

PROBABILISTIC-FUZZY KNOWLEDGE-BASED SYSTEM FOR MANAGERIAL APPLICATIONS

Katarzyna Rudnik¹, Anna Walaszek-Babiszewska²

¹ *Opole University of Technology, Faculty of Production Engineering and Logistics, Poland*

² *Opole University of Technology, Faculty of Electrical Engineering, Automatic Control and Computer Science, Poland*

Corresponding author:

Katarzyna Rudnik

Opole University of Technology

Faculty of Production Engineering and Logistics

Institute of Processes and Products Innovation

Ozimska 75, 45-370 Opole, Poland

phone: +48 77 4234035

e-mail: k.rudnik@po.opole.pl

Received: 27 August 2011

Accepted: 17 November 2011

ABSTRACT

The paper deals with an inference system with probabilistic-fuzzy knowledge base as a tool which can help users in analyzing complete uncertainty of real problems in the company using fuzzy sets and probability. In the mentioned system, knowledge is saved in the weighted IF-THEN fuzzy rules, where the weights constitute marginal probabilities of the fuzzy events in the antecedents and conditional probabilities of the fuzzy events in the consequents. Moreover, the paper presents a proposal of the use of fuzzy association rules as a method of automatic knowledge base extraction in the inference system. For this purpose a modification of the PF-Growth algorithm is described. A numerical example is analyzed by using wind speed prediction process. The correct estimation of wind speed, as the potential energy resource, is necessary for control of the wind turbine work and it is important for the localization process of wind turbines, production planning and estimating cost-effectiveness of such investments.

KEYWORDS

knowledge-based system, fuzzy system, probability of fuzzy events, fuzzy association rules, manager decision.

Introduction

Within a wide scope of enterprises management, there are a lot of tasks that involve limited knowledge and uncertainty as to the course of events and activities concerning the management of facilities and objects. Such circumstances result from the pace, range, grade and depth of changes propagation in the globalized economy which is constituted by seemingly local endeavours [1], but, which is also an outcome of natural phenomena that still remain uncontrollable for the minds of scientists. Hence, one may distinguish the following types of knowledge deficiency: incompleteness of information - also referred to as subjective uncertainty, and objective uncertainty resulting from the characteristics of the analyzed processes and objects. To describe and systematise

the knowledge concerning these issues, fuzzy logics models have been applied as discussed in references [1–4].

At the same time, in many financial and decision-making situations, companies have to deal with uncertainty due to randomness, i.e. the random nature of the values of social phenomena (sick rate among employees), financial ratios (Stock Exchange indices) or even geological phenomena (speed and direction of wind measured in the course of estimating the profitability of investments in wind power plants) that are used in numerous decision-making processes in companies, as far as operational, tactical and strategic management is concerned. For the modelling of such processes the mathematical models are generally applied in consideration of the theory of probability.

Based on inspiring Zadeh's idea of Soft Computing (1994), using fuzzy sets theory and fuzzy logic, as well as theory of probability and stochastic processes, the methodology of the creation of the knowledge representation and the procedures of reasoning of probabilistic and stochastic SISO, MISO and MIMO systems has been formulated (e.g. [5, 6]).

It seems natural that the combination of the two methods of analysis – the fuzzy logics theory and the probability theory, should provide a complete description of the uncertainty of real problems that occur in processes of company management (Fig. 1), (e.g. [7, 8]). Accordingly, this was the reason for the origin of the concept of an inference system based on the probability-fuzzy knowledgebase [9, 10].

Type of uncertainty	Subjective uncertainty resulting from human ignorance	Objective uncertainty resulting from object's characteristic	Uncertainty resulting from randomness
Method of uncertainty modelling	Modelling of uncertainty using the theory of fuzzy logic		Modelling of uncertainty using the theory of probability
Complete description of uncertainty which occur in enterprise			

Fig. 1. The concept of the uncertainty description in managerial situations.

In the discussed system, the linguistic knowledge is contained within weighted IF-THEN rules, constituting the boundary conditions and conditional probability of fuzzy events in the antecedent and consequent of the rules. On principle, a fuzzy system should enable a simplified reconstruction of a complex research problem. However, in consideration of multiple variables of the system and a big number of the identified fuzzy sets, the creation of the said model involves complex calculation procedures. Furthermore, considering the total distribution of the probability of fuzzy events, the number of elementary rules of the database is N^m , where N is a number of model variables, m – number of fuzzy sets of the variable (assuming the equal number of the fuzzy sets for each variable). A big number of rules exerts an impact not only on the time of model identification, difficulties in reasoning with the use of the created database, but also on possible implementation in real objects. It will be shown in the paper that it is possible to use the concept of association rules search (one of the data mining methods) for obtaining the parameters of a model of an inference system with probabilistic-fuzzy database. The algorithm shall make it possible to directly find credible fuzzy rules with their corresponding weights, enabling the inference on the grounds of the constructed model.

Concept of inference system with a probability-fuzzy knowledge base

Inference fuzzy systems are knowledge-based systems that use the linguistic approach in modeling and reasoning (inferring), also often labeled as fuzzy inference and modeling [2, 11–13]. In the linguistic approach, variables assume the values described by the categories of a natural language, for example: high, medium, low costs, adequate, inadequate quality [14]. Verbal assessment makes it possible to describe uncertain knowledge of the analyzed variables, but, at the same time, it is semantically dependent on real domains.

In a classic approach of the theory of fuzzy logics fuzzy set A defined in non-empty space \aleph , is determined by the characteristic function, also called membership function in the following form [15]:

$$\mu_A : \aleph \rightarrow [0, 1], \tag{1}$$

where $[0,1]$ denotes the range of real numbers from 0 to 1.

The discussed inference system is based on an analogical approach to the definition of fuzzy sets, however, instead of the membership function, constant grades of membership are used, defined for separable ranges of variable values (for example: [7] and [8]). The legitimacy of applying the grades of membership in modelling the system was demonstrated in [10]. The discretization of the space of the variable values has a slight effect on the structure of the model, yet, it shortens the time of rules generation and inference. A longer time of rules generation and inference (reasoning) in view of the database defined by the membership function results from the impossibility of introducing vector-based calculations, which is important in the course of implementing the system in Matlab computational environment.

The inference system with a probabilistic-fuzzy knowledge base is a Multiple Input-Single Output (MISO) type of system, with many inputs but just one single output. The structure of such system and its connections with the decision-making environment are shown in Fig. 2.

The discussed inference system is composed of the following parts (compare: [2, 11–13, 17, 18]):

- knowledge base that contains information essential for a given problem,
- fuzzification block that transforms quantitative data into qualitative data represented by fuzzy sets on the bases of membership grades entered in the database,

- inference block that utilizes the database and the implemented aggregation methods and final inference (reasoning) to solve specialized problems,
- defuzzification block that calculates the crisp value (defuzzified value) at the system output on the bases of the resulting membership grades.

