diagnostic rules, fuzzy sets,
the Dempster-Shafer theory

Ewa STRASZECKA[1]

THE BASIC PROBABILITY ASSIGNMENT AS A MEASURE OF DIAGNOSTIC RULES SIGNIFICANCE

Diagnostic rules are usually IF-THEN rules, but they should satisfy specific requirements of a diagnosis. Thus, not always the classical methods of rules determination are applicable. In the present paper it is suggested to find out the set of rules by an elimination of superfluous rules from the maximal rule set or adding rules that improve inference to the minimal set of rules. It is shown that the basic probability assignment determined in the Dempster-Shafer theory of evidence can be used as a measure indicating symptoms that are the most significant for a diagnosis and should create rules. A set of IF-THEN rules with fuzzy premises and crisp conclusions can be built in this way. The proposed method is illustrated by determining rules allowing for diagnostic inference for a database of thyroid gland diseases.

1. DIAGNOSTIC RULES

1.1. GENERAL FORM OF THE RULE AND IMPRECISION OF ITS PREMISE

Diagnostic rules usually have an abductive form: IF symptoms THEN disease [3]. They describe a heuristic mapping between symptoms and diagnostic hypotheses. Yet, the rules are never certain and their abductive form involves difficulties in application of a certainty measure as a factor of the rule significance. The classical probability is difficult to handle as it requires values of conditional probabilities of all combinations for symptoms of a given disease. Only under this condition the probability of a disease given symptoms can be calculated from Bayes formula [4]. Obviously, the calculation is easy for one rule, but quite difficult for their collection that makes a knowledge base of a diagnosis support system.

A fuzzy formulation of a premise, for instance 'high cholesterol level' or 'low heart rate', can indicate the use of a fuzzy rule for the diagnostic rule representation. Yet, the rule is not entirely fuzzy, since its conclusion is crisp (disease). Fuzzyfication of the conclusion is awkward and requires an introduction of an additional variable that resembles a disease risk. Still, the risk domain must be assumed without any clear indications concerning its scale or units.

These reasons made the author of this paper to search for another solution for a diagnostic rule representation and for a diagnostic inference. It was found in the framework of the Dempster-Shafer theory and the fuzzy set theory. The rule representation is proposed as [6]:

$$\text{IF } X_1 \text{ is } X_1^l, \text{ and, } \ldots, \text{ and } X_n \text{ is } X_n^l \text{ THEN diagnosis is } D_l, \tag{1}$$

where X_i is a linguistic variable (e.g. a laboratory test), X_i^l – an i-th linguistic value (e.g. 'high') used in a symptom description in the D_l diagnosis. The X_j^l can be represented by a membership function.

[1]Institute of Electronics, Silesian University of Technology, 16 Akademicka St., 44-100 Gliwice, ewa.straszecka@polsl.pl.

Since the membership function can be particularly a characteristic function or a singleton, the formula (1) represents a variety of diagnostic rules.
Thus, imprecision of a symptom in the diagnostic rule is expressed by a fuzzy set, but certainty of the rule remains undetermined. The latter can be represented by the value of the basic probability defined in the Dempster-Shafer theory of evidence.

1.2. UNCERTAINTY OF THE RULE

The premise of the rule can be considered as a focal element in the Dempster-Shafer theory of evidence (DST). Focal elements in the DST are predicates to which a probability value is assigned. Thus, they can represent symptoms. The DST can be extended for fuzzy focal elements [5]. In the DST probability is defined in a manner that differs from the classical definition. It is the basic probability assignment (BPA) denoted by m [2]:

$$m(f) = 0, \sum_{s \in S} m(s) = 1. \tag{2}$$

where f is the false predicate and S is the set of focal elements, in our case symptoms s. Thus, s resembles the predicate 'X_i is X_i^l' or 'X_i is X_i^l and ... and X_n is X_n^l'. In the latter case the focal element will be called complex.
In the current work a method that join the Dempster-Shafer and fuzzy set theories with some aspects of rule based system is proposed. Therefore, it is necessary to explain the accurate meaning of concepts that are used. The rule premise consists of conditions. Linguistic variables that are used in the conditions are represented by fuzzy sets. Each rule condition is a symptom. If rules that concern one diagnosis are gathered in a separate collection, we may consider only their premises during inference. Only when reasoning is performed, the conclusion is assigned with some certainty. The set of premises is the set of focal elements. The basic probability values are assigned to focal elements. After reasoning for all diagnoses, their certainty is compared and the final diagnosis is chosen. Thus, the diagnoses are often called diagnostic hypotheses in the present work.
In (2) the dependence of focal elements is not considered, hence calculations of the BPA values are easier in comparison to classical probability. Moreover, the same symptoms may be included in different focal elements: for instance a laboratory test is mentioned in the premise of one rule, but also in conjunction with another test in another rule. The BPA value stands for the certainty of the rule. It can be calculated as a normalized frequency of occurrence of rule premises for the chosen diagnostic hypothesis. However, rule premises are fuzzy, so a presence of a symptom, i.e. the focal element, must be considered in imprecise categories. Thus, a threshold η_{BPA} is assumed and focal elements which memberships are greater than the threshold are presumed to be present. The BPA for fuzzy focal elements are calculated as [6]:

$$m_D(f) = 0, \sum_{\substack{s_i \in S, i=1,\ldots,n \\ \eta_i > \eta_{BPA}}} m_D(s_i) = 1. \tag{3}$$

where n is the number of focal elements for the D diagnostic hypothesis and $s_i \cap s_j \neq \emptyset$ is possible. The η_i level of symptom's precision is determined as the conjunction of membership functions of the premise condition ($\mu_i(x)$) and evidence, i.e. patient's observation. If the observation is a measurement x_i^*, the latter membership is the $\delta_{x_i}^*$ singleton and the level is calculated as [7]:

$$\eta_i = \sup_{x \in \mathbf{X}} \left[\mu_i(x) \wedge \delta_{x_i}^* \right] = \mu_i \left(x_i^* \right). \tag{4}$$

The (4) formula is here given for a single focal element, for complex focal elements it is the minimal value of levels of included symptoms [7].

2. CERTAINTY OF DIAGNOSIS

The level of symptom's precision is used not only for BPA, but also for the belief (Bel) and plausibility (Pl) calculation [1]. These measures are in the classical Dempster-Shafer theory the measures of certainty of an evidence that is characterized by the $[Bel(D), Pl(D)]$ interval. The interval becomes a single value if focal elements are separate:$\forall_{i,j} s_i \cap s_j = \emptyset$. In such a case $Bel(D) = Pl(D) = P(D)$, where $P(D)$ is the (classical) probability of the diagnosis. Still, this situation does not concern our problem, hence the measures for fuzzy focal elements must be defined as [6]:

$$Bel(D, \eta_T) = \sum_{\substack{s_i \in S \\ \eta_i > \eta_T}} m(s_i), \tag{5}$$

$$Pl(D, \eta_T) = \sum_{\substack{s_i \in S \\ \theta_i > \eta_T}} m(s_i), \tag{6}$$

where D stands for the diagnosis and η_T is the precision threshold of reasoning. In (6) the θ_i precision level is used which is calculated as maximum value of precision of symptoms for the complex focal element [7]. The η_T threshold does not need to depend on η_{BPA}, but it is reasonable and confirmed by the experiments [6] that $\eta_T \geq \eta_{BPA}$.
The medical diagnosis is always very cautious, thus the Bel measure seems to be more appropriate for its estimation. In the present method this measure is used for differential diagnosis. Values of the Bel measure are calculated for all diagnostic hypotheses and the one with the greatest value wins. If the maximal value is not unique, the final conclusion cannot be driven.

