g}(i\Delta t)\sin n\omega(i\Delta t), \tag{50}$$

(3) actual errors of estimates of the coefficients of Fourier series a_n^*, b_n^* from the following expression:

$$\lambda_{a_n} = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{\varepsilon}(i\Delta t)\cos n\omega(i\Delta t), \tag{51}$$

$$\lambda_{b_n} = \frac{2}{N}\sum_{i=1}^{N} \overset{\circ}{\varepsilon}(i\Delta t)\sin n\omega(i\Delta t); \tag{52}$$

(4) robust estimates of the coefficients of Fourier series of the noisy signals, using the unconventional algorithms (8)–(20);

(5) relative errors of estimates of the coefficients of Fourier series of the noisy signals and robust estimates of the coefficients of Fourier series:

Table 4 Results of the calculation of the robust estimate of the coefficient $\mathbf{a}_n$ of spectral decomposition of the noisy signal

n	ϖ	a_n	$\overset{*}{a_n}$	a_n^R	$N_{b_n}^+ - N_{b_n}^-$	$N\,P_{a_n}$	$P_{a_n}^+$	$P_{a_n}^-$
0	0	0.000	−0.000	0.000	0	0.00	0.00	0.00
25	$\pi/100$	10.010	11.176	9.837	450	27940.402	28.926	22.367
50	$2\pi/100$	0.020	0.954	0.013	256	2383.933	24.692	26.352
75	$3\pi/100$	20.000	23.977	18.982	1622	59942.184	30.398	24.101
100	$4\pi/100$	−0.130	−1.549	−0.189	−550	−3873.065	28.699	24.406
125	$5\pi/100$	30.000	35.628	28.853	2174	89070.086	30.764	15.059

Table 5 Results of the calculation of the robust estimate of the coefficient $\mathbf{b}_n$ of spectral decomposition of the noisy signal

n	ϖ	b_n	$\overset{*}{b_n}$	b_n^R	$N_{b_n}^+ - N_{b_n}^-$	$N\,P_{b_n}$	$P_{b_n}^+$	$P_{b_n}^-$
0	0	0.000	−0.000	0.000	0	0.00	0.00	0.00
25	$\pi/100$	15.000	17.75	14.913	1024	44367.03	27.31	19.03
50	$2\pi/100$	25.000	29.46	24.093	1847	73,641.44	28.67	15.37
75	$3\pi/100$	−0.130	−0.41	−0.174	−108	−1028.90	22.29	21.75
100	$4\pi/100$	−20.00	−24.30	−19.66	−1665	−60740.66	18.25	27.47
125	$5\pi/100$	−0.050	1.48	0.040	513	3694.92	22.60	26.19

$$\Delta a = \left|a_n - a_n^*\right| / a_n * 100,\ \Delta b = \left|b_n - b_n^*\right| / b_n * 100\%; \tag{53}$$

$$\Delta a^R = \left|a_n - a_n^R\right| / a_n * 100,\ \Delta b^R = \left|b_n - b_n^R\right| / b_n * 100\%. \tag{54}$$

Given below are the results of the computational experiment with the input useful signal in the form of the sum of harmonic oscillations

$X(i\Delta t) = 10\cos(i\Delta t) + 20\cos(3 \cdot i\Delta t) + 30\cos(5 \cdot i\Delta t) + 15\sin(i\Delta t) + 25\sin(2 \cdot i\Delta t) - 20\sin(4 \cdot i\Delta t) + 35$, with the noise $\varepsilon(t)$ complying with the normal distribution law with mathematical expectation $m_\varepsilon \approx 0$ and variance $D(\varepsilon) \approx 130$, the value of the relative error of the readings $g(i\Delta t)$ is $\lambda_{rel} = D^*(\varepsilon)/D(g) = 0.253$ (Tables 4 and 5).

Thus, application of the robust technology of spectral analysis allows one to ensure that the equality (34) holds true.

A technology was developed and the results of computational experiments are given for improving the conditionality of correlation matrices in solving the static identification problems of high-rise buildings, construction objects and strategic objects in seismically active regions. For this purpose, the following were calculated:

(1) estimates of the correlation functions of the useful and noisy input and output signals $R_{\overset{\circ}{X}_k \overset{\circ}{X}_l}(\mu)$, $R_{\overset{\circ}{X}_k \overset{\circ}{y}}(\mu)$, $R_{\overset{\circ}{y}\overset{\circ}{y}}(\mu)$, $R_{\overset{\circ}{g}_k \overset{\circ}{g}_l}(\mu)$, $R_{\overset{\circ}{g}_k \overset{\circ}{\eta}}(\mu)$, $R_{\overset{\circ}{\eta}\overset{\circ}{\eta}}(\mu)$ $k, l = \overline{1, n}$;

(2) robust estimates of auto-correlation functions of noisy input and output signals $R^R_{\overset{\circ}{X}_k \overset{\circ}{X}_k}(\mu)$, $R^R_{\overset{\circ}{X}_k \overset{\circ}{y}}(\mu)$ from the expression (7);

(3) matrices $\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0)$, $\vec{R}_{\overset{\circ}{X}\overset{\circ}{y}}(0)$, $\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0)$, $\vec{R}_{\overset{\circ}{g}\overset{\circ}{\eta}}(0)$ of the useful and noisy signals, the robust matrix $\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)$ were formed from the expression (21), their determinants $\Delta_{\overset{\circ}{X}\overset{\circ}{X}}(0)$, $\Delta_{\overset{\circ}{g}\overset{\circ}{g}}(0)$, $\Delta^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)$ and the condition numbers $H\left(\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0)\right)$, $H\left(\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0)\right)$, $H\left(\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)\right)$ were calculated;

(5) coefficients of the mathematical model of the useful and noisy signals, robust estimates were calculated from the expressions (22)–(23).

A comparative analysis was carried out. For that purpose, the following were calculated:

(1) the values of relative errors for the estimates of the elements of correlation matrices of noisy signals and robust estimates $\Delta R_{\overset{\circ}{g}_k\overset{\circ}{g}_l}(0)$, $\Delta R_{\overset{\circ}{g}_k\overset{\circ}{\eta}}(0)$, $\Delta R^R_{\overset{\circ}{X}_k\overset{\circ}{X}_l}(0)$, $\Delta R^R_{\overset{\circ}{x}_k\overset{\circ}{y}}(0)$ from the expressions (42), (43);

(2) matrices of relative errors for the elements of correlation matrices of noisy signals and robust correlation matrices $\Delta\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0)$, $\Delta\vec{R}_{\overset{\circ}{g}\overset{\circ}{\eta}}$, $\Delta\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)$, $\Delta\vec{R}^R_{\overset{\circ}{X}\overset{\circ}{y}}(0)$;

(3) relative errors of the determinants $p\Delta_{\overset{\circ}{g}\overset{\circ}{g}}(0)$ and $p\Delta^R_{\overset{\circ}{X}\overset{\circ}{X}}(0)$ of the correlation matrices of noisy signals and robust correlation matrices;

(4) relative errors of the coefficients of the mathematical model of noisy signals and robust coefficients of the model $pc_{k_{\overset{\circ}{g}\overset{\circ}{g}}}$, $pc^R_{k_{\overset{\circ}{x}\overset{\circ}{x}}}$.

The results of the computational experiment with violated classical conditions are given. Three useful input signals were formed $x_1(i\Delta t) = 30\sin(i\Delta t) + 13\cos(1.2\cdot i\Delta t - 3) - 35\sin(1.1\cdot i\Delta t) + 100$,$x_2(i\Delta t) = 50\sin(i\Delta t - 0.5) + 120$, $x_3(i\Delta t) = 70\sin(i\Delta t + 0.3) - 45\cos(0.4\cdot i\Delta t) + 150$ and the output signal $y(i\Delta t) = 120 + 10x_1(i\Delta t) + 15x_2(i\Delta t) - 10x_3(i\Delta t)$. The noise $\varepsilon_1(t)$ complies with the exponential distribution with mathematical expectation $m_{\varepsilon_1} \approx 7$ and mean square deviation $\sigma_{\varepsilon_1} \approx 7$. The noise $\varepsilon_2(t)$ complies with the gamma-distribution with mathematical expectation $m_{\varepsilon_2} \approx 9$ and mean square deviation $\sigma_{\varepsilon_2} \approx 5$. The noise $\varepsilon_3(t)$ complies with the normal distribution with mathematical expectation $m_{\varepsilon_3} \approx 0$ and mean square deviation $\sigma_{\varepsilon_3} \approx 25$. The noise $\phi(t)$ complies with the normal distribution with mathematical expectation $m_\phi \approx 0$ and mean square deviation $\sigma_\phi \approx 100$.

Based on the analysis of the obtained results, we conclude that the values of coefficients of the mathematical model of noisy signals differ substantially from those of coefficients of useful signals, and the values of robust coefficients of the mathematical model practically match those of coefficients of useful signals:

$$c_{k_{\overset{\circ}{g}\overset{\circ}{g}}} \neq c_{k_{\overset{\circ}{X}\overset{\circ}{X}}}, \tag{55}$$

Table 6 Results of the calculation of the robust correlation matrix in solving of the statics identification problems

$\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = \begin{bmatrix} 465.9 & 674 & 383.8 \\ 674 & 1250 & 1107.5 \\ 383.8 & 1107.5 & 3251.9 \end{bmatrix}$	$\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0) = \begin{bmatrix} 539.2 & 692 & 375.9 \\ 692 & 1288 & 1082.3 \\ 375.9 & 1082.3 & 3806 \end{bmatrix}$	$\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = \begin{bmatrix} 486 & 692 & 375.9 \\ 692 & 1262.7 & 1082.3 \\ 375.9 & 1082.3 & 3257.5 \end{bmatrix}$
	$\Delta\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0) = \begin{bmatrix} 15.7437 & 2.6673 & 2.0647 \\ 2.6673 & 3.0434 & 2.2712 \\ 2.0647 & 2.2712 & 17.0386 \end{bmatrix}$	$\Delta\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = \begin{bmatrix} 4.3188 & 2.6673 & 2.0647 \\ 2.6673 & 1.0150 & 2.2712 \\ 2.0647 & 2.2712 & 0.1720 \end{bmatrix}$
$\vec{R}_{\overset{\circ}{X}\overset{\circ}{y}}(0) = \begin{bmatrix} 10930 & 14420 & -12070 \end{bmatrix}^{T}$	$\vec{R}_{\overset{\circ}{g}\overset{\circ}{\eta}}(0) = \begin{bmatrix} 11350 & 14650 & -12400 \end{bmatrix}^{T}$	$\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{y}}(0) = \begin{bmatrix} 11350 & 14650 & -12400 \end{bmatrix}^{T}$
	$\Delta\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0) = \begin{bmatrix} 3.8577 & 1.6525 & 2.7756 \end{bmatrix}$	$\Delta\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{y}}(0)$ $\begin{bmatrix} 3.8577 & 1.6525 & 2.7756 \end{bmatrix}$
$H\left(\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0)\right) = 66.8700$	$H\left(\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(0)\right) = 39.6712$	$H\left(\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0)\right) = 63.8403$
$\Delta_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 2.3387e+008$	$\Delta_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 5.7031e+008$	$\Delta^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 2.5448e+008$
$\Delta 1_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 2.3387e+009$	$\Delta 1_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 6.4381e+009$	$\Delta 1^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 2.9239e+009$
$\Delta 2_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 3.5081e+009$	$\Delta 2_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 6.7348e+009$	$\Delta 2^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 3.4545e+009$
$\Delta 3_{\overset{\circ}{X}\overset{\circ}{X}}(0) = -2.3387e+09$	$\Delta 3_{\overset{\circ}{g}\overset{\circ}{g}}(0) = -4.4097e+009$	$\Delta 3^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = -2.4542e+09$
	$p\Delta_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 143.8550$	$p\Delta^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 8.8119$
	$p\Delta 1_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 175.2828$	$p\Delta 1^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 25.0222$
	$p\Delta 2_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 91.9798$	$p\Delta 2^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 1.5276$
	$p\Delta 3_{\overset{\circ}{g}\overset{\circ}{g}}(0) = 88.5518$	$p\Delta 3^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(0) = 4.9366$
$c0_{\overset{\circ}{X}\overset{\circ}{X}} = 120;\ c1_{\overset{\circ}{X}\overset{\circ}{X}} = 10;$	$c0_{\overset{\circ}{g}\overset{\circ}{g}} = -143.32; c1_{\overset{\circ}{g}\overset{\circ}{g}} = 11.2888$	$c0^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = -108.2; c1^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = 11.5$
$c2_{\overset{\circ}{X}\overset{\circ}{X}} = 15; c3_{\overset{\circ}{X}\overset{\circ}{X}} = -10$	$c2_{\overset{\circ}{g}\overset{\circ}{g}} = 11.81; c3_{\overset{\circ}{g}\overset{\circ}{g}} = -7.7321$	$c2^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = 13.6; c3^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = -9.64$
	$pc0_{\overset{\circ}{g}\overset{\circ}{g}} = 219.4335; pc1_{\overset{\circ}{g}\overset{\circ}{g}} = 12.8879$	$pc0^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = 190.18$
	$pc2_{\overset{\circ}{g}\overset{\circ}{g}} = 21.273; pc3_{\overset{\circ}{g}\overset{\circ}{g}} = 22.6787$	$pc1^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = 14.8976$
		$pc2^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = 9.5021$
		$pc3^{R}_{\overset{\circ}{X}\overset{\circ}{X}} = 3.5614$

$$c^{R}_{k_{\overset{\circ}{X}\overset{\circ}{X}}} \approx c_{k_{\overset{\circ}{X}\overset{\circ}{X}}}. \tag{56}$$

It means that the adequacy of the mathematical model is ensured (Table 6).

A technology was developed and the results of computational experiments are given for improving the conditionality of correlation matrices in solving the dynamic identification problems of high-rise buildings, construction objects and strategic objects in seismically active regions. For this purpose, the following were calculated:

(1) estimates of the correlation functions $R_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$, $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ from the expressions (24)–(28);
(2) matrices$\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ of the useful and noisy signals and the robust matrix $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ were formed from the expression (29);
(3) determinants $\Delta_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $\Delta_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$, $\Delta^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ and the condition numbers $H\left(\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right)$, $H\left(\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)\right)$, $H\left(\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right)$ of the matrices $\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$, $\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$, $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$.

The following were calculated for the purpose of comparative analysis:

(1) the values of relative errors for the estimates of the elements $\Delta R_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$, $\Delta R^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ of the correlation matrix of noisy signals and robust correlation matrix;
(2) matrices of relative errors for the elements of the correlation matrix of noisy signal$\Delta\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$ and robust correlation matrix $\Delta\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$.

Given below are the results of the computational experiments for solving the dynamic identification problem when the classical conditions do not hold true for the useful signal and the noise. The input useful signal $X(t)$ is obtained through interpolation by Fourier series downsampled with the parameter $r = 100$ of the normally distributed random sequence. The noise $\varepsilon(t)$ complies with the exponential distribution with mathematical expectation $m_{\varepsilon} \approx 10$ and mean square deviations $\sigma_{\varepsilon} \approx 10$. The correlation matrix was built for the values $\mu = 0, 1,\ldots, 9$. The results of the calculations are given in Table 7.

The analysis of the obtained results allows us to conclude that the correlation matrix of the noisy signal differs from the correlation matrix $\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ of the useful signal (Table 7, Rows 1, 2). The robust correlation matrix $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ practically matches the correlation matrix $\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ of the useful signal (Table 7, Rows 1, 3). The value of the determinant $\Delta^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ and the condition number $H\left(\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right)$ of the robust correlation matrix $\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ is closer to the value of the determinant $\Delta_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ and the condition number $H\left(\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right)$ of correlation matrix $\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$ of the useful signal (Table 7, Row 6).

Table 7 Results of the calculation of the robust correlation matrix in solving of the dynamics identification problems

$\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	773.6555	773.4497	772.9718	772.2215	771.1990	769.9044	768.3380	766.5004	764.3921	762.0139
	773.4497	773.6555	773.4497	772.9718	772.2215	771.1990	769.9044	768.3380	766.5004	764.3921
	772.9718	773.4497	773.6555	773.4497	772.9718	772.2215	771.1990	769.9044	768.3380	766.5004
	772.2215	772.9718	773.4497	773.6555	773.4497	772.9718	772.2215	771.1990	769.9044	768.3380
	771.1990	772.2215	772.9718	773.4497	773.6555	773.4497	772.9718	772.2215	771.1990	769.9044
	769.9044	771.1990	772.2215	772.9718	773.4497	773.6555	773.4497	772.9718	772.2215	771.1990
	768.3380	769.9044	771.1990	772.2215	772.9718	773.4497	773.6555	773.4497	772.9718	772.2215
	766.5004	768.3380	769.9044	771.1990	772.2215	772.9718	773.4497	773.6555	773.4497	772.9718
	764.3921	766.5004	768.3380	769.9044	771.1990	772.2215	772.9718	773.4497	773.6555	773.4497
	762.0139	764.3921	766.5004	768.3380	769.9044	771.1990	772.2215	772.9718	773.4497	773.6555
$\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	898.1711	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532	781.4268	780.9344	770.9113
	786.5662	898.1711	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532	781.4268	780.9344
	784.1673	786.5662	898.1711	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532	781.4268
	786.6325	784.1673	786.5662	898.1711	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532
	784.1410	786.6325	784.1673	786.5662	898.1711	786.5662	784.1673	786.6325	784.1410	784.4166
	784.4166	784.1410	786.6325	784.1673	786.5662	898.1711	786.5662	784.1673	786.6325	784.1410
	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	898.1711	786.5662	784.1673	786.6325
	781.4268	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	898.1711	786.5662	784.1673
	780.9344	781.4268	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	898.1711	786.5662
	770.9113	780.9344	781.4268	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	898.1711
$\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	788.9652	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532	781.4268	780.9344	770.9113
	786.5662	788.9652	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532	781.4268	780.9344
	784.1673	786.5662	788.9652	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532	781.4268
	786.6325	784.1673	786.5662	788.9652	786.5662	784.1673	786.6325	784.1410	784.4166	777.9532
	784.1410	786.6325	784.1673	786.5662	788.9652	786.5662	784.1673	786.6325	784.1410	784.4166
	784.4166	784.1410	786.6325	784.1673	786.5662	788.9652	786.5662	784.1673	786.6325	784.1410
	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	788.9652	786.5662	784.1673	786.6325
	781.4268	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	788.9652	786.5662	784.1673
	780.9344	781.4268	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	788.9652	786.5662
	770.9113	780.9344	781.4268	777.9532	784.4166	784.1410	786.6325	784.1673	786.5662	788.9652

(continued)

Table 7 (continued)

$\Delta\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	16.0945	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514	1.9473	2.1641	1.1676
	1.6958	16.0945	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514	1.9473	2.1641
	1.4484	1.6958	16.0945	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514	1.9473
	1.8662	1.4484	1.6958	16.0945	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514
	1.6782	1.8662	1.4484	1.6958	16.0945	1.6958	1.4484	1.8662	1.6782	1.8849
	1.8849	1.6782	1.8662	1.4484	1.6958	16.0945	1.6958	1.4484	1.8662	1.6782
	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	16.0945	1.6958	1.4484	1.8662
	1.9473	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	16.0945	1.6958	1.4484
	2.1641	1.9473	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	16.0945	1.6958
	1.1676	2.1641	1.9473	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	16.0945
$\Delta\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	1.9789	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514	1.9473	2.1641	1.1676
	1.6958	1.9789	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514	1.9473	2.1641
	1.4484	1.6958	1.9789	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514	1.9473
	1.8662	1.4484	1.6958	1.9789	1.6958	1.4484	1.8662	1.6782	1.8849	1.2514
	1.6782	1.8662	1.4484	1.6958	1.9789	1.6958	1.4484	1.8662	1.6782	1.8849
	1.8849	1.6782	1.8662	1.4484	1.6958	1.9789	1.6958	1.4484	1.8662	1.6782
	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	1.9789	1.6958	1.4484	1.8662
	1.9473	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	1.9789	1.6958	1.4484
	2.1641	1.9473	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	1.9789	1.6958
	1.1676	2.1641	1.9473	1.2514	1.8849	1.6782	1.8662	1.4484	1.6958	1.9789
$H\left(\vec{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right)$		$H\left(\vec{R}_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)\right)$		$H\left(\vec{R}^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)\right)$		$\Delta_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$		$\Delta_{\overset{\circ}{g}\overset{\circ}{g}}(\mu)$	$\Delta^{R}_{\overset{\circ}{X}\overset{\circ}{X}}(\mu)$	
2.158e + 05		77.5210		2.322e + 04		2.741e-04		2.577e + 22	4.759e + 09	

5 Conclusion

The developed noise technologies allow one to identify the early period of changes in the technical condition and seismic stability of high-rise buildings, building structures and strategic objects. The use of those technologies in monitoring systems makes it possible to warn the relevant services to take the necessary preventive measures at the initial stage of failure origin. This will lead to considerable reduction of the number of disastrous failures caused by delayed diagnostics in the modern monitoring systems.

References

1. Suschev, S.P.: Monitoring of stability and residual life of high-rise buildings and structures with application of "Strela" mobile diagnostics complex. Unique and special technologies in construction (UST-Build 2005). In: M.: CNTCMO, pp. 68–71 (2005)
2. Aliev, T.: Digital Noise Monitoring of Defect Origin, 235 p. Springer, London (2007)
3. Aliev, T.: Robust Technology with Analysis of Interference in Signal Processing, 199 p. Kluwer Academic/Plenum Publishers, New York (2003)
4. Aliev, T.A., Musaeva, N.F., Guluyev, G.A., Sattarova, U.E.: Noise indication of change of dynamical condition of production objects. Mechatronics, automation, control, No. 8, pp. 2–5 (2011)
5. Aliev T.A., Musaeva N.F., Guluyev G.A., Sattarova U.E. Noise Technology of Indication and Identification of the Latent Period of Transition of an Object from a Normal Condition to an Emergency One. Mechatronics, automation, control, No. 9, pp. 13–18 (2010)
6. Aliev, T.A., Guluyev, G.A., Rzayev, A.H., Pashayev, F.H.: Correlated indicators of microchanges in technical state of control objects. Cybernetics and systems analysis, No. 4. Springer, New York (2009)
7. Aliev, T.A., Abbasov, A.M., Guluyev G.A., Rzayev, A.H., Pashayev, F.H.: Positionally-binary and spectral indicators of microchanges in technical conditions of objects of the control. Automatic Control and Computer Sciences, No. 3. Allerton Press, Inc., New York (2009)
8. Musaeva, N.F.: Methodology of Calculating Robustness as an Estimator of the Statistical Characteristics of a Noisy Signal, Automatic Control and Computer Sciences, vol. 39, no. 5, pp. 53–62. Allerton Press Inc., New York (2005)

Analytical Structure Characterization and Stability Analysis for a General Class of Mamdani Fuzzy Controllers

Hao Ying and Dimitar Filev

Abstract Stability of a fuzzy control system is closely related to the analytical structure of the fuzzy controller, which is determined by its components such as input and output fuzzy sets and fuzzy rules. We first characterize the mathematical input–output structure of fuzzy controllers and then utilize the structure characteristics to advance stability analysis. We study how the components of a general class of Mamdani fuzzy controllers dictate the controller's input–output relationship. The controllers can use input fuzzy sets of any types, arbitrary fuzzy rules, arbitrary inference methods, either Zadeh or the product fuzzy logic AND operator, singleton output fuzzy sets, and the centroid defuzzifier. We theoretically prove that regardless of the choices for the other components, if and only if Zadeh fuzzy AND operator and piecewise linear (e.g., trapezoidal or triangular) input fuzzy sets are used, the fuzzy controllers become a peculiar class of nonlinear controllers with the following interesting characteristics: (1) they are linear with respect to input variables; (2) their control gains dynamically change with the input variables; and (3) they become linear controllers with constant gains around the system equilibrium point. These properties make the fuzzy controllers suitable for analysis and design using conventional control theory. This necessary and sufficient condition becomes a sufficient condition if the product AND operator is employed instead. We name the fuzzy controllers of this peculiar class type-A fuzzy controllers. Taking advantage of this new structure knowledge, we have established a necessary and sufficient local stability condition for the type-A fuzzy control systems. It can be used not only for the stability determination, but also for

H. Ying (✉)
Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA
e-mail: hao.ying@wayne.edu

D. Filev
Research and Advanced Engineering, Ford Motor Company, Detroit, MI, USA
e-mail: dfilev@ford.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_15,

practically designing a type-A fuzzy control system that is at least stable at the equilibrium point even when model of the controlled system is mathematically unknown. Three numerical examples are provided to demonstrate the utility of our new findings.

1 Introduction

Efforts have been made to rigorously derive and study analytical structure of fuzzy controllers. By analytical structure, we mean the mathematical relationship between the input and output of a fuzzy controller. Precise understanding of the structure is fundamentally important because it can enable one to analyze and design fuzzy control systems more effectively with the aid of conventional control theory [2, 7, 14]. The analytical structure is determined by a fuzzy controller's components including input fuzzy sets, output fuzzy sets, fuzzy rules, fuzzy inference, fuzzy logic operators, and defuzzifier. Different component choices obviously result in different analytical structures. The explicit structures of some fuzzy controllers have been derived [3, 11] (recently, the explicit structures of some type-2 fuzzy controllers have been investigated [1, 15]). They are related to classical controllers such as PID control [4, 10, 13] and sliding mode control [5]. No work, however, has been reported in the literature that characterizes analytical structures into different types with respect to component choices.

Our motivations for the current study are two folds—we first characterize the mathematical input–output structure of fuzzy controllers of a broad class by class and then utilize the structure characteristics and class to advance system stability analysis. Structurally speaking, such classification can produce structure information that is broader than what an individual analytical structure can because one class can cover many different fuzzy controllers with various nonlinear input–output structures. Subsequently, the structure classification can provide useful guidelines for the controller developer to choose appropriate types of the components (e.g., triangular fuzzy sets instead of Gaussian ones) at the early development stage, reducing time and effort on design and analysis in practice. Up to date, triangular and trapezoidal fuzzy sets have been most widely used for input variable fuzzification whereas other types (e.g., Gaussian and bell-shape fuzzy sets) are utilized to a lesser extent. As to fuzzy logic AND operators, Zadeh type and the product type are used far more often than any other types. As a matter of fact, the remaining types have hardly been used. Historically, these current preferred choices were determined mainly based on the results of a great deal of the trial-and-error effort, computer simulation study and the empirical development of successful fuzzy control applications. No analytically rigorous reasons in the context of conventional control theory have been given in the literature. In this paper, we attempt to provide a more rigorous answer and show that these popular choices are indeed theoretically sensible.

Another important benefit of classifying fuzzy controller structures is to make stability analysis more tractable and effective for a wide class of fuzzy controllers as one could concentrate on one class of fuzzy controllers a time instead of on all fuzzy controllers at once. Consequently, the stability results could be less conservative and more practically meaningful. In this paper, we will develop an analytical stability criterion that can be used to judge (local) stability of a broad class of fuzzy control systems regardless of the availability of the mathematical model of the system under control. The criterion can also be utilized to analytically design a stable fuzzy control systems not only when the system model is given, but also when it is unknown.

2 Configuration of a General Class of Mamdani Fuzzy Controllers

This class of Mamdani fuzzy controllers has m input variables, designated as $x_i(n), i = 1, 2, \ldots, m$, where n signifies sampling instance. $x_i(n)$ is computed using the current and/or historical output of a dynamic plant to be controlled (e.g., $y(n)$ and $y(n-1)$) as well as target output signal $S(n)$. This means the input space to be m-dimensional. $x_i(n)$ is multiplied by a scaling factor k_i, resulting in the scaled input variable $X_i(n)$. The universe of discourse for $X_i(n)$ is $[a_i, b_i]$. Assuming that each of the intervals is divided into M_i-1 subintervals, all the subintervals of the m intervals produce a total of

$$\Phi = \prod_{i=1}^{m} (M_i - 1)$$

m-dimensional "blocks" (i.e., divisions), each is designated as $\Theta_i, i = 1, \ldots, \Phi$.

M_i input fuzzy sets are defined over $[a_i, b_i]$. Like most fuzzy controllers in the literature, each subinterval has two fuzzy sets defined over it. The j-th fuzzy set for $X_i(n)$ is designated as $\tilde{A}_{i,j}$ whose membership function is denoted $\mu_{\tilde{A}_{i,j}}(x_i)$. The fuzzy sets can be any types. Of particular interest to this study is the piecewise linear fuzzy sets. Two examples of such fuzzy sets are illustrated in Fig. 1 (these examples are hypothetical and similar fuzzy sets are rarely used by fuzzy control in practice. We show this type of more dramatic examples to ensure a broad coverage). Note that the trapezoidal and triangular types, the most widely-used types among all the fuzzy set types available, are the special cases of the piecewise linear type. If $\mu_{\tilde{A}_{i,j}}(x_i)$ is of the trapezoidal type, it can be represented by

$$\mu_{\tilde{A}_{i,j}}(x_i) = \beta_{ij} \cdot x_i(n) + \lambda_{ij}$$

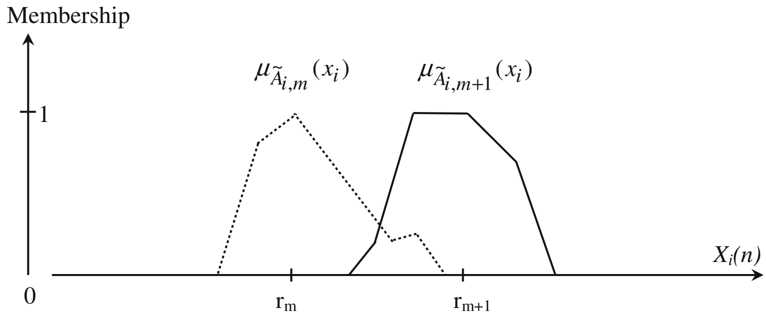


Fig. 1 Illustrative definition of piecewise linear fuzzy sets. Shown are two hypothetical ones. The widely-used *triangular* and *trapezoidal* fuzzy sets are their special cases

where λ_{ij} and β_{ij} are constants. β_{ij} is a constant taking different values in the different segments of $[a_i,\ b_i]$. When $x_i(n) \in [a_i, a_i^{'}], \beta_{ij} \geq 0$ and when $x_i(n) \in [b_i^{'}, b_i], \beta_{ij} \leq 0$, where $a_i^{'} < b_i^{'}$. When $x_i(n) \in [a_i^{'}, b_i^{'}], \mu_{\tilde{A}_{i,j}}(x_i) = 1$. Obviously, the trapezoidal type becomes the triangular type if $a_i^{'} = b_i^{'}$.

The fuzzy controllers use a total of $\Omega = \prod_{i=1}^{m} M_i$ fuzzy rules, each of which is in the following format:

$$\text{IF } X_1(n) \text{ is } \tilde{A}_{1,I_1} \text{ AND}\ldots\text{AND } X_m(n) \text{ is } \tilde{A}_{m,I_m} \text{ THEN } u(n) \text{ is } \tilde{V}_k$$

where $\tilde{V}_k$ is an output fuzzy set of the singleton type for the output variable $u(n)$ whose universe of discourse is $[U_L,\ U_H]$. That is, $\tilde{V}_k$ is nonzero only at one location in the interval and the nonzero value is designated as V_k. The fuzzy AND operator can be Zadeh type (i.e., the minimum operator) or the product type, but not both at the same time for the rule base. As for reasoning, any fuzzy inference method may be used in the rules. It will produce the same inference outcome because the output fuzzy sets are of the singleton type. The popular centroid defuzzifier is employed to combine the inference outcomes of the individual rules:

$$u(n) = \frac{\sum_{h=1}^{\Omega} \mu_h(\mathbf{x}) \cdot V_h}{\sum_{h=1}^{\Omega} \mu_h(\mathbf{x})} \tag{1}$$

Here, $\mathbf{x} = [x_1(n) \cdots x_m(n)]$ is the input vector and $\mu_h(\mathbf{x})$ is the resulting membership of executing all the fuzzy logic AND operations in the h-th rule whereas V_h signifies the nonzero value of the singleton output fuzzy set in the rule.

3 Structure Characterization and Local Stability Determination

3.1 Structure Characterization

In control theory, a general nonlinear controller is described by $u(n) = f(\mathbf{x})$ and a controller is linear if f is linear, that is

$$u(n) = \zeta_1 x_1(n) + \cdots + \zeta_m x_m(n) + \zeta_0 \tag{2}$$

where ζ_i, $i = 1, \ldots, m$, is a constant gain and ζ_0 is a constant control offset term. We use this formalism as a base to classify the fuzzy controllers into two types.

Definition 1 A fuzzy controller of the general class is defined as *local type-A fuzzy controller* in an m-dimensional region of input space if its input–output relation satisfies

$$u(n) = c_1(\mathbf{x})x_1(n) + \cdots + c_m(\mathbf{x})x_m(n) + c_0(\mathbf{x}) \tag{3}$$

$c_i(\mathbf{x})$, gain for $x_i(n)$, must be either a constant or a fractional expression whose numerator is constant. All the terms, from $x_1(n)$ to $x_m(n)$, must be present. $c_0(\mathbf{x})$ must be either a constant (including 0) or a fractional expression whose numerator does not contain $a_i x_i(n), i = 1, \ldots, m$, or their linear combination (a_i is constant).

Definition 2 A fuzzy controller is defined as *local type-B fuzzy controller* in the region if it is not of local type-A.

Definition 3 If a fuzzy controller is of local type-A in every region of input space, it is defined as a global type-A fuzzy controller. Otherwise, it is defined as a global type-B fuzzy controller.

A type-A fuzzy controller, global or local, possesses interesting properties—it is linear with respect to its input variables and nonlinear in terms of the gains, $c_i(\mathbf{x})$, which vary with $\mathbf{x}$. On the other hand, a type-B fuzzy controller cannot be linear with respect to input variables.

We are now ready to formally characterize the structure of the fuzzy controllers.

Theorem 1 A *necessary and sufficient* condition for a Mamdani fuzzy controller of the general class that uses Zadeh AND operator to be of global type-A is that all of its input fuzzy sets are piecewise linear.

Proof We first prove the necessity of the condition—if at least one of the input fuzzy sets involving a particular m-dimensional region of the input space is not piecewise linear, the fuzzy controller cannot be of global type-A.

Without loss of generality, assume that $\mu_{\tilde{A}_{i,j}}(x_i)$, a non-piecewise linear function, is the membership function of the j-th fuzzy set for $x_i(n)$. To derive the input–output analytical relationship for the fuzzy controllers, the input space must be divided into many m-dimensional regions in such a way that in each of the regions one and only one of the membership functions in a fuzzy rule will always be the smallest [13]. As a result, that membership function will be the resulting membership function for that rule after Zadeh AND operation is carried out. The total number of such regions depends on m and the shape of the input fuzzy sets. Inevitably, $\mu_{\tilde{A}_{i,j}}(x_i)$ will be the smallest membership function in at least one region (or this fuzzy set would not be useful and hence should be removed used—cannot be true). After this step of considering the individual rules, the obtained regions must be put together (i.e., superimposed) so that all the rules can be considered at the same time. This process will create the final m-dimensional regions. Let us assume that there are a total of Φ regions, in each of which, up to 2^m fuzzy rules will be involved (here we assume, without loss of generality, that the input fuzzy set for $x_i(n)$ intersects only with its adjacent (two) fuzzy sets and intersects only once with each of them). Let Ψ_i be the number of the rules involved in the i-th final region and suppose that in the region the input fuzzy sets for $x_1(n)$ is the smallest for α_1 rules, $x_2(n)$ for α_2 rules,…, and $x_m(n)$ for α_m rules, where $\sum_{p=1}^{m} \alpha_p = \Psi_i$ (some α_p may be 0). For the i-th final region, we get the input–output relation:

$$u(n) = \frac{\sum_{j=1}^{m} \sum_{k=1}^{\alpha_j} \mu_{h_k}(x_j) \cdot V_{h_k}}{\sum_{j=1}^{m} \sum_{k=1}^{\alpha_j} \mu_{h_k}(x_j)} \tag{4}$$

where $\mu_{h_k}(x_j)$ represents the resulting smallest membership function and $V_{h_k}$is the singleton output fuzzy set involved in the fuzzy rule. Because V_{h_k} represents the rule consequent and all the rule consequents should not have the same singleton fuzzy sets, thus the above equation should not/cannot be equal to V_{h_k} for all the final regions. Hence, if at least one input fuzzy set, say $\mu_{\tilde{A}_{i,j}}(x_i)$, is not linear or piecewise linear in one of the final regions, $u(n)$ in that region will not be able to be written in the form of (3) because $x_i(n)$ cannot be factored out of $\mu_{\tilde{A}_{i,j}}(x_i)$. That is to say that the fuzzy controller cannot be a local type-A controller in that region and therefore the fuzzy controller cannot be a global type-A controller.

Now let us prove the condition to be sufficient. When all the input fuzzy sets are linear or piecewise linear, the smallest membership function in every final region will always be a linear function of $x_i(n)$ (note that a piecewise linear membership function can always be decomposed into a series of linear membership functions). Suppose that the linear fuzzy sets are

$$\mu_{h_k}(x_j) = \beta_{h_k} x_j(n) + \lambda_{h_k}.$$

As a result,

$$u(n)=\frac{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)\cdot V_{h_k}}{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)}=\sum_{j=1}^{m}\frac{\sum_{k=1}^{\alpha_j}\beta_{h_k}\cdot V_{h_k}}{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)}x_j(n)+\frac{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\lambda_{h_k}\cdot V_{h_k}}{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)}$$
$$=c_1(\mathbf{x})x_1(n)+\cdots+c_m(\mathbf{x})x_m(n)+c_0(\mathbf{x})$$

where

$$c_j(\mathbf{x})=\frac{\sum_{k=1}^{\alpha_j}\beta_{h_k}\cdot V_{h_k}}{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)},\quad j=1,2,\ldots,m \text{ and } c_0(\mathbf{x})=\frac{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\lambda_{h_k}\cdot V_{h_k}}{\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)}$$

Unless the fuzzy sets are such chosen that $\sum_{j=1}^{m}\sum_{k=1}^{\alpha_j}\mu_{h_k}(x_j)$ is a constant, the value of $c_j(\mathbf{x})$ changes with $x_j(n)$ and hence is variable gain for $x_j(n)$. According to Definition 3, the fuzzy controller is a global type-A fuzzy controller. ■

We can rigorously show that Theorem 1 also holds for the fuzzy controllers that use the product AND operator when the popular input fuzzy sets in the literature are utilized. They include triangular, trapezoidal, Gaussian, Bell, Generalized Bell, Sigmoid, and so on. This kind of result, however, is not comprehensive enough. We know that the necessity portion of the theorem will not hold if the input fuzzy sets are allowed to be any nonlinear functions. Subsequently, we decided to present the following sufficient condition. This result can be extended to become a necessary and sufficient condition if some (mild) mathematical constraint is applied to the fuzzy sets. The constraint will most likely not affect the practicality of the fuzzy controllers, if at all.

In the next section, we will conduct local stability analysis. For that purpose, we only need to know the properties of the relevant part of the fuzzy controllers—the local fuzzy controller covering the system equilibrium point. The following results are obvious in light of Theorem 1 and Definitions 1 to 3.

Corollary 1 A necessary and sufficient condition for a Mamdani fuzzy controller of the general class that uses Zadeh AND operator to be of local type-A around the system equilibrium point is that all the input fuzzy sets covering the point are linear.

Note that the input fuzzy sets have to be linear, not piecewise linear, because the latter is not sensible when we consider only a point.

Corollary 2 A Mamdani fuzzy controller of the general class that uses Zadeh AND operator is a global type-B fuzzy controller if at least one input fuzzy set is not piecewise linear.

Theorem 2 A *sufficient* condition for a Mamdani fuzzy controller of the general class that uses the product AND operator to be of global type-A is that all of its input fuzzy sets are piecewise linear.

Proof Due to the product fuzzy AND operator, in each m-dimensional "block", the combined membership function for the h-th rule can be represented as

$$\mu_h(\mathbf{x}) = \mu_{\tilde{A}_{1,I_1}}(x_1) \times \cdots \times \mu_{\tilde{A}_{m,I_m}}(x_m)$$

and hence

$$\begin{aligned} u(n) &= \frac{\sum_{h=1}^{\Omega} \mu_h(\mathbf{x}) \cdot V_h}{\sum_{h=1}^{\Omega} \mu_h(\mathbf{x})} = \sum_{h=1}^{\Omega} \frac{V_h}{\sum_{h=1}^{\Omega} \mu_h(\mathbf{x})} \mu_h(\mathbf{x}) \\ &= \sum_{h=1}^{\Omega} \theta_h(\mathbf{x}) \times \mu_h(\mathbf{x}) = \sum_{h=1}^{\Omega} \theta_h(\mathbf{x}) \times \mu_{\tilde{A}_{1,I_1}}(x_1) \times \cdots \times \mu_{\tilde{A}_{m,I_m}}(x_m) \end{aligned} \tag{5}$$

where

$$\theta_h(\mathbf{x}) = \frac{V_h}{\sum_{h=1}^{\Omega} \mu_h(\mathbf{x})}.$$

Note that due to the use of the product AND operation, $\mu_h(\mathbf{x})$ here differs from the case above that uses Zadeh AND operator. Assume that in the entire input space, all the input fuzzy sets are linear or piecewise linear. Then in any Θ_i, carrying out all the multiplication operations in (5) and simplifying the resultant expression gives

$$\begin{aligned} u(n) &= \sum_{i=1}^{m} a_i(\mathbf{x}) \cdot x_i(n) + \sum_{i=1}^{m} \sum_{j>i}^{m} a_{ij}(\mathbf{x}) \cdot x_i(n) \cdot x_j(n) + \cdots + a_{1\ldots m}(\mathbf{x}) x_1(n) \cdots x_m(n) + \delta(\mathbf{x}) \\ &= \sum_{i=1}^{m} a_i(\mathbf{x}) \cdot x_i(n) + \Gamma(\mathbf{x}) \end{aligned}$$

where

$$\Gamma(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j>i}^{m} a_{ij}(\mathbf{x}) \cdot x_i(n) \cdot x_j(n) + \cdots + a_{1\ldots m}(\mathbf{x}) x_1(n) \cdots x_m(n) + \delta(\mathbf{x})$$

The denominators of $a_i(\mathbf{x})$ and $\Gamma(\mathbf{x})$ are the same—$\sum_{h=1}^{\Omega} \mu_h(\mathbf{x})$. The numerators of $a_i(\mathbf{x})$ are constants determined by the slops and y-intercepts of the fuzzy sets. The numerator of $\Gamma(\mathbf{x})$ contains the cross-products of $x_1(n), \ldots, x_m(n)$ as well as a constant term, all determined by the slopes and y-intercepts of the fuzzy sets. Note that none of the terms in the numerator can be expressed as $a_i x_i(n), i = 1, \ldots, m$, or their linear combination (a_i is constant). By Definition 1, the fuzzy controller is a global type-A fuzzy controller. ■

For the same reason as for obtaining Corollary 1, we state Corollary 3 for Theorem 2.

Corollary 3 A sufficient condition for a Mamdani fuzzy controller of the general class that uses the product AND operator to be of local type-A around the system equilibrium point is that all the input fuzzy sets covering the point are linear.

Note that a Mamdani fuzzy controller of the general class that uses the product AND operator may (very likely) be a global type-B fuzzy controller if at least one input fuzzy set is not piecewise linear.

In a sense, type-A fuzzy controllers are most similar to the linear controller as far as the controller structure is concerned. These fuzzy controllers can be treated as nonlinear controllers with variable gains. The variable gains are mathematically complicated in general. Deriving their explicit expressions is possible for some configurations [12] but impossible for others. The variable gains can empower the fuzzy controllers to outperform the linear controller (e.g., PID control), especially when the system under control is nonlinear or with time delay. Our previous study of Mamdani as well as TS fuzzy controllers with variable gains has shown this point [12, 13]. This may provide one theoretical justification/explanation for the dominant use of triangular/trapezoidal input fuzzy sets and Zadeh or product fuzzy AND operator in the current practice of fuzzy control. There are many other choices available, but most of them have hardly been used. Such strong preferences have existed in the literature for many years with little theoretical support. The prior justifications are based on perceived component simplicity and empirical observation of good control performance attributed to these particular preferred selections.

3.2 Local Stability Analysis

We now turn our attention to local stability determination of the fuzzy controllers regulating nonlinear dynamic systems. Without loss of generality, assume that when a system to be controlled is at the equilibrium point of our interest (i.e., $y(n) = 0$), $\mathbf{x} = \mathbf{0}$. We want to study two related issues: (1) conditions for the fuzzy control system to be stable at least in the area around the equilibrium point, and (2) design of the fuzzy control system that will be at least stable in the area.

At $\mathbf{x} = \mathbf{0}$, a local fuzzy controller satisfying Corollary 1 or 3 becomes a local linear controller:

$$u(n) = c_1(\mathbf{0})x_1(n) + \cdots + c_m(\mathbf{0})x_m(n) + c_0(\mathbf{0}) \tag{6}$$

where $c_0(\mathbf{0}), \ldots, c_m(\mathbf{0})$ are the values of $c_i(\mathbf{x})$ and are the constant gains. The term of $c_0(\mathbf{0})$ does not affect system stability; it will be dropped from the stability study below. If both the system to be controlled and the local fuzzy controller are linearizable at the equilibrium point, then the system stability at that point can be decided by applying Lyapunov's linearization method [6] to the linearized local fuzzy controller (i.e. (2)) and the linearized system. Thus, we obtain the following result.

Theorem 3 Suppose that a fuzzy controller of the general class is used to control a nonlinear system that is linearizable at the equilibrium point. If the fuzzy controller is a local type-A controller at the equilibrium point that is also linearizable, the fuzzy control system is locally stable (or unstable) at the equilibrium point if and only if the linearized control system is strictly stable (or unstable) at the equilibrium point.

Proof The conclusion can be obtained using Lyapunov's linearization method (see [10] for similar proof). The method assumes the nonlinear control system to be continuously differentiable at the equilibrium point. In essence, it states that if the linearized system is strictly stable (or unstable) at the equilibrium point, then the equilibrium point is locally stable (or unstable) for the original nonlinear system. ■

The linearizability test must be met before the stability condition can be used because it is the precondition for the theorem. Our previous study indicates that the use of Zadeh fuzzy logic AND operator results in more than one control structure to cover $\mathbf{x} = \mathbf{0}$, which can sometimes fail the test. This is usually not true for the fuzzy controllers using the product AND operator because in most cases there is only one control structure for the entire area around $\mathbf{x} = \mathbf{0}$ [12]. A test failure only means inapplicability of the theorem; it does not imply system instability.

Theorem 3 offers some practically important advantages. First, it is a necessary and sufficient condition. Unlike sufficient conditions or necessary conditions, it is not conservative and is the "tightest" possible stability condition. Second, only explicit structure of the local fuzzy controller covering the equilibrium point is required. As long as all the input fuzzy sets covering the equilibrium point are linear, the theorem is usable. Third, the theorem can be used not only when the system model is available, but also when it is unavailable but is known linearizable at the equilibrium point. (Most physical systems are linearizable.) In the latter case, one can devise a linear controller and use it to control the system. If the resulting control system is observed to be locally stable (unstable), then the same system controlled by a linearized fuzzy controller whose gains at the equilibrium point equal to the gains of the linear controller will be locally stable (unstable). This design approach can be attractive as in practice, physical systems are often too complex and/or costly to be precisely modeled.

4 Numerical Examples

Since Theorems 1 and 2 and their corollaries are straightforward to apply, no numerical examples are needed. We now use three examples to illustrate the utility of Theorem 3 and its above-mentioned advantages in system analysis and design.

Example 1 (Local stability determination when system model is known) Assume that a continuous-time system to be controlled is

$$y''(t) + 20y'(t) + 5\sin(y(t)) = 4u(t)$$

and the sampling period for the system is 0.1. Suppose that a type-A fuzzy controller has been designed. It uses the product fuzzy AND operator and its input variables are $x_1(n) = S(n) - y(n)$ and $x_2(n) = y(n-1) - y(n)$, where $S(n)$ is the reference input signal. The controller's output variable is $\Delta u(n) = u(n) - u(n-1)$. There can be numerous fuzzy sets for the scaled input variables $X_1(n)$ and $X_2(n)$; but to determine the local stability one only needs to know whether the fuzzy sets covering the area around $\mathbf{x} = \mathbf{0}$ are linear and, if so, what their mathematical expressions are. Supposedly they are

$$\mu_{\tilde{A}_{1,1}}(x_1) = -0.95x_1(n) + 0.4, \quad \mu_{\tilde{A}_{1,2}}(x_1) = x_1(n) + 0.5,$$

for $x_1(n) \in [-0.5,\ 0.5]$ and

$$\mu_{\tilde{A}_{2,1}}(x_2) = -1.1x_2(n) + 0.5, \quad \mu_{\tilde{A}_{2,2}}(x_2) = 0.9x_2(n) + 0.2$$

for $x_2(n) \in [-0.5,\ 0.5]$.

The four fuzzy rules resulted from the four combinations of these fuzzy sets are

IF $X_1(n)$ is $\tilde{A}_{1,1}$ AND $X_2(n)$ is $\tilde{A}_{2,1}$ THEN $\Delta u(n)$ is $\tilde{V}_1$
IF $X_1(n)$ is $\tilde{A}_{1,1}$ AND $X_2(n)$ is $\tilde{A}_{2,2}$ THEN $\Delta u(n)$ is $\tilde{V}_2$
IF $X_1(n)$ is $\tilde{A}_{1,2}$ AND $X_2(n)$ is $\tilde{A}_{2,1}$ THEN $\Delta u(n)$ is $\tilde{V}_3$
IF $X_1(n)$ is $\tilde{A}_{1,2}$ AND $X_2(n)$ is $\tilde{A}_{2,2}$ THEN $\Delta u(n)$ is $\tilde{V}_4$

where $\tilde{V}_1, \tilde{V}_2, \tilde{V}_3$, and $\tilde{V}_4$ are singleton fuzzy sets whose $\Delta u(n)$ values corresponding to the nonzero memberships of these sets are at 1, 0.71, -1.083, and -1, respectively.

The question is: Is this fuzzy control system stable at $\mathbf{x} = \mathbf{0}$?

Solution Due to the use of the linear fuzzy sets, the local fuzzy controller covering the region around $\mathbf{x} = \mathbf{0}$ satisfies Corollary 3 and hence is a local type-A fuzzy controller. For the stability determination, we need to have the explicit structure of this local fuzzy controller, which can be derived by plugging the four fuzzy sets and the four fuzzy rules into (1). The result is

$$\Delta u(n) = c_1(\mathbf{x})x_1(n) + c_2(\mathbf{x})x_2(n) + c_0(\mathbf{x})$$

where

$$c_1(\mathbf{x}) = \frac{1.23578}{0.035x_1(n) - 0.18x_2(n) - 0.01x_1(n)x_2(n) + 0.63},$$
$$c_2(\mathbf{x}) = \frac{0.109605}{0.035x_1(n) - 0.18x_2(n) - 0.01x_1(n)x_2(n) + 0.63},$$
$$c_0(\mathbf{x}) = -\frac{0.00914327}{0.035x_1(n) - 0.18x_2(n) - 0.01x_1(n)x_2(n) + 0.63}.$$

The controlled system is obviously linearizable at $\mathbf{x} = \mathbf{0}$ and the linearized system is

$$y''(t) + 20y'(t) + 5y(t) = 4u(t).$$

With the sampling period 0.1, the discrete-time pulse transfer function is

$$H(z) = \frac{Y(z)}{U(z)} = \frac{0.01131\ \mathrm{z} + 0.005913}{\mathrm{z}^2 - 1.114\ \mathrm{z} + 0.1353}.$$

The local fuzzy controller is obviously linearizable and the resultant controller is

$$\Delta u(n) = 1.96155x_1(n) + 0.173977x_2(n) - 0.0145131. \tag{7}$$

For the stability determination, one only needs to consider $\Delta u(n) = 1.96155e(n) + 0.173977r(n)$. Because

$$\begin{aligned} u(n) &= u(n-1) + \Delta u(n) = u(n-1) + 1.96155x_1(n) + 0.173977x_2(n) \\ &= u(n-1) + 2.135527x_1(n) - 0.173977x_1(n-1), \end{aligned}$$

the transfer function of the linearized local fuzzy controller is

$$C(z) = \frac{U(z)}{X_1(z)} = \frac{2.135527z - 0.173977}{z - 1}.$$

The closed-loop control system at $\mathbf{x} = \mathbf{0}$ is

$$\frac{H(z)C(z)}{1 + H(z)C(z)} = \frac{0.024159\ (\mathrm{z} + 0.5226)\ (\mathrm{z} - 0.08146)}{(\mathrm{z} - 0.1376)\ (\mathrm{z}^2 - 1.952\mathrm{z} + 0.9913)}.$$

The poles are $z = 0.1376$, and $z = 0.9760 \pm 0.1965\mathrm{i}$, all of which are inside the unit circle. Therefore, the linearized control system is stable at $\mathbf{x} = \mathbf{0}$ stable, so is the local fuzzy control system.

In next example, we show how to determine the local stability even when system model is unknown.

Example 2 (Local stability determination when system model is unknown) Suppose that the system model in Example 1 is unavailable but is known to be linearizable at $\mathbf{x} = \mathbf{0}$. Also, assume that we have devised a linear PI controller $\Delta u(n) = 1.96155x_1(n) + 0.173977x_2(n)$ and found it to be able to control the system stably at $\mathbf{x} = \mathbf{0}$. Note that this PI controller is exactly the same as the local fuzzy controller (4) without the constant offset term. With these modifications, can the local stability of the fuzzy control system in Example 1 be determined ?

Solution According to Example 1, the new fuzzy controller system is linearizable. Thus Theorem 3 is applicable. Since the PI control system containing $\Delta u(n) = 1.96155x_1(n) + 0.173977x_2(n)$, which is the linearized local fuzzy controller, is known to be locally stable, the fuzzy control system is logically locally stable too.

Example 3 (Design of at least locally stable fuzzy control system when system model is unknown) Suppose that the $\Delta u(n)$values for $\tilde{V}_1, \tilde{V}_2, \tilde{V}_3$, and $\tilde{V}_4$ are not given in the above example, but all the other conditions remain the same. How should one design their values so that the resulting fuzzy control system is stable at least at $\mathbf{x} = \mathbf{0}$?

Solution It can be derived that the local fuzzy controller is $\Delta u(n) = c_1(\mathbf{x})x_1(n) + c_2(\mathbf{x})x_2(n) + c_0(\mathbf{x})$ where

$$c_1(\mathbf{x}) = \frac{0.2V_1 + 0.5V_2 - 0.19V_3 - 0.475V_4}{0.035x_1(n) - 0.18x_2(n) - 0.01x_1(n)x_2(n) + 0.63},$$
$$c_2(\mathbf{x}) = \frac{0.45V_1 - 0.55V_2 + 0.36V_3 - 0.44V_4}{0.035x_1(n) - 0.18x_2(n) - 0.01x_1(n)x_2(n) + 0.63},$$
$$c_0(\mathbf{x}) = \frac{0.1V_1 + 0.25V_2 + 0.08V_3 + 0.2V_4 + x_1^2(n)(0.9V_1 - 1.1V_2 - 0.855V_3 + 1.045V_4)}{0.035x_1(n) - 0.18x_2(n) - 0.01x_1(n)x_2(n) + 0.63}.$$

Thus,

$$c_1(\mathbf{0}) = 0.3175V_1 + 0.7937V_2 - 0.3019V_3 - 0.7540V_4,$$
$$c_2(\mathbf{0}) = 0.7143V_1 - 0.8730V_2 + 0.5714V_3 - 0.6984V_4,$$
$$c_0(\mathbf{0}) = 0.1587V_1 + 0.3968V_2 + 0.1270V_3 + 0.3175V_4.$$

We know from Example 2 that the linear PI controller $\Delta u(n) = 1.96155x_1(n) + 0.173977x_2(n)$, which is the same as the local fuzzy controller (4) without the constant term, can control the system stably at least at $\mathbf{x} = \mathbf{0}$. Thus, the fuzzy control system of interest will be locally stable too if the values of the design parameters satisfy the following simultaneous equations:

$$\begin{cases} 0.1587V_1 + 0.3968V_2 + 0.1270V_3 + 0.3175V_4 = 0 \\ 0.3175V_1 + 0.7937V_2 - 0.3019V_3 - 0.7540V_4 = 1.96155 \\ 0.7143V_1 - 0.8730V_2 + 0.5714V_3 - 0.6984V = 0.173977. \end{cases}$$

The number of solution set is infinite, and every set achieves the local stability. One set, for instance, is: $V_1 = 1, V_2 = -0.40, V_3 = -1.25$, and $V_4 = -0.5$.

5 Conclusion

We have achieved two objectives: (1) to establish the conditions for a subset of a general class of Mamdani fuzzy controllers to be a specific type of nonlinear controllers described in (3), and (2) to utilize these conditions and establish a tight local stability criterion for analyzing or designing the fuzzy control systems even when the controlled system model is mathematically unavailable. This type of controllers has some desirable characteristics suitable for analysis and design using conventional control theory.

Based on our results, we recommend that the trapezoidal and triangular fuzzy sets, the only two widely-used piecewise linear types, be used for input fuzzy sets as the first choice. As we have demonstrated theoretically and through examples, the benefits of doing so include (1) clearer connection between the fuzzy controllers and conventional control (2) easier (local) stability analysis, and (3) more practically meaningful system design.

References

1. Du, X., Ying, H.: Derivation and analysis of the analytical structures of the interval type-2 fuzzy PI and PD controllers. IEEE Trans. Fuzzy Syst. **8**, 802–814 (2010)
2. Farinwata, S.S., Vachtsevanos, G.: Stability analysis of the fuzzy logic controller designed by the phase portrait assignment algorithm. In: Proceedings of the 2nd IEEE International conference Fuzzy Systems, San Francisco, pp. 1377–1382 (1993)
3. Langari, R.: A nonlinear formulation of a class of fuzzy linguistic control algorithms. In: Proceedings of the 1992 American Control Conference, Chicago, IL (1992)
4. Lewis, F.L., Liu, K.: Towards a paradigm for fuzzy logic control. Automatica **32**, 167–181 (1996)
5. Li, H.X., Gatland, H.B.: A new methodology for designing a fuzzy logic controller. IEEE Trans. Syst. Man Cybern. **25**, 505–512 (1995)
6. Slotine, J.-J.E., Li, W.: Applied Nonlinear Control. Prentice Hall, Englewood Cliffs (1991)
7. Wang, L.X.: Adaptive Fuzzy Systems and Control. PTR Prentice Hall, Upper Saddle River (1994)
8. Wang, H.O., Tanaka, K., Griffin, M.F.: An approach to fuzzy control of nonlinear systems: stability and design issue. IEEE Trans. Fuzzy Syst. **4**, 14–23 (1996)
9. Yager, R.R., Filev, D.P.: Essentials of Fuzzy Modeling and Control. Wiley, New York (1994)

10. Ying, H.: The simplest fuzzy controllers using different inference methods are different nonlinear proportional-integral controllers with variable gains. Automatica **29**, 1579–1589 (1993)
11. Ying, H.: Practical design of nonlinear fuzzy controllers with stability analysis for regulating processes with unknown mathematical models. Automatica **30**, 1185–1195 (1994)
12. Ying, H.: Fuzzy Control and Modeling: Analytical Foundations and Applications. IEEE Press, New York (2000)
13. Ying, H., Siler, W., Buckley, J.J.: Fuzzy control theory: a nonlinear case. Automatica **26**, 513–520 (1990)
14. Zeng, X.J., Singh, M.G.: Decomposition property of fuzzy systems and its applications. IEEE Trans. Fuzzy Syst. **4**, 167–181 (1996)
15. Zhou, H., Ying, Hao: A method for deriving the analytical structure of a broad class of typical interval type-2 Mamdani fuzzy controllers. IEEE Trans. Fuzzy Syst. **21**, 447–458 (2013)

Fuzzy Control as the Entrance Door to Control Theory

António Dourado

Abstract Control Systems have a more and more important role in our lives. The future of human kind depends largely on the space exploration, and autonomous intelligent controlled systems will be the key of its success. However the control field has become largely unpopular in universities and it is more and more difficult to attract young bright students to research in the field. The first university course on control is decisive to fix attraction level of the field. Fuzzy logic can be used to face control as a challenging game well adapted to the way of thinking of young generations. A proposal is developed here and some experience is reported.

1 Introduction

It is frequent to hear Professors of Control complaining about the difficulty of recruiting young students to control area. The situation is such that in most cases students have a mandatory course on systems and control theory based on excellent traditional books dated from 50 or more years ago (developing in detail the "classical control theory" in complex domain), and after that they run away from the area and the optional courses remain emptied. In the author's opinion this is because these first courses, usually at second or third year, have a dense heavy analytical mathematical formulation not properly adapted to the brains of young generations. Moreover, this complex formulation hides the practical meaning of control concepts.

Young students are as bright as the previous generations, but they have much more developed skills. They do many different things and have a multiplicity of interests. Their brains have been trained in a different way, oriented by the digital world of interfaces, imaging and games. It is our challenge to teach them systems and control in a way that fits with their way of thinking and acting.

A. Dourado (✉)
Department of Informatics Engineering, University of Coimbra,
3030-290 Coimbra, Portugal
e-mail: dourado@dei.uc.pt

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_16,

Can control be taught as a game, with high interaction with the students? Yes, it can with the support of fuzzy logic and using mainly graphical inference techniques.

A proposal is developed here and some experiences are reported.

Firstly an historical perspective of control is proposed in Sect. 2 leading to the concept of several levels of learning presented in Sect. 3. The possible role of fuzzy logic is proposed in Sect. 4 and conclusions in Sect. 5.

2 A Perspective of Control History

The Automatic Control History can be divided into three phases: the artists phase, the pre-scientific phase and the scientific phase [1–3].

In the artists phase, from Classic Antiquity to the end of Middle Age (Sect. 14), the inventions were produced by the ability, creativity and genius of some exceptional individuals. The pre-scientific phase goes from 15th to 18th centuries, during Renaissance and Baroque times [1].

The scientific phase started in control after 1867 by the works of the pioneers around the stability theory, passed by the classic theory, the modern theory, the neo-classic theory. Classical control theory was developed from 1920 to 1955 by three important schools: (i) the industrial instrumentalists, (ii) the communications engineers (Black, Nyquist, Bode), and (iii) the MIT Servomechanism Laboratory and Radiation Laboratory. By the end of the Second World War, the Classical Control Theory emerged as a body of knowledge and was completed with the work of Evans in 1950. These achievements may be considered the first great synthesis in Control (see Fig. 1). The information processing was basically in electrical analogue form.

The digital computer lead to the then called Modern Control Theory in time-domain state space. The microprocessor (in 1972) and real-time operating systems from Computer Science stimulated identification and adaptive control (see Table 1).

However part of control community did not accept this time domain approach. The Neo-Classic school [2] continued to use frequency domain tools and concepts that lead to the development of robust control in the frequency domain. This may be seen as the first schism in Control community. The introduction into robust control of the concepts and tools of state space, leads to the unification of the two approaches (in frequency and in time domain).

The potentialities of digital computers and the ubiquity of microprocessors allowed processing several types of information and in very big quantities with the methods and tools developed by Computer Science and Artificial Intelligence communities. A part of the control community looked in that direction and new types of control systems were developed during the last 30 years, founding the so called “Intelligent” Control discipline, or data-based control (as alternative to model-based control).

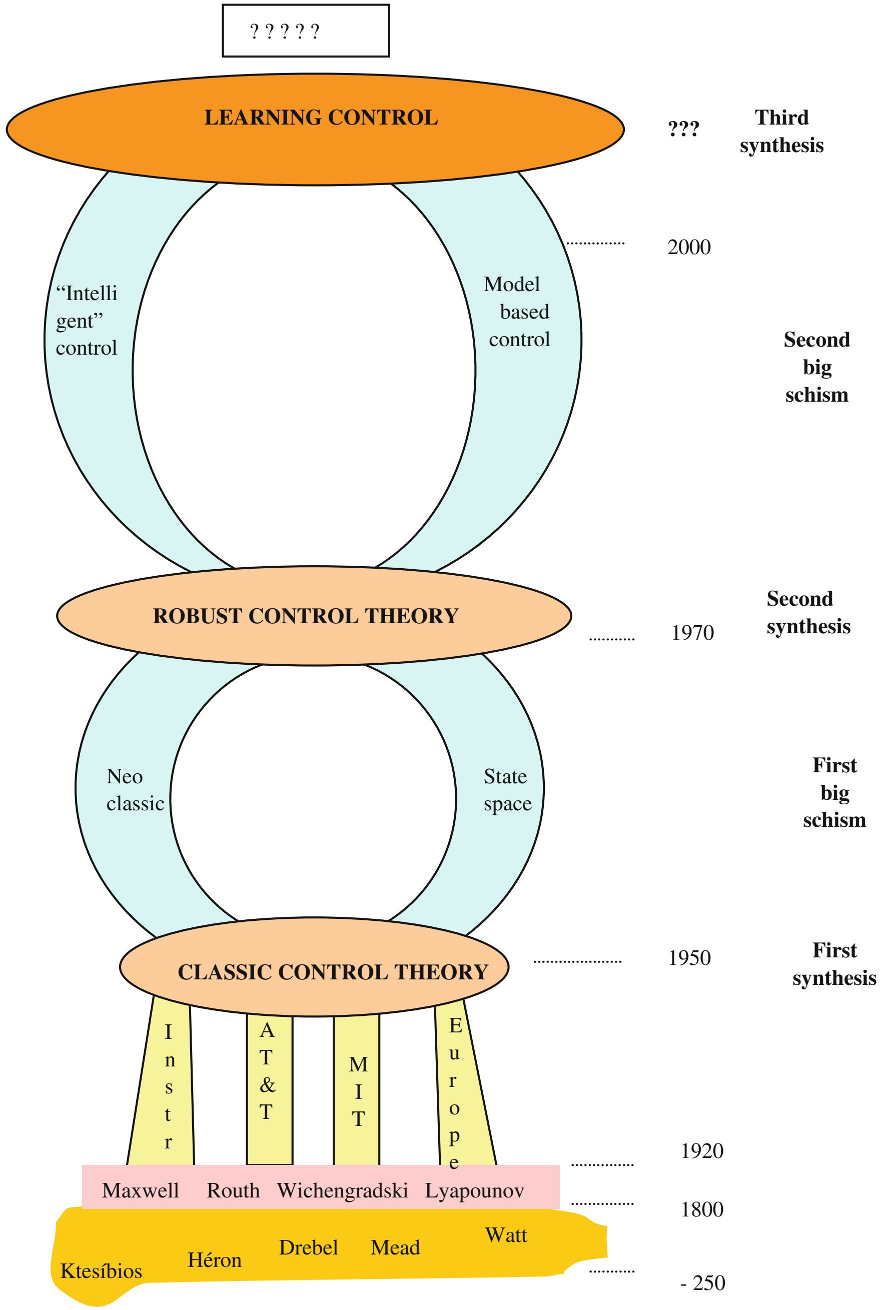


Fig. 1 Review of the automartic control history. The situation today indiciates already the coming of a new synthesis, the third one in control history. This synthesis will be the learning control theory (or theory of learning and acting) [5]

Another part of Control community continued the development of the usually called model-based control approach and further important progress have been achieved. This lead to the second schism in Control.

As a consequence it can be said that Automatic Control is actually being developed in the framework of three paradigms:

- the integral-differential paradigm, including all control methods using models starting from differential equations (transfer function and state-space),
- the data paradigm, including all control methods based on experimental and empirical data (black-box models, neural networks models),
- the linguistic paradigm, using qualitative representations like fuzzy systems and controllers.

3 Learning Control

After the digital computer dominated the analogue one, process models are implemented in digital computers, using some programming language. After compilation in the digital computer, and independently of the starting knowledge used to arrive to it, the model is translated into a series of 0's and 1's. They are just pieces of machine code to process information. The computer receives information and processes it (and eventually takes a decision). This is in the author's opinion the main characteristic of a model (in the present context). The concept of **Model** should then be defined in the information space, as [4]:

Definition 1 Model. A model is a representational tool reducible to a piece of computer code enabling the computer to represent some part of the world. With that code it can process perceived information from that world and produce other information about that world with some usefulness.

