Probabilistic – Fuzzy Knowledge Bases for Diagnostic Systems

ANNA WALASZEK-BABISZEWSKA
Department of Control and Computer Engineering
Opole University of Technology
Sosnkowski str., 31, 45-272 Opole
POLAND
a.walaszek-babiszewska@po.opole.pl  http://www.po.opole.pl

Abstract: - In this work we present the methods of creating the knowledge bases by using the theory of fuzzy systems as well as the probability and stochastic processes theory. We show that such knowledge-based systems can be applied in different diagnostic tasks. The structure of the reason-result fuzzy model is predefined by experts at the beginning of the task. The notion of probability of fuzzy events is used to formulate probabilities of the occurrence of the linguistic values of input and output variables. Collected data serve for computing the empirical probability distributions of linguistic variables. The calculated probabilities of fuzzy events have been included into inference procedures. Also, the knowledge base of the stochastic process with fuzzy states is presented and the forecast procedure.

Key-Words: - Knowledge-based systems, Fuzzy sets, Fuzzy logic reasoning, Probability of fuzzy events, Linguistic random variable

1 Introduction

Diagnostic expert systems belong to the most numerous group, besides advisory expert systems used for planning, forecasting or control.

The first applications of computer systems into medical activities were the main inspiration for creating the first expert systems, e.g. MYCIN [1], CASNET [2], and CADUCEUS [3]. Since the first medical expert systems have been made, the number of projects, publications and computer programs of the type of experts systems significantly grew up [4]. Last two decades, development of expert systems, or more generally knowledge based systems, supporting medical activities has been focused mainly on the reasoning procedures with uncertain data and on knowledge exploration for adjusting the knowledge base of the system. Two of such systems: GlucoNotify and FuzzyARDS have been presented in [5]. The first one, GlucoNotify is a fuzzy knowledge-based system, working as a real-time hyperglycaemia control program in an Intensive Care Unit. FuzzyARDS is an intelligent on-line monitoring program for the intensive care data of patients with Acute Respiratory Distress Syndrom (ARDS). In such complex intelligence systems, data from monitoring devices are prepared and used for adjusting medical knowledge of the system [5].

There are knowledge based systems accessible for users by World Wide Web, e.g. HEPAXPERT [6], and mentioned FuzzyARDS [5]. The knowledge based system HEPAXPERT is the system interpreting the results of serologic tests for infection with hepatitis A and B viruses.

According to Schuh, 'uncertainty is the central, critical fact in medical reasoning' [7]. Many applications of knowledge based medical systems confirm, that fuzzy set theory is very suitable for expressing uncertainty of symptoms, reasons, and for creating the inference, diagnostic procedure [7], [8], [9].

In this paper, first, we remind the mathematical foundations of fuzzy knowledge bases and inference procedures, taking into account possibility of applications in various medical diagnostic systems. Next, we propose mathematical description of the probabilistic-fuzzy knowledge base, as well as, the inference procedure. The proposed system takes into account also randomness occurring in real system.

2 Fuzzy Knowledge Based System

In expert systems the knowledge base usually has a form of a set of rules

\[
\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B,
\]

given by experts. In diagnostic systems, \( x \) is a variable representing reasons, and \( y \) represents recognized results concerning the diagnosed medical object.
The proposition \( x \ is \ A \), in compliance with Zadeh, have a semantic sense, represented by the membership function \( \mu_A \) of the fuzzy set \( A \), restricting the possible value of \( x \) [10], [11]. According to the theory of approximate reasoning, restrictions for the possible values of \( y \), for the given value of \( x \), are determined by the implication operators. The generalized modus ponens allows deducing from the rules an imprecise conclusion, while imprecise premises do not exactly match the values of antecedents. The approximate reasoning is based on a fuzzy logic and fuzzy sets and especially expresses an imprecision of statements in natural language [10], [11].

### 2.1 Fuzzy Knowledge Base for a Stabilizing System

Let us consider a simple hypothetic fuzzy knowledge-based system for stabilizing a certain measurable health parameter, \( \xi(t) \). The parameter should be hold at the desirable, constant level \( \xi_d \). Let the block for parameter monitoring be connected by feedback with a dosage block, determining the dose value \( u(t) \) of the needed substance. The parameter \( \xi(t) \) is measured at discrete moments \( t = t_k \), \( k = 1,2,\ldots, K \) and the following deviations are determined:

\[
e(t_k) = \xi_d - \xi(t_k),
\]

\[
\Delta e(t_k) = e(t_k) - e(t_{k-1}) = \xi(t_{k-1}) - \xi(t_k).
\]