3. DATABASE

The method presented in the previous two sections is tested for the database [8]. The database concerns thyroid diseases and includes data for 3 diagnoses: euthyroidism (the state of normal level of thyroid hormones), hyperthyroidism and hypothyroidism. In the present paper the diagnoses are denoted as D_1, D_2 and D_3, respectively. The diagnosis is the first element in the data sample, thus parameters that follow it are here denoted as variables v_2–v_6. The number of cases for the three diagnoses (in the corresponding order) are 150, 35 and 30. During experiments the whole database is the train and the test set as in the current work a correctness of approach rather than an efficiency of solution is tested.

4. DETERMINATION OF MEMBERSHIP FUNCTIONS AND BPA VALUES

The rule (1) require defining membership functions. A general approach to this problem is provided in [5] and [7]. Now, only the functions that are constructed for the present application are described. Examples of the functions are presented in Fig. 1. They are trapezoid functions with two crucial points determined as crossover points of distributions that estimate frequency of occurrence of a symptom with different diagnoses. The other two points are quartiles of the samples [7]. If quartiles do not correspond to the supposed crossover points, the former are admitted more reliable and the slope of the trapezoid is assumed as 'steep' [5]. This is the case observed in the left diagram of Fig. 1 for the right slope of the leftmost function. During calculations the functions are modified toward triangular and toward characteristic functions to find the best shape for an application. In Fig. 1 three functions correspond to the three diagnoses. The membership function for D_1 is depicted by solid line, for D_2 by dashed line and for D_3 by dotted line. It should be pointed out that the final membership functions are not symmetrical and their values do not need to sum up to 1, i.e. $\exists_x \sum_{i=1,2,3} \mu_{Di}(x) \neq 1$. The shapes are not similar to that used by experts, but experiments confirm [5] that such data-driven membership functions are more efficient in diagnosis support than functions for which the sum is 1 and which are used in the fuzzy control.

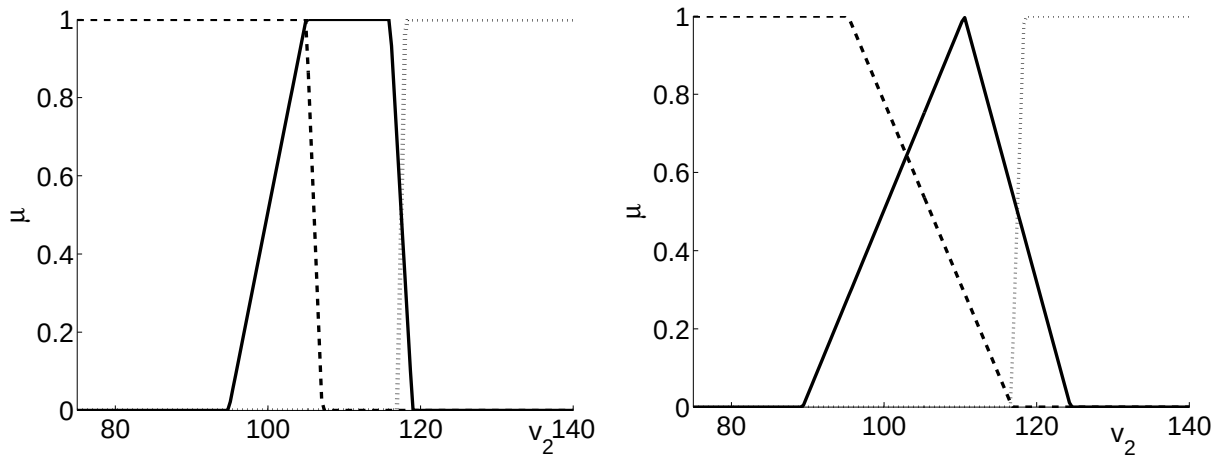

Fig. 1. Membership functions for v_2 variable of the database [8]: the shape at start (left) and the best shape after a modification (right).