Two keywords appear in this definition: code and information. They are connected to the two main properties of a model: *computability* and *granularity* (of information) [5].

Computability measures the adequacy of the representation tool to be translated into a piece of computer code. So this definition of model includes in the same way the conventional, the neural networks, the fuzzy or any other kind of representation that can be programmed in computers. It does not include analogue models (like architectural maquetes).

Granularity of information expresses the type of information the model processes:

- high granularity: it expresses and processes qualitative information (like fuzzy and possibilistic models, for example),
- low granularity: it expresses quantitative information (given by numbers in a given numeric scale).

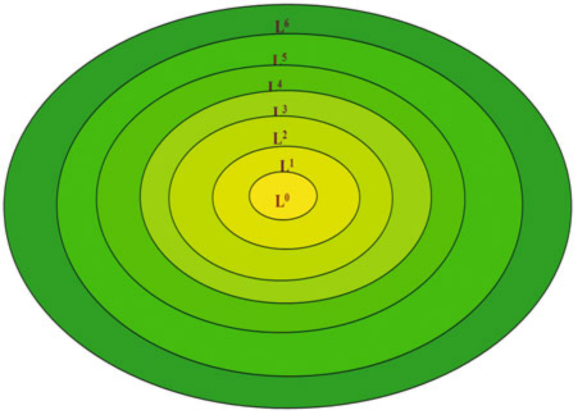


Fig. 2 The several levels of learning

Granularity is the additional dimension (other than time and space) that transforms the state space into the information space. The classic state space is in this context the sub-space of quantitative information in the information space. Granularity may be tentatively defined as the minimum difference between two pieces of information such they can be distinguished.

The term “intelligent control” is intensively used nowadays in control community and outside it. However there is not yet an agreement about what intelligence should be considered in this context. This fact creates walls to communication, and sometimes different people use the same term to express different things or use different terms to say the same thing.

There have been discussions about what should be an Intelligent Systems (see for example Antsaklis [6]). If one can accept the general simple definition that *a system is intelligent if it can sense its environment, detect its changements and adapt its behavior and goals to these changements (to pursue its own goals)* then it is easy to arrive to a definition of intelligent control.

Definition 2 Intelligent controller. A controller is intelligent if it can percept changements in the controlled system or in the environment of the controlled system and adapts itself to those changements in order to maintain the performance of the control system.

Adaptation in this definition is a consequence of intelligence. How can the controllers developed during the last decades be included in this definition? How broad and unifying is this definition?

Usually adaptation, in the control community, is associated with parameter estimation. In order to give to it a more general meaning, *learning* is proposed as the extension (in the information space) of adaptation [5].

Definition 3 Learning. Learning is the procedure used by an intelligent controller to change its own behavior as a consequence of the changements in the behavior of the controlled system or its environment.

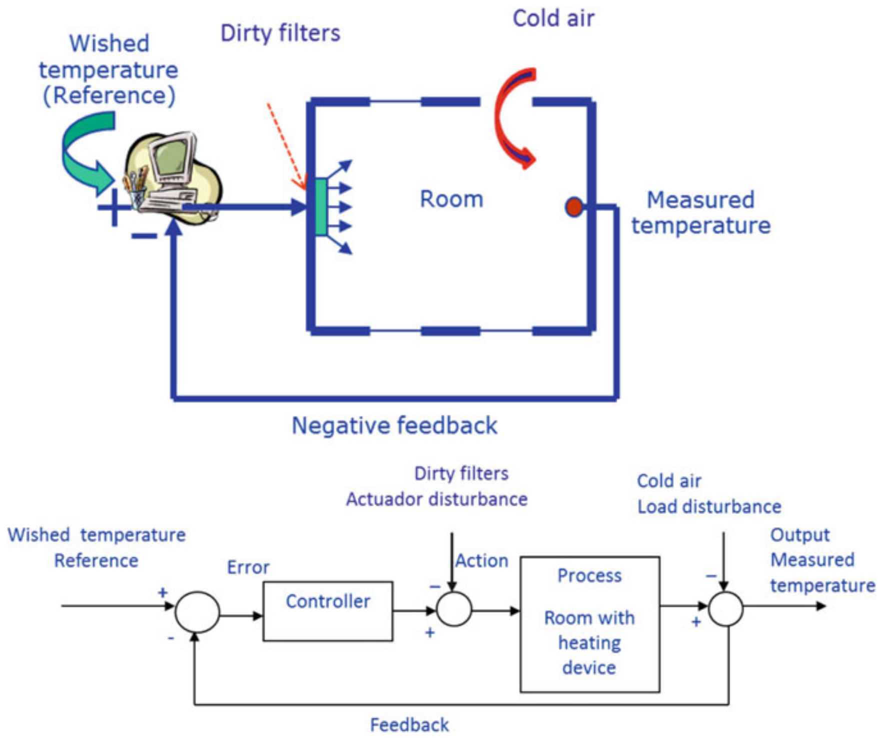


Fig. 3 The air conditioning to illustrate the closed-loop control problem

This very general concept of learning can be analysed in more detail and a taxonomy of controllers can be proposed, as illustrated in Fig. 2, according to the level of their learning ability [4].

Level L^0 corresponds to the fixed control theories, including the several levels of robustness controllers:

- H^0 no robustness
- H^1 gain and phase margin controllers
- H^2 optimal controller

… until H^{inf} controllers.

Level $\mathbf{L^1}$ parameter learning—controllers with on-line parameter estimation in linear case, in neural networks (connection weights and parameters of activation functions), in fuzzy systems (scale factors for fuzzification and defuzzification, centers and widths of membership functions).

Level $\mathbf{L^2}$ structure learning—gain-scheduling controllers, switching controllers, on-line order estimation, on-line fuzzy rule base construction (number of antecedents, association of antecedents and consequents, number of rules), on-line pruning and growing techniques in neural networks and neuro-fuzzy systems, etc. Controller reconfiguration in fault diagnosis could also be included here.

Level $\mathbf{L^3}$ trajectory learning—for example optimisation methods at supervisory level for process control, robot path planning, etc.

Level $\mathbf{L^4}$ task learning—short-term production planning in process control, autonomous agents (robot) task planning.

Level $\mathbf{L^5}$ goal learning—long-term production planning in process control, the capability of a system to find its own goals in a complex multi-system structure (for example in a set of autonomous robots working together).

Level $\mathbf{L^6}$ learning organizations—including the concepts of medium and long term learning in a multi-system changing organization of agents (systems) with complete autonomy in a dynamic environment.

Actually the levels $\mathbf{L^1}$, $\mathbf{L^2}$, $\mathbf{L^3}$, $\mathbf{L^4}$, are well developed, although still subject of intensive and extensive research. Levels $\mathbf{L^5}$ and $\mathbf{L^6}$ are still in an exploratory phase, particularly in mobile robotics.

In the highest levels of learning one meets Artificial Intelligence and Machine Learning. In this framework, control is mainly information processing and decision making, and as such can be defined as [4]:

Definition 4 Control. Control is the art and the science of information processing in order to make a decision (by an artificial controller or an agent) to act over mass and energy to reach some target defined by the system builder.

This definition involves the three bases of modern science: mass, energy, and information. It allows including all existent approaches for controller synthesis and those to come yet.

4 The Role of Fuzzy Control and the Challenge of Control Education

Teaching control in the proposed new concept space is a challenge that has not yet received the appropriate effort from the control community. In fact, in many universities control is taught by the same books as in was 50 years ago (the main difference is that the new editions of books have colored figures instead of black and white ones).

Fuzzy Logic [7, 8] can give a very important contribution to this aim. The first course in control should be fuzzy control [13], but using mainly graphical tools.

One important aspect of control is a clear practical understanding of the basic concepts: open loop, feedback, closed loop, reference, error, input, output, load disturbance, actuator disturbance and the dynamic behavior of the control system (speed of response, over/undershoot, settling time, etc.). These concepts can be illustrated by simulated examples, without using complex plane representation, and connecting them to real life. One interesting example is the air conditioning of a room with window, illustrated in Fig. 3, which everybody knows quite well.

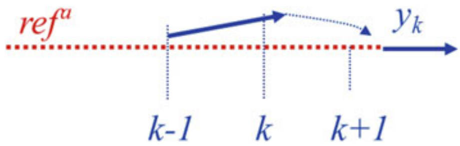


Fig. 4 Rule: IF error is Negative Small AND variation of error is Negative Small THEN control signal change is Negative Big

It is easy to illustrate the several disturbances that can affect the system behavior and need to be controlled. Students can put themselves in the role of manual controllers of heating/cooling the room in the several realistic situations that can happen.

After defining the control system the control game starts. The objective is to compute the appropriate action for the actual state of the process. It is not enough to know the error itself at instant k, e_k, but also how is it evolving. Did it increase from the previous instant or not? The increase/decrease of the error between the actual instant and the previous one is defined as $\Delta \mathrm{e_k} = \mathrm{e_k} - \mathrm{e_{k-1}}$ is approximated by the derivative of the error. It is more intuitive for students to work with the so-called differential control, where at each instant the controller gives the change in action and not the action itself. The action itself is the integral of the changes since the start. Moreover, this corresponds to the inclusion of integral control assuring null steady state error for constant (step) references and load disturbances.

A verbal control strategy can be developed. For example it is intuitive to say that if it is a bit hot (error is Negative Small) and the temperature is going up slowly (change of error is Negative Small), then heating must be quite negative: give substantial cold air to low down the temperature towards the reference. So a verbal control rule can be stated as in Fig. 4.

Similarly, if the room is very cold (error is Positive Big) and it is getting colder fastly (change of error is Positive Big), then warm up as much as you can (change of action must be Positive Big). This is illustrated in Fig. 5.

Facing the other possible situations, the well-known table of Fig. 6 is obtained. In (a) the granularity of information is lower (more detail) than in (b). The number of rules increases exponentially with the inverse of the granularity.

Using fuzzy logic it is quite simple to pass from these verbal control rules to effective control signals, i.e., to compute with words [9].

Firstly the fuzzy sets for NB, NS, ZE (Zero), PS, PB must be defined. Then the rules must be written and putted to work.

The fuzzy logic development is based on graphical tools more than on the formal mathematical fuzzy relations, exploring the intuitive thinking of students.

The Fuzzy Logic Toolbox of Matlab [10] has a very interesting and useful user interface, the GUI *fisedit*, allowing to implement all these steps. Granularity, shape of membership functions, inference operators, defuzification techniques, are easily implemented in this GUI. The GUI produces a *controller.fis* file. The toolbox has also two blocks for Simulink allowing to run this *controller.fis* file, by just indicating the name *controller.fis* in these blocks dialog window. One of the blocks has a "rule viewer" allowing to observe in real-time the firing of the rules. This dynamic simulation is very interesting for students.

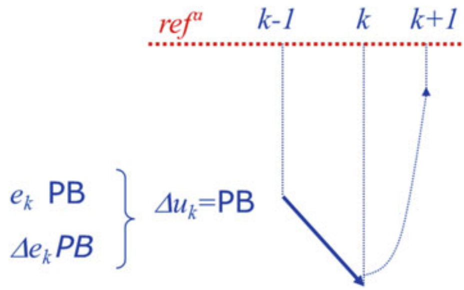


Fig. 5 Rule: IF error is Positive Big AND variation of error is Positive Big THEN control signal change is Positive Big

Fig. 6 The well known fuzzy rule bases of a PI fuzzy control **a** with 25 rules, **b** with 9 rules

(a)

e_k \ Δe_k	NB	NS	ZE	PS	PB
NB	NB	NB	NB	NS	ZE
NS	NB	NB	NS	ZE	PS
ZE	NB	NS	ZE	PS	PB
PS	NS	ZE	PS	PB	PB
PB	ZE	PS	PB	PB	PB

(b)

e_k \ Δe_k	N	ZE	P
N	N	N	Z
ZE	N	Z	P
P	Z	P	P

The theory underlying the firing of the rules is studied mainly by graphical means, as in Fig. 7 for the firing of two rules using minimum in firing strength (conjunction of antecedents) and the product for Mamdani implication. The aggregation of disjunctive rules is made by max-prod.

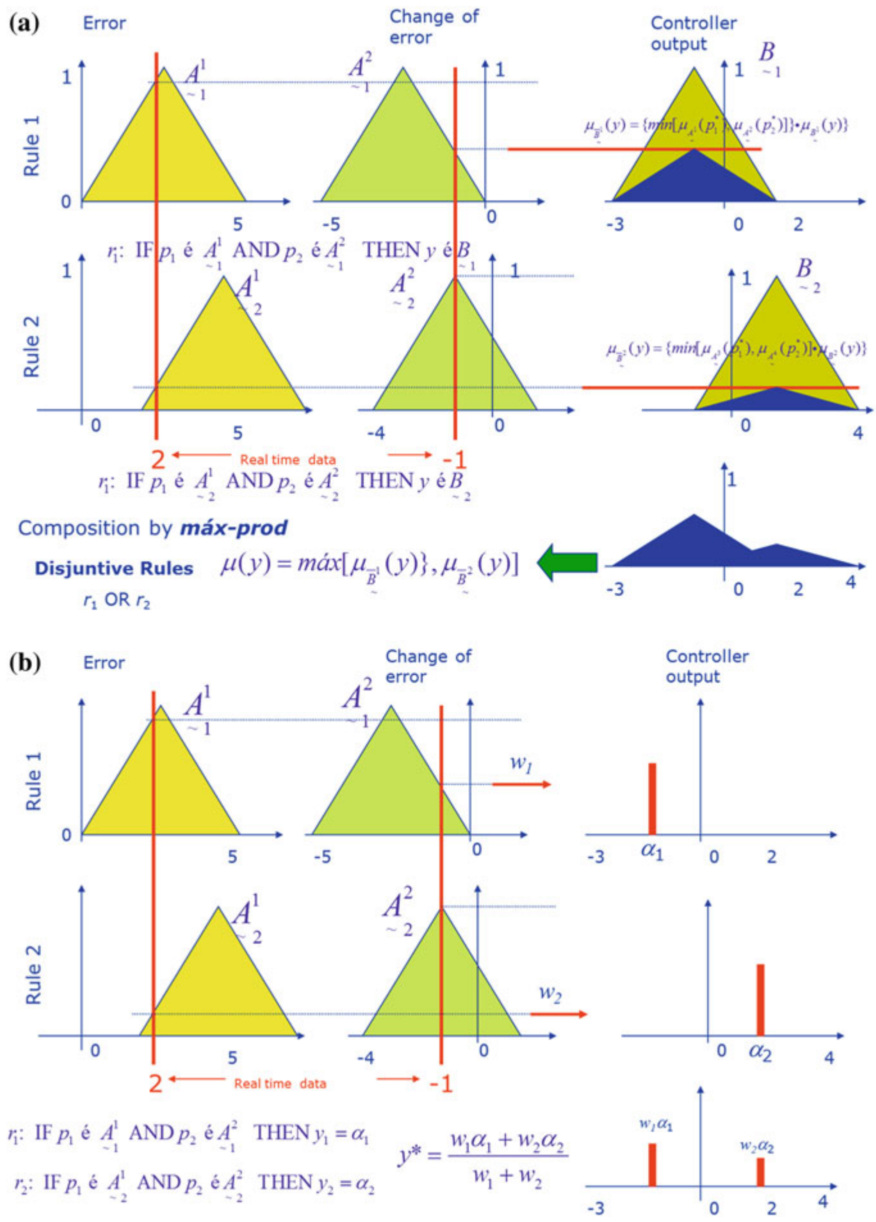


Fig. 7 Rule firing and aggregation: **a** Mamdani type [11], **b** Sugeno type [12]

The defuzzification is made usually by the centre of gravity method, since it is the more intuitive and there is actually no problem with its fast computation. The Simulink block diagram is completed with the reference input, load and actuator disturbances, visualization tools, and blocks to compute the control system

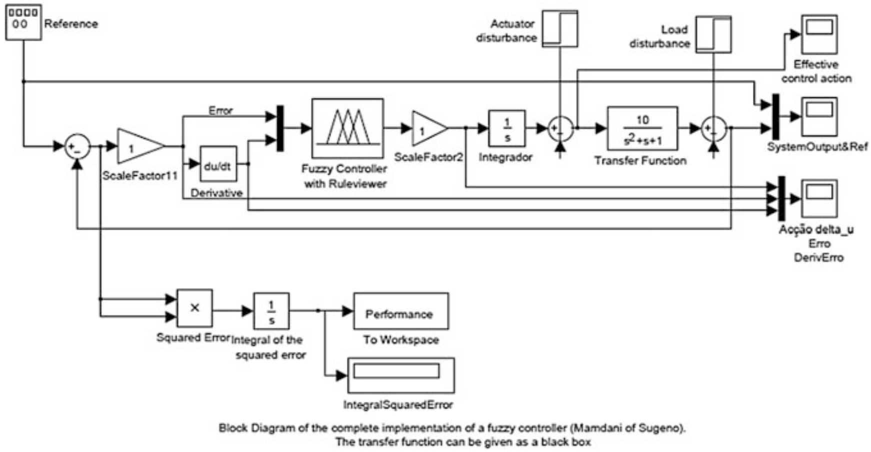


Fig. 8 Complete Simulink block diagram for fuzzy control implementation

performance. Figure 8 shows an example of the complete block diagram. Note that the controller is running a *fis* file that must be in the same directory. There are two scale factors, one in the error signal and another in the controller output trying to maintain these signals in the normalized interval [−1 1].

Several types of reference signals are used (step, sinusoidal, etc.) and the students are able to tune the fuzzy rules in order to obtain acceptable performance. Tuning the fuzzy rules is performed by shaping the membership functions of the antecedents and consequents and by fixing the scale factors. A typical result for a square wave reference is given in Fig. 8. By this exercising the student develop their capability to think about closed loop systems behaviour.

5 Some Experience and Conclusions

This approach is being followed for some years in a Soft Computing course at MSc level in Informatics Engineering and Biomedical Engineering, at the University of Coimbra, in which artificial neural networks and neuro-fuzzy systems are also studied. Fuzzy Logic is taught in two hours, fuzzy rule based systems in two hours, and the students have 4 h in class plus homework to develop and implement their fuzzy controller for a system given to them as a black box with a transfer function inside (Biomedical students know what a transfer function represents, but Informatics students do not). That is the first contact all of them have with control.

A learning-by-doing approach is followed. It is interesting to observe in the classroom how students amuse themselves putting the fuzzy controller working, improving the performance by trial and error, and how happy they are when they finally find a good control system, which most of them succeed. Typical results are shown in Figs. 9 and 10.

Table 1 Overview of the historical relation between technology and control [5]

Type of control	Information processing	Source of knowledge	Dominant discipline
Intelligent control	Super-microcomputer	Computer science artificial intelligence	Information engineering
Adaptive control on-line identification	Digital microcomputer	Recursive estimation real time operating syst.	Computer engineering
State space theory	Digital computer	Numerical methods	Electrical/Comp. engineering
Classical theory (transfer function)	Electronic valve and analogue computer	Laplace transform and cables	Electrical engineering
Empirism and art	Mechanical parts	Levers, hammer and screw-driver	Mechanical engineering

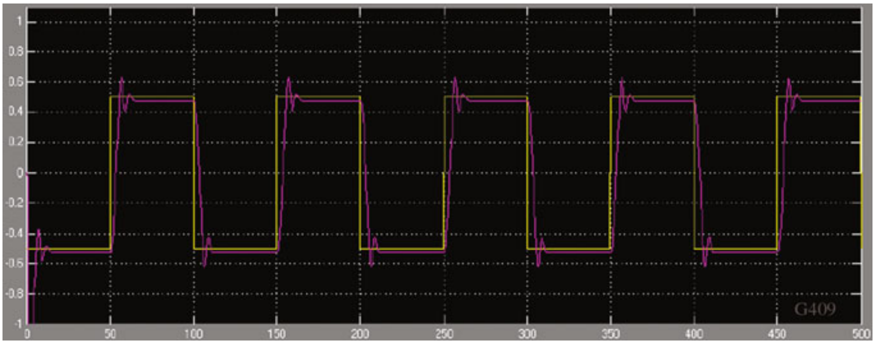


Fig. 9 A typical result for a Sugeno type fuzzy controller with 9 rules, for the regulator problem. In yellow the reference and in Cian the obtained system response with the fuzzy controller

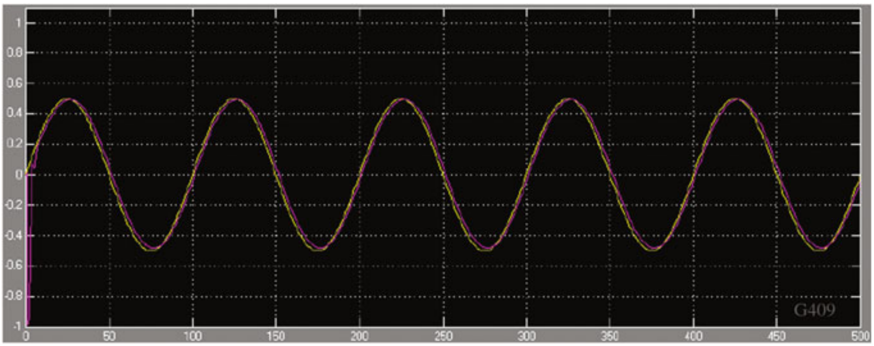


Fig. 10 A typical result for a Mamdani type Fuzzy Controller with 25 rules, for the tracking problem. In yellow the reference and in Cian the obtained system response with the fuzzy controller

In the author's opinion and experience students learn the basic concepts of control systems and develop their understanding of the set of problems involved in their synthesis for a good performance, stimulating them to further study with more formal mathematical tools (Figs. 9 and 10).

References

1. Bennett, S.: A History of Control Engineering (1800–1930). IEE and Peter Peregrinus, London (1979)
2. Bennett, S.: A History of Control Engineering (1930–1955). IEE and Peter Peregrinus, London (1993)
3. Mayr, O.: The Origins of Feedback Control. The MIT Press, Cambridge (1970)
4. Dourado, A.: Four definitions as a contribution to the third synthesis in systems and control. In: Proceedings of European Control Conference, ECC2001, Porto, 9–11 Sept 2001
5. Dourado, A.: A challenge for a new organization in systems and control curricula. In: Proceedings of the 17th World Congress of the International Federation of Automatic Control Seoul, Korea, 6–11 July 2008, pp. 30337–10342
6. Antsaklis, P.: Final Report of task Force in Intelligent Control. Technical Committee on Intelligent Control, IEEE Control Systems Society (1993)
7. Zadeh, L.A.: Fuzzy algorithms. Inf. Control **12**(2), 94–102 (1968)
8. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
9. Zadeh, L.A.: From computing with numbers to computing withwords—from manipulation of measurements to manipulation of perceptions. Int. J. Appl. Math. Comput. Sci. **12**(3), 307–324 (2002)
10. The Mathworks Inc, 3 Apple Hill Drive Natick, MA 01760-2098 United States. www.mathworks.com
11. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man Mach. Stud. **7**, 1–13 (1975)
12. Terano, T., Asai, K., Sugeno, M.: Applied Fuzzy Systems. Academic Press Inc, Boston (1994)
13. Yager, R., Filev, D.: Essentials of Fuzzy Modeling and Control. Wiley, NewYork (1994)

Nonlinear Control for Multiple-Input and Multiple-Output Nonlinear Systems with PB Models Based on I/O Linearization

Tadanari Taniguchi, Luka Eciolaza and Michio Sugeno

Abstract This chapter proposes a servo control of multiple-input multiple-output (MIMO) nonlinear systems approximated by piecewise bilinear (PB) models. The approximated model is found to be fully parametric. The input-output (I/O) feedback linearization is applied to stabilize PB control systems. The controller can be represented as a Look-Up-Table (LUT). We apply the servo control based on PB models to the Caltech ducted fan model. Although the controller is simpler than the conventional I/O feedback linearizing controller, the control performance based on PB model is the same as the conventional one. The example is shown to confirm the feasibility of our proposals by computer simulations.

1 Introduction

In recent years, piecewise linear (PL) systems which are fully parametric have been intensively studied in connection with nonlinear systems [1–4]. We are interested in the parametric piecewise approximation of nonlinear control systems based on the original idea of PL approximation. The PL approximation has general approximation capability for nonlinear functions with a given precision. However, it is difficult to handle some PL system such as simplexes in the rectangular coordinate system. To overcome this difficulty, one of the authors suggested to use

T. Taniguchi (✉)
Tokai University, Hiratsuka, Kanagawa 2591292, Japan
e-mail: taniguchi@tokai-u.jp

L. Eciolaza · M. Sugeno
European Centre for Soft Computing, 33600 Mieres, Asturias, Spain
e-mail: luka.eciolaza@softcomputing.es

M. Sugeno
e-mail: michio.sugeno@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_17,

the piecewise bilinear (PB) approximation [5]. We note that a bilinear function as a basis of PB approximation is, as a nonlinear function, the second simplest one after a linear function.

This chapter deals with the PB approximation of nonlinear control systems and discusses the modeling and the servo controller design based on I/O linearization. Our model has the following features. (1) The PB model is derived from fuzzy if-then rules with singleton consequents. (2) It is built on piecewise hyper-cubes partitioned in the state space. (3) It has general approximation capability for nonlinear systems. (4) It is a piecewise nonlinear model, the second simplest after a PL model. (5) It is continuous and fully parametric. So far we have shown the necessary and sufficient conditions for the stability of PB systems with respect to Lyapunov functions in the two dimensional case [6, 7] where membership functions are fully taken into account. We derived the stabilizing conditions [8, 9] based on the feedback linearization, where [8] applies the input-output linearization and [9] applies the full-state linearization. In the feedback linearization, we design a state feedback controller which transforms a nonlinear system into an equivalent linear system. In the PB approximation of nonlinear control systems, we studied a regulation control based on I/O linearization [9] and full feedback linearization [8], nonlinear model predictive control [10], design of look-up-table (LUT) controllers [11] and robust controller [12]. These control systems are the single-input and single-output (SISO) nonlinear systems. In the past studies, an approximation procedure considering the MIMO nonlinear systems was not discussed.

This chapter proposes a servo control with PB models for the MIMO nonlinear system. We also design an LUT controller, where the LUT-controllers are widely used for industrial control, in particular, for vehicle control because of its simplicity and visibility. We apply the method for the servo control based on PB models to the Caltech ducted fan model. Although the controller is simpler than the conventional I/O feedback linearization controller, the control performance based on PB model is the same as the conventional one.

This chapter is organized as follows. Section 2 presents the canonical form of PB models. Section 3 proposes the servo control for MIMO nonlinear systems with PB modeling and I/O linearization. Section 4 shows the example demonstrating the feasibility of the proposed methods, and Sect. 5 gives conclusions.

2 Canonical Form of Piecewise Bilinear Models

2.1 Open-Loop Systems

In this section, we introduce the PB models suggested in [5]. We deal with the two dimensional case without loss of generality. Define a vector $d(\sigma_1, \sigma_2)$ and a rectangle $R_{\sigma_1\sigma_2}$ in the two-dimensional space as, respectively,

$$
\begin{aligned}
&d(\sigma_1,\sigma_2) \equiv (d_1(\sigma_1), d_2(\sigma_2))^T, \\
&R_{\sigma_1\sigma_2} \equiv [d_1(\sigma_1), d_1(\sigma_1+1)] \times [d_2(\sigma_2), d_2(\sigma_2+1)].
\end{aligned} \tag{1}
$$

σ_1 and σ_2 are integers: $-\infty < \sigma_1, \sigma_2 < \infty$ where $d_1(\sigma_1) < d_1(\sigma_1+1)$, $d_2(\sigma_2) < d_2(\sigma_2+1)$ and $d(0,0) \equiv (d_1(0), d_2(0))^T$. The superscript ^{T} denotes *transpose* operation.

For $x \in R_{\sigma_1\sigma_2}$, the PB system is expressed as

$$
\begin{cases}
\dot{x} = \sum\limits_{i_1=\sigma_1}^{\sigma_1+1} \sum\limits_{i_2=\sigma_2}^{\sigma_2+1} \omega_1^{i_1}(x_1)\omega_2^{i_2}(x_2) f(i_1, i_2), \\
x = \sum\limits_{i_1=\sigma_1}^{\sigma_1+1} \sum\limits_{i_2=\sigma_2}^{\sigma_2+1} \omega_1^{i_1}(x_1)\omega_2^{i_2}(x_2) d(i_1, i_2),
\end{cases} \tag{2}
$$

where

$$
\begin{cases}
\omega_1^{\sigma_1}(x_1) = (d_1(\sigma_1+1) - x_1)/(d_1(\sigma_1+1) - d_1(\sigma_1)), \\
\omega_1^{\sigma_1+1}(x_1) = (x_1 - d_1(\sigma_1))/(d_1(\sigma_1+1) - d_1(\sigma_1)), \\
\omega_2^{\sigma_2}(x_2) = (d_2(\sigma_2+1) - x_2)/(d_2(\sigma_2+1) - d_2(\sigma_2)), \\
\omega_2^{\sigma_2+1}(x_2) = (x_2 - d_2(\sigma_2))/(d_2(\sigma_2+1) - d_2(\sigma_2)),
\end{cases} \tag{3}
$$

and $\omega_1^{i_1}(x_1), \omega_2^{i_2}(x_2) \in [0,\ 1]$. In the above, we assume $f(0,0) = 0$ and $d(0,0) = 0$ to guarantee $\dot{x} = 0$ for $x = 0$. Due to lack of space, we use $\omega_1^{i_1}$ and $\omega_2^{i_2}$ in $\omega_1^{i_1}(x_1)$ and $\omega_2^{i_2}(x_2)$ from the following pages.

A key point in the system is that the state variable x is also expressed by a convex combination of $d(i_1, i_2)$ with respect to $\omega_1^{i_1}$ and $\omega_2^{i_2}$ just as in the case of $\dot{x}$. As is seen in Eq. (3), x is located inside $R_{\sigma_1\sigma_2}$ which is a rectangle: a hypercube in general. That is, the expression of x is polytopic with four vertices $d(i_1, i_2)$. The model of $\dot{x} = f(x)$ is built on a rectangle including x in the state space and it is also polytopic with four vertices $f(i_1, i_2)$. We call this form of the canonical model (2) parametric expression.

Representing $\dot{x}$ with x in Eqs. (2) and (3), we can obtain the state space expression of the model which is found to be bilinear (bi-affine) [5]. Therefore, the derived PB model has simple nonlinearity. In the case of the PL approximation, a PL model is built on simplexes partitioned in the state space, triangles in the two dimensional case. Note that any three points in the three dimensional space are spanned with an affine plane: $y = a + bx_1 + cx_2$. A PL model is continuous. It is, however, difficult to handle simplexes in the rectangular coordinate system.

Also we can see that any four points in the three dimensional space can be spanned with a bi-affine plane: $y = a + bx_1 + cx_2 + dx_1x_2$. In contract to a PL model, a PB model as such is built on rectangles with the four vertices $d(i_1, i_2)$, on hyper-cubes in a general dimensional space, partitioned in the state space; it well matches the rectangular coordinate system. Therefore, PB models would be applicable to control purpose.

2.2 Closed-Loop Systems

We consider a two-dimensional nonlinear control system.

$$\begin{cases} \dot{x} = f_o(x) + g_o(x)u(x), \\ y = h_o(x). \end{cases} \tag{4}$$

The PB model (5) can be constructed from the nonlinear system (4).

$$\begin{cases} \dot{x} = f(x) + g(x)u(x), \\ y = h(x), \end{cases} \tag{5}$$

where

$$\begin{cases} f(x) = \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \omega_1^{i_1} \omega_2^{i_2} f_o(i_1, i_2), \\ g(x) = \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \omega_1^{i_1} \omega_2^{i_2} g_o(i_1, i_2), \\ h(x) = \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \omega_1^{i_1} \omega_2^{i_2} h_o(i_1, i_2), \\ \quad x = \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \omega_1^{i_1} \omega_2^{i_2} d(i_1, i_2). \end{cases} \tag{6}$$

The modeling procedure in the region $R_{\sigma_1\sigma_2}$ is as follows.

Algorithm 2.1
Piecewise bilinear modeling procedure

(1) Assign vertices $d(i_1, i_2)$ for $x_1 = d_1(\sigma_1), d_1(\sigma_1 + 1)$, $x_2 = d_2(\sigma_2), d_2(\sigma_2 + 1)$ of the state vector x, then the state space is partitioned into piecewise regions, see also Fig. 1.
(2) Compute the vertices $f_o(i_1, i_2)$, $g_o(i_1, i_2)$ and $h_o(i_1, i_2)$ in Eq. (6), by substituting the values of $x_1 = d_1(\sigma_1), d_1(\sigma_1 + 1)$ and $x_2 = d_2(\sigma_2), d_2(\sigma_2 + 1)$ into original nonlinear functions f_o, g_o and h_o in the system (4). Figure 1 illustrates the expression of $f_1(x)$, where $f(x) = (f_1(x), f_2(x))^T$ and $x \in R_{\sigma_1\sigma_2}$.

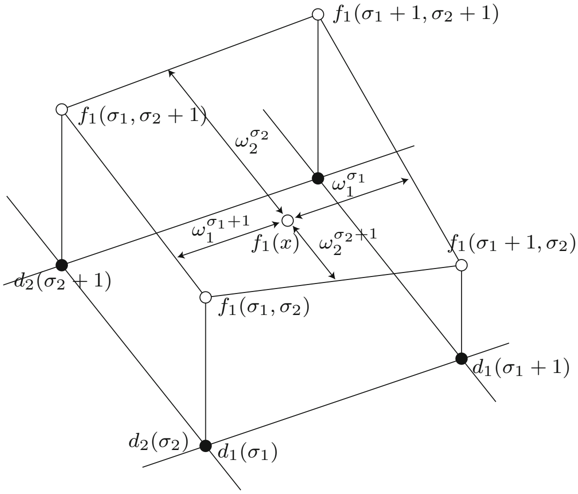


Fig. 1 Piecewise region $(f_1(x),\ x \in R_{\sigma_1\sigma_2})$

The overall PB model can be obtained automatically when all the vertices are assigned. Note that $f(x)$, $g(x)$ and $h(x)$ in the PB model coincide with those in the original system at the vertices of all the regions.

3 Servo Control for MIMO Nonlinear Systems with PB Modeling and I/O Linearization

This section deals with the I/O linearization of nonlinear control systems approximated with PB models. We consider, in particular, nonlinear systems, and show their I/O linearization based on PB models in detail. Next we give a brief introduction to the I/O linearization [13] for PB control systems.

3.1 I/O Linearization

Consider the PB system for MIMO nonlinear system:

$$\begin{cases} \dot{x} &= f(x) + g_1(x)u_1(x) + \cdots + g_p(x)u_p(x), \\ y_1 &= h_1(x), \ldots, y_p = h_p(x), \end{cases} \tag{7}$$

where

$$
\begin{aligned}
f(x) &= \sum_{i_1=\sigma_1}^{\sigma_1+1} \cdots \sum_{i_n=\sigma_n}^{\sigma_n+1} \omega_1^{i_1} \cdots \omega_n^{i_n} f(i_1, \ldots, i_n), \\
g_j(x) &= \sum_{i_1=\sigma_1}^{\sigma_1+1} \cdots \sum_{i_n=\sigma_n}^{\sigma_n+1} \omega_1^{i_1} \cdots \omega_n^{i_n} g_j(i_1, \ldots, i_n), \\
h_k(x) &= \sum_{i_1=\sigma_1}^{\sigma_1+1} \cdots \sum_{i_n=\sigma_n}^{\sigma_n+1} \omega_1^{i_1} \cdots \omega_n^{i_n} h_k(i_1, \ldots, i_n), \\
x &= \sum_{i_1=\sigma_1}^{\sigma_1+1} \cdots \sum_{i_n=\sigma_n}^{\sigma_n+1} \omega_1^{i_1} \cdots \omega_n^{i_n} d(i_1, \ldots, i_n), \\
& \qquad j, k = 1, \ldots, p, \\
\omega_i^{\sigma} &= (d_i(\sigma+1) - x_i)/(d_i(\sigma+1) - d_i(\sigma)), \\
\omega_i^{\sigma+1} &= (x_i - d_i(\sigma))/(d_i(\sigma+1) - d_i(\sigma)), \\
& \qquad i = 1, \ldots, n, \ \sigma = \sigma_1, \ldots, \sigma_n,
\end{aligned}
$$

$x \in \Re^n$, $u \in \Re^p$, $y \in \Re^p$ and f, g_j are assumed to be smooth vector fields and h_k to be smooth functions.

$$
\begin{aligned}
\dot{y}_k &= \frac{\partial h_k}{\partial x} \big(f(x) + g_1(x) u_1 + \cdots + g_p(x) u_p \big) \\
&= L_f h_k(x) + \sum_{j=1}^{p} L_{g_j} h_k(x) u_j.
\end{aligned}
$$

If $L_{g_j} h_k(x) = 0$, then $\dot{y}_k = L_f h_k(x)$ is independent of u_j. Define γ_k to be the smallest integer such that at least one of the inputs appears in $y_k^{\gamma_k}$, that is,

$$
y_k^{\gamma_k} = L_f^{\gamma_k} h_k(x) + \sum_{j=1}^{p} L_{g_j} L_f^{\gamma_k - 1} h_k(x) u_j. \tag{8}
$$

with an least one of the $L_{g_j} L_f^{\gamma_k - 1} h_k(x) \neq 0$, for some x. Define the $p \times p$ matrix $G(x)$ as

$$
G(x) = \begin{pmatrix} L_{g_1} L_f^{\gamma_1 - 1} h_1(x) & \ldots & L_{g_p} L_f^{\gamma_1 - 1} h_1(x) \\ \vdots & \ddots & \vdots \\ L_{g_1} L_f^{\gamma_p - 1} h_p(x) & \ldots & L_{g_p} L_f^{\gamma_p - 1} h_p(x) \end{pmatrix}.
$$

There exists vector relative degree $\gamma_1, \ \gamma_2, \ \ldots, \ \gamma_p$ of the system (7) at x_0 if

$$L_{g_j}L_f^l h_k(x) = 0, \quad 0 \le l \le \gamma_k - 2$$

for $j = 1, \cdots, p$ and the matrix $G(x_0)$ is nonsingular. If the system (7) has well defined vector relative degree, then (8) is written as

$$\begin{pmatrix} y_1^{\gamma_1} \\ \vdots \\ y_p^{\gamma_p} \end{pmatrix} = \begin{pmatrix} L_f^{\gamma_1} h_1 \\ \vdots \\ L_f^{\gamma_p} h_p \end{pmatrix} + G(x)\begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}.$$

Since $G(x_0)$ is non-singular, it follows that $G(x) \in \Re^{p \times p}$ is bounded away from nonsingularity for $x \in U$ a neighborhood U of x_0, meaning that $G^{-1}(x)$ and has bounded norm on U. Then the state feedback control law

$$u = A + Bv = -G^{-1}(x)\begin{pmatrix} L_f^{\gamma_1} h_1 \\ \vdots \\ L_f^{\gamma_p} h_p \end{pmatrix} + G^{-1}(x)v \tag{9}$$

yields the linear closed loop system

$$\begin{pmatrix} y_1^{\gamma_1} \\ \vdots \\ y_p^{\gamma_p} \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}.$$

A coordinate transformation vector $z = (z_1, \ldots, z_p)^T$, where

$$z_i = \left(h_i(x), L_f h_i(x), \ldots, L_{fs}^{\gamma_i - 1} h_i(x)\right)^T,$$

so that the system is transformed into the form

$$\dot{z} = Az + Bv, \tag{10}$$

which is linear and controllable.

Note that all the feedback linearizable PB systems (7) are transformed into the linear system (10). Therefore it is easy to design the stabilizing controller and analyze stability of the PB systems.

3.2 Servo Control

We apply a servo control to the PB system (7). According to the internal model principle, we consider the servo control for the extended model of (10):

$$\begin{cases} \dot{z} = Az + Bv, \\ \dot{\xi} = -Cz + r \end{cases} \tag{11}$$

where ξ is the extended state, C is the output matrix, $v = -Fz + K\xi$, F is the regulation gain, K is the gain of the servo control, r is the reference signal. The extended system (11) is transformed into

$$\dot{\bar{z}} = \bar{A}\bar{z} + \bar{r}, \tag{12}$$

where

$$\bar{z} = \begin{pmatrix} z \\ \xi \end{pmatrix}, \quad \bar{A} = \begin{pmatrix} A - BF & BK \\ -C & 0 \end{pmatrix}, \quad \bar{r} = \begin{pmatrix} 0 \\ r \end{pmatrix}.$$

The gains of F and K are designed such that $\bar{A}$ is Hurwitz. The derived linear controller v is substituted to the linearized controller (9). Figure 2 illustrates the block diagram of the servo control system. T means the coordinate transformation block.

4 Numerical Examples

In this section, we consider the Caltech ducted fan model [14] which is considered to be a Harrier aircraft (VTOL fighter) in hover mode or a thrust vectored aircraft in forward mode. We use a simplified model of the Harrier aircraft to demonstrate the feasibility of the proposed method. The aircraft model is given by

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -g \sin x_5 - \frac{d}{m} x_2 + \frac{\cos x_5}{m} u_1 - \frac{\sin x_5}{m} u_2 \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = g(\cos x_5 - 1) - \frac{d}{m} x_4 + \frac{\sin x_5}{m} u_1 + \frac{\cos x_5}{m} u_2 \\ \dot{x}_5 = x_6 \\ \dot{x}_6 = \frac{r_d}{J} u_1 \\ y_1 = h_1 = x_1, \quad y_2 = h_2 = x_3, \end{cases} \tag{13}$$

where outputs y_1 and y_2 are the position of the aircraft center of mass, x_5 is the angle of a point on the mass axis of the fan. We choose the system parameters as $g = 9.81$, $m = 8.5$, $d = 0.5$, $r_d = 0.26$ and $J = 0.05$, while the initial condition is $x(0) = (5.0, 5.0, 5.0, 0, -0.9\pi/2, 0)^T$.

The state-space is divided by the following vertices.

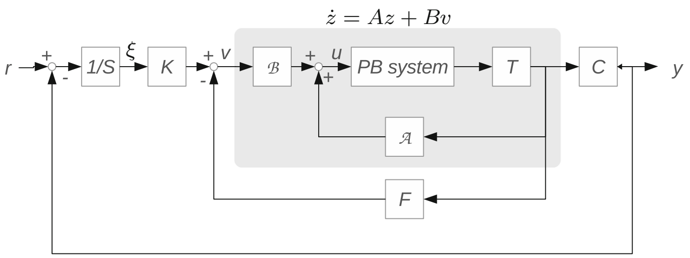


Fig. 2 Block diagram of the servo control system

$$\begin{aligned} &x_1 \in \{-10, 0, 10\}, \quad x_2 \in \{-10, 0, 10\}, \quad x_3 \in \{-10, 0, 10\},\\ &x_4 \in \{-10, 0, 10\}, \quad x_5 \in \{-2\pi, -5\pi/3, \quad -4\pi/3, \cdots, 2\pi\},\\ &x_6 \in \{-20, 0, 20\}, \quad x_7 \in \{-10, 0, 10\}, \quad x_8 \in \{-10, 0, 10\}, \end{aligned}$$

then the PB model is constructed as

$$\begin{cases} \dot{x} = f(x) + g_1(x)u_1 + g_2(x)u_2,\\ y_1 = h_1 = x_1,\\ y_2 = h_2 = x_3, \end{cases}$$

where

$$\begin{aligned} f(x) &= \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_4=\sigma_4}^{\sigma_4+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \sum_{i_6=\sigma_6}^{\sigma_6+1} \omega_2^{i_2}\omega_4^{i_4}\omega_5^{i_5}\omega_6^{i_6} f(i_2, i_4, i_5, i_6),\\ f(x) &= (f_1, f_2, f_3, f_4, f_5, f_6)^T,\\ g_1(x) &= \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_5^{i_5} g_1(i_5), g_2(x) = \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_5^{i_5} g_2(i_5),\\ g_1(x) &= (g_{11}, g_{12}, g_{13}, g_{14}, g_{15}, g_{16})^T,\\ g_2(x) &= (g_{21}, g_{22}, g_{23}, g_{24}, g_{25}, g_{26})^T,\\ x &= \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_3=\sigma_3}^{\sigma_3+1} \sum_{i_4=\sigma_4}^{\sigma_4+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \sum_{i_6=\sigma_6}^{\sigma_6+1} \omega_1^{i_1}\omega_2^{i_2}\omega_3^{i_3}\omega_4^{i_4}\omega_5^{i_5}\omega_6^{i_6} d(i_1, i_2, i_3, i_4, i_5, i_6),\\ x &= (d_1, d_2, d_3, d_4, d_5, d_6)^T. \end{aligned}$$

Tables 1, 2 and 3 show the PB models of $f(x)$, $g_1(x)$ and $g_2(x)$, respectively. The derivatives $\dot{y}_1$ and $\dot{y}_2$ are obtained as

Table 1 PB models of $f(x) = (f_1, f_2, f_3, f_4, f_5, f_6)^T$ at $x = (x_1, d_2(1), x_3, d_4(1), d_5(i_3), d_6(1))^T$

$d_5(i_3)$	f_1	f_2	f_3	f_4	f_5	f_6
$d_5(1) = -2\pi$	−10.0	0.588	−10.0	0.588	−20.0	0
$d_5(2) = -5\pi/3$	−10.0	−7.91	−10.0	−4.32	−20.0	0
$d_5(3) = -4\pi/3$	−10.0	−7.91	−10.0	−14.1	−20.0	0
$d_5(4) = -\pi$	−10.0	0.588	−10.0	−19.0	−20.0	0
$d_5(5) = -2\pi/3$	−10.0	9.08	−10.0	−14.1	−20.0	0
$d_5(6) = -\pi/3$	−10.0	9.08	−10.0	−4.32	−20.0	0
$d_5(7) = 0$	−10.0	0.588	−10.0	0.588	−20.0	0
$d_5(8) = \pi/3$	−10.0	−7.91	−10.0	−4.32	−20.0	0
$d_5(9) = 2\pi/3$	−10.0	−7.91	−10.0	−14.1	−20.0	0
$d_5(10) = \pi$	−10.0	0.588	−10.0	−19.0	−20.0	0
$d_5(11) = 4\pi/3$	−10.0	9.08	−10.0	−14.1	−20.0	0
$d_5(12) = 5\pi/3$	−10.0	9.08	−10.0	−4.32	−20.0	0
$d_5(13) = 2\pi$	−10.0	0.588	−10.0	0.588	−20.0	0

Table 2 PB models of $g_1(x) = (g_{11}, g_{12}, g_{13}, g_{14}, g_{15}, g_{16})^T$ at $x = (x_1, d_2(1), x_3, d_4(1), d_5(i_3), d_6(1))^T$

$d_5(i_3)$	g_{11}	g_{12}	g_{13}	g_{14}	g_{15}	g_{16}
$d_5(1) = -2\pi$	0	0.118	0	0	0	5.20
$d_5(2) = -5\pi/3$	0	0.0588	0	0.102	0	5.20
$d_5(3) = -4\pi/3$	0	−0.0588	0	0.102	0	5.20
$d_5(4) = -\pi$	0	−0.118	0	0	0	5.20
$d_5(5) = -2\pi/3$	0	−0.0588	0	−0.102	0	5.20
$d_5(6) = -\pi/3$	0	0.0588	0	−0.102	0	5.20
$d_5(7) = 0$	0	0.118	0	0	0	5.20
$d_5(8) = \pi/3$	0	0.0588	0	0.102	0	5.20
$d_5(9) = 2\pi/3$	0	−0.0588	0	0.102	0	5.20
$d_5(10) = \pi$	0	−0.118	0	0	0	5.20
$d_5(11) = 4\pi/3$	0	−0.0588	0	−0.102	0	5.20
$d_5(12) = 5\pi/3$	0	0.0588	0	−0.102	0	5.20
$d_5(13) = 2\pi$	0	0.118	0	0	0	5.20

$$\dot{y}_1 = \frac{\partial h_1}{\partial x}\frac{dx}{dt} = L_f h_1 = f_1 = x_2,$$
$$\dot{y}_2 = \frac{\partial h_2}{\partial x}\frac{dx}{dt} = L_f h_2 = f_3 = x_4.$$

Since $L_{g_j} h_k = 0$, where $j, k = 1, 2$, we continue to calculate the derivatives $\ddot{y}_1$ and $\ddot{y}_2$ as

Table 3 PB models of $g_2(x) = (g_{21}, g_{22}, g_{23}, g_{24}, g_{25}, g_{26})^T$ at $x = (x_1, d_2(1), x_3, d_4(1), d_5(i_3), d_6(1))^T$

$d_5(i_3)$	g_{21}	g_{22}	g_{23}	g_{24}	g_{25}	g_{26}
$d_5(1) = -2\pi$	0	0	0	0.118	0	0
$d_5(2) = -5\pi/3$	0	−0.102	0	0.0588	0	0
$d_5(3) = -4\pi/3$	0	−0.102	0	−0.0588	0	0
$d_5(4) = -\pi$	0	0	0	−0.118	0	0
$d_5(5) = -2\pi/3$	0	0.102	0	−0.0588	0	0
$d_5(6) = -\pi/3$	0	0.102	0	0.0588	0	0
$d_5(7) = 0$	0	0	0	0.118	0	0
$d_5(8) = \pi/3$	0	−0.102	0	0.0588	0	0
$d_5(9) = 2\pi/3$	0	−0.102	0	−0.0588	0	0
$d_5(10) = \pi$	0	0	0	−0.118	0	0
$d_5(11) = 4\pi/3$	0	0.102	0	−0.0588	0	0
$d_5(12) = 5\pi/3$	0	0.102	0	0.0588	0	0
$d_5(13) = 2\pi$	0	0	0	0.1176	0	0

$$\begin{aligned}\ddot{y}_1 &= \frac{d}{dt}(L_f h_1) = L_f^2 h_1 + L_{g_1} L_f h_1 u_1 + L_{g_2} L_f h_1 u_2 \\ &= f_2 + g_{12} u_1 + g_{22} u_2, \\ &= \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_2^{i_2} \omega_5^{i_5} f_2(i_2, i_5) + \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_5^{i_5} \{g_{12}(i_5) u_1 + g_{22}(i_5) u_2\},\end{aligned}$$

$$\begin{aligned}\ddot{y}_2 &= \frac{d}{dt}(L_f h_2) = L_f^2 h_2 + L_{g_1} L_f h_2 u_1 + L_{g_2} L_f h_2 u_2 \\ &= f_4 + g_{14} u_1 + g_{24} u_2 \\ &= \sum_{i_4=\sigma_4}^{\sigma_4+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_4^{i_4} \omega_5^{i_5} f_4(i_4, i_5) + \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_5^{i_5} \{g_{14}(i_5) u_1 + g_{24}(i_5) u_2\}.\end{aligned}$$

Eqs. $\ddot{y}_1$ and $\ddot{y}_2$ are written as

$$\begin{aligned}\begin{pmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{pmatrix} &= \begin{pmatrix} f_2 \\ f_4 \end{pmatrix} + G \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_2^{i_2} \omega_5^{i_5} f_2(i_2, i_5) \\ \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_4=\sigma_4}^{\sigma_4+1} \omega_2^{i_2} \omega_4^{i_4} f_4(i_4, i_5) \end{pmatrix} + \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_5^{i_5} \begin{pmatrix} g_{12}(i_5) & g_{22}(i_5) \\ g_{24}(i_5) & g_{24}(i_5) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.\end{aligned}$$

Since $G(x)$ is nonsingular, the linearized control law

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = -G^{-1}(x)\begin{pmatrix} f_2 \\ f_4 \end{pmatrix} + G^{-1}(x)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{14}$$

yields the linear system:

$$\begin{pmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

The coordinate vector z is represented as

$$z = (h_1, L_f h_1, h_2, L_f h_2)^T = (h_1, f_1, h_2, f_3)^T,$$

then the linearized system is written as

$$\dot{z} = Az + Bv \tag{15}$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

The linear controller v is calculated as

$$v = -Fz$$

so that the linearized system (15) is stable.

We consider the extended model (12) for the servo control.

$$\begin{pmatrix} \dot{z} \\ \dot{\xi} \end{pmatrix} = \begin{pmatrix} A - BF & BK \\ -C & 0 \end{pmatrix}\begin{pmatrix} z \\ \xi \end{pmatrix} + \begin{pmatrix} 0 \\ r \end{pmatrix} \tag{16}$$

where

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

the extended state $\xi = (x_7, x_8)^T$, $x_7 \in \{-10, 0, 10\}$ and $x_8 \in \{-10, 0, 10\}$. The feedback gains of the linear controller $v = -Fz + K\xi$ are calculated as

$$F = \begin{pmatrix} 5.273 & 2.788 & 0 & 0 \\ 0 & 0 & 5.273 & 2.788 \end{pmatrix},$$
$$K = \begin{pmatrix} 2.121 & 0 \\ 0 & 2.121 \end{pmatrix}$$

so that the extended system (16) is stable. The linear controller v is substituted for the controller (14). Finally, the piecewise nonlinear controller is obtained as

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = -G^{-1}(x)\begin{pmatrix} f_2 \\ f_4 \end{pmatrix} + G^{-1}(x)\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \tag{17}$$

where

$$G^{-1} = \left(\sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_5^{i_5} \begin{pmatrix} g_{12}(i_5) & g_{22}(i_5) \\ g_{24}(i_5) & g_{24}(i_5) \end{pmatrix} \right)^{-1},$$

$$\begin{pmatrix} f_2 \\ f_4 \end{pmatrix} = \begin{pmatrix} \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \omega_2^{i_2} \omega_5^{i_5} f_2(i_2, i_5) \\ \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_4=\sigma_4}^{\sigma_4+1} \omega_2^{i_2} \omega_4^{i_4} f_4(i_4, i_5) \end{pmatrix},$$

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = -F \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_3=\sigma_3}^{\sigma_3+1} \sum_{i_4=\sigma_4}^{\sigma_4+1} \omega_1^{i_1} \omega_2^{i_2} \omega_3^{i_3} \omega_4^{i_4}\, d(i_1, d_2, i_3, i_4)$$
$$+ K \sum_{i_1=\sigma_7}^{\sigma_7+1} \sum_{i_2=\sigma_8}^{\sigma_8+1} \omega_7^{i_7} \omega_8^{i_8} d(i_7, d_8).$$

Note that the controllers (14) and (17) based on PB model are simpler than the conventional feedback linearizing controllers. Since the nonlinear terms of the controllers (14) and (17) are not the original nonlinear terms (e.g., $\sin x_5,\ \cos x_5$) but the piecewise approximation models.

The piecewise nonlinear controllers are applied to the original nonlinear system (13). Figure 3 shows the trajectories of $y_1 = x_1$ and $y_2 = x_3$ in the case of the reference signal $r = (0, 0)^T$. This case is the same as a regulation problem. Figure 4 shows the trajectories of $y_1 = x_1$ and $y_2 = x_3$ in the case of the reference signal $r = (2, 1)^T$.

We show that if piecewise nonlinear controllers (14) are designed with PB systems, the LUT controller can be easily obtained when all the vertices of the piecewise nonlinear controllers are assigned. The LUT-controller is widely used for industrial control, in particular, for vehicle control because of its simplicity and visibility.

In this example, the LUT controller (18) is constructed using the vertices of the piecewise nonlinear controllers (17).

Fig. 3 Output responses using the piecewise nonlinear controller $(r = (0, 0)^T)$

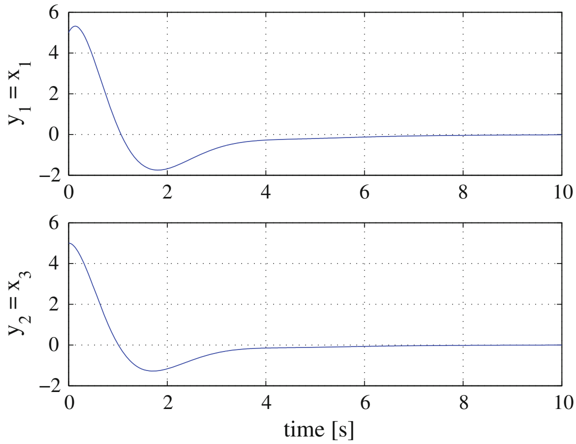
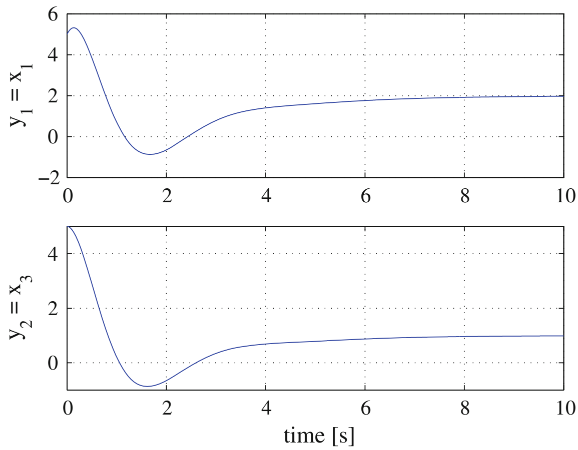


Fig. 4 Output responses using the piecewise nonlinear controller $(r = (2, 1)^T)$



$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \sum_{i_1=\sigma_1}^{\sigma_1+1} \sum_{i_2=\sigma_2}^{\sigma_2+1} \sum_{i_3=\sigma_3}^{\sigma_3+1} \sum_{i_4=\sigma_2}^{\sigma_4+1} \sum_{i_5=\sigma_5}^{\sigma_5+1} \sum_{i_7=\sigma_7}^{\sigma_7+1} \sum_{i_8=\sigma_8}^{\sigma_8+1} \times \omega_1^{i_1} \omega_2^{i_2} \omega_3^{i_3} \omega_4^{i_4} \omega_5^{i_5} \omega_7^{i_7} \omega_8^{i_8} u(i_1, i_2, i_3, i_4, i_5, i_7, i_8) \tag{18}$$

Table 4 shows a part of the LUT controller $u = (u_1, u_2)$ at $x = (d_1(1), d_2(1), d_3(1), d_4(1), x_5, d_7(1), d_8(1))$, where $x_5 = -2\pi, -5\pi/3, \ldots, 2\pi$. The LUT controller is also applied to the original nonlinear system (13) in the same initial conditions as the previous example. Figures 5 and 6 show the regulation and servo control results, respectively. From the simulation results of Figs. 5 and 6, the LUT controller (18) has same control performance as the piecewise nonlinear controllers (17).

Table 4 LUT controllers $u_1(d_1(1), d_2(1), d_3(1), d_4(1), x_5, d_7(1), d_8(1))$ and $u_2(d_1(1), d_2(1), d_3(1), d_4(1), x_5, d_7(1), d_8(1))$

$d_5(i_5)$	u_1	u_2
$d_5(1) = -2\pi$	319.6	319.6
$d_5(2) = -5\pi/3$	508.7	−158.7
$d_5(3) = -4\pi/3$	189.2	−561.6
$d_5(4) = -\pi$	−319.6	−486.3
$d_5(5) = -2\pi/3$	−508.7	−8.116
$d_5(6) = -\pi/3$	−189.2	394.8
$d_5(7) = 0$	319.6	319.6
$d_5(8) = \pi/3$	508.7	−158.7
$d_5(9) = 2\pi/3$	189.2	−561.6
$d_5(10) = \pi$	−319.6	−486.3
$d_5(11) = 4\pi/3$	−508.7	−8.116
$d_5(12) = 5\pi/3$	−189.2	394.8
$d_5(13) = 2\pi$	319.6	319.6

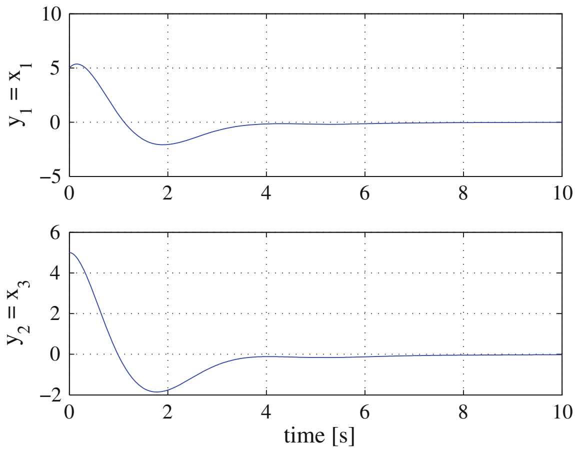


Fig. 5 Output responses using the LUT controller ($r = (0, 0)^T$)

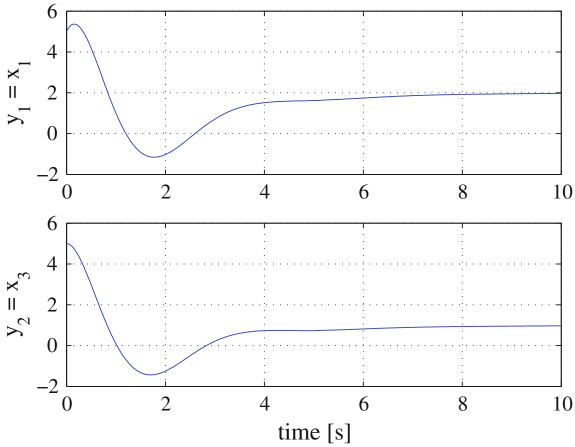


Fig. 6 Output responses using the LUT controller ($r = (2, 1)^T$)

5 Conclusions

This chapter has proposed a servo control of multiple-input multiple-output (MIMO) nonlinear systems approximated by piecewise bilinear (PB) models. The approximated model is found to be fully parametric. The input-output (I/O) feedback linearization has been applied to stabilize PB control systems. The controller can be represented as a Look-Up-Table (LUT). We have applied the servo control based on PB models to the Caltech ducted fan model. Although the controller is simpler than the conventional I/O feedback linearizing controller, the control performance based on PB model is the same as the conventional one. The example has been shown to confirm the feasibility of our proposals by computer simulations.

Acknowledgments This project was supported by a URP grant from Ford Motor Company which the authors thankfully acknowledge. Also, the authors would like to thank Dr. Dimitar Filev and Dr. Yan Wang of Ford Motor Company for his valuable comments and discussions. In addition, this work was supported by Grant-in-Aid for Young Scientists (B: 23700276) of The Ministry of Education, Culture, Sports, Science and Technology in Japan.

References

1. Sontag, E.D.: Nonlinear regulation: the piecewise linear approach. IEEE Trans. Autom. Control **26**, 346–357 (1981)
2. Johansson, M., Rantzer, A.: Computation of piecewise quadratic Lyapunov functions of hybrid systems. IEEE Trans. Autom. Control **43**(4), 555–559 (1998)
3. Imura, J.I., der Schaft, V.: Characterization of well-posedness of piecewise-linear systems. IEEE Trans. Autom. Control **45**, 1600–1619 (2000)
4. Feng, G., Lu, G.P., Zhou, S.S.: An approach to hinfinity controller synthesis of piecewise linear systems. Commun. Inf. Syst. **2**(3), 245–254 (2002)
5. Sugeno, M.: On stability of fuzzy systems expressed by fuzzy rules with singleton consequents. IEEE Trans. Fuzzy Syst. **7**(2), 201–224 (1999)
6. Sugeno, M., Taniguchi, T.: On improvement of stability conditions for continuous mamdani-like fuzzy systems. IEEE Tran. Syst. Man Cybern. Part B **34**(1), 120–131 (2004)
7. Taniguchi, T., Sugeno, M.: Stabilization of nonlinear systems based on piecewise Lyapunov functions. In: FUZZ-IEEE 2004, pp. 1607–1612 (2004)
8. Taniguchi, T., Sugeno, M.: Piecewise bilinear system control based on full-state feedback linearization. In: SCIS & ISIS 2010, pp. 1591–1596 (2010)
9. Taniguchi, T., Sugeno, M.: Stabilization of nonlinear systems with piecewise bilinear models derived from fuzzy if-then rules with singletons. In: FUZZ-IEEE 2010, pp. 2926–2931 (2010)
10. Taniguchi, T., Sugeno, M.: Nonlinear model predictive control of piecewise bilinear systems based on feedback linearization. In: ISIS 2011, pp. 131–134 (2011)
11. Taniguchi, T., Sugeno, M.: Design of LUT-controllers for nonlinear systems with PB models based on I/O linearization. In: FUZZ-IEEE 2012, pp. 1589–1596 (2012)
12. Taniguchi, T., Sugeno, M.: Robust stabilization of nonlinear systems modeled with piecewise bilinear systems based on feedback linearization. In: IPMU 2012, pp. 111–120 (2012)
13. Sastry, S.: Nonlinear Systems. Springer, New York (1999)
14. Murray, R.M.: Modeling of the caltech ducted fan. Control and Dynamical Systems. California Institute of Technology. Tech. Rep. (1996)

Part V
Neural Networks: Construction and Application

Improving the Model Convergence Properties of Classifier Feed-Forward MLP Neural Networks

Annamária R. Várkonyi-Kóczy, Balázs Tusor and József Bukor

Abstract Recently, the application of Artificial Neural Networks (ANNs) has become very popular. Their success is due to the fact that they are able to learn complex input-output mappings and are able to find relationships in unstructured data sets. Further, neural nets are relatively easy to implement in any application. In the last years, classification has become one of the most significant research and application area of ANNs because these networks have proved to be very efficient in the field. Unfortunately, a big difficulty of the usage of feed-forward multilayer perceptron (MLP) neural nets with supervised learning is that in case of higher problem complexity, the NN model may not converge during the training or in better cases needs a long training time which scales with the structural parameters of the networks and the quantity of input data. However, the training can be done off-line, this disadvantage may limit the usage of NN models because the training has a non-negligible cost and further, can cause a possibly non-tolerable delay in the operation. In this chapter, to overcome these problems, a new training algorithm is proposed which in many cases is able to improve the convergence properties of NN models in complex real world classification problems. On one hand, the accuracy of the models can be increased while on the other hand the training time can be decreased. The new training method is based on the well-known back-propagation algorithms, however with a significant difference: instead of the original input data, a reduced data set is used during the teaching phase. The reduction is the result of a complexity optimized classification procedure. In the

A. R. Várkonyi-Kóczy (✉)
Institute of Mechatronics and Vehicle Engineering, Óbuda University, Budapest, Hungary
e-mail: varkonyi-koczy@uni-obuda.hu

B. Tusor
Integrated Intelligent Systems Japanese-Hungarian Laboratory, Óbuda University, Budapest, Hungary
e-mail: balazs.tusor@gmail.com

J. Bukor
Department of Mathematics and Informatics, J. Selye University, Komarno, Slovakia
e-mail: bukorj@selyeuni.sk

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_18,

resulted new, reduced input data set, each input sample is replaced by the center of the cluster to which it belongs and these cluster centers are used during the training (each element once). As result, new, complex ambiguous classification problems can be solved with acceptable cost and accuracy by using feed-forward MLP NNs.

1 Introduction

Most of the engineering problems involve some kinds of classification tasks. Classification deals with finding a structure in a collection of unlabeled data. The procedure assigns a set of objects into groups whose members are similar in some way and are dissimilar to the objects belonging to other groups (so called clusters). Classification can be defined as the probably most important unsupervised learning problem and in most cases it leads to a (usually iterative) multi-objective optimization task. Many difficulties may occur during the classification, caused by e.g. the multi-dimensional spaces, time/data complexity, finding an adequate distance measure, non-unambiguous interpretation of the results, overlapping of the clusters, etc.

Such problems like probability estimation, feature selection, feature extraction, data mining, statistical data analysis, pattern recognition, machine learning etc. can strongly be related to classification. Although, there are many kinds of well-known traditional classification procedures based on statistical methods and Bayesian decision theory, in the past years soft computing based techniques have emerged as promising alternatives to conventional classification methods. Probably the most successful candidates of soft computing based classifiers are Artificial Neural Networks.

Feed-forward Artificial Neural Networks (*ANNs*) are systems based on the operation of biological neural networks. They are networks built of many simple processing units (neurons) arranged into two or more layers. Each unit can be connected to all units in the next layer. The most important ability of *ANNs* is that they are able to find relationship (structure) among the elements of unstructured data sets and can learn complex functions from input/output data-pairs, which means that they can adjust their parameters using iterative training algorithms in order to achieve accurate input-output mapping. Another advantage is that they are relatively easy to implement in any application.