The dose increments

\[
\Delta u(t_k) = u(t_k) - u(t_{k-1})
\]

are determined according to the changes of the tested health parameter

\[
\Delta u(t_k) = F[e(t_k), \Delta e(t_k)].
\]

Treating the human body reaction as a reaction of the inertial system with delay, we can describe the stabilizing process by rules, according to the formulated by Mamdani and Assilian, PI Fuzzy Logic Controller [12], [13]. Assuming, that numeric values of \( \Delta u(t_k) \), \( e(t_k) \), \( \Delta e(t_k) \) belongs to \([-1,1]\) and the linguistic values are: Negative (\( N \)), Zero (\( Z \)) and Positive (\( P \)), then we can write the first rule of the created knowledge base as follows:

\[
R_1: \text{IF } e(t_k) \text{ is } P \text{ AND } \Delta e(t_k) \text{ is } Z \text{ THEN } \Delta u(t_k) \text{ is } P.
\]

The whole hypothetic rule base can be presented in a simple form, as a table of decision (Table 1).

#### Table 1 Table of decision for stabilizing process

<table>
<thead>
<tr>
<th>( e(t_k) )</th>
<th>( \Delta e(t_k) )</th>
<th>( \Delta u )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>( Z )</td>
<td>( P )</td>
</tr>
<tr>
<td>( Z )</td>
<td>( Z )</td>
<td>( Z )</td>
</tr>
<tr>
<td>( N )</td>
<td>( P )</td>
<td>( P )</td>
</tr>
</tbody>
</table>

The actual value of dosage is calculated as

\[
u(t_k) = u(t_{k-1}) + \Delta u(t_k),
\]

where \( \Delta u(t_k) \) is the numeric result of approximate reasoning from the fuzzy rule base.

### 2.2 General Form of the Fuzzy Model for Diagnostic Systems

General form of the fuzzy knowledge base, also called a knowledge representation is usually a MISO fuzzy model (multi-input, single-output).

The model is described by the input linguistic vector \( x \), representing symptoms and the output linguistic variable \( y \), representing a diagnosis. The collection of linguistic rules is predefined by experts, in a form

\[
R_i: \text{IF } A_{1,i} \text{ AND } A_{2,i} \sem dashes \ldots \text{AND } A_{p,i} \text{ THEN } y \in B_i
\]

with words ALSO between rules, where:

\( i = \text{rule number}, \ i = 1,\ldots, I; \)

\( A_{1,i},\ldots, A_{p,i} \)-fuzzy sets representing linguistic values \( L(x_1),\ldots,L(x_p) \) of the particular input variables, defined by membership functions \( \mu_{A_{1,i}}(u_1),\ldots,\mu_{A_{p,i}}(u_p) \) in the domains \( U_1,\ldots,U_p \); \n
\( B_i \)-fuzzy sets representing the set of linguistic values \( L(y) \) of the output variable (fuzzy states of diagnosis); fuzzy sets are defined by membership functions \( \mu_{B_i}(v) \) on the output numeric space \( V \).

Conditional statement (6) describes a fuzzy relation

\[
R_i: (A_{1,i} \times \ldots \times A_{p,i}) \rightarrow B_i
\]

determined by the membership function

\[
\mu_{A_{1,i} \times \ldots \times A_{p,i} \rightarrow B_i}(u_1,\ldots,u_p,v) = \mu_{A \rightarrow B_i}(u,v)
\]
on the Cartesian product of the universes \( U_1 \times \ldots \times U_p \times V \).

Sometimes, the output variable (diagnosis) is a non-fuzzy, crisp value \( y \). In such situation, the rule base (6) is the Takagi-Sugeno type fuzzy model.

### 2.3 Fuzzy Logic Reasoning

Essentials of approximate reasoning from a fuzzy knowledge base, have been presented in many works of the object literature, e.g. in [10] – [15]. According to that, for the given new proposition
\[
x_i \text{ is } A_i^*, \ldots, x_p \text{ is } A_p^* \tag{9}
\]
representing the new linguistic values of the input variables, the fuzzy conclusion from the elementary \( i \)-th rule is determined as (\( y \text{ is } B_i^* \)). Fuzzy set \( B_i^* \) on \( V \) is derived according to the generalized modus ponens [10], [11]
\[
\mu_{B_i^*}(v) = \sup_{u \in U} T(\mu_{A_i^*}(u), \mu_{A_i \rightarrow B_i}(u, v)) \tag{10}
\]
where \( T \) denotes t-norm fuzzy operator. In (10), the fuzzy relation (7), (8), representing the conditional statement may be interpreted by the conjunctive or implicator operators, as follows [15]:
\[
\mu_{A \rightarrow B}(u, v) = T(\mu_A(u), \mu_B(v)) \tag{11}
\]
\[
\mu_{A \rightarrow B}(u, v) = I(\mu_A(u), \mu_B(v)) \tag{12}
\]
Aggregation of fuzzy conclusions (10) from particular rules is realized by s-norm, in particular as the union operation of fuzzy sets \( B_i^* \):
\[
\mu_{B^*}(v) = \sup(\mu_{B_1^*}(v), \ldots, \mu_{B_J^*}(v)), \forall v \in V. \tag{13}
\]