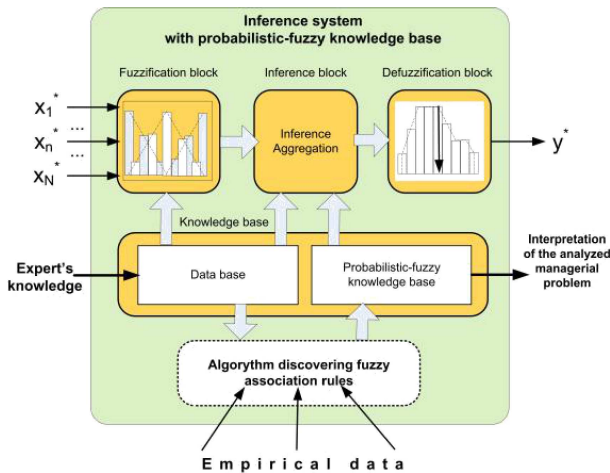


Fig. 2. The structure of the inference system with the probabilistic-fuzzy knowledge base [16].

The probabilistic-fuzzy knowledgebase contains two components: database and rule base (compare: [19]). The database contains information defined by experts on a given application field containing linguistic values $\{A_m^n, B_j/m, j = 1, \dots, J, n = 1, \dots, N, m = 1, \dots, M\}$ (compare model (4)) of the variables accounted in the rule base and definitions of fuzzy sets identified with these values. On the other hand, the probabilistic-fuzzy rules database, as the name itself indicates, contains a set of linguistic rules in the form of (4), which are created on the grounds of a modified algorithm generating fuzzy association rules. The algorithm makes it possible to adjust the model to measurement data. The characteristic form of the rules, exposing an empirical probability distribution of fuzzy events enables a simple interpretation of the knowledge contained in the model and additional analysis of the considered problem. Detailed information on particular blocks of the system shall be described in the successive sub-chapters of the paper.

Probabilistic-fuzzy representation of knowledge

Among a wide range of formal representations of the knowledge of inference systems, one of most easily accessible methods of recording human knowledge is the IF-THEN rule, derived from mathematical logics that describe inference or decision-making rules

and consisting of conditional part p_r , referred to as antecedent or premise and decision part q_r , referred to as consequent. Hence, the general form of the rule is as follows:

$$\text{IF } p_r \text{ THEN; } q_r \tag{2}$$

where the terms IF and THEN are the key words preceding the antecedent and the conclusion of the rule, respectively.

Detailed form of the rules database is dependent on the applied model. The simplicity of notation and easiness of interpretation and reasoning (inference) on the grounds of IF-THEN models have been very popular in fuzzy models and neuron-fuzzy models applications. One of their advantages is the option of devising models on the grounds of little information on the system in comparison with mathematical models [12]. An attempt to achieve better accuracy of representing various real systems, in consideration of different degree of accessibility of information and its form, has led to intense development of the structure of models [13].

The fuzzy model is a set of fuzzy conditional rules of the type

$$\text{IF } x \text{ is } A_i \text{ THEN } y \text{ is } B_j, \tag{3}$$

determining the reason-result relationship between linguistic variables [14] of the system: x – an input linguistic variable, y – an output linguistic variable and A_i, B_j – linguistic values of the variable. The linguistic values A_i, B_j are represented by fuzzy sets determined on the spaces of the model input and output.

The inference system based on probabilistic-fuzzy knowledge base discussed in this paper is based on a model which in professional literature, it is often referred to as: ‘fuzzy model’ or ‘fuzzy knowledge representation with probability measures of fuzzy events’ [5]. A typical feature of the knowledge base of the discussed model for MISO systems is the representation of knowledge in the form of the collection of file rules [7]:

$$\begin{aligned} &\text{IF } x_1 \text{ is } A_o^{(1)} \text{ AND...} \\ &\text{AND } x_N \text{ is } A_o^{(N)}, \text{ with weight } w_o \\ &\text{THEN } y \text{ is } B_{1/o}, \text{ with weight } w_{1/o} \\ &\dots \\ &\text{THEN } y \text{ is } B_{l/o}, \text{ with weight } w_{l/o} \\ &\dots \\ &\text{THEN } y \text{ is } B_{L/o}, \text{ with weight } w_{L/o}, \end{aligned} \tag{4}$$

where o – number of the file rule, N – number of the model input variable, $A_o^{(1)}, \dots, A_o^{(N)}, B_{l/o}$ – fuzzy sets representing the values of linguistic input variables x_1, \dots, x_N and the output variable y , in the l -th

elementary rule of the o -th file rule, w_o – weight of the o -th file rule, representing the probability of the joint fuzzy event $(x_1 \text{ is } A_o^{(1)}) \cap \dots \cap (x_N \text{ is } A_o^{(N)})$ in the antecedent, $w_{l/o}$ – weight of the l -th elementary rule, representing the conditional probability of the fuzzy event $(y \text{ is } B_{l/o})$ in the consequent, subject to $(x_1 \text{ is } A_o^{(1)}) \cap \dots \cap (x_N \text{ is } A_o^{(N)})$.

We can write the weight w_o as a joint probability of the fuzzy event $A_o = A_o^{(1)} \times \dots \times A_o^{(N)}$ in the antecedent:

$$P(\mathbf{x} \text{ is } A_o) = P((x_1 = A_o^{(1)}) \cap \dots \cap (x_N = A_o^{(N)})), \quad (5)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_N]$.

The weight of the consequent part of the elementary rule $w_{l/o}$, according to [20], as the conditional probability of fuzzy events, is calculated as follows

$$\begin{aligned} & P((y \text{ is } B_{l/o}) / (\mathbf{x} \text{ is } A_o)) \\ &= \frac{P((y = B_{l/o}) \cap (\mathbf{x} = A_o))}{P(\mathbf{x} = A_o)}. \end{aligned} \quad (6)$$

Probability of fuzzy event A , according to Zadeh, is defined as [20]:

$$P(A) = \sum_{i=1}^n p(x_i) \mu_A(x_i), \quad (7)$$

where $p(x_i) \in [0, 1]$ is a probability in the sense of the theory of probability, x_i is an element of the discrete space of consideration $\aleph = \{x_1, x_2, \dots, x_n\}$, and $\sum_{i=1}^n p(x_i) = 1$.