Nowadays, because of the above advantages, the usage of *ANNs* has become quite widespread. They are generally used in three different types of problems: classification, function approximation, and data processing, among which the first is the possibly most important field. In this chapter, we also concentrate on *ANNs* used for classification.

To mention a couple of latest examples taken from different fields of research for the successful application of *ANNs* in solving classification problems, in [1] *ANNs* are used to map out the current state of mangrove trees. Today, this topic has become very important, because mangrove forests are unique and natural ecosystems that can be used to produce forestry product such as charcoal, timber, supply food to their surrounding marine life, and protect the inland from disturbances like erosion, flood, and tsunami. In [2] authors use *ANNs* to estimate the elevation of a sound-source, in order to achieve a robotic system that can sense in acoustics, specifically in elevation localization. In [3] *ANNs* are applied to recognize handwritten characters for English alphabets without feature extraction, etc.

As it can be seen from the above examples as well, the application area of *ANNs* is wide, covering possibly the whole engineering field. Although, besides the advantages, a big drawback of their usage is that they usually need a significant amount of time to be trained, which scales with the structural parameters of the networks and the quantity of input data. However, this can be done offline; the training has a non-negligible cost and further, can cause a delay in the operation.

In this chapter, we introduce a new training procedure for feed-forward *ANNs* used for classification. According to our experiences, the developed method improves the convergence properties of *NN* models, i.e. both the speed of the training and the accuracy of the *ANNs* are increased. The technique applies the well-known back-propagation algorithms, however with a significant difference: instead of the original input data, a reduced data set is used during the teaching. The reduction is the result of a complexity optimized classification procedure and during the training, each input training sample is substituted by the center of the cluster to which it belongs. For the determination of the training set, a reduced complexity clustering method, developed also by the authors of this chapter, is used. As result, new, complex ambiguous classification problems can be solved with acceptable cost and accuracy by using feed-forward MLP *NNs*.

The chapter is organized as follows: In Sect. 2 the proposed new clustering and training algorithms are discussed in details while in Sect. 3 the efficiency of the new training method is analyzed via different classification problems and the results are compared to that of the classical back-propagation training procedures. Finally, Sect. 4 is devoted to the conclusions and further works.

2 The New Training Algorithm

To increase the speed of the training of feed-forward *ANNs* used for classification, we have developed a new training procedure. The main idea is that instead of directly using the training data in the training phase, the data is clustered first and the *ANNs* are trained by using the centers of the obtained clusters. Figure 1 shows the block diagram of the general supervised learning scheme extended with a clustering step. The goal of the training is to tune the model (in our case the *ANN*) in order to make the output of the model (y) approximate the desired output (d) of the examined unknown system using the value (c) determined by the criteria function (typically a function of the approximation error). The model input is the cluster center (u') of the cluster which the actual input (u) belongs to. This means that the input data set has to be clustered before the training.

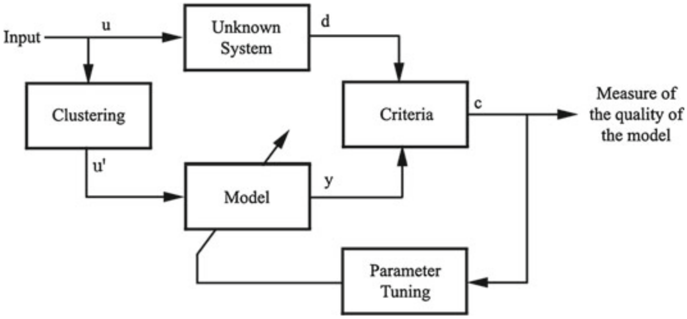


Fig. 1 The new supervised learning scheme extended with clustering

The role of the inserted clustering step is to reduce the quantity of the input data (u) used during the training and parallel with it to preserve the information contained in the original data set as much as possible, thus making the time required for learning the mapping of the unknown system to be modeled shorter while preserving the accuracy of the system's performance as if it was trained with the original data set. The results of the clustering step are the centers of the appointed clusters (u').

Using the cluster centers instead of the original input data has a further, very important effect. In many of the cases, the engineering or scientific problem in which we have to execute the classification can be very complex and it easily may happen that we try to solve the classification based on the knowledge of only a subset of the important influencing parameters. I.e., because of the lack of any exhaustive knowledge about the problem, we try to solve the task using only partial information, i.e. based on a subset of the characteristic parameters. This can happen either because of our insufficient knowledge about the problem or because of some complexity reduction purposes. As a consequence, it may easily happen that in the resulting (reduced) space of the training data, the clusters become overlapping. The separation could be done according to further parameters (dimensions), we are not aware of or we do not care about. The overlapping is possibly the most serious problem in classification and in most of the cases checkmates the successful training. In our case, the application of the cluster centers may eliminate this problem and can make possible to reach convergence.

For the training procedure, we have also developed a novel clustering method fundamentally based on the well-known k-means clustering method [4]. The principle of the new technique is similar to the original algorithms except the assignment procedure which has been modified (time reduced). While in the original k-means algorithms, the samples are compared to all of the existing clusters and the sample is assigned to the best fitting (nearest) cluster, in our method, the given sample is compared to the existing clusters one by one and the procedure stops if a "near enough" cluster is found. I.e., the sample gets assigned to the first cluster where the distance between the cluster's center and the given sample is less than a predefined, arbitrary value (distance factor). If there is no such cluster then a new cluster is appointed with the given sample assigned to it. Figure 2 shows the flowchart of the algorithm. The clustering is used on only one class at one time, so the incidental similarity between patterns of different classes will not cause any problem during the training.

The most important parameter of the new pre-clustering method is the clustering distance. Its value has a direct effect on the complexity of the training and on the accuracy of the classification. Choosing it too low results in too many clusters (nearing the complexity of the original training data set) while big clustering distances result in a fewer number of clusters, but less accurate classification.

The new training method has been first proposed for *Circular Fuzzy Neural Networks* (*CFNNs*) in [5]. According to the efficiency analyses, in case of *CFNNs*, the introduction of the pre-clustering step has reduced the training-time in average by more than 32 % together with a reduction of the number of input samples by approximately 54 %. (The difference between *CFNNs* and regular feed-forward *ANNs* is that the weights, biases, and outputs of *CFNNs* are fuzzy numbers (based on *Fuzzy Neural Networks* proposed in [6]), further their topology is realigned to a circular topology, and the connections between the input and hidden layers are trimmed).

In this chapter, authors extend the reduced complexity training method previously applied on CFNNs to the model fitting of feed-forward classifier MLP neural networks. In the next section the performance of the training and the convergence properties of the models are analyzed in details.

3 Examples

The effectiveness of the new training and clustering technique applied in feed-forward neural network models has been measured by the speed increase achieved in the training session. The experiments have been conducted for *ANNs* who had to learn to classify 2 dimensional inputs. Four different networks have been trained and compared in each experiment: one using the original data set and three using clustered data sets (generated by different clustering distances). The original data set in all cases consists of 500 samples (though the classes do not necessary contain exactly the same amount of samples; however their ratio has been set to more or less equal). After the training, each network has been tested on a separate test data set (not used during the training) containing 1,000 samples.

Let us denote the networks trained using the clustered data sets by A_i, B_i, and C_i with clustering distances set to 0.05, 0.1, and 0.2, respectively.

All the experiments have been conducted on an average PC (Intel Pentium 4 CPU 3.00 GHZ, 3 GB RAM, Windows XP 32-bit operating system). The experimental options for the training session of the *ANNs* in all experiments have been set to the following:

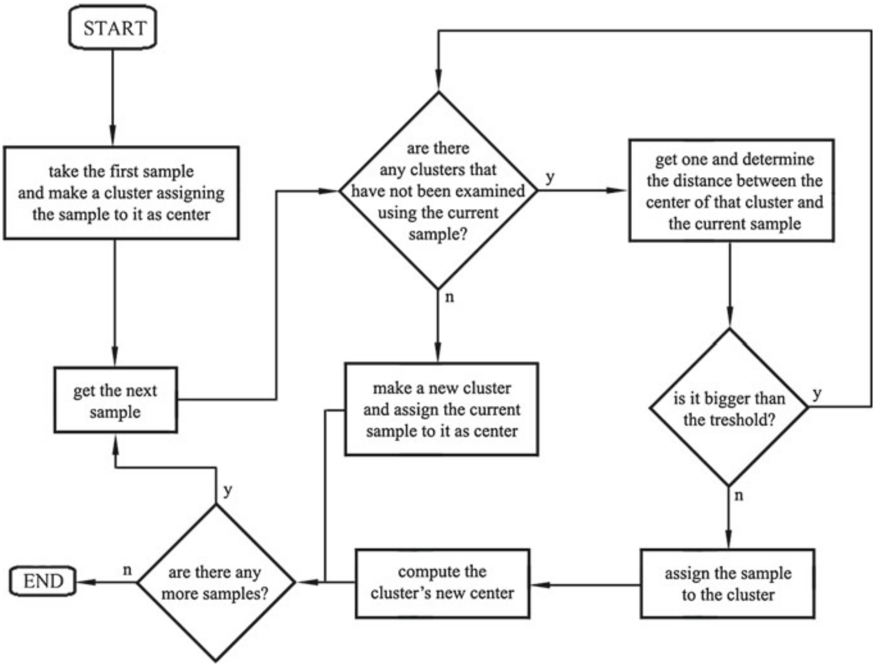


Fig. 2 The algorithm of the time reduced clustering step

- Learning rate: 0.8.
- Coefficient of the momentum method: 0.1.
- Number of hidden layer neurons: 10.
- Error threshold: 0.01 (except for networks A3 and A3', where 0.05 was used as threshold).

In the following, three typical examples, representing different problem complexities, are presented. The *first example* covers a relatively simple problem where three easily separable, non-overlapping classes are to be separated. In Fig. 3 the original, unclustered data set can be seen. The first neural network (with parameters corresponding to the description above) has been trained with this (unclustered) data set. Figure 4 shows the responses of the trained network on the whole input domain (0...1 $\times$ 0...1), by using a resolution of 0.05. As it can be well seen, the trained network can solve the problem successfully and is able to classify the key regions correctly.

Figure 5 shows the acquired cluster centers got by using the three, above defined clustering distances on the input data. As it is foreseeable, the clustering done with the smallest clustering distance results in the most and the clustering done with the largest clustering distance results in the least number of clusters (and thus cluster centers). To provide a better opportunity for the comparison, Fig. 6 presents the responses of the networks trained by the clustered data, over the training domain.

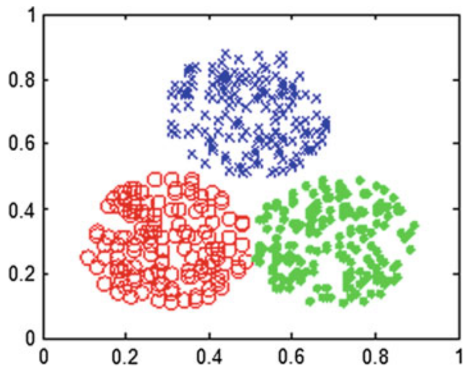


Fig. 3 Ex. 1: The original training data set

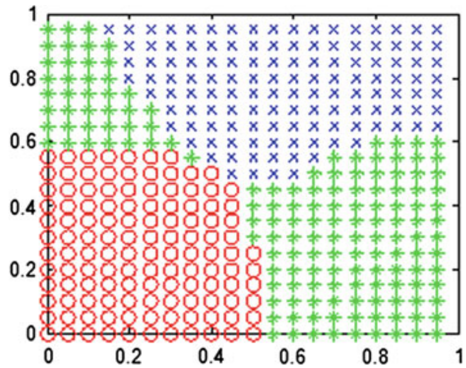


Fig. 4 Ex. 1: The performance of the network trained with the original training data set on the whole domain

Table 1 shows the number of clusters, as result of the clustering, the time required to train the neural networks, and the speed increase relative to the time required to train the network with the original, unclustered data set. It also shows the accuracy of all the trained networks on the test data set. Compared to the original 500 data points, the clustering that uses the largest clustering distance results in only 7 clusters (and thus 7 cluster centers), while the clustering using the smallest clustering distance results in 75 clusters. The network trained with the original data set can achieve 100 % accuracy, while among the networks trained with the clustered data sets, the worst accuracy equals 89.8 % (in case of the largest clustering distance), and the best accuracy reaches 99.7 % (in case of the smallest clustering distance). The training using the original data set needs 2 min and 38 s, while the training of network C_1 takes only 6 s (96 % speed increase compared to the original). B_1 takes 22 s (86 % speed increase) and A_1 takes 44 s (72 % speed increase).

Table 2 shows the accuracy of the trained networks over the training and the test data sets. Out of the three results, A_1 proves to be the most accurate, achieving almost 100 % accuracy on both data sets.

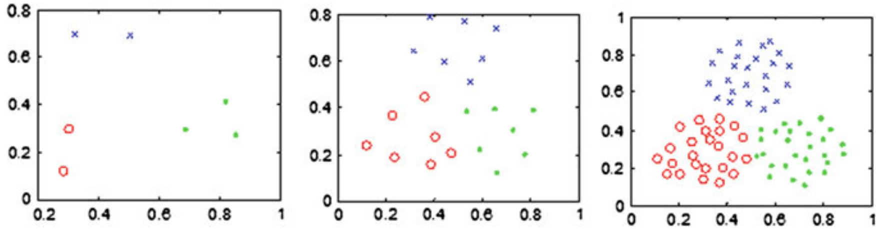


Fig. 5 Ex. 1: The centers of the resulted clusters

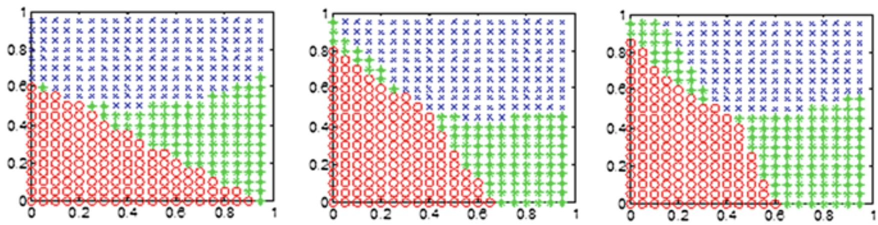


Fig. 6 Ex. 1: The classification results of the networks trained with the clustered training data sets on the whole domain

Table 1 The speed of the training in the first example

Network		Accuracy on the test data set (%)	Number of samples	Required time for training	Relative speed increase (%)
Original		100	500	2 min 38 s	–
Clustered	A_1	99.7	75	44 s	72
	B_1	95.6	21	22 s	86
	C_1	89.8	7	6 s	96

The *second example* illustrates a much harder problem in which there are three classes, one in the middle of the input domain and the other two at the corners overlapping with the first one (see Fig. 7).

The traditional training of the neural network (with parameters corresponding to the description above) has failed, i.e. the problem proved to be too complex for this network. On the other hand, the training based on the new method has led to success. Figure 8 shows the acquired cluster centers (similarly to the previous example) while in Fig. 9 the responses of the trained networks can be followed on the whole domain.

Table 3 shows the number of resulting clusters, the time required to train the neural networks, and the accuracy of the trained networks measured over the test data set.

Table 2 The accuracy of the training in the first example

Network		Accuracy of the training on the original training data set	Accuracy of the training on the test data set
Clustered	A_1	499/500 (99.8 %)	997/1000 (99.7 %)
	B_1	481/500 (99.8 %)	956/1000 (95.6 %)
	C_1	450/500 (90 %)	898/1000 (89.8 %)

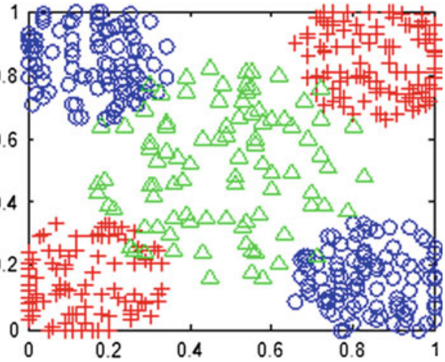


Fig. 7 Ex. 2: The original input (training) data set

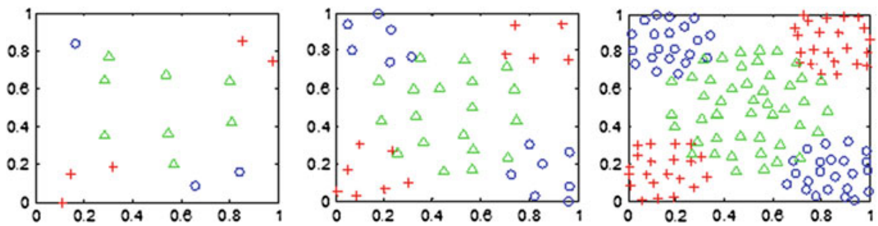


Fig. 8 Ex. 2: The centers of the resulted clusters

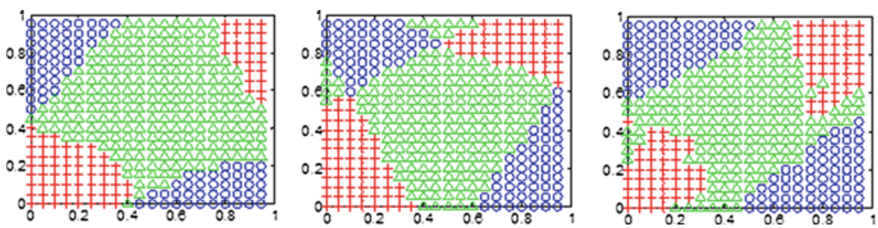


Fig. 9 Ex. 2: The classification results of the networks trained with the clustered training data sets on the whole domain

Table 3 The speed of the training in the second example

Network		Accuracy on the test data set (%)	Number of samples	Required time for training
Clustered	C_2	95.2	134	17 min 37 s
	B_2	91	44	1 min 47 s
	A_2	80.6	16	35 s

Table 4 The accuracy of the training in the second example

Network		Accuracy of the training on the original training data set	Accuracy of the training on the test data set
Clustered	A_2	476/500 (95.2 %)	971/1000 (97.1 %)
	B_2	455/500 (91 %)	977/1000 (97.7 %)
	C_2	403/500 (80.6 %)	888/1000 (88.8 %)

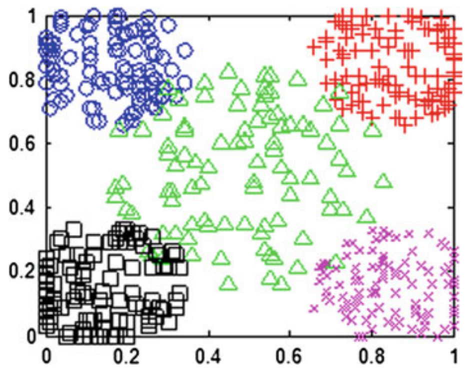


Fig. 10 Ex. 3: The original training data set

For comparison, Table 4 shows the accuracy of the trained networks on the training and the test data sets. Out of the three results, as in the previous example, A_2 proved to be the most accurate, achieving 95.2 % accuracy on both data sets after 17 min and 37 s of training. B_2 can achieve an accuracy of 91 %, after less than 2 min of training; and last but not least C_2 can achieve an accuracy of 80.6 % after only 35 s of training. This (in harmony with the results of the other experiments) clearly implies that there is an inverse correlation between the clustering distance and the time required for training. If the chosen clustering distance is too small, the number of the gained clusters converges to the quantity of the original data set, providing better coverage of the original data set however reducing the time reduction we gained from the clustering. On the other hand, if the chosen clustering distance is too big, it significantly reduces the number of clusters and thus, the time required for training, but with a burden of potentially providing worse coverage of the original data set, possibly resulting in less accuracy.

Finally, the *third example* presents a similar structure but more complex problem as is shown in example 2. In this case, the problem includes 5 clusters.

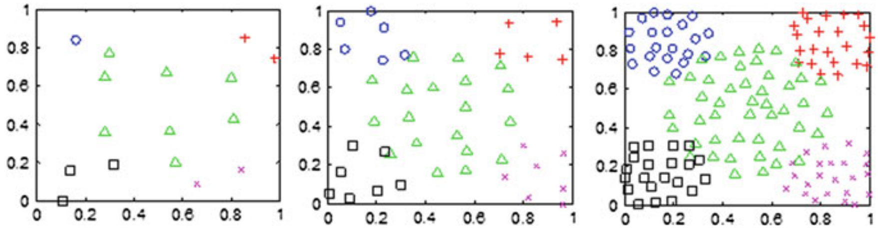


Fig. 11 Ex. 3: The centers of the resulted clusters

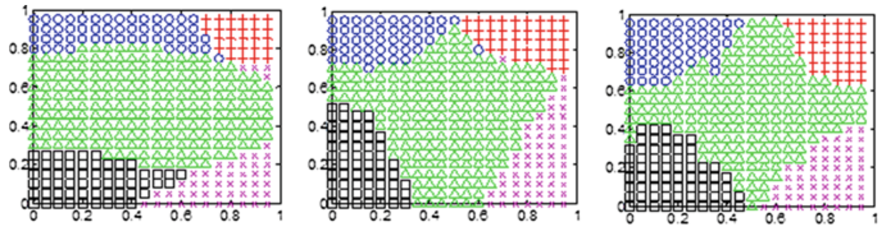


Fig. 12 Ex. 3: The performance of the networks trained with the clustered training data sets on the whole domain

Figure 10 shows the original, unclustered data set. The training of the neural network (with parameters corresponding to the description above) proved to be unsuccessful in this case, as well; the problem is too complex for the used network with the given parameters. Figure 11 presents the acquired cluster centers, similarly to the previous two examples and Fig. 12 shows the responses of the trained networks on the whole domain.

Table 5 shows the number of resulting clusters, the time required to train the neural network, and the accuracy of the trained networks measured over the test data set. Despite of the failed attempt of training the NN with the original data set, the network has again successfully learned the classification when the three clustered data sets have been used. The C_3 network achieves an accuracy of 96.6 % in 53 min and 11 s and B_3 can also achieve a relatively high, 89.8 % accuracy in less than 3 min.

For comparison, Table 6 shows the accuracy figures of the networks on the original training data set as well as on the test data set. It is interesting to remark that C_3 can achieve even better results on the test data set (97.8 %), while B_3 proves to be only slightly less accurate (96.7 %). Further, despite of A_3 providing only 71.6 % accuracy on the training data set, it achieves a 82.3 % accuracy over the test data set.

Table 5 The speed of the training in the third example

Training sets		Number of samples	Required time for training	Accuracy on the test data set (%)
Original		500	Can not learn	Can not learn
Clustered	C_3	134	53 min 11 s	96.6
	B_3	44	2 min 39 s	89.8
	A_3	16	27 s	71.6

Table 6 The accuracy of the training in the third example

Network		Accuracy of the training on the original training data set	Accuracy of the training on the test data set
Clustered	C_3	483/500 (96.6 %)	978/1000 (97.8 %)
	B_3	443/500 (89.8 %)	967/1000 (96.7 %)
	A_3	359/500 (71.6 %)	823/1000 (82.3 %)

4 Conclusions and Future Work

In this chapter, authors present a novel training method for regular feed-forward MLP classifier *ANNs* by which the convergence properties of the neural network models can be improved.

The algorithm applies a time-reduced pre-clustering procedure. This step reduces the quantity of the used training data (and by this the complexity of the training procedure) and parallel with it keeps the training ability of the models. Further, especially when facing "hard" classification problems, it results in an improvement on the convergence properties of the fitted model.

By tuning the clustering distance applied during the input data set reduction, an appropriate trade-off between the complexity of the training (iterative model fitting) and the accuracy of the model can also be set.

Our future work includes research aiming to give an algorithm for the determination of the "optimal" clustering distance for given circumstances. We also would like to generalize to presented technique towards other types of neural networks and problems.

Acknowledgment This work was sponsored by the Hungarian National Scientific Fund (OTKA 78576).

References

1. Chun, B.B., Jafri, M.Z.M., San, L.H.: Mangrove mapping in Penang Island by using Artificial Neural Network technique. In: Proceedings of the IEEE Conference on Open Systems (ICOS), pp. 245–249, 25–28 Sept 2011
2. Murray, J.C., Erwin, H.R.: A neural network classifier for notch filter classification of sound-source elevation in a mobile robot. In: Proceedings of the 2011 International Joint Conference on Neural Networks (IJCNN), pp. 763–769, 31 July–5 Aug 2011

3. Pradeep, J., Srinivasan, E., Himavathi, S.: Neural network based handwritten character recognition system without feature extraction. In: Proceedings of the International Conference on Computer, Communication and Electrical Technology (ICCCET), pp. 40–44, 18–19 March 2011
4. Aloise, D., Deshpande, A., Hansen, P., Popat, P.: NP-hardness of Euclidean sum-of-squares clustering. Mach. Learn. **75**, 245–249 (2009)
5. Várkonyi-Kóczy, A.R., Tusor, B.: Improving the supervised learning of neural network based classification. In: Proceedings of the IEEE International Symposium on Intelligent Signal Processing, WISP2011, Floriana, Malta, 19–21 Sept 2011
6. Ishibushi, H., Tanaka, H.: Fuzzy neural networks with fuzzy weights and fuzzy biases. In: Proceedings of the IEEE Neural Network Conference, San Francisco, USA, vol. 3, pp. 1650–1655 (1993)

Chaotic Systems Predictability Using Neuro-Fuzzy Systems and Neural Networks with Bred Vectors

**Pettras Leonardo Bueno dos Santos,
Haroldo Fraga de Campos Velho, Rosangela Cintra
and Sandra Sandri**

Abstract The predictability of the behavior of chaotic systems is of great importance because many real-world phenomena have some type of chaotic regime. In chaotic systems, small changes in the initial conditions can lead to very different results from the original system trajectory. The prediction of chaotic systems behavior is usually very difficult, particularly in practical applications in which initial conditions are obtained by measurement instruments, very often subject to acquisition errors. Here we use "bred vectors" methodology to generate pairs of input/output that are then used to train Neural Networks and Neuro-Fuzzy Systems. We apply the approach to predict regime change for Lorenz strange attractors and the nonlinear coupled three-waves problem from solar physics.

1 Introduction

The ability to predict the future state of a system, given its present state, is a non-trivial problem. It is also of capital importance; a correct prediction is the main factor in the prevention (or minimization of impact) of major natural disasters.

P. L. B. dos Santos (✉) · H. F. de Campos Velho · R. Cintra · S. Sandri
The Brazilian National Institute for Space Research (INPE), São José dos Campos, SP 12227-010, Brazil
e-mail: pettrasleonardo@yahoo.com.br

H. F. de Campos Velho
e-mail: haroldo@lac.inpe.br

R. Cintra
e-mail: rosangela.cintra@lac.inpe.br

S. Sandri
e-mail: sandra.sandri@inpe.br

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_19,

Due to their own nature, prediction in the context of chaotic systems is particularly challenging, because small changes in the initial conditions can lead to very different results from the original trajectory of a chaotic system. The problem is aggravated in practical applications, in areas such as geophysical and astronomical sciences, in which initial conditions are obtained by measuring instruments, very often subject to errors in precision. Earth climate, for instance, is a prototypical chaotic system [1] of fundamental importance. Its behavior is hard to predict and is nowadays obtained from the aggregation of an ensemble of results, derived from running a computational model with a set of different initial conditions.

In recent years, the breeding vector technique has been used with success to predict behavioral aspects of several chaotic systems. This technique was developed as a method to generate initial perturbations for ensemble forecasting in numerical weather prediction at the American National Center for Environmental Prediction (NCEP) [2]. Nowadays, it is a well-established and computationally inexpensive method for generating perturbations to be used on ensemble integrations [3].

The breeding method involves simply running the nonlinear model twice at each time step, one with the original data (control run) and the other with a small perturbation added to it. After a fixed number of time steps, the obtained results are subtracted and the difference is rescaled so that it has the same size as the original perturbation. The rescaled difference, called a bred vector, is added to the control run and the process is repeated.

Bred vectors have been used to predict the behavior of several chaotic systems [3–5]. In particular, the authors of [4] (respec. [3]) derived rules for predicting behavior of the Lorenz attractor (respec. three-waves problem), upon observation of the bred-vector growth. The accuracy of the prediction using these rules have been extremely good in both experiments, showing that not only bred vectors provide a good tool for the prediction of behavior of chaotic systems, but also that the knowledge produced in the process is interpretable. However, obtaining rules directly from observation can be painstaking.

The main contribution of the present work is an investigation on the use of neuro-fuzzy system ANFIS [6] to evaluate the predictability of chaotic systems. This task is formulated as a classification problem, where classes of dynamics are identified. The results furnished by ANFIS are compared to those obtained through the use of a standard neural network [7]. We have performed experiments on Lorenz strange attractor [1] and the nonlinear coupled three-waves model [3, 8]. Neural networks have already been used with bred vectors to predict the behavior of Lorenz attractor [5] but here we explore different types of experiments.

This chapter is organized as follows. In Sect. 2 we briefly present chaotic systems and bred vectors, as well as the systems used in our applications: the Lorenz attractor and the three-waves problem. In Sect. 3 we briefly address neural networks, fuzzy systems and neuro-fuzzy systems. Sections 4 and 5 respectively bring the experiments and the conclusion.

2 Chaotic Systems and Bred Vectors

Chaotic systems are extremely sensitive to initial conditions: a slight deviation from a trajectory in the state space can lead to dramatic changes in future behavior [9]. The prediction of the future state of a system knowing its initial conditions is a fundamental problem with obvious applications in geophysical flows. The predictability of weather and climate forecasts, for instance, is determined by the projection of uncertainties in both initial conditions and model formulation onto flow-dependent instabilities of the chaotic climate attractor. Since it is essential to be able to estimate the impact of such uncertainties on forecast accuracy, no weather or climate prediction can be considered complete without a forecast of the associated flow-dependent predictability [3].

The breeding method [2] was developed as a technique to generate initial perturbations for ensemble forecasting in numerical weather prediction. The method consists in running the nonlinear model twice, one with the original data (control) and the other with a small perturbation added to it. The control solution is subtracted from the perturbed solution after a fixed amount of time steps and the rescaled difference (a bred vector) is then added to the control run and the process is repeated. The difference between the control and perturbed solutions are rescaled so the difference has the same size as the original perturbation. The growth rate of the bred vectors is a measure of the local instability of flow [3]. Figure 1 illustrates bred vectors growth.

Bred vectors are a nonlinear generalization of leading Lyapunov exponents, that rates the differences of two initially close trajectories of chaotic dynamical systems. Studies on the stability properties of evolving flows can be found in [10] for Lyapunov vectors and in [11] for bred vectors.

The bred-vectors algorithm is given as follows [12].

(1) Calculate the initial perturbation $A = \delta f(x, t)$, using an arbitrary norm.
(2) Add the perturbation calculated in the previous step to the basic solution, integrate the perturbed condition with the nonlinear model for a fixed number of time steps, and subtract the original unperturbed solution from the perturbed nonlinear integration

$$\delta A = \delta f(x, t + \Delta t) = f(x, t + \Delta t) - f(x, t). \tag{1}$$

(3) Measure the size $A + \delta A$ of the evolved perturbation $\overline{\delta f(x, t + \Delta t)}$, and divide the perturbation by the measured amplification factor so that its size remains equal to A:

$$\delta f(x, t + \Delta t) = \overline{\delta f(x, t + \Delta t)} \sim A/(A + \delta A) \tag{2}$$

The initialization step (1) is executed only once. Steps 2 and 3 are repeated for each time interval (identified as a fixed number of time steps).

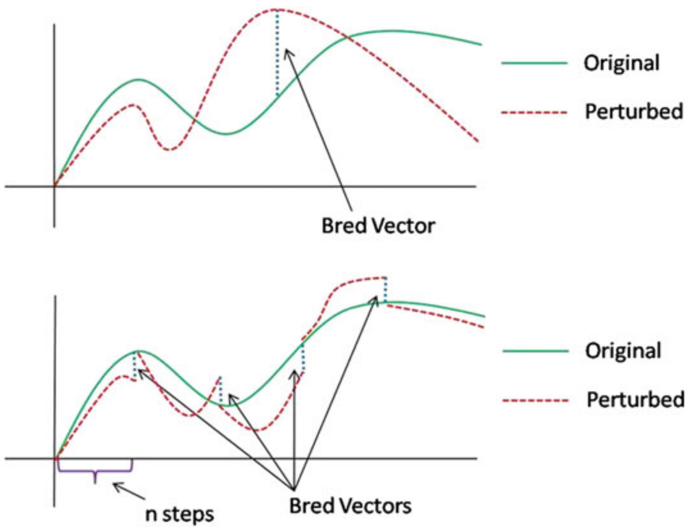


Fig. 1 Bred vectors growth illustration

Bred vectors have been used with success to predict the behavior of chaotic systems such as the Lorenz strange attractor [4] and the three-waves systems [3], described below.

2.1 Lorenz Attractor

A strange attractor is a particular kind of chaotic dynamic system, in which the trajectory is confined in a finite region, and where orbits are never repeated. The Lorenz strange attractor has been very widely used as a prototype of chaotic behavior [1]. It has two regimes, which could represent two possibilities for any given season of the year (e.g. "warm winter" and "cold winter"). Although apparently simple, in such a system it is hard to identify when a regime change will happen and how long it will last. The Lorenz Model equations are given as:

$$dx/dt = -\sigma x - y \tag{3}$$

$$dy/dt = -\rho x - y - xz \tag{4}$$

$$dz/dt = xy - \beta z \tag{5}$$

with $\sigma = 10, \rho = 28, \beta = 8/3$ as parameters, the resulting system is a strange attractor [1] (see Fig. 2).

In [3], the breeding method was applied on the Lorenz model, integrated with time steps $\Delta t = 0.01$, and a second run started from an initial perturbation $\delta x_0 = (\delta x_0, \delta y_0, \delta z_0)$ added to the control at time t_0. The difference δx between the

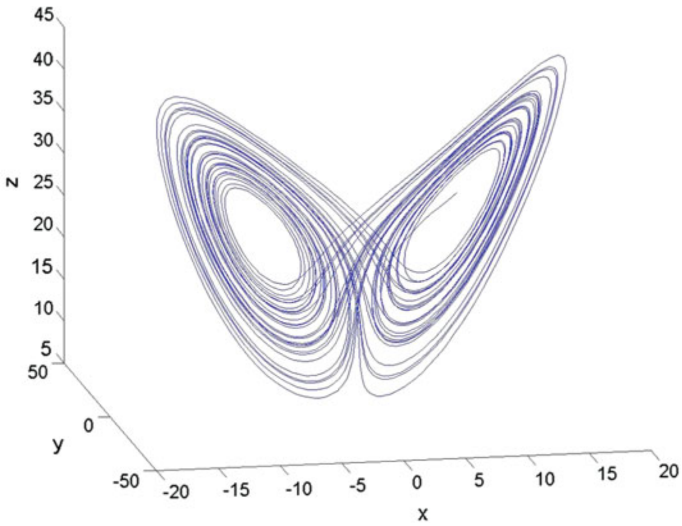


Fig. 2 Solutions of the Lorenz model equations showing two chaotic regimes

perturbed and the control run was taken at every 8 times steps. The growth rate of the perturbation was measured per time step as [4]:

$$g = \frac{1}{n} * (|\delta x| / |\delta x_0|).$$

Figure 3 illustrates the attractor built under these conditions, using 4 colors to indicate bred vector growth intervals. The presence of a red star indicates that the bred vector growth in the previous 8 steps was greater than 0.064, the blue stars indicate a negative growth rate, green stars denotes the interval greater than 0 and less than 0.032, and yellow stars mean the interval greater or equal 0.032 and less or equal than 0.064. Figure 4 depicts the same system, considering only one axis: the abscissa (time) separates the two regimes and each inflection point indicates the beginning of a new orbit of the system (similar graphics can be drawn for the other axes).

From the observation of the system depicted in Figs. 3 and 4, the following rules have been proposed in [4]:

Rule Ev.1: When the growth rate exceeds 0.064 over a period of 8 steps, as indicated by the presence of one (or more) red stars, the current regime will end after it completes the current orbit.
Rule Ev.2: The length of the new regime is proportional to the number of red stars. For example, the presence of 5 or more stars in the old regime, indicating sustained strong growth, implies that the new regime will last 4 orbits or more.

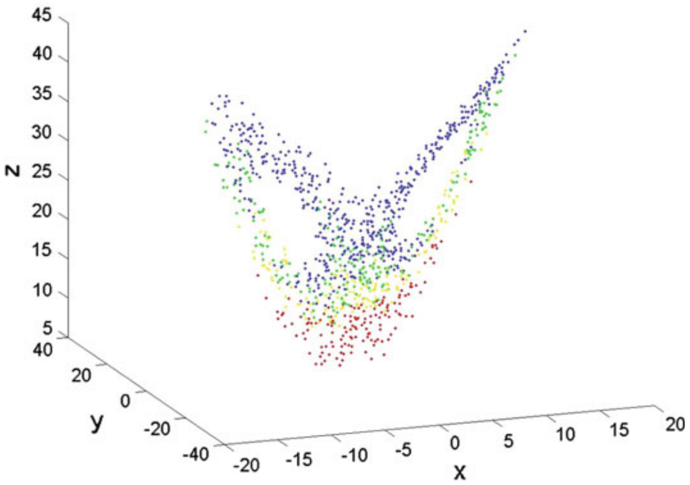


Fig. 3 Lorenz attractor colored with the bred vector classes

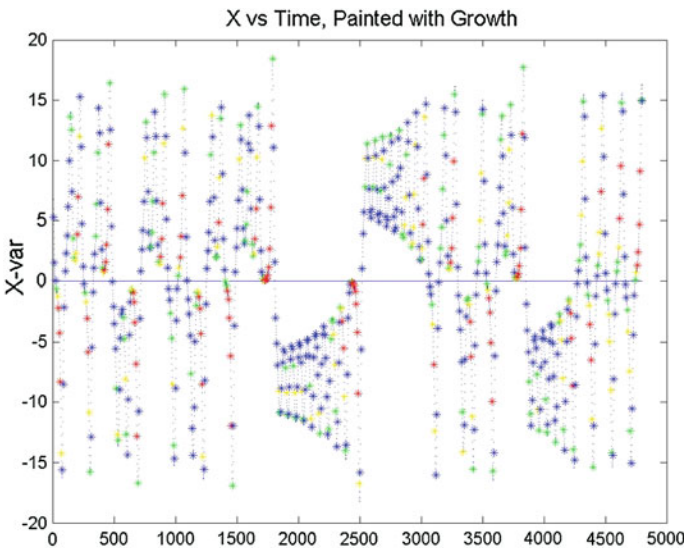


Fig. 4 $X(t)$ for Lorenz attractor colored with the bred vector classes

The Lorenz attractor was then executed for 40,000 time steps (with 187 changes of regime) using these rules, obtaining 91.4 % prediction accuracy (see [4] for more details).

2.2 Nonlinear Coupled Three-Waves Model

Nonlinear three-waves coupling is of general interest in many branches of physics [3]. It is for instance responsible for the generation and modulation of plasma waves in the planetary magnetosphere and solar winds [8].

In the simplest model for describing the temporal dynamics of resonant nonlinear coupling of three waves, one assumes the waves to be monochromatic, with the electric fields written in the form: $E_\alpha(x,t) = \frac{1}{2}A_\alpha(x,t)\exp\{i(k_\alpha x - \omega t)\}$, where $\alpha = 1, 2, 3$, and the time scale of the nonlinear interactions is much longer than the periods of the linear (uncoupled) waves.

In order for three-waves interactions to occur, the wave frequencies ω_α and wave vectors k_α must satisfy the resonant conditions

$$\omega_3 \cong \omega_1 - \omega_2, \quad k_3 = k_1 - k_2 \tag{6}$$

Under these circumstances, the nonlinear temporal dynamics of the system can be governed by the following set of three first-order autonomous differential equations written in terms of the complex slowly varying wave amplitude [13]:

$$dA_1/d\tau = v_1 A_1 + A_2 A_3 \tag{7}$$

$$dA_2/d\tau = i\delta A_2 + v_2 A_2\, A_1 A_3^* \tag{8}$$

$$dA_3/d\tau = v_3 A_3\, A_1 A_2^* \tag{9}$$

where the variable $\tau = \chi t$ (t is time, and χ is a characteristic frequency), $\delta = (\omega_1 - \omega_2 - \omega_3)/\chi$ is the normalized linear frequency mismatch, and $v_\alpha = \bar{v}_\alpha/\chi$ ($\alpha = 1, 2, 3$) gives the linear wave behaviors on the long time scale. In order to have interactions among the three-waves, the wave frequencies ω_α must satisfy the resonant condition: $\omega_3 \sim \omega_1 - \omega_2$.

In the experiments, following [3], wave A_1 is assumed to be linearly unstable ($v_1 > 0$) and the other two waves, A_2 and A_3, are linearly damped ($v_2 = v_3 \equiv -v < 0$) and thus $\chi = v_1$ [13, 14]. The system admits both periodic and chaotic waves. For the chaotic dynamics, a strange attractor is found (see Fig. 5). As in the Lorenz system dynamics, the coupled three-waves system has two *seasons* in the strange attractor. However, contrary to what happens with Lorenz attractors, the seasons (or regimes) have no symmetry. One regime is identified as a line formed by a curve on XY plane followed by another curve on the YZ plane (Fig. 5). The other regime is characterized by the straight line in the intersection between the XY and YZ planes.

In [3], the breeding method was applied on the three-waves model with $\Delta t = 0.001$ and initial perturbation δx_0 added to the control at time t_0. The bred vector was calculated at every 8 time steps (using the same bred vector classification as in the Lorenz model). Figure 5 illustrates the attractor built under these conditions, and Fig. 6 depicts the system, considering a single axis (similar graphics can be

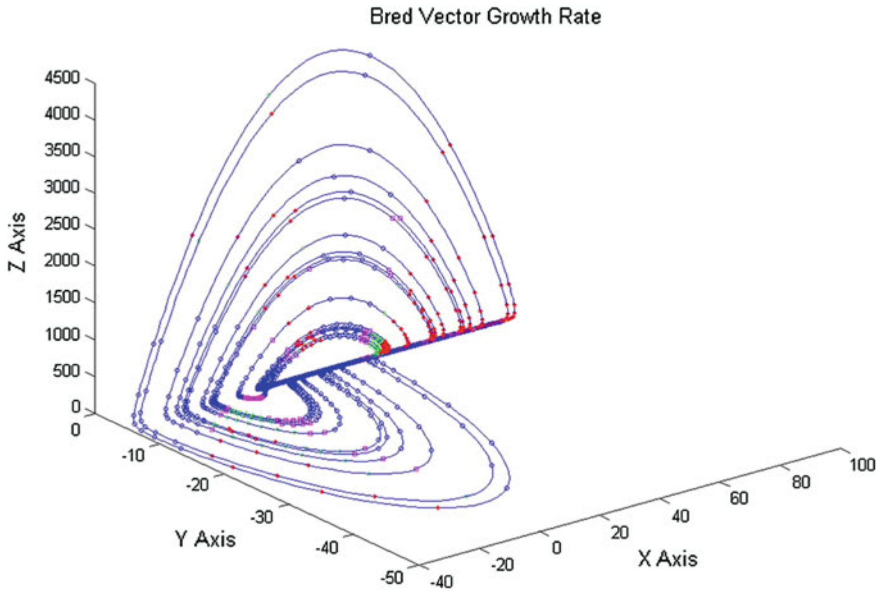


Fig. 5 The three-waves model attractor colored with the bred vector classes

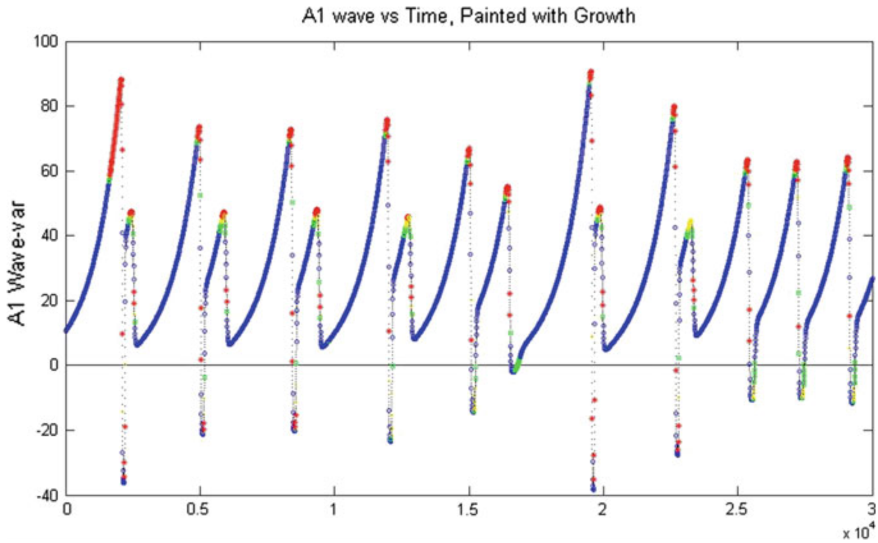


Fig. 6 $A1(t)$ for the three-waves model model

drawn for the other axes). We can see that here the regimes alternate steadily, contrary to what happens with the Lorenz attractor. The question here is: at which time step the "straight" regime will change to the "curve" regime?

From the observation of the system depicted in Figs. 5 and 6, the following rules have been proposed for each wave:

Rule Ci.1 The presence of a red star indicates that the bred vector growth in the previous 8 steps was greater than 0.35. A red star indicates that the regime is going to change: if the trajectory is on the straight (or curve) regime, the dynamics is going to continue on the curve (or straight) regime.
Rule Ci.2 Furthermore, the presence of 2 or 3 red stars for regime-curve indicates that the next regime will be on the straight line, with a long duration trajectory on this regime. More than 3 red stars into regime-curve implies a changing to the regime-stright with long duration (following on YZ and XY planes).

The three-waves system was executed for 20,000 time steps (with 23 changes of regime) using these rules, with several thresholds, obtaining 100 % prediction accuracy with the best threshold for Rule 1 and 95.6 % for Rule 2 (see [3] for more details).

3 Neural Networks, Fuzzy Systems and Neuro-Fuzzy Systems

Artificial Neural Networks (ANNs) [7] aim at emulating the learning behavior of the human brain. The network is composed of nodes (with some calculation abilities) and weighted edges between nodes. Usually, for a supervised ANN, a backpropagation method is used to adjust the weights according to the difference between the desired and the calculated output. In the experiments reported here, a simple neural network with a single hidden layer have been used, taken from the WEKA platform [15].

Fuzzy Systems [16] aim at emulating some of the human capacity of reasoning with vague information. Membership to a fuzzy set is measured by a number between 0 and 1 in the real scale, instead of simply 0 or 1 as in its classical counterpart. Most of the systems created using fuzzy sets theory are based on rules of thumb of the type “If condition then conclusion”, where the variables in both the condition and conclusion parts are associated to fuzzy sets. The two most well-known kinds of fuzzy systems are the Sugeno and Mamdani models. Both frameworks use fuzzy terms to model the conditions on the left-hand side of the rules, but differ on the modeling of the conclusions on the right-hand side of the rules. In Sugeno fuzzy systems, a conclusion of a rule is modeled as a (usually linear) function of the values of the input variables. In Mamdami systems, fuzzy sets of the output variables are used in the conclusions of rules, a characteristic that make these systems interpretable, contrary to what happens with Sugeno systems.

The term *neuro-fuzzy* emerged in the mid-1980s and corresponds a mixture of artificial neural networks with fuzzy rule-based systems. A neuro-fuzzy system [17] is used to derive a fuzzy rule based system, whose defining parameters (fuzzy terms and rules) are learnt through training performed in a neural network-like learning process upon the presentation of a set of pairs (input, desired output).

Artificial neural networks are very appropriate to deal with problems related to pattern recognition, but ill-suited to explain how the answers are obtained, functioning as a sort of "black box". On the other hand, fuzzy rule based systems, even working with inaccurate information, are appropriate to explain how results are obtained, but needs an expert to create the inference rules. Neuro-fuzzy systems, using neural networks learning apparatus to generate the inference rules of a fuzzy system, aim at maximizing the advantages of each model while minimizing their disadvantages.

Neuro-fuzzy systems are classified by the type of fuzzy system they produce. The two most noteworthy neuro-fuzzy system types are those derived from Sugeno and Mamdani fuzzy systems. Neuro-fuzzy systems can be thought of as a series of successive processing layers, but, contrary to what usually happens with neural networks, each layer is composed by its own specific kind of nodes and represents a particular stage in the processing of a fuzzy system. The initial layers of Sugeno and Mamdani neuro-fuzzy systems coincide, but they diverge greatly in the last ones. In the following we present the coinciding layers.

(1) In the first layer, the current value of each input fuzzy variable is compared with the fuzzy terms associated with that variable, resulting in a compatibility degree for each term.
(2) In the second layer, the compatibility degrees from the different input variables are combined, resulting in the overall compatibility degree of the potential rules with the data.

In the last layers, a fuzzy set (respect. a set of constants) is learned for each rule in a Mamdani (respec. Sugeno) systems.[1]

In the present work, the experiments have been made with ANFIS (Adaptive-Network-Based Fuzzy Inference System) [6]. It uses the backpropagation algorithm to calculate the parameters of the fuzzy terms on the rule left-hand side the LMS algorithm (Least Square Means) for calculating the parameters on the right-hand side of the rule in a Sugeno fuzzy system.

4 Experiments

Our experiments view the prediction problem as a classification task. Here we have addressed the prediction of two chaotic systems: the Lorenz attractor and the three-waves problem. We have made 3 types of experiments for the Lorenz attractor and a single one for the three-waves problem. On the experiment on the three-waves problem and for the Experiment 1 on Lorenz attractor problem, two

[1] As a matter of fact, in Mamdani neuro-fuzzy systems, the set of possible output fuzzy terms are learned a priori and the association between rules right-hand sides and fuzzy herms are learned in a competitive manner.

kinds of training were made, one with only two classes (dichotomic) and another with a larger number of classes (multi-classes). The testing was performed for these two kinds of training plus an extra one, in which we condensed the results of the multi-classes experiment to the two classes used in the dichotomic experiment. In all experiments, the size of the bred vectors obtained at each time step t, denoted as $s(t)$ were classified by colors, as in [4]: red $(s(t) > 0.064)$, yellow $(0.032 \leq s(t) < 0.0064)$, green $(0 \leq s(t) < 0.032)$ and blue $(s(t) < 0)$.

For both problems addressed here, the same ANN and ANFIS configuration were used. The neural network (ANN) and the ANFIS experiments were produced using platforms WEKA and MATLAB, respectively. Training with ANFIS was performed considering 3 triangular fuzzy terms for each input variable and constant output, with 300 epochs. We have chosen the Multilayer Perceptron with 3 layers as the ANN architecture. The ANNs used the following configuration: learning rate = 0.3, momentum = 0.2, epochs = 500, 4 neurons in the input layer, n neurons on the output layer and $(4 + n)/2$ neurons in the hidden layer, where n is the number of classes.

4.1 Three-Waves System

Here we describe our best results for the three-waves system and the configuration of the predictors (ANN and ANFIS) used to obtain those results. In the three-waves experiment we have used 170 samples, divided into training (100), validation (30) and test (40). The input is the number of bred vectors in each class (color) found in the preceding straight line regime.

In the two-classes experiment, the classes are described as follows:

- A: the straight line trajectory will last up to 1,200 time steps;
- B: the straight line trajectory will last more than 1,200 time steps.

In the multi-classes experiment, we have used 10 classes, numbered from 0 to 9. Each class is associated to the interval bounding the number of time steps in which the straight line trajectory is predicted to last:

- 0: [0, 1200]
- 1: [1201, 1400]
- 2: [1401, 1600]
- 3: [1601, 1800]
- 4: [1801, 2000]
- 5: [2001, 2200]
- 6: [2201, 2400]
- 7: [2401, 2600]
- 8: [2601, 2800]
- 9: >2800

Table 1 Confusion matrices for the three-waves problem experiments with multi-classes

ANFIS	0	1	2	3	4	5	6	7	8	9
0	14	2	1	0	0	0	0	0	0	0
1	2	1	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	2	2	0	0	0	0
5	0	0	0	0	0	3	1	0	0	0
6	0	0	1	0	0	0	0	0	0	0
7	0	0	0	0	0	1	2	3	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	1	1	1	0
ANN	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	16	0	0	0	0	0	0	0	1	0
1	1	1	1	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	0	4	0	0	0	0	0
5	0	0	0	0	3	0	0	1	0	0
6	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	1	1	0	2	2	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	2	0

The confusion matrices obtained by the use of ANFIS and the ANN on the test set for the multi-class experiments are given in Table 1. The lines correspond to the real class and the column to system prediction. Table 2 brings the confusion matrices obtained by the use of ANFIS and the ANN on the two-classes experiments. In the confusion matrices in Table 3, class *A* remains unchanged but class *B* has been obtained by merging classes 1 to 9, from the multi-classes experiment.

Table 4 brings the accuracy results for the three-waves experiments. We can see that ANN and ANFIS have good overall performances with a significant advantage for ANN. Moreover, they both perform poorly with a large number of classes but much better in the dichotomic cases. Finally, the performance of ANN when the training was done with a large number of classes was improved when the set of classes were merged into only two classes.

4.2 Lorenz Attractor

Here we describe our best results for the Lorenz attractor and the configuration of the predictors (ANN and ANFIS) used to obtain those results. In the Lorenz attractor experiment we have used 1275 samples, divided into training (800),

Table 2 Confusion matrices for the three-waves problem experiments with two-classes

ANFIS	*A*	*B*
A	15	2
B	3	20
ANN	*A*	*B*
A	15	2
B	1	22

Table 3 Confusion matrices for the three-waves problem experiments with two classes merged from the multi-classes experiments

ANFIS	*A*	*B*
A	14	3
B	2	21
ANN	*A*	*B*
A	16	1
B	1	22

validation (200) and test (275). The input is the number of bred vectors in each class (color) found in each orbit.

We have performed 3 types of experiments using the Lorenz attractor. In Experiment T1, we are only interested in the prediction of when the regime will change. In Experiment T2 we are interested in knowing whether the regime will go unchanged in the next orbit and, if it is supposed to change, in which orbit the change will take place. Experiment T3 is a mix of Experiments 1 and 2. Experiment 1 is similar to what can be found in the literature (see [4, 5]). Experiments 2 and 3 intend to verify whether extra information may be useful in the prediction process. This is relevant in certain applications, in which this kind of prediction is an indicator of intensity of a phenomenon, as in climate prediction for instance.

4.2.1 Experiment T1

In the two-classes experiment, the classes are described as follows:

- *A*: the current regime will not change up to 5 orbits;
- *B*: the current regime will change after 5 orbits.

In the multi-classes experiment, we have 10 classes, numbered from 0 to 9. Each class is associated to the number of orbits that are predicted to occur before regime change; for instance, class 9 means that the number of orbits before regime change is larger than 8.

The confusion matrices obtained by the use of ANFIS and the ANN on the test set for the multi-class experiments are given in Table 5. The lines correspond to the real class and the columns to system prediction. Table 6 brings the confusion matrices obtained by the use of ANFIS and the ANN on the two-classes

Table 4 Accuracies for the the three-waves problem experiments

Accuracy	ANFIS (%)	ANN (%)
10-classes	62.5	60
2-classes (original)	87.5	92.5
2-classes (merged)	87.5	95

Table 5 Confusion matrices for the Lorenz attractor: experiments T1 with multi-classes

ANFIS	0	1	2	3	4	5	6	7	8	9
0	104	14	4	0	0	0	0	0	0	0
1	9	46	7	2	0	0	0	0	0	0
2	0	13	17	2	0	3	0	0	0	0
3	0	2	7	5	7	3	0	0	0	0
4	0	0	0	3	4	3	2	0	0	0
5	0	0	1	1	1	3	0	0	0	0
6	0	0	0	0	3	1	0	0	0	0
7	0	0	0	0	1	1	0	0	0	0
8	0	0	0	0	0	2	0	0	0	0
9	0	1	0	0	1	1	1	0	0	0
ANN	0	1	2	3	4	5	6	7	8	9
0	115	4	1	0	2	0	0	0	0	0
1	5	47	11	1	0	0	0	0	0	0
2	0	9	21	2	3	0	0	0	0	0
3	0	1	8	10	5	0	0	0	0	0
4	0	0	1	4	6	1	0	0	0	0
5	0	0	1	2	3	0	0	0	0	0
6	0	0	0	1	3	0	0	0	0	0
7	0	0	0	1	1	0	0	0	0	0
8	0	0	0	0	2	0	0	0	0	0
9	1	0	0	0	2	0	0	0	0	1

experiments. Table 7 brings the confusion matrices from ANFIS and ANN, with class *A* corresponding to class 0 (i.e. the regime will change in the next orbit), and class *B* obtained by merging classes 1 to 9, from the multi-classes experiment. Classes *A* and *B* thus mean that "the regime will change in the next orbit", and "the trajectory will stay in the current regime for more than one orbit".

Table 8 brings the accuracy results for the Lorenz attractor T1 experiments. We can see that ANN and ANFIS have approximate overall performances with a slight advantage for ANN. Moreover, they both perform reasonably well with a large number of classes but much better results have been obtained in the dichotomic cases. Finally, ANN obtained improved results when the training was done with a large number of classes, which were then merged into only two, contrary to ANFIS, which had its performance decreased.

Table 6 Confusion matrices for the Lorenz attractor: experiments T1 with two-classes

ANFIS	*A*	*B*
A	111	11
B	9	144
ANN	*A*	*B*
A	106	16
B	5	148

Table 7 Confusion matrices for the Lorenz attractor: experiments T1 with two classes merged from the multi-classes experiments

ANFIS	*A*	*B*
A	104	18
B	9	144
ANN	*A*	*B*
A	115	7
B	6	147

Table 8 Accuracies for the Lorenz attractor: experiments T1

Accuracy	ANFIS (%)	ANN (%)
10-classes	65.09	72.72
2-classes (original)	92.72	92.36
2-classes (merged)	90.18	95.27

4.2.2 Experiments T2 and T3

Experiments T2 and T3 are fragments of a series of experiments reported in [18].

Experiment T2 is multi-classes with 13 classes numbered from 0 to 12, described as:

- 0: The trajectory will stay in the current regime.
- *k*: The regime will change in the next orbit and will then stay in the new regime for *k* orbits.

The confusion matrices obtained by the use of ANFIS and ANN for Experiment T2 are given in Table 9.

Experiment T3 is multi-classes with 6 classes numbered from 1 to 6 described as:

- 1: The trajectory will stay in the current regime for more than 2 orbits.
- 2: The trajectory will stay in the current regime for 2 more orbits.
- 3: The trajectory will stay in the current regime for 1 more orbit.
- 4: The regime will change in the next orbit and will then stay in the new regime for 1 more orbit.
- 5: The regime will change in the next orbit and will then stay in the new regime for 2 more orbits.

Table 9 Confusion matrices for the Lorenz attractor: experiments T2

ANFIS	0	1	2	3	4	5	6	7	8	9	10	11	12
0	140	7	2	3	0	0	0	0	1	0	0	0	1
1	5	41	10	0	1	0	0	0	0	1	0	0	0
2	0	0	27	2	0	0	0	0	0	0	0	0	0
3	0	0	2	6	3	0	0	0	0	0	0	0	0
4	0	0	1	2	5	1	0	0	0	0	1	0	1
5	0	0	0	0	3	3	0	0	0	0	0	0	0
6	0	0	0	0	0	2	0	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	1
ANN	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	148	5	0	0	1	0	0	0	0	0	0	0	0
1	14	36	8	0	0	0	0	0	0	0	0	0	0
2	0	0	27	1	1	0	0	0	0	0	0	0	0
3	0	0	2	7	2	0	0	0	0	0	0	0	0
4	0	0	0	7	2	1	0	1	0	0	0	0	0
5	0	0	0	4	1	1	0	0	0	0	0	0	0
6	0	0	0	0	0	2	0	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	0	0	0	0

- 6: The regime will change in the next orbit and will then stay in the new regime for more than 2 orbits.

The confusion matrices obtained by the use of ANFIS and ANN for Experiment T3 is given in Table 10. The accuracy in Experiments T2 and T3 are given in Table 11.

We see that for Experiments T2 and T3, both learning systems produced very good results, with a slight advantage of ANFIS over ANN. Considering that these experiments address long-term events, the ordering of their performance is as expected: for both ANFIS and ANN, the accuracy in Experiment T1 is higher than in both T2 and T3.

Table 10 Confusion matrices for the Lorenz attractor: experiments T3

ANFIS	1	2	3	4	5	6
1	44	8	1	1	0	0
2	8	13	12	2	0	0
3	4	7	48	3	0	2
4	1	1	4	41	9	2
5	0	0	0	0	27	2
6	0	0	0	0	3	32
ANN	1	2	3	4	5	6
1	43	9	2	0	0	0
2	5	17	13	0	0	0
3	1	11	52	0	0	0
4	2	6	15	26	9	0
5	0	0	0	3	24	2
6	0	0	0	0	2	33

Table 11 Accuracies for the Lorenz attractor: experiments T2 and T3

Accuracy	ANFIS (%)	ANN (%)
Experiment T2	81.45	80.36
Experiment T3	74.54	70.90

5 Conclusions and Future Work

Predictability using bred vector methodology was first used to analyze the the well-known chaotic Lorenz system in [12]. The same system was evaluated with breeding and neural network, where the dynamics is identified into different classes [5]. Bred vector methodology was also used for the coupled three-waves mode from solar physics, showing similar efficiency as an analysis dynamics tool, as already verified by Cintra and Campos Velho [3].

The ability of ANFIS neuro-fuzzy system was investigated to evaluate the predictability of chaotic systems using the bred-vector technique, for both the Lorenz attractor and the coupled three-waves model. This task is formulated as a classification problem, where classes of dynamics are identified. The results obtained with ANFIS neuro-fuzzy approach were compared with an Artificial Neural Network (multilayer perceptron) codified in the Weka package.

For the Experiment T1 associated with the three-waves problem and from the Lorenz attractor, the learning systems were trained using: (i) two classes, and (ii) a larger number of classes (10). A post-classification strategy was also employed to obtain two classes from the multi-classes experiments. Very good results were obtained, specially in the dichotomic case, both when training was performed directly with two classes, as well as when the two classes were obtained from merging classes in the multi-classes experiments. In Experiments T2 and T3 for the Lorenz attractor, a possibility was explored of using the data for making predictions in problems in which other factors than the main prediction, such as intensity, play a role.

In what regards the performance of ANFIS compared to ANN, we see that the latter produced (in general) better results. However, in the Lorenz attractor experiments, ANFIS outperformed ANN for long term-events.

In our experiments, training with a large number of classes that were then merged produced better results when ANN was used, contrary to what happened when ANFIS was used instead. Therefore, in what regards post-processing, experimenting is needed to tell whether it is better to train the systems with the original intended number of classes or to train them with a larger number of classes and then merge some of them to obtain the final classification (special care would have to be taken for the class that characterizes immediate regime change).

The results obtained so far show that neuro-fuzzy systems of the Sugeno type (ANFIS) are useful for the prediction of chaotic systems. The use of other systems (neuro-fuzzy and otherwise) is under investigation, with the potential for producing fuzzy systems by learning, in particular those based in the Mamdani paradigm [16]. The ultimate goal is to use the derived fuzzy systems as a basis for the automatic production of interpretable rules, such as those created by observation for the Lorenz attractor [4] and the three-waves system [3].

We also intend to investigate the use of asymmetrical means of quantifying the accuracy of predictions. In some contexts, it is very harmful that an event occurs before expected. On others, the largest damage is produced when an event occurs ahead of what is expected. For instance, a crop may be lost when the rain comes either too early or too late, but the costs are different for distinct types of stakeholders.

In the future we also intend to quantify the confidence on the predictions and to explore the tradeoff between imprecise but trustful predictions versus precise but less trustful ones.

Acknowledgment The authors thank FAPESP and CNPq for research support.

References

1. Lorenz, E.N.: Deterministic non-periodic flow. J. Atmos. Sci. **20**(130–141), 1963 (1963)
2. Toth, Z., Kalnay, E.: Ensemble forecasting at NCEP and the breeding method. Mon. Weather Rev. **126**, 3292–3302 (1997)
3. Cintra, R.S., Campos Velho, H.F.: Predictability for a chaotic solar plasma system. In: Symposia: Advanced Modeling on Computational Fluid Dynamics (2008)
4. Evans, E., Bathi, N.K., Kinney, J., Pann, L., Peña, M., Yang, S.-C., Kalnay, E., Hansen, J.: Rise undergraduates find that regime changes in Lorenz's model are predictable. Bull. Amer. Meteor. Soc. **85**, 521–524 (2004)
5. Pasini, A., Pelino, V.: Can we estimate atmospheric predictability by performance of neural network forecasting? The toy case studies of unforced and forced lorenz models. In: Proceedings of CIMSA'2005 (2005)

6. Shing, J., Jang, R.: ANFIS: adaptive-network-based fuzzy inference systems. IEEE Trans. Syst. Man Cybern. **23** 714–723 (1993)
7. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall, New Jersey (1999)
8. Chian, A.C.-L., Lopes, S.R., Alves, M.V.: Nonlinear excitation of Langmuir and Alfvén waves by auroral whistler waves in the planetary magnetosphere. Astron. Astrophys. **288**, 981–984 (1994)
9. Guégan, D., Leroux, J.: Predicting chaos with Lyapunov exponents: zero plays no role in forecasting chaotic systems. In: Tlelo-Cuautle, E. (Ed.) Chaotic Systems. InTech. http://www.intechopen.com/articles/show/title/predicting-chaos-with-lyapunov-exponents-zero-plays-no-role-in-forecasting-chaotic-systems Accesseed 04 mar 2011
10. Alligood, K.T., Sauer, T.D., Yorke, J.A.: Chaos: An Introduction to Dynamical Systems. Springer, New York (1996)
11. Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge (2005)
12. Kalnay, E., Peña, M., Yang, S.-C., Cai, M.:Breeding and predictability in coupled Lorenz models. In: Proceedings of the ECMWF (2002)
13. Meunier, C., Bussac, M.N., Laval, G.: Intermittency at the onset of stochasticity in nonlinear resonant coupling processes. Physica **D4** 236–243 (1982)
14. Lopes, S.R., Chian, A.-L.: A coherent nonlinear theory of auroral Langmuir-Alfvén-whistler (LAW) events in the planetary magnetosphere. s **365**, 669–676 (1996)
15. WEKA Đ Waikato Environment for Knowledge Analysis. http://www.cs.waikato.ac.nz/ml/weka/ Acessed 25 jul 2012
16. Gomide, F., Pedrycz, W.: An Introduction to Fuzzy Sets. MIT Press, Cambridge (1998)
17. Lin, C.-T., Lee, C.S.G.: Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems. Prentice Hall, Upper Saddle River (1996)
18. dos Santos, P.L.B., Sandri, S., de Campos Velho, H.F.: Previsibilidade em sistemas ca—óticos utilizando o sistema Neuro-Difuso ANFIS (in Portuguese). In: II Congresso Brasileiro de Sistemas Fuzzy (CBSF), pp. 113–126. Natal, Brazil (2012)

Part VI
Applications

Creation the Model of Educational and Methodical Support Based on Fuzzy Logic

Shahnaz N. Shahbazova

Abstract This article is devoted to the design and construction of automated an educational complex architecture, researched and specified basic, necessary and sufficient, models and methods capable of qualitatively improve the efficiency of the educational process. The original system of representation of educational materials will upgrade the quality of training programs, plans, and appropriate teaching and learning materials to meet the light of the above requirements. It imposes more stringent requirements for their content in terms of the feasibility of monitoring tasks of understanding and mastering. On the basis of the results of research and development can be efficiently and effectively reorganized the existing and, if necessary, set up corresponding new intelligent systems and information infrastructure of individual universities and the entire education system as a whole in compliance with the principles of systematization, templates, and modularity.

1 Introduction

The level of development of new technologies and methods of automation control enables us to reach a level of intellectualization of complex processes, that also include the learning process and, when highly intellectual task of receive education and knowledge can be qualitatively resolved with minimal participation of the teacher. Automated system, in this case, will take on not only information, but also pedagogical function [1].