The BPA can be calculated as the normalized frequency of occurrence using the functions, providing that for complex elements minimum precision level must be greater than the η_{BPA} in the (3) formula. In this way a separate BPA is found for each diagnosis. The focal elements of different diagnoses are equivalent in this sense that they include the same X_i medical parameters (1), but individual significance of these elements is expressed by different BPA values. Let us give as instances three rules - one for each diagnosis, concerning three focal elements - $s_i^1 \in S_{D1}$, $s_i^2 \in S_{D2}$ and $s_i^3 \in S_{D3}$:

$$\begin{aligned}
&\texttt{IF } V_3 \texttt{ is } \mu_3^1\texttt{, and, } V_6 \texttt{ is } \mu_6^1 \texttt{ THEN diagnosis is } D_1\texttt{, significance: } m_{D1}(s_i^1);\\
&\texttt{IF } V_3 \texttt{ is } \mu_3^2\texttt{, and, } V_6 \texttt{ is } \mu_6^2 \texttt{ THEN diagnosis is } D_2\texttt{, significance: } m_{D2}(s_i^2);\\
&\texttt{IF } V_3 \texttt{ is } \mu_3^3\texttt{, and, } V_6 \texttt{ is } \mu_6^3 \texttt{ THEN diagnosis is } D_3\texttt{, significance: } m_{D3}(s_i^3).
\end{aligned} \tag{7}$$

The rules can be rewritten in a more compact form as:

$$\begin{aligned}
&\texttt{IF } s_i^1 \texttt{ THEN diagnosis is } D_1\texttt{, significance: } m_{D1}(s_i^1);\ s_i^1 \equiv \{V_3^1, V_6^1\};\\
&\texttt{IF } s_i^2 \texttt{ THEN diagnosis is } D_2\texttt{, significance: } m_{D2}(s_i^1);\ s_i^2 \equiv \{V_3^2, V_6^2\};\\
&\texttt{IF } s_i^3 \texttt{ THEN diagnosis is } D_3\texttt{, significance: } m_{D3}(s_i^1);\ s_i^3 \equiv \{V_3^3, V_6^3\}.
\end{aligned} \tag{8}$$

The notation in (7) and (8) is considerably simplified, but does not violate the general form of (1). The membership functions μ_3^j, μ_6^j as well as values of probability assignments $m_{Dj}(s_i)$, $j = 1, 2, 3$, differ. The only things they have in common are domains of the membership functions and crossover points (if they match quartiles).

The BPA, and Bel values depend on the shape of membership functions as well as on η_{BPA} and η_T. In the experiments performed in the present work shapes of membership functions are changed with a modification coefficient adjusted in the $[-0.9, 0.9]$ interval with 0.1 step. Negative coefficient means approaching characteristic function which would be reached with the coefficient equal to -1. The modification with positive coefficient move a slope toward the triangular function. The η_{BPA} and η_T thresholds are both changed in the $[0.05, 1]$ interval with 0.05 step. Yet, the present experiment confirms the previous observations that the best solutions are found for $\eta_{BPA} \leq \eta_T$. Each of algorithms and every set of rules are tested with the mentioned change of membership functions and thresholds to find the best solution.

5. DETERMINATION OF THE RULE SET

In theory we are free to define any focal element, but in practice it is not an easy task. Since the DST has few limitations, the set of focal elements decides of an application effectiveness, so much attention must be payed to rule creating. The first idea is to make the complete set of all combinations