Very popular method of reasoning is the Simplified Method of Fuzzy Reasoning (SMFR) [13]. The algebraic multiplication as a t-norm (Larsen’s rule) in (11) and algebraic sum as an aggregation operator in (13), give the simple analytical form of inferred fuzzy output. The numeric value of output \( v^* \) has a form of weighted sum
\[
v^* = \frac{\sum_{j=1}^{J} \tau_j v_j^*}{\sum_{j=1}^{J} \tau_j}, \tag{14}
\]
where \( v_j^* \) are numeric values of centroids of fuzzy sets \( B_j \) of the antecedent variable. Changing on the input of the systems needs only calculating the levels of switching \( \tau_j \) for antecedents of particular rules.

## 3 Probabilistic - Fuzzy Knowledge Base

Some of the remained medical experts systems have an option of updating their fuzzy data bases on the grounds of the collected data. The approach presented in this paper is a response to such need, facilitating the designation of the probability of simple facts as well as entire logical connections formulated in terms of fuzzy categories.

### 3.1 Fuzzy Model with Weights – Diagnostic Knowledge Base

In this paragraph a linguistic fuzzy model of a MISO system (6) is considered as the set of weighted rules. The exemplary \( i \)-th file rule has the form [16], [17]:
\[
R_i: w_i (IF x_1 \text{ is } A_{1i} AND x_2 \text{ is } A_{2i}...AND x_p \text{ is } A_{pi}) \THEN y \text{ is } B_i (w_{ij})
\]
\[
ALSO y \text{ is } B_j (w_{jj})
\]
\[
ALSO y \text{ is } B_j (w_{jj})
\]
\[
ALSO y \text{ is } B_j (w_{jj})
\]
\[
\]}

The structure of the model is created with cooperation of experts of the diagnostic process. Input variables \( x_1, \ldots, x_p \) represent assumed symptoms and output variable \( y \) represents recognized diagnosis. The variable can be determined in the discrete or continuous domains, accordingly to that, fuzzy sets \( A_{1i}, \ldots, A_{pi}, B_j \) are defined. The rule weights \( w_i \) and \( w_{jj} \) \( i=1,\ldots,I, \) \( j=1,\ldots,J \) represent probabilities of fuzzy events occurring in the antecedents and consequence of the rule, as follows:
\[
w_i = P(A_i), \ w_{jj} = P(B_j / A_i) \tag{16}
\]
where \( A_i \) is a joint fuzzy event occurring in the antecedent. According to Zadeh’s definition, and according to [16], probability of that event is calculated as follows
\[
P(A_i) = \sum_{u \in U} p(u) T(\mu_{A_{1i}}(u_1), \ldots, \mu_{A_{pi}}(u_p)) \tag{17}
\]
where \( p(u) \) is the empirical probability distribution calculated by using the set of data
\[
\{x_{1,i}, \ldots, x_{p,i}, y_i\}_{i=1,...,N}, \tag{18}
\]
as the t-norm in (18), multiplication of membership functions is usually applied.

Weights of file rules, \( w_j \), \( i=1,\ldots, I \) as a probability distribution of input linguistic variables, should fulfill the relationship

\[
\sum_{i=1}^I w_i = \sum_{i=1}^I P(A_i) = 1. \quad (19)
\]

Elementary rule weights \( w_{ji} , \quad i=\text{const}, \quad j=1,\ldots, J \) state the conditional probabilities of the events \( (Y \text{ is } B_j) \), given \( (X \text{ is } A_i) \). They can be calculated from Bayesian formula

\[
w_{ji} = P(B_j / A_i) = \frac{P(A_i \times B_j)}{P(A_i)} , \quad (20)
\]

where

\[
P(A_i \times B_j) = \sum_{(u,v) \in U \times V} p(u,v) T(\mu_{A_i}(u), \mu_{B_j}(v)) \quad (21)
\]

is the joint probability distribution of linguistic input-output vector

\[
P(X,Y) = \{ P(A_i \times B_j) \}, \quad i=1,\ldots, I, \quad j=1,\ldots, J . \quad (22)
\]