The successful development of information technology provides the ability to conduct research and experiments in the field of educational systems for the

S. N. Shahbazova (✉)
Department of Information Technology and Programming, Azerbaijan Technical University,
25 H. Cavid Ave, Baku AZ1073, Azerbaijan
e-mail: shahbazova@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_20,

implementation of purposeful educational programs on the basis of information systems endowed with the functions of the expert system. The effectiveness of the implementation of which depends largely on the level of development and readiness analysis and information models.

It should be clarified that no intelligent learning system compares on quality of teaching learning material and evaluation of knowledge, with a teacher (I would use the word instructor) very seriously suitable to his duties. In its turn intelligent information system is able to offer independent unified standards of learning and knowledge control, providing pre-defined profile organizations requirements on the level of factual knowledge and skills of students in the specialty.

This chapter analyzes the architecture and designed intelligent information systems learning and testing. It identifies the key components and modules; functional design is carried out, the methods of interaction for the implementation of the task.

2 Structuring of Educational Material

The initial stage of designing complex systems planned for the study of an educational course should be split into separate educational materials (EM). Under the EM, we understand the objects and phenomena, concepts, methods of operation, selected from the appropriate science and included in the program educational discipline or topic of discipline to study them.

The aggregate of EM will represent in block diagram form—tree graph, which will be functional directed graph of the course of content, and construct by a hierarchical principle. Nodes (vertices) of the graph are the EM ribs—the hierarchical relationships between them [17].

In the design of the graph must respect the rules of hierarchical tree structures. For example, the presence of unacceptable individual peaks unrelated to the parent the hierarchy of EM, except the root, at the same time, EM group at the same level should be carried out by any common basis.

In parallel with the construction of the graph need to create a table EM, which will need to make the name of EM, which is similar to compiling a table of contents of educational material, when the content pre-crushed into chapters, paragraphs, etc. However, the construction of the graph content of the material, as opposed to drawing up a table of contents, there is no need to worry about the consistency of the EM [13]. It is important to display a hierarchical structure of educational material.

After the structure and content of educational material selection is necessary to formulate the requirements on the level presentation (clarity), the level of ease of learning, degree of automation (if it is necessary, for example for practical and laboratory classes). In this case, for each EM_ is necessary to generate the required

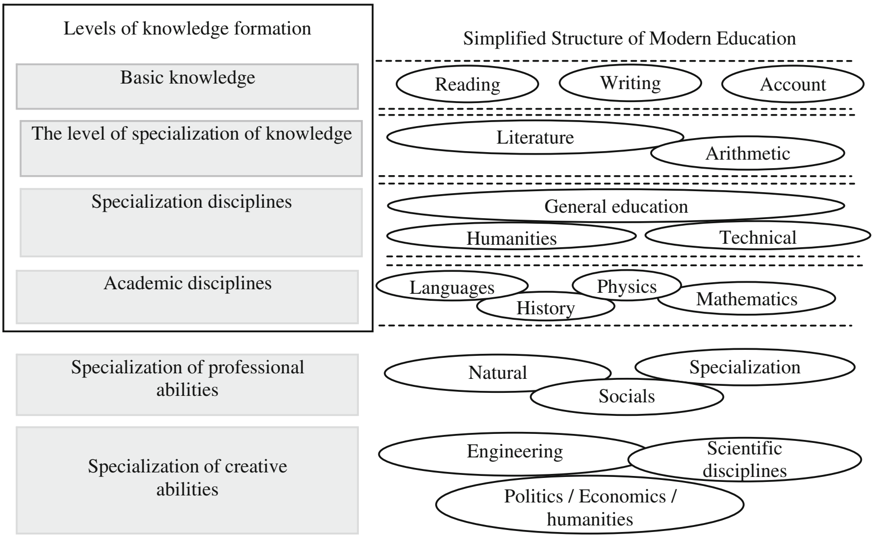


Fig. 1 General simplified model of information structure of modern education system

level of initial knowledge, which should have student a result of previous learning in other disciplines or subjects, and the level that must be achieved for the successful mastering of this EM. Educational material, so built into common directed graph, so set a clear continuity and the relationship of various educational courses or specific topics in one academic discipline [14, 18].

This model allows you to [20]:

- Clearly define the content of educational material and learning goal;
- Present the content in a clear and visible form;
- To attract experts to discuss the completeness of the contents and goal indicators at the initial stage;
- Provide a clear continuity and of academic disciplines;
- Go to a machine model representation forms of content;
- Determine the composition of the educational complex;
- Form a holistic representation of the contents of educational material as the developers and the users of the complex (teachers and students);
- Formulate the requirements for the type, number and sequence of exercises for understanding and consolidation of theoretical material [11].

Model content of educational material does not contain answers to question such as, in what sequence should the EM be studied and what logical connections are between them. These issues are discussed in the formation model of the development of educational material (Fig. 1).

3 Model Mastering of Educational Material

One of the important points of the modeled system is an automated system lies the task of creating a completely individual learning environment. As is the case with other intellectual tasks, we take as the basis of the algorithm of system behavior imitation teacher investigate similar problems. In this case, an example of the behavior of the teacher is quite laconic and useful and not in need of simplification or complication in an automated environment.

Suppose that during the student learning, the teacher notices a lack of previously conveyed knowledge or no stability control procedures knowledge of the results, which reflect the level of achievement, or lack of understanding [26]. The logical conclusion by the teacher will be to return to the previous academic subjects, for repetition of the educational process in order to eliminate uncertainties and gaps in the knowledge that the student has forgotten or overlooked. However, during the procedure the control of knowledge has received a formal evaluation, which did not correspond to the actual volume, which was not detected earlier.

With streaming form of education in the group, this method is not acceptable, so there is a regular decrease in the value of evaluations of student knowledge, which at the time did not learn the necessary amount of material due to the weak capacity or because of skipping a few classes [12]. The student will have to correct himself—identifying training topics, the absence or lack of knowledge on which the cause of deterioration in the ability to understand and mastering all further material.

This procedure requires from a student a sufficiently high level of motivation and discipline, which is not available to everyone. Delaying the solution of this problem leads to an avalanche growth of educational material required for reconsideration. Therefore, occurrence in practice a situation where we see the emergence of a stream, a large group of students, showing consistently weak results, balancing on the verge of evaluations “bad” and “satisfactory”, which is not rare graduate education institution, presenting a rather weakly educated citizens with vague perspectives finding jobs in their field and be useful to society [8].

The composition of model of mastering includes relationship matrix of priority and the logical connections EM, sequence of study of EM, the graph of logical connections EM [4]. Construction of the model is produced in four stages:

- Formation of the relations priority EM;
- Relations priorities and construction of the sequence studying of educational material in the form of a list EM;
- Formation logical connections of EM;
- Construction of a graph logical connections of EM.

The first and third stages are informal and are executed based on the analysis of educational material of experts, is educational courses (Fig. 2).

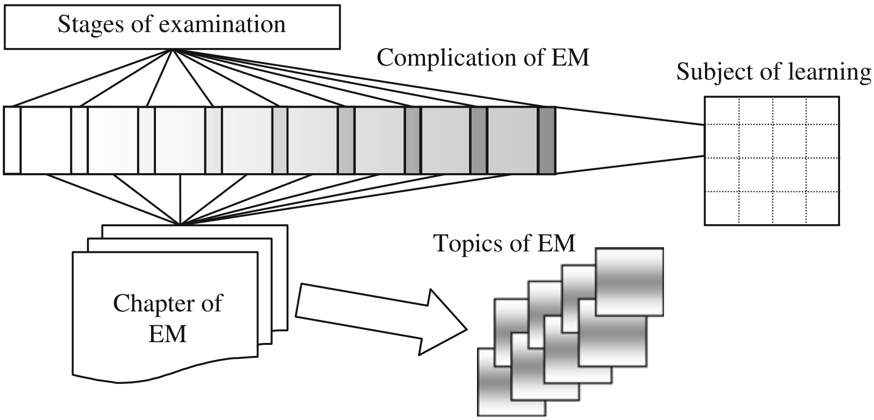


Fig. 2 Block diagram of the educational process

Thus, of EM constructed by the following order:

- Relations of priority of EM;
- Logical connections of EM;
- Studying sequence of studying of EM.

The process of filling the plan of EM is appropriate to maintain, based on texts with the educational material for all EM, if they exist. Analysis of the content of educational material allows more objectively to identify relations of priority and logical connection between the EM [7, 9].

The choice of the sequence (for example, from general to specific or vice versa), may not be obvious for all EM so the relations of priority and logical connections, and therefore in the future on the presentation of educational material, have an impact not only objective, but also subjective factors. These include plan of skilled developer of EM, his teaching experience requirements profile organization, etc.

Model mastering of educational material of the complex determines the sequence of its presentation in the manual, variants of the trajectories of its mastering determines the success of the educational process. It depends on the results shown by student knowledge in the way of control procedures, after which the curriculum is reconstructed individually [10].

4 The Model of Educational and Methodology Support

Quality education depends on many factors, but is largely determined by the state of methodological support of disciplines. A prerequisite for improving the quality of methodological support is the organization to improve educational-methodological

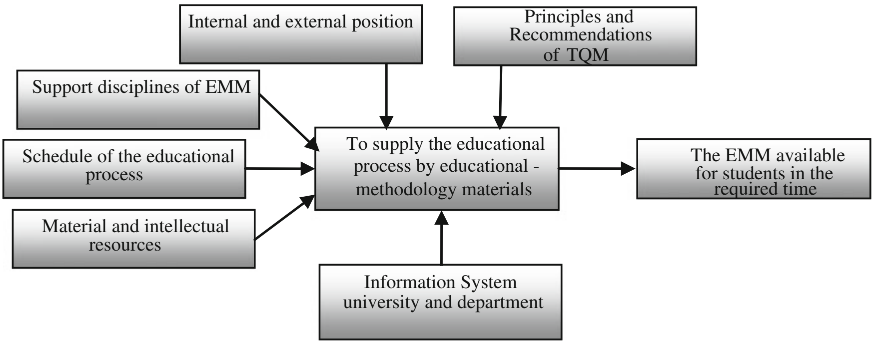


Fig. 3 The process support of educational-methodical materials

works and the broad involvement of the teaching staff creating educational-methodological development of modern high-level.

To improve efficiency of functioning educational and methodical support it is necessary to use the process approach. Process is called set of the interconnected and interacting activities which transforms inputs into outputs [21].

Selecting production educational and methodological materials in a separate process may be represented the formation of educational and methodical support of disciplines as shown in Fig. 3.

The main goal of the process is the support of educational process all necessary educational and methodical materials. At the beginning of the process graph of the educational process (disciplines of the curriculum), are indicators of educational and methodical supply of disciplines and resources of the process, which includes, in addition to the material (paper and other materials), and intellectual resources (educational-methodical work of university teachers) [25].

At the bottom is the mechanism by which the process is functioning—the information system chairs and the university. Above the process are the regulatory provisions and principles. In our case, it is the internal and external standards of educational and methodical support of the university, as well as the principles of total quality management (TQM) [24].

At the output of the production process-are educational and methodical material in the hands of the student an appropriate academic schedule time.

Process of ensuring is broken into four blocks. "It's a sub process" to determine the need the university in EMM, planning of the educational process, "implementation plans" and "process control of the educational process."

Determine the need of university in EMM is one of the most important sub-process in the effectiveness of educational and methodical support. It is presented in Fig. 4.

At the entrance—supplying disciplines of EMM and the schedule of the educational process, the output—the need for EMM by a certain deadline, and indicators of formation needs for management purposes [6].

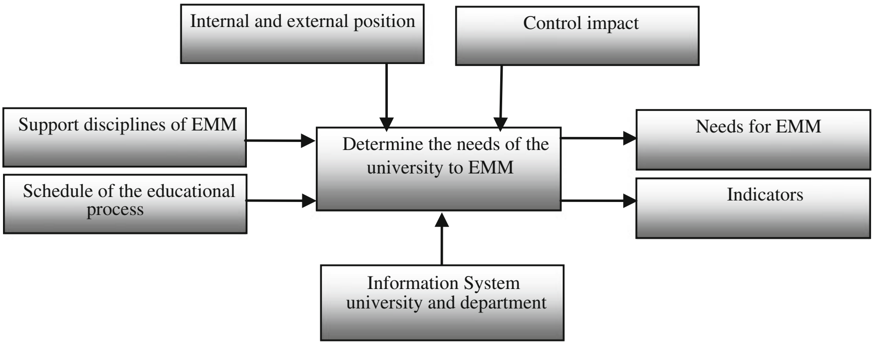


Fig. 4 The process of determining the needs of university in EMM

The most important factor in the effectiveness of this step is informational support. Really even for the average university having about 20–30 specialties (training directions, a general list of disciplines than 1,000 [22]. Considering that each discipline can be up to 5–6 educational-methodical development of different types (work program, study guide, lectures, workshops and problem books, guidelines, etc.), a general list of educational-methodical development, component of methodological software all specialties, will be calculated for several thousand. Monitoring and ensuring the balance and harmony of the array development is a very complicated process. Software based on the analysis curriculum educational and methodical condition fund the student body and displays information about the needs of the university in the EMM [31]. And here it is necessary take into account such factors as the lack of material availability in manuscript form, in the form of the original layout, in electronic form and there is a certain circulation of stock, library, branches, etc.

Next under the process of "planning of the educational process." At the beginning, the need for EMM (output of the previous sub process) and the schedule of the educational process. At the exit of the plans for the educational process, which includes a plan educational and methodical work of teachers in preparing manuscripts plan, prepublication work of editorial calving Multiplier plan, distribution plan for EMM to destinations (representative offices, branches, sending students). Also, the output indicators of planning for management purposes. The process is shown in Fig. 5.

The efficiency of this block is also largely dependent on information support, as harmonious balancing of a number of plans, given that they are interdependent, is rather laborious.

Followed by sub "implementation plans." At the entrance of the plans for the educational process (the output of the previous sub process) and the resources of the university. At the output of the finished product—EMM in the hands of students in the corresponding period [19]. Also, at the end the output indicates execution of plans for management purpose. It is shown in Fig. 6.

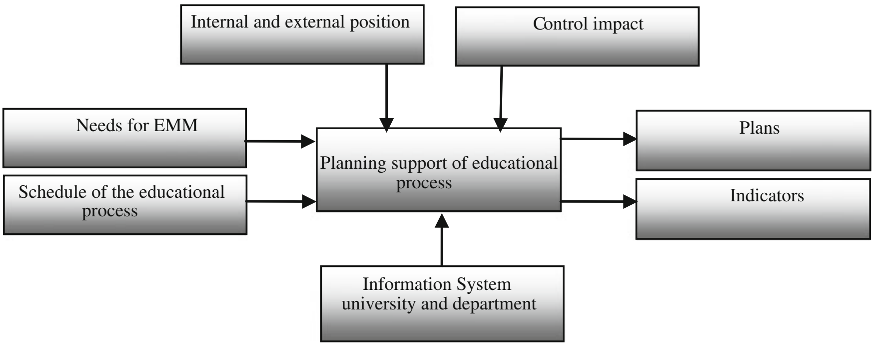


Fig. 5 The planning process of the support of educational process

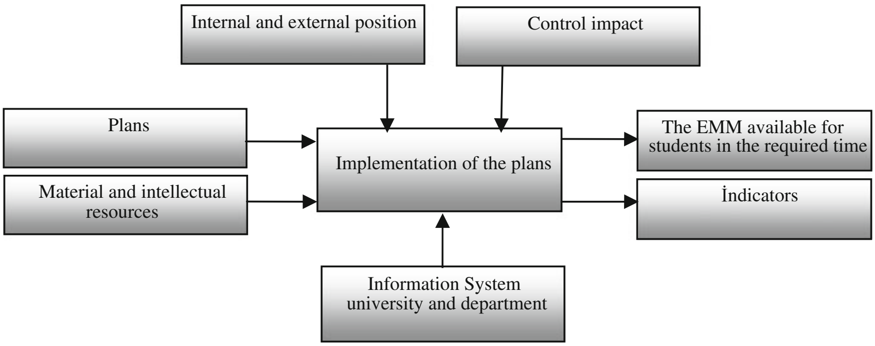


Fig. 6 The process of implementing plans

At this stage, the material product is directly created. The process consists of the following steps: preparation of the manuscript by the author, review and examination of the manuscript, prepress and the creation of the layout, the creation of electronic books and printing on paper.

Here exactly appears fully a functional barrier of organization. For the effective functioning of the whole process of educational and methodological support necessary quality management theory, it is necessary to identify the "owner" or "host" of the process, who shall assume all the powers and functions of management [27]. Typically, this "owner" is a scientific and methodical department, which is fully responsible for the educational and methodological support of the educational process.

Management is allocated a separate process. At the entrance of all the above figures come three blocks. On this basis and in accordance with internal and external regulations as well as the principles of quality management theory formed the administrative decision regarding a particular sub process. The process is shown in Fig. 7.

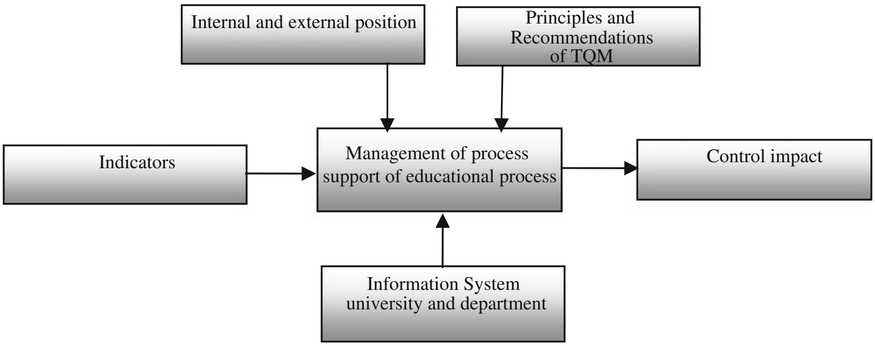


Fig. 7 Management of process of the educational process

The effective operation this block is also largely dependent on information support [28].

The advantages of this system are the integration of expert system in a distributed network of the university, of the possibility accumulation and development of the database, as well as capacity of knowledge base constructed on a sufficiently powerful and flexible rules [5]. Original representation system of educational materials allow to qualitatively upgrade educational programs, plans, and corresponding educational methods materials to meet the in light of the of satisfaction above mentioned, as presents imposes more stringent requirements for their content in terms of feasibility control tasks of understanding and mastering [35].

5 Method for Determining the Level of Knowledge

Simplified method of learning is bring to making decisions about selection the level of complexity of the question, which will be given to the student as follows, based on the result of the response to the previous question [15].

The solution to this problem depends on many parameters, most of which are unknown in an intellectual system. However, rather the exact answer can be found using the mathematical apparatus of fuzzy logic.

In this chapter, to analyze the current situation, the following events play a role.

Algorithms of Decision Making are based on the results of the following tasks [29]:

- Preliminary analysis of the student's knowledge—is used to assess the level of knowledge of the student, to make a decision on the choice of the first question (lagging behind student will first be asked from the group of simple, while qualified students will be asked more complex);
- The previous question the student answered correctly—in this case, the student is asked a question of increased complexity;

- The student answered the previous question incorrectly—in this case, the student is asked the question of reduced complexity;
- Analysis of the results and a decision making on the final evaluation or of continuing testing—number of correct answers multiplied by the complexity in relation to errors and multiplicity of correct and incorrect answers are input subprograms decision making to issue evaluation, or if there is a high probability uncertainty, testing continues [3].

This list reflects the real computational tasks facing the analysis program. Decision-making in the choice of questions, which, according to the program, the respective abilities of the student. When an incorrect answer is produced revaluation data about the student and the follow will be asked less complicated question. With the correct answer, the program asks question with progressive complexity [36].

This method of decision-making of questions selection sequence allows the individual to make the process of learning or testing, and at the end, to give the most accurate evaluation of student abilities [34].

At the end of the test—it can be a decision of a student and teacher (for example, due to the end of time), summing up the results on, the number of correct answers and their complexity. An analysis program produces a change in the relevant database records of the student, and then begins the process of analysis, which aims to provide information about the student.

This information may include: the current level of intelligence of the student and the comparison with the previous values, the analysis of questions for which the student answered incorrectly and displays the correct answers with commentary. Comments are included by teacher when entering questions in the database.

The question of value for the evaluation test may be accepted as a program and the by teacher or their joint decision. This method allows decision making to conduct training or testing of any level of complexity, from the individual to the state test exam [30].

Consider the block of intelligent system, which has the responsibility of testing [23].

As has been previously mentioned due to the large number of external parameters analysis and issuance of decisions are made using the mathematical apparatus of fuzzy logic. The duties of the subsystem also includes testing and meeting the following requirements:

1. Protection of correct answers from unauthorized access;
2. Protection of the number of correct answers to the questions from the student editing;
3. Ensuring equal conditions for making a test.

Choice of questions carried out by performance multiple calculations set of fuzzy expressions whose ultimate goal is to transfer decision-making subsystem their results [16].

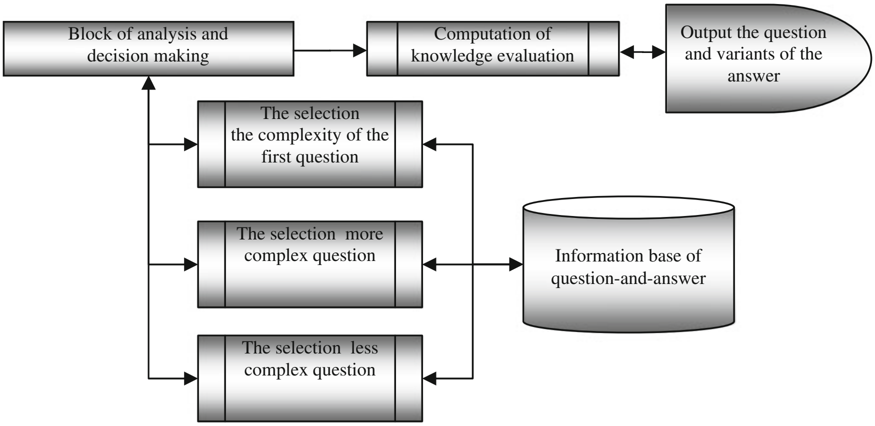


Fig. 8 A simplified diagram of the system of control knowledge

These expressions can be represented in the form of conditional statements complicated structure (Fig. 8).

As a very simple example is the expression of the form [33]:

If the PreviousAnswer = Right,
THEN CorrectAnswers = CorrectAnswers +1

The next complication is keeping weight questions (complexity):

If the PreviousAnswer = right,
Then WeightCorrectAnswer = WeightCorrectAnswer +
TableWeight (IndexCurrentAnswer)

Adding records of the elapsed time and several other parameters, and complicating the logic expressions, a level of intelligence of subsystem of testing can be achieved that can be compared with the real teacher survey [32]. Additional parameters in mathematical expressions serve the level values of the following characteristics: the ability to remember, attentiveness, reaction speed, speed of decision-making, speed reading, etc.

In the process of working with the program, the student cannot only test his or her knowledge, but also to learn. This is achieved not only when the question is asked, but also with the presence of all of the comment and explanations given by the teacher.

On reaching the stable results in a certain group of questions, the student can go to the next level of question complexity [2]. This transition will enable the student to not stop at the achieved results but to develop further.

Gradual in learning will give the necessary time to complete mastering and fastening of the material, and then transition to a new, more complex material. Each transition is accompanied by a small test on the previous material with an analysis of its mastering.

6 Conclusion

In this chapter, we solve the problem effectively integrate traditional learning processes with the processes of open education, with the possibility of learning without leaving the job, directly to the student residence, the ability to produce quality and competitive of courses and will gradually lead to the standardization of various countries. Displayed continuously increasing need for intelligence information system is solved first, the development of more efficient algorithms and methods for an electronic educational material. Secondly, the application of the developed models and methods will improve the hardware, software, and information components of third-party automation systems learning and control of knowledge. In conclusion, this chapter finalizes the analysis and projects the architecture of future system. It also defines the basic components and modules, conducted functional design, methods of interaction for the implementation of the presented task.

References

1. Abbasov A.M.: Information Boom: new trends and expectations. Studies in fuzziness and soft computing, soft computing: state of the art theory and novel application, pp. 1–12. Springer (2012)
2. Badami, V., Chiang, K.H., Khedkar, P.S., Marcelle, K.W., Schutten, M.J.: Industrial applications of fuzzy logic at general electric. Proc. IEEE. **83**(3), 450–465 (1995)
3. Barsky, A.B.: Neural Networks: recognition, management, decision-making. Finance and statistics, pp. 30–63. Moscow (2004)
4. Bellman, R., Zadeh, L.A.: Decision-making in ambiguous circumstances, issues analysis and decision-making, pp. 180–199. Springer, New York (1976)
5. Bernshteyn, L.S., Bojenyuk, A.V.: Fuzzy Models of Decision Making: Deduction, Induction, Analogy, pp. 78–99. Univ Tsure, Taganrog (2001)
6. Borisov, V.V., Kruglov, V.V., Fedulov, A.S.: Fuzzy Models and Networks, pp. 224–284. Hot line–Telecom, Moscow (2007)
7. Bouchon-Meunier, B., Yager, R.R.: Fuzzy Logic and Soft Computing (Advances in fuzzy systems: application and theory), pp. 84–93, 103–119. World Scientific, Singapore (1995)
8. Bouchon-Meunier, B., Valverde, L., Yager, R.R.: Uncertainty in Intelligent Systems. Elsevier, Amsterdam (1993)
9. Bouchon-Meunier, B., Yager, R.R., Zadeh, L.A.: Uncertainty in Knowledge Bases. Springer, Heidelberg (1991)
10. Cordon, O., Herrera, F.: Linguistic modeling by hierarchical systems of linguistic rules. Technical Report # DECSAI–990 114, Department of Computer Science and A. I, pp. 187–215. University of Granada July 1999

11. Dubois, D., Prade, H., Yager, R.R.: Readings in Fuzzy Sets. Morgan Kaufman, San Mateo (1993)
12. Gorbunova, L.G.: On the realization of the rating system in pedagogical high schools. In: Proceedings of 2nd international technical conference (Part 1), pp. 105–106. University Education, Penza (1998)
13. Hanss, M.: Applied Fuzzy Arithmetic: An introduction with engineering applications. 1 edn, pp. 100–116, 139–147. Springer (2004)
14. Heaton, J.: Introduction to Neural Networks for C. 2nd edn, pp. 224–231. Heaton Research Inc, St. Louis (2008)
15. Kandel, A., Langholz, G.: Fuzzy Control Systems, pp. 159–187. CRC Press LLC, Boca Raton (1993)
16. Laurene, V.F.: Fundamentals of Neural Networks: Architectures, Algorithms and Applications. US edn, pp. 103–121. Prentice Hall, Englewood Cliffs (1993)
17. Ledeneva, T.M.: Fuzzy Information Processing: a tutorial/TM Ledeneva, pp. 212–233. Voronezh State University, Voronezh (2006)
18. Nikravesh, M., Zadeh, L.A., Kacprzyk, J.: Soft Computing for Information Processing and Analysis, pp. 93–99. Springer, New York (2005)
19. Novak, V., Perfilieva, I., Mochkorzh, I.: Mathematical Principles of Fuzzy Logic. In: Averkina, M. (ed.) FIZMATLIT, pp. 209–252 (2006)
20. Shahbazova, S.H., Freisleben, B.: A network-based intellectual information system for learning and testing. In: 4th international conference on application of fuzzy systems and soft computing, pp. 308–313. Siegen (2000)
21. Shahbazova, S.H.: Decision-making in determining the level of knowledge of students in the learning process under uncertainty. Informatica Int. J. Comput. Inf. **37**(3), 339–344 (2013) (Print edition ISSN: 0350–5596)
22. Shahbazova, S.H.: Application of Fuzzy Sets for Control of Student Knowledge. Appl. Comput. Math. Int. J. **10**(1), pp. 195–208 (2011) (Special issue on fuzzy set theory and applications, ISSN 1683–3511)
23. Shahbazova, S.H.: Functional design of the control of knowledge on base of fuzzy logic. In: International conference on application of information and communication technology and statistics in economy and education, pp. 24–31. Sofia, Bulgaria (2012)
24. Shahbazova, S.H.: Application of fuzzy sets for control of student knowledge. Appl. Comput. Math. Int. J. **10**(1), 195–208 (2011) (Special issue on fuzzy set theory and applications, ISSN 1683–3511)
25. Shahbazova, S.H.: Development of the knowledge base learning system for distance education. Int. J. Intell. Syst. 27(4), 343–354 (2012)
26. Shahbazova, S.H.: Applied research in the field of automation of learning and knowledge control: Studies in fuzziness and soft computing. Soft computing: state of the art theory and novel application, pp. 223–240. (2012)
27. Shahbazova, Sh: Simulating the behavior of the teacher, the use of expert systems in the field of educational systems, control systems and machines. J. Inst. Cybern. Glushkov Nat. Acad. Sci. Ukraine **3**, 68–75 (2012)
28. Shahbazova, Sh: Investigation of the basic problems and trends of traditional education systems. Int. J. Technol. Manage. Inf. Probl. Ukraine **3**, 110–117 (2013)
29. Yager, R.: Fuzzy sets and theory of possibilities: Recent advances. (ed.) Radio and Communications. 196, pp. 391–405
30. Yager, R.R., Zadeh, L.A.: Fuzzy Sets, Neural Networks and Soft Computing. Van Nostrand Reinhold, New York (1994)
31. Yager, R.R., Zadeh, L.A.: An Introduction to Fuzzy Logic Applications in Intelligent Systems. Kluwer Academic Publisher, Norwell (1992)
32. Yager, R.R., Fedrizzi, M., Kacprzyk, J.: Advances in the Dempster-Shafer Theory of Evidence. Wiley, New York (1994)
33. Zadeh, L.A., Kacprzyk, J.: Fuzzy Logic for the Management of Uncertainty, pp. 75–84. Wiley, New York (1992) (1st Printing edition)

34. Zadeh, L.A., Klir, G.J., Yuan, B.: Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh, pp. 60–69 (1996)
35. Zadeh, L.A.: The new approach to the analysis of difficulty systems and decision processes, pp. 23–37. Knowledge Mathematics Today, Gottingen (1974)
36. Zadeh, L.A.: The Concept of Linguistic Variable and its Application to the Adoption of Approximate Solutions, pp. 140–164. Springer, New York (1976)

Decision-Making Based on Fuzzy Estimation of Quality Level for Cargo Delivery

Yuriy P. Kondratenko and Ievgen V. Sidenko

Abstract This chapter presents the proposed approach and algorithms for designing hierarhical decision support systems (DSS) based on fuzzy logic with flexible rule base. Special case of changing the structure of the input data's vector for DSS in transport logistics is considered by authors. The main idea is a correction of fuzzy rule base of fuzzy DSS when different decision-makers can decrease dimension of the vector of DSS's input coordinates according to their own priorities and criteria. Simulation results confirm the effectiveness and appropriateness of editing fuzzy knowledge bases rules for DSS which solve the problems of transport logistics.

1 Introduction

Decision support systems (DSSs) are widely used in economics, enterprise organization, medicine, agriculture, technical diagnostics and others [1–3] for improving significantly the quality of decision-maker's (DM) choice of optimal decision from a set of alternative variants in difficult or extreme situations.

One of the important areas of DSS appliance is transport logistics [2, 4, 5]. Formation and organization of work of cargo delivering chains is connected with intense and operational information exchange between participants of transport process, fast reaction and high demands to the quality of cargo delivering [6]. The process of making effective decisions is to choose the best alternative variant

Y. P. Kondratenko (✉) · I. V. Sidenko
Department of Intelligent Information Systems, Petro Mohyla Black Sea State University, 68-th Desantnykiv str. 10, Mykolaiv 54003, Ukraine
e-mail: y_kondratenko@rambler.ru

I. V. Sidenko
e-mail: emoty@mail.ru

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_21,

among the existing ones on a particular system of criteria and preferences, and set of evaluative dimensions (input system coordinates).

Customers of transport services often are not satisfied with the quality of services, as there take place violations of delivery terms, spoiling and loosing cargo, problems with flow of documents, considering a difficult cooperation process of big amount of forwards, carriers and logistic companies, complexity of building rational routes of cargo transportation, absence of universal program systems, that accompany processes of cargo transportation in real time [5, 7].

Particularly acute are problems of designing fuzzy DSS with flexible structure, i.e. systems, the amount of input parameters of which can change during the decision-making process by DM [6].

2 General Problem Statement

Last decade in control systems and DSS that functioning in conditions of uncertainty are widely used intellectual technologies, in particular, based on fuzzy logic [8, 9]. For different DM the part of input coordinates can be not important according to their own priorities [6, 9]. For example, for DSS which can support the process of sensor choosing from a set of alternative sensors the 5 input coordinates are mostly important: power consumption, price, embedded functionality, size and noise $(N = 5)$. But in some cases a specific DM may be interested in and he can only ask, for example, just 3 input signals $(N_r = 3)$, and other 2 input signals $(N_{NI} = 2)$ are not interested (not important) for DM. That is the dimension of the vector of the input coordinates reduces from 5 to 3. In this case the correction of the fuzzy rule bases is needed in the interactive DSS's modes.

This research deals with automatic correction of fuzzy DSS's rules base for variable structure of input coordinates vector and modeling of actual situations to approve working ability and effectiveness of proposed method of antecedents and consequents correction. Main focus is concentrated on the modern DSS that will provide choice of optimal decisions when solving problems of transport logistics with big amount of input parameters, multicriterion structure, the lack in many cases of sufficiently complete prior information [2, 4, 5].

3 The Analysis of Related Publications

Improvement of cargo delivering quality by different companies-carriers is done by means of collecting the consolidated cargo from several shippers, developing optimal routes of cargo transferring, and also connecting more efficient type of transport on certain stage of transportation [1, 2]. Irrational transportations lead to increase of logistic and first of all transport expenses, and also additional workload of transport routs.

Today there exist a lot of publications on the research of DSS based on the fuzzy logic [6, 9, 10], which examine methods of the theory of fuzzy sets for modeling, analysis and synthesis of intelligent systems. Researches, which are conducted in different countries, have proved that for many subjects to management, parameters of which change in the process of operation, it is appropriate to use fuzzy computerized automatic control systems [3, 9, 11]. Learning procedure of the fuzzy Mamdani-type model presents a nonlinear optimization problem, within a framework of which the main attention is paid to achieving the maximum accuracy of learning of the fuzzy DSS [11].

Own priorities of DM essentially influence on formation in each actual situation of the input coordinates vector's dimension. Herein some of input coordinates for one DM can be important, and for another DM—non important. Among famous approaches to correction in such cases rules bases of fuzzy DSS is usage of weighting coefficients for fuzzy rules. A change of weighting coefficients vector for corresponding rules of fuzzy knowledge bases allows reducing the influence of input parameters, which by the choice of DM in some situations may not participate in the process of decision-making, on the result of system work. However, at the same time there appears a need to re-configure corresponding coefficients at each change of the input data structure [9].

The work [12] considers the problem of developing effective methods and algorithms for optimization of fuzzy rules bases of Sugeno-type fuzzy controllers that can be applied to control of dynamic objects, including objects with non-stationary parameters. Proposed in [12] approach based on calculating the coefficient of influence of each of the rules on the formation of control signals for different types of input signals provides optimization of a linguistic rule bases by using exclusion mechanism for rules with negligible influence. The publications [9, 13–16] also deals with solving problem of fuzzy rule base reduction which based on the different approach.

The main difficulty in the process of setting multidimensional fuzzy models is the large number of parameters of such systems. Moreover, their number rapidly increases with enlargement of the number of inputs and fuzzy linguistic terms needed to assess the relative values. One of the ways of its effective solution is to move from the regular division of input space to irregular [9], which consists of rectangular segments. Another way is to avoid a grid partitioning of input space and to use a non-grid partitioning, in particular, rectangular (k-d tree partition) and quadratic partition (quad tree partition). The purpose of a non-grid partitioning is to reduce the number of fuzzy segments [9].

Based on publications [9, 17] devoted to the study of simplification of fuzzy models proved the feasibility and effectiveness of the method of local models.

Considered methods for correcting fuzzy rule bases can't be directly applied for optimizing fuzzy hierarchical DSS with flexible structure of input vector that caused the feasibility researches authors in this direction.

4 Fuzzy DSS's Multi-level Structure

Let's consider a fuzzy system which modeling dependence:

$$y = F(x_1, x_2, \ldots, x_N), \tag{1}$$

where x_i, $i = 1\ldots N$—input linguistic variables; y—output variable.

Fuzzy systems by hierarchic properties are divided into one-level and multi-level. To simulate multidimensional fuzzy dependences (1) should be used hierarchical approach to building DSS based on fuzzy logical derivation [7]. In such systems, the output of one subsystem is given to the input of another with a higher level of hierarchy. The experience of creating such systems shows that their efficiency is achieved in that case when the number of separate subsystem inputs does not exceed five ($N \leq 5$). Therefore, with a large number of input variables is needed their hierarchical structuring by the general (common in the frames of subsystem) properties [8].

Let's consider, for example, fuzzy single-level system, which models the dependence (1), while the number of input parameters is $N = 9$. Relations between input variables and output variable are described by fuzzy rules of a knowledge base F. As a result of hierarchic structuring fuzzy one-level system can be represented as, for example, three-level hierarchic system. Each of subsystems f_1, f_2, f_3, f_4, f_5 of this system represents corresponding fuzzy dependences: $y_1 = f_1(x_1, x_2, x_3, x_4)$, $y_2 = f_2(y_1, y_3)$, $y_3 = f_3(x_5, x_6)$, $y_4 = f_4(x_7, x_8, x_9)$ and $y = f_5(y_2, y_4)$.

The results of the research [9, 12] of fuzzy hierarchical organized systems proved that in the process of developing fuzzy DSS with hierarchical structuring of the input parameters, the number of rules is greatly reduced, and consequently takes place a reduction of fuzzy knowledge base (rules).

Method of actual correction of fuzzy DSS rule bases (RB) by levels of hierarchy can simplify RB, reducing the number of rules with decreasing the dimension of the input signal vector X. This is achieved by reducing the number of rules at appropriate levels when excluding the least important for the DM input parameters, on the basis of which subsystems of each hierarchical level are pre-formed. Lack of interest of DM, for example, in the input parameters $\{x_1, x_2, x_3, x_4\}$ of the subsystem f_1 will significantly reduce the number of rules of the system, because in the future they will not take part in the process of decision-making $y_1 = f_1(x_1, x_2, x_3, x_4) = NI$.

5 Rule Base of Fuzzy DSS for Estimation of Cargo Delivery Quality

Analysis of review of the transport logistics tasks and a method of evaluation of the quality of cargo delivery using fuzzy DSS are presented in the article [7]. For fuzzy DSS of such type (with 19 input coordinates) particularly actual is a problem

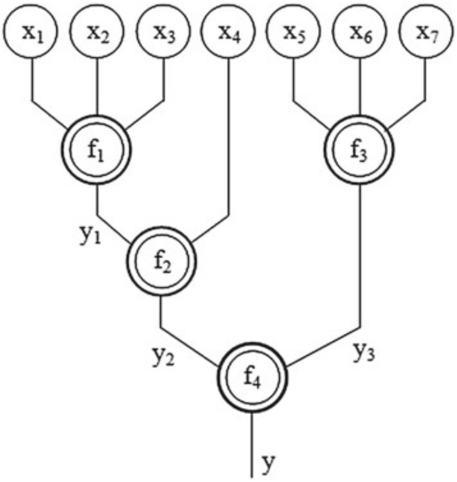


Fig. 1 Hierarchical structure of fuzzy DSS with 7 inputs

of automatic adjustment of fuzzy RB in situations when a specific DM is interested in real N_r input coordinates, herewith $N_r < N$. Hereinafter we will in more detail way observe an proposed approach that allows with the help of DSS of aforementioned class providing a creation of optimal decisions (according to the content of previously selected criteria) by correction of the RB at a prior uncertainty of input information, that is $N_r < N$.

For illustration of the proposed approach as an example we will observe fuzzy DSS that has one output y and 7 input parameters x_i, $i = 1, \ldots, 7$; $N = 7$. Among them are x_1—tracking cargo in process of cargo transferring, x_2—level of cargo security protection, x_3—insurance of cargo, x_4—monitoring of flow of traffic, x_5—price of transport services, x_6—assortment of cargo for delivery, x_7—alignment of freight participants actions, y—evaluation of the quality of transport services.

As a result of structuring of the input data as a part of the three-level DSS 4 fuzzy subsystems were formed that implement the following dependences [6, 7]:

$$y_1 = f_1(x_1, x_2, x_3);\ y_3 = f_3(x_5, x_6, x_7);$$
$$y_2 = f_2(y_1, x_4);\ y = f_4(y_2, y_3).$$

The structure of the hierarchical fuzzy DSS, that consists of 7 input linguistic variables $\{x_1, x_2, \ldots, x_7\}$, 4 RB with fuzzy rules $\{f_1, f_2, f_3, f_4\}$ and one output linguistic variable y, is presented in Fig. 1.

In the process of structuring the input variables are combined by common characteristics that are principal (important) for a particular fuzzy subsystem. Such hierarchical approach allows reducing the amount of fuzzy rules of the RB and thus increasing the sensitivity of the system to the operations of the input variables (factors) [3, 18, 19].

Table 1 Rules for the first subsystem $y_1 = f_1(x_1, x_2, x_3)$

No of rule	1	2	3	4	5	6	7	8	9
x_1	L								
x_2	L			M			H		
x_3	L	M	H	L	M	H	L	M	H
y_1	L	L	LM	L	LM	M	LM	M	MH
No of rule	10	11	12	13	14	15	16	17	18
x_1	M								
x_2	L			M			H		
x_3	L	M	H	L	M	H	L	M	H
y_1	L	LM	M	LM	M	MH	M	MH	H
No of rule	19	20	21	22	23	24	25	26	27
x_1	H								
x_2	L			M			H		
x_3	L	M	H	L	M	H	L	M	H
y_1	LM	M	MH	M	MH	H	MH	H	H

While describing the linguistic variables for DSS that was presented in Fig. 1, a diapason of variation values, the number of terms and the form of the MF (triangular) were defined. Corresponding the structure of DSS (Fig. 1) the authors has created three fuzzy RB with fuzzy rules of production of "IF—THEN" type, the first one of which $y_1 = f_1(x_1, x_2, x_3)$ will be further observed in more detail.

A set of rules from the RB of the first subsystem $y_1 = f_1(x_1, x_2, x_3)$ of fuzzy DSS is presented in Table 1. Herewith for evaluation of all input variables $\{x_1, x_2, \ldots, x_7\}$ in three linguistic terms are used (L—"low", M—"medium", H—"high"), and for evaluation of output variable—5 corresponding terms (L—"low", LM—"lower than mean", M—"medium", MH—"higher than mean", H—"high").

The project of hierarchically organized DSS for the evaluation of the quality of cargo delivery, whose structure is presented in Fig. 1, can be synthesized, for example, in computing environment MatLab or in a computing environment FuzzyTECH [8, 20, 21].

6 Preliminary Fuzzy Rules Reducing

The purpose of non-grid partitioning is to reduce the number of fuzzy segments. Partition of the input space will be denser in areas where the simulated system for surface display changes more rapidly and less dense in areas with a smooth surface. Within each segment is only one rule, so here it is advisable to use the Takagi-Sugeno model. In these models, the output of each rule is not a fuzzy set. It is the function (usually linear). An example of such a rule may be an expression of the next form [22]:

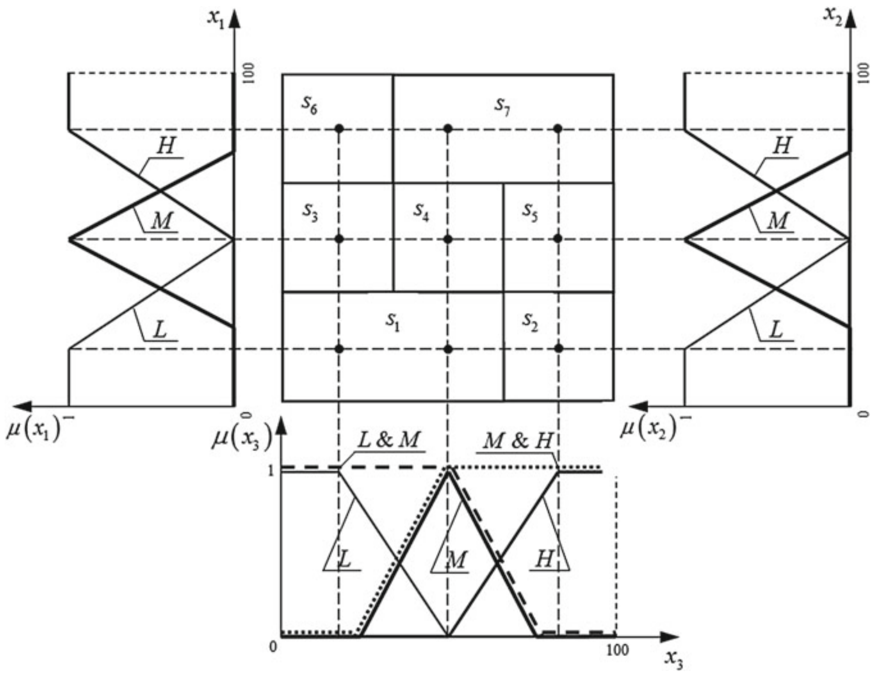


Fig. 2 The non-grid partitioning the input space to the first fuzzy subsystem $y_1 = f_1(x_1, x_2, x_3)$

$$\text{IF } x_1 = A_{11} \text{ AND } x_2 = A_{21} \quad \text{THEN y} = a_{11}x_1 + a_{21}x_2 + a_{01},$$

where A_{11}, A_{12} C {L, M, H}; a_{11}, a_{21}, a_{01}-consequent coefficients

Nevertheless can be used the model of Mamdani-type [23].

In research [9] is considered an example of non-grid partitioning of the input space into three segments using Gaussian membership functions of fuzzy sets.

Let's consider the approach of non-grid partitioning the input space to the first fuzzy subsystem $y_1 = f_1(x_1, x_2, x_3)$. Graphical representation of non-grid partitioning in seven segments $s_1, \ldots, s_7$ is presented in Fig. 2.

Each segment can match the one rule that specifies the portion of the surface model associated with this sector. Thus, instead of 27 rules, the model contains 25 rules (two of which are modified). Some of them are presented below.

$$\text{R1:IF } x_1 = L \quad \text{AND } x_2 = L \quad \text{AND } x_3 = L\&M \quad \text{THEN } \text{y}_1 = L(s_1),$$
$$\text{R3:IF } x_1 = L \quad \text{AND } x_2 = L \quad \text{AND } x_3 = H \quad \text{THEN } \text{y}_1 = LM(s_2),$$

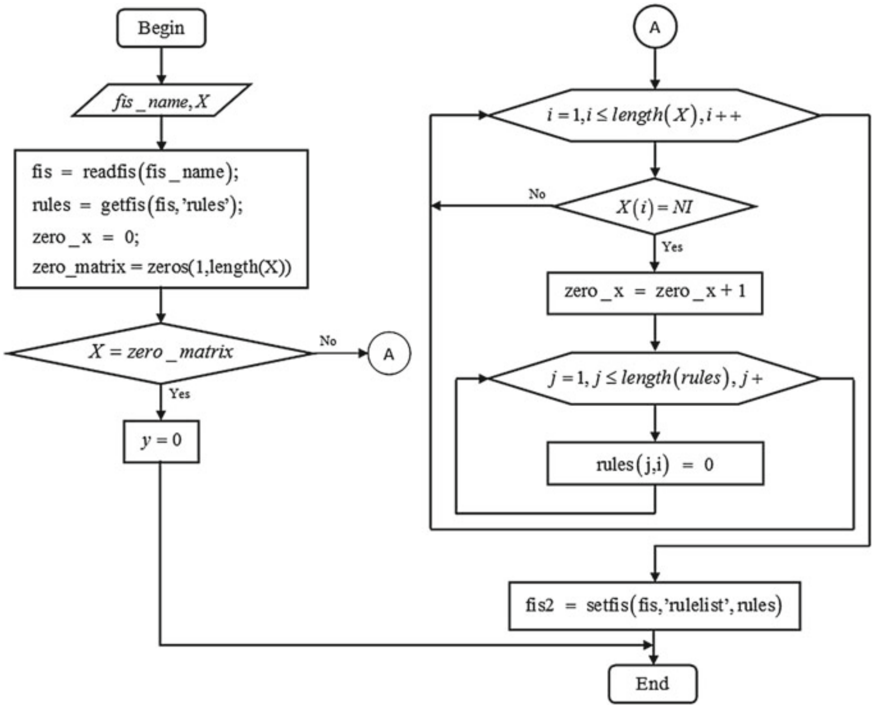


Fig. 3 Block-diagram of the editing algorithm of rules antecedents of DSS

$$\begin{aligned}
&\text{R13:IF } x_1 = M \quad \text{AND } x_2 = M \quad \text{AND } x_3 = L \quad \text{THEN } \mathrm{y}_1 = LM(s_3),\\
&\text{R14:IF } x_1 = M \quad \text{AND } x_2 = M \quad \text{AND } x_3 = M \quad \text{THEN } \mathrm{y}_1 = M(s_4),\\
&\text{R15:IF } x_1 = M \quad \text{AND } x_2 = M \quad \text{AND } x_3 = H \quad \text{THEN } \mathrm{y}_1 = MH(s_5),\\
&\text{R25:IF } x_1 = H \quad \text{AND } x_2 = H \quad \text{AND } x_3 = L \quad \text{THEN } \mathrm{y}_1 = MH(s_6),\\
&\text{R26:IF } x_1 = H \quad \text{AND } x_2 = H \quad \text{AND } x_3 = M\&H \quad \text{THEN } \mathrm{y}_1 = H(s_7)
\end{aligned}$$

Ability to define segments of the input space (Fig. 2) due to the fact that it uses a membership functions «*L&M*» and «*M&H*». Their kernels by the length cover practically the whole of the segment length [9]. The use of these membership functions is possible only if the consequents of two or more rules are the same. In this case, the antecedents of the rules differ only in one coordinate.

In our example, the consequents of rules No 1 and No 2 are the same. In this case, the antecedents of coordinates x_1 and x_2 corresponds to linguistic term «-L» and coordinate x_3 changes from linguistic term «L» to «M». Therefore, we can specify an additional term «*L&M*», which combines two linguistic terms «L» and «M» input coordinate x_3. The same thing happens with the rules No 26 and No 27 with using additional linguistic term «*M&H*». The use of such membership functions is one of the methods to reduce the number of rules.

In some cases [9] it is appropriate to use the compression method for fuzzy rule processing. The number of fuzzy rules of each subsystem can be reduced using an approach based on the compression rules. According to this principle, as example for first fuzzy subsystem (Fig. 1), rule No 1 and No 10 (Table 1) can be combined to one resulting rule, the structure of which is as follows:

$$\text{IF } x_1 = L \quad \text{AND } x_2 = L \quad \text{AND } (x_3 = L \text{ OR } x_3 = M) \quad \text{THEN } \mathrm{y}_1 = L$$

7 Method of Antecedents and Consequents Correction for Variable Structure of the DSS Inputs

In the process of fuzzy DSS work with a fixed structure of the RB and at a variable structure of the vector of input data $N_r < N$, the results of making decisions y undergo deformation. This is due to the fact that the values of the input parameters (signals) that do not take part in modeling of fuzzy DSS $(x_i = 0, i \in \{1, 2, \ldots, N\})$, carry out negative impact on the result y through the appropriate fuzzy rules. To solve this problem the authors have developed an approach (based on two algorithms of editing of rules antecedents and consequents), which consists in correction of the rules of fuzzy RB at variation of input parameters that allows not to take into account the values of input signals $x_i = NI$, $i \in \{1, 2, \ldots, N\}$, which are not important for DM in the process of decision making.

A block-diagram of the editing algorithm of rules antecedents of fuzzy DSS is presented in Fig. 3.

With the implementation [20] of the algorithm (Fig. 3) all the rules of RB of the first subsystem $y_1 = f_1(x_1, x_2, x_3)$ are processed (27 rules from Table 1). If one of the input signals (x_1, x_2, x_3) is not interesting for DM, for example, $x_1 = NI$, then the algorithm of editing antecedent of rules will change the value of input coordinate x_1 terms to the value of zero signal. Since the input parameter x_1 will not take part in decision-making process, then the value of its antecedent is also excluded from consideration. Herein the values of coordinates antecedents x_2 and x_3 remain unchanged in 27 rules (Table 1), but all possible situations (complete fuzzy model [9]) of values of their antecedents have been considered in rules 1–9 and are repeated in blocks of rules 10–18, 19–27.

Therefore, if the antecedents are repeated, then only one rule with the appropriate antecedent is left that indicates about reduction in the RB from 27 (Table 1) to 9 rules.

A block-diagram of the editing algorithm of rules consequents of fuzzy DSS is presented in Fig. 4.

These algorithms can be applied in the process of designing of fuzzy DSS with a variety structure of the input data.

After correcting the antecedents of RB of fuzzy DSS it is needed to edit the consequent of RB according to modified antecedents. This is necessary because the value of coordinate antecedents, for example x_1, with zero meaning $x_1 = 0$ through appropriate fuzzy rules undermine the result of the output variable y_1 (Table 1). The value of the zero signal $x_1 = 0$ corresponds to term «L» of antecedent of each rule (rules from the Table 1). In addition, let's consider, for example, the situation when the value of antecedents coordinates x_2 and x_3 correspond to term «M», then the value of consequent output variable y_1 will respond to term «LM» (rule number 5 from the Table 2). If we exclude from consideration antecedent coordinate x_1, i.e. $x_1 = NI$, then the rule number 5 will simplify to a single variable, the formula takes the following form IF $x_2 = M$ AND $x_3 = M$ THEN $\mathrm{y}_1 = LM$ that does not correspond to reality. The value of consequent of output coordinate in such situation should meet the term «M», as input variables x_2 and x_3 correspond to term «M». So setting to zero the input coordinate $x_1 = 0$ leads to deformation of results of fuzzy DSS work, unlike actual correction of consequents of rules RB.

Let's illustrate realization of method on different actual sets of values: I—$y_1 = f_1(x_1, x_2, x_3)$; II—$y_1 = f_1(x_1, x_3)$; III—$y_1 = f_1(x_1)$; IV—$y_1 = f_1(x_2, x_3)$; V—$y_1 = f_1(x_2)$. Here in, the actual sets of values characterize different-type actual situations, for example, for the first subsystem $y_1 = f_1(x_1, x_2, x_3)$(Fig. 1, Table 1), which can appear in the process of decision making.

For abovementioned actual sets of values when implementing the first stage of editing rules on the basis of algorithms, which is presented in Fig. 3, RB of the first subsystem $y_1 = f_1(x_1, x_2, x_3)$ (Table 1) is transformed into reduced by the amount of RB rules. At the I set of values, the rules with number 1 to 27 are remained in RB, at the II set—rules No 1, 2, 3, 10, 11, 12, 19, 20, 21, at the III set—rules No 1, 10, 19, at the IV set—rules No 1–9, at the V set—rules No 1, 4, 7.

Let's consider in more detail a stage of editing consequents of reduced RB rules (Fig. 4) for fours actual set of values. Evaluation and correction of consequent of output coordinate's rules y_1 of the first subsystem $y_1 = f_1(x_1, x_2, x_3)$ for IV actual set of values we will consider on the example of the 3rd rule (Fig. 4). For presented in Table 1 third rule the value of antecedent of input (x_1, x_2, x_3) and consequent of output y_1 coordinates can be represented as follows:

$$\text{IF } x_1 = L \quad \text{AND } x_2 = L \quad \text{AND } x_3 = H \quad \text{THEN } \mathrm{y}_1 = LM,$$

where (L, LM, H) are linguistic terms of corresponding subsystem coordinates. For the fourth actual situation (IV) in the decision making process two input variables take part, i.e. $y_1 = f_1^{\mathrm{IV}}(x_2, x_3)$.

In this case, antecedent of coordinate x_2 corresponds to linguistic term «L», and of coordinate x_3—«H», at the same time, with the consequent of output variable y_1 remains unchanged and corresponds to term «LM» (Table 1). The actual situation requires correction of consequent value as according to expert evaluation for $x_2 = L$ and $x_3 = H$ the value of consequent of output variable y_1 has to match term «M».

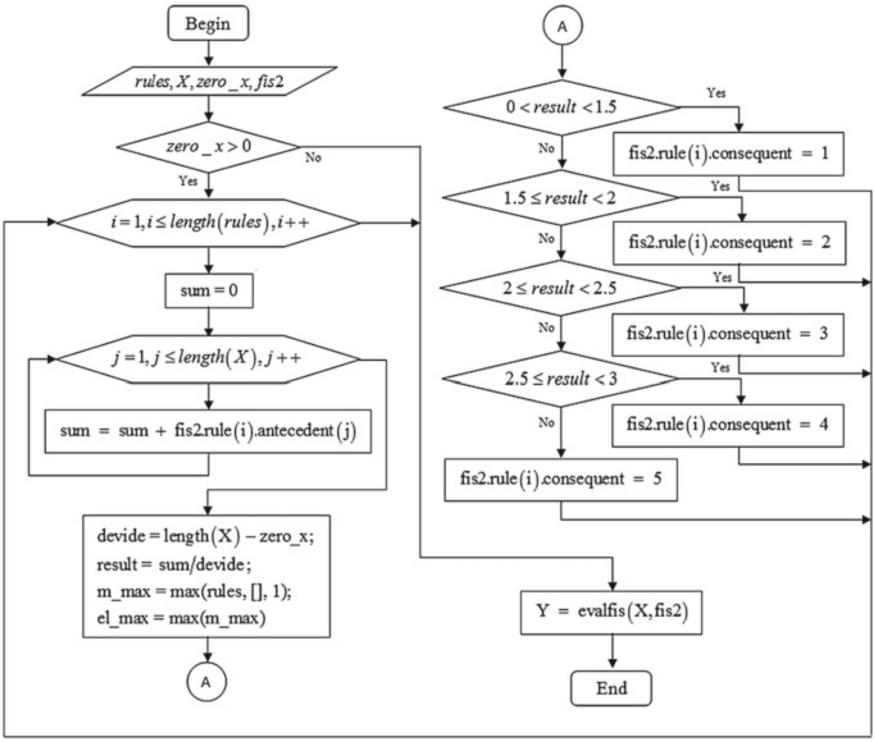


Fig. 4 Block-diagram of the editing algorithm of rules consequents of DSS

Table 2 Modeling Results of the Fuzzy DSS

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
I	30	NI	60	90	1,200	NI	30	40,1
II	50	NI	35	70	5,700	NI	80	65,7
III	90	NI	85	85	10,000	NI	90	77,1

Thus for all relevant sets of input values I–V, RB of second subsystem $y_2 = f_2(y_1, x_4)$ (Table 1) remains the same as with any dimension of input vector X the component y_1 will be formed. Editing of RB (Table 1) should be done only in case when $x_4 = NI$.

8 Modeling Results

Modeling results of fuzzy DSS for different types of sets (I, II, III) of input data $\{x_1, \ldots, x_7\}$ are presented in Table 2.

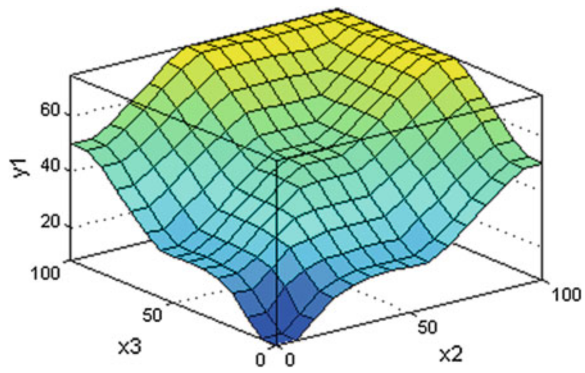


Fig. 5 Characteristic surface of the subsystem $y_1 = f_1(x_1, x_2, x_3)$

In Fig. 5 there is shown characteristic's surface for the rule base of the first subsystem $y_1 = f_1(x_1, x_2, x_3)$ for two components: x_2—level of cargo security protection, x_3—insurance of cargo, when $x_1 = const$.

From the simulation results it is obvious that changing the values of one or a group of input parameters (factors) $\{x_1, \ldots, x_7\}$ of hierarchical fuzzy DSS in different degrees qualitatively and quantitatively impact on the output variable y, which proves the efficiency of the entire system and all its subsystems with the corresponding fuzzy rule bases. In Table 2 there is highlighted one of the relevant sets of input values $x_1 = 30$, $x_2 = NI$, $x_3 = 60$, $x_4 = 90$, $x_5 = 1200$, $x_6 = NI$, $x_7 = 30$, to which will be applied algorithm of editing RB rules. According to this set there will be held a comparative evaluation of the effectiveness of using the algorithm of editing fuzzy DSS rules with flexible hierarchical structure.

Therefore, we can conclude that the indicators value x_2 and x_6 almost does not influence on the result of the system work with application of the editing algorithm of the rules.

9 Conclusions

As a result of the research, authors proposed antecedent/consequent method of current correction of RB rules of fuzzy DSS, which allow increasing of the decision-making effectiveness when solving different problems, including problems of transport logistics.

Synthesized on the basis of the method of actual correction of fuzzy rule bases the intellectual information model of DSS can be widely used for evaluation of the level of quality and efficiency of cargo delivery for variable structure of input factors in the rail, marine and air sectors of cargo transportation. Proposed by authors approach for rule base correction allows to structure and to configure developed DSS to solve different specific problems of transport logistics.

The approach of editing the rule base of fuzzy DSS can be used without limits to the number of DSS inputs and number of linguistic terms for each input and output coordinates.

References

1. Gil-Lafuente, A.M.: Fuzzy Logic in Financial Analysis. Studies in Fuzziness and Soft Computing. Springer, Berlin (2005)
2. Kondratenko, Y.P.: Optimisation problems in marine transportation. Incidencia de las relaciones economicas internacionales en la recuperacion economica del area mediterranea. In: Real Academia de Ciencias Economicas y Financieras, Barcelona, pp. 43–52 (2011)
3. Rotshtein, A.P.: Intellectual Technologies of Identification: Fuzzy Logic, Genetic Algorithms, Neuron Networks. Universum, Vinnitsa (1999). (In Russian)
4. Encheva, S., Kondratenko, Y., Solesvik, M., Tumil, S.: Decision Support Systems in Logistics. In: International Electronic Conference on Computer Science, pp. 254–256 (2007)
5. Mirotin, L.B.: Transport Logistics. Publishing House Examen, Moscow (2005). (In Russian)
6. Kondratenko, Y.P., Sidenko, Ie.V.: Correction of the knowledge database of fuzzy decision support system with variable structure of the input data. In: International Conference MS’12, pp. 56–61, Minsk (2012)
7. Kondratenko, Y.P., Encheva, S.B., Sidenko, Ie.V.: Synthesis of intelligent decision support systems for transport logistic. In: IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 642–646, Prague (2011)
8. Leonenkov, A.V.: Fuzzy Simulation in the Environment MatLab and FuzzyTECH. BHV-Petersburg, Sanct-Petersburg (2005). (In Russian)
9. Piegat, A.: Fuzzy Modeling and Control. Springer, Heidelberg (2001)
10. Gil-Aluja, J., Gil-Lafuente, A.M., Klimova, A.: The optimization of the economic segmentation by means of fuzzy algorithms. J. Comput. Optim. Econ. Finan **1**(3), 169–186 (2011)
11. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
12. Kondratenko, Y.P., Klymenko, L.P., Al Zu’bi, E.Y.M.: Structural optimisation of fuzzy controllers base of intelligent rules. Lect. Model. Simul. **11**(1), 99–108 (2010)
13. Oleynik, A.A., Subbotin, S.A.: The fuzzy rule base reduction based on multiagent approach. In: Herald of the National Technical University KhPI, Kharkov, 2009, no. 43, pp. 126–137 (In Ukrainian)
14. Jantsen, J.: A robustness study of fuzzy control rules. In: Proceedings of the International Conference on EUFIT’97, pp. 1223–1227, Aachen (1997)
15. Sanchez, V.L., Otero, J.: Boosting fuzzy rules in classification problems under single-winner inference. Int. J. Intell. Syst. 1021–1035 (2007)
16. Yager, R., Filev, D.: Essential of Fuzzy Modeling and Control. Wiley, New York (1994)
17. Ching-Chang, W., Nin-Shen, L.: Rule extraction for fuzzy modelling. Fuzzy Sets Syst **2**, 23–30 (1997)
18. Miller, C.A.: The magic number seven plus or minus two: some limits on our capacity for proceeding information. Psychol. Rev. **63**, 81–97 (1956)
19. Kondratenko, Y.P., Sidenko, Ie.V.: Method of actual correction of the knowledge database of fuzzy decision support system with flexible hierarchical structure. J. Comput. Optim. Econ. Finan. **4**(2–3), 57–76 (2012)
20. Fuzzy Logic Toolbox. User’s Guide, Version 2, The MathWorks, Inc. (1999)

21. Shtovba, S.D.: The design of fuzzy systems by means of MatLab. Hot line—Telecom, Moscow (2007) (In Russian)
22. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. In: IEEE Transactions on Systems, Man and Cybernetics. **15**(1), 116–132 (1985)
23. Mamdani, E.H.: Application of fuzzy logic to approximate reasoning using linguistic synthesis. In: IEEE Transactions on Computers. **26**(12), 1182–1191 (1977)

Efficacy of Selected Soft Computing Techniques in Ranking of Sites for Hazardous Industrial Installation

Kalyani Salla, Sanjay Kadam and Ashok Deshpande

Abstract Environmental impact assessment (EIA) is a process of identifying impact and risks that a proposed project (e.g., nuclear power plant installation) may have on the environment. The EIA methods require measurement of specific parameters and variables to estimate the values of impact indicators. However, many parameters and impact indicators in EIA cannot be measured precisely (e.g., lifestyle quality, social acceptance, etc.), and are sometimes very subjective in nature. In order to process this inaccurate and subjective information, we have used soft computing techniques to model the EIA process. In the present study we have implemented two well defined soft computing methods for EIA, namely, Fuzzy Indexing and Artificial Neural Networks (ANNs). The chapter presents a comparative evaluation of these methods with the existing BEES method.

K. Salla (✉)
Department of Computer Science, University of Pune, Maharashtra, India
e-mail: profkalyanis@gmail.com

S. Kadam
Evolutionary Computing and Image Processing Group, Centre for Development of Advanced Computing, Pune, Maharashtra, India
e-mail: sskadam@cdac.in

A. Deshpande
Berkeley Initiative Soft Computing, Special Interest Group, Environmental Management System, Berkeley, CA, USA
e-mail: ashok_deshpande@hotmail.com

A. Deshpande
College of Engineering, Pune, India

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_22,

1 Introduction

Environmental impact assessment is a process of identifying impact and risks that a proposed project may have on the environment. The project could be an installation of a thermal, nuclear or hydro electric power plant, or setting up of a hazardous industrial unit at some site. The EIA methods require measurement of specific parameters and variables to estimate the values of impact factors or indicators [6]. These parameters and indicators, for example, may represent willingness for resettlement, response from public for an upcoming power plant, accessibility to community resources, response for public involvement programs, or impact on human health and economy.

However, many of these environment impact indicators and governing parameters cannot be measured precisely (e.g., lifestyle quality, social acceptance, etc.) and are sometimes very subjective. Also, emotional feelings and values cannot be measured quantitatively, but are important in expressing knowledge or information in EIA. Additionally, ways to express information and values held by different groups also varies, which further increases the subjectivity. Because of the subjective nature of the parameters and indicators in EIA, we believe that soft computing; specifically fuzzy logic [7], could be an effective approach to model the imprecise and subjective environmental conditions and impacts.

The EIA process has currently attracted significant research interests. Presently, addressing issues related to hazardous industrial installations has assumed vital importance across the globe. For example, the issues of rehabilitation, provision of safe drinking water, providing good infrastructural facilities, and studying the impact on human health are now being given serious consideration in the EIA process. The imprecision in EIA encourages the use of experts to determine the scope and significance of a project [10], who reach agreement on what they believe is important without consulting the investors. In a typical EIA, there is a need to maintain transparency with the general public and consider them as one of the stakeholders in the decision making process. Often, promises made are seldom fulfilled, which results into public outcry and anger.