of conditions, which for 5 variables makes 31 rules. Still, for many variables this is hardly possible and some of the rules certainly will be superfluous. Thus, the second idea is to join variables for which data show correlation [5]. This approach is partly successful, but fails in case of categorical variables [6]. The method which is now proposed is based on the assumption that the basic probability values may indicate focal elements, i.e. symptoms, which are the most promising in a diagnosis.
An experiment is performed in the following way. Firstly, the complete set of rules is made and BPAs for 3 diagnoses are calculated. Next all values of the three BPAS are sorted and the stage of rule elimination begins. Each time the rule with the smallest BPA value is eliminated. It is deleted from each of the 3 sets regardless in which BPA the minimum is found. Afterwards, the new set of focal elements is used while BPA determining and the knowledge is tested for the database. A classification error is calculated. In the present experiment its value is just an information about the robustness of the method, but in future a considerable increase of the error may indicate the end of the elimination. Now, the elimination process is continued until only 6 rules are left. Premises of five rules concern single variables v_2–v_6 and one rule has a complex focal element in the premise. Such an end of calculation is reasonable for the method, because if single variables in premises are sufficient, there is no need to use the Dempster-Shafer theory. This part of the experiment should check the possibility to eliminate superfluous rules from a knowledge base.
The aim of the second part of the experiment is to examine if rules can be constructed from 'promising' focal elements, i.e. focal elements of the greatest values. At the beginning frequencies of occurrence are found for single variables. These with the greatest frequencies are used to make focal elements. All possible combinations of symptoms concerning the variables are made. The BPAs for the sets of focal elements are calculated and next rules are tested. Then, the elimination algorithm is used to decrease the number of rules. This algorithm can be of use when many medical parameters have to be considered and it would be difficult to build the complete set of focal elements.
It is possible and probably more efficient to use the elimination algorithm in such a way that it deletes the rule that is the least significant in the last calculated BPA. Yet, this way of elimination make calculations longer. It is easier to test once many rules than to repeat calculations many times for decreasing number of rules. Thus, the latter manner of elimination can be used only if the set of rules is relatively small.

6. RESULTS

Methods described in sections 1 and 2 with membership functions found in a manner given in section 4 are used to diagnose cases of the database from section 3. At first the elimination algorithm from section 5 is used. Fig. 2 shows BPAs calculated for all possible, i.e. 31 rules and membership functions that result in the lowest classification error.

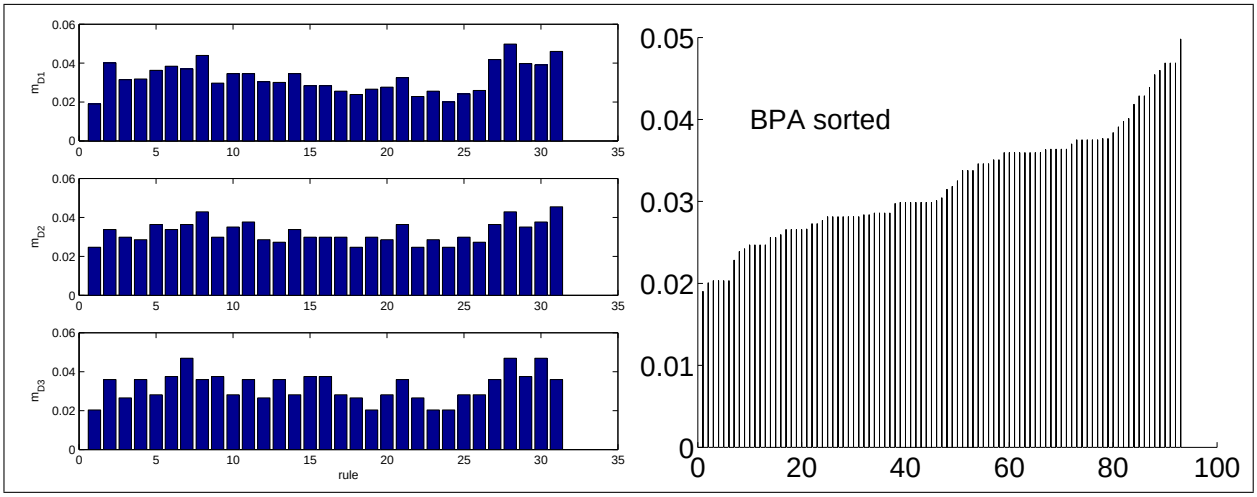

Fig. 2. BPA values: for the three diagnoses (left) and altogether, sorted (right).