According to [5], weights are calculated for linguistic values $A_o^{(n)}$ and $B_{l/o}$ determined on real sets X_n , $n = 1, \dots, N$ and Y of values of systems variables. The spaces X_n , $n = 1, \dots, N$ and Y are divided into disjoint intervals $a_k^{(n)}$ and b_k , respectively, and the fuzzy sets $A_o^{(n)}$, $B_{l/o}$ can be presented as the following sums:

$$\begin{aligned} A_o^{(n)} &= \sum_{k=1, \dots, K} \mu_{A_o^{(n)}}(a_k^{(n)}) / a_k^{(n)}, \\ B_{l/o} &= \sum_{k=1, \dots, K} \mu_{B_{l/o}}(b_k) / b_k, \end{aligned} \quad (8)$$

$$\begin{aligned} & P((y \text{ is } B_{l/o}) / (\mathbf{x} \text{ is } A_o)) \\ &= \frac{\sum_{a_k^{(1)}, \dots, a_k^{(N)}, b_k \in X_1 \times \dots \times X_N \times Y} [p(a_k^{(1)}, \dots, a_k^{(N)}, b_k) \cdot T(\mu_{A_o^{(1)}}(a_k^{(1)}), \dots, \mu_{A_o^{(N)}}(a_k^{(N)}), \mu_{B_o}(b_k))]}{\sum_{a_k^{(1)}, \dots, a_k^{(N)} \in X_1 \times \dots \times X_N} [p(a_k^{(1)}, \dots, a_k^{(N)}) \cdot T(\mu_{A_o^{(1)}}(a_k^{(1)}), \dots, \mu_{A_o^{(N)}}(a_k^{(N)}))]} \end{aligned} \quad (11)$$

where the following relationships must be fulfilled

$$\begin{aligned} \sum_{A_o^{(n)}} \mu_{A_o^{(n)}}(a_k^{(n)}) &= 1, \quad \forall a_k^{(n)}, \quad n = \text{const} \\ \sum_{B_{l/o}} \mu_{B_{l/o}}(b_k) &= 1, \quad \forall b_k. \end{aligned} \quad (9)$$

The exemplary definitions of fuzzy sets, according to the above rules, are presented at Fig. 3.

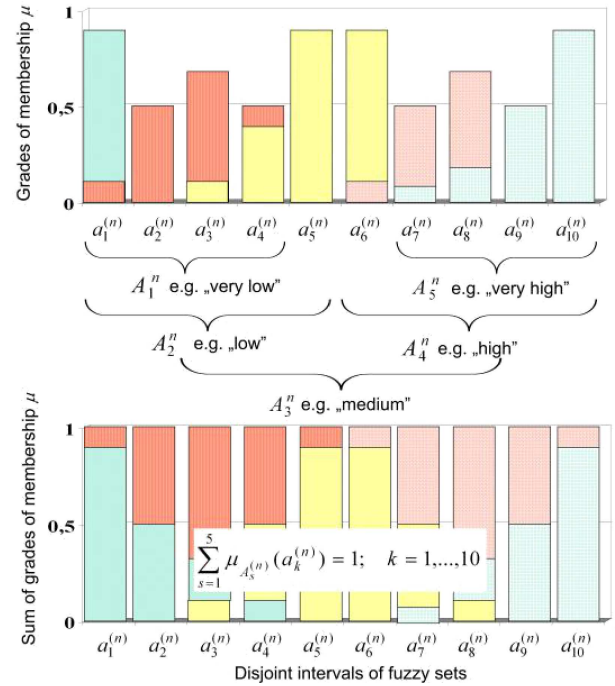


Fig. 3. Example of the fuzzy sets defined on disjoint intervals of the space X_n [16].

According to [5] and using the notation in formulas (5)–(8) we can write the following relationships for weights w_o and $w_{l/o}$, respectively:

$$\begin{aligned} & P(\mathbf{x} \text{ is } A_o) \\ &= \sum_{a_k^{(1)}, \dots, a_k^{(N)} \in X_1 \times \dots \times X_N} [p(a_k^{(1)}, \dots, a_k^{(N)}) \\ &\quad \cdot T(\mu_{A_o^{(1)}}(a_k^{(1)}), \dots, \mu_{A_o^{(N)}}(a_k^{(N)}))], \end{aligned} \quad (10)$$

More often the product t-norm is used.

Assuming that each measurement x_i is equally probable, the probability of fuzzy event A may be also calculated by means of membership coefficients of particular elements to fuzzy set A , in the following way:

$$P(A) = \frac{1}{I} \sum_{i=1}^I \mu_A(x_i). \quad (12)$$

Methods of fuzzy knowledge discovery

The IF-THEN rules that constitute the knowledge bases of the fuzzy system may be defined in two ways:

- as logical rules constituting subjective definitions created by experts on the grounds of experience and knowledge of the investigated phenomenon,
- as physical rules constituting objective knowledge models defined on the grounds of observations and natural research into the analyzed process (object) and its regularities.

In the case of fuzzy modeling there were initially logics rules, yet, in consideration of machine learning a hybrid of rules was gradually implemented according to which initial assumptions concerning fuzzy sets and the associated rules are defined following the experts' conviction, whereas other parameters are adjusted to measurement data. The objective of automatic data discovery is to obtain the smallest set of IF- THEN rules enabling as accurate representation of the modeled object or phenomenon as possible.

Methods of knowledge discovery for fuzzy systems of Mamdani type include [13]:

- Wang-Mendel method,
- Nozaki-Ishibuchi-Tanaki method,
- Sugeno-Yasukawa method,
- template-based method of modelling fuzzy systems [2].

In order to obtain databases for fuzzy systems, data mining methods have also been applied.

Data mining, considered as the main stage in knowledge discovery [21] is focused on non-trivial algorithms of searching "hidden", so far unknown and potentially required information [22] and its records in the form of mathematical expressions and models. Some of the data mining methods identify zones in the space of system variables, which, consequently, create fuzzy events in the rules. This may be accomplished by searching algorithm clusters or covering algorithms, also called separate and conquer algorithms. Other methods, for example: fuzzy association rules, are based on constant division for each

attribute (fuzzy grid) and each grid element is regarded as a potential component of the rule. As far as the first approach is concerned, each identified rule has its own fuzzy sets [23]. Therefore, from the point of view of rules interpretation, the second approach seems more applicable [24].

Association rules as ways of fuzzy knowledge discovery

Irrespective of automatic knowledge discovery, rules of the fuzzy model are obtained on the bases of their optimal adjustment to experimental data. In view of this, the generation of the rules may be understood as a search for rules with high occurrence frequency, where, the frequency parameter influences the optimal rules adjustment. In such case, rules in the form of (4) may be analyzed as the co-existence of fuzzy variable values in experimental data, i.e.: fuzzy association rules.

The issue of association rules was first discussed in [25]. Nowadays it is one of the most common data mining methods. In a formal approach, the association rules have the form of the following implications:

$$X \Rightarrow Y(\text{sup}\%, \text{conf}\%), \quad (13)$$

where X and Y are separable variable sets (attributes) in the classic approach to mathematical sets, often referred to as: X – conditioning values set, Y – conditioned values set.

Considering the fuzzy rules of association, the following may be derived:

$$A_1 \cap \dots \cap A_n \Rightarrow A_{n+1} \cap \dots \cap A_m (\text{sup}\%, \text{conf}\%), \quad (14)$$

where A_1, A_n are shortened notations of: variable - fuzzy set in the rule antecedent (i.e. $A_n \cong x$ is A_n , where x is a variable, A_n – fuzzy set, A_{n+1}, A_m pairs, variable - fuzzy set of the rule consequent.