For an upcoming project, the challenge in EIA today is to identify impacts, predict future environmental conditions, evaluate role of environmental parameters and address societal values [10]. To make the EIA process more realistic, scoping is the first step that is carried out [10]. It aims at including only the relevant parameters that accurately depict the existing environmental conditions. The subjectivity aspect in the EIA can lead to disagreement amongst the decision makers. It may result in delay of the evaluation process and increase the EIA cost without resolving the underlying concerns of all the parties. Since societal values and ecological attributes cannot be measured objectively, the final decision statement may fail to depict the actual conditions.

The appropriate choice of an Industrial site is an important and at times an irreversible decision. A poor choice of site may lead to overheads such as increased cost of transportation, lack of skilled labor or degradation of existing environmental conditions. On the other hand, the application of reliable and accurate methods for site selection may lead to efficient utilization of resources, reduced transportation cost, and satisfied stakeholders with minimal ill effects caused to the existing environment.

We believe that research into newer computing methods and techniques could alleviate some of the above-mentioned problems in EIA. Although, fuzzy logic has been used in the past to model the EIA process [2, 4, 13, 14, 15], it has not been applied to rank industrial sites. In our opinion, the use of soft computing methods would reduce the subjectivity aspect in EIA and also address the imprecise nature of the governing parameters and impact indicators. In the present research work, we have implemented two well defined soft computing methods for EIA, namely, Fuzzy Indexing and Artificial Neural Networks.

Fuzzy Indexing has been used to rank the potential industrial sites. This has been done by calculating the product of the membership grade based on fuzzy sets for each parameter and the value of the corresponding triplets (that gives index for Air, Water, Land, Socioeconomic and Ecological attributes). The overall score has been computed as extended product of fuzzy numbers [1].

The second approach explored in this chapter is the use of artificial neural networks, specifically, the linear vector quantization (LVQ) algorithm to map the input patterns to the corresponding output patterns (classes) [10]. Each input pattern consists of fuzzy values for the 38 attributes or parameters (such as Air Quality, Water Quality, Land Availability, Socioeconomic factor, and Ecological factor). The corresponding output pattern consists of class labels (such as ‘Poor’, ‘Fair’, ‘Good’ and ‘V. Good’) representing suitability of the site. The LVQ uses the supervised version of the Kohonen learning rule to adjust the connection weights during the training process.

The chapter is organized as follows: Sect. 2 describes the Methodology, where an overview of the methods used for ranking the industrial sites is discussed, Sect. 3 presents a Case study, Sect. 4 presents Results and Discussion, and Sect. 5 presents Conclusions and Future work.

2 Methodology

The three methods for ranking sites for setting up hazardous industrial power plants have been briefly discussed.

2.1 Fuzzy Indexing

The Fuzzy indexing method uses the Triangular membership function distribution comprising a four point scale pattern, i.e. Poor, Fair, Good and V. Good. The first number in the Triplet denotes the value at which the membership function is 1.

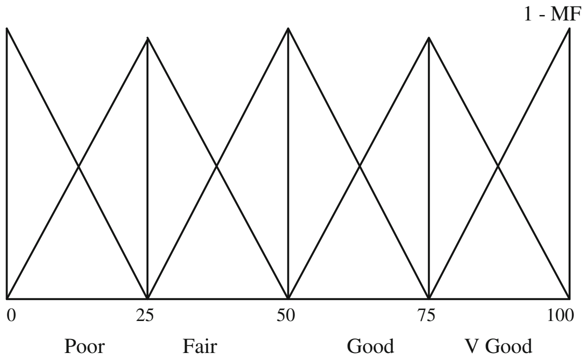


Fig. 1 Triangular membership function distribution pattern [1]

The second and the third number in the triplet designate the distance to the left and right, respectively, of first number where the membership value is zero.

The values of the Triplets are as follows: Poor [0 0 25], Fair [0 25 50], Good [25 50 75], and Very Good [75 100 100]. The '1' in Fig. 1 is the value of Membership function indicating full membership in the respective fuzzy set.

Following are the steps to compute the rank of a Site based on different attributes. We give expressions only for the 'Air attribute'. The expressions for the other attributes are similar.

Step 1: The Site Index for each major attribute is calculated as the product of membership grades in Table 2 with the corresponding values of Triplets. For Air, the SIA (Site Index Air) is given as follows [7]

$$\mathrm{SIA} = \Sigma \mu_A * \mathrm{T} \tag{1}$$

where, l_A is a vector of membership grades of the attribute 'Air' and 'T' indicates the value of Triplets.

Step 2: Calculate the preference or importance of each major attribute. The preference of importance QA for 'Air' attribute is computed as follows:

$$\mathrm{QA} = \Sigma \mathrm{R} * \mathrm{T} \tag{2}$$

where, 'R' represents a vector of relative weightages as shown in Table 3 and 'T' is the value of Triplets.

Step 3: Normalize the triplets of preference (for each attribute) by dividing each by Qsum. The Qsum is sum of the first components of QA (Air), QW (Water), QL (Land), QS (Socioeconomic), and QE (Ecological). For the 'Air' attribute, the normalized value QArel is given by

$$\mathrm{QArel} = \mathrm{QA}/\mathrm{Qsum} \tag{3}$$

Step 4: The overall score of a Site SO1 is given as

$$\begin{aligned} \text{SO1} = {} & \text{SIA} * \text{QArel} + \text{SIW} * \text{QWrel} + \text{SIL} * \text{QLrel} \\ & + \text{SIS} * \text{QSrel} + \text{SIE} * \text{QErel} \end{aligned} \tag{4}$$

Here, SIA, SIW, SIL, SIS, SIE are the site Index values for major attributes Air, Water, Land, Socioeconomic and Ecological and QArel, QWrel, QLrel, QSrel and QErel represent the relative weightages of the criteria or the relative weighted Triplets as calculated in Step 3.

The extended sum and product is calculated as

$$(\text{a b c}) * (\text{d e f}) = (\text{ad ae} + \text{db af} + \text{dc}) \tag{5}$$

where, * represents extended product of two fuzzy numbers, and + represents the extended sum of two fuzzy numbers.

The product of membership grade calculated based on fuzzy sets for each parameter and value of Triplets gives the Index for Air, Water, Land, Socioeconomic and Ecological attributes. In a similar way, Triplets for individual preference to the importance of each criterion can be created. The overall score can be computed as extended product of the fuzzy numbers.

Step 5: The overall score X is calculated by considering horizontal distance of centroid of a triangle from its origin. The overall quality which is expressed by the Triplet (a b c) can be represented by a triangle ABC

The first digit of the Triplet gives an idea about the quality or suitability of the site. The higher the value, the better is the suitability. The triangles ABD and BDC are right angled triangles and the centroid is located at a distance 1/3 from the base. The horizontal distance of the centroid from the origin (0) is calculated as follows (that gives the overall score X):

$$\text{X} = \text{a} - ((\text{b} - \text{c})/3) \tag{6}$$

2.2 Artificial Neural Networks

The ANN used in our study is the linear vector quantization (LVQ) neural network model, which is a pattern classifier where each output unit represents a particular class. The weight vector (called as reference vector) of an output unit represents a particular class. During the training process, the output units are positioned by adjusting their weights through supervised learning. A supervised version of Kohonen learning rule is applied in the LVQ ANN.

The LVQ ANN consists of a single layer (competitive layer) of neurons, where each neuron is fed with the values of the input vector. The software simulation of a LVQ network may involve introduction of an input layer to feed the inputs to the competitive layer and an output layer to get the user-defined output classes. The classes learned by the competitive layer are referred to as subclasses and the classes of the output layer as target classes [4]. The training of the linear vector quantization ANN performed on the competitive layer of neurons is as follows:

Step 1: Initialize the weights (reference) vectors and the learning rate α.
Step 2: Repeat Steps 3 to 7, until a stopping criteria is met.
Step 3: For each input pattern $\mathbf{X} = (x_1, x_2, \ldots, x_n)$, repeat Steps 4 to 5.
Step 4: Compute D(j), the square of the Euclidean distance

$$D(j) = \sum \left(w_{ij} - x_i\right)^2$$

Find the 'j'th neuron where D(j) is the minimum.

Step 5: Update weight vector w_j of the jth neuron as follows
If $(T = c_j)$, then

$$w_j(\text{new}) = w_j(\text{old}) + \alpha\left[x - w_j(\text{old})\right]$$

(i.e., move weight vector w_j towards input vector $\mathbf{X}$)
If $(T\ != c_j)$ then

$$w_j(\text{new}) = w_j(\text{old}) - \alpha\left[x - w_j(\text{old})\right]$$

(i.e., move w_j away from $\mathbf{X}$)
where, c_j denotes pre-assigned class represented by the jth neuron, and T is the correct class for input $\mathbf{X}$.

Step 6: Reduce the learning rate α.
Step 7: Stopping condition: This could be a fixed number of iterations or learning rate reaching some small value.

3 Case Study

In the case study we use the above-mentioned methods to rank different sites for setting up hazardous industrial power plants, based on various attributes. The candidate sites considered in this research work are located in the states of Arunachal Pradesh, Himachal Pradesh and Sikkim of India. The major attributes which have been used in the study include Air, Water, Land, Socioeconomic and Ecological. These attributes in turn comprise sub-attributes. For example, Air has

Table 1 Numeric data for 17 sub-attributes for three sites

Attributes	Site 1—Arunachal Pradesh	Site 2—Himachal Pradesh	Site 3—Sikkim
Air			
CO (in mg/m^3)	6	5	3
Lead (in $\mu g/m^3$)	1.50	1	1.0
NO_2 (in $\mu g/m^3$)	24	80	9.0
Ozone (in $\mu g/m^3$)	80	60	90
PP (in $\mu g/m^3$)	40	35	20
Sox (in $\mu g/m^3$)	50	98	6.8
PM_{10} (in $\mu g/m^3$)	56	140	79.0
$PM_{2.5}$ (in $\mu g/m^3$)	21	60	19.0
Ammonia air (in mg/m^3)	60	60	80
Water			
Ph	7.0	7.3	7.19
Total coliform (MPN/100 ml)	510	250	150
DO (mg/l)	5	9.6	8.81
BOD (mg/l)	3	9	3
Ammonia W (in mg/l)	1.0	0.01	0.6
Electrical conductivity (μmhos/cm)	2350	3700	44.333
Sodium absorption Ratio	26	25	20
Boron (mg/l)	1	1.36	0.61

sub-attributes such as CO, Lead, and Ozone, and Particle pollution, which depict air quality.

Amongst these attributes, some (e.g., Air and Water) have numeric values for their sub-attributes, while others (e.g., Land, Socioeconomic and Ecological) have linguistic descriptions for their corresponding sub-attributes. Table 1 depicts numeric values (as reported in [9, 8, 12]) for different sub-attributes for the three sites. There are 17 such sub-attributes.

Besides these, we also consider 21 descriptive or linguistic sub-attributes (for ranking of the sites) such as availability of land, proximity to water resource, feasibility of transportation, and availability of infrastructure. The values for these descriptive sub-attributes have been taken from the concerned reports [9, 8, 12].

The next step is to assign fuzzy values to these 38 inputs or sub-attributes (17 numeric and 21 descriptive) based on the fuzzy membership grades assigned to their corresponding major attributes (i.e., Air, Water, Land, Socioeconomic and Ecological). Table 2 depicts the fuzzy membership grades assigned to the major attributes for different sites. These grades have been assigned by the environmental experts in reference to the guidelines specified by the Central Pollution Control Board, Ministry of Environment and Forest, Government of India. The median of all the values was calculated to arrive at the values specified in Table 2.

We also need to specify the preference of importance for each of the major attributes. We assign membership grades for each such attribute. Table 3 depicts

Table 2 Membership grades for major attributes

Criteria	Poor	Fair	Good	V Good
Air				
Site 1—(Arunachal)	0.6	0.3	0	0
Site 2—(Himachal)	0.3	0.4	0	0
Site 3—(Sikkim)	0	0	0.3	0.6
Water				
Site 1—(Arunachal)	0	0.5	0.3	0
Site 2—(Himachal)	0	0.2	0.2	0.4
Site 3—(Sikkim)	0	0	0.3	0.5
Land				
Site 1—(Arunachal)	1	0	0	0
Site 2—(Himachal)	0	1	0	0
Site 3—(Sikkim)	0	0	1	0
Socioeconomic				
Site 1—(Arunachal)	1	0	0	0
Site 2—(Himachal)	0	1	0	0
Site 3—(Sikkim)	0	0	0	1
Ecological				
Site 1—(Arunachal)	1	0	0	0
Site 2—(Himachal)	0	0	1	0
Site 3—(Sikkim)	0	0	0	1

Table 3 Preference or importance given for criteria

Attributes	Poor	Fair	Good	V. Good
Air	0	0	0.4	0.6
Water	0	0	0.4	0.6
Land	0	0	0.4	0.6
Socioeconomic	0	0	0.4	0.6
Ecological	0	0	0.4	0.6

the four membership grades assigned to each of the five major attributes. Since we would like to give importance to each of these attributes in decision making, we assign a membership grade of 0.6 for 'V.Good' and 0.4 for 'Good'.

The triangular membership function is represented by a set of three numbers called as 'Triplets'. The first number of 'Triplet' denotes the value at which the membership function is 1. The second and the third number of the triplet designate the distance to the left and right, respectively, from the first number, where the value of membership function is 0. Table 4 depicts the values of these triplets.

We now discuss the application of Fuzzy indexing used for ranking of the sites. The following steps describe the detailed calculations in accordance to the method discussed in Sect. 2.1. We describe calculation performed for each of the three sites.

Table 4 Values of the triplets

Attributes	Number where MF = 1		
Poor	0	0	25
Fair	0	25	50
Good	25	50	75
V. Good	75	100	100

3.1 Fuzzy Indexing Calculations for Site 1

We present a detailed description to compute the rank of Site 1 (Arunachal Pradesh). The calculations for the other sites can be performed by similar method.

Step 1: The site index for Air quality (SIA) is calculated as the sum of the products of the membership grades (Table 2) with the values of the triplets (Table 4). It is given as (Eq. 1):

$$\mathrm{SIA} = [91.6666 \quad 25.0000 \quad 8.3333].$$

Similarly, we can calculate the site index for other major attributes. The calculated values for the major attributes are as follows: SIA (Air) = [91.6666 25.0000 8.3333], SIW (Water) = [90.63 25.0 9.375], SIL (Land) = [25.0 25.0 25.0], SIS (Socioeconomic) = [100.0 25.0 0.0], and SIE (Ecological) = [100.0 25.0 0.0].

Step 2: We next calculate the preference or importance of each major attribute. The detailed calculations (Eq. 2) for 'Air' attribute are as given as follows (Refer Tables 3 and 4)

$$\mathrm{QA} = (90.0 \quad 25.0 \quad 10.0)$$

The preference of importance QW (Water), QL (Land), QS (Socioeconomic) and QE (Ecological) for other attributes can similarly be calculated using the above method.

Step 3: We then take sum of the first value of all the triplets obtained in Step 2 and denote it by *Qsum*. We normalize the triplets of preference (for each attribute) by dividing each by Qsum (Eq. 3). For example, the calculations for the 'Air' attribute (represented by QArel), with a Qsum = 450, are depicted as follows:

$$\begin{aligned}\mathrm{QArel} &= (90.0 \quad 25.0 \quad 10.0)/450\\ &= (0.2000 \quad 0.0555 \quad 0.0222)\end{aligned}$$

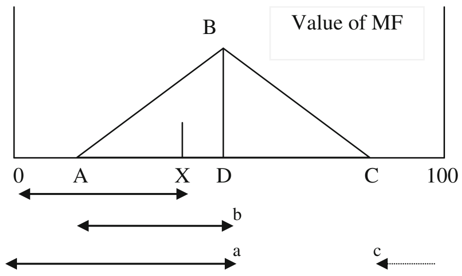


Fig. 2 Overall score as triangle ABC and Triplet abc

Similarly, we calculate triplets for other major attributes. Their values are as follows QWrel (Water) = (0.2000 0.0555 0.0222), QLrel (Land) = (0.2000 0.0555 0.0222), QSrel (Socioeconomic) = (0.2000 0.0555 0.0222), and QErel (Ecological) = (0.2000 0.0555 0.0222).

Step 4: Therefore, the overall score SO1 for Site 1 can be computed as extended product of the fuzzy numbers as follows (Eq. 4):

$$\begin{aligned} \mathrm{SO1} &= \mathrm{SIA} * \mathrm{QArel} + \mathrm{SIW} * \mathrm{QWrel} + \mathrm{SIL} * \mathrm{QLrel} \\ &\quad + \mathrm{SIS} * \mathrm{QSrel} + \mathrm{SIE} * \mathrm{QErel} \\ &= [33.5416 \;\; 15.9837 \;\; 28.7268] \end{aligned}$$

Step 5: In reference to Fig. 2 and the procedure to compute the centroid (for defuzzified value) as explained in Sect. 2.1, we calculate the defuzzified value XS1 for Site 1 as follows (Eq. 5):

$$\begin{aligned} \mathrm{XS1} &= 33.5416 - ((15.9837 - 28.7268)/3) \\ &= 33.5416 - (-12.7431)/3) \\ &= 33.5416 - (-4.2477) \\ &= 37.7893 \end{aligned}$$

3.1.1 Fuzzy Indexing Calculations for Other Sites

Using similar method as described for Site 1, we can compute the extended products and defuzzified values for other sites. The results for Site 2 (SO2 and XS2) and Site 3 (SO3 and XS3), respectively, are depicted below:

$$\begin{aligned}
SO2 &= [54.1071 \quad 37.8869 \quad 33.5119] \\
XS2 &= 54.1071 - ((37.8869 - 33.5119)/3) \\
&= 54.1071 - (1.4583) \\
&= 52.6488
\end{aligned}$$

$$\begin{aligned}
SO3 &= [81.4583 \quad 47.6273 \quad 22.5925] \\
XS3 &= 81.4583 - ((47.6273 - 22.5925)/3) \\
&= 81.4583 - (8.3449) \\
&= 73.11
\end{aligned}$$

Finally, the ranking of sites can be done in the descending order of XSi values ($i = 1, 2, 3$). The higher the value of XSi, more is the suitability or rank of the site for setting up of hazardous industrial power plants. In the present case study, the individual sites had the following ranking (in descending order of their ranks):

Site 3 (Sikkim)	Score 73.11	Rank 1
Site 2 (Himachal Pradesh)	Score 52.6488	Rank 2
Site 1 (Arunachal Pradesh)	Score 37.7893	Rank 3

The ranking of the above sites for setting up of hazardous industrial power plants has also been done with the existing Battelle Environmental System (BEES). The ranks computed by the Fuzzy Indexing method and those computed by the standard Battelle Environmental System concur with each other. In the Fuzzy Indexing method many of the non-numeric attributes were also considered to enhance the accuracy of the method in comparison to the existing methods.

3.2 LVQ Neural Network Model for Ranking of Sites

The learning vector quantization neural network model used in our case study has 38-10-4 architecture. That is, it consists of three layers: An input layer consisting of 38 neurons, a hidden layer comprising 10 neurons, and an output layer with 4 neurons. The 38 neurons in the input layer represent various sub-attributes of the major attributes (Air, Water, Land, Socioeconomic and Ecology). The 10 neurons in the hidden layer represent the subclasses, while the 4 neurons in the output layer represent the main output classes. The output classes considered in our case study were: 1—‘V. Good’, 2—‘Good’, 3—‘Fair’ and 4—‘Poor’. The weight vector of each neuron in the output layer represents reference vector of some main output class.

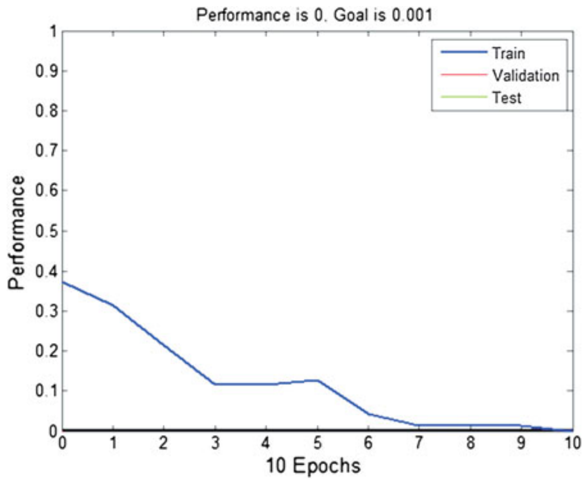


Fig. 3 Training of LVQ ANN with 35 patterns

The examples of some sub-attributes considered in the study include amount of CO, Lead, NO_2, Ozone, particle pollution, SO_2, SPM, RPM, Ammonia content in Air. Others include electrical conductivity, Sodium absorption, Boron content, resettlement, suitable infrastructure, proximity to water source, unloading space, transport facility, land away from sanctuaries, thick population, seacoast regulations, effects on human health, economy, legal rights, access to community services, household facilities, public involvement programs, response from public, adequate sunlight, latitude, altitude and chances of fire.

A total of number 50 patterns were used in the LVQ simulations, out of which 70 % (35 patterns) were used as training data and the remaining 30 % (15 patterns) were used as test data. We have used the MATLAB software for our simulations [4]. The patterns were created by referring to various EIA reports on setting up hazardous industrial power plants at the different sites across the nation. Each pattern comprises values for the above-mentioned 38 sub-attributes. The output class for each pattern has been assigned by referring to the guidelines provided by the Central Pollution Control Board, Ministry of Environment and Forest, Government of India. For some patterns, expert opinion was also considered to decide the output class of these patterns.

The ANN was trained with the following attributes/parameter values (in bracket): Learning (Supervised), Learning Rate (0.01), Learning function (Learning Vector Quantization 1), Network performance or Mean squared error (0.01), Distance function (Euclidean), Transfer function in the hidden layer (compet), and Transfer function in the output layer (purelin).

Before training, the input patterns were normalized in each of the 38 sub-attributes so that their values ranged in the interval [0, 1]. Figure 3 illustrates the convergence of the network during the training process. The error in the network predictions and the observed values gradually decreases with execution of epochs; where one epoch represents a training cycle that feeds each pattern in the training

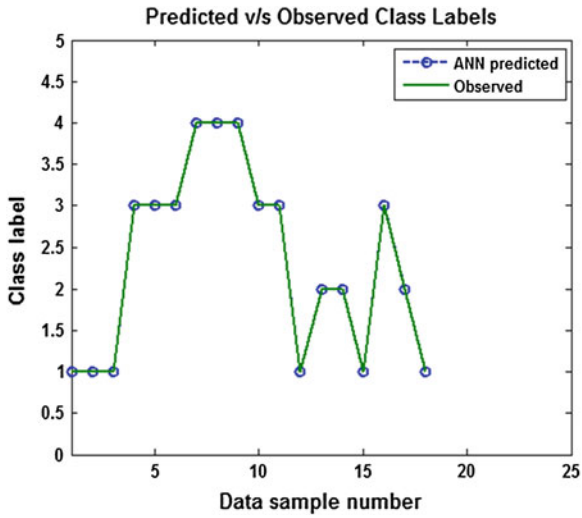


Fig. 4 Predicted v/s Observed class labels

set to the LVQ network. The network required 10 epochs (10 training cycles, each comprising 35 patterns) for convergence as shown in Fig. 3.

The trained LVQ network was then used to predict the classes for the 15 test samples. The trained LVQ network predicted the classes for these 15 samples accurately (with 100 % accuracy). The trained LVQ network was then used to predict the classes (representing the ranks) for the three sites discussed in this chapter. Figure 4 displays the prediction results of trained LVQ network for the 15 test samples and the 3 samples for the three sites (last three data samples). From Fig. 4 it can be seen that the network gave correct class labels for all the 18 unseen patterns (15 test patterns and 3 site patterns).

4 Results and Discussion

Table 5 depicts the results of the various methods used to rank the sites for setting up of hazardous industrial power plants at different locations. The existing method for analyzing impacts of an upcoming Plant is the Battelle Environmental Evaluation System also called BEES [3]. The BEES method computes the positive and negative score with reference to all the 38 attributes considered in the study.

From Table 5 it can be observed that the ranks computed by the Fuzzy Indexing method, the LVQ neural network model and those computed by the standard Battelle Environmental Evaluation System match exactly with each other. The accuracy of the Fuzzy indexing method could be attributed to our consideration of the partial memberships of each of the sub-attributes, which represent the actual situation. The LVQ ANN was trained and tested with crisp normalized data, which accurately predicted the ranks of the three sites discussed in the chapter.

Table 5 Ranking of sites by three different methods

Sites	Ranking method		
	BEES[a]	Fuzzy indexing	LVQ ANN
Site 1 (Arunachal Pradesh)	III	III	III
Site 2 (Himachal Pradesh)	II	II	II
Site 3 (Sikkim)	I	I	I

[a] Existing standard Battelle environmental evaluation system method

An illustrative example in the case study elaborates how linguistic parameters like availability of land with least resettlement, adequate unloading space, suitable infrastructure, land suitability with geological aspects, etc. have been considered to assign the membership grade for land suitability. Similarly, the Socioeconomic attributes like consequences to economy, accessibility to community services, willingness for resettlement, and response to public involvement programs have been considered to assign membership grades for the Socioeconomic aspect described in Table 2.

Environmental experts were consulted and a median of the values assigned by these experts has been considered in our present work. A total of seventeen numeric and twenty one non-numeric attributes have helped to accurately portray the existing environmental conditions by understanding the quantitative description provided in the EIA reports. The assignment of fuzzy membership grades and the value or weight age of different components with respect to Air, Water, Land, Socioeconomic, and Ecological quality have helped in comparing the existing alternative sites objectively and precisely by including the most important parameters in the assessment process. This distinguishes our approach with the previous attempts discussed in this chapter.

The results can be further verified by comparison with results obtained with more number of methods. The methods were tested for three individual sites located in Sikkim, Himachal Pradesh and Arunachal Pradesh for upcoming hazardous industrial power plants (like hydro electric power plant) as part of the case study. The results obtained were satisfactory and better than the existing methods. To the best of our knowledge, this is the first attempt to apply fuzzy indexing for the positive and negative impact score of attributes that are considered to decide the suitability of upcoming hazardous industrial plants. The computer software for the above methods has been developed in JAVA, which was used to verify the results. The results obtained by the existing standard method (i.e., BEES) and the ones generated by the software were compared to calculate the accuracy of the methods.

5 Conclusions and Future Work

The chapter presented two methods, namely, Fuzzy Indexing and LVQ ANN for ranking sites for setting up of hazardous industrial power plants. The ranks computed by these methods and those computed by the standard Battelle Environmental System concur with each other. In the Fuzzy Indexing method many of the non-numeric attributes were also considered to enhance the accuracy of the method in comparison to existing methods. In Future work, more number of Fuzzy ranking methods would be applied and the results obtained from all the methods would be compared. The performance evaluation of the methods in terms of speed would also be done. An attempt would also be made to extend the pure ANN approach with a Neuro-Fuzzy model, which would embed the linguistic description of parameters, and partial membership values in fuzzy sets (i.e., 'V. Good', 'Good', 'Fair' and 'Poor').

Acknowledgement We would like to express sincere gratitude to Dr. Mrs. S. S. Tikle, Environmental Science Department, University of Pune, India and other Experts recommended by her for their valuable opinion in assessing Environmental parameters and Reports.

References

1. Das, H.: Food Processing Operation Analysis, pp. 383–394. Asian Books Private Limited, Chennai (2005)
2. KaiNing, Y.U., BingXu, L.V.: Research of the environmental impact assessment methods caused by foundation pit dewatering. In: 2011 International Symposium on Water resource and environmental protection (ISWREP), vol. 4, pp. 2777–2780, May 2011
3. Lohani, B.N., Evans, J.W., Everitt, R.R., Ludwig, H., Richard, A.C., Tu, S.L.: Environmental Impact Assessment for developing countries in Asia, Asian Development Bank (1997)
4. MathWorks Matlab Help, Learning Vector Quantization Networks (2011)
5. Mohamed, M.Y.: Artificial neural network model to assess the impacts of land development on the river flow. The 2nd international conference on water resources and arid environment, King Saud University, Riyadh, Saudi Arabia (2006)
6. Omrani, H., Ion-Boussier, L., Trigano, P.: An approach for environmental impact assessment. In: Proceedings of the 5th International Conference on Computational Intelligence, Man Machine Systems and Cybernetics, Italy (2006)
7. Ross, T.J.: Fuzzy Logic with Engineering Application. McGraw Hill, New Mexico (1995)
8. R.S Envirolink Technologies, TT Energy. Environmental Impact Assessment of H.E Project (2010)
9. R.S Envirolink Technologies, Hydro Power Plant. Environmental Impact Assessment Report (2010)
10. Shepherd, R.B.: Quantifying Environmental Impact Assessments using Fuzzy Logic. Springer, New York (2005)
11. Sivanandam, S.N., Sumathi, S., Deepa, S.N.: Introduction to Neural Networks using Matlab. Tata Mc Graw Hill, Delhi (2006)
12. Smec India. Environmental Impact Assessment and Environmental Management Plan for NAFRA Hydro Electric Power Project (2009)
13. Tayebi, M.H., Tangestani, M.H., Roosta, H. Environment impact assessment using neural network model: a case study of the Jahani, Konarsiah and Kohe Gach salt plugs, SE Shiraz, Iran. In: Wagner, W., Székely, B. (eds.) ISPRS TC VII Symposium—100 Years ISPRS, vol. 35, part 7B. IAPRS, Vienna, Austria, 5–7 July 2010
14. Teresa, M.V, Manuel J.F., Carlos L.G.: Application of fuzzy logic to qualify the environmental impact in abandoned mining sites, Water Air Soil Pollut. **217**, 303–315 (2011). doi: 10.1007/s11270-010-0587-6
15. Varshosaz, K.: Application of fuzzy logic in environmental impact assessment modeling of a man-made lake in Western Tehran (Iran). The 2nd international conference on environmental science and technology IPCBEE, vol.6, IACSIT Press, Singapore (2011)

Fuzzy Resolution with Similarity-Based Reasoning

Banibrata Mondal and Swapan Raha

Abstract Resolution is an useful tool for mechanical theorem proving. Resolution models the refutation proof procedure, which is mostly used in constructing a 'proof' of a 'theorem'. In this chapter, an attempt is made to derive a fuzzy resolvent from imprecise information expressed as standard rule using similarity based inverse approximate reasoning methodology. For complex clauses, we investigate similarity based ordinary approximate reasoning to derive a fuzzy resolvent. The proposal is well-illustrated with artificial examples and a real life problem.

1 Introduction

Man had for long wanted to find a general decision procedure to prove theorems. Since 1930, consistent efforts were made by Herbrand and others in this respect. It was in 1965 that Robinson [1] made a major breakthrough. He introduced a

This research has been partially supported by the UGC SAP (DRS) Phase-II project under the Department of Mathematics, Visva-Bharati and UGC major research project no. 36-293/2008 (SR);26.03.09.

B. Mondal (✉) · S. Raha
Department of Mathematics, Visva Bharati, Santiniketan, West Bengal, India
e-mail: mbanibrata@gmail.com

S. Raha
e-mail: swapan.raha@visva-bharati.ac.in

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_23,

machine-oriented logic based on resolution principle to check the (un)satisfiability of a set of formulae represented in some standard form.

Mechanical theorem proving techniques were first applied to deductive question answering systems and then to problem solving, program synthesis, program analysis and many others. In automated theorem proving, resolution acts as a rule of inference leading to a refutation based theorem-proving technique.

Now, given some formulae modelling vague knowledge, can we define an automated procedure which decides whether a given conclusion is logically entailed by the knowledge? A first step towards formalization of automated deduction in fuzzy logic was taken by Lee and Chang [2]. Lee's works [2, 3] were continued and implemented by other researchers. Lee's fuzzy formulae are syntactically defined as classical first-order formulae, but they differ semantically as the formulae have a truth value in [0,1]. Mukaidono [4, 5] generalized Lee's [3] result. Dubois and Prade [6] established fuzzy resolution principle in the case of uncertain propositions. In Kim et al. [7], antonym-based fuzzy hyper-resolution was introduced and its completeness was proved. Fontana and Formato [8] introduced, a fuzzy resolution rule based on an extended most general unifier provided by the extended unification algorithm. Sessa [9] proposed a methodology to manage uncertain and imprecise information in the frame of the declarative paradigm of logic programming considering, a similarity relation between function and predicate symbols in the language of a logic program. Habiballa and Novak [10] presented refutational theorem proving system for fuzzy description logic based on the general resolution rule. Raha and Ray [11] investigated a generalized resolution principle that handles the inexact situation effectively and is applicable for both well-defined and undefined propositions. They associated a truth value to every proposition and used Zadeh's concept of approximate reasoning based on possibility theory to model a deductive process. In this research, we extend our knowledge to present, a generalised resolution principle that deals with fuzzy propositions and uses the technique of similarity based reasoning. Similarity is inherent in approximate reasoning, but none used it in fuzzy resolution. Indeed, human reasoning is often performed on the basis of analogy and/or similarity between entities. We, therefore, accomplish the task of choosing an almost complementary pair in fuzzy resolution on the basis of similarity/dissimilarity measure of fuzzy sets instead of contradictory degree of the pair. The advantage of our method is that, it reduces significant and important drawback of compositional rule of inference and executes effective resolution to show its flexibility for automated reasoning. In this article, we have avoided the generic problem in generalised modus ponens (GMP) by using similarity-based inverse approximate reasoning method [12] in fuzzy resolution. However, for complex set of clauses the method is not suitable. In this case, we investigate similarity-based ordinary approximate reasoning [13] to deduce fuzzy resolution.

2 Preliminaries

We start by summarizing a few basic concepts of fuzzy logic connected with fuzzy resolution.

Definition 1 A resolvent of two clauses C_1, C_2 containing the complementary literals p and $\neg p$ respectively, is defined as

$$res(C_1, C_2) = \{C_1 - \{p\}\} \cup \{C2 - \{\neg p\}\},$$

U is understood as the disjunction of the literals present in them.

It is also a logical consequence of C_1 and C_2.

Definition 2 A resolution deduction of a clause C from a set S of clauses is a finite sequence of clauses $C_1, C_2, \ldots, C_n = C$ such that, each C_i is either a member of S or is a resolvent of two clauses taken from S. From the resolution principle in propositional logic we deduce that, if S is true under some truth valuation v, then $v(C_i) = TRUE$ for all i, and in particular, $v(C) = TRUE$ [14].

A simple resolution scheme is:

$$\begin{array}{r} premise1 : \mathrm{a} \vee \mathrm{b} \\ premise2 : \quad \neg \mathrm{b} \\ \hline conclusion : a \end{array}$$

In first order logic, resolution condenses the traditional syllogism of logical inference down to single rule. To recast the logical inference using the resolution technique, the formulae are represented in conjunctive normal form.

Definition 3 C_1 and C_2 being two clauses in fuzzy logic, $R(C_1, C_2)$ being a classical resolvent of C_1 and C_2, and A being the literal C_1 and C_2 are resolved upon (i.e., $C_1 \in A$ and $C_2 \in \sim A$ or conversely), then fuzzy resolvent $R(C_1, C_2)_{cd(A)}$ of C_1 and C_2, where $cd(A)$ is the contradictory degree of keyword A or the confidence associated with the resolvent, is defined by

$$R(C_1, C_2)_{cd(A)} \triangleq R(C_1, C_2) \ \mathrm{V}(A \wedge \ \sim A)$$

An important theorem which was proved in [4, 5] is given in the following [15]:

Theorem 1 A set S of fuzzy clauses is unsatisfiable if and only if, there is a deduction of empty clause with its confidence of resolvent $cd \neq 0$ from S.

Fuzzy Implication:

Typically, a fuzzy rule 'If X is A then Y is B' (A and B are fuzzy sets) is expressed as $I(a, b)$, where I is a fuzzy implication and a and b are membership grades of A and B respectively. From an algebraic point of view, some implication operators basically identified in [16] are classified with four families I_i^T, $i \in \{1, 2, 3, 4\}$, where T and T^* denote the specific t-norm and correlated t-conorm respectively

Table 1 T-Norms and T-Conorms

T/T *	T(a, b)	T*(a, b)
M	min(*a*, b)	max(*a*, b)
P	a.b	a + b − a.b
B	max(0, a + b − 1)	min(1, a + b)

Table 2 Expression of fuzzy implication operators

I_i^T	M	P	B
I_1^T	max(min(a, b), 1 − a)	1 − a + a^2.b	max(1 − a, b)
I_2^T	max(1 − a, b)	1 − a + a.b	min(1 − a + b, 1)
I_3^T	1 if $a \le b$ b otherwise	1 if $a \le b$ b/a otherwise	min(1 − a + b, 1)
I_4^T	1 if $a \le b$ $1 - a$ otherwise	1 if $a \le b$ $\frac{1-a}{1-b}$ otherwise	min(1 − a + b, 1)

(see, Table 1) to implement the implication operators. Here, we use only two families I_2 and I_3 of implication operators. Family I_2, often named S-implication, derives from classical logic form $a \rightarrow b = \neg a \vee b$. Family I_3, known as R-implication, reflects a partial ordering on propositions. With reference to the t-norms and t-conorms in Table 1, the explicit expressions of fuzzy implication operators I_i^T are presented in Table 2.

Similarity Index:

The notion of similarity plays a fundamental role in theories of knowledge and behaviour and has been dealt with extensively in psychology and philosophy. If we study the behaviour pattern of children we find that, children have a natural sense in recognizing regularities in the world and to mimic the behavior of competent members of their community. Children thus make decisions on similarity matching. The similarity between two objects suggests, the degree to which properties of one may be inferred from those of the other. The measure of similarity provided, depends mostly on the perceptions of different observers. Emphasis should also be given to different members of the sets, so that no one member can influence the ultimate result. Many measures of similarity have been proposed in the existing literature [17, 18]. A careful analysis of the different similarity measures reveals, that it is impossible to single out one particular similarity measure that works well for all purposes.

Suppose U be an arbitrary finite set, and $\mathcal{F}(U)$ be the collection of all fuzzy subsets of U. For $A, B \in \mathcal{F}(U)$, a similarity index between the pair {*A*, *B*} is denoted as *S*(*A*, *B*; *U*) or simply *S*(*A*, *B*) which can also be considered as a function $S: \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$. In order to provide a definition for similarity index, a number of factors must be considered. We expect a similarity measure *S*(*A*, *B*) to satisfy the following axioms :

P1. *S*(*B*, *A*) = *S*(*A*, *B*), *S*(*not A*, *not B*) = *S*(*A*, *B*), *not A* being some negation of *A*.
P2. 0 ≤ *S*(*A*, *B*) ≤ 1.

P3. S(*A*, *B*) = 1 if and only if *A* = *B*.
P4. For two fuzzy sets *A* and *B*, simultaneously *not null,* if *S*(*A*, *B*) = 0 then $\min(\mu_A(u), \mu_B(u)) = 0$ for all u ∈ *U*, i.e., $A \cap B = \Phi$.
P5. If either $A \subseteq B \subseteq C$ or $A \supseteq B \supseteq C$ then S(*A*, *C*) ≤ min{S(A, *B*), *S*(*B*, *C*)}.

We now consider a definition of measure of similarity which has been proposed in [13, 19].

Definition 4 *Similarity Indices* Let A and B be two fuzzy sets defined over the same universe of discourse U. The similarity index S(A, B) of pair {A, B} is defined by

$$S(A,B) = 1 - \left(\frac{\sum_u |\mu_A(u) - \mu_B(u)|^q}{n}\right)^{\frac{1}{q}},$$

where *n* is the cardinality of the universe of discourse and $q \geq 1$ is the family parameter.

Measure of dissimilarity is another measure of comparison of objects in literature. Many authors like Meunier et al. [17] have defined measure of dissimilarity in different way. However, we use the dissimilarity measure in the context of similarity measure and consider measure of dissimilarity of two fuzzy subsets *A* and *B* defined over $\mathcal{F}(U)$, denoted by *D*(*A*,*B*) as *D*(*A*, *B*) = *1* − *S*(*A*, *B*). Moreover, we assume *D*(*A*, *B*) = *S*(*not A*, *B*) = *S*(*A*, *notB*). Through out the chapter, we use this concept of dissimilarity.

In the next section, principle of fuzzy resolution is discussed based on the method of inverse approximate reasoning [12].

3 Fuzzy Resolution Based on Inverse Approximate Reasoning

Instead of complementary literal in the set of clauses in fuzzy logic, we introduce here a concept of similar/dissimilar literal. Let us consider two clauses $C_1 = P \vee C_1{}'$ and $C_2 = P' \vee C_2{}'$. Resolvent of C_1 and C_2 denoted by $\mathrm{res}(C_1, C_2) = C_1{}' \vee C_2{}'$ if and only if similarity between *P* and *P*´ is less than Q or equivalently, dissimilarity between *P* and *P*´ is greater than 1 − Q, Q being pre-defined threshold.

The argument form of simple fuzzy resolution is as follows.

$$\begin{array}{c} A \quad \vee \quad B \\ \quad\quad\quad not\ B \\ \hline A \end{array}$$

The scheme for Generalised Fuzzy Resolution is given in Table 3.

Table 3 Generalised fuzzy resolution

Rule: X is A or	Y is B
Fact:	Y is B'
Conclusion: X is A'	

Table 4 Inverse approximate reasoning

Rule(p): if X is A then	Y is B
Fact(q):	Y is B^*
Conclusion: X is A^*	

In this case, we can say that the Disjunctive Syllogism holds if B' is close to *notB*, A' is close to A.

The scheme in inverse approximate reasoning is as given in the following Table 4. Here, fuzzy sets A and A^* are defined over the universe of discourse $U = \{u_1, u_2, \ldots, u_m\}$ and fuzzy sets B and B* are defined over the universe of discourse $V = \{v_1, v_2, \ldots, v_n\}$.

We, now, transform the disjunction form of rule into fuzzy implication or fuzzy relation and apply the method of inverse approximate reasoning to generate the required resolvent.

We proposed two algorithms SIAR and INAR for inverse approximate reasoning based on similarity given in [12]. We apply these methods to obtain a resolvent in fuzzy resolution.

In classical logic, we have

$$a \to b \equiv \neg a \vee b, \ \forall a, b \in \{0, 1\}. \tag{1}$$

Let us extend this classical logic equivalence to fuzzy logic, by interpreting the disjunction and negation as a fuzzy union (t-conorm) and a fuzzy complement, respectively. Fuzzy implication thus obtained is usually referred to in the literature as S-implication.

We now consider the classical logic tautology which is obtained from (1).

$$a \vee b \equiv \neg a \to b, \ \forall a, b \in \{0, 1\}. \tag{2}$$

Extending the classical equivalence (2) into fuzzy logic, we find that the fuzzy union is transformed to fuzzy implication. In fuzzy resolution, we deal with rule of type 'X is A or Y is B'. Like classical logic, we may transform the rule into 'If X is *notA* then Y is B' under fuzzy logic. Hence, the equivalent scheme of Table 3 that conforms fuzzy resolution is considered in Table 5. It is noted that A^* is similar to A, whenever B* is similar to *notB*.

We have demonstrated in [12] that, if the given data is sufficiently dissimilar to the consequent part of a given rule then we conclude that the resulting fuzzy set is sufficiently dissimilar to the antecedent part of the rule. Applying this method, in the scheme given in Table 5, we get the required resolvent, which establishes the fuzzy resolution principle. The algorithm is as follows:

Table 5 Equivalent scheme conforms fuzzy resolution

Rule: If X is *notA* then	Y is B
Fact:	X is B^*
Conclusion: X is A^*	

ALGORITHM—FRIAR:

Step 1. Translate the rule into fuzzy implication as $\mu_R(u,v) = I(\mu_{notA}(u),\ \mu_B(v))$, where I is an implication operator;

Step 2. Take cylindrical extension of B^* in V on $U \times V$, say R', defined by $R' = \sum_{U\times V} \mu_B(v)/(u,v)$;

Step 3. Compute $R^* = R \cap R'$, where $\cap$ denotes any fuzzy conjunction operator;

Step 4. Obtain $A^* = projR^*$ on U defined by

$$projR^*\text{on U} = \sum\nolimits_u \operatorname*{Sup}_v \mu_R(u,v)/u.$$

Mathematically, we get

$$\begin{aligned}\mu_{A*}(u) &= proj_{v\in V} R^*(u,v)\\ &= \sup_{v\in V} T(\mu_R(u,v), \mu_{R'}(u,v))\\ &= \sup_{v\in V} T(\mu_R(u,v), \mu_{B*}(v))\end{aligned} \tag{3}$$

where T is a t-norm used to describe fuzzy conjunction operator.

It is expected that, for the observation 'Y is *notB*' and the given premise 'X is A or Y is B' we can conclude 'X is A' by fuzzy resolution. However, for the the observation 'Y is B' no conclusion can be drawn. We establish the above criteria by the following theorem.

Theorem 2 Let B^* = *not B* be normal and R be interpreted by any S-implication satisfying (3). Then $A^* \supseteq A$ for any t-norm T. (consistency)

Corollary Let B^* = *not B* be normal and R be interpreted by any S-implication satisfying (3). Then $A^* = A$ for Lukasiewicz t-norm T.

Example 1 Consider the premises

$$\begin{array}{llll} p: & X \text{ is } LARGE & or & Y \text{ is } SMALL\\ q: & & & Y \text{ is not } SMALL\end{array}$$

in which X and Y are defined over the universes $U = \{u_1,\ldots, u_4\}$ and $V = \{v_1,\ldots, v_4\}$ respectively and fuzzy sets labelled by *LARGE, SMALL* and *not SMALL* are defined by

$$A \triangleq \text{LARGE} = 0/u_1 + 0.45/u_2 + 0.95/u_3 + 1/u_4 \tag{4}$$

Table 6 A* for Reichenbach S-implication

T/μA*	u_1	u_2	u_3	u_4	A^*	S(A, A*)
M	0.35	0.53	0.95	1.0	⊃ A	0.820
P	0.23	0.45	0.95	1.0	⊃ A	0.886
B	0.0	0.45	0.95	1.0	A	1.0

Table 7 A* for Kleene-Dienes S-implication

T/μA*	u_1	u_2	u_3	u_4	A^*	$S(A, A^*)$
M	0.35	0.45	0.95	1.0	⊃ A	0.825
P	0.23	0.45	0.95	1.0	⊃ A	0.886
B	0.0	0.45	0.95	1.0	A	1.0

Table 8 A* for Lukasiewicz S-implication

$T/\mu A^*$	u_1	u_2	u_3	u_4	A^*	$S(A, A^*)$
M	0.35	0.60	0.95	1.0	⊃ A	0.809
P	0.23	0.51	0.95	1.0	⊃ A	0.882
B	0.0	0.45	0.95	1.0	A	1.0

$$B \triangleq \text{ SMALL} = 1/v_1 + 0.65/v_2 + 0.15/v_3 + 0.0/v_4 \tag{5}$$

$$\text{B}^* \triangleq not\,\text{SMALL} = 0/v_1 + 0.35/v_2 + 0.85/v_3 + 1/v_4 \tag{6}$$

The similarity between fuzzy sets B and B^* is 0.0, i.e., fuzzy set B^* in observation is dissimilar to fuzzy set B in the disjunctive form of rule.

Again, by INAR, we study the shape of the resolvent A^* for data given in (4), (5) and (6) with different S-implications and different t-norms, which is described in the Tables 6, 7 and 8, where M, p and B indicate minimum, product and bounded product respectively for t-norms.

The result shows that the dissimilarity between B^* and B assures the similarity between A^* and A when the reasoning mechanism is handled using inverse approximate reasoning. Thus the proposition 'given a disjunction and the negation of one of the disjuncts, the other may be inferred' is established in fuzzy logic.

Example 2 Now, we consider the scheme and data of *Example 1* except B^*. Consider $B^* = 0.0/v_1 + 0.1225/v_2 + 0.7225/v_3 + 1.0/v_4$ in (6). We shall observe the results for the given premise '*p*' and data in (4) and (5).

In this case, $S(B,B^*) = 0.1304$, i.e., fuzzy sets B and B* are dissimilar.

Let us execute the reasoning mechanism by INAR. The results are shown in Table 9, 10 and 11 respectively for different implications and t-norms.

Table 9 A* for Reichenbach S-implication

$T/\mu A^*$	u_1	u_2	u_3	u_4	A^*	$S(A, A^*)$
M	0.15	0.53	0.95	1.0	⊃A	0.9144
P	0.11	0.45	0.95	1.0	⊃A	0.9458
B	0.0	0.45	0.95	1.0	A	1.0

Table 10 A* for Kleene-Dienes S-implication

$T/\mu A^*$	u_1	u_2	u_3	u_4	A^*	$S(A,A^*)$
M	0.15	0.45	0.95	1.0	⊃ A	0.9250
P	0.11	0.45	0.95	1.0	⊃ A	0.9458
B	0.0	0.45	0.95	1.0	A	1.0

Table 11 *A** for Lukasiewicz S-implication

$T/\mu A^*$	u_1	u_2	u_3	u_4	A^*	$S(A, A^*)$
M	0.15	0.60	0.95	1.0	⊃ A	0.8939
P	0.11	0.45	0.95	1.0	⊃ A	0.9458
B	0.0	0.45	0.95	1.0	A	1.0

That is, if B^* is not exactly match with *notB* but these are dissimilar, the fuzzy resolvent can be obtained through inverse approximate reasoning method. The technique is new.

Theorem 3 Let $B^* = B$ be normal and R be interpreted by any implication satisfying (3). Then $A^* = UNKNOWN$ for any t-norm T.

We prove the theorem for Reichenbach S-implication and $T = min$ only, but the above theorem can be proved for any other implications and any other t-norms in the similar way.

Example 3 In *Example* 1, if we take $B^* = B \triangleq SMALL = 1.0/v_1 + 0.65/v_2 + 0.15/v_3 + 0.0/v_4$ in (6) then either by SIAR or by INAR we get $A^* = 1.0/u_1 + 1.0/u_2 + 1.0/u_3 + 1.0/u_4 = UNKNOWN$, for all of the cases.

Theorem 4 Let $B^* = not\ B$ be normal and R be interpreted by Rescher-Gaines R-implication satisfying (3). Then $A^* \subseteq A$ for any t-norm T.

Example 4 For the data given in *Example 1,* applying INAR for Rescher-Gaines R-implication combined with any t-norm T, we get the fuzzy resolvent A^* as

$A^* = 0.0/u_1 + 0.35/u_2 + 0.85/u_3 + 1.0/u_4$ which is a subset of A and $S(A, A^*) \approx 0.929$. It establishes *Theorem 4.*

We, now, apply another method SIAR [12] to obtain fuzzy resolvent for the scheme given in Table 3. Let us consider another classical logic equivalence

$$a \vee b \equiv b \vee a \equiv \neg b \rightarrow a \tag{7}$$

The classical logic equivalence (7) can be extended in fuzzy logic with implication and negation function. Then we transform the rule in Table 3 into its equivalent form 'p^1 : If Y is *notB* then X is A' over the domain of $[0,\underline{1}]^{V \times U}$. A fuzzy rule may be defined by means of a conjunction for defining a fuzzy Cartesian product rather than in terms of a multivalued logic implication [12, 20].

Therefore, the rule in p^1 is transformed into fuzzy relation R as

$$\mu_R(v,u) = T(1 - \mu_B(v), \mu_A(u)), \tag{8}$$

where T is a t-norm describing a fuzzy conjunction.

Now, we can apply our method SIAR described in [12]. The algorithm is as follows :

ALGORITHM—FRSIAR:

Step 1. Translate given premise p^1 and compute R(*not B, A*) by (8);
Step 2. Compute similarity measure S(*not B, B**) using some suitable definition;
Step 3. Modify R(*not B, A*) with S(*not B, B**) to obtain the modified conditional relation R(*not B*, $A|B^*$) using *scheme* in [12];
Step 4. Use *sup-projection* operation on R(*not B*, $A|B^*$) to obtain A^* as

$$\mu_{A*}(u) = \sup_v \mu_{\mathrm{R}(not\,B,A|B*)}(v,u). \tag{9}$$

We now illustrate the method applied here by some suitable examples.

Example 5 Let us consider the data of *Example 1*. For completely dissimilar B^* with B and for different t-norms T, the shapes of fuzzy resolvent A^* in U are studied here, when we apply SIAR. In each case, it turns out exactly the fuzzy set A which corresponds '*LARGE*'.

Example 6 Let us consider the data in *Example 1* where B^* is not completely dissimilar with B, but dissimilarity exceeds certain threshold. Applying SIAR we observe the shapes of A^* and compare it with given A for different t-norms. Since $S(A, A^*) \approx 0.92$, i.e., A^* is almost similar to A, it establishes fuzzy resolution in reasoning.

In the above methods, we applied INAR or SIAR when the disjunctive knowledge can be transformed into fuzzy implication. However, it may not always be the case. Moreover, when the expert knowledge is in complex form of disjunction it is difficult to apply INAR or SIAR. So, we extend our method in such a way that can deal with complex premises.

Table 12 Generalised fuzzy resolution-extended form

p:	X is A or	Y is B
q:	Y is B´ or	Z is C
r:	X is A´ or	Z is C´

4 Fuzzy Resolution with Complex Clauses

In this section, we extend the scheme given in Table 3. Let *X*, *Y* and *Z* be three linguistic variables that take values from the domain *U, V* and *W* respectively. We consider the derivation of an inexact conclusion '*r*' from two typical knowledge (premises) '*p*' and '*q*' according to the scheme given in Table 12, where *A*'s, *B*'s and *C*'s are approximations of possibly inexact concepts by fuzzy sets over *U, V* and *W* respectively. In 1993, Raha and Ray [11] applied Zadeh's [21] concept of approximate reasoning with the application of possibility theory to model a deductive process 'Generalised Disjunctive Syllogism'. They used projection principle and conjunction principle to deduce fuzzy resolvent. Here, we investigate another method which is described in the following algorithm.

ALGORITHM—FRCEP:

Step 1. Translate the premise *p* into fuzzy relation

$$R_1 \subseteq \mathcal{F}(U \times V) \text{ as } \mu_{R_1}(u,v) = \min(\mu_A(u), 1 - \mu_B(v));$$

Step 2. Translate the premise *q* into fuzzy relation

$$R_2 \subseteq \mathcal{F}(V \times W) \text{ as } \mu_{R_2}(v,w) = \min(\mu_{B\prime}(v), \mu_C(w));$$

Step 3. Take cylindrical extension of R_1 in $U \times V$ on $U \times W$, say $R_1{}'$, defined by

$$R_1' = \sum_{U \times V \times W} \mu_{R_1}(u,v)/(u,v,w);$$

Step 4. Take cylindrical extension of R_2 in $V \times W$ on $U \times V \times W$, say $R_2{}'$, defined by

$$R_2' = \sum_{U \times V \times W} \mu_{R_2}(v,w)/(u,v,w);$$

Step 5. Construct $R' = R_1{}' \cap R_2{}'$, where $\cap$ denotes any fuzzy conjunction operator;

Step 6. Compute *S*(not*B*, *B*´) and, say, *s*;

Step 7. Modify *R*´ with *s* by a *Scheme* in [12] and, say, *R**;

Step 8. Obtain *R* = *projR** on $U \times W$ defined by

$$projR^* \text{ on } U \times W = \sum_{U \times W} \sup_v \mu_{R*}(u,v,w)/(u,w);$$

Step 9. Obtain *A*' and *C*' by projecting *R* separately on *U* and *W* such that

$$A' = proj_U R = \sum_U \sup_w \mu_R(u, w)/u \text{ and}$$
$$C\prime = proj_W R = \sum_W \sup_u \mu_R(u, w)/w)$$

Symbolically, the fuzzy resolvent R is obtained by, for $u \in U$, $v \underline{\in} V$ and $w \underline{\in} W$, $\mu_R(u, w)$

$$= \sup_v \mu_{R*}(u, v, w)$$
$$= \sup_v \{s \rightarrow \mu_{R\prime}(u, v, w)\},$$
$$= s \rightarrow \sup_v \mu_{R\prime}(u, v, w)$$
$$= s \rightarrow \sup_v T\left(\mu_{R_1\prime}(u, v, w), \mu_{R_2\prime}(u, v, w)\right)$$
$$= s \rightarrow \sup_v T\left(\mu_{R_1}(u, v), \mu_{R_2}(u, v)\right)$$
$$= s \rightarrow \sup_v T(\min(\mu_A(u), 1 - \mu_B(v)), \min(\mu_{B\prime}(\mathrm{v}), \mu_C(w)))$$
$$= s \rightarrow T\left(\min\left(\mu_A(u), 1 - \inf_v \mu_B(v)\right), \min\left(\sup_v \mu_{B\prime}(v), \mu_C(w)\right)\right)$$
$$= s \rightarrow T(\mu_A(u), \mu_C(w)), \text{ iff } 1 - \inf_v \mu_B(v) = \sup_v \mu_{B\prime}(v) = 1.$$

This derivation can be achieved if there is a $v_0 \in V$ such that $\mu_B(v_0) = 0$ and $\mu_{B'}(v_0) = 1$ which is possible if the fuzzy sets B and B' are dissimilar, i.e., *notB* and B are similar for any implication → in derivation. We observe two criteria here.

Criterion 1: Taking $x \rightarrow y = \min(1, y/x)$, we get

$$\mu_R(u, w) = \min(1, T(\mu_A(u), \mu_C(w))/s);$$

Criterion 2: Taking $x \rightarrow y = 1 - x + xy$, we get

$$\mu_R(u, w) = 1 - s + T(\mu_A(u), \mu_C(w)).s.$$

From the two criteria above, we observe that when $s = S(notB, B') = 0$, i.e., when B and B' are completely similar, $R = U \times W = UNKNOWN$. Therefore, fuzzy resolvent could be anything. However, if s is close to unity, i.e., if B and B' are almost dissimilar we have R is close *to* $\sum_{U \times W} T(\mu_A(u), \mu_C(w))/(u, w)$ which, after re-translation, gives 'If X is A or Z is C'. Again, we observe that a small change in B' produces a small change in fuzzy resolvent—which ensures our method is reasonable one.

Let us consider a scheme given in Table 13 where variables X_i ($i = 1, 2,\ldots,m$) and the respective fuzzy subsets A_i ($i = 1, 2,\ldots,m$) are defined on universe U_i ($i = 1,2,\ldots, m$) respectively; variables Y_j ($j = 1, 2,\ldots,n$) and the respective fuzzy subsets B_j ($j = 1, 2,\ldots, n$) are defined on universe V_j ($j = 1, 2, \ldots, n$) respectively. (A_k, B_l) is almost dissimilar over the same universe U_k (= V_l) with the degree of confidence of keyword A_k is *cd*

Table 13 Generalised fuzzy resolution-another extension

C_1:X_1 is A_1 or X_2 is A_2 or ··· or X_m is A_m;
C_2:Y_1 is B_1 or Y_2 is B_2 or ··· or Y_n is B_n;
R(C_1,C_2): X_1 is A_1´ or ··· or X_m is A_m´
or Y_1 is B_1´ or ··· or Yn is B_n´

$(A_k) = 1 - S(A_k, B_l)$ and the corresponding variables X_k, Y_l, defined over U_k (= V_l) respectively, assign same linguistic variable.

The algorithm is as follows:

ALGORITHM—FRAE:

Step 1. Check for pair of literals (A_i, B_j), $\forall$ i, j from clauses C_1 and C_2 whether these are defined over same universe and are assigned by same linguistic variable; Otherwise, resolution is not possible;

Step 2. If dissimilarity $D(A_k, B_l)$ is high, i.e., $D(A_k, B_l) > 1 - Q$, Q is pre-defined threshold then go to the next step and say, A_k is keyword; Otherwise, there is no fuzzy resolvent;

Step 3. Modify B_j either by $B_j' = \min(1, B_j/D(A_k, B_l))$ or by $B_{j'} = 1 - (1 - B_j).S(A_k, B_l))$;

Step 4. Fuzzy resolvent is $R(C_1, C_2) = A_1' \vee \cdots \vee A_m' \vee B_1' \vee \cdot \cdot \cdot \vee B_n'$ and $R(C_1, C_2)_{cd} = cd(A_k)$ which is measured as $cd(A_k) = D(A_k, B_l)$;

Step 5. Repeat the process until empty clause, with the confidence $cd \neq 0$, is derived for more than two clauses.

Hence, we prove the (un)satisfiability of a theorem by the deduction of empty clause from a set of fuzzy clauses.

Let us consider an example to illustrate the method. Suppose variables that range over finite sets or can be approximated by variables ranging over such sets.

Example 7 Let us consider the premises

$$p : X \text{ is LARGE} \quad \textit{or} \;\; Y \text{ is SMALL } \textit{or} \; Z \text{ is LARGE;}$$
$$q : X \text{ is not LARGE;}$$
$$r : Y \text{ is not SMALL;}$$

in which X, Y and Z are defined over the universes $U = \{u_1,\ldots, u_4\}$, $V = \{v_1,\ldots, v_4\}$ and $W = \{w_1, w_2, w_3, w_4\}$ respectively. Fuzzy sets labelled by LARGE, SMALL and not SMALL defined over the universes U, V and W respectively are given in *Example 1* and fuzzy set LARGE defined over W is given by

$$C \triangleq LARGE = 0/w_1 + 0.20/w_2 + 0.75/w_3 + 1/w_4.$$
$$A' \triangleq not\ LARGE = 1.0/u_1 + 0.55/u_2 + 0.05/u_3 + 0.0/u_4.$$

We execute the following steps with the given data, using ALGORITHM-FRCEP.

Execution:

Step 1. Compute the fuzzy relation $R_1 = A \vee B \vee C$ by

$$\mu_{R'}(u,v,w) = \min(1-\mu_A(u), 1-\mu_B(v), \mu_C(w));$$

Step 2. Extend A' in U cylindrically on $U \times V \times W$ as

$$R_2 = \sum_{U\times V\times W} \mu_{A'}(u)/(u,v,w);$$

Step 3. Compute $R' = R_1 \cap R_2$ by $\mu R'(u,v,w) = T(\mu R_1(u,v,w), \mu R_2(u,v,w))$, taking fuzzy conjunction $\cap$ as t-norm T;
Step 4. Modify R' with $S(1-\mu_A(u), \mu_{A'}(u)) = s_1$ to get R^* as $\mu_{R^*}(u,v,w) = 1-(1-\mu_{R'}(u,v,w))\cdot s_1$;
Step 5. Project R^* on $V \times W$ such that

$$R = \sum_{V\times W} \sup_u \mu_{R*}(u,v,w) = \begin{pmatrix} 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.20 & 0.45 & 0.45 \\ 0.00 & 0.20 & 0.75 & 0.85 \\ 0.00 & 0.20 & 0.75 & 1.00 \end{pmatrix}$$

Step 6. Extend B' in V cylindrically on $V \times W$ as $R_3 = \sum_{V\times W} \mu_{B\prime}/(v,w)$
Step 7. Compute $R'' = R \cap R_3$ by $\mu_{R''}(v, w) = \mathrm{T}(\mu_R(v, w), \mu_{R3}(v, w))$, taking fuzzy conjunction $\cap$ as t-norm T;
Step 8. Modify R'' with $S(1 - \mu_B(v), \mu_{B'}(v)) = s_2$ to get R^{**} as $\mu_R{}^{**}(v, w) = 1-(1-\mu_{R''}(v, w))\tilde{n}s_2$;
Step 9. Project R^{**} on W such that

$$C' = \sum_w \sup_v \mu_{R**}(v,w) = 0/w_1 + 0.20/w_2 + 0.75/w_3 + 1/w_4,$$

which is completely similar to C, i.e., $S(C', C) = 1$.

Even if A' and B' in the respective premises p and q, are not completely dissimilar to A and B respectively, but the dissimilarity measures attain values greater than certain predefined threshold then we can get a fuzzy resolvent C' which is almost similar to C, using FRCEP.

Suppose,

$$A' = 1.0/\mu_1 + 0.3025/\mu_2 + 0.0025/\mu_3 + 0.0/\mu_4;$$
$$B' = 0.0/v_1 + 0.1225/v_2 + 0.7225/v_3 + 1.0/v_4.$$

$S(notA, A') = 0.873992$; $S(notB, B') = 0.869604$ are computed.

Then, it yields $C' = 0.24/w_1 + 0.39/w_2 + 0.81/w_3 + 1.0/w_4$ with $S(C, C') = 0.843441$.

Therefore, it is possible to get a fuzzy resolvent from a set of fuzzy clauses if there is a pair of dissimilar literals contained in the respective clauses.

5 Example in Real Life

The statements in real life are mostly fabricated by vague, imprecise, uncertain and incomplete information which can be dealt with using fuzzy logic. In daily life, we have to often decide whether one statement follows from some other statements. Fuzzy resolution can be a way to deal with such phenomena. Let us consider the following example.

Example 8 Suppose, the stock price is low if the prime interest rate is high. Suppose, also, people are not happy when stock price is low. Assume that the prime interest rate is high. Can we conclude that people are not happy? Again, if we assume that the interest rate is very high, how can we conclude about the happiness of people?

For this problem, we construct the rule base in Table 14.

(i) **For Observation 1**: The above scheme may be transformed into the equivalent scheme given in Table 15. in which *prime interest rate, stock price* and *happiness* are defined over the respective universes U = [0, 20]%, V = Rs. [10,000, 22,000] and W = [0,100]. Fuzzy sets defined over the respective universes U, V and W are given in Table 16.

The membership grade of fuzzy complement is defined by $1 - a$ here, where a is the membership grade of a fuzzy set.

The similarity between complement of fuzzy set *LOW* in p and fuzzy set *not LOW* in q (denoted by s_1) is 1.0. Therefore, in two clauses, fuzzy sets are dissimilar. Hence, we can resolve upon these dissimilar pair and apply ALGORITHM—FRCEP to get the resolvent. Here,

$$R_1 = \sum_{U\times V} \min(\mu_{not}\, HIGH(u), 1 - \mu_{LOW}(v))/(u, v)$$
$$R_2 = \sum_{V\times W} \min(\mu_{not}\, LOW(u),\ \mu_{not}\, HAPPY\ (w))/(v, w)$$

We construct $R' = R_1' \cap R_2'$, where $\cap$ denotes any fuzzy conjunction operator (here, $\cap$ = min) and

$$R_1\prime = \sum_{U\times V\times W} \mu_{R_1}(u, v)/(u, v, w),$$
$$R_2\prime = \sum_{U\times V\times W} \mu_{R_2}(v, w)/(u, v, w).$$

We modify R' with dissimilarity $s_1 = 1.0$ as we get $\mu_{R*} = (u, v, w) = 1 - (1 - \mu_{R'})\cdot s_1$ and we get

$$R = projR^{*} \text{ on } U \times W$$

Let us extend the fuzzy sets 'HIGH' cylindrically in the premise r on $U \times W$ as

$$R_3 = \sum_{U\times W} \mu_{HIGH}/(u, w);$$

Table 14 Rule base and observations

Rule 1:	IF the prime interest rate is HIGH THEN the stock price is LOW;
Rule 1:	IF the stock price is LOW THEN people are not HAPPY;
Observation 1:	The prime interest rate is HIGH;
Observation 2:	The prime interest rate is very HIGH;

Table 15 Equivalent scheme for observation 1

p:	The prime interest rate is not HIGH OR the stock price is LOW;
q:	The stock price is not LoW OR the people are not HAPPY;
r:	The prime interest rate is HIGH;

Table 16 Fuzzy sets in scheme

HIGH	=	0.0/0 + 0.0/2 + 0.1/4 + 0.3/6 + 0.5/8 +0.6/10 + 0.7/12 + 0.9/14 + 1.0/16 ($\subset U$)
LOW	=	1.0/10, 000 + 0.8/12,000 + 0.75/14, 000 +0.5/16,000 + 0.3/18, 000 + 0.1/20,000 +0.0/22,000 ($\subset V$)
HAPPY	=	0.0/0 + 0.25/25 + 0.5/50 + 0.75/75 +1.0/100 ($\subset W$)

and compute $R^{''} = R \cap R_3$ by

$$\mu_{R''}(u, w) = T(\mu_R(u, w), \mu_{R_3}(u, w)),$$

taking fuzzy conjunction $\cap$ as t-norm T (here, $\cap$ = min);

Next, we calculate the dissimilarity measure of the fuzzy set *not HIGH* in p and the fuzzy set *HIGH* in r (denoted by s_2) and $R^{''}$ is induced by the dissimilarity measure $s_2 = 1$ to get R^{**} as

$$\mu_{R^{**}}(u, w) = 1 - (1 - \mu_{R''}(u, w)) \cdot s_2;$$

At last, we project R^{**} on W such that

$$C\prime = \sum_W \sup_v \mu_{R^{**}}(u, w)$$
$$= 0.50/0 + 0.50/25 + 0.50/50 + 0.25/75 + 0.0/100$$

and $S(C', \textit{not HAPPY}) = 0.75$, which is rather similar to *not HAPPY*.