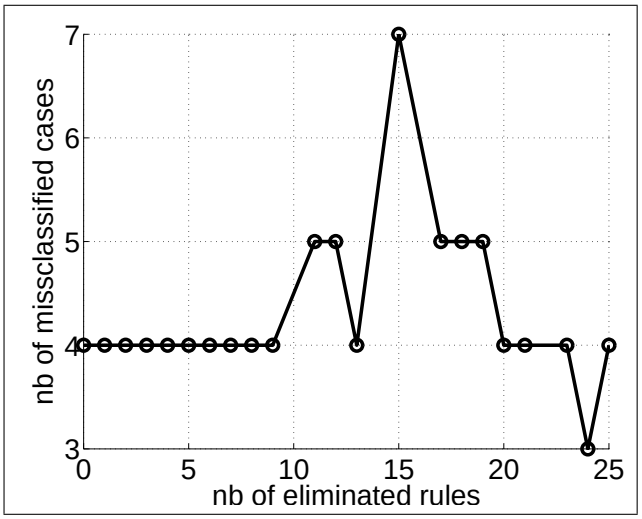

Fig. 3. The number of wrongly diagnosed cases according to the number of eliminated rules.

The BPA values for the three diagnoses are on the left, while their values, taken together and sorted are on the right. It is observable in the right diagram that the values are changed gradually, thus rules cannot be easily divided for the most and least important. Therefore, results of the elimination must be estimated by means of the classification error. The number of wrongly diagnosed cases according to the number of eliminated rules are presented in Fig. 3. First of all, it must be noticed that a determination of rules by means of the proposed method is more efficient than the method used in [6], based on correlation of variables, which resulted in 9 rules and 6 wrongly diagnosed cases. In the present method, the number of incorrectly classified cases at the beginning remains the same which means that several rules are superfluous. Next the number increases, but afterwards it decreases to the same and subsequently even lower number than at the beginning. Thus, for 7 rules only 3 cases are wrongly diagnosed. An increase and then a reduction of the error can be intuitively explained as a consequence of removing rules responsible for rare diagnoses followed by an effect of removing rules for which the diagnosis is too obvious. Anyway, it must be admitted that this algorithm is better than the approach based on correlation coefficients - the error is lower and the number of rules is smaller. The both approaches are better than reference methods [5].

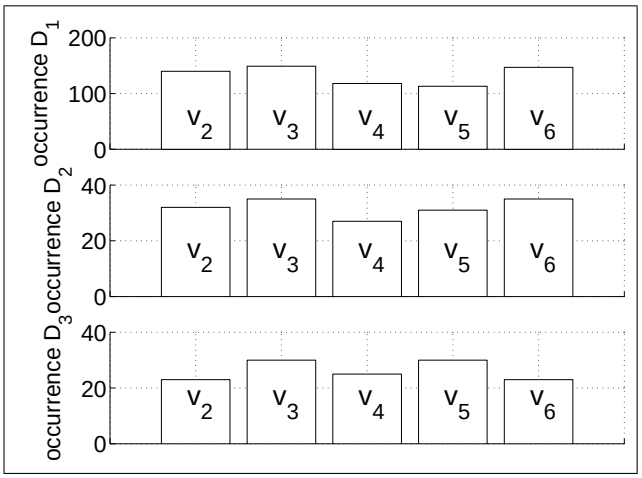

Fig. 4. The number of occurrence for single-variable focal elements.

The second method of determining the knowledge base is illustrated in Fig. 4 and Fig. 5. Firstly, the occurrence of focal elements, each concerning a single variable is found. In Fig. 4 it is shown that for the D_1 and D_2 diagnoses focal elements with variables v_3 and v_6 occur must often, while for D_2 - elements with variables v_3 and v_5 are the most frequent. Thus, focal elements with variables v_3, v_5

and v_6 are created. As the result the set S of focal elements $S^l = s_i^l$, $i = 1, ..., 9$ is obtained for each of the D_l $l = 1, 2, 3$ diagnosis, in which the following focal elements are included: $s_1^l = \{V_3^l, V_5^l\}$, $s_2^l = \{V_3^l, V_6^l\}$, $s_3^l = \{V_5^l, V_6^l\}$, $s_4^l = \{V_3^l, V_5^l, V_6^l\}$, $s_5^l = \{V_2^l\}$, $s_6^l = \{V_3^l\}$, $s_7^l = \{V_4^l\}$, $s_8^l = \{V_5^l\}$, $s_9^l = \{V_6^l\}$. The BPAs for these focal elements are in the upper diagram in Fig. 5.

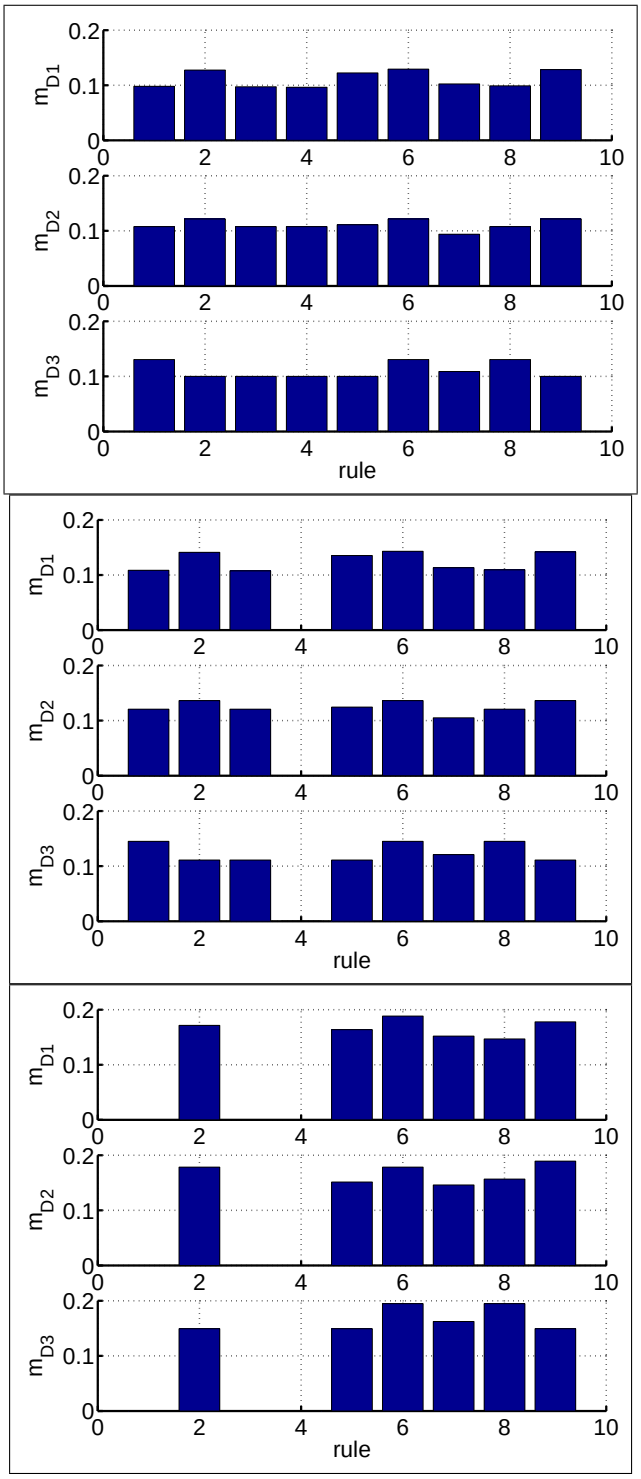

Fig. 5. BPAs for 9, 8 and 6 rules (upper, middle and lower diagram, respectively) .

The smallest value is $m_{D1}(s_4) \approx 0.096$, so s_4^l focal elements are eliminated. The next BPAs, for 8 focal elements: s_j^l, $j = 1, 2, 3, 5, 6, 7, 8, 9$ are in the middle diagram of Fig. 5. Their smallest value (excluding single-variable focal elements) is $m_{D1}(s_3) \approx 0.11$, so now s_3^l are eliminated. Next BPAs are determined for s_j^l, $j = 1, 2, 5, 6, 7, 8, 9$. The focal element with the smallest basic probability is $s_1^l \approx 0.12$, so at last the set of focal elements is: $s_2^l = \{V_3^l, V_6^l\}$, $s_5^l = \{V_2^l\}$, $s_6^l = \{V_3^l\}$, $s_7^l = \{V_4^l\}$,

$s_8^l = \{V_5^l\}$, $s_9^l = \{V_6^l\}$. BPAs for these focal elements (shown in lower diagram of Fig. 5) result in correct diagnoses except for only 2 cases, which is the best result obtained during the experiments.
The BPA diagrams in Fig. 5 do not show great changes in BPA values when rules are eliminated. Small BPA changes together with the decreasing error indicate monotonous diagnosis which is an important advantage of diagnosis support systems.

7. CONCLUSIONS

The paper suggests using the Dempster-Shafer theory extended for focal elements to diagnostic rule representation and proposes the method of building sets of necessary rules if a database of numerous medical parameters is given. In this way a knowledge base can be derived from the database. Two manners of choice of the most appropriate rules are presented, both based on an analysis of the basic probability assignment. This assignment must be determined for any application of the suggested method, so much time and effort can be saved on a separate algorithm of rules search. It is also possible that the proposed way of rules determination could be applicable in other problems than the diagnosis support. The experiments show that the elimination of rules by means of the proposed method does not disturb monotonicity of a diagnosis which is an important advantage of diagnosis support systems. The significance of remaining rules is also gradually changed, which is intuitively right.
The proposed method can be used for an automatic creation of diagnostic rules and an estimation of their significance on the basis of population data. This facility would be very important for diagnosis support systems because they are difficult to adapt to different hospital information systems and diverse medical procedures employed in various countries. An automatic adjustment of diagnostic rules can make a knowledge transfer easier.
Basic probability assignments can be combined in the framework of the Dempster-Shafer theory. Moreover, focal elements of the combined assignments do not need to be the same. Thus, a knowledge base from an expert can be combined with a data-driven rules which could improve a diagnosis support. The tests were performed for only one database and obviously the method should be further examined, but primary results are interesting and might be of use for other researchers, thus the author decided to present them in the paper.

BIBLIOGRAPHY

[1] DEMPSTER A. P., A generalization of Bayesian inference, J. Royal Stat. Soc., J. Wiley, New Jersey, USA, 1968, Vol. 30, No. 2, pp. 205-247.
[2] KACPRZYK J., FEDRIZZI M. (eds.), Advances in Dempster-Shafer Theory of Evidence, J. Wiley, New York, USA, 1994.
[3] LUCAS P. J. F., Model-based diagnosis in medicine, Artificial Intelligence in Medicine, Elsevier, 1997, Vol. 10, pp. 201-208.
[4] SCHORK M. A., REMINGTON R. D., Statistics with Applications to the Biological and Health Sciences, Prentice Hall Inc., Upper Saddle River, New Jersey, USA, 2000.
[5] STRASZECKA E., Combining uncertainty and imprecision in models of medical diagnosis, Information Sciences, Elsevier, 2006, Vol. 176, pp. 3026-3059.
[6] STRASZECKA E., Measures of uncertainty and imprecision in medical diagnosis support, Wyd. Pol. Sl., Gliwice, 2010.
[7] STRASZECKA E., Uncertainty and imprecision in medical diagnosis support, J. Medical Informatics & Technologies, Univ. of Silesia, Katowice, Poland, 2012, Vol. 19, pp. 11-22.
[8] database online: ftp.ics.uci.edu/pub/machine-learning-databases/thyroid-disease, files new-thyr.*, 5.06.2013.