Each association rule is connected with two statistical measures that determine the validity and power of the rule: *support* (sup%) – support, probability of concurrent/simultaneous incidence of set ($X \cap Y, A_1 \cap \dots \cap A_n \cap A_{n+1} \cap \dots \cap A_m$) in the set collection and *confidence* (conf%) – also called *credibility* which is conditional probability ($P(Y|X), P(A_{n+1} \cap \dots \cap A_m | A_1 \cap \dots \cap A_n)$). The issue of discovering fuzzy association rules involves finding, in a given database, all support and trust values that are higher than the association rules the support and trust of which are higher than the defined minimal values of support and trust given by users.

The first application of the association rules was in basket analysis. However, taking into account the fact that the rules may include variables that

are derived from diverse variables expressed in a natural language, the ranges of the application of the discussed method may be extended to decision-making, planning, control, forecasting, etc. Fuzzy association rules, as shown in this paper, may be also applied in acquiring data for fuzzy inference systems.

Algorithm of knowledge discovery based on fuzzy association rules

A basic algorithm of association rules discovery is an Apriori iterative algorithm [25]. It has been subject of many modifications aimed at improving its efficiency (for example: AprioriTid, AprioriHybrid). There are also other algorithms to be found in professional publications: SETM, FreeSpan, Eclat, Partition. The FP-Growth algorithm is effective due to its calculations complexity [26], but, it generates association rules only in non-fuzzy version (13). The discovery of fuzzy association rules (14) is possible by means of algorithms described in, for example [27–29]. The combination of data exploration by means of fuzzy association rules with genetic algorithms was discussed in [30]. This paper presents the authors' own version of the said algorithm created on the assumptions of FP-Growth, to be applied for generating the database with probabilistic-fuzzy rules for multiple input and single output.

The input of the proposed algorithm:

- set I of measurements used for model identification,
- predefined database, linguistic values of variables considered in the model and definitions of fuzzy sets identified with the values,
- threshold value of minimal support ($min w$).

The output: rule base of a probabilistic-fuzzy knowledge base.

Notations used in the presentation of the algorithm:

I – number of measurements used for the identification of the knowledge model,

$N + 1$ – total number of variables (N input data, one output data),

K – number of separable ranges with equal width in variable ranges,

x_n – model input variables, $x_n \in X_n \subset R$, $n = 1, \dots, N$,

y – model output variable, $y \in Y \subset R$,

$|A^{(n)}|$ ($|B|$) – number of the linguistic values for the n -th input variable (output variable),

$A_j^{(n)}$ – j -th linguistic value of n -th input variable, $j = 1, \dots, |A^{(n)}|$, $n = 1, \dots, N$,

B_j – j -th linguistic value of output variable, $j = 1, \dots, |B|$,

$\mathbf{a}^{(n)} = (a_1^{(n)}, \dots, a_k^{(n)}, \dots, a_K^{(n)})$ – separable ranges of the n -th input variable x_n , $n = 1, \dots, N$,

$\mathbf{b} = (b_1, \dots, b_k, \dots, b_K)$ – separable ranges of output variable y ,

w – calculated support value for frequent sets candidates,

$min w$ – assumed minimal support value,

C_r^* – set consisting of candidates of frequent fuzzy events of r -elements ($1 \leq r \leq N + 1$) of the system variables,

F_r^* – set consisting of frequent fuzzy events of r -elements ($1 \leq r \leq N + 1$) of the system variables,

D^i – i -th set of empirical values of the model $\{x_1^i, \dots, x_N^i, y^i\}$, $i = 1, \dots, I$ (i -th measurement),

J_i^{**} – number of the created combinations of fuzzy events from the i -th set of experimental data D^i ,

J^* – number of unrepeatable combinations of single fuzzy events of the variables in the antecedent and the consequent of the rule.

A frequent set is a set of which the probability of occurrence is bigger than the value of the assumed minimal support $min w$.

The algorithm for generating the rules for the inference system with probabilistic-fuzzy knowledge base is shown in Fig. 4.

An attempt at adopting the Apriori algorithm has been made to enable a direct search of fuzzy association rules. However, this version of the algorithm has turned out to be very complex in calculations in comparison with the modified FP-Growth algorithm. The time of generating the rules of this algorithm is longer, even though the same results are obtained (Fig. 5).

In paper [16] inverse results of the comparison are presented, yet, this is not contradictory- as the papers consider different versions of FP-Growth algorithms. Paper [16] considers the calculations of fuzzy probability on the bases of (10), but, in this article the calculations based on the power of fuzzy sets were used (12).

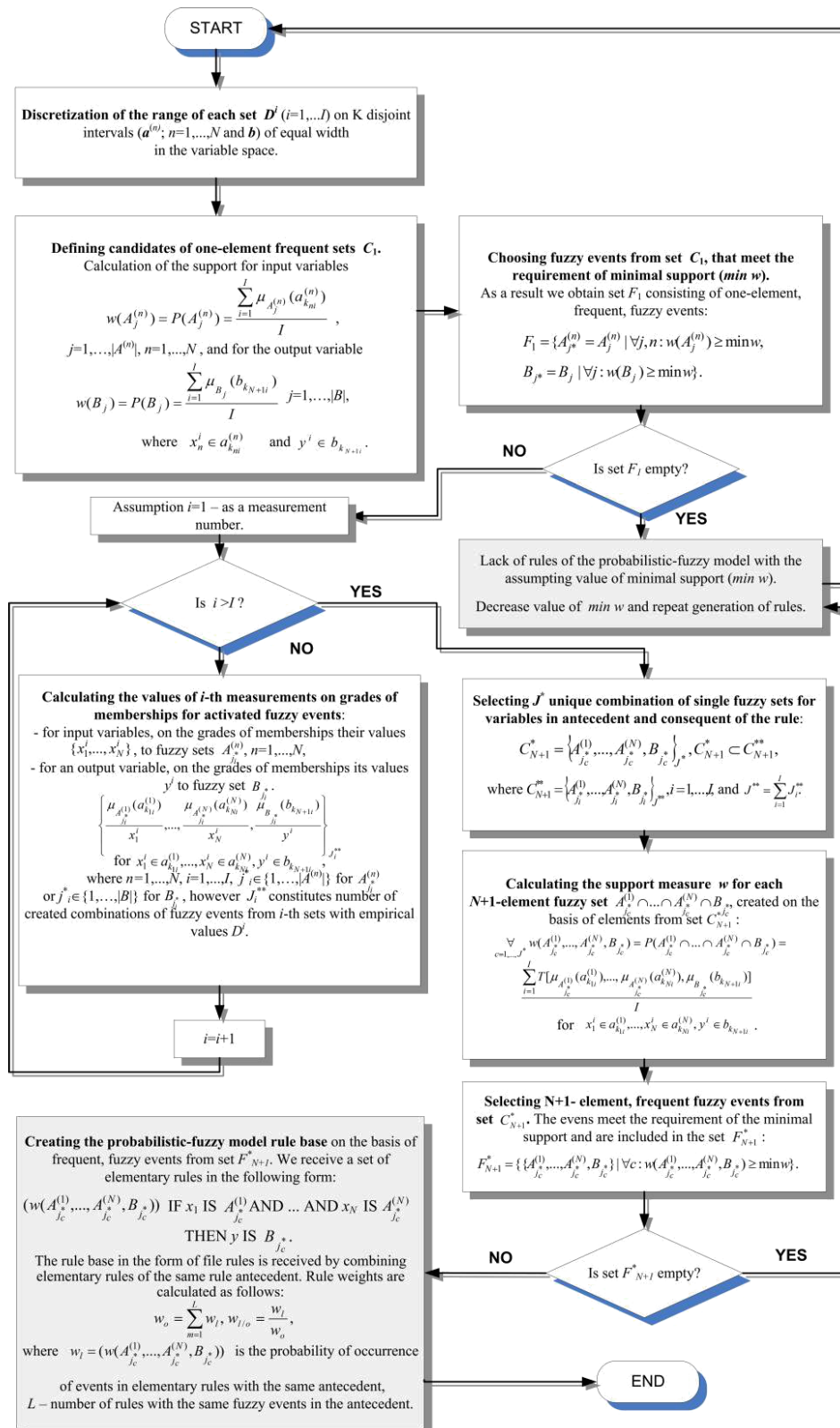


Fig. 4. The scheme of generating the rules for the inference system with probabilistic-fuzzy knowledge base, based on modified FP-Growth algorithm.

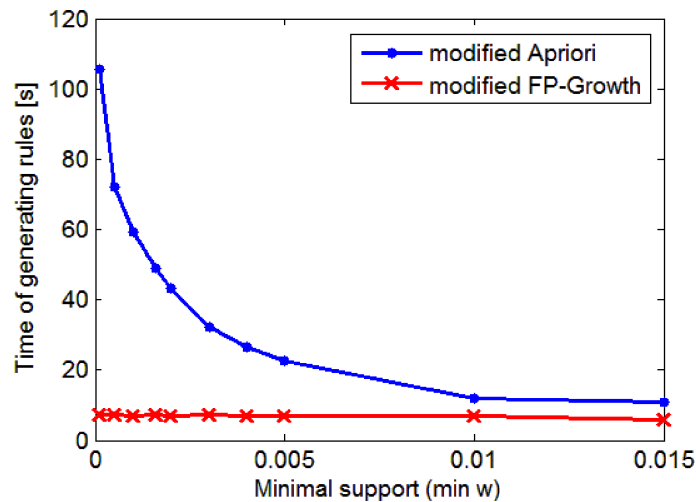


Fig. 5. The time of generating the fuzzy association rules as the function of the minimal support value for the modified Apriori and FP-Growth algorithms (3 input variables, one output variable, 7 fuzzy sets for each variable, 1000 learning data).

Inference based on the constructed model

The fuzzy model inference mechanism with multiple inputs and a single output enables the calculation of the membership function of the conclusion, on the bases of the crisp input data, and, in consequence, the defuzzified value of the model output.

For the system with the rules base in the form of (4), there are many possible ways of obtaining non-fuzzy results y^* [6]. One way of inference for an exemplary file rule is presented in Fig. 6. The inference takes advantage of the following parameters: the minimum as t-norm operator, the algebraic product as the inference operator and the Centre of Area as the defuzzification method.

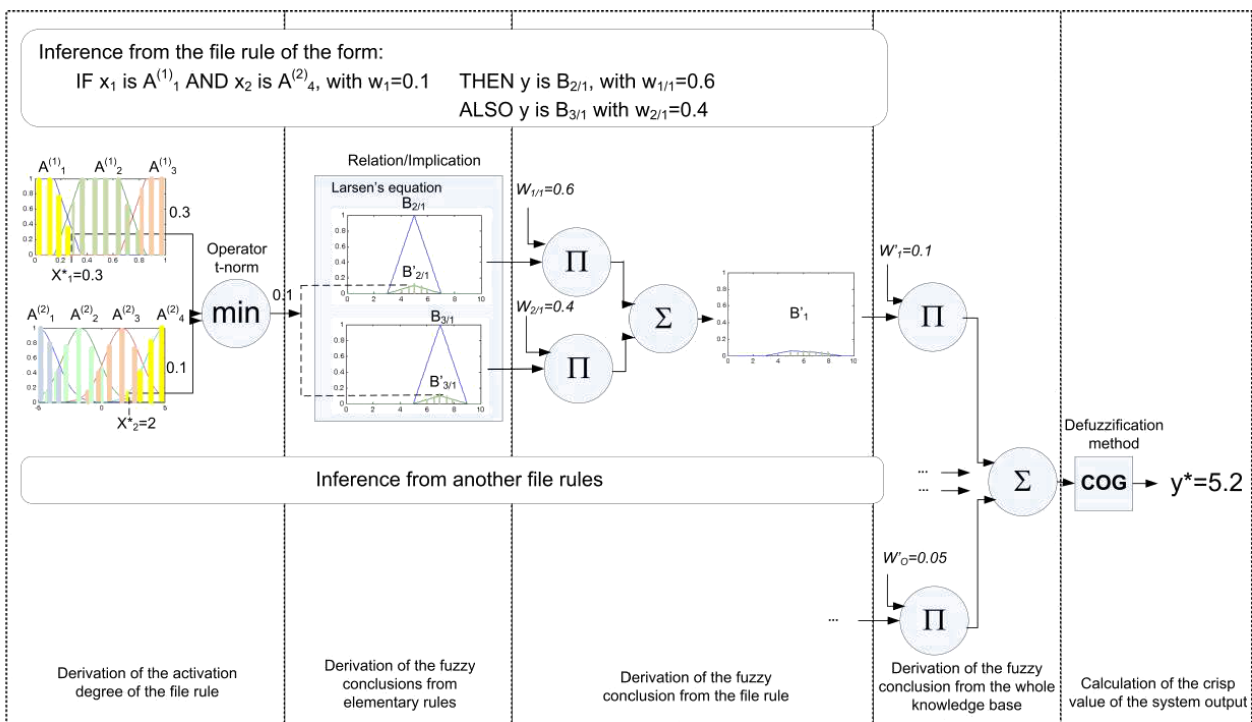


Fig. 6. Scheme of reasoning in the discussed system with probabilistic-fuzzy knowledge base (cf. [16]).

Example of applying the discussed model for forecasting the wind speed in wind power plants

Wind power is an alternative and renewable energy source. Thanks to wind turbines it is possible to transform wind energy into electric power. The wind resources all over the world are abundant. It is estimated that the amount of wind power feasible to utilize from the technological point of view is about 53 thousand TWh/year – i.e. for times more than the annual global demand for power. Surely, it is impossible to utilize the full potential of the wind, but, it is possible to use the wind within the speed range from 3–4 to 25–30 m/s [31].

Wind is a movement of air mass emerged in the course of irregular pressure distribution caused by uneven warming of the Earth by sun rays. Warm and heated air is lighter and it flows upwards making space for cooler air masses and creating air circulation. Some of the most important wind parameters used in wind power plants are wind speed and wind direction. The dependence of wind on land surface configuration leads to local whirling- changes of the wind power and direction. Studies carried out for many years prove the changeability of these parameters in time (in successive years, as well as in months of the year- according to the seasons) in 24-hour cycles and in minutes (even seconds) where wind changeability has a random nature.

Hence, the changeability of wind parameters, which is a source of energy, may cause disruption of the electric energy generation process [32]. Therefore, from the point of view of wind power energy management, it is very important to subjugate the wind by the identification of its parameters in due time- in order to control the parameters of a wind power plant. This is especially important during regular operation of the mechanism of setting the blades and the plant and in cases of emergency to predict sudden, hurricane-like wind blows and to minimise unfavourable effects of disruptions with foresight. Accurate estimation of wind energy resources is of crucial importance to the location of wind power plants, production planning, cost and feasibility assessment. The capacity of wind power plants is directly proportional to the surface of the rotor blades moved by the wind and to the third power of the wind speed, which may be inferred from the following equation [32]:

$$P = \frac{1}{2} \cdot \rho \cdot A \cdot V^3 \quad [\text{W}], \quad (15)$$

where P – wind power plant capacity [W], A – rotor blades surface [A], ρ – air density [kg/m³], V – wind speed [m/s].

The discussed inference system has been applied to predict wind speed. From 01-01-2010 to 09-01-2010, 11 000 measurements of the value of wind power were recorded at 1-minute samplings. The averages of measurements from 4 next minutes were re-researched. First 2000 measurements were learning data, the remaining ones – test data. The forecasts of wind speed $v(t)$ have been made on the grounds of the last three measurements of wind speed denoted as $v(t - 3)$, $v(t - 2)$, $v(t - 1)$. For each variable 9 fuzzy sets have been defined (with the linguistic values describing the wind as: “very light”, “light”, “mild”, “moderate”, “fairly strong”, “strong”, “very strong”, “squally”, “very squally”) assuming 45 disjoint intervals of the variables values. Exemplary values of the membership for variable $v(t - 3)$ are shown in Fig. 7. The membership grades for other variables have been analogically defined.

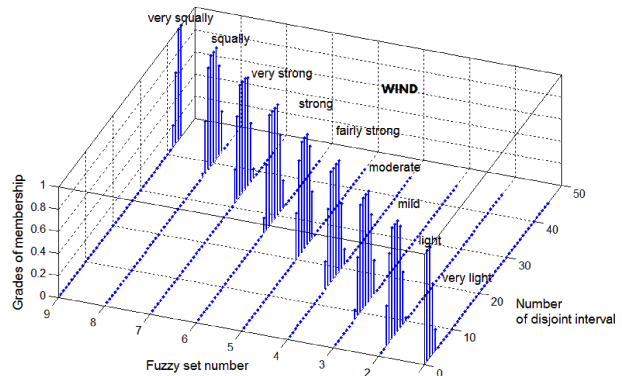


Fig. 7. An example of the definition of fuzzy sets for wind speed.

The probabilistic-fuzzy knowledge base has been created with the use of altered P-Growth algorithm (Fig. 4) for the modified Apriori algorithm [16]. At 2000 items of learning data the time of the rules generation is shorter in the case of the modified Apriori algorithm (Fig. 8).

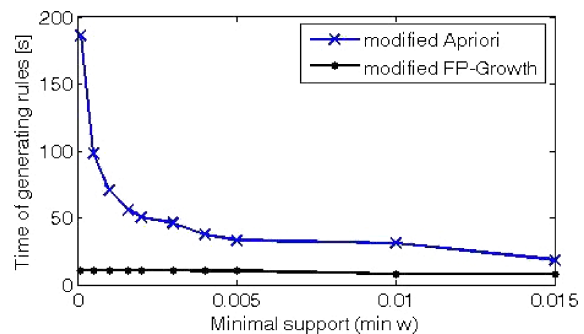


Fig. 8. Comparison of the time of generating rules using training data for different algorithms.

The accuracy of forecasting the wind parameters has been tested on the grounds of the root mean square error (RMSE) which, in this case, has been calculated as:

$$RMSE = \sqrt{\frac{\sum_{t=4}^N [v(t) - \hat{v}(t)]^2}{N - 3}}, \quad (16)$$

where N – set size, $\hat{v}(t)$ – forecasted value at time, $v(t)$ – real value at time.

The number of elementary model rules and values of RMSE for learning and testing data, depending on the minimal support parameter ($min w$) have been compiled in Figs. 9 and 10. As observed, it seems feasible to limit the number of rules down to 92, in order to simplify the model complexity and, at the same time, to observe the same accuracy of representation at the error level of 0.55 m/s. It is only after a certain value of the minimal support that the prediction error significantly increases, which shows that the model is too simple and incapable of representing relevant forecasted wind speeds v . Following such assumptions, the optimal model structure is derived at the minimal support value equal to $min w = 0.001$, then, the root mean square error for the learning data is 0.5514 m/s, whereas for the testing data it is 0.6434 m/s and the derived model consists of 92 elementary rules (47 file rules). The most important file rules are:

1: IF ($v(t-3)$ IS 'moderate') AND ($v(t-2)$ IS 'moderate') AND ($v(t-1)$ IS 'moderate') [0.1337] THEN ($v(t)$ IS 'moderate') [0.6989]

ALSO ($v(t)$ IS 'mild') [0.1665]

ALSO ($v(t)$ IS 'fairly strong') [0.1346]

2: IF ($v(t-3)$ IS 'fairly strong') AND ($v(t-2)$ IS 'fairly strong') AND ($v(t-1)$ IS 'fairly strong') [0.0973] THEN ($v(t)$ IS 'fairly strong') [0.6827]

ALSO ($v(t)$ IS 'moderate') [0.2253]

ALSO ($v(t)$ IS 'strong') [0.0920]

3: IF ($v(t-3)$ IS 'mild') AND ($v(t-2)$ IS 'mild') AND ($v(t-1)$ IS 'mild') [0.0749] THEN ($v(t)$ IS 'mild') [0.6683]

ALSO ($v(t)$ IS 'moderate') [0.2131]

ALSO ($v(t)$ IS 'light') [0.1186]

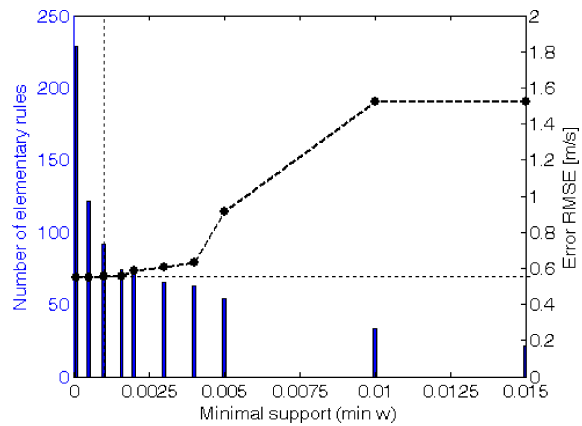


Fig. 9. The number of rules and values of RMSE errors for the training data depending on the minimal support parameter ($min w$).

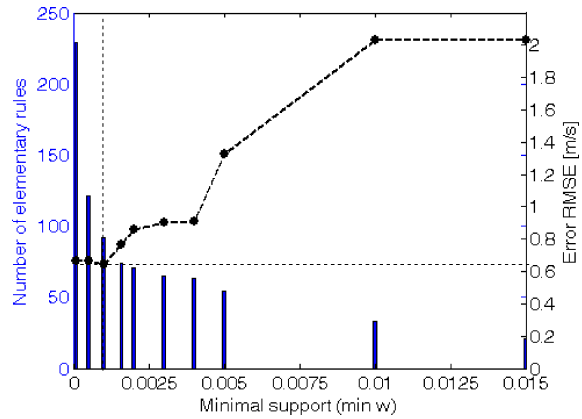


Fig. 10. The number of rules and values of RMSE errors for the testing data depending on the minimal support parameter ($min w$).

The comparison of the forecasts with real values for the learning data and the test data is shown in Figs. 11 and 12. One of the advantages of the system is an opportunity of detecting the structure of the model that enables an estimation of the characteristics of the process by the probability distribution of the occurrence of simultaneous events recorded in a human natural language (Table 1).

Wind speed is a parameter that is difficult to forecast, hence the insignificant discrepancy between the real and forecasted values. A similar level of results is rendered by forecasts made with the use of fuzzy-neuron systems (ANFIS). The discussed forecast considers a short time span, hence, it may be applied to wind power plant facilities, whereas, in the planning of electric energy production and estimating the costs of its efficiency, long time forecast should be applicable.

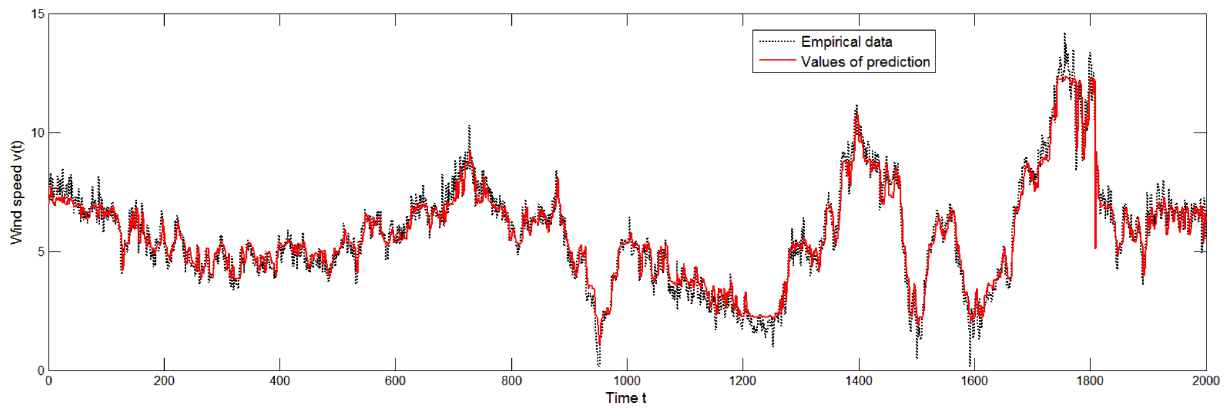


Fig. 11. Comparison of the prediction values and empirical data of wind speed for training data.

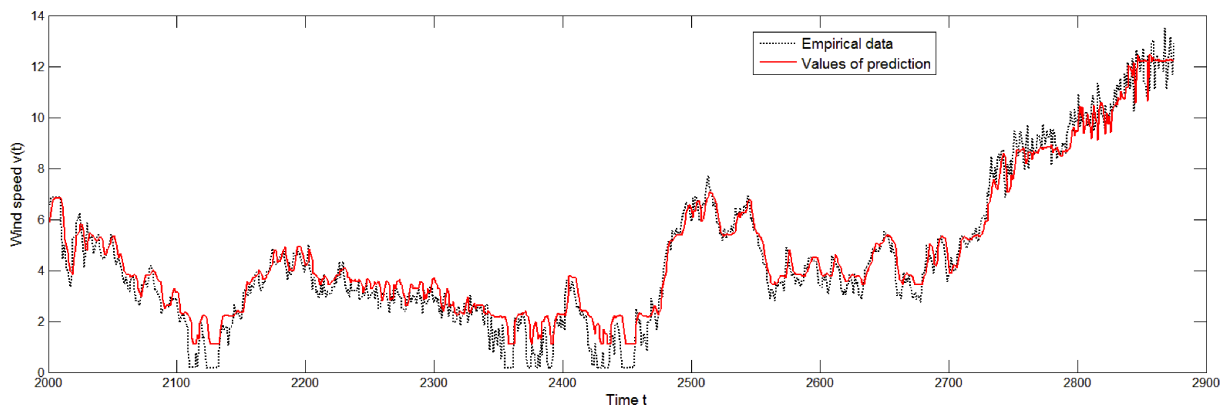


Fig. 12. Comparison of the prediction values and empirical data of wind speed for testing data.

Table 1

The joint probability distribution of fuzzy events, on the basis of wind speed prediction.

To assume, that wind $v(t - 3)$ is 'mild' and wind $v(t - 2)$ is 'mild'

$v(t - 1)$	$v(t)$								
	very light	light	mild	moderate	fairly strong	strong	very strong	squally	very squally
very light	0	0	0	0	0	0	0	0	0
light	0	0.0064	0.0077	0.0000	0	0	0	0	0
mild	0	0.0089	0.0500	0.0160	0	0	0	0	0
moderate	0	0.0000	0.0127	0.0147	0	0	0	0	0
fairly strong	0	0	0	0	0	0	0	0	0
strong	0	0	0	0	0	0	0	0	0
very strong	0	0	0	0	0	0	0	0	0
squally	0	0	0	0	0	0	0	0	0
very squally	0	0	0	0	0	0	0	0	0

Conclusions

The elaboration of the inference system with probabilistic-fuzzy knowledge database opens new possibilities in modeling processes that should take into account uncertainty levels in probabilistic and fuzzy categories.

The use of fuzzy logics with file system database makes it possible to express incomplete and uncer-

tain information in a natural language, typical of human beings. In addition, the application of the probability of events in linguistic categories enables the adjustment of the model on the grounds of numerical information derived from the data stored in the course of the operation of a given company. The created model becomes easier for interpretation by its users, which is important in the strategic decision-making process.

The discussed fuzzy association rules may be used as a method of knowledge discovery in the inference system. The search of the inference space by means of fuzzy rules association algorithms (including the modified FP-Growth algorithm) shortens the time of model creation and enables the reduction of its complexity. Thanks to the use of different parameters (fuzzy implication operators and t-norm operators) it is possible to individually adjust the model to a given analyzed process or decision problem.

References

- [1] Koźmiński A.K., *Management under the Uncertainty Conditions*, Wydawnictwo Naukowe PWN, Warszawa 2005 (in Polish).
- [2] Yager R.R., Filev D.P., *Essentials of fuzzy modeling and control*, WNT, Warszawa 1995 (in Polish).
- [3] Bellman R.E., Zadeh L.A., *Decision-making in a fuzzy environment*, Management Sciences, 17, 141–164, 1970.
- [4] Piech H., *Inference on the base of fuzzy strategies*, Wydawnictwa Politechniki Częstochowskiej, Częstochowa 2005 (in Polish).
- [5] Walaszek-Babiszewska A., *Fuzzy knowledge representation using probability measures of fuzzy events*, in: Automation and Robotics, Juan Manuel Ramos Arreguin (Ed.), In-Tech Education and Publishing, Vienna, 2008, pp. 329–342.
- [6] Walaszek-Babiszewska A., *Fuzzy modeling of the stochastic systems*, Theory, models, knowledge bases, OW Politechnika Opolska, Studia i Monografie, z. 261. Opole 2010 (in Polish).
- [7] Walaszek-Babiszewska A., *Fuzzy sets as an instrument to formalize of experts' knowledge in computer systems*, Zeszyty Naukowe Politechniki Śląskiej, Organizacja i Zarządzanie. Mat. Krajowej Konferencji Naukowej Wiedza – Informacja – Marketing, Szczyrk 2004 (in Polish).
- [8] Walaszek-Babiszewska A., Chudzicki M., *Fuzzy model for the information and decision making support system for the CFM branch company*, Applied Computer Science, Vol. 2, No. 1, 2006, Decision Support Engineering, Banaszak Z., Matuszek J. (Eds.), pp. 110–121.
- [9] Błaszczuk K., *Implementation of a probabilistic-fuzzy modelling system in Matlab*, X Międzynarodowe Warsztaty Doktoranckie, OWD, Wisła, 2008, pp. 74–78.
- [10] Błaszczuk K., *Notes on defining fuzzy sets in the created inference system with probabilistic-fuzzy knowledge base*, IV Środowiskowe Warsztaty Doktorantów PO, Zeszyty Naukowe, Politechnika Opolska, z. 63, Elektryka, Nr 335/2010, OW, Opole-Pokrzywna 2010, pp. 9–10.
- [11] Rutkowska D., Piliński M., Rutkowski L., *Neural network, genetic algorithms and fuzzy systems*, PWN, Warszawa-Łódź 1997 (in Polish).
- [12] Piegat A., *Fuzzy modeling and control*, EXIT, Warszawa 2003 (in Polish).
- [13] Łęski J., *Neural-Fuzzy Systems*, WNT, Warszawa 2008 (in Polish).
- [14] Zadeh L.A., *The concept of a linguistic variable and its application to approximate reasoning*, Part I. Information Sciences, 8, pp. 199–249, 1975; Part II, 8, pp. 301–357, 1975; Part III, 9, 1975, pp. 43–80.
- [15] Zadeh L.A., *Fuzzy sets*, Information and Control, 8, 338–353, 1965.
- [16] Rudnik K., Walaszek-Babiszewska A., *The fuzzy inference system with a model based on the association rules*, Zarządzanie Przedsiębiorstwem, Nr 2 (2010), Polskie Towarzystwo Zarządzania Produkcją (in Polish).
- [17] Kacprzyk J., *Multistage Fuzzy Control*, Wydawnictwo Naukowo-Techniczne, Warszawa 2001 (in Polish).
- [18] Nowicki R.K., *Fuzzy decision-making systems in the tasks with limited knowledge*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2009 (in Polish).
- [19] Herrera F., *Genetic fuzzy systems: status, critical considerations and future directions*, Intern. Journal of Computational Intelligence Research, 1, 1, 59–67, 2005.
- [20] Zadeh L.A., *Probability Measures of Fuzzy Events*, Journal of Mathematical Analysis and Applications, 23, 2, 421–427, 1968.
- [21] Fayyad U., Piatetsky-Shapiro G., Smyth P., *From data mining to knowledge discovery in databases*, AI Magazine, 37–54, 1996.
- [22] Frawley W., Piatetsky-Shapiro G., Matheus, C., *Knowledge Discovery in Databases: An Overview*, AI Magazine, pp. 57–70, 1992.
- [23] Walaszek-Babiszewska A., Błaszczuk K., *A modified Apriori algorithm to generate rules for inference system with probabilistic-fuzzy knowledge base*, 7th Int. Workshop on Advanced Control and Diagnosis 19–20 November 2009, Zielona Góra, CD-ROM.
- [24] Hüllermeier E., *Fuzzy methods in machine learning and data mining: status and prospects*, Fuzzy Sets and System, 156, 3, 387–406, 2005.
- [25] Agrawal R., Imielinski T., Swami A., *Mining association rules between sets of items in large databases*,

- ACM Sigmod Intern. Conf. on Management of Data, Washington D.C., May 1993, pp. 207–216.
- [26] Han J., Pei H., Yin Y., *Mining Frequent Patterns without Candidate Generation*, Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX), ACM Press, New York, NY, USA, 2000.
- [27] Chen G., Wei Q., *Fuzzy Association Rules and the Extended Mining Algorithm*, Information Sciences, 147, 201–228, 2002.
- [28] Hong T.P., Kuo C.S., Chi S.C., *Trade-off between Computation Time and Number of Rules for Fuzzy Mining from Quantitative Data*, Intern. Journal of Uncertainty, Fuzziness and Knowledge-Based System, 9, 5, 587–604, 2001.
- [29] Kuok C.M., Fu A., Wong M.H., *Mining Fuzzy Association Rules in Database*, ACM Sigmod VI Intern. Conf. on Information and Knowledge Management, 27, Issue I, 41–46, 1998.
- [30] Alcalá-Fdez J., Alcalá R., Gacto M.J., Herrera F., *Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms*, Fuzzy Sets and System, 160, 905–921, 2009.
- [31] *Aspects of wind energy*, Baza danych odnawialnych źródeł energii województwa podkarpackiego, <http://www.baza-oze.pl> (in Polish).
- [32] *Methods for assessing wind energy resources*, Baza danych odnawialnych źródeł energii województwa podkarpackiego, <http://www.baza-oze.pl> (in Polish).