(ii) **For Observation 2**: In this case, the observation is changed into

$$r : \text{The prime interest rate is } \textit{very HIGH}$$

to the scheme given in Table 15.

Describing the membership grades of the fuzzy set '*very HIGH*' in q as a^2, where a is the membership grades of the fuzzy set '*HIGH*' defined over U, defining other parameters in the same way and executing in a similar way, we can derive a fuzzy set

$$C' = 0.46/0 + 0.46/25 + 0.46/50 + 0.37/75 + 0.16/100,$$

the similarity of which is 0.71 with the fuzzy set *not HAPPY* in the premise q, whenever dissimilarity between '*not HIGH*' $\subset U$ in premise p and '*HIGH*' $\subset U$ in observation r is 0.84.

Hence, in either case, we can derive a conclusion like 'people are not so happy' for the given observations. Therefore, our resolution method is effective in deriving a conclusion from some given statements in real life.

6 Conclusion

This chapter presented a resolution principle for fuzzy formulae based on similarity and approximate reasoning methodology. Similarity is inherent in approximate reasoning and resolution deduction can be used as a rule of inference to generate new clause from a given set of clauses. The essential idea of resolution of two clauses is to search for a literal in a clausal formula that is almost complementary to a literal in the other form. The clause formed by the disjunction of the remaining literals and subsequent removal of the pair of almost complementary literals is a logical consequence. If we put the resolvent in the set of clauses its behaviour (satisfiability) never changes. It can be applied directly to any set S of clausal formulae (not necessarily to ground clauses) to test the (un)satisfiability of S. To test the unsatisfiability it checks whether S contains the empty clause, (as a resolution deduction). This could be a powerful technique in constructing a proof of a theorem using refutation procedure. Examples cited in the chapter attempted to demonstrate how resolution can be effectively used to construct a proof of a theorem or to make a decision in real life. Inverse approximate reasoning may be applied to model different goal-directed search techniques. We apply inverse approximate reasoning method to avoid the inherent problem of GMP. Instead of testing complementary literals we use dissimilarity concept of fuzzy literals.

References

1. Robinson, J.A.: A machine oriented logic based on the resolution principle. J.ACM **12**(1), 23–41 (1965)
2. Lee, R., Chang, C.L.: Some properties of fuzzy logic. Inf. Control **19**(1), 417–431 (1971)
3. Lee, R.: Fuzzy logic and the resolution principle. J. ACM **19**(1), 109–119 (1972)
4. Shen, Z., Ding, L., Mukaidono, M.: Fuzzy resolution principle. In: Proceedings of the Eighteenth International Symposium on Multivalued Logic, Spain, May 24–26, 1988

5. Mukaidono, M.: Fuzzy inference of resolution style. In: Yager, R.R. (ed.) Fuzzy Set and Possibility Theory, pp. 224–231. Pergamon Press, New York (1982)
6. Dubois, D., Prade, H.: Necessity measure and the resolution principle. IEEE Trans. Syst. Man Cybern. **17**(3), 474–478 (1987)
7. Kim, C., Kim, D.S., Park, J.: A new fuzzy resolution principle based on the antonym. Fuzzy Sets Syst. **113**(2), 299–307 (2000)
8. Fontana, F.A., Formato, F.: A similarity-based resolution principle. Int. J. Intell. Syst. **17**(9), 853–872 (2002)
9. Sessa, M.I.: Approximate reasoning by similarity-based sld resolution. Theor. Comput. Sci. **275**, 389–426 (2002)
10. Habiballa, H., Novak, V.: Fuzzy general resolution. In: Proceedings of International Conference Aplimat 2002, pp. 199–206. Slovak Technical University, Bratislava (2002)
11. Raha, S., Ray, K.S.: Approximate reasoning based on generalised disjunctive syllogism. Fuzzy Sets Syst. **61**(2), 143151 (1994)
12. Mondal, B., Raha, S.: Similarity based inverse approximate reasoning. IEEE Trans. Fuzzy Syst. **19**(6), 10581071 (2011)
13. Mondal, B., Mazumdar, D., Raha, S.: Similarity in approximate reasoning. Int. J. Comput. Cognit. **4**(3), 46–56 (2006)
14. Kelly, J.J.: The Essence of Logic. Prentice-Hall, New Delhi (1997)
15. Zadeh, L.A.: The concept of linguistic variable and its application to approximate reasoning: Part I and part II. Inf. Sci. **8**, 301–357 (1975)
16. Lazzerini, B., Marcelloni, F.: Some considerations on input and output partitions to produce meaningful conclusions in fuzzy inference. Fuzzy Sets Syst. **113**(2), 221–235 (2000)
17. Bouchon-Meunier, B., Rifqi, M., Bothorel, S.: Towards general measures of comparison of objects. Fuzzy Sets Syst. **84**(2), 143–153 (1996)
18. Zwick, R., Carlstein, E., Budescu, D.V.: Measures of similarity among fuzzy concepts: a comparative analysis. Int. J. Approx. Reason. **1**(2), 221–242 (1987)
19. Raha, S., Pal, N.R., Ray, K.S.: Similarity based approximate reasoning: methodology and application. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **32**(4), 541547 (2002)
20. Ughetto, L., Dubois, D., Prade, H.: Implicative and conjunctive fuzzy rule -a tool for reasoning from knowledge and examples. In: Proceedings of AAAI/IAAI'1999, Menlo Paru, CA, USA, pp. 214–219 (1999)
21. Zadeh, L.A.: A theory of approximate reasoning. In: Hayes, J.E., Michie, D., Mikulich, L.I. (eds.) Machine Intelligence, vol. 9, pp. 149–194. Elsevier, New York (1979)

Authors' Biography

Banibrata Mondal received B.Sc. and M.Sc. degrees in Mathematics from Visva-Bharati, Santiniketan, West Bengal, India. He is currently working toward the Ph.D. degree with the Department of Mathematics, Visva Bharati. His current research interest includes formulation of inverse approximate reasoning based on similarity.

Swapan Raha received B.Sc. and M.Sc. in Mathematics and M.Tech. (Computer Science) from Calcutta University, Kolkata, India. He received the Ph.D. degree from Indian Statistical Institute, Kolkata. He is currently a Professor with the Department of Mathematics, Visva Bharati, Santiniketan, West Bengal, India. His research interests include fuzzy logic and approximate reasoning, fuzzy automata and fuzzy control.

Development of Environment Friendly Air Conditioner *Using* Fuzzy logic

Roshan Kshirsagar, Chetankumar Patil and Ashok Deshpande

Abstract Air Conditioner (AC), Heating Ventilation and Air Conditioning (HVAC) are the basic needs of modern society. Some parameters in AC system (other than cooling inside room area), if not controlled, may cause discomfort to the occupants. These parameters could be excess of CO_2, humidity and alike, accumulated in the room. Two major concerns about the conventional controls include: (1) Wastage of energy due to variation in load changes and external disturbances, and (2) Accumulation of CO_2 inside closed room. If the refrigeration load and capacity of plant remains constant then there may be no need to opt for a new control strategy. However in practice, the two parameters do fluctuate. In order to resolve the above two concerns, efforts were made by the authors to design and implement Type 1 Fuzzy Inference System (Type 1 FIS)—commonly known as Fuzzy Expert System. Environment Friendly—the term used in the chapter relates to reduction in CO_2 level in the room resulting in partial indoor air pollution control. The detailed control instrumentation used in fuzzy logic control is the integral part of the chapter.

1 Introduction

In modern society, air conditioners are commonly installed in homes and in public enclosed spaces due to the natural demand for thermal comfort. The task of dehumidification, air conditioner (AC) seizes to function like a dehumidifier in a

R. Kshirsagar (✉)
Instrumentation and Control Department, College of Engineering, Pune, India
e-mail: roshankshirsagar@yahoo.com

C. Patil · A. Deshpande
College of Engineering, Pune, India
e-mail: cyp.instru@coep.ac.in

A. Deshpande
Berkeley Initiative Soft Computing (BISC), Special Interest Group (SIG), Environmental Management Systems (EMS), Berkeley, CA, USA
e-mail: ashok_deshpande@hotmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_24,

conventional controller. Also complex interactions between user preferences, actual room temperature and humidity level are very difficult to model mathematically. In summary, conventional controllers such as feedback control, a proportional–integral–derivative controller (PID control), and alike have a few but important limitations such as: dead band, hysteresis, nonlinearity among others. Therefore, there is a need to look for some other formalism in order to model such a dynamic system. Fuzzy logic application has gained and will continue to gain its importance in the application of various areas of science and technology including fuzzy controllers, in the near future.

Wealth of information is available on the concept of fuzzy sets and fuzzy logic. In Computing With Words, fuzzy logic is labeled as the Level 1 complexity. The concept of linguistic variable, fuzzy sets, and approximate reasoning play a central role in fuzzy logic.

Figure 1 shows Fuzzy patches which cover all the regions in a non-linear system model. The key is: *Fuzzy rules are fuzzy patches.*

The use of fan and damper are the two important equipments in air conditioners. Figure 2 exhibits non-linear properties of fans and dampers for different working points. Even a well-tuned PID controller may not be able to achieve the desired performance for all set points and process variations and shows oscillations and more time to settle down.

The first success story of fuzzy logic controller goes back to 1973 when Professor Ebrahim Mamdani and Assilian, implemented fuzzy rule base system to stablise the speed of a small steam engine. Figure 3 presents the output of a fuzzy controller developed by Mamadani and Assilian—a departure from the conventional controls, demonstrating the strength of the concept of linguistic variable and approximate reasoning. Thereafter, there are over 1,000 fuzzy logic based smart products patented.

Major advantages of Fuzzy logic formalism are:

1. Massive parallelism.
2. Distributed representation and hybrid computation.
3. Low energy consumption.
4. Exact control without any overshoot.
5. High speed calculations.

2 Fuzzy Logic Based Environment Friendly Air Conditioner

Performance of an air conditioning system by a variable speed fan for load matching and thermal comfort has been one of the objectives of the chapter using fuzzy logic based algorithm. In this approach, cooling fan speed is controlled using Fuzzy logic Air Controller (FLC). However, the emphasis of the study is on Dilution of CO_2 accumulated in a room so as to reduce indoor air pollution, apart from the energy savings to make the fuzzy logic based controller Environment Friendly.

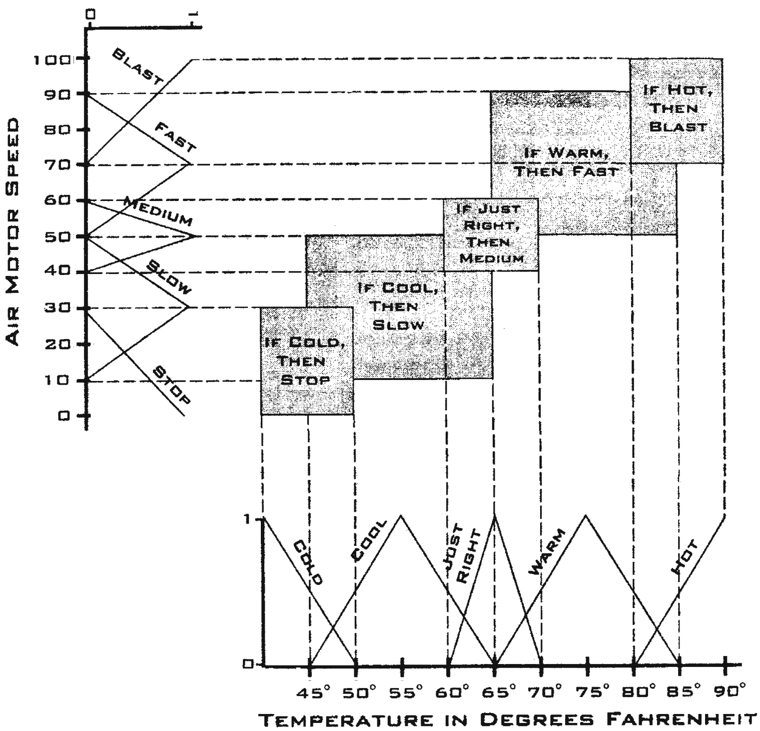


Fig. 1 Fuzzy rules are patches

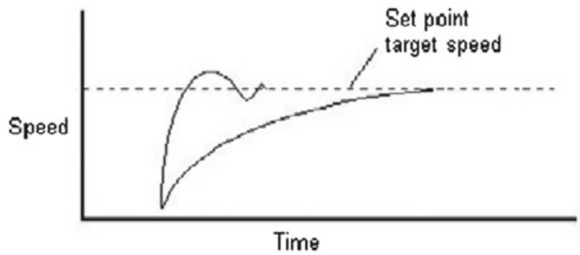


Fig. 2 System response without fuzzy controller

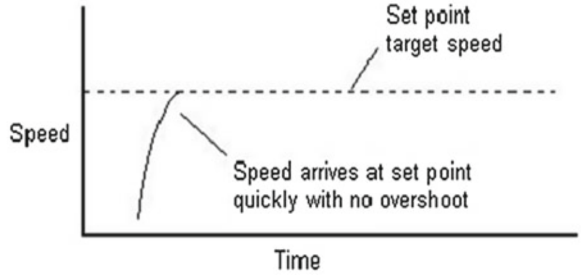


Fig. 3 System response with fuzzy controller

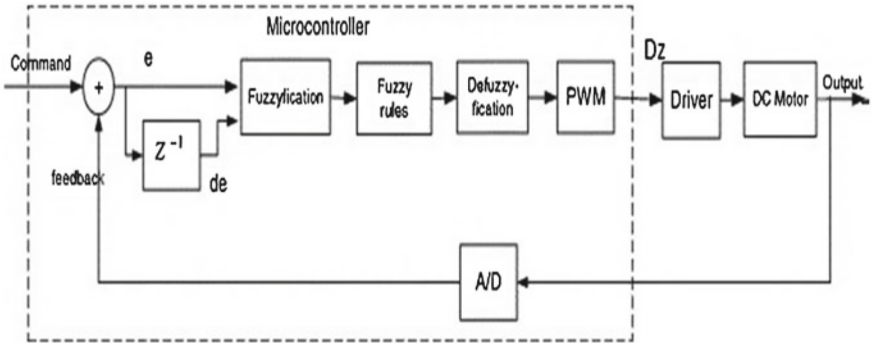


Fig. 4 Model fuzzy logic air conditioner

Figure 4 represents the main components of Fuzzy logic Air FLC viz: input and output variables, fuzzification, inference mechanism, fuzzy rule base and defuzzification. FLC involves receiving input signal and converting the signal into fuzzy variable (fuzzifier). The fuzzy control rules relate the input fuzzy variables to an output fuzzy variable which is called fuzzy associative memory (FAM), followed by defuzzification to obtain fuzzy to crisp conversion (defuzzifier). In this FLC, there are two fuzzy input variables and one output fuzzy variable. The fuzzy input variables are the error, 'e' which is the error between the reference and the measured temperature and the error difference(De) which is the rate of change of the error can be approximated as the difference between the present error and the previous error. The output fuzzy variable, DZ is the voltage signal to the motor.

What is Environment Friendly?

Accumulation of CO_2 inside a closed room causes discomfort to the occupants due to its increased levels. This is usually experienced in office, homes and more importantly, in cinema theatres and malls. There is no provision in the existing air-conditioner to remove CO_2 and air flow in the room. We provide necessary facilities to detect and dilute CO_2 and fresh air flow in order to avoids indoor air pollution, ensuring safety of the occupants. Experiments have shown that when CO_2 level exceeds 350 ppm, the person inside the room complains of suffocation.

Response of human bodies with the increase in CO_2 [google document]

1. 0.04 %: Suit for breath,
2. 0.07 %: Few people who are sensitive to gas feel discomfort.
3. 0.1 %: Feel discomfort generally.
4. 3 %: Deepness of breath increases.
5. 4 %: Headache, blood pressure rises.
6. 8–10 %: Obvious unconsciousness, so as to stop breath.
7. 30 %: Death.

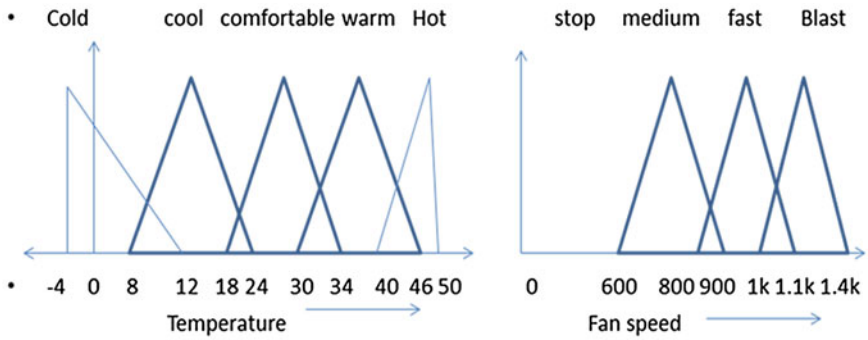


Fig. 5 Fuzzy membership for temp and fanspeed

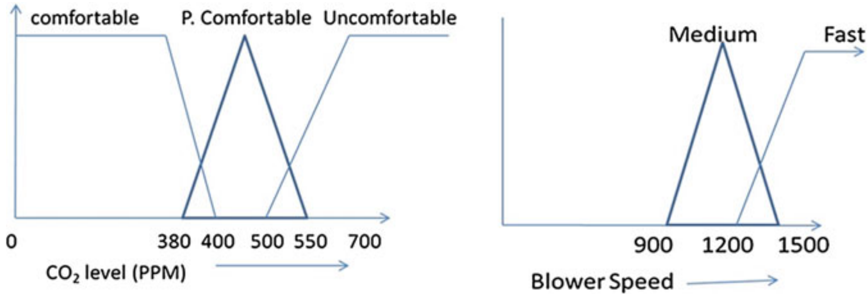


Fig. 6 Fuzzy membership for CO_2 and blowerspeed

2.1 *Formulation of Fuzzy in Temperature Control and Environment Friendly Fuzzy If then Rules*

Fuzzy rules for Temperature and Fan Speed (Fig. 5):

1. If temp. is Cold then fan Speed is Stop.
2. If temp. is Cool then fan Speed is Stop.
3. If temp. is Comfortable then fan Speed is Medium
4. If temp. is Warm then fan Speed is Fast.
5. if temp. is Hot then fan Speed is Blast.

Fuzzy rule for Carbon dioxide and Blower Speed (Fig. 6):

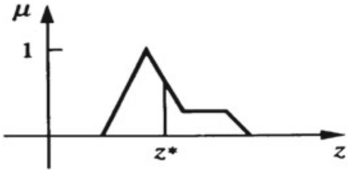


Fig. 7 Defuzification

1. If CO_2 level is Comfortable then Blower speed is Stop.
2. If CO_2 level is Partially Comfortable then Blower speed is Medium.
3. If CO_2 level is Uncomfortable then Blower speed is Fast.

2.2 Defuzzification Process

Centroid most prevalent and physically appealing of all defuzzification methods (Fig. 7) but it is the disadvantage is that it is computationally intensive, which is summarized [1]:

$$z^* = \frac{\int \mu_C(z) \cdot z dz}{\int \mu_C(z) dz}$$

where $\int$ denotes an algebraic integration.

3 Control Instrumentation

Following is the system description [2]:

The FLC model is being implemented using Micro controller PIC16f73 with the following properties:

1. High performance RISC CPU.
2. Two Capture, Compare, PWM modules.
3. Analog to Digital converter. Many electrical signals around us are Analog in nature. That means a quantity varies directly with some other quantity. The first quantity is mostly voltage while that second quantity can be anything like temperature, pressure, light, force or acceleration. For example in LM35 temperature sensor the output voltage varies according to the temperature, so if we could measure voltage, we can measure temperature. But most of our computer (or Microcontrollers) are digital in nature. They can only differentiate between HIGH or LOW level on input pins. For example if input is more than 2.5 V it will be read as 1 and if it is below 2.5 then it will be read as 0 (in case of 5 V systems). So we cannot measure voltage directly from MCUs. To solve this problem most modern MCUs have an ADC unit. ADC stands for analog to

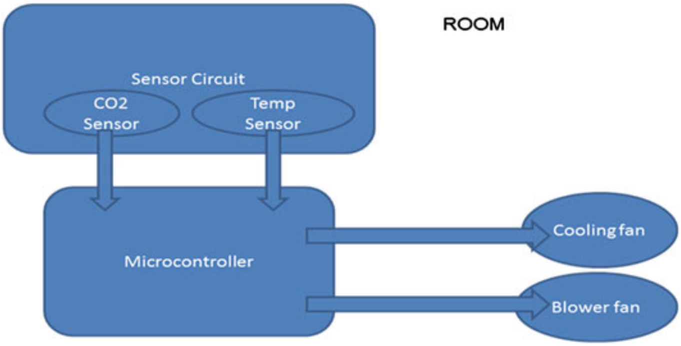


Fig. 8 Block diagram of the working model

digital converter. It will convert a voltage to a number so that it can be processed by a digital systems like MCU.

Figure 8 depicts the block schematic of working model as well working principles. Temperature and CO_2 sensors control the fan speed and the blower speed, respectively. A cooling fan is fitted in Air Conditioner AHU (Air Handling Unit) which describes the flow of cooled air according to Temperature signal received from sensor. The blower starts when CO_2 level exceeds the set point value, which allows air to enter inside the room until CO_2 level again goes down to suitable breathing level.

We select option of replacing existing air from the room by fresh air with the following considerations:

1. CO_2 is a heavier gas, it settles down very quickly and hence it is difficult to remove.
2. While removing CO_2 from room, large amount of cooled air may get exhausted, resulting in affecting thermal comfort index.

Figure 9 shows the working model of air conditioner. The refrigeration cycle uses four essential elements to create a cooling effect. The system refrigerant start its cycle in a gaseous state. The compressor pumps the refrigerant gas up to a high pressure and temperature. From there it enters a heat exchanger (sometimes called a "condensing coil" or condenser) where it loses energy (heat) outside. In the process the refrigerant condenses into a liquid. The liquid refrigerant is returned indoors to another heat exchanger ("evaporating coil" or evaporator). A metering device allows the liquid to flow in at a low pressure at the proper rate. As the liquid refrigerant evaporates it absorbs energy (heat) from the inside air, returns to the compressor, and repeats the cycle. In the process heat is absorbed from indoors and transferred outdoors, resulting in cooling.

On controller board. Two sensors: Temperature and CO_2. The signal received to micro controller is in fuzzy form. Output of micro controller controls speed of AC cooling fan and blower fan.

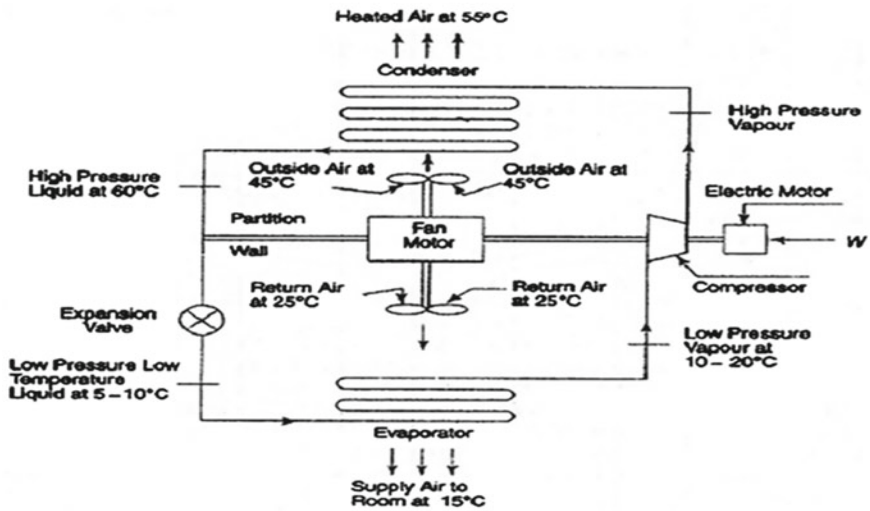


Fig. 9 Working model air conditioner

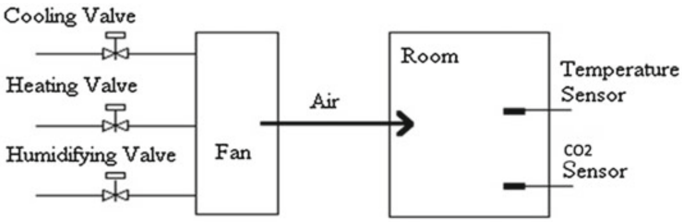


Fig. 10 Block implemented block diagram of air conditioner

Descriptions of Sensors:

1. Temperature sensor LM35 is Calibrated directly in degree Celsius (Centigrade) means output given is directly converted into degrees by converting factor. Linear at 10.0 mV/°C scale factor, 0.5 % accuracy (at 25 °C) as well [3, 4].
2. CO_2 sensor Selected-TGS4161 Range 350–5000 ppm (Parts per Million), Suitable for Indoor air Quality control. High selectivity to carbon dioxide Fig. 10 conceptual Air conditioner.

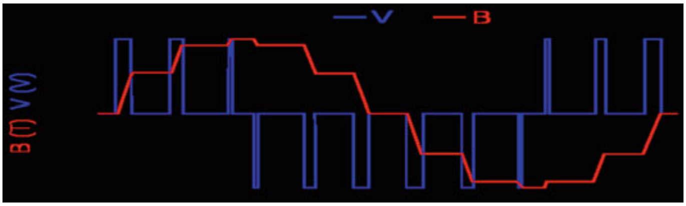


Fig. 11 Pulse width modulation

3.1 Pulse Width Modulation

Pulse-width modulation (PWM), or pulse-duration modulation (PDM), is a commonly used technique for controlling power to inertial electrical devices, made practical by modern electronic power switches [5].

The average value of voltage (and current) fed to the load is controlled by turning the switch between supply and load on and off at a fast pace. The longer the switch is on compared to the off periods, the higher the power supplied to the load. The PWM switching frequency has to be much faster than what would affect the load, which is to say the device that uses the power. Typically switching have to be done several times a minute in an electric stove, 120 Hz in a lamp dimmer, from few kilohertz (kHz) to tens of kHz for a motor drive and well into the tens or hundreds of kHz in audio amplifiers and computer power supplies.

The term duty cycle describes the proportion of 'on' time to the regular interval or 'period' of time; a low duty cycle corresponds to low power, because the power is off for most of the time. Duty cycle is expressed in percentage, 100 % being fully on [6].

The main advantage of PWM is that power loss in the switching devices is very low. When a switch is off there is practically no current, and when it is on, there is almost no voltage drop across the switch. Power loss, being the product of voltage and current, is thus in both cases close to zero. PWM also works well with digital controls, because of their on/off nature can easily set the needed duty cycle (Fig. 11).

Generally we kept duty cycle of PWM 10–90 % means for full speed PWM is 90 % for both, Cooling fan and Blower fan duty cycle is 90 % for full speed (Fig. 12).

$$\text{Duty Cycle} = t \backslash T$$

where: 't' is the duration that the function is active.
'T' is the period of the function.

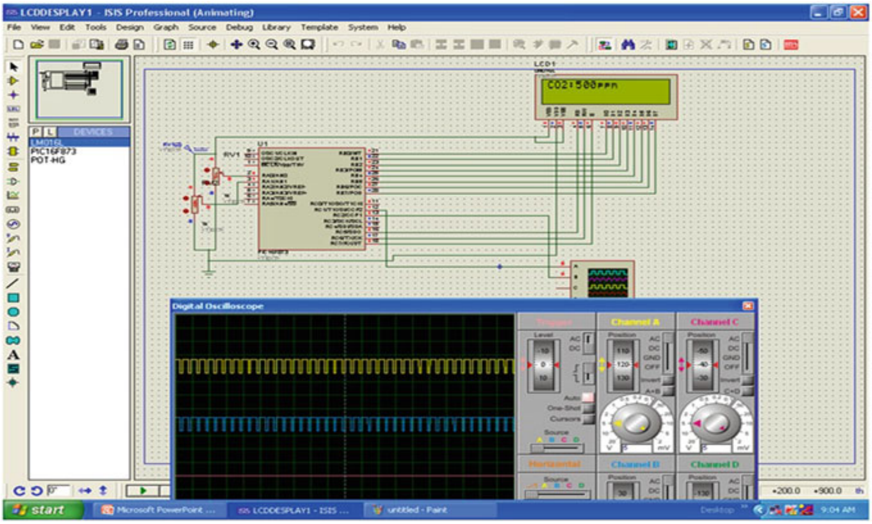


Fig. 12 PWM generated from the hardware

Description of Circuit Diagram

The implementation of FLC (Fig. 13) is described for fuzzy logic air conditioner [7]

1. Micro controller used is PIC because it has in built ADC and PWM module also it has RISC processor makes system less complex and easy for programming.
2. Temperature Sensor: The LM35 series are precision integrated-circuit temperature sensors, whose output voltage is linearly proportional to the Celsius (Centigrade) temperature. The LM35 thus has an advantage over linear temperature sensors calibrated in °K, as the user is not required to subtract a large constant voltage from its output to obtain convenient Centigrade scaling.
3. CO_2 TGS4161 is a new solid electrolyte CO_2 sensor which offers miniaturization and low power consumption. A range of 350–5,000 ppm of carbon dioxide can be detected by TGS4161, making it ideal for indoor air control applications.
4. Optoisolator MCT2E: the devices consist of a gallium arsenide infrared emitting diode optically coupled to a monolithic silicon phototransistor detector. Applications such as General Purpose Switching, Circuits Interfacing and coupling systems of different potentials and impedances and I/O Interfacing Solid State Relays
5. Transistors: Transistors used to amplification of PWM signal. This PWM signal needs boost for increase resolution of signal after Optoisolators. driver circuit should understand value of output to actuate motor on and off.
6. Pulse transformers: In addition to industrial applications of variable-speed motor drives, home appliances is another major area where PWM motor control drives are finding increasing applications. Thus, the design of a low-cost, reliable, efficient, variable speed motor has become a prime focus for both appliance designers and electronic component manufacturers. The components

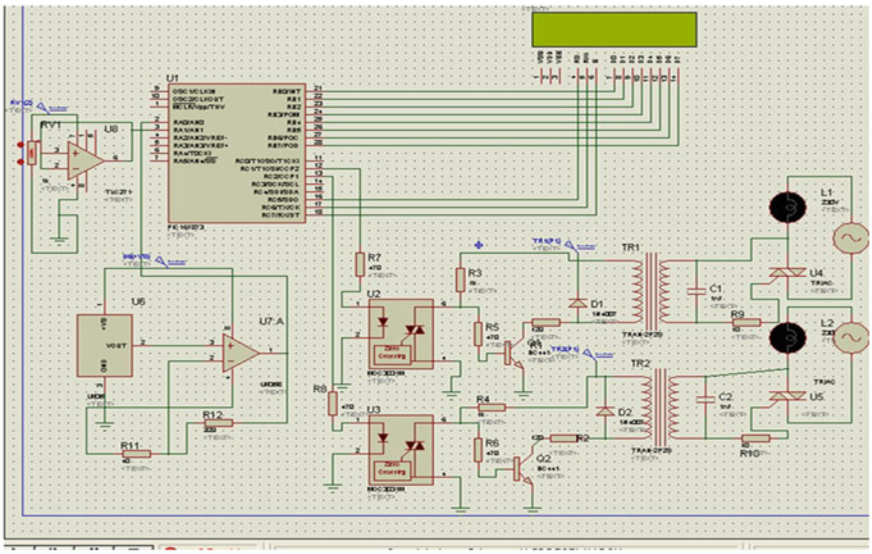


Fig. 13 Details of circuit is shown

needed for the three-phase motor electronics include IGBTs, gate drivers, inverters, microcontroller units, analog current and voltage sensors among others. It is in the area of optically isolated gate drivers and optically isolated analog current and voltage sensors that modern, state of the art, low cost, and reliable optocouplers (optoisolators) are becoming the component of first preference among the designers.

7. Triacs: Universal motors are mostly operated in AC current mode and are controlled by means of TRIACS. This widespread solution leads to a cheap electronic controller board but has some drawbacks. In particular the high peak to peak current gives poor motor efficiency and the consequential high brush temperature leads to limited motor lifetime (Fig. 13).

The closed-loop system is shown in Fig. 4. The control system for the cooling fan and blower speed consists of ICs temperature and CO_2 sensors, an inverter and an electrical motor coupled to the fans. The ICs CO_2 sensors monitor the CO_2 level of the room and emit electrical signals.

Proportional to the state of the conditioned space same function performed by temperature sensor. This signal is filtered using low pass filters and optoisolator before it reaches the controller and inverter, thus minimizing the noise. The output signal is supplied to the controller, which output a control signal that is a function of the error. The on/off and FLC signal output is supplied to the inverter, which modulates the frequency of electrical supply to the motor such that it is linearly proportional to the control signal. Electricity of 50 Hz is supplied to the inverter, which supplies electricity to turn on/off

Fuzzy logic controller is designed and developed in both Matlab and Hardware (Fig. 14).

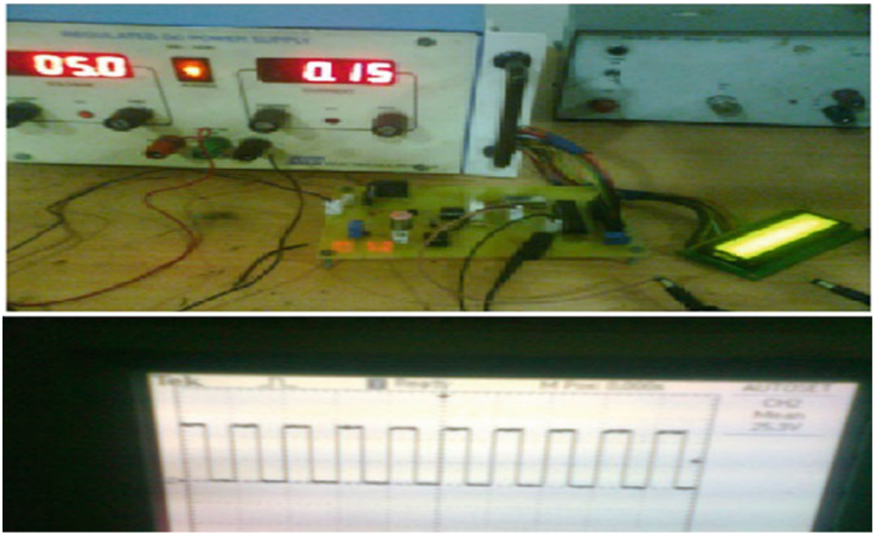


Fig. 14 FLC hardware results

4 Scope for Further Research

4.1 Introduction to Artificial neural network (ANN) in Air Conditioner

Advantages of ANN are well documented and there are many commercial products using this soft computing technique. An attempt will be made to compare the performance of ANN model and fuzzy model for [8]:

4.2 Additional Sensors

In order to enhance the credibility of fuzzy logic controls, we propose to introduce a few sensors for moisture control and humidity control in the existing split air conditioner.

5 Concluding Remarks

This was possibly the first attempt to develop Environment Friendly fuzzy logic based air conditioner. We are aware of several limitations in the prototype model and there is need for improvement. The development of Environment friendly Neuro—Fuzzy Air conditioner with additional sensors for moisture and humidity control is underway.

References

1. Ross, T.J.: Fuzzy Logic with Engineering Applications, 2nd edn. John Wiley and Sons, USA
2. PIC16F7X Data Sheet
3. LM35 Data Sheet
4. TGS4162 Data Sheet
5. Mazidi, M.A., Mazidi, J.G.: The 8051 Microcontroller and Embedded Systems. 2nd edn (Low price edition) (2002)
6. Arthur, J.H., Beard, J.T., Bolton, C.: Simplified analytical modeling of an air-conditioner with a positive displacement compressor. IEEE publications (2002)
7. Zain', M.D., Abdullah2, N.H., Hanafi3, A.H.M.: Microcontroller-based energy saving control for air—conditioning system using fuzzy logic approaching: an Overview. In: Proceedings of 4"h Student Conference on Research and Development (SCOReD 2006), Shah Alam, Selangor, Malaysia, 27–28 June 2006
8. Chin, L., Mital, D.P.: Fuzzy logic and neural networks. IEEE (2000)

A Fuzzy Stochastic Programming Approach for Multi-level Capacitated Lot-Sizing Problem Under Uncertainty

N. Sahebjamnia and S. A. Torabi

Abstract This chapter develops a fuzzy stochastic multi-objective linear programming model for a multi-level, capacitated lot-sizing problem in a mixed assembly shop. The proposed model aims to minimize the total cost consisting of total variable production cost, inventory cost, backorder cost and setup cost while maximizing the resource utilization rate simultaneously. To cope with inherent mixed fuzzy-stochastic uncertainty associated with objective functions coefficients, e.g., setup, holding, and backorder costs, they are treated as fuzzy stochastic parameters with discrete probability density function. To validate the proposed model and the expediency of the proposed solution method, a number of randomly generated test problems of different sizes are solved. The results demonstrate the usefulness of the proposed model and its solution approach.

1 Introduction

The Multi-Level Capacitated Lot-Sizing Problem (ML-CLSP) is a well-known formulation for big-bucket multiple periods, multiple items lot-sizing problems in capacitated production systems often faced in a practical production planning setting. Setup times/costs occur in such a system when changing from one item to another one. The objective is to find an optimal production plan consisting of the required setup periods and the time-phased production and inventory quantities at each level of a complex bill-of-material (BOM) structure over a finite planning horizon while minimizing the total cost and ultimately meeting all customer demands by the end of concerned horizon (see, Refs. [1–3]).

N. Sahebjamnia (✉) · S. A. Torabi
Department of Industrial Engineering, University of Tehran, Tehran, Iran
e-mail: n.sahebjamnia@ut.ac.ir

S. A. Torabi
e-mail: satorabi@ut.ac.ir

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_25,

There are a considerable amount of works in the context of ML-CLSP. Among them, Ref. [4] propose a hybrid heuristic method to solve the dynamic ML-CLSP. The proposed technique combines the capability of the Lagrangean relaxation technique to decompose the original hard-to-solve problem into a number of smaller sub-problems, and the intensive search capability of the simulated annealing to solve the problem efficiently. To show the performance of the proposed approach, the results are compared by using the benchmark problems available in the literature.

Helber and Sahling [5] present an optimization-based solution approach for the dynamic ML-CLSP with positive lead times. They obtain flexible, accurate and relatively fast results by solving a series of mixed-integer programs in an iterative fix-and-optimize algorithm.

By considering more criteria and limitations, the CLSP family becomes even much more difficult. Consequently, various solution techniques especially those of meta-heuristics have been proposed in the past decades (see, Refs. [6–8]).

Recently, Wu et al. [3] propose two new mixed integer programming models for capacitated multi-level lot-sizing problems with backlogging, whose linear programming relaxations provide good lower bounds on the optimal solution value. A new effective optimization framework is also proposed that achieves high quality solutions in reasonable computational time. Computational results show that the proposed optimization framework is superior to other well-known approaches on several important performance dimensions.

The dynamic nature of ML-CLSP imposes a high degree of uncertainty which significantly increases the complexity of the problem. Hence, there has been a growing attention in recent years on developing new CLSP-related formulations and solution methods while taking into account the uncertain nature of the input data through fuzzy or stochastic programming techniques leading to more real production planning setting. From this viewpoint, Lee et al. [9] propose three different lot sizing algorithms with fuzzy demands. Lee et al. [10] also incorporate fuzzy demands into a part-period balancing lot-sizing algorithm. Pappis and Karacapilidis [11] derive an appropriate number of production runs and lot sizes for multi-product lot-sizing problem with fuzzy demands. Lee and Yao [12] use the fuzzy concept to deal with economic production quantity while demand and production quantity are uncertain. Chang et al. [13] modify the economic reorder point problem by introducing a fuzzy backorder quantity. The results offer a better sense of using economic fuzzy quantities. Pai [14] applies the fuzzy set theory to solve the CLSP with fuzzy capacity for which the well-known CLSP heuristic of Ref. [15] is modified to deal with fuzzy capacity, Meaningful and practical interpretations of fuzzy lot size are provided through the use of fuzzy integrals and possibility theory. Yan et al. [16] formulate a lot-sizing problem with fuzzy unit profits, fuzzy capacities and fuzzy demands as a credibility measure based fuzzy programming model. They derive the crisp equivalent model when the fuzzy parameters are characterized by trapezoidal fuzzy numbers.

In the case of stochastic CLSP, Ref. [17] solve a variant of the stochastic dynamic CLSP, where item and period specific backorder costs as well as extendible production capacities are considered. Brandimarte [18] considers the stochastic CLSP where the uncertainty of the demand is represented through a scenario tree. In this case, demands are modeled as discrete random variables. Tempelmeier and Herpers [19] propose a formulation of the dynamic CLSP under random demand, when the performance is measured in terms of a fill rate per cycle which is a popular performance measure in industry. Tempelmeier [20] develop a column generation heuristic to deal with the dynamic multi-item CLSP under random period demands (SCLSP).

In this chapter, a new multi-item ML-CLSP model is formulated for shops with mixed assembly products. For more information about the configuration of a mixed assembly shop, the interested reader may consult with Ref. [21]. More importantly, to counter ambiguous and imprecise nature of the input data in a mixed assembly shop, the parameters of the proposed model are treated as Fuzzy Stochastic Variables (FSVs) which are usually characterized experimentally or mentally (i.e., by using the available objective data as well as the subjective knowledge of the field experts) in practice. Accordingly, a bi-objective fuzzy stochastic multi objective linear programming (FSMOLP) model is formulated for the concerned ML-CLSP to cope with these ill-known parameters appropriately. For converting the original FSMOLP model to a crisp equivalent one, since FSVs exist in both objective functions and constraints, the imprecise constraints and objective functions are transformed to their crisp equivalent forms, respectively. Finally, by applying an effective aggregate function, the crisp equivalent of original bi-objective FSMOLP model is changed to a single objective model. As such, the main contributions of this chapter can be highlighted as follows:

1. Formulating a new FSMOLP model for ML-CLSP in a mixed assembly shop environment which accounts for the ambiguous and imprecise nature of the input data.
2. Proposing a novel algorithm to find the crisp equivalent of original FSMOLP model.
3. Defining a new fuzzy stochastic ranking membership to avoid from the complexity of defuzzifying and derandomizing processes.

2 Fuzzy Stochastic (Random) Programming

In a fuzzy stochastic programming problem, fuzziness and randomness could happen at the same time because the nature of two uncertainties is different. Fuzzy number represents the incomplete, imprecise information. The stochastic variable represents randomness or chance of events. Fuzzy numbers can vary randomly in real life. For example, the estimation of tolerance of assembling items can be estimated as a fuzzy number. These values can vary from time to time, period-by-period in the production lot. Thus, values could be modeled as FSV. In real case, fuzzy stochastic linear programming arises in several circumstances. The parameters of linear programs could be FSV, because they depend on many features. However, it is hard to determine exactly the values of these parameters. Moreover, the features, which are fluctuating due to uncertain environment, could make these parameters vary. These circumstances often happen in big-bucket planning, development strategies [22] engineering design [23], and financial modeling [24], in which the described conditions (objectives, constraints, coefficients) cannot be determined precisely and certainly.

Several methods have been developed in the literature to deal with the fuzzy stochastic models involving both ambiguousness and randomness of the coefficients and parameters in objective functions and constraints. Hop [25] converts the FSMOLP into the corresponding deterministic MOLP problem by using superiority and inferiority measures of the FSV. Two issues are notable in his work. First he defines FSV as fuzzy number with discrete distribution function while we consider a fuzzy number with stochastic parameter (with discrete or continues distribution function). As follow, Converting FSMOLP model into crisp single objective LP model would be increased the complexity of the models. Hop adds uncertain constraints into objective function and change the structure of the model completely. Here, we propose a novel FSMOLP approach to convert the proposed FSMOLP into an equivalent crisp single objective programming model. To this end, we propose a hybrid approach inspiring by the approaches developed in Refs. [25–27].

Jimenez [27] general ranking method which can be applied to different kinds of membership functions such as triangular, trapezoidal and nonlinear ones in both symmetric and a symmetric forms and it is based on the strong mathematical concepts such as expected interval and expected value of fuzzy numbers.

- Torabi and Hassini [26] provided a novel fuzzy approach (TH method) which can find an efficient compromise solution to solve multi-objective possibilistic mixed integer linear program models.
- Hop [25] shows by converting the fuzzy stochastic linear program into the conventional deterministic linear program, fuzzy stochastic models could be solved easily by standard optimization packages.

However a well-known approach to convert the fuzzy linear program (FLP) into the conventional deterministic linear program (LP) is the method of ordering fuzzy numbers. Similarly, the fuzzy stochastic linear program (FSLP) has also been converted by defuzzifying and derandomizing FSV in two manners i.e. sequentially or simultaneously. In the sequential approaches, the defuzzifying process is performed first. Then the derandomizing process is done later while in the simultaneously approach both derandomizing and defuzzifying done concurrently by calculating the expected value of FSV (see Ref. [28]). The defuzzifying process often utilizes ranking operations of fuzzy sets to defuzzify partly FSV. The derandomizing process implements stochastic programming techniques, such as, the chance constrained programming approach or the two stage programming approach (see Ref. [29]). The main disadvantage of the sequential method is to create a large number of additional constraints and variables.

An illustrated example of FSMOLP could be the case of the lot-sizing (e.g. total cost and resource utilization). These objective functions consist of FSV since available resources, demand and coefficients (both in objective and constraint) can also be modeled as FSV because the ambiguous insights with statistical data in different environmental conditions contribute to the parameters of the model however, production output depends on process parameters (for example, worker speeds) and resource accessibility time. These parameters are fluctuating and hard to approximate precisely. Although the computation process of the expected value of the FSV might be complicated, we propose a simultaneously method that reduce significantly number of additional constraints and variables in the obtained LP. Moreover, the main advantage of the proposed approach is a new ranking method that decreases the cost of our method for defuzzifying FSV. These aspects are the main motivations of our work to simplify proposed solution process.

3 Proposed Model

In production planning problem, we seek optimal decisions for production activities (assembly or manufacturing procedure) that transform raw materials into final products. The standard production planning problem assumes that it is possible for customer demand to be satisfied completely and on time. And the objective is to fulfill customer demand at minimum total (production, inventory and setup) cost. Here the model was developed such that at the end of each period backorder could be happened for one or more final product, while at the end of the last period all of the demand should be satisfied. The following assumptions are considered in the formulation of proposed CLSP model:

A. *Assumptions*

1. Each item needs just an identical resource for assembling.
2. Available resource will be retrieved at the end of each level.
3. Several products are produced, each of which is assembled from two or more item from upper level.
4. The considered CLSP problem is big-time bucket problem that the planning horizon is finite and split into a number of periods.
5. When all tasks of a level are finished, the tasks of the lower level could be started.
6. The demand of the product is dynamic.
7. Demand of products should be satisfied at the end of planning horizon completely.
8. The backorder and inventory holding is allowed, while both of them are equal to zero at the begging and ending of the planning horizon.

9. The production system has been designed as multi-level that in each level various item should be assembled.
10. In terms of number of products/item, we consider a multi-item, multi-product production planning problem.
11. The capacity constraint is considered on available resource.
12. To assembling an item a setup cost and time is occurred. Setup time is considered as constraint in model and setup cost appeared as cost in objective function.
13. Input data are treated as fuzzy stochastic parameters each of which with identical fuzzy stochastic membership function during the planning horizon.

The following notations are used in the formulation of proposed CLSP model:

B. *Notations*

(1) *Indices*:

i Product, $i = 1,\ldots,P$
v Item, $v = 1,\ldots,V$
l level. $l = 1,\ldots,L$
j Machine, $j = 1,\ldots,J$
h Period (time bucket), $h = 1,\ldots,H$

(2) *Parameters*:

$\tilde{\tilde{C}}_{ilvj}^{Sh}$ Setup cost for item v of level l of product i on machine j in period h
$\tilde{\tilde{C}}_{i}^{Hh}$ Holding cost of product i in period h
$\tilde{\tilde{C}}_{i}^{Bh}$ Back order cost of product i in period h
$\tilde{\tilde{d}}_{i}^{h}$ Demand of product i at the end of period h
RT_{ilvj} The production rate of item v of level l of product i on machine j
t_{ilvj}^{h} The processing time of item v of level l of product i on machine j at period h
b_{ilv} The number of needed for item v of level l of product i for producing a product i
ST_{ilvj} Setup time for item v of level l of product i on machine j
$\tilde{\tilde{a}}_{ilv}$ The amount of resource needed to produce one unit of item v of level l of product i
$\tilde{\tilde{R}}_{j}$ The accessible amount of resource j
λ_{ilvj} 1, If item v of level l of product I is need machine j, 0: otherwise

(3) *Decision variable*:

I_{i}^{h} Inventory level of product i at the end of period h
B_{i}^{h} Back order level of product i at the end of period h
Q_{i}^{h} The lot size of product i at period h

C. *Fuzzy Stochastic ML-CLSP model*

In practice, imprecision parameters of the proposed ML-CLSP complicate decision-making. So, in this chapter, demand and process parameters have been assumed to be fuzzy stochastic parameters. Formally, we are given:

- A set of products according to the customers demand should be delivered at the end of each period. Producers, market, customers and competitors situation might effect on the amount of demand domain during the planning horizon. Hence, the demand considered as imprecise value in recent researches.
- Due to the ambiguousness of real environmental situation and complicated structure of the mixed assembly shops, determining the amount of needed resource to assemble during the assembling process is not possible exactly. Consequently, consuming and available resource at each level is considered as fuzzy stochastic parameter. (It could be considered as assembling disruptions or risks)

Complicated structure of the shop with mixed assembly product and dynamic nature of production planning problem, impose high degree of uncertainty in CLSP decisions. Ho [30] classified the uncertainty affecting the real world production systems into two groups: (1) environmental uncertainty and (2) system uncertainty. In the context of production planning, environmental uncertainty is related to uncertainties in demand and backorder penalty. System uncertainty includes the uncertainties within the production system such as, uncertainty in production cost and actual capacity or consumption of resources. In addition big-bucket nature of the CLSP model, increase the impact of uncertainty in this problem.

However, setting precise amount of the parameters for the production planning models such as lot-sizing is not reasonable or possible. So, researchers tend to develop the models and solutions to tackle the imprecision and ambiguousness situation of the real case of lot-sizing problem (see Refs. [31–33]).

Although both of stochastic and fuzzy approaches separately are worthy techniques, applying fuzzy stochastic methods is more desirable. We explore four main issues as the major reasons of using the FSV instead of deterministic, fuzzy or stochastic variable in the proposed CLSP model as follows:

1. *Lack of Experimental data*: In many real cases there is not enough experimental data to obtain the actual and exact distribution characteristics of parameters. Consequently, by predicting a stochastic distribution large amount of data would be losing.
2. *System uncertainty*: System parameters could be determined subjectively (i.e. fuzzy parameters with identical membership function) or objectively (i.e. stochastic parameters with identical distribution). While in both way, by changing unpredictable feature of the system, we will miss some values of the parameters having main effect on our decision.

3. *Nature of parameter*: Determining ambiguous production planning parameters is not possible essentially. For example, separating production variation cost and setup cost is not possible when using the same workers.
4. *Environmental uncertainty*: It is not clear to distinguish disruption and uncertainty in production planning problems. Environmental factors sometimes is interpreted as unreliable and uncertain parameters creating production planning interruption, whereas uncertainty can be explained as matching environmental factors between producers and customers (as two definite parameters categories). We believe that two issues are important in discussing the production panning parameters: the outcome of environmental factors impact and expectation of shop floors factors on the production planning parameters. Fuzzy stochastic parameters with both fuzzy and stochastic characteristics could have applied as effective tools to deal with these two issues.

Consequently, due to the unavailability or incompleteness of data in real case, especially in big-time bucket production planning problem (with long-term horizon), the parameters embedded in such CLSP model have an imprecise and ambiguous nature. So, in order to model the lack of knowledge about these parameters we use appropriate fuzzy stochastic membership function. We also consider a decision horizon including multiple periods in the proposed model and so the amount of the lot sizes, backorder and holding are determined according to fuzzy stochastic parameters such as demand, four types of cost parameters and resource utility at each period. The fuzzy stochastic ML-CLSP model has been developed as follows:

$$\begin{aligned}\text{Min}\ &\sum_{h=1}^{H}\sum_{i=1}^{P}\sum_{l=1}^{L}\sum_{v=1}^{V}\tilde{\tilde{C}}_{ilv}^{Sh}\cdot Y_i^h+\sum_{h=1}^{H}\sum_{i=1}^{P}\tilde{\tilde{C}}_i^{Hh}\cdot I_i^h\\&+\sum_{h=1}^{H}\sum_{i=1}^{P}\tilde{\tilde{C}}_i^{Bh}\cdot B_i^h+\sum_{h=1}^{H}\sum_{i=1}^{P}\sum_{l=1}^{L}\sum_{v=1}^{V}\tilde{\tilde{C}}_{ilv}^{Ph}\cdot Q_i^h\cdot b_{ilv}\end{aligned} \tag{1}$$

$$\text{Min}\ \sum_{h=1}^{H}\sum_{l=1}^{L}\sum_{j=1}^{J}\left(\tilde{\tilde{R}}_j-\sum_{i=1}^{P}\sum_{v=1}^{V}b_{ilv}\cdot Q_i^h\cdot\tilde{\tilde{a}}_{ilv}\cdot\lambda_{ilvj}\right) \tag{2}$$

Subject to:

$$I_i^{h-1}-B_i^{h-1}+Q_i^h=I_i^h-B_i^h+\tilde{\tilde{d}}_i^h\quad\forall i,h\geq 2,3,\ldots,H \tag{3}$$

$$Q_i^1-I_i^1+B_i^1-\tilde{\tilde{d}}_i^h=0\quad\forall i \tag{4}$$

$$I_i^H=B_i^H=I_i^0=B_i^0=0\quad\forall i \tag{5}$$

$$\sum_{i=1}^{P}\sum_{v=1}^{V} b_{ilv}\cdot Q_i^h\cdot \tilde{\tilde{a}}_{ilv}\cdot \lambda_{ilvj}\leq \tilde{\tilde{R}}_j \quad \forall h,l,j \tag{6}$$

$$Q_i^h \leq M\cdot Y_i^h \quad \forall i,h \tag{7}$$

$$Y_i^h = \{0,1\} \quad \forall i,h \tag{8}$$

$$Q_i^h, B_i^h, I_i^h \in \text{Integer} \quad \forall i,h \tag{9}$$

The objective of the proposed fuzzy stochastic ML-CLSP is to determine the lot sizes of the products whereas satisfying the cost and utility of the mixed assembly shop. There are two objectives associated with the model: one is to minimize the costs related to the assembling operations, resource set ups, backorder and inventory; the other is to maximize the utilization of the available resource in the mixed assembly shop. In Eq. (1), four terms indicate, respectively, resource setup cost, holding cost, backorder cost and production variation cost. In the first term, $\tilde{\tilde{C}}_{ilvj}^{Sh}$ is the setup cost of resource j and Y_i^h is an auxiliary 0–1 integer variable which assumes a value of one if product i is produced in period, and zero otherwise. In the second and third term, $\tilde{\tilde{C}}_i^{Hh}$ and $\tilde{\tilde{C}}_i^{Bh}$ are the holding and backorder cost for a unit of product i that holding or non-delivery at the end of period. In the fourth term, $\tilde{\tilde{C}}_{ilv}^{Ph}$ is per unit cost of performing operation item v of level l of product i. This includes both the assembling and consumed resource cost. In the second objective function $\tilde{\tilde{R}}_{lj}^h = \sum_{i=1}^{P}\sum_{v=1}^{V} b_{ilv}\cdot Q_i^h\cdot \tilde{\tilde{a}}_{ilv}\cdot \lambda_{ilvj}$. The $\tilde{\tilde{R}}_j$ is the total amount of resource j that is available during planning horizon. We split the mixed assembly shop into a number of levels including various item from different products. When any level is started, all related item will be assembled until the next level begin. At the start of next level all resources will be retrieved and $\tilde{\tilde{R}}_j$ is available again. In the second objective, we try to close the amount of total available resource with consuming resource $(\tilde{\tilde{R}}_{lj}^h)$.

The equation constraint sets (3)–(9) are in effect. The Eqs. (3)–(5) balance the amount of the backorder, holding and lot sizes of all products from $h = 2$ to end of the planning horizon and $h = 1$ respectively. The third formulation shows that at the start and end of the planning horizon backorder and holding is equal to zero. The fuzzy stochastic capacity constraint, ensure that the overall consumption must remain lower than the available capacity. The available resource $\tilde{\tilde{R}}_j$ would be retrieve at the end of each level for each resource. Constraint (7) production equation, grantee that if any product is produced in each period $(Q_i^h > 0)$ the decision variable Y_i^h is become one, and zero otherwise. And variable constraints (8)–(9), shows the type of each decision variable.

4 Objective Function with Fuzzy Stochastic Coefficient

In this section we try to convert the fuzzy stochastic coefficient of imprecise objective functions into the crisp ones such that using maximum performance of the fuzzy stochastic number. Assumed that P is a fuzzy stochastic programming problem as follow:

$$\begin{aligned} P:\min \mathrm{Z}_h &= \tilde{\tilde{c}}_h x \ \text{ for } \ h = 1,\ldots,H \\ &\text{s.t.} \\ &\mathrm{x} \in \mathrm{X} \end{aligned}$$

$\tilde{\tilde{c}}_h$ is a fuzzy stochastic coefficient of the objective function (h-th) with discrete probability density function and X is feasible space of the problem.

Definition 1 For fuzzy stochastic number ($\tilde{\tilde{c}}_h$) we define the upper and lower bound of $\tilde{\tilde{c}}$ ($\tilde{c}_l$ and $\tilde{c}_u$) as follow:

Lower bound of $\tilde{\tilde{c}}(\tilde{c}_l)$: The maximum movement that $\tilde{\tilde{c}}$ could move to left hand side according to the distribution deviation and multiple γ that determine by decision makers ($\gamma\sigma^2_{c_h}$).

$$\tilde{c}_l = \begin{cases} 0 & (-\infty,(c_1-\gamma\sigma^2_{c_1})] \\ \frac{x-(c_1-\gamma\sigma^2_{c_1})}{(c_2-\gamma\sigma^2_{c_2})-(c_1-\gamma\sigma^2_{c_1})} & [(c_1-\gamma\sigma^2_{c_1}),(c_2-\gamma\sigma^2_{c_2})] \\ 1 & [(c_2-\gamma\sigma^2_{c_2}),(c_3-\gamma\sigma^2_{c_3})] \\ \frac{(c_4-\gamma\sigma^2_{c_4})-x}{(c_4-\gamma\sigma^2_{c_4})-(c_3-\gamma\sigma^2_{c_3})} & [(c_3-\gamma\sigma^2_{c_3}),(c_4-\gamma\sigma^2_{c_4})] \\ 0 & [(c_4-\gamma\sigma^2_{c_4}),+\infty) \end{cases} \tag{10}$$

Upper bound of $\tilde{\tilde{c}}(\tilde{c}_u)$: The maximum movement that $\tilde{\tilde{c}}$ could move to right hand side according to the distribution deviation and multiple γ that determine by decision makers ($\gamma\sigma^2_{c_h}$).

$$\tilde{c}_u = \begin{cases} 0 & (-\infty,(c_1+\gamma\sigma^2_{c_1})] \\ \frac{x-(c_1+\gamma\sigma^2_{c_1})}{(c_2+\gamma\sigma^2_{c_2})-(c_1+\gamma\sigma^2_{c_1})} & [(c_1+\gamma\sigma^2_{c_1}),(c_2+\gamma\sigma^2_{c_2})] \\ 1 & [(c_2+\gamma\sigma^2_{c_2}),(c_3+\gamma\sigma^2_{c_3})] \\ \frac{(c_4+\gamma\sigma^2_{c_4})-x}{(c_4+\gamma\sigma^2_{c_4})-(c_3+\gamma\sigma^2_{c_3})} & [(c_3+\gamma\sigma^2_{c_3}),(c_4+\gamma\sigma^2_{c_4})] \\ 0 & [(c_4+\gamma\sigma^2_{c_4}),+\infty) \end{cases} \tag{11}$$

According to $\tilde{c}_l$ and $\tilde{c}_u$ the membership function of each objective function would be obtained (12) and (13) and shown in Fig. 1 for minimizing and maximizing respectively.

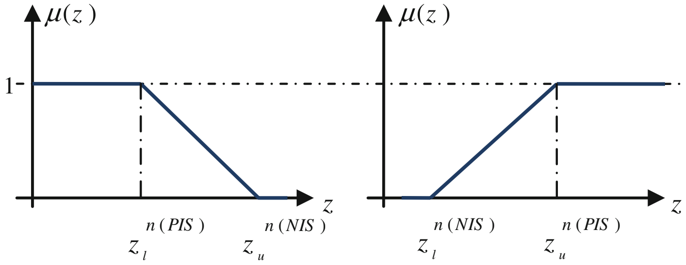


Fig. 1 Membership linear functions of the objectives

$$\mu_n^z(z) = \begin{cases} 0 & Z > z_u^{n(NIS)} \\ \frac{z_l^{n(PIS)} - Z}{z_u^{n(NIS)} - z_l^{n(PIS)}} & z_l^{n(PIS)} \le Z \le z_u^{n(NIS)} \\ 1 & Z \le z_l^{n(PIS)} \end{cases} \tag{12}$$

$$\mu_n^z(z) = \begin{cases} 1 & Z > z_u^{n(PIS)} \\ \frac{Z - z_u^{n(PIS)}}{z_u^{n(PIS)} - z_l^{n(NIS)}} & z_l^{n(NIS)} \le Z \le z_u^{n(PIS)} \\ 0 & Z \le z_l^{n(NIS)} \end{cases} \tag{13}$$

Often the researchers to find the positive ideal solution (PIS) and negative ideal solution (NIS) solve the MOLP problem $2N$ time (N number of the objective functions). To avoid iterative procedure and reduce running time, we define $z_l^{n(PIS)}$ and $z_u^{n(NIS)}$ and $z_u^{n(PIS)}$ and $z_l^{n(NIS)}$ for minimization and maximization of objective function respectively as Definitions 2 and 3.

Definition 2 Define $z_l^{n(PIS)}$ and $z_u^{n(NIS)}$ for minimization: $z_l^{n(PIS)} = \min z_l^n$ (x) (For each n individually)

$$z_u^{n(NIS)} = \max_{\hat{x}_n} \{ z_u^n(\hat{x}_n) \}$$

($\hat{x}$ is the solution of z_u^n, for maximizing each n individually).

Definition 3 Define $z_u^{n(PIS)}$ and $z_l^{n(NIS)}$ for maximization: $z_u^{n(PIS)} = \max z_u^n$ (x) (For each h individually)

$$z_l^{n(NIS)} = \min_{\hat{x}_n} \{ z_l^n(\hat{x}_n) \}$$

($\hat{x}$ is the solution of z_l^n, for minimizing each n individually) where z_l^n and z_u^n is formulated as (14) and (15).

$$\begin{aligned} Z_l^n &= \min EV(\tilde{\tilde{c}}_l^n) \cdot x \\ &\text{s.t.} \\ &x \in X \end{aligned} \tag{14}$$

$$\begin{aligned} Z_u^n &= \min EV(\tilde{\tilde{c}}_u^n) \cdot x \\ &\text{s.t.} \\ &x \in X \end{aligned} \tag{15}$$

Torabi and Hassini [26], proposed a novel interactive fuzzy approach to solve multi objective linear programming problem and finding a preferred compromise solution. We applied their aggregate function to solve P as follow:

$$\begin{aligned} &\max \eta(z) = \varsigma \eta_0 + (1-\varsigma) \sum_n \theta_n \mu_n(z) \\ &\text{s.t.} \\ &\eta_0 \le \mu_n(z)\; n = 1, \ldots, N \\ &z \in F(z) \\ &\varsigma \in [0, 1] \end{aligned} \tag{16}$$

where $\mu_n(z)$ and $\eta_0 = \min_n\{\mu_n(z)\}$ denote the satisfaction degree and minimum satisfaction degree of objective functions, respectively. This formulation has a new achievement function defined as a convex combination of the lower bound for satisfaction degree of objectives (η_0), and the weighted sum of these achievement degrees $(\mu_n(z))$ to ensure yielding an adjustable balanced compromise solution. Moreover, θ_n and ς indicate the relative importance of the nth objective function and the coefficient of compensation, respectively. The θ_n parameters are determined by the decision maker based on her/his preferences such that $\sum_n \theta_n = 1,\ \theta_n > 0$. Also, ς controls the minimum satisfaction level of objectives as well as the compromise degree among the objectives implicitly. That is, the proposed formulation is capable of yielding both unbalanced and balanced compromised solutions for a given problem instance based on the decision maker's preferences through adjusting the value of parameter ς [26].

By defining upper and lower bound of fuzzy stochastic coefficient (Definition 1) we convert the imprecision objective function into crisp one for single or multi objective fuzzy stochastic programming problem.

5 The Proposed Fuzzy Solution Approach

To solve the FSMOLP, several methods have been proposed in the literature. By considering the imprecision situation of the production environment as fuzzy stochastic parameters, both experimentally nature of the data and satisfaction level of decision makers are take into account simultaneously. But it would increase the complexity of the solution method. However, in this chapter to solve the proposed fuzzy stochastic CLSP model we propose an interactive fuzzy stochastic solution approach by developing a new approach that derandomizing and defuzzifying the FSV simultaneously. The proposed method can be highlighted as follows.

Step 1: Determine the appropriate fuzzy membership function and probability distributions for imprecise parameters and formed the fuzzy stochastic *CLSP* model.

Step 2: Convert the vague constraint into the crisp ones by using the expected value [27] of corresponding imprecise parameters by using following equation:

$$\chi(1+\psi-\xi)\cdot E_1^{\tilde{\bar{A}}_1}+(1-\chi)(1-\psi+\xi)\cdot E_2^{\tilde{\bar{A}}_1} \leq \chi(1+\psi-\xi)\cdot E_1^{\tilde{\bar{A}}_2}+(1-\chi)(1-\psi+\xi)\cdot E_2^{\tilde{\bar{A}}_2} \quad (17)$$

Step 3: Convert fuzzy objective functions into the crisp ones by using the expected value of the upper and lower bound coefficient (Eqs. (10) and (11)) and find the Z_l^n, Z_u^n (Eqs. (14) and (15)).

Step 4: Determine the membership function of the objective functions by using (12) and (13) and according to the Definitions 2 and 3.

Step 5: Convert the equivalent crisp MOLP model into a single ones using Ref. [26] aggregation functions (16) which resulting the efficient solutions.

To run the proposed algorithm the decision maker should specify the value of the: feasibility degree of the constraints (χ); the deviation of the objective function coefficient (γ); coefficient of compensation (ς); relative importance of the fuzzy goals θ_n. Jimenez [34] proposed a three steps procedure to maximize the satisfaction (feasibility) degree. The deviation of the objective function coefficient (γ) is set by the DM. Also, the coefficient of compensation and relative importance of the fuzzy goals are specified by decision makers generally. If the decision makers are not satisfied with the obtained results, another solution will be provided by changing these values [35].

6 Conclusion

This study propose a new fuzzy stochastic multi-objective linear programming (FSMOLP) model to formulate the capacitated lot-size production planning problem in multi-item, multi-level, multi-product and multi-period mixed assembly shop. To deal with imprecision and vagueness nature of data in real cases, we have used appropriate fuzzy stochastic membership function for the CLSP formulation. An interactive fuzzy stochastic approach has been developed. First, imprecise constraints convert into crisp one and then convert fuzzy stochastic objective functions into the crisp ones by using the expected value of the coefficient upper and lower bound. Finally, transfer the equivalent crisp MOLP model into a single ones using appreciate aggregation function which resulting the efficient solutions.

The numerical experiments indicate that the proposed method is very promising fuzzy stochastic approach to solve fuzzy stochastic CLSP model with discrete probability density function. Moreover, it would results efficient solutions based on the decision maker's preferences along with offering appropriate effective parameters to provide different solutions. This approach can also be used for solving other fuzzy stochastic MOLP models due to its computationally advantages.

References

1. Absi, N., Kedad-Sidhoum, S.: The multi-item capacitated lot-sizing problem with setup times and shortage costs. Eur. J. Oper. Res. **185**, 1351–1374 (2008)
2. Toledo, F.M.B., Armentano, V.A.: A Lagrangian-based heuristic for the capacitated lot-sizing problem in parallel machines. Eur. J. Oper. Res. **175**, 1070–1083 (2006)
3. Wu, T., Shi, L., Geunes, J., Akartunali, K.: An optimization framework for solving capacitated multi-level lot-sizing problems with backlogging. Eur. J. Oper. Res. **214**, 428–441 (2011)
4. Ozdamar, L., Barbarosoglu, G.: An integrated Lagrangean relaxation-simulated annealing approach to the multi-level multi-item capacitated lot sizing problem. Int. J. Prod. Econ. **68**, 319–331 (2000)
5. Helber, H., Sahling, F.: A fix-and-optimize approach for the multi-level capacitated lot sizing problem. Int. J. Prod. Econ. **123**, 247–256 (2010)
6. Okhrin, I., Richter, K.: An O(T3) algorithm for the capacitated lot sizing problem with minimum order quantities. Eur. J. Oper. Res. **211**, 507–514 (2011)
7. Narayanan, A., Robinson, P.: Efficient and effective heuristics for the coordinated capacitated lot-size problem. Eur. J. Oper. Res. **203**, 583–592 (2010)
8. Jans, R., Degraeve, Z.: Meta-heuristics for dynamic lot sizing: a review and comparison of solution approaches. Eur. J. Oper. Res. **177**, 1855–1875 (2007)
9. Lee, Y.Y., Kramer, B.A., Hwang, C.L.: A comparative study of three lot-sizing methods for the case of fuzzy demand. Int. J. Oper. Prod. Manage. **11**, 72–80 (1991)
10. Lee, Y.Y., Kramer, B.A., Hwang, C.L.: Part-period balancing with uncertainty: a fuzzy sets theory approach. Int. J. Oper. Prod. Manage. **28**, 1771–1778 (1990)
11. Pappis, C.P., Karacapilidis, N.I.: Lot size scheduling using fuzzy numbers. Int. Trans. Oper. Res. **2**, 205–212 (1995)
12. Lee, H.M., Yao, J.S.: Economic production quantity for fuzzy demand quantity and fuzzy production quantity. Eur. J. Oper. Res. **109**, 203–211 (1998)
13. Chang, S.C., Yao, J.S., Lee, H.M.: Economic reorder point for fuzzy backorder quantity. Eur. J. Oper. Res. **109**, 183–202 (1998)
14. Pai, P.F.: Capacitated lot size problems with fuzzy capacity. Math. Comput. Model. **38**, 661–669 (2003)
15. Florian, M., Klein, M.: Deterministic production planning with concave cost and capacity constraint. Manage. Sci. **18**, 12–20 (1971)

16. Yan, W., Zhao, J., Cao, Z.: Fuzzy programming model for lot sizing production planning problem. In: Fuzzy Systems and Knowledge Discovery, pp. 285–294. Springer, Berlin (2005)
17. Sox, C., Muckstadt, J.: Optimization-based planning for the stochastic lot-scheduling problem. IIE Trans. (Institute of Industrial Engineers) **29**, 349–357 (1997)
18. Brandimarte, P.: Multi-item capacitated lot-sizing with demand uncertainty. Int. J. Prod. Res. **44**, 2997–3022 (2006)
19. Tempelmeier, H., Herpers, S.: ABCβ-a heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint. Int. J. Prod. Res. **48**, 5181–5193 (2010)
20. Tempelmeier, H.: A column generation heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint. Omega **39**, 627–633 (2011)
21. Mahdavi, I., Shirazi, B., Solimanpur, M., Sahebjamnia, N.: Lot size approximation based on minimising total delay in a shop with multi-assembly products. Int. J. Prod. Res. **47**, 2685–2703 (2009)
22. Luhandjula, M.K., Gupta, M.M.: On fuzzy stochastic optimization. Fuzzy Sets Syst. **81**, 47–55 (1996)
23. Shih, C.J., Wangsawidjaja, R.A.S.: Mixed fuzzy-probabilistic programming approach for multi-objective engineering optimization with random variables. Comput. Struct. **59**, 283–290 (1996)
24. Zmeškal, Z.: Value at risk methodology under soft conditions approach (fuzzy-stochastic approach). Eur. J. Oper. Res. **161**, 337–347 (2005)
25. Van Hop, N.: Solving fuzzy (stochastic) linear programming problems using superiority and inferiority measures. Inf. Sci. **177**, 1977–1991 (2007)
26. Torabi, S.A., Hassini, E.: An interactive possibilistic programming approach for multiple objective supply chain master planning. Fuzzy Sets Syst. **159**, 193–214 (2008)
27. Jiménez, M.: Ranking fuzzy numbers through the comparison of its expected intervals. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **4**, 379–388 (1996)
28. Iskander, M.G.: A suggested approach for possibility and necessity dominance indices in stochastic fuzzy linear programming. Appl. Math. Lett. **18**, 395–399 (2005)
29. Ruszczynski, A., Shapiro, A.: Stochastic Programming. Handbooks in Operations Research and Management Science, vol. 10. Elsevier, Amsterdam (2003)
30. Ho, C.-J.: Evaluating the impact of operating environments on MRP system nervousness. Int. J. Prod. Res. **27**, 1115–1135 (1989)
31. Guan, Y., Liu, T.: Stochastic lot-sizing problem with inventory-bounds and constant order-capacities. Eur. J. Oper. Res. **207**, 1398–1409 (2010)
32. Xie, J., Zhao, X., Lee, T.S.: Freezing the master production schedule under single resource constraint and demand uncertainty. Int. J. Prod. Econ. **83**, 65–84 (2003)
33. Dellaert, N.P., Jeunet, J.: Randomized multi-level lot-sizing heuristics for general product structures. Eur. J. Oper. Res. **148**, 211–228 (2003)
34. Jiménez, M., Arenas, M., Bilbao, A., Rodríguez, M.V.: Linear programming with fuzzy parameters: an interactive method resolution. Eur. J. Oper. Res. **177**, 1599–1609 (2007)
35. Pishvaee, M.S., Torabi, S.A.: A possibilistic programming approach for closed-loop supply chain network design under uncertainty. Fuzzy Sets Syst. **161**, 2668–2683 (2010)

Context-Dependent Interpretation of Medical Data

M. Kwiatkowska and N. T. Ayas

Abstract Medical data are intrinsically context-dependent, and cannot be properly interpreted outside of their specific contexts. Therefore, data analysis, especially, secondary data analysis, such as data mining, must incorporate contextual information. This chapter discusses the need for an explicit context representation in medical data mining. It focuses on five contextual dimensions: goal orientation, interdependency of data, time sensitivity, source validity, and absent value semantics. It demonstrates context-dependent modeling based on examples of clinical data used for screening, diagnosis, and research of a serious respiratory disorder, obstructive sleep apnea (OSA). In particular, the chapter describes context-dependent interpretation for three OSA risk factors: large neck circumference, snoring, and smoking. Furthermore, it presents a conceptual framework for representation of the contextual information. This framework is based on a semiotic approach to represent multiple interpretations of data and a fuzzy-logic approach to represent vagueness of data.

1 Introduction

The notion of context and its significance in human communication have been studied by various fields, such as linguistics, artificial intelligence, cognitive science, mobile computing, software engineering, and information retrieval [1–3]. In medicine, context is central for correct diagnosis, accurate prognosis, and appropriate treatment. Medical data are context-bound, and cannot be properly

M. Kwiatkowska (✉)
Department of Computing Science, Thompson Rivers University, Kamloops, BC, Canada
e-mail: mkwiatkowska@tru.ca

N. T. Ayas
Department of Medicine, University of British Columbia, Vancouver, BC, Canada
e-mail: nayas@providencehealth.bc.ca

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_26,

interpreted outside of their context. Therefore, data analysis, especially secondary analysis such as data mining of electronic patient records (EPRs), must recognize the highly contextual nature of medical information [4]. Patients' data collected for screening, diagnosis, and evaluation of treatment are viewed as valuable resource, and they could and should be reused in data mining; however, this secondary utilization of data must incorporate an explicit representation and analysis of their contextual information.

Context is not a single and uniform notion; in fact, medical data are collected and interpreted in several contextual dimensions. This chapter focuses on five main aspects of contextual information:

(1) *Goal Orientation*. As stated in Ref. [4], "data are always produced with a given purpose and their hardness and specificity is directly tailored to that purpose." All medical data are collected in a specific clinical context for a specific purpose (e.g., screening, diagnosis, medical research). For example, data regarding habitual snoring are collected differently for screening, diagnosis, and research of obstructive sleep apnea. Screening data include a yes/no answer to a simple question: "Do you snore?" The data collected for diagnostic purposes comprise two values: frequency and intensity. The questionnaire has two specific questions: "How often do you snore?" and "How loud do you snore?" Data collected for research include the actual sound recording from overnight study. Thus, in these three contexts data have different precision and different source.
(2) *Interdependency of Data*. Medical records cannot be simply divided into a set of independent pieces of data (atoms), which can be independently interpreted. In many cases, medical data elaborate and "explain" each other [4]. For example, the recording of a blood pressure should also include information about current treatment for high blood pressure and use of antihypertensive medication. Thus, patient who takes antihypertensive medication and has arterial blood pressure measured as 135/65 (systolic/diastolic in mmHg) has a "normal" reading by is still classified as a patient with hypertension (high blood pressure).
(3) *Time Sensitivity*. In medical information systems, patient's data are collected at specific time points, which correspond to particular diagnostic and treatment steps. Thus, interpretation of specific symptoms and signs depends on the specific stage of the treatment [4]. All medical data have a temporal dimension. This means that each datum is taken at specific point of time. Temporal context is important for the interpretation of data in two ways. First, many readings may be interpreted differently depending on the time of the day or time of the year. For example, arterial blood pressure changes during the day, with characteristic "lower" values in the afternoon. Thus, the interpretation of data must consider the time of the day: morning or afternoon measure. Secondly, many readings are interpreted as sequences with specific tendencies. For example, several consecutive measures of body temperatures are compared to see if the temperature is falling or increasing.

(4) *Source Validity*. All data are obtained from specific sources, which means data may differ in regards to their credibility and validity. As stated in Ref. [5]: "The assessment of the quality of a data source is context dependent, i.e. the notions of 'good' or 'poor' data cannot be separated from the context in which the data is produced or used." In general, the sources of medical data can be divided into subjective and objective. Typical subjective data are collected from self-administered questionnaires. The objective data include laboratory results, medical images, clinical examination results, and health practitioners' notes. The objective data have usually higher credibility than subjective data. However, even a highly credible source of objective data, such as echocardiogram (ECG) monitoring, may produce invalid data when, for example, the sensors are incorrectly connected.

(5) *Absent-Value Semantics*. Medical records are often incomplete. The EPR records, self-reported questionnaires, and various medical databases may have several missing values. There are various reasons for absent values, and the "empty fields" in EPRs have different meanings depending on their context. For example, reference [4] describes the data entry practice used by ICU staff. When monitoring the stable patients, the nurses enter only the clinically significant data (significant changes) and leave the other entry fields empty (not significantly changing values). In this specific context, the sparse data entry saves time, and the data are easily interpretable by other co-workers. Thus, the absence of data in this case means insignificant information or no-changes since last entry. This type of omission requires careful analysis in the preprocessing phase of data mining. As noted in Ref. [6], missing values have different semantics, and the interpretation of absent data is highly contextual.

In our research on contextual information, we focus on the reasons for absent values. We have identified six types of omissions: logical exclusion of not applicable data (e.g., data specific to female gender is omitted from a record of a male patient), omission of insignificant data (e.g., nurses in ICU enter only the clinically significant data), intentional omission of sensitive data (e.g., patients not reporting smoking), lack of knowledge omission (e.g., patients are not aware of their snoring; therefore, they report 'not sure'), discontinuation of the data entry (e.g., patients drop out of study or treatment, and their new data are not being entered), and erroneous omission (e.g., random omission by the patient, computer system, or health practitioner).

This chapter presents a conceptual model for the representation of contextual information. We identify five aspects of context: goal-orientation, interdependency-of-data, time-sensitivity, source-validity, and absent-value semantics. Specifically, this model focuses on the modeling of imprecision within specific contexts. It concentrates on the interpretation of imprecision, and the analysis of the required precision or the allowable imprecision of data. Imprecision is highly contextual and interpretative, i.e., a statement "high body temperature" may be sufficiently precise in a specific situation or a more precise value such as "40.5 °C rectal" is needed. Thus, imprecision is a quality of specific data used in a specific context. Often, imprecise values are (or must be) sufficient, since obtaining precise data may be impossible, impractical, or too expensive.

The main motivation for our research on context-dependent interpretation of medical data is the automated and semi-automated creation of predictive models for clinical decision support systems (CDSS). Predictive models play an important role in all aspects of diagnosis, prognosis, and treatment. With the recent availability of EPRs and access to various medical databases, the creation of predictive models can be supported by machine-learning methods [7, 8]. This data-driven approach takes advantage of the availability of vast amounts of medical data. Since the data have been previously collected, processed, and stored for their primary use, the cost of their re-use is minimal. Thus, on the one hand, the secondary use of medical data reduces the cost of data acquisition. Yet, on the other hand, the secondary use of data is encumbered by inherent context-dependency of medical data. Therefore, we argue that (1) the secondary use of data requires modeling of contextual information and (2) the creation of predictive models requires contextual analysis of data. To address these two requirements, we propose a conceptual model based on a semiotic-approach for modeling contextual interpretation and a fuzzy-logic approach for modeling imprecision.

This chapter is structured as follows. Section 2 discusses various perspectives on context, and presents three examples of contextual information for medical data. Section 3 provides a representational framework for contextual interpretation. Section 4 provides conclusions, and describes an ongoing and future work.

2 Context Modeling

The term "context" has a very broad definition. For example, the Merriam-Webster dictionary defines context as "the interrelated conditions in which something exists or occurs." Depending on a discipline, the term "context" has been used with various meanings. For example, in software engineering, context is understood as any information that might be used to specify the situation of an entity. An entity is a person, a place, an object that is relevant in the interaction between the user and the software system [1]. In database systems, specific aspects of context have been represented using models such as, for example, object-oriented model, multidimensional data representation and fuzzy relational data model [7–9]. Reference [8] describes a conceptual model for the interpretation of fuzzy terms in context of the values of other related attributes. In mobile computing, context has been represented using an ontology-based approach and a case-based reasoning [9].

In medicine, the notion of context is central to all tasks. Reference [9] describes two types of contextual information: "situational" and "set-of-beliefs." The situational context provides three perspectives: (1) patient: patient's history, the type of patient's disorder, patient's response to treatment, (2) temporal: the course of the patient's disorder, and (3) clinical: specific clinical guidelines, expertise, and experience. The "set-of-beliefs" context provides a set of underlying assumptions made by the clinicians. For example, the "set-of-beliefs" may exclude a specific disorder based on the absence of specific symptoms. Contextual interpretation of medical data has been applied, for example, in orthopedics [10]. Reference [10] describes a fuzzy-based system to combine objective biomechanical data with subjective medical data.

In this section, we describe how the medical data and their contextual information are used for three purposes: screening, diagnosis, and medical research. We demonstrate these three goal-orientations using and an example of a serious respiratory disorder: obstructive sleep apnea (OSA). We believe that the contextual modeling of medical data requires somewhat deeper understanding of the underlying domain. Therefore, we present the background medical knowledge, and we give examples of specific medical predictors. Furthermore, we exemplify our discussion using various clinical data.

2.1 OSA and Its Predictors

OSA is a common, serious respiratory disorder afflicting, according to conservative studies, 2–4 % of the adult population. "Apnea" means "without breath," and occurs only during sleep, and is, therefore, a condition that might go unnoticed for years. OSA is caused by collapse of the soft tissues in the throat as the result of the natural relaxation of muscles during sleep. The soft tissue blocks the air passage and the sleeping person literally stops breathing (apnea event) or experiences a partial obstruction (hypopnea event). OSA is associated with significant risk for hypertension, congestive heart failure, coronary artery disease, myocardial infarction, stroke and arrhythmia. Furthermore, patients with OSA have higher risk during and after anesthesia, since their upper airway may be obstructed as a consequence of sedation. Since a typical symptom of OSA is daytime sleepiness, the untreated OSA patients have higher rates of vehicle accidents. In summary, OSA has multiple consequences not only for the patients, but also for their families and co-workers. Although an early diagnosis of OSA is critical, it is estimated that 82 % of men and 92 % of women with moderate-to-severe OSA have not been diagnosed.

The gold standard for the diagnosis of OSA is an overnight in-laboratory polysomnography (PSG). The most important score for OSA diagnosis is the apnea-hypopnea index (AHI), which is calculated as a number of apnea and hypopnea events per hour of sleep [11]. In the diagnosis based solely on the AHI index, apnea is classified as mild for AHI between 5 and 14.9, moderate for AHI between 15 and 29.9, and severe for AHI $\geq$ 30.

Medical literature describes several risk factors (predictors) for OSA. For example, Ref. [12] lists seven groups of risks: daytime symptoms (e.g., excessive daytime sleepiness, morning fatigue), nocturnal symptoms (e.g., habitual snoring, witnessed apneas), anatomical signs (e.g., obesity, large neck circumference), life style factors (e.g., sleep deprivation, smoking), demographic factors (e.g., age, gender, and ethnicity), coexisting medical conditions (e.g., hypertension, diabetes), and physiological factors (e.g., blood oxygen saturation). These various predictors are used in screening for OSA, diagnosis of OSA, and in medical research of OSA.

(1) *OSA Screening*. The screening process often utilizes short self-reporting questionnaires, for example, STOP-bang questionnaire [13]. This self-reporting questionnaire is used to stratify patients for unrecognized OSA before surgeries and to triage patients for diagnosis and treatment. The STOP-bang consists of eight questions considering eight risk factors: snoring, tiredness (excessive day-time sleepiness), observed apnea, blood pressure (hypertension), BMI (obesity), age (older age), neck circumference (thick neck), and gender (male). Each positive answer is counted as 1, and a score ≥ 3 indicates a higher risk of OSA. The scores are evaluated, and the patients at risk obtain special perioperative care, and later the patients are referred for further evaluation by family doctors or respiratory specialists.
(2) *OSA Diagnosis*. The OSA diagnostic process typically includes initial assessment by a family doctor, referral to sleep disorder clinic, overnight in-clinic study, and diagnosis by a respiratory specialist. The initial assessment process involves analysis of the risk factors. The final diagnosis of OSA is based on the results from overnight PSG combined with the results of a sleep questionnaire, physical examination, and the patient's medical history. The OSA diagnostic process uses data of diverse credibility and validity. The subjective data are based on self- and other-reporting using instruments such as sleep questionnaires, standardized scales, a sleep diary, and bed-partner reporting. The objective data are based on medical examinations, overnight PSGs, and home sleep studies (e.g., oximetry, actigraphy).
(3) *OSA Research*. OSA was discovered and described in 1965. Since then numerous OSA risk factors, predictors, and predictive rules have been studied by many research groups. We limit our discussion to three predictors, and describe the most relevant methods of data definition and acquisition. We exemplify our discussion using the clinical data. For the contextual analysis of neck circumference data, we use the data set from Ref. [14]. For the contextual analysis of snoring data, we use the data obtained from patients' records, and base our discussion on Refs. [12, 15–17]. For the contextual analysis of smoking data, we use the data obtained from patients' records, and base our discussion on Refs. [11, 12].

2.2 *Example 1: Neck Circumference*

Several studies indicate a short and fat neck as a characteristic sign of obstructive sleep apnea [14]. Large neck circumference is related to an upper-body fat distribution; therefore, men have larger necks than women for the same BMI values.

Moreover, taller people have usually larger necks. Some studies define specific values for neck circumference indicating risk for OSA, for example, a neck circumference above 43.2 cm for men and greater than 40.6 cm for women are often used as OSA predictors in screening [18].

We describe the following five aspects of contextual information for the concept of "thick neck:"

(1) *Goal Orientation*. The concept of thick neck is operationalized differently for the three different tasks: screening, diagnosis, and research. For the screening purposes, data is obtained from self-reported questionnaires, and the values are binary (yes/no). For example, the STOP-Bang questionnaire has a question: "Neck circumference >40 cm?" (yes/no). In the diagnostic process, there are two approaches: (1) neck circumference (collar size) is reported by the patient and (2) neck circumference is measured by a health practitioner during physical examination. In both approaches, the values are numeric. In research, the measurements are taken by the specialists conducting the studies, and the measuring process follows a well-defined procedure. For example, the study described in [14] uses a tape measure and measures the neck circumference "at the level of the cricothyroid membrane."

(2) *Interdependency of Data*. We show the interdependency between NC, gender and ethnicity based on a data set collected for clinical research [14]. This set has 239 records of adult patients: 199 males (83 %) and 40 females (17 %), 164 Asian patients (69 %), and 75 white patients (31 %). The mean NC for all patient is 39.9 cm, a mean NC for women is 36.3 (N = 40), and mean NC for men 40.7 cm (N = 199). Figure 1 uses modified boxplots to show the difference between the distribution of NC values for females and males. The modified boxplot shows in the central box the first quartile, the median, and the third quartile. The lines extend from the box to the smallest and largest observations that are not suspected outliers. The suspected outliers are plotted as individual data points, which are more than $1.5 \times IQR$ outside of the box (IQR is measured as a distance between the first and the third quartiles).
The difference in the distribution of values is even more prominent when the data are divided by gender and ethnicity, as shown in Fig. 2. The mean value for the White males is 41.5 cm, and the mean value for the Asian females is 35.3 cm. Thus, we conclude that values representing anatomical measures such as neck circumference should be analyzed separately for different subgroups.

(3) *Time Sensitivity*. Since the patients are adults, the possible changes in NC are related to gain or loss of weight, but are relatively slow and not significant.

(4) *Source Validity*. Data for NC are obtained in two ways: self-reported (subjective) and clinically measured (objective). Thus, the validity of the data significantly varies, and, as a result the secondary analysis should not combine the subjective data with the objective data.

(5) *Absent-Value Semantics*. We show the problems with missing data using 644 patients' records obtained from a respiratory clinic. The neck circumference is self-reported by the patients. However, many patients do not know their NC.

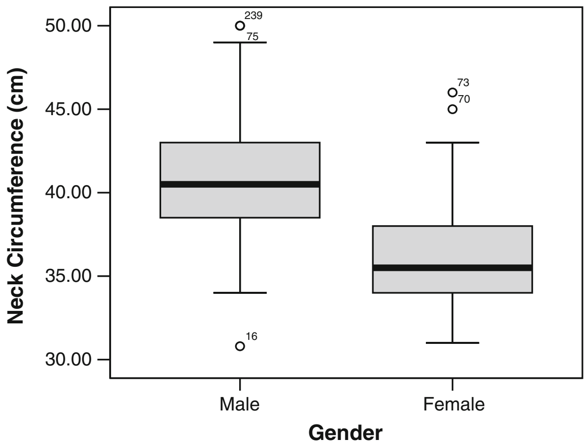


Fig. 1 Boxplots for Neck Circumference (NC) by gender and ethnicity

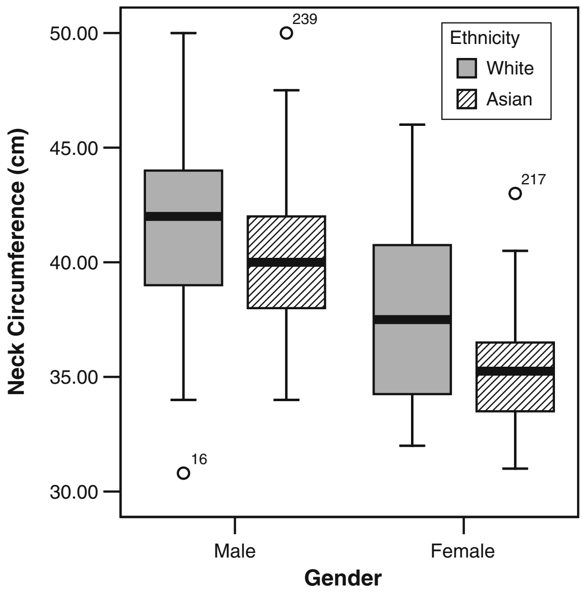


Fig. 2 Boxplots for Neck Circumference (NC) by gender and ethnicity

In the analyzed data set, 37 % of the records (240/644) have the answer "not sure" and 4 % of the records (25/644) have missing values. Thus, in total, 41 % (265/644) patients did not provide the NC data. A question about neck circumference was even more difficult for women: 75 % (141/188) female patients answered "not sure," 7 % (13/188) answers were missing. Thus, in total, 82 % (154/188) female patients did not provide the NC data. Therefore, considering that almost 50 % data are missing, the self-reported NC data should be excluded from secondary analysis.

2.3 Example 2: Snoring

Snoring is a noise produced by a sleeping individual in which the soft palate and the uvula vibrate during breathing. Loud snoring is a sign that the breathing is strained and that the airway is not completely open. The narrower the throat becomes during the inhalation, the more effort is needed to draw in air, leading to more vibration. Snoring may have several aspects, and the two most often studied are *frequency* and *intensity*. Although there are cases of OSA without snoring and many snorers do not have OSA, loud and habitual snoring is recognized as one of the most important predictors of OSA [12].

(1) *Goal Orientation*. The concept of “snoring” is operationalized differently for the three different tasks: screening, diagnosis, and research. For the purpose of screening, data are obtained from self-reported questionnaires, and the values are binary (yes/no). For example, the STOP-Bang questionnaire has a question: “Do you snore loudly (loud enough to be heard through closed door)?” (yes/no). For the purpose of diagnosis, data are obtained from self-reported questionnaires and, if available, from an overnight PSG recording. The questionnaires ask for the frequency and intensity of snoring, and use rating scales (e.g., scale from 1 to 5) for the answers. For the purpose of research, data are obtained using an overnight sound recording.
(2) *Interdependency of Data*. Patients who are undergoing treatment for OSA should not snore. Thus, in research, the patients undergoing treatment must be evaluated separately.
(3) *Time Sensitivity*. Snoring may be habitual or sporadic. Thus, the data must be clearly time-stamped.
(4) *Source Validity*. Data for snoring are obtained in two ways: self-reported (subjective) and clinically recorded (objective). In the first case, the patients may be not aware of their snoring or report incorrectly the intensity or frequency or snoring. In the second case, the recordings for “non-habitual snorers” must be done for at least 6 hours or even repeated for 2–3 nights since the presence of absence of snoring depends on many factors (e.g., sleeping position, sleep stage). Thus, the validity of the data varies between different sources and even different recording methods. As a result, the secondary analysis should clearly distinguish between the subjective data and the objective data.
(5) *Absent-Value Semantics*. We analyzed 447 male patients’ records obtained from a respiratory clinic. In this data set, snoring is self-reported using a rating scale (numeric scale from 1 to 5) corresponding to values: “never”, “rarely”, “sometimes”, “frequently”, “almost always”, and “not sure.” There are no missing values, which means that the patients answered the question or specified “not sure.” Figure 3 shows the frequency of each answer. The answer “not sure” is relatively frequent (13 %), and the number of answers “not sure” is larger than the “never” and “rarely” combined together.

Fig. 3 Histogram for snoring for males (N = 447)

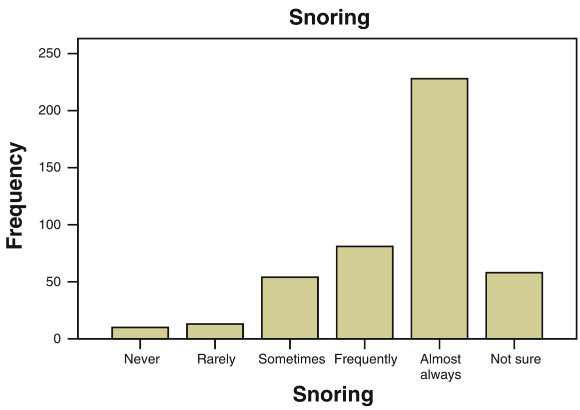
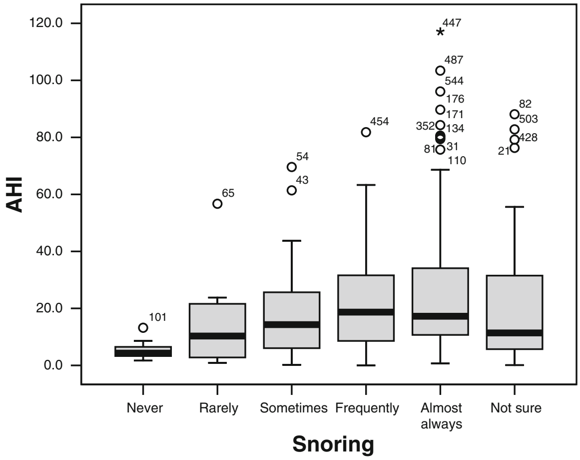


Fig. 4 Box plots of AHI for five frequencies of snoring and "not sure" value



The high percentage of "not sure" answers for snoring creates a problem for the data interpretation in secondary data analysis. For example, if we want to study a relationship between snoring and OSA, we should decide how we will treat absent values. Since absence of snoring decreases the probability of OSA, we have to be particularly careful about an assumption that the answer "not sure" corresponds to "absence of snoring." To show the complexity of the interpretation of absent value, we have examined the association between AHI (an indicator of presence and severity of OSA described in Sect. 2.1) and the reported frequency of snoring. Figure 4 shows the spread of AHI values for different frequencies of snoring. In Fig. 4, the boxplots visualize the different distributions of AHI for two answers: "never" and "not sure." For the "never" answer, all records except one outlier have AHI < 10. For the "no sure" answer, the median is higher than the median for the "never" answer, third quartile represents AHI > 10, and the four outliers have AHI > 50, which indicates that these patients have extremely severe OSA, and they should immediately start their treatment.

From the analysis of the clinical data, we found that (1) more than 10 % of the patients are "not sure" about their snoring and (2) more than half of the patients answering "not sure" have moderate to severe (and extremely severe) OSA. We observed that the answers regarding snoring are based on two types of logic. In screening, the answers are based on binary logic: yes/no. In diagnosis, the answers are based on three-valued logic, and they allow for the "third" unknown value. Therefore, secondary analysis of the data should consider the differences in the types of questions (binary vs. ternary answers), and consider the negative answers as including the "no sure" values.

2.4 Example 3: Smoking

Smoking is a self-reported risk factor used in many OSA questionnaires. Operationalization of the concept "smoking" is very complex. In general, smoking can be defined as "history of smoking" and "current smoking." Furthermore, the questions about smoking can use a categorical variable with binary values "yes" and "no" and a continuous variable indicating the numbers of smoked cigarettes (packs) per day. For the sake of brevity, we will limit our discussion on contextual information to one aspect: absent-value semantics.

We analyzed 795 patients' records from a respiratory clinic, in which "smoking" is operationalized as a self-reported question about a "current smoking" with dichotomous values "yes" or "no." We should emphasize here that the "smoking" question does not provide "not sure" option since the patients should know their "smoking status." The data set has a very high percentage of missing values for "smoking" question: 23 % (182/795). Smoking has been reported by 15 % of the patients (118/795) and non-smoking by 62 % of the patients (495/795). Interestingly, the number of missing values (182) is higher than the number of "yes" answers (118).

Furthermore, we analysed the AHI distribution for the three groups of answers: missing values, "no," and "yes", as shown in Fig. 5. The AHI distribution for "missing values" is similar to the groups reporting "yes" and "no. The mean and median value of AHI for the "no" answer is higher than mean and median value for the "yes" answer. However, an interpretation of these findings is impossible. First, 23 % of the patients did not answer the question "Do you currently smoke?" This group of patients with missing data has a fairly high AHI mean value of 24.5. Second, many patients with serious respiratory problems who quit smoking will report "no" for the question "Do you currently smoke?" We also compared the ratio of missing answers among females and males, and the proportion of missing values is slightly higher for females (27 %) than for males (21 %).

The high percentage of missing data can be explained by the fact that data on smoking are sensitive in social context of respiratory clinic. Other examples of even more sensitive data would be "use of alcohol or recreational drugs." Although questions on alcohol/recreational drug were included in the studied

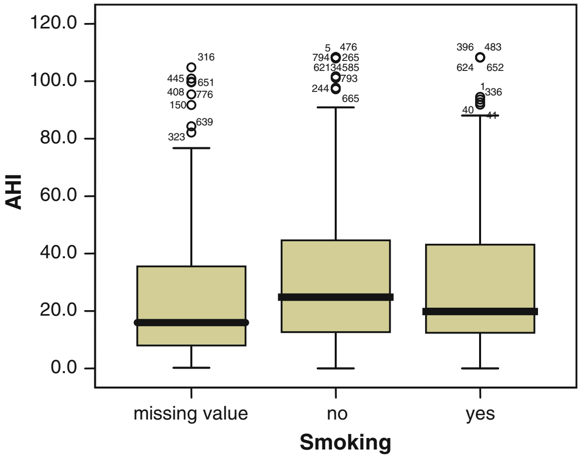


Fig. 5 Boxplots of AHI for three groups of answers

questionnaire, the data are too sensitive for the reporting in the context of this chapter.

In a self-reporting questionnaire, sensitive data are especially prone to intentional omissions. Thus, the contextual interpretation of the absent values should include: (1) analysis of the other absent values for the same patient (pattern of answers/omissions), (2) analysis of the omissions/answers to the same question for all records, and (3) overall level of absent values. From the analysis of the clinical data, we found that 23 % of patients did not answer the sensitive question about smoking; however, at the same time, only 4 % missed the NC answer and all patients answered the question about snoring. Nonetheless, the investigation of intentionality or randomness of omissions requires further research. At this phase of our study, we can conclude that the answers to sensitive questions are highly contextual and must be cautiously approached in secondary analysis. Also, the level of sensitivity of data varies among deferent cultural groups and organizations.

3 Fuzzy-Semiotic Representation

This section presents an outline of a conceptual framework for the representation of contextual information. This framework has its theoretical foundations in semiotics and fuzzy logic. In our discussion, we concentrate on the modeling of predictors and risk factors. Our fuzzy-semiotic framework addresses the contextualization of the predictor interpretation and the imprecision of predictors. For the purpose of this research, a predictor is defined as an established or suspected symptom, sign, correlate, or co-morbid condition. Therefore, predictors may vary from quantitatively measurable ones such as neck circumference to qualitative (subjective) predictors such as sleepiness.

3.1 Semiotic-Based Approach

We use a semiotic-based approach to define the contextual information. Semiotics is a discipline which can be broadly defined as the study of signs. Since signs, meaning-making, and representations are all present in every part of human life, the semiotic approach has been used in almost all disciplines, from mathematics through literary studies to information sciences and computing science. A semiotic paradigm is, on one hand, characterized by its universality and transdisciplinary nature, but, on the other hand, it is associated with different traditions and with a variety of empirical methodologies. We are using a framework based on the Peircean model of sign and semiosis as a process. Peirce defined "sign" as any entity carrying some information and used in a communication process. Peirce, and later Charles Morris, divided semiotics into three categories: syntax (the study of relations between signs), semantics (the study of relations between signs and the referred objects), and pragmatics (the study of relations between the signs and the agents who use the signs to refer to objects in the world). This triadic distinction is represented by a Peirce's semiotic triangle: the representamen (the form which the sign takes), object, and interpretant. The notion of "interpretant," is somewhat simplified in our research, and we propose to use six dimensions: agent, purpose, social and cultural context, bias, time, and view.

Our conceptual framework describes predictors at three levels: semantic (conceptualization), syntactic representation (operationalization), and pragmatic interpretation (utilization of measurements). The conceptualization level defines the medical concept (ontology) in terms of its general semantics. The representation level defines the possible measures of the medical concept. The interpretation level has three aspects: a diagnostic value of the predictor, a practical utility of the predictor in clinical setting, and a knowledge base (KB) containing medical facts and rules related to the predictor.

Figure 6 shows the semiotic triangle representing the concept of "neck thickness." The concept is connected with one of its possible representations (Representamen), the measure of NC during the physical examination (representation). The interpretant has six dimensions, for example, it is described by agent = health practitioner, purpose = diagnosis, SC context = low sensitivity, bias = (gender, ethnicity), time = timestamp, view = measuring procedure.

(1) *Predictor Definition*. We define a predictor as a quadruple: $P = \langle C, M, I, KB \rangle$. The predictor conceptualization is represented by C, the set of applicable measures for the concept by M, and the set of possible interpretations by I. A knowledge base for the predictor is represented by KB.

(2) *Knowledge Base*. KB contains five types of facts regarding interpretation: agents (e.g., patients, health professionals, and medical sensors), purposes (e.g., screening, diagnostic, research, prognostic, treatment evaluation), social and cultural contexts (e.g., sensitivity of data), biases (e.g., dependencies between predictors), and views (e.g., diagnostic criteria used by the clinics).

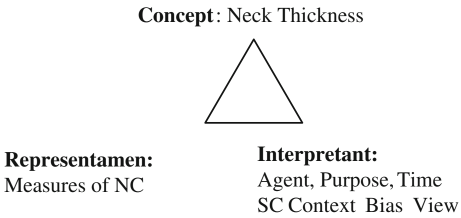


Fig. 6 Semiotic triangle for neck thickness as OSA predictor

We generated the KB for neck thickness from published medical studies using MEDLINE. We list here only four facts. The description of our KB and references can be found in [19].

KB1 A large neck is a characteristic sign of OSA

KB2 In general, the neck circumference (NC) is significantly larger in men than in women

KB3 NC ranges from 25 to 65 cm for adults. Typical male NCs range from 42 to 45 cm. Mean NC for women 36.5 ± 4.2 (based on N = 1,189) and 41.9 ± 3.8 for men (based on N = 2,753)

KB4 Male NCs can be divided into three groups (the ranges were selected for the diagnosis of OSA): small to normal (<42), intermediate (42–45), and large (>45).

3.2 Fuzzy-Logic Based Approach

We present the fuzzy-logic approach using one of OSA predictor described in Sect. 2: large neck circumference. Our fuzzy model is based on three KB facts: KB2, KB3, and KB4. The predictor NC is defined as a linguistic variable *L* = *<NC, {small, typical, large, atypical}*, [25, 65], $\{\mu_{small}, \mu_{typical}, \mu_{large}, \mu_{atypical}\}$>. The linguistic variable *L* has a set of four possible terms: {*small, typical, large, atypical*}. The universe of discourse has the range from 25 to 65 cm (based on KB3), and the set of the membership functions, MFs, $\{\mu_{small}, \mu_{typical}, \mu_{large}, \mu_{atypical}\}$ defines the mapping for the fuzzification function. Figure 7 illustrates four MFs for the neck circumference for males and Fig. 8 illustrated MFs for females. The MFs are constructed based on KB3 and KB4.

4 Conclusions and Future Work

In this chapter, we have presented contextual modeling and context-dependent interpretation of medical data, which we believe, is mandatory for a meaningful secondary analysis of data. We have discussed five contextual dimensions: goal

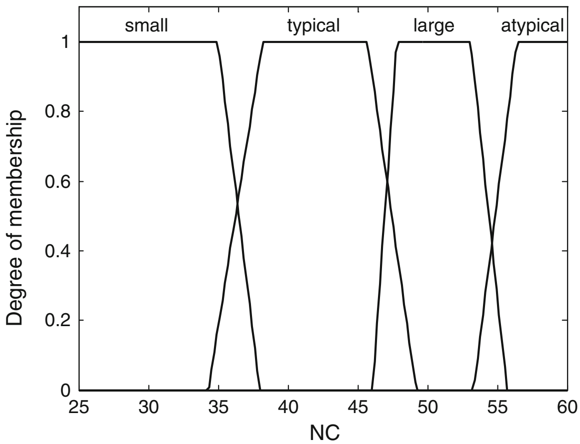


Fig. 7 Membership functions for NC for males

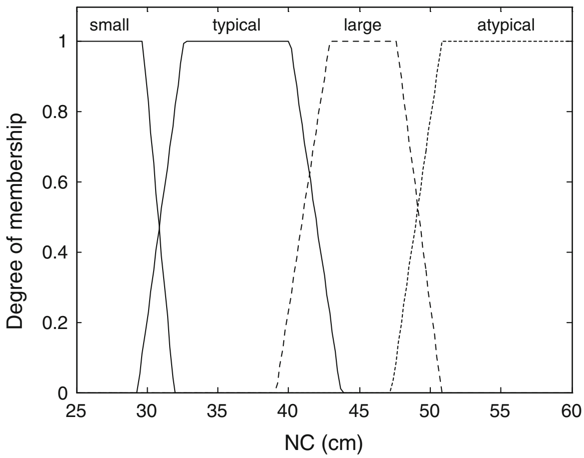


Fig. 8 Membership functions for NC for females

orientation, interdependency of data, time sensitivity, source validity, and absent-value semantics. We have situated our discussion in the context of real clinical situations such as screening, diagnosis, and clinical research of obstructive sleep apnea (OSA); and we used data from a respiratory clinic and data sets obtained from the authors of research publications. Throughout our chapter, we have argued that contextual information cannot be "decoupled" from data, and, what's more, it must be explicitly represented for effective secondary data analysis. We have presented extensive background knowledge and detailed description of medical data acquisition. We believe that this deep knowledge is mandatory for correct interpretation of data. Once the data are de-contextualized, they lose their primary meaning, and could be interpreted in misleading ways. The validity of the data mining process depends on correct re-use of data which have been already collected and used for other purposes (different than data mining). The integrity of "data miners" depends on careful analysis, modeling, interpretation, and reporting of the subtle differences in meaning and usage of primary data.

We have described contextual information for three OSA risk factors: neck circumference, snoring, and smoking. They have been selected to represent three categories of risk factors: an anatomical sign (thick neck), a nocturnal symptom (habitual snoring), and a life-style behavior (smoking). We have used these specific examples of risk factors to demonstrate that (1) interpretation of medical data is highly contextual (2) absence of value has many possible interpretations, and (3) imprecision of data depends on its context. We described three examples of missing data pertaining to three reasons: omission of sensitive data (smoking), unknown but easily obtainable data (self-reported neck circumference); and unknown but obtainable data (habitual snoring), all of which require careful analysis of context for appropriate handling. We observed that the secondary data analysis must involve investigation of the reasons for missing values. For example, in the case of sensitive data, such as smoking, the reason for the high percentage of missing values must be examined. The data on smoking have 23 % (!) missing values, which is higher rate than the answer "yes" (15 %). Therefore, any further analysis of correlation between smoking and severity of OSA may lead to spurious results and may indicate, for example, that there is no significant difference between the mean value of AHI for smokers (mean = 30.9) and non-smokers (mean = 32.9). Furthermore, we discussed the difference between binary answers (yes/no) and three-valued answers (yes/no/not sure) and their significance for the interpretation of missing values.

Also, we have outlined a conceptual framework for representing contextual information. To address the complex dimensionality of contextual information, we used semiotic approach. The semiotic approach provides an analytical tool to compare predictors in terms of their operationalization and interpretation. Without such a tool, data mining does not differentiate between subjective and fuzzy measurements and objective and precise measurements. In order to meaningfully interpret the data mining models and patterns, we must first capture the meaning of the data (semantics) and the usage of the data (pragmatics), not only the actual numeric values (syntax). To address the underlying imprecision of data, we used fuzzy-logic approach. We have illustrated our conceptual model using an example of neck circumference.

We realize that the described conceptual framework is rather sketchy and it requires many refinements. We are planning to continue our work on the proposed framework. We intend to extend the contextual model and create a complete representation of contextual information for the core OSA predictors. Also, we have started a related project on data analysis from STOP-bang questionnaire (used by the TRU students for OSA screening in geographically remote areas). We plan to utilize the proposed model for the analysis of patients' data from screening. We believe that the explicit modeling of contextual information will allow us to analyze the data and to compare the results from screening with diagnostic data from clinical records.

References

1. Mylonas, P., Vallet, D., Castells, P., Fernandez, M., Avrithis, Y.: Personalized information retrieval based on context and ontological knowledge. Knowl. Eng. Rev. **00**, 1–24 (2004)
2. Barsalou, L.W.: Context-independent and context-dependent information in concepts. Mem. Cogn. **10**, 82–93 (1982)
3. Dey, A.K.: Understanding and using context. Pers. Ubiquit. Comput. **5**(1), 125–134 (2001)
4. Berg, M., Goorman, E.: The contextual nature of medical information. Int. J. Med. Inform. **56**, 51–60 (1999)
5. Bertossi, M., Rizzolo, F., Lei, J.: Data quality is context dependent. In: Proceedings of WS. Enabling Real-Time Business Intelligence (BIRTE'10), pp. 52–67. Springer LNBIP 48 (2011)
6. Latkowski, R.: On indiscernibility of relations for missing attribute values. Fundam. Inf. **67**, 131–147 (2005)
7. Bolchini, C., Curino, C., Quintarelli, E., Schreiber, F., Tanca, L.A.: A data-oriented survey of context models. SIGMOD Rec. **36**(4), 19–26 (2007)
8. Zhang, W., Yu, C., Reagan, B., Nakajima, H.: Context-dependent interpretation of linguistic terms in fuzzy relational databases. In: Proceedings of the Eleventh International Conference on Data Engineering, pp. 139–146 (1995)
9. Dojat, M., Pachet, F.: Representing medical context using rule-based object-oriented programming techniques. In: Barahona, P., Stefanelli, M., Wyatt, J. (eds.) Artificial Intelligence in Medicine, pp. 423–424. Springer, Berlin (1995)
10. Roux, E., Godillon-Maquinghen, A.-P., Caulier, P., Bouilland, S., Bouttens, D.: A support method for the contextual interpretation of biomechanical data. IEEE Trans. Inf. Technol. Biomed. **10**(1), 109–118 (2006)
11. Douglas, N. J.: Clinicians' Guide to Sleep Medicine. Arnold, London (2002)
12. Fairbanks, D.N.F.: Snoring: a general overview with historical perspective. In: Fairbanks, D.N.F., Mickelson, S.A., Woodson, B.T. (eds.) Snoring and Obstructive Sleep Apnea, 3rd edn., pp. 1–17. Lippincott Williams & Wilkins, Philadelphia, PA (2003)
13. Chung, F., Subramanyam, R., Liao, P., Sasaki, E., Shapiro, C., Sun, Y.: High STOP-Bang score indicates a high probability of obstructive sleep apnoea. Br. J. Anaesth. **108**, 768–775 (2012)
14. Lam, B., Tench, E., Ryan, C.F.: Craniofacial profile in Asian and white subjects with obstructive sleep apnea. Thorax **60**(6), 504–510 (2005)
15. Shin, J.J., Berry, R.B., Khoo, M.C.: Fuzzy assessment of sleep-disordered breathing during continuous positive airway pressure therapy. Sleep **21**(8), 817–828 (1998)
16. Czopek, K.: Evaluation of breathing dynamics using the correlation of acoustic and ECG signals. Comput. Cardiol. **38**, 33–36 (2011)
17. Collop, N.A., Cassell, D.K.: Snoring and sleep-disordered breathing. In: Lee-Chiong, T.L., Sateia, M.J., Carskadon, M.A. (eds.) Sleep Medicine, pp. 349–355. Hanley & Belfus, Philadelphia Inc. (2002)
18. Woodson, B.T.: Obstructive sleep apnea: Evaluation by physical examination and special studies. In: Fairbanks, D.N.F., Mickelson, S.A., Woodson, B.T. (eds.) Snoring and Obstructive Sleep Apnea, 3rd edn., pp. 51–67. Lippincott Williams & Wilkins, Philadelphia (2003)
19. Kwiatkowska, M., Atkins, M.S., Ayas, N.T., Ryan, C.F.: Knowledge-based data analysis: first steps towards the creation of clinical prediction rules, using a new typicality measure. IEEE Trans. Inf. Technol. Biomed. **11**(6), 651–660 (2007)

Part VII
Time Series and Data Analysis

Time Series Trend Extraction and Its Linguistic Evaluation Using F-Transform and Fuzzy Natural Logic

Vilém Novák, Viktor Pavliska, Martin Štěpnička and Lenka Štěpničková

Abstract In this chapter, we contribute to the innovative method of time series analysis and forecasting using fuzzy transform and fuzzy natural logic. We will demonstrate that the F-transform is a powerful technique for extraction of the trend-cycle of time series. Further step is automatic linguistic evaluation of the course (tendency) of time series in a specified time slot. The main tool is the first degree F-transform (F^1-transform) which makes it possible to estimate average tangent of the given function. We thus obtain an objective result even if the trend is not visually apparent from the graph of the time series.

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

V. Novák (✉) · V. Pavliska · M. Štěpnička · L. Štěpničková
University of Ostrava, Institute for Research and Applications of Fuzzy Modeling,
NSC IT4Innovations, 701 03 Ostrava 1, Czech Republic
e-mail: Vilem.Novak@osu.cz

V. Pavliska
e-mail: Viktor.Pavliska@osu.cz

M. Štěpnička
e-mail: Martin.Stepnicka@osu.cz

L. Štěpničková
e-mail: Lenka.Stepnickovag@osu.cz

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_27,

1 Introduction

In a sequence of papers (see [1–3]) we suggested an advanced application of two special soft computing methods, namely the *fuzzy transform (F-transform)* and *fuzzy natural logic*[1] to analysis and forecasting of time series. The forecast precision is fully comparable with precision of top professional systems such as ForecastPro®. In comparison with them, however, we provide also automatically generated linguistic comments to the way, how forecast of the trend-cycle of the given time series has been obtained. These comments are based on application of methods of fuzzy natural logic to the forecast. Let us mention that we developed a special software system LFL Forecaster.[2]

In [5], we proposed the idea to generate automatically also linguistic evaluation of the trend of time series that is, its general tendency in a given time slot. Notice that in practice, the trend of the time series is often not clear, even when viewing the graph. The situation can be even more difficult if we are to characterize many (tens to thousands) time series. This may be useful, for example, when managers are to make decision about future directing of their company, for global economical analysis, and in many other occasions. In [3], we proposed to generate the comments on the basis of the first degree F–transform (F^1-transform) because it makes possible to estimate average value of a first derivative of a given function over a specified area. In this chapter, we elaborate this idea in more detail from linguistic point of view.

The *fuzzy natural logic* (FNL) is an extension of mathematical fuzzy logic in narrow sense. Its goal is to develop a *formal theory of natural human reasoning*. This requires a mathematical model of the semantics of natural language.[3] At present, it consists of formal theories of evaluative linguistic expressions, intermediate and generalized quantifiers and their syllogisms and the formal theory of the meaning of fuzzy/linguistic IF-THEN rules and approximate reasoning (cf., e.g., [6]).

By a fuzzy set, we understand a function $A{:}U \to [0, 1]$ where U is a universe and [0,1] is a support set of some standard algebra of truth values (cf. [7]). The set of all fuzzy sets over U is denoted by $\mathcal{F}(U)$. If A is a fuzzy set in U then we will sometimes write $A \underset{\sim}{\subset} U$.

2 Fuzzy Transform

The concept of fuzzy transform (F-transform) was in detail described in [8]. It has a lot of interesting properties (approximation ability, optimality and others) and great potential in various kind of applications, for example, image processing,

---

[1] This logic originates from fuzzy logic in broader sense (FLb-logic) introduced in [4] and developed as extension of the mathematical fuzzy logic.

[2] A testing version of LFL Forecaster is available on the web page http://irafm.osu.cz.

[3] Of course, at present stage we confine only to a small part of natural language expressions.

mining dependencies from numerical data, signal processing, special numerical methods and other ones (see [9, 10] and elsewhere). In general, we distinguish F-transforms of zero and higher degrees.

2.1 The Principle of F-Transform

The F-transform can be applied to a continuous real function f:$[a, b] \to R$. In the first step, we must define a *fuzzy partition* of the domain $[a, b]$. This consists of a finite set of fuzzy sets $\mathcal{A} = \{A_1, \ldots, A_n\}$, $n \geq 3$, defined over nodes $a = c_1, \ldots, c_n = b$. Properties of the fuzzy sets from $\mathcal{A}$ are specified by five axioms, namely: *normality, locality, continuity, unimodality, and orthogonality*. A fuzzy partition $\mathcal{A}$ is called *h-uniform* if the nodes $c_1, \ldots, c_n$ are *h-equidistant*, i.e., for all $k = 1, \ldots, n - 1$, $c_{k+1} = c_k + h$, where $h = (b - a)/(n - 1)$ and the fuzzy sets $A_2, \ldots, A_{n-1}$ are shifted copies of a *generating function* A_0:$[-1, 1] \to [0, 1]$ such that for all $k = 2, \ldots, n - 1$

$$A_k(x) = A_0\left(\frac{x - x_k}{h}\right), \quad x \in [c_{k-1}, c_{k+1}]$$

(for $k = 1$ and $k = n$ we consider only half of the function A_0, i.e. restricted to the interval $[0, 1]$ and $[-1, 0]$, respectively). We also put $s_0 = \int_{-1}^{1} A_1(x)dx$.

The membership functions $A_1, \ldots, A_n$ in a fuzzy partition $\mathcal{A}$ are called *basic functions*. Once the basic functions $A_1, \ldots, A_n \in \mathcal{A}$ are selected, we define a direct *F-transform* of a continuous function f as a vector $\mathbf{F}^{\mathcal{A}}[f] = (F_1[f], \ldots, F_n[f])$, where each k-th *component* $F_k[f]$ is equal to

$$F_k[f] = \frac{\int_a^b f(x)A_k(x)\,dx}{\int_a^b A_k(x)\,dx}, \quad k = 1, \ldots, n.$$

The meaning of the $F_k[f]$ component is *weighted average* of the functional values $f(x)$ where weights are the membership degrees $A_k(x)$. The *inverse F-transform* of f with respect to $\mathbf{F}^{\mathcal{A}}[f]$ is a continuous function[4] $\hat{f} : [a, b] \to R$ such that

$$\hat{f}(x) = \sum_{k=1}^{n} F_k[f] \cdot A_k(x), \quad x \in [a, b].$$

It is also shown that the function $\hat{f}$ differs from f (unless f is a constant function) but, under certain conditions, the sequence $\{\hat{f}_n\}$ obtained by increasing the number of basic functions in a fuzzy partition uniformly converges to f. All the details and full proofs can be found in [8, 11].

[4] By abuse of language, we call by direct F-transform both the procedure as well as its result $\hat{f}$.

2.2 F^1-Transform

In the previous subsection, we introduced the original concept of F-transform that is F^0-transform (i.e., zero-degree F-transform). Its components are real numbers. If we replace them by polynomials of arbitrary degree $m \geq 0$, we arrive at F^m transform. This generalization has been in detail described in [11].

Let us remark that extension of the F-transform to higher degree is not autotelic. First of all, we can achieve better approximation properties. Other nice property is the possibility to estimate also derivatives of the given function f as average values over wider area. In this chapter, we need only F^1-transform whose brief description follows.

Definition 1 Let f:[a, b] $\to$ $\mathbb{R}$ be a continuous function and $\mathcal{A} = \{A_1, \ldots, A_n\}$, $n \geq 3$ be a fuzzy partition of [a, b]. The vector of linear functions

$$F^{1,\mathcal{A}}[f] = (\beta_1^0 + \beta_1^1(x - c_1), \ldots, \beta_n^0 + \beta_n^1(x - c_n)) \tag{1}$$

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

$$\beta_k^0 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)A_k(x)dx}{hs_0}, \tag{2}$$

$$\beta_k^1 = \frac{\int_{c_{k-1}}^{c_{k+1}} f(x)(x - x_k)A_k(x)dx}{\int_{c_{k-1}}^{c_{k+1}} (x - c_k)^2 A_k(x)dx} \tag{3}$$

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

The fuzzy partition in the above definition needs not be uniform. If it is uniform, then the following simplification holds true.

Theorem 1 Let an h-uniform partition of [a, b] be given by the triangular-shaped basic functions $A_1, \ldots, A_n \in \mathcal{A}$ with the generating function $A_0 = 1 - |x|$. Then the coefficients β_k^0 and β_k^1 in the representation (1) of $\mathbf{F}^{1,\mathcal{A}}[f]$ are given by

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

$$\beta_k^1 = \frac{12 \int_{c_{k-1}}^{c_{k+1}} f(x)(x - c_k)A_k(x)dx}{h^3}, \tag{5}$$

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

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

Theorem 2 Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be an h-uniform partition of $[a, b]$, let functions f and $A_k \in \mathcal{A}$, $k = 1, \ldots, n$, be four times continuously differentiable on $[a, b]$. Finally, let $\mathbf{F}^{1,\mathcal{A}}[f]$ be the F^1-transform (1) of f. Then, for every $k = 1, \ldots, n$, the following estimation holds true:

$$\beta_k^1 = f'(c_k) + O(h). \tag{6}$$

According to Theorem 2, the coefficient β_k^1 is a convenient average estimation of the first derivative of f in the interval $[c_{k-1}, c_{k+1}]$. We will use this result in the evaluation of time series trend.

2.3 *Inverse* $\mathbf{F}^1$*-Transform*

Similarly as in the ordinary F-transform, the inverse F^1-transform of a function f is defined as a linear combination of basic functions with "coefficients" given by the F^1-transform components.

Definition 2 Let $f : [a, b] \to \mathbb{R}$ be a given function and $\mathbf{F}^{1,\mathcal{A}}[f] = (F_1^1[f], \ldots, F_n^1[f])$ be the F^1-transform of f with respect to $\mathcal{A} = \{A_1, \ldots, A_n\}$. Then, the function $\hat{f}^1 : [a, b] \to \mathbb{R}$ defined by

$$\hat{f}^1(x) = \sum_{k=1}^{n} F_k^1[f] A_k(x) \tag{7}$$

is called *inverse* F^1*-transform* of f with respect to $\mathbf{F}^{1,\mathcal{A}}[f]$.

Similarly as the F^0 transform, we can introduce a sequence $\{\hat{f}\}$ of inverse F^1-transforms of f that uniformly converges to f (see [11] for the details).

3 Analysis of Time Series Using F-Transform

3.1 *Decomposition of Time Series*

Let us consider a stochastic process (see [12, 13])

$$X : [a, b] \times \Omega \to \mathbb{R} \tag{8}$$

where $[a, b] \subset \mathbb{R}$ is an interval of reals and $\langle \Omega, \mathcal{A}, P \rangle$ is a probabilistic space. For simplicity, we will usually suppose that $a = 0$. By a *time series* we understand a stochastic process where $[a, b]$ is replaced by a finite set $Q = \{0, \ldots, p\} \subset \mathbb{N}$.

It follows from (8), each $X(t, \omega)$ for $t \in [a, b]$ and $\omega \in \Omega$ is a random variable. If we fix $\omega \in \Omega$ then we obtain one realization of (8) and in this case, we will write $X(t)$ only. Our basic assumption is that $X(t, \omega)$ can be decomposed into three constituents (cf. [13, 14]), namely

$$X(t, \omega) = TC(t) + S(t) + R(t, \omega), \quad t \in [a, b], \omega \in \Omega, \tag{9}$$

where $TC(t)$ is a *trend-cycle* and $S(t)$ is a seasonal constituent of the time series $X(t)$. Both $TC(t)$ and $S(t)$ are ordinary real functions. The $R(t, \omega)$ is a random noise, i.e. each $R(t)$ for $t \in [a, b]$ is a random variable with the mean value $\mathbf{E}(R(t)) = \mu \approx 0$ and variance $\mathbf{D}(R(t)) = \sigma > 0$.

3.2 Extraction of the Trend-Cycle and Trend

The seasonal component $S(t)$ is a periodic function and so, we will suppose that it can be expressed as a mixture of sine and cosine functions

$$S(t) = \sum_{i=1}^{r} B_i \sin(\omega_i t + \varphi_i) + \sum_{i=j}^{s} C_j \cos(\omega_j t + \varphi_j) \tag{10}$$

where ω_i, ω_j are frequencies, φ_i, φ_j phase shifts. Hence, realization of (9) can be in more detail written as

$$X(t) = TC(t) + \sum_{i=1}^{r} B_i \sin(\omega_i t + \varphi_i) + \sum_{i=j}^{s} C_j \cos(\omega_j t + \varphi_j) + R(t) \tag{11}$$

where $t \in [a, b]$.

Let $\bar{T}$ be the largest of the periodicities T_i, T_j, $i \in \{1, \ldots, r\}$, $j \in \{1, \ldots, s\}$ occurring in the seasonal constituent $S(t)$ in (11) and let us consider a set of h-equidistant nodes $c_1, \ldots, c_n$ with the distance

$$h = d\bar{T} \tag{12}$$

where $d \in N$ is such that $n \geq 3$ is assured. Finally, let us denote

$$o = \sum_{i=1}^{r} \left| \frac{2B_i}{d_i^2 \pi^2} \sin \varphi_i \right| + \sum_{j=1}^{s} \left| \frac{2C_j}{d_j^2 \pi^2} \cos \varphi_i \right| + 2\mu \tag{13}$$

where $d_i = \frac{h}{T_i}$, $i = 1, \ldots, r$ and $d_j = \frac{h}{T_j}$, $j = 1, \ldots, s$. It can be proved (see [15]) that each ith or jth summand in (14) is equal to 0 in case that $d_i \in \mathbb{N}$ or $d_j \in \mathbb{N}$, respectively.

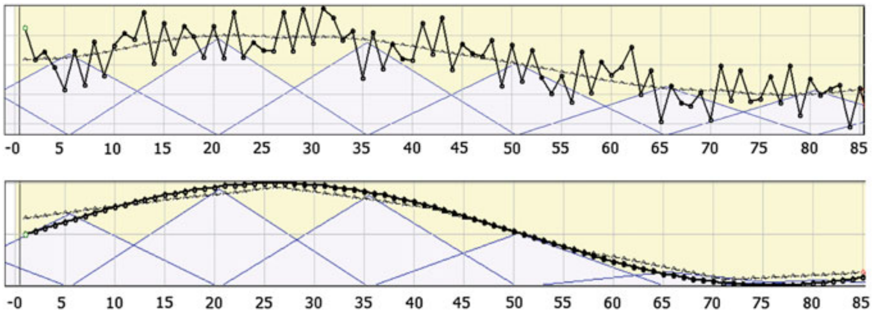


Fig. 1 Fuzzy transform of the artificial time series. The upper part shows the given time series, its trend-cycle and the fuzzy partition with triangular basic functions of the width 30. For comparison, the original trend-cycle $TC(t) = 20\sin(0.06t)$ and its fuzzy transform are depicted in the lower part

Theorem 3 [15] Let $X(t)$ be realization of the stochastic process (9) considered over the interval $[0, b]$. If we construct a fuzzy partition over the set of h-equidistant nodes $c_1, \ldots, c_n$ with the distance (13) then the corresponding inverse F-transform of $\hat{X}$ provides estimation of the trend-cycle $TC(t)$ with the following precision:

$$|\hat{X}(t) - TC(t)| \leq 2\omega(h, TC) + o, \qquad t \in [c_2, c_{n-1}] \tag{14}$$

where $\omega(h, TC)$ is a modulus of continuity of TC and o is the error (13).

Let us emphasize that o is a small number. Furthermore, the trend-cycle is generally a function with very small changes (low frequency) and so, its modulus of continuity $\omega(h, TC)$ is also small. Hence, we may conclude from (14) that the inverse F-transform $\hat{X}$ of the time series is approximately equal to the trend cycle TC.

To demonstrate the above theorem, we constructed an artificial time series

$$\begin{aligned} X(t) = {} & 20\sin(0.06t) + 5\sin(0.63t + 1.5) \\ & + 5\sin(1.26t + 0.35) + 15\sin(2.7t + 1.12) \\ & + 7\sin(0.41t + 0.79) + R(t) \end{aligned} \tag{15}$$

where the first member of (15) is trend-cycle having periodicity $T = 100$ and the other four sine members form the seasonal constituent $S(t)$. Their periodicities are $T_1 = 10$, $T_2 = 5$, $T_3 = 2.3$, $T_4 = 15.4$, respectively. Therefore, we set $\bar{T} = T_4$ and $d = 1$. This means that the distance h in (12) between nodes is $h = 15$ and so, width of the basic functions is $2\ h = 30$. The $R(t)$ is a random noise with the mean value $\mu = 0.1$. Demonstration of the result is in Fig. 1.

The recognizable feature of the trend-cycle is its *trend* (tendency). This may be global but it can manifest itself also in certain local time slots (for example, quarter of year, finishing of a construction, etc.). The time series, however, can contain large variations so that its trend need not be well recognizable. It would thus be quite useful to have an objective tool using which the trend can be clearly recognized. Such a tool is the F^1-transform. On the basis of Theorem 2, we will formulate the following definition.

Definition 3 Let $\mathbf{F}^{1,\mathcal{A}}[X]$ be F^1-transform of the time series X in (9). Then trend $T(A_k)$ of X in the area characterized by the fuzzy set $A_k \in \mathcal{A}$ is

$$T(A_k) = \beta_k^1 \tag{16}$$

where β_k^1 is the coefficient (3) (or, alternatively (5)).

By this definition, the value $T(A_k)$ is a weighted average tangent of the function $X(t)$ over the area determined by the fuzzy set (basic function) $A_k \in \mathcal{A}$.

4 Linguistic Evaluation of the Behavior of Time Series

4.1 *Evaluative Linguistic Expressions*

One of methods of the fuzzy natural logic applied in [2] is *perception-based logical deduction* (PbLD). It is applied to forecasting of the trend cycle. The basic theory behind is the formal theory of special natural language expressions called *evaluative linguistic expressions* which was in detail described in [16]. Recall that the latter are expressions such as *small, very big, rather medium, extremely strong*, etc. On the basis of our theory a mathematical model of the meaning of evaluative expressions was developed.

One of essential concepts is that of (linguistic) *context*. In our case, it is determined by a triple of real numbers $\langle v_L, v_S, v_R\rangle$ where $v_L < v_S < v_R$ $(\in \mathbb{R})$. These numbers represent the smallest, typically medium, and the largest thinkable values, respectively. The context is thus a set

$$w = \{x | v_L \le x \le v_R\} \tag{17}$$

together with the three distinguished points $DP(w) = \langle v_L, v_S, v_R\rangle$. By W we denote the set of all contexts (17) and by $EvExpr$ the set of all considered evaluative expressions. Each evaluative expression $Ev \in EvExpr$ is assigned the meaning which is a function (called *intension*)

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

Thus, the intension $\mathrm{Int}(Ev)$ assigns to each context $w \in W$ a fuzzy set $\mathrm{Ext}_w(Ev) \underset{\sim}{\subset} w$. The latter is called *extension* of the expression Ev in the context $w \in W$.

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

$$\langle noun \rangle is\ \langle simple\ evaluative\ expression \rangle \tag{18}$$

where

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

where <*hedge*> is a linguistic hedge, for example *very, rather, extremely, more or less, roughly*, etc. and <*TE—adjective*> is a trichotomous evaluative adjective, for example *small, medium, big, large, weak, good*, etc. The "is" takes here the role of a copula assigning property to objects and not as a genuine verb.

4.2 Linguistic Evaluation of the Time Series Trend

Our theory also enables to define a special function of *local perception*

$$\text{LPerc} : w \times W \to EvExpr \tag{19}$$

that assigns to each value $x \in w$ for a given context $w \in W$ an evaluative expression of the form (18). The function is constructed in such a way that given a linguistic context $w \in W$ and a value $x \in w$, the result of (19) is the *most plausible* evaluative expression which characterizes x in the given context w. This function makes it possible to *learn* linguistic description (a set of fuzzy/linguistic IF-THEN rules) characterizing behavior of the trend cycle so that its course can be predicted.

Using (19), we can also generate linguistic evaluation of time series trend (cf. Definition 3). There are some specificities regarding such evaluation. First of all, we must say, what does it mean "extreme increase (decrease)". In practice, it is determined by the largest acceptable difference of time series values with respect to a given (basic) time interval. Hence, mathematically we speak about the tangent. The usual basic time interval is 12 months, 31 days, etc. depending on the kind of the time series. Thus, the context is determined by the three distinguished values v_L, v_S, v_R of the tangent. The largest tangent v_R is determined in the way mentioned above while the smallest one is usually $v_L = 0$. The typical medium value v_S is determined analogously as v_R. The result is the context $w_{tg} = \langle v_L, v_S, v_R \rangle$. Now, we can linguistically characterize the trend $T(A_k)$ in (16) with respect to the context w_{tg}, i.e. we will automatically generate evaluative linguistic expressions using the function of local perception (19):

$$\text{LPerc}(T(A_k), w_{tg}). \tag{20}$$

This is justified by the fact that $T(A_k)$ is an average tangent over an area covered by the basic function $A_k \in \mathcal{A}$.

Predications using which we linguistically evaluate time series trend have specific form. *Sign* of the trend is characterized by a special word, namely $+$ is expressed by the word *increasing* (or *increase*) and $-$ by *decreasing* (or *decrease*). This can further be completed by special expressions characterizing *tilt of the trend*. Moreover, the obtained expressions are apparently subject to ordering that is similar to the natural ordering of the "standard" evaluative expressions. We conclude that the general syntactic form of expressions characterizing trend is either of (21) or (24) below:

$$\langle \textit{Trend} \rangle \ \textit{is} \ \langle \text{tilt characterization} \rangle \tag{21}$$

where

$$\langle \text{tilt characterization} \rangle := \text{stagnating} | \langle \text{hedge} \rangle \langle \text{sign} \rangle, \tag{22}$$

$$\langle \text{sign} \rangle := \text{increasing} | \text{decreasing} \tag{23}$$

and

$$\begin{aligned} \langle \text{hedge} \rangle := \ & \text{negligibly} | \text{slightly} | \text{somewhat} | \\ & \text{clearly} | \text{roughly} | \text{sharply} | \text{significantly}. \end{aligned}$$

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

$$\langle \text{sign of trend} \rangle \ \text{is} \ \langle \text{special hedge} \rangle \tag{24}$$

where <*sign of trend*> := increaseldecrease and

$$\begin{aligned} \langle \text{special hedge} \rangle : = \ & \text{negligible} | \text{slight} | \text{small} | \text{clear} | \\ & \text{rough} | \text{large} | \text{fairlylarge} | \text{quite large} | \text{significant} | \text{huge}. \end{aligned}$$

It is clear that, in fact, the feature *increase (decrease)* (of trend) is evaluated in both cases (21) (24). The difference in their use depends on the syntactic specificities but not on their semantics. Some cases, for example "trend is slightly increasing" and "increase of trend is slight" are even syntactic variations of the same expression.

This suggest the idea that the above special evaluative predications (21) and (24) are semantically tantamount[5] to the standard form

$$\langle \text{sign of trend} \rangle \ \text{is} \ \mathcal{B} \tag{25}$$

[5] It is difficult to state that these expressions are synonymous in the strict sense. This requires further linguistic research.

where $\mathcal{B}$ is an evaluative expression generated by the function (19) according to the following tables:

Case (a)

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

Case (b)

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

5 Examples

As mentioned, practical realization of the idea presented above is based on Definition 3. Depending on the definition of the fuzzy partition, the generated evaluation may concern either the whole time series or an arbitrary part of it.

The way how evaluation is generated is demonstrated on the case of smooth curve of monthly inflation measure in Fig. 2. Note that Slot 2 is wider than Slot 5 and so, the program generated the comment *fairly large decrease* for Slot 2 and *huge decrease* for Slot 5. The evaluations are generated on the basis of objectively

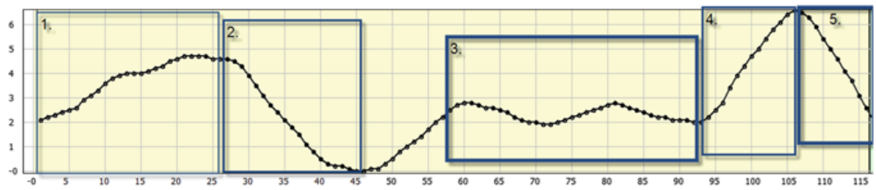


Fig. 2 Automatically generated linguistic evaluations of monthly trend of inflation measure over 10 years. The context is $\langle v_L = 0, v_S = 2/12, v_R = 5/12\rangle$. Evaluation of trend in the marked areas is the following: Slot 1.: *clear increase*, Slot 2.: *fairly large decrease*, Slot 3. *stagnating*, Slot 4. *significant increase* and Slot 5. *huge decrease*

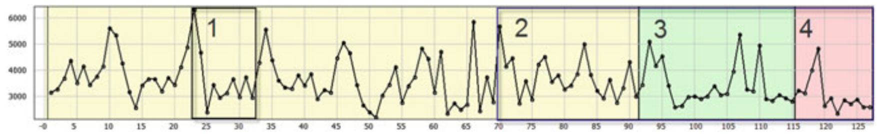


Fig. 3 Demonstration of evaluation of trend of various parts of a complicated time series. Trend of the whole series is *stagnating*. Slot 1 (time 23–32): *clear decrease*, Slot 2 (time 70–127): *negligible decrease*, Slot 3 (time 92–115): *small increase*, Slot 4 (time 116–127): *fairly large decrease*

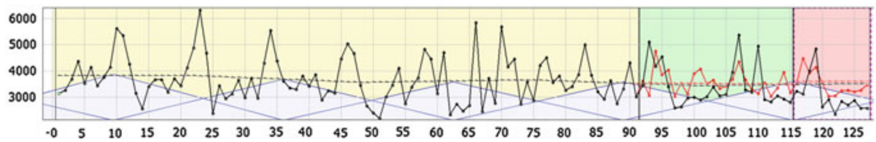


Fig. 4 Analysis and forecast of the time series from Fig. 3. Slot 3 (time 92–115—validation part) and Slot 4 (time 116–127—testing part) contain computed and predicted trend cycle, and also real and predicted values the time series itself. The generated comments in Slot 4 are: *rough decrease* for the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

computed average tangent and we argue that they comply well with the course of the time series.

More complicated is demonstration of comments generated to one time series taken from *M3-Competition* provided by the *International Institute of Forecasters* (the real content of the time series is not known). The time series is depicted in Fig. 3. One can see that its trend is by no means clear. Slot 3 (time 92–115) of this time series is validation part, on which the quality of the forecast is tested and the best one is chosen. Slot 4 (time 116–127) is testing part that is not used for computation of the forecast but only for comparison of the forecast with the real data. In Fig. 4 the same time series is analyzed and its forecast is computed.

The linguistic context for the trend evaluation was set to $\langle v_L = 0, v_S = 1200/12, v_R = 3000/12\rangle$ since the time series demonstrates clear periodicity of $T = 12$ (this was obtained using periodogram—cf., e.g., [12]). The generated evaluation of trend of the predicted values in the testing part is *rough decrease* while evaluation of trend of the real data is *fairly large decrease*. Both evaluations are in good agreement which is another support for the quality of our forecasting method. Thus, instead of presenting concrete predicted data, the manager might be satisfied with the information that "rough decrease" is expected.

6 Conclusion

In this chapter, we presented two special soft computing methods that can be used for analysis and forecasting of time series. These methods are F-transform (both zero and first degree) and the theory of fuzzy natural logic. The F-transform is used especially for extraction of trend-cycle of the time series while methods of the fuzzy natural logic are applied to prediction of the latter. Furthermore, combination of the F-transform and fuzzy natural logic is applied to forecast the trend cycle and also to generate linguistic comments to the trend (tendency) of the time series in arbitrary time slots. We believe that such comments can be useful in situations, for example, when it is difficult to see the global trend because the time series is too much varying.

One of the directions for further research is application of the theory of *intermediate quantifiers* (see [17, 18]). The latter are linguistic expressions such as *most, many, almost all, few, some, a large part of*, etc. Using this theory, we can model the meaning of expressions, such as "most (many, few) analyzed time series stagnated recently but their future trend is slightly increasing", "huge decrease of trend of almost all time series in the recent quarter of the year", and similar. Moreover, we can also apply syllogistic reasoning with such expressions, for example

Many analyzed time series are stagnating
Almost all analyzed time series are from car industry
Some time series from car industry are stagnating

It is important to realize that this valid generalized Aristotle's syllogism is true in all models and the same holds for all the other ones. In [18] we proved validity of over one hundred of intermediate syllogisms.

Other direction of research is focused on mining of interesting information over the given set of time series. We start with analysis and forecasting of all the time series. Then we generate comments to their trend in interesting time slots, summarize them using the intermediate quantifiers and derive further properties on the basis of valid syllogisms. Mining information from time series will be the topic of the future papers.

References

1. Novák, V., Štěpnička, M., Perfilieva, I., Pavliska, V.: Analysis of periodical time series using soft computing methods. In: Ruan, D., Montero, J. Lu, J., Martnez, L., D'hondt, P., Kerre, E. (eds.) Computational Intelligence in Decision and Control, pp. 55–60. World Scientific, New Jersey (2008)
2. Novák, V., Štěpnička, M., Dvořák, A., Perfilieva, I., Pavliska, V., Vavřčková, L.: Analysis of seasonal time series using fuzzy approach. Int. J. Gen. Syst. **39**, 305–328 (2010)
3. Novák, V., Perfilieva, I., Pavliska, V.: The use of higher-order F-transform in time series analysis. In: World Congress IFSA 2011 and AFSS 2011, Surabaya, Indonesia, pp. 2211–2216 (2011)
4. Novák, V.: Towards formalized integrated theory of fuzzy logic. In: Bien, Z., Min, K. (eds.) Fuzzy Logic and Its Applications to Engineering, Information Sciences, and Intelligent Systems, pp. 353–363. Kluwer, Dordrecht (1995)
5. Novák, V., Perfilieva, I., Jarushkina, N.G.: A general methodology for managerial decision making using intelligent techniques. In: Rakus-Anderson, E., Yager, R., Ichalkaranje, N., Jain, L. (eds.) Recent Advances in Fuzzy Decision-Making, pp. 103–120. Springer, Heidelberg (2009)
6. Novák, V.: Reasoning about mathematical fuzzy logic and its future. Fuzzy Sets Syst. **192**, 25–44 (2012)
7. Novák, V., Perfilieva, I., Močkoř, J.: Mathematical Principles of Fuzzy Logic. Kluwer, Boston (1999)
8. Perfilieva, I.: Fuzzy transforms: theory and applications. Fuzzy Sets Syst. **157**, 993–1023 (2006)
9. Perfilieva, I., Novák, V., Dvořák, A.: Fuzzy transform in the analysis of data. Int. J. Approx. Reason. **48**, 36–46 (2008)
10. Perfilieva, I.: Fuzzy transforms: A challenge to conventional transforms. In: Hawkes, P. (ed.) Advances in Images and Electron Physics, vol. 147, pp. 137–196. Elsevier Academic Press, San Diego (2007)
11. Perfilieva, I., Daňková, M., Bede, B.: Towards a higher degree F-transform. Fuzzy Sets Syst. **180**, 3–19 (2011)
12. Anděl, J.: Statistical Analysis of Time Series. SNTL, Praha (1976) (in Czech)
13. Hamilton, J.: Time Series Analysis. Princeton University Press, Princeton (1994)
14. Bovas, A., Ledolter, J.: Statistical Methods for Forecasting. Wiley, New York (2003)
15. Novák, V., Perfilieva, I., Krejnovich, V.: F-transform in the analysis of periodic signals. In: Inuiguchi, M., Kusunoki, Y., Seki, M. (eds.) Proceedimgs of 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, Osaka, Japan: Osaka University, pp. 150–158 (2012)
16. Novák, V.: A comprehensive theory of trichotomous evaluative linguistic expressions. Fuzzy Sets Syst. **159**(22), 2939–2969 (2008)
17. Novák, V.: A formal theory of intermediate quantifiers. Fuzzy Sets Syst. **159**(10), 1229–1246 (2008)
18. Murinová, P., Novák, V.: A formal theory of generalized intermediate syllogisms. Fuzzy Sets Syst. **186**, 47–80 (2012)

Using Artificial Neural Networks in Fuzzy Time Series Analysis

Cagdas Hakan Aladag

Abstract In recent years, fuzzy time series have been drawn great attention due to their potential for use in time series forecasting. In many studies available in the literature, fuzzy time series have been successfully used to forecast time series contain some uncertainty. Studies on this method still continue to reach better forecasting results. Determination of fuzzy relations between observations is an important phase of fuzzy time series approaches which directly affect the forecasting performance. In order to establish fuzzy relations, different techniques have been utilized in the literature. One of these techniques is artificial neural networks method. In this study, it is shown that how different artificial neural networks models can be used to determine fuzzy relations with real time series applications.

1 Introduction

The real time series data such as temperature and share prices of stockholders contain some uncertainty in itself. Using conventional time series methods cannot give satisfactory results for such time series [1]. It would be wiser to utilize fuzzy time series approaches instead of conventional methods for such time series. Besides, when fuzzy time series are employed in time series analysis, there is no need to have at least 50 observations and to satisfy the linearity assumption as the conventional ones do [2]. Therefore, fuzzy time series forecasting models has received a significant amount of attention from both academicians and practitioners in recent years [3]. Various fuzzy time series forecasting models have been proposed in the literature [4]. Also, there have been some studies in order to improve this approach so that more accurate forecasts can be obtained [5].

C. H. Aladag (✉)
Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey
e-mail: chaladag@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_28,

Fuzzy time series was firstly introduced by Song and Chissom [6–8], which were based on the theory of fuzzy set as introduced by Zadeh [9]. All of the studies [6–8] were about fuzzy time series approaches and all of them were mentioned three basic stages of fuzzification, defining fuzzy relations and defuzzification. Later on, various fuzzy time series approaches have been proposed by improving these three stages in order to get better forecasting results [10].

The observations of a time series in fuzzy time series approaches represent discrete fuzzy sets differently from the conventional time series approaches in which they are real numbers. In the fuzzification stage, the observations, which are real numbers, of a time series in a certain period are converted as discrete fuzzy sets [10]. For the fuzzification stage, various approaches have been suggested in the literature [5]. Song and Chissom [6–8], Chen [11, 12], Singh [13], Egrioglu et al. [2, 14] used fuzzification methods based on partition of universe of discourse. The operation of partition of universe of discourse is to divide universe of discourse into equal or unequal sub intervals according to a predefined interval length [5]. The effect of the predefined interval length on the forecasting performance is examined by Huarng [15]. Some other studies based on length of interval are as follows: Egrioglu et al. [10], Huarng and Yu [16], Yolcu et al. [17], Davari et al. [18], Kuo et al. [19], Hsu et al. [20], Huang et al. [21], and Lee et al. [22, 23].

After fuzzy relations are established and fuzzy forecasts are produced, the obtained fuzzy forecasts are transformed into crisp values in the defuzzification phase. Then, a performance measure which is based on the difference between observations and forecasts is calculated to evaluate the fuzzy time series forecasting model. In the literature, the studies made by Aladag et al. [4], Cheng et al. [24], Jilani and Burney [25], Yu [26], and Liu et al. [27] are based on the stage of defuzzification.

Another important phase of fuzzy time series approaches is determination of fuzzy relations between observations. In the literature, various approaches have been proposed to improve this phase. The related literature was reviewed by Aladag et al. [5] as follows:

"Sullivan and Woodall [28] used transition matrices based on Markov chain instead of using fuzzy logic relation matrix. Chen [11] proposed an approach that makes the phase of determination of fuzzy relations easier. In the method proposed by Chen [11], instead of using fuzzy logic relation matrix, fuzzy logic group relationships tables are used to establish fuzzy logic relationships. Also, fuzzy logic group relationships tables have been used in many studies such as Huarng [15], Yu [26], Huarng and Yu [16], Cheng et al. [29], and Egrioglu et al. [10]. Huarng and Yu [30] suggested a method in which feed forward neural networks are used to determine fuzzy relations. A high order fuzzy time series forecasting method in which feed forward neural networks are employed to establish fuzzy relations was proposed by Aladag et al. [31]. In Aladag et al. [32], Elman recurrent neural networks were also used to determine fuzzy relations. Yu and Huarng [33], and Yolcu et al. [17] proposed a different approach in which membership function values are employ to use feed forward neural networks in the phase of determination of fuzzy relations.

In the analysis of time invariant fuzzy time series, the phase of determination of fuzzy logic relationships has important impact on forecasting performance. By establishing fuzzy logic relations, Song and Chissom [6, 7] obtained a relationship matrix consists of these fuzzy logic relations. Obtaining a relationship matrix requires complex operations. Thus, Chen [11] proposed to use fuzzy logic group relationships tables in the phase of determination of fuzzy logic relationships differently from the method proposed by Song and Chissom [6, 7]. Instead of complex matrix operations, fuzzy logic group relationships tables have been generally preferred in the literature to establish fuzzy relations."

When fuzzy logic group relationships tables are used, membership values of fuzzy sets are ignored and fuzzy sets' elements whose membership value is 1 are only taken into consideration. This situation causes information loss and decrease in the explanation power of the model. In order to overcome these problems, Aladag et al. [5] used particle swarm optimization method to calculate membership values in the fuzzy relationship matrix. Also, Aladag [34] proposed a fuzzy time series forecasting approach in which multiplicative neuron model is utilized to define fuzzy relations in order to increase forecasting accuracy. In Aladag's [34] method, particle swarm optimization algorithm was employed to train multiplicative neuron model.

To sum up, determination of fuzzy relations is an important stage that directly affects the forecasting performance of fuzzy time series approaches. In the literature, there are various methods for this phase. In recent years, artificial neural networks (ANN) have been successfully utilized to establish fuzzy relations. Huarng and Yu [30] firstly suggested using feed forward neural networks to determine fuzzy relations. Then, Aladag et al. [31] proposed a high order fuzzy time series forecasting method in which feed forward neural networks are employed to determine fuzzy relations. Also, Aladag [34] use multiplicative neuron model to define fuzzy relations. In these studies, it was observed that using ANN in the phase of determination of fuzzy relations produces accurate forecasts.

In this study, two recent fuzzy time series forecasting methods based on ANN are introduced and these methods are applied to TAIFEX data. The next section presents the fundamental definitions about fuzzy time series. Section 3, gives brief information about the methods proposed by Aladag et al. [31] and Aladag [34]. The implementation is given in Sect. 4. The last section concludes the chapter.

2 Fuzzy Time Series

Aladag et al. [4] presented basic definitions of fuzzy time series as follows:

Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_b\}$. A fuzzy set A_i of U is defined as $A_i = f_{A_i}(u_1)/u_1 + f_{A_j}(u_2)/u_2 + \cdots + f_{A_i}(u_b)/u_b$, where f_{A_i} is the membership function of the fuzzy set $A_i; f_{A_i} : U \to [0, 1]$. u_a is a generic element of fuzzy set $A_i; f_{A_i}(u_a)$ is the degree of belongingness of u_a to $A_i; f_{A_i}(u_a) \in [0, 1]$ and $1 \leq a \leq b$.

Definition 1 Fuzzy time series, let $Y(t)(t = \ldots, 0, 1, 2, \ldots)$ a subset of real numbers, be the universe of discourse by which fuzzy sets $f_j(t)$ are defined. If $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$ then $F(t)$ is called a fuzzy time series defined on $Y(t)$.

Definition 2 Let $F(t)$ be a fuzzy time series. If $F(t)$ is a caused by $F(t-1)$, then this fuzzy logical relationship is represented by

$$F(t-1) \rightarrow F(t)$$

and it is called first order fuzzy time series forecasting model.

Definition 3 Let $F(t)$ be a fuzzy time series. If $F(t)$ is a caused by $F(t-1)$, $F(t-2)$, $\ldots$, $F(t-m)$, then this fuzzy logical relationship is represented by

$$F(t-m), \ldots, F(t-2), F(t-1) \rightarrow F(t)$$

and it is called the mth order fuzzy time series forecasting model.

3 The Methods Based on ANN

3.1 The Method Proposed by Aladag et al. [31]

Chen [11] proposed a first order fuzzy time series method which is simple since it does not require matrix operations unlike the method introduced by Song and Chissom [6, 7]. In this method, Chen [11] suggested to use fuzzy logic group relationship tables to define fuzzy relations. Since first-order fuzzy time series models have a simple structure, they can generally be insufficient to explain more complex relationships. For this reason, Chen [12] proposed a new method which analyzes a high order fuzzy time series forecasting model. In the method proposed by Chen [12], fuzzy logic group relationship tables is also used for determination of fuzzy relations. However, implementation of Chen's [12] approach becomes more difficult when the order of fuzzy time series increases. Therefore, Aladag et al. [31] proposed a new high order fuzzy time series forecasting method in which fuzzy relations are established using feed forward neural networks. In addition, it was observed that forecasting performance of the method suggested by Aladag et al. [31] is better. The stages of the method based on neural networks can be given as follows [31]:

Stage 1. Define and partition the universe of discourse.
The universe of discourse for observations, $U = [starting, ending]$, is defined. After the length of intervals, l, is determined, the U can be partitioned into equal-length intervals $u_1, u_2, \ldots, u_b$, $b = 1, \ldots$ and their corresponding midpoints $m_1, m_2, \ldots, m_b$, respectively.

Table 1 Notations for second order

Observation no	F_{t-2}	F_{t-1}	F_t	Input-1	Input-2	Target
1	–	–	A_6	–	–	–
2	–	A_6	A_2	–	–	–
3	A_6	A_2	A_3	6	2	3
4	A_2	A_3	A_7	2	3	7
5	A_3	A_7	A_4	3	7	4
6	A_7	A_4	A_2	7	4	2

$$\begin{aligned} u_b &= [starting + (b-1) \times l, starting + b \times l], \\ m_b &= \frac{[starting + (b-1) \times l, starting + b \times l]}{2} \end{aligned}$$

Stage 2. Define fuzzy sets.
Each linguistic observation, A_i, can be defined by the intervals u_1, u_2, …, u_b.

$$A_i = f_{A_i}(u_1)/u_1 + f_{A_i}(u_2)/u_2 + \cdots + f_{A_i}(u_b)/u_b$$

Stage 3. Fuzzify the observations.
For example, a datum is fuzzified to A_i, if the maximal degree of membership of that datum is in A_i.
Stage 4. Establish the fuzzy relationship with feed forward neural network.
An example will be given to explain Stage 4 more clearly for the second order fuzzy time series. Because of dealing with second order fuzzy time series, two inputs are employed in neural network model, so that lagged variables F_{t-2} and F_{t-1} are obtained from fuzzy time series F_t. These series are given in Table 1. The index numbers (i) of A_i of F_{t-2} and F_{t-1} series are taken as input values whose titles are Input-1 and Input-2 in Table 1 for the neural network model. Also, the index numbers of A_i of F_t series are taken as target values whose title is Target in Table 1 for the neural network model. When the third observation is taken as an example, inputs values for the learning sample [A_6, A_2] are 6 and 2. Then, target value for this learning sample is 3.
Stage 5. Defuzzify results.
The defuzzified forecasts are middle points of intervals which correspond to fuzzy forecasts obtained by neural networks in the previous stage.

3.2 The Method Proposed by Aladag [34]

In order to reach high accuracy level, Aladag [34] proposed another high order fuzzy time series forecasting approach by improving the method suggested by Aladag et al. [31]. In this approach [34], the multiplicative neuron model is utilized to define fuzzy relations and particle swarm optimization is used to train this network model. In [34], it was showed that the proposed approach produces very accurate forecasts. Pseudo code of the algorithm of Aladag's fuzzy time series forecasting approach can be given as follows [34]:

Stage 1. Define and partition the universe of discourse.
Stage 2. Define fuzzy sets.
Stage 3. Fuzzify the observations.
Stage 4. Establish the fuzzy relationship with MNM model.
Stage 5. Defuzzify results.

The detailed information about the related method can be found in Aladag [34].

4 The Implementatıon

In the implementation two recent fuzzy time series approaches proposed by Aladag et al. [31] and Aladag [34] are applied to TAIFEX data whose observations are between 03.08.1998 and 30.09.1998. The time series, which is also used in [35], has 47 observations. The first 31 and the last 16 observations are used for the training and the test sets, respectively.

When the approach proposed by Aladag et al. [31] is employed, the best forecasts are obtained for 10th order model. When TAIFEX data is forecasted by using the approach suggested by Aladag [34], the best forecasts are obtained from 10th order fuzzy time series model. The forecasting results obtained from the both methods are summarized in Table 2. In Table 2, the obtained forecasts, root mean square error (RMSE), and mean absolute percentage error (MAPE) values are presented. The formulas of RMSE and MAPE are given as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (foreCast_t - actual_t)^2}{n}}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{foreCast_t - actual_t}{actual_t} \right|$$

where t represents the time and n is the number of observations in the test set. $forecast_t$ is the forecast at t from any mentioned model and $actual_t$ is the observed value at t.

5 Conclusion

In this study, it is focused on using artificial neural networks in fuzzy time series analysis. It is a well-known fact that the phase of determination of fuzzy relations has important effect on the forecasting performance of fuzzy time series approaches. Therefore, various methods have been used to define fuzzy relations in the literature. Among these approaches, one of the most promising ones is artificial

Table 2 The obtained results

TAIFEX test data	Aladag et al. [31]	Aladag [34]
6709.75	6,850	6,750
6726.5	6,850	6,750
6774.55	6,850	6,850
6,762	6,850	6,850
6952.75	6,850	6,850
6,906	6,850	6,850
6,842	6,850	6,850
7,039	6,850	6,850
6,861	6,850	6,950
6,926	6,850	6,850
6,852	6,850	6,850
6,890	6,850	6,850
6,871	6,850	6,850
6,840	6,850	6,850
6,806	6,850	6,850
6,787	6,850	6,750
RMSE	83.58	72.55
MAPE	0.96 %	0.82 %

neural networks. In recent studies, different artificial neural network models have been used in the phase of determination of fuzzy relations to increase the forecasting performance of fuzzy time series method. In this study, two recent fuzzy time series approaches proposed by Aladag et al. [31] and Aladag [34] are presented. While feed forward neural networks are employed in the approach suggested by Aladag et al. [31], the multiplicative neuron model is utilized in the approach proposed by Aladag [34]. It is also shown that how artificial neural networks can be used for defining fuzzy relations. In the implementation, the obtained forecasting results are presented when the both approaches are applied to TAIFEX data.

References

1. Aladag, C.H.: Using artificial neural networks in fuzzy time series forecasting. In: Applied Statistics 2010 International Conference (AS 2010). Ribno, Bled, Slovenia (2010)
2. Egrioglu, E., Aladag, C.H., Basaran, M.A., Yolcu, U., Uslu, V.R.: A new approach based on the optimization of the length of intervals in fuzzy time series. J. Intell. Fuzzy Syst. **22**, 15–19 (2011)
3. Aladag, S., Aladag, C.H., Egrioglu, E.: Analyzing Ankara air quality data with fuzzy time series. In: International 7th Statistics Congress Proceedings, pp. 218–219. Ankara, Turkey, May 2011
4. Aladag, C.H., Yolcu, U., Egrioglu, E.: A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural networks. Math. Comput. Simul. **81**, 875–882 (2010)

5. Aladag, C.H., Egrioglu, E., Yolcu, U., Dalar, A.Z.: A new time invariant fuzzy time series forecasting method based on particle swarm optimization. Appl. Soft Comput. **12**(10), 3291–3299 (2012)
6. Song, Q., Chissom, B.S.: Fuzzy time series and its models. Fuzzy Sets Syst. **54**, 269–277 (1993)
7. Song, Q., Chissom, B.S.: Forecasting enrollments with fuzzy time series—Part I. Fuzzy Sets Syst. **54**, 1–10 (1993)
8. Song, Q., Chissom, B.S.: "Forecasting enrollments with fuzzy time series—Part II. Fuzzy Sets Syst. **62**(l):1–8 (1994)
9. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**:338–353 (1965)
10. Egrioglu, E., Aladag, C.H., Yolcu, U., Uslu, V.R., Erilli, A.: Fuzzy time series forecasting method based on Gustafson-Kessel fuzzy clustering. Expert Syst. Appl. **38**, 10355–10357 (2011)
11. Chen, S.M.: Forecasting enrollments based on fuzzy time-series. Fuzzy Sets Syst. **81**, 311–319 (1996)
12. Chen, S.M.: Forecasting enrollments based on high order fuzzy time series. Cybern. Syst. **33**, 1–16 (2002)
13. Singh, S.R.: A simple method of forecasting based on fuzzy time series. Appl. Math. Comput. **186**, 330–339 (2007)
14. Egrioglu, E., Aladag, C.H., Yolcu, U., Uslu, V.R., Basaran, M.A.: Finding an optimal interval length in high order fuzzy time series. Expert Syst. Appl. **37**, 5052–5055 (2010)
15. Huarng, K.: Effective length of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst. **123**, 387–394 (2001)
16. Huarng, K., Yu, T.H.K.: "Ratio-based lengths of intervals to improve fuzyy time series forecasting, IEEE Trans. Syst. Man Cybern. Part B Cybern. **36**, 328–340 (2006)
17. Yolcu, U., Aladag, C.H., Egrioglu, E., Uslu, V.R.: Time series forecasting with a novel fuzzy time series approach: an example for Istanbul stock market. J. Stat. Comput. Simul. **83**(4), 597−610 (2013)
18. Davari, S., Zarandi, M.H.F., Turksen, I.B.: An improved fuzzy time series forecasting model based on particle swarm intervalization. In: The 28th North American Fuzzy Information Processing Society Annual Conferences (NAFIPS 2009), Cincinnati, June 2009
19. Kuo, I.H., Horng, S.J., Chen, Y.H., Run, R.S., Kao, T.W., Chen, R.J., Lai, J.L., Lin, T.L.: Forecasting TAIFEX based on fuzzy time series and particle swarm optimization. Expert Syst. Appl. **37**, 1494–1502 (2010)
20. Hsu, L.Y., Horng, S.J., Kao, T.W., Chen, Y.H., Run, R.S., Chen, R.J., Lai, J.L., Kuo, I.H.: Temperature prediction and TAIFEX forecasting based on fuzzy relationships and MTPSO techniques. Expert Syst. Appl. **37**, 2756–2770 (2010)
21. Huang, Y.L., Horng, S.J., Kao, T.W., Run, R.S., Lai, J.L., Chen, R.J., Kuo, I.H., Khan, M.K.: An improved forecasting model based on the weighted fuzzy relationship matrix combined with a PSO adaptation for enrollments. Int. J. Innovative Comput. Inf. Control **7**(7), 4027–4046 (2011)
22. Lee, L.W., Wang, L.H., Chen, S.M.: Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques. Expert Syst. Appl. **34**, 328–336 (2008)
23. Lee, L.W., Wang, L.H., Chen, S.M.: Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms. Expert Syst. Appl. **33**, 539–550 (2007)
24. Cheng, H.C., Chen, T.L., Teoh, H.J., Chiang, C.H.: Fuzzy time-series based on adaptive expectation model for TAIEX forecasting. Expert Syst. Appl. **34**, 1126–1132 (2008)
25. Jilani, T.A., Burney, S.M.A.: A refined fuzzy time series model for stock market forecasting. Phys. A **387**, 2857–2862 (2008)
26. Yu, T.H.-K.: Weighted fuzzy time series models for TAIEX forecasting. Phys. A **349**, 609–624 (2005)
27. Liu, J.W., Chen, T.L., Cheng, C.H., Chen, Y.H.: Adaptive-expectation based multi attribute FTS model for forecasting TAIEX. Comput. Math. Appl. **59**(2), 795–802 (2010)

28. Sullivan, J., Woodall, W.H.: A comparison of fuzzy forecasting and Markov modeling. Fuzzy Sets Syst. **64**(3), 279–293 (1994)
29. Cheng, C.H., Cheng, G.W., Wang, J.W.: Multi-attribute fuzzy time series method based on fuzzy clustering. Expert Syst. Appl. **34**, 1235–1242 (2008)
30. Huarng, K., Yu, T.H.K.: The application of neural networks to forecast fuzzy time series. Phys. A **363**, 481–491 (2006)
31. Aladag, C.H., Basaran, M.A., Egrioglu, E., Yolcu, U., Uslu, V.R.: Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations. Expert Syst. Appl. **36**, 4228–4231 (2009)
32. Aladag, C.H., Egrioglu, E., Gunay, S., Yolcu, U.: High order fuzzy time series forecasting model and its application to IMKB. Anadolu Univ. J. Sci. Technol. **11**(2), 95–101 (2010)
33. Yu, T.H.K., Huarng, K.H.: A neural network- based fuzzy time series model to improve forecasting. Expert Syst. Appl. **37**, 3366–3372 (2010)
34. Aladag, C.H.: Using multiplicative neuron model to establish fuzzy logic relationships. Expert Syst. Appl. **40**(3), 850−853 (2013)
35. Cheng, S.M., Wei, L.Y., Chen, Y.S.: Fusion ANFIS models based on multi-stock volatility causality for TAIEX forecasting. Neurocomputing **72**, 3462–3468 (2009)

Using Zadeh's Granulation Concept: Granular Logics and Their Application to Sensor Data Analysis

Valery B. Tarassov and Maria N. Svyatkina

Abstract The main components of Zadeh's scientific contribution are shown. In particular, his information granulation theory is considered in the context of sensor data mining. The concept of granule and granular structures are presented, mereological approach to generating granules is developed, some ways of constructing and using granules in multi-valued logics are shown. Basic ideas of granular computing are discussed. Here the problem of sensor data mining is considered from the viewpoint of granular computing. The notion of logical pragmatics is clarified. The investigation of mathematical structures for data pragmatic analysis is associated with interpreting multi-valued logics and bilattices as formal pragmatics techniques. The concept of Belnap's sensor is proposed, the combinations of Belnap's sensors are discussed. Logical-algebraic models of communication between two sensors are constructed on the basis of bilattices. As a result, the map for dealing with different sensor data inconsistencies is introduced and algorithm for eliminating faults in sensor networks is suggested.

This work is partially supported by RFBR Grant No 10-01-00844-a, No 13-07-13133-ofi-m-RRLW and No 14-07-31317-mol-a.

V. B. Tarassov (✉) · M. N. Svyatkina
CIM Department, Bauman Moscow State Technical University, Moscow,
2nd Baumanskaya st. 5, 105005 Russia
e-mail: tarasov@rk9.bmstu.ru

M. N. Svyatkina
e-mail: maria.svyatkina@gmail.com

L. A. Zadeh et al. (eds.), *Recent Developments and New Directions in Soft Computing*, Studies in Fuzziness and Soft Computing 317,
DOI: 10.1007/978-3-319-06323-2_29,

1 Introduction

The contribution of Lotfi Zadeh to modern systems theory, information science, computer science and artificial intelligence is extremely broad and valuable (see, for instance [1–5]). He is the founder of Fuzzy Set Theory [6] and Linguistic Variable Concept [7], Possibility Theory [8] and Approximate Reasoning [9], Local and Fuzzy Logics [10], Soft Computing [11] and Computing with Words [12, 13]. In the beginning of 21st century such initiatives of the Father of Fuzzy Logic as Information Granulation Theory and Granular Mathematics development program are of special concern.

In 1979 his work entitled «Fuzzy Sets and Information Granularity» [14] was published, in which *information granules* were introduced in the form ((X is g) is α), where g is a granule represented by fuzzy constraint, and α is the possibility (plausibility) degree. Then in 1997 L. Zadeh formulated some fundamentals of the Theory of Fuzzy Information Granulation (TFIG) in his seminal paper «Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic» [15]. According to Zadeh, granulation is one of three basic operations that underlie human cognition: granulation, organization and causation. In particular, he specified *granule* as a collection of objects which are drawn together by indistinguishability, similarity, proximity or functionality. As a result, the well-known Zadeh's incompatibility principle was completed by the following Zadeh's granulation principle: "To exploit the tolerance for imprecision, employ the coarsest level of granulation which is consistent with the allowable level of imprecision" [16]. The arrival of information granulation means the transition from the ordinary machine-centric to human-centric approach in information gathering and knowledge discovery [17]. We face the problem of representing information on some required level of specificity-abstraction or precision-approximation.

Also in 1997 Lin coined the term "granular computing" [18] widely used nowadays. It is considered as an emerging conceptual and computational paradigm of information/knowledge processing, or as an umbrella term for theories, methodologies, techniques and tools that make use of information granules in the process of problem-solving [19]. Here information granules are seen as complex dynamic information entities which are formed by intelligent agents in the process of data abstraction and derivation of knowledge from information in order to achieve some goal. The level of granulation may be given as the number of objects in the granule related to the total number of granules. To differ from convenient data-driven numeric computing, granular computing is often referred as knowledge-oriented methodology [18].

By selecting different levels of granulation we can obtain different layers of knowledge. According to Zadeh, granular computing is a suitable foundation for computing with words, i.e. computation based on information described in natural language.

The implementation of granular computing requires both the construction of granulation theory and the development of nontraditional computations.

Granulation theory includes generation, interpretation, classification and representation of granules and granular structures.

Nontraditional computations are related to granular constraints and granular reasoning. Generally, granulation procedures play a leading part in the design and implementation of modern intelligent systems, for example, intelligent agents [20–22].

2 On the Concept of Granule and Granular Structures

The term *"granule"* is originated from Latin word *granum* that means grain, to denote a small particle in the real or imaginary world. The granules differ one from another by their nature, complexity, sizes, abstraction level.

Typical interpretations of granules are: part of the whole, cluster, sub-problem of the problem, piece of knowledge, variable constraints. For instance, decomposition may be seen as a transition to a lower level of granulation and composition inversely.

There exist various classifications of granules: physical and conceptual granules, crisp and fuzzy granules, one-dimensional and multidimensional granules, data and knowledge granules, time and space granules, etc.

Conventional formal approaches aimed to construct granules are based on: subsets, intervals, relations, partitions, hierarchies, clusters, neighborhoods, bags, graphs, semantic networks. Granulation is often induced by non-standard sets. Classical sets satisfy two basic postulates: (1) membership postulate; (2) distinguishability postulate. Non-standard sets do not meet these conditions. Here various approximations (sub-definite set theory, rough set theory) and distributions (probability theory, fuzzy set theory, Dempster-Shafer theory) arise.

2.1 Predecessors of Granularity Theory: Lesniewski's Mereology

The concept of Zadeh's granule is based on equivalence or, at least, tolerance relation that is a reflexive and symmetric binary relation. It is not the only way to specify granules: here order relations are worth considering too. In this context Lesniewski's mereology [23] may be viewed as a direct precursor of modern granulation theory. *Mereology* (also called partonomy) is the theory of parthood relations: both relations of part to whole and relations of part to part within a whole are studied.

Let us point out that mereology falls outside the scope of classical set theory. In Cantor's set theory the discernibility postulate and empty set are widely used. To differ from Cantor's approach the advocates of mereology: (a) make emphasis to the wholistic nature of set as a collective class; (b) employ only «part_of» relation; (c) do not use empty set.

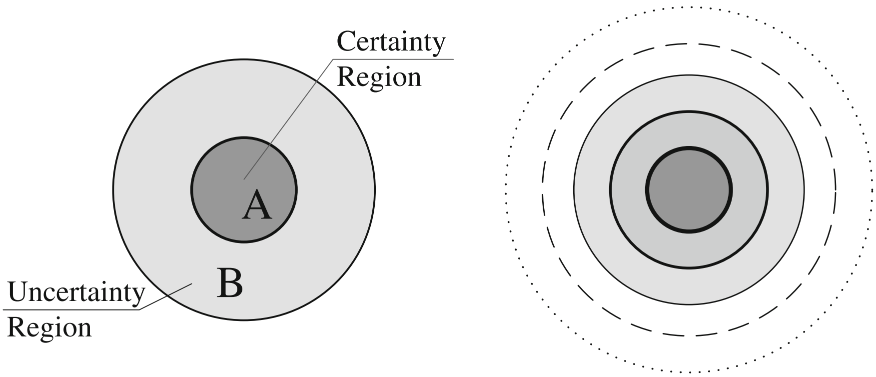


Fig. 1 Examples of non-standard sets: **a** Simple non-standard set given by set inclusion, **b** Flou set

Parthood relation is based on the following axioms:

1. Everything is part of itself (reflexivity axiom).
2. Two distinct things cannot be part of each other: if *A* is a part of *B,* then *B* is not a part of *A* (anti-symmetry axiom).
3. Any part of any part of the thing is itself part of this thing: if *A* is a part of *B* and *B* is a part of *C,* then *A* is a part of *C* (transitivity axiom).

Thus, the parthood relation is a partial ordering. Below we shall take mereological approach to granulation by considering non-standard sets (Fig. 1). The main characteristics of a set are its boundary, measure and power. Non-standard sets are relative objects with flexible boundaries which strongly depend on observer's information level. They usually have a non-strict boundary zone interpreted as uncertainty or imprecision area. The procedure of information granulation is reduced to specifying two regions: certainty region *A* and uncertainty region $B \backslash A$.

The transition from 2 to n nested sets $A_i \subset U$, i = 1,.., n, $A_1 \subseteq \ldots \subseteq A_n$ brings about granulation by flou set (ensemble flou) $A_{fl} = (A_1, \ldots, A_n)$, introduced by Gentilhomme [24]; if $n \to \infty$, then we obtain an infinite family of mappings $[0, 1] \to 2^U$. In Negoita and Ralescu [25] have proved the Stone-like theorem that the set of fuzzy subsets $[0, 1]^U$ in *U* and the family of ordinary (crisp) sets are isomorphic regarding to intersection and union operations.

2.2 Models of Granules and Granular Structures Based on Subsets

In this section we apply mereological approach presented above and illustrated in Fig. 1 to consider a rather natural way of specifying granules as subsets (crisp or fuzzy) of some universal set (see also [26, 27]).

Definition 1 Let *U* be a universal set. We call crisp granule any subset $g \in 2^U$, where 2^U is a power set of *U*.

So we power set 2^U consists of all possible crisp granules generated on the universe U.

Definition 1* Fuzzy granule is a fuzzy subset $g_f \in [0, 1]^U$, where $[0, 1]^U$ is a set of fuzzy subsets on U.

Here the set $[0, 1]^U$ consists of all possible fuzzy granules generated on the universe U.

Definition 2 For any two granules g, $g' \in 2^U$, if $g \subseteq g'$, then g is called subgranule of the granule g', and, in its turn, g' is a supergranule of g.

Granular structures may be generated by binary relations, for instance, by partially ordered or by equivalence relations.

Let us recall some necessary definitions. A set U provided with a relation $\leq$ that is reflexive, anti-symmetric and transitive is called partially ordered set (poset) $POS = \langle U, \leqslant \rangle$. Let X and Y be two posets. The direct product $X \times Y$ of two posets X and Y is specified as a set of pairs (x, y), where $x \in X$, $y \in Y$, such that $(x_1, y_1) \leq (x_2, y_2)$ iff $x_1 \leq x_2$ in X and $y_1 \leq y_2$ in Y.

Definition 3 Let G be a family of subsets g of the set U. A pair $GS = \langle G, \subseteq \rangle$ is called granular structure if $\subseteq$ is a set inclusion. From this definition it is clear that flou sets form granular structures.

Definition 3* Let G_f be a family of fuzzy subsets g_f of the set U, and $\subseteq_f$—fuzzy inclusion. Then a pair $\langle G_f, \subseteq_f \rangle$ is referred as fuzzy granular structure.

Now let us take an equivalence relation E that is reflexive, symmetric and transitive. The equivalence relation may be given by a mapping $[\bullet]_E : U \to 2^U$, $[x]_E = \{y \in U \,|xEy\}$. The subset $[x]_E$ is the equivalence class containing x.

Definition 4 An equivalence class $[\mathrm{x}]_E \in 2^U$ which contains non-distinguishable elements is a granule.

A family of all equivalence classes called *quotient set* and denoted by $U/E = \{[x]_E | x \in U\}$ generates granular structure.

Generally we shall specify the concept of granular algebraic system GAS in the framework of Maltsev's algebraic systems theory.

Definition 5 A granular algebraic system is a triple

$$GAS = \langle G, \ \Omega, \ \Pi \rangle, \tag{1}$$

where G is a non-empty set of granules called the basis of GAS, Ω is a set of operations over granules from G, Π is a predicate set expressing relations between granules.

In case of $\Pi = \varnothing$ GAS is reduced to granular algebra, and for $\Omega = \varnothing$ we obtain granular model or granular relational system.

We shall take these definitions to specify logical granules.

3 Granules in Multi-Valued Logics

Last years, we observe some interesting suggestions to bridge the gap between granular computing and multi-valued logics. Specifically, the attempts to construct granular logics have been performed [28, 29]. Below we shall discuss about granules and granular structures in logical semantics. The early predecessor of logical granulation was Vasiliev [30] who proposed two-leveled logical structure and introduced a well-known logical triangle with three values: (1) X is A; (2) X is not A; (3) X (is and is not) A simultaneously.

One of the first approaches to construct non-standard granular semantics was introduced by Dunn [31]. He proposed a new strategy of constructing logical semantics, where both gaps and gluts were possible. Here the term "*glut*" stands for granular truth value "both true and false" (it means the refusal from singular truth values of classical logic) and the gap is viewed as none (neither true nor false)—it extends bivalence principle of classical logic.

Let V be the set of logical truth values. Following Dunn, we may take as truth values not only elements $v \in V$, but also any subsets of V, including empty set $\varnothing$. In other words, a direct transition from the set V to its power set 2^V is suggested.

In his turn, Zadeh proposed further extension of this approach by introducing fuzzy truth values $v_f \in [0, 1]^V$ and linguistic truth labels [7, 10].

So the concept of granular logical matrix as a result of rethinking Lukasiewicz-Tarski's formalism may be introduced as follows.

Definition 6 Granular logical matrix is a triple

$$GLM = \langle 2^V,\ \Omega_\Lambda,\ D \rangle, \tag{2}$$

where 2^V is the power set of truth values, Ω_Λ is the set of logical operations, $\Omega_\Lambda = \{\rceil,\ \wedge,\ \vee,\ \rightarrow\}$, $D \in 2^V$ is the subset of designated truth values.

It is obvious from definitions 1 and 1* that any subset $g_\Lambda \in 2^V$ (or more generally $g*_\Lambda \in [0, 1]^V$) viewed as an extended truth value is a granule. In a simplest case for $|V| = 2$ we obtain basic Dunn-Belnap's logical semantics $2^V = V_4 = \{\mathrm{T, B, N, F}\}$, where T is true, F is false, B = both = {T, F}, N = none = {∅} [32].

In this context the generalized matrix introduced by R. Wojcicki may be rewritten in the form $LM_\mathrm{G} = \langle L,\ G \rangle$, where L is a lattice and $G \subseteq 2^L$ is a granular structure obtained as a family of sublattices of the lattice L.

Now let us apply Birkhoff-Jaskovski's approach to constructing product lattices and product logics as a way of generating logical granules. It is worth stressing that in the framework of algebraic logic various product logics are often given by direct product lattices. A simple example is the construction of granular logical values as $\mathrm{C}_1 \times \mathrm{C}_2 = \mathrm{C}_4$, viewed as the product of trivial lattices {0, 1} × {0, 1} (Fig. 2).

The applications of this semantics vary from Rescher's modalized truth to concerted truth for two experts (or sensors).

While constructing his four-valued logic, Belnap considered two order relations: truth order and set inclusion order {∅, {F}, {T}, {T, F}}. In the first case we

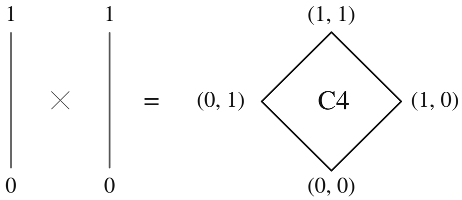


Fig. 2 Construction of minimal granular logical semantics C4 as a product of two classical semantics C_1 and C_2

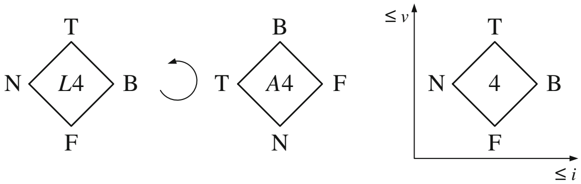


Fig. 3 Hasse diagrams for logical lattice L4, Scott lattice A4 and basic bilattice "four"

have F < N < T and F < B < T, but B and N placed between T and F are incomparable. In the second case, $\varnothing \subset \{T\} \subset \{T, F\}$ and $\varnothing \subset \{F\} \subset \{T, F\}$, i.e. T and F are incomparable. So truth order induces Belnap's logical lattice L4, and information order corresponds to Scott's lattice A4 (interpreted here as information lattice). Ordinary Hasse diagrams for these lattices, as well as double Hasse diagram for minimal bilattice are depicted in Fig. 3.

A natural way to generate information/truth granules consists in constructing product lattices and bilattices. The concepts of bilattice and prebilattice were introduced by Ginsburg [33] and Fitting [34] respectively. They already found a wide application in AI, for instance, in logic programs, truth maintenance systems, non-monotonic reasoning, partial and inconsistent information processing. In this chapter we propose a way of using bilattice theory to generate granules in sensor data mining.

Let L and K be lattices. Let us take the set $L \times K$ of ordered pairs (x, y), where $x \in L$, $y \in K$. The meet $\wedge$ and join $\vee$ operations are defined coordinate wisely:

$$(x_1, y_1) \wedge (x_2, y_2) = (x_1 \wedge x_2, (y_1 \wedge y_2),$$
$$(x_1, y_1) \vee (x_2, y_2) = (x_1 \vee x_2, (y_1 \vee y_2).$$

As a result we obtain a new lattice $L \times K$ called direct lattice product of L and K. Here the important special cases (for $L = K$) are multi-lattices $ML = L^m$ and, in particular, bilattices $BL = L^2$.

Informally, bilattice may be seen as a product lattice with two different ordering relations, where a link between these orders is of special concern. More detailed presentation of bilattice begins with the concept of bi-ordered set.

A bi-ordered set is a triple $BOS = \langle U, \leqslant_V, \leqslant_I \rangle$, where $U = X \times X$ is a product set, and $\leqslant_V, \leqslant_I$ are two order relations, namely truth order and information order. This structure becomes a pre-bilattice PBL if the components $\langle X, \leqslant_V, \rangle$ and $\langle X, \leqslant_I \rangle$ form complete lattices. For any PBL we can define operations $\wedge$ and $\vee$ as a greatest lower bound and a least upper bound on the lattice $\langle X, \leqslant_V, \rangle$ and the operations $\otimes$ and $\oplus$ as a greatest lower bound and a least upper bound on the lattice $\langle X, \leqslant_I \rangle$.

If two different order relations are linked by Ginsburg's negation operation $\neg_G$ that satisfies the following conditions

1. $\forall x, y \in U, x \leqslant_V y \Rightarrow \neg_G x \geqslant_V \neg_G y$ (anti-monotonicity by $\leq_V$);
2. $\forall x, y \in U, x \leqslant_I y \Rightarrow \neg_G x \geqslant_I \neg_G y$ (monotonicity by $\leq_I$);
3. $\forall x, y \in U, \neg(\neg x) = x$ (involution),

then pre-bilattice becomes a bilattice.

So the bilattice may be given as a quadruple [33]

$$BL = \langle U, \leqslant_V, \leqslant_I, \neg_G \rangle. \quad (3)$$

Finally, a logical bilattice [35] is a pair $LBL = \langle BL, BF \rangle$, where BL stands for bilattice and BF is a prime bifilter on BL.

Another way of logical granulation related to Vasiliev-Dunn's ideas consists in constructing vector semantics, for instance, for any proposition p from $\wp$ we have $v: \wp \rightarrow [0, 1]^2$, $v(p) = (\mathrm{T}(p), \mathrm{F}(p))$ (truth and falsity are considered here independently) or more generally $v: \wp \rightarrow [0, 1]^3$, $v(p) = (\mathrm{T}(p), \mathrm{F}(p), B(p))$. In the latter case, each proposition is characterized by degree of truth, degree of falsity and degree of inconsistency (or contradiction).

4 Logical Pragmatics in Sensor Data Analysis and Its Generalization

4.1 Interpreting Data in Sensor Networks

Sensor network is a group of specialized sensing devices with some communication infrastructure intended to collect heterogeneous data in a distributed (often decentralized) way. Recent advances in wireless communications and sensor technology have enabled the development of Wireless Sensor Networks (WSN) [36]. A typical WSN consists of a large number of low-cost and low-energy sensors, which are scattered in an area of interest to collect observations and preprocess them. Each sensor node has its own communication hardware that allows to link with other sensor nodes in a wireless fashion.

On the one hand, many applications of wireless sensor networks involve data exchange and collaboration among sensor nodes to achieve a common task. Moreover, here sensor negotiation and coalition formation processes arise. On the other hand, the problem of online sensor mining is faced in order to make intelligent decisions promptly.

In wireless network sensors have serious resource constraints; it requires the granulation of data stream for communicating sensors. Such a granulation may be seen as a mapping from continuous data set to a knowledge lattice. Three main operations of online mining in sensor networks were mentioned in [37]: (1) detection of sensor data irregularities; (2) clustering of sensor data; (3) discovery of sensory attribute correlations.

In the next sections we shall introduce logical approach to sensor data analysis and interpretation based on logical pragmatics and combination of Belnap's sensors.

4.2 A Few Words about Prescriptions, Norms and Logical Pragmatics

Nowadays the concept of logical truth is subject of hot discussions. Some logicians tend to conserve truth as a universal logical category by putting together three different truth theories: (1) correspondence theory; (2) coherence theory; (3) pragmatic theory. Following correspondence theory, truth statements are viewed as corresponding to actual state of affairs, coherence theory makes emphasis on logical consistency between generated hypotheses and existing knowledge, and pragmatic theory associates truth with practical utility.

Other logicians, for instance Ivin [38], stress on non-universality of truth and suggest to complement truth by other logical categories. Below we shall explain this viewpoint by taking agent-oriented approach in AI.

The activities of any agents are based both on descriptions and prescriptions which are associated with different information pieces. Descriptions contain information about external world's states perceived or represented by agent, whereas prescriptions are referred to possible agent's actions to this world (given by norms, orders, demands, and so on). Here descriptions are truth-valued, and prescriptions are characterized by some utility values.

The difference between descriptions and prescriptions is shown in Fig. 4. Here the truth value v is seen as a correspondence between world's object or process and its logical description (the primacy of real world object is supposed), whereas the utility u of any prescription is related to normative model (the primacy of norm is established and its applicability to some class of agents is viewed).

Moreover, in semiotics the term «semantics» expresses the relations between a message and its sender and «pragmatics»—between a message and its receiver (user). In our opinion, the development of logical approach to sensor data mining requires the introduction of logical pragmatics. To differ from logical semantics logical pragmatics makes emphasis on truth-values interpretation from the viewpoint of norms and preferences.

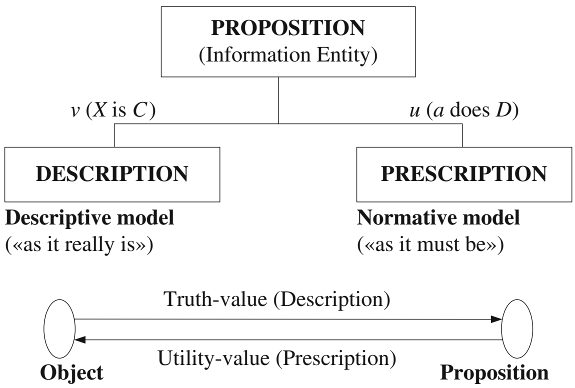


Fig. 4 On the difference between logical descriptions and prescriptions

4.3 Belnap's Sensors and Their Combination

In this chapter we introduce the concept of Belnap's sensor in the following way. Let us consider the sensor able to measure some parameter of inspected process and to interpret obtained information. Such a sensor equipped with four-valued data mining pragmatics will be referred as Belnap's sensor. The data obtained from this sensor are granulated according to four values:

- T—«pragmatic truth» («norm»—sensor data are located in a «green zone»);
- F—«pragmatic falsehood» («out of norm»—sensor data are located in «red zone»);
- B—«pragmatic ambiguity» («partial fault»—sensor data are located in «yellow zone»);
- N—«total uncertainty» (sensor resources are exhausted or sensor is «sleeping»).

The logical matrix for this Belnap's sensor may be written in the following way

$$LM_{\mathrm{V4}} = \langle \{\mathrm{T}, \mathrm{B}, \mathrm{N}, \mathrm{F}\}, \{\neg,\ \wedge,\ \vee\}\ \{\mathrm{F}\} \rangle, \tag{4}$$

where T, B, N, F are pragmatic truth, pragmatic ambiguity, total uncertainty and pragmatic falsehood respectively, $\neg$, $\wedge$, $\vee$ are main logical operations over truth values—negation, conjunction and disjunction, F is anti-designated value.

We can use product lattices and multi-lattices to construct the logic of sensor network. The logic of Belnap's sensor network is based on the expression 4^n, where n is an integer, $n > 1$. A simple combination of two Belnap's sensors gives $4^2 = 16$ pragmatic values, a system of three sensors—$4^3 = 64$, etc.

Now let us interpret basic granular pragmatic truth values for two communicating sensors:

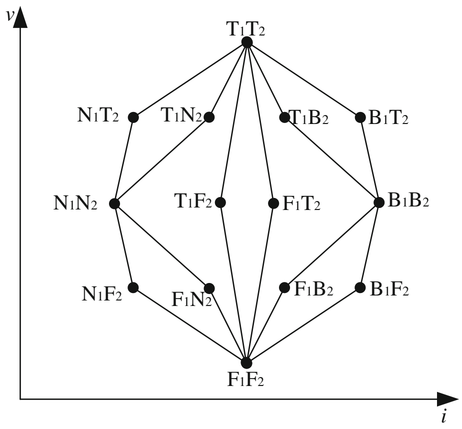


Fig. 5 Bilattice representation of two Belnap's sensors combination pragmatics

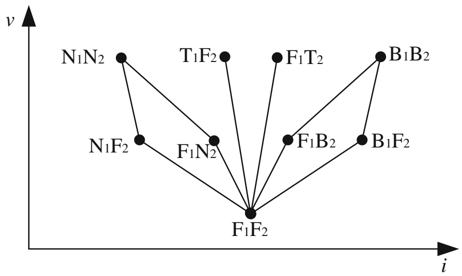


Fig. 6 A map of fault values for two communicating Belnap's sensors

- T_1T_2—«concerted truth» (data issues from both sensors take norm values);
- F_1F_2—«concerted falsehood» (both sensors data are out of norm that witnesses for failure state);
- $T_1B_2 \sim B_1T_2$—«partial contradiction as the first-order fault» (the first sensor shows a norm value and the second sensor indicates partial fault, and vice versa);
- $T_1N_2 \sim N_1T_2$—«partial truth with uncertainty» (the first sensor indicates a norm value and the second sensor sleeps, and vice versa);
- $T_1F_2 \sim F_1T_2$—«full contradiction» (the first sensor indicates a norm value, and the second sensor informs about failure state, and vice versa);
- B_1B_2—«concerted ambiguity» (the data from both sensors inform about first order fault);
- N_1N_2—«total uncertainty» (the resources of both sensors are exhausted or both sensors sleep;

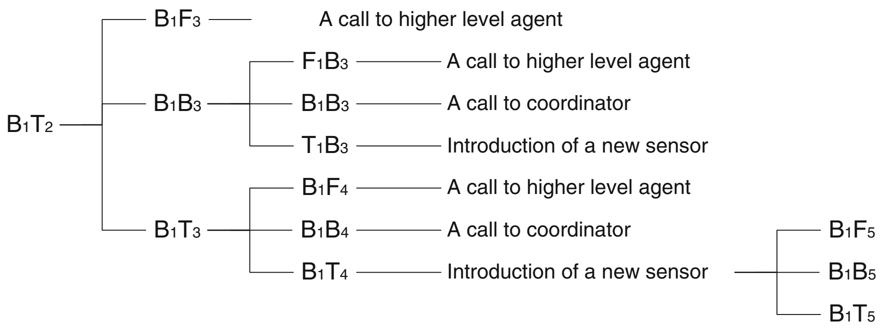


Fig. 7 Algorithm of dealing with faults in sensor networks

- $F_1B_2 \sim B_1F_2$—«partial contradiction as the second-order fault» (the first sensor indicates an out of norm value and the second sensor indicates a fault), and vice versa;
- $F_1N_2 \sim N_1F_2$—«partial falsehood with uncertainty» (the first sensor indicates a failure value and the second sensor sleeps), and vice versa.

The appropriate bilattice to illustrate the communication between two Belnap's sensors is shown in Fig. 5.

In Belnap's logic the values B and N are considered independently, hence for its extension the values B_1N_2 и N_1B_2 are forbidden.

Below we present the map of extended anti-designated values for two Belnap's sensors combination (Fig. 6).

4.4 Algorithm of Dealing with Faults in Sensor Networks

As a result the following procedures of solving partial contradictions in sensor networks are suggested (Fig. 7). The discovery of the second order fault supposes an immediate call to higher level decision agent. The solution in the situation of concerted ambiguity may be found by the coordinator (for instance, by replacing a pair of sensors) and elimination of the first order fault requires the adding of a new sensor into basic network.

Acknowledgment The authors are grateful to Alexey N. Averkin, Ildar Z. Batyrshin, Janusz Kacprzyk and Witold Pedrycz for useful discussion and valuable comments.

References

1. Kerre, E.: The Impact of Fuzzy Set Theory on Contemporary Mathematics. Int. J. Appl. Comput. Math. **10**(1), 20–34 (2011)
2. Turksen, B.: Scientific and Philosophical Contributions of L.A.Zadeh. In: Proceedings of the First International Conference on Soft Computing and Computing with Words in System Analysis, Decision and Control (ICSCCW-2001, Antalya, Turkey, 6–8 June 2001), b-Quadrat Verlag, Siegen, pp. 17–23 (2001)
3. Averkin, A.N., Batyrshin, I.Z., Tarassov, V.B.: Lotfi Zadeh & Russia. Artif. Intell. News **2–3**, 16–20 (2001). (in Russian)
4. Preface to the Proceedings of the International Conference on Integrated Models and Soft Computing in Artificial Intelligence (IMSC'2011, Kolomna, Russia, 16–19 May 2011), Physmathlit, Moscow, pp. 14–23 (2011) (in Russian)
5. Zadeh, L.A.: "Fuzzy Sets and Fuzzy Information Granulation". Key selected. In: Huan, Chonghu, Ruan, Da (eds.) Advances in Fuzzy Mathematics. Beijing Normal University Press, Beijing (2009)
6. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
7. Zadeh, L.A.: The concept of linguistic variable and its application to approximate reasoning. Inf. Sci. **8(**Parts 1 and 2), 199–249, 301–357 (1975)
8. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. **1**(1), 3–28 (1978)
9. Zadeh, L.A.: Theory of Approximate Reasoning. In: Hayes, J., Michie, D., Mikulich, L.I. (eds.) Machine Intelligence, pp. 149–194. Halstead Press, New York (1979)
10. Bellman, R., Zadeh, L.A.: Local and Fuzzy Logics. In: Dunn, J.M., Epstein, G. (eds.) Modern Uses of Multiple-Valued Logics, pp. 105–165. D. Reidel, Dordrecht (1977)
11. Zadeh, L.A.: Fuzzy logic, neural network and soft computing. Commun. ACM **37**(3), 77–84 (1994)
12. Zadeh, L.A.: Fuzzy logic = computing with words. IEEE Trans. Fuzzy Syst. **4**, 103–111 (1996)
13. Zadeh, L.A.: From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions. In: Wang P.P. (ed.) Computing with Words, pp. 35–68. Wiley and Sons, New York (2001)
14. Zadeh, L.A.: Fuzzy sets and information granularity. In: Gupta, M.M., Ragade, R.K., Yager, R.R. (eds.) Advances in Fuzzy Sets Theory and Applications, pp. 3–20. North-Holland, Amsterdam (1979)
15. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst. **90**, 111–127 (1997)
16. Zadeh, L.A.: Toward a logic of perception based on fuzzy logic. In: Novak, V., Perfilieva. I. (eds.) Discovering the World with Fuzzy Logic, pp. 4–28. Physica-Verlag, Heidelberg (2000)
17. Bargiela, A., Pedrycz, W.: Granular Computing: An Introduction. Kluwer Academic Publishers, Dordrecht (2003)
18. Lin, T.Y.: Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems. In: Skowron, A., Polkowski, L. (eds.) Rough Sets in Knowledge Discovery, pp. 107–140. Physica-Verlag, Heidelberg (1998)
19. Yao, Y.: Granular Computing: Basic Issues and Possible Solutions. In: Proceedings of the 5th Joint Conference on Information Sciences
20. Wooldridge, M.: An Introduction to Multi-Agent Systems, 2nd edn. Wiley, Chichester (2009)
21. Tarassov, V.: From Multi-Agent Systems to Intelligent Organizations. Editorial URSS, Moscow (2002) (in Russian)
22. Kalutskaya, A.P., Tarassov, V.B.: Information Granules and Theit Construction Techniques: Application to the Development of Intelligent Agents. In: Proceedings of the 10th International Conference on Intelligent Information Analysis (IIA'2010, Kiev, Ukraine, 18–21 May 2010). Prosvita Editions, Kiev, pp. 291–297 (2010) (in Russian)

23. Surma, S.J., Srzednicki, J., Barnett, D.I., Rickey, F.V.: St. Leśniewski, Collected Works, vol. 1, pp. 129–173. Kluwer Academic Publishers, Dordrecht (1992)
24. Gentilhomme, Y.: Les sous-ensembles flous en linguistique. Cahiers de Linguistique Theorique et Appliquee **5**, 47–63 (1968)
25. Negoita, C.V., Ralescu, D.A.: Application of Fuzzy Sets to System Analysis. Birkhauser Verlag, Basel (1975)
26. Tarassov, V.B.: Information granulation by cognitive agents and non-standard fuzzy sets. In: Proceedings of the 6th International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control (ICSCCW'2011, Antalya, Turkey, 1–2 Sept 2011). b-Quadrat Verlag, Kaufering, pp. 59–74 (2011)
27. Yao, Y., Zhang, N., Miao, D., Xu, F.: Set-theoretic approaches to granular computing. Fundamenta Informaticae **115**(2–3), 247–264 (2012)
28. Zhou, B., Yao, Y.: A logic approach to granular computing. Int. J. Cogn. Inform. Nat. Intell. **2**(2), 63–79 (2008)
29. Liu, Q., Liu, L.: A description language based on granular computing—granular logic. In: Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (Moscow, Russia, 25–27 June 2011). Lecture Notes in Artificial Intelligence, vol. 6743, pp.91–94. Springer, Heidelberg (2011)
30. Vasiliev, N.A.: Imaginary Logic. Nauka, Moscow (1989) (in Russian)
31. Dunn, J.M.: Intuitive semantics for first degree entailment and coupled trees. Philos. Stud. **29**, 149–168 (1976)
32. Belnap, N.: A useful four-valued logic. In: Dunn, J.M., Epstein, G. (eds.) Modern Uses of Multiple-Valued Logic, pp. 8–37. D. Reidel, Dordrecht (1977)
33. Ginsberg, M.: Multivalued Logics: a Unified Approach to Reasoning in AI. Comput. Intell. **4**, 256–316 (1988)
34. Fitting, M.: Bilattices and the theory of truth. J. Philos. Logic **18**, 225–256 (1989)
35. Arieli, O., Avron, A.: Reasoning with logical bilattices. J. Logic Lang. Inform. **5**(1), 25–63 (1996)
36. Vinyals, M., Rodriguez-Aguilar, J.A., Cerquides, J.: A survey on sensor networks from a multiagent perspective. Comput. J. **54**(3), 455–470 (2011)
37. Ma, X., Yang, D., Tang, Sh., Luo, Q., Zhang, D., Li, Sh.: Online mining in sensor networks. Netw. Parallel Comput. (Lecture Notes in Computer Science) **3222**, 544–550 (2004)
38. Ivin, A.A.: Argumentation Theory. Vysshaya Shkola Publishers, Moscow (2007) (in Russian)