Let us note, that formula (21) determines a probability of the joint fuzzy event occurring in the elementary rule. The weight \( P(A_i \times B_j) = w_{ji} \) tells us how often, the combination of symptoms \( A_{i,j}, \ldots, A_{p,j} \) and the result \( B_j \) occurred together. And the weights \( w_{ji} \) determine the conditional probabilities of the particular states \( B_{ij} \), \( i=1,\ldots, I, \quad j=1,\ldots, J \) of the diagnostic variable \( y \), when the combination \( A_{i,j}, \ldots, A_{p,j} \) of symptoms occurred in the \( i \)-th rule.

### 3.2 Inference and Aggregation Procedure

When the proposition \( (x \text{ is } A_i) \) is given, then from the \( ij \)-th elementary rule of the model (15), the conclusion \( (y \text{ is } B_{ji}) \) can be computed. The membership function \( \mu_{B_{ji}}(v) \) of the inferred fuzzy output is given by formulas (10) – (12).

Aggregation operation on fuzzy conclusions \( B_{ji}^* \), \( j=1,\ldots, J \), derived from all elementary rules of \( i \)-th file rule, is executed by the following weighted sum

\[
\mu_{B_i^*}(v) = \sum_{j=1}^J w_{ji} \mu_{B_{ji}}(v) . \quad (23)
\]

The set \( B_i^* \) determines the conditional mean fuzzy value of the diagnostic variable \( y \). The condition is formulated by the antecedent of the \( i \)-th rule.

The conclusion \( B^* \) from the whole knowledge base is the following weighted sum

\[
B^* = \sum_{i=1}^I \sum_{j=1}^J w_{ji} B_{ji}^* \quad (24)
\]

Formula (24) determines the mean fuzzy value of the diagnostic variable \( y \). The numeric value can be calculated by using one of the methods of defuzzyfication, according to the type of the domain \( V \).

According to mentioned the Simplified Method of Fuzzy Reasoning, the model of knowledge can have a form

\[
R_i: w_i (IF \; x_i \text{ is } A_{i,j} \; AND \; x_j \text{ is } A_{x,j} \ldots \text{AND} \; x_k \text{ is } A_{x,j} \;

\text{THEN} \; y_i \; \left( w_{1ji} \right) \]

\[
\text{ALSO} \; y_j \; \left( w_{2ji} \right) \]

\[
\text{ALSO} \; y_k \; \left( w_{3ji} \right) \]

(25)

where the set \( \{ y_1, \ldots, y_j, \ldots, y_J \} \) represents particular discrete results of recognized diagnosis and weights \( w_{ji} \) state for the conditional probabilities of these occurrence.

### 4 Fuzzy Knowledge Base with Stochastic Character of Variables

In many diagnostic situations variables representing symptoms are time–varying stochastic variables.

Let us consider the task of a forecast of one variable, the stochastic process \( X(t) \). The process is observed at moments

\[
t_1, t_2, \ldots, t_n \in T; \quad t_n > t_{n-1} > \ldots > t_1 .
\]

The data of observations of the process values \( (X(t_1), \ldots, X(t_n)) \) serve for calculating the \( nD \) empirical probability distributions of the process.

Also, the linguistic random variables have been determined with their collection of linguistic values, represented by fuzzy sets \( A_{k,i_k} \), where \( k=1,\ldots, n; \quad i_k=1,\ldots, I \) in the domain \( \chi \) of process values.

The proposed fuzzy model, as the knowledge representation of the stochastic fuzzy-valued process, has the form of the collection \( \{ R(j) \} \) of file rules [17]:

---

Recent Advances in Applied & Biomedical Informatics and Computational Engineering in Systems Applications

5 Conclusion
A formalized description of knowledge bases and inference procedures has been introduced with the aim of the applications in diagnostic systems. It is assumed that symptoms and the recognized diagnosis are represented by linguistic variables, taking uncertain, linguistic values represented by fuzzy sets. Knowledge of experts, concerning the tested processes, can be described in the form of knowledge representation, as the fuzzy rule-based model. Also, it would be important in diagnostic systems, the probability of the fuzzy events occurring in the fuzzy models has been introduced. Probability distributions of the variables of the reason-result fuzzy models determine the probabilistic or stochastic structure of the fuzzy models, as well as, allow calculating the probability of the inferred diagnosis. In future works, the generally formulated probabilistic-fuzzy models will be tested in real diagnostic situations.

References:


