diagnosis support, fuzzy sets,
Dempster-Shafer theory

Ewa STRASZECKA *

REPRESENTATION OF SYMPTOMS AND EVIDENCE IN DIAGNOSIS SUPPORT

Suitability of a diagnosis support and its intuitive transparency depend on medical knowledge representation. An appropriate representation should solve the following problems: a consistent interpretation of symptoms of different nature, formulation of rules that is close to diagnostic heuristics and implementation of a credibility measure that estimates diagnostic hypotheses. The paper suggests a description of symptoms by means of classical and fuzzy sets and a model of diagnostic reasoning in the framework of the Dempster-Shafer theory. Thus, credibility of the diagnosis is defined by the imprecision measure (i.e. a characteristic or a membership function) and by the certainty of the rule (i.e. a basic probability assignment value). A unified approach is proposed for all kinds of symptoms and evidence that are used during the diagnosis. Models for symptoms and evidence that are difficult to define (e.g. pain), are suggested.

1. INTRODUCTION

Multiple parameters of different nature have to be represented in medical diagnosis support. The parameters can be either explicitly defined (e.g. pregnancy), or described by a fuzzy statement (e.g. obesity). A diagnosis support algorithm should represent all kinds of diagnostic information and interpretation of the parameters should be consistent. Moreover, knowledge representation should stick to the intuitive meaning of symptoms, as inference in diagnostic problems follows heuristic rules. Thus, conditions of rules in a diagnosis support system have to represent both crisp and fuzzy parameters. Yet, such symptoms as ‘pain’ or ‘lack of appetite’ are difficult to define. They can be hardly described by crisp concepts. On the other hand, for the majority of such symptoms a domain is not determined which cause troubles in their representation by fuzzy sets. Still, diagnosticians struggle for accuracy in their interpretation, as the example of pain (measured in the VAS: Visual Analog Scale) [2] shows. Not only symptoms, but also evidence are difficult to model. For instance, when a patient says that he has noticed a sign of a disease about three weeks ago, his statement is hardly crisp. It should be also noticed that representation of symptoms is not always the same as interpretation of evidence. For instance, in the rule a symptom is formulated as: “quick heart rate” and the observation is 90/min. Thus a method of matching rule condition (symptom) with the observation (evidence) has to be chosen. The method should also define a measure of matching accuracy. The paper suggests consistent representation of symptoms,

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

recommends fuzzy sets for difficult symptoms description and introduces a measure of matching symptoms with evidence. A method of determination of membership functions is presented. The diagnostic inference is done in the framework of the extended Dempster-Shafer theory.

2. REPRESENTATION OF SYMPTOMS

Heuristic rules that are commonly used in the diagnosis support have the following form:

$$\text{IF } X^1 \text{ is } X_l^1, \text{ and}, \ldots, \text{and } X^n \text{ is } X_l^n \text{ THEN } D_l \tag{1}$$

where X^i is a medical parameter (e.g. a laboratory test), $X^i{}_l$ is a linguistic variable that describes a value of the parameter (e.g. “elevated”), and D_l is a diagnosis. The linguistic variable $X^i{}_l$ may be represented by the characteristic function of the classical set $\chi^i{}_l(x)$ (in case of “present” or “absent” symptoms), or by a membership function of a fuzzy set $\mu^i{}_l(x)$ (for instance: “low test result”). Thus, the symptoms are defined either precise or by means of a measure of imprecision that is the membership function. Yet, the diagnosis is usually crisp, as physicians do not use such statements as “light pneumonia”. Though the diagnosis is precise, it can be uncertain, as sometimes disease manifestations are not clear. This can be expressed by the risk of the disease defined by a fuzzy set of a conclusion. Still, it seems more natural to define a credibility of the rule. It is possible in the framework of the Dempster-Shafer theory (DST) [3]. Hence, the rule (1) can be represented as:

$$\text{IF } X^{j1} \text{ is } X_l^{j1}, \text{ and}, \ldots, \text{and } X^{jn} \text{ is } X_l^{jn} \text{ THEN } D_l \quad \text{with certainty } m_l(a_j) \tag{2}$$

where j is the index of the rule, m_l is the basic probability assignment (bpa) [3] determined for the D_l diagnosis, and the set of focal elements [3] include all conditions of rule premises. The focal element may be single, i.e. $a^i{}_l \equiv$'X^i is $X^i{}_l$' or complex, i.e. $a^k{}_l \equiv$'X^{k1} is $X^{k1}{}_l$ and, …, and X^{kn} is $X^{kn}{}_l$'. The set of focal elements of the l-th diagnosis is $A_l=\{ a^j{}_l\}$, $j=1, \ldots, N$ and for this set the DST principles [3]:

$$m_l(f)=0, \qquad \sum_{\substack{a^l_j \in A_l \\ j=1,\ldots,N}} m(a^l_j)=1, \tag{3}$$

hold true. Hence, the m_l bpa is the measure of certainty of the rules that refers to the D_l diagnosis.

3. MATCHING SYMPTOMS WITH EVIDENCE

The symptoms are estimated by characteristic or membership functions, still, observations of patient's findings are not necessary represented in the same way as the premise conditions. Several cases of matching observed evidence and symptoms could be listed. They are:

- Symptom is described by the characteristic function and evidence is an argument of the function. Let’s assume that “struma” is the X^j symptom and during an examination a physician observes the struma. Then: χ_{struma}(“present”) = 1 and χ_{struma} (“absent”) = 0 and evidence is χ_{struma} (“present*”) for which accuracy of matching equals 1.

- Symptom is represented by a fuzzy set and evidence is a value of its domain, for example the symptom is “low laboratory test result” described by the $\mu_{low}(x)$ and evidence is the test result for a patient denoted as x^*. Then the $\mu_{low}(x^*)$ is the accuracy of matching the evidence and the symptom.

- Evidence and symptom are both described by fuzzy sets, and the domain of these sets is either known or assumed, but generally accepted. For instance the symptom “gradual exacerbation of the disease manifestations” can be defined by means of the fuzzy set over the time domain and the evidence “the disease manifestations started about 3 weeks ago” may create another fuzzy set on the same domain. An example of the assumed scale is VAS (Visual Analog Scale [2]). In this approach the symptom “the worst pain” ($\mu_{worst}(x)$) can be matched with the observation “mild pain” ($\mu_{mild}*(x)$).

In case of the characteristic function, evidence either entirely matches the symptom or excludes it from inference. On the contrary, when the symptom or evidence is fuzzy, a partial accuracy of matching is possible. This partial accuracy will be defined as the imprecision level η. Hence, for the j-th focal element:

$$\eta_j = \min_{i=1,\dots,n} \eta_{ji}; \quad \eta_{ji} = \chi_{ji}(x_i^*) \text{ or } \eta_{ji} = \mu(x_i^*) \text{ or } \eta_{ji} = \max(\mu_{ij}(x_i) \wedge \mu_{ij}^*(x_i)) \qquad (4)$$

Now, an inference may concern only these symptoms that are sufficiently reliable. Namely, if the imprecision level of a focal element is greater than a threshold, then the focal element takes part in the inference. For such symptoms the belief in the diagnosis can be determined as:

$$Bel(D_l) \sum_{\substack{a_j^l \in A_l \\ \eta_j \ge \eta_T}} m(a_j^l), \qquad (5)$$

in analogy to the original definition of belief in DST [3]. When belief values of all possible diagnoses are calculated, then the final diagnosis can be determined as a result of their comparison. A detailed description of the algorithm of diagnosis support is given in [4].

4. MEMBERSHIP FUNCTIONS

Let us assume that a symptom takes part in an inference only when it is supported by evidence with the imprecision level greater than the threshold η_T. If evidence is evaluated by the η_T criterion, then a membership function shape has an essential influence on the results of the diagnostic inference. Trapezoidal and triangular functions have been chosen for the research. The crucial points for determination of a trapezoid have been quartiles of an empirical data distribution and cross-points of theoretical distributions (see fig.1). The points can be found if training sets of at least two competing diagnostic hypotheses are given. Generally, statistical features of training data sets are used to determine membership functions [5]. Still an application of the cross-points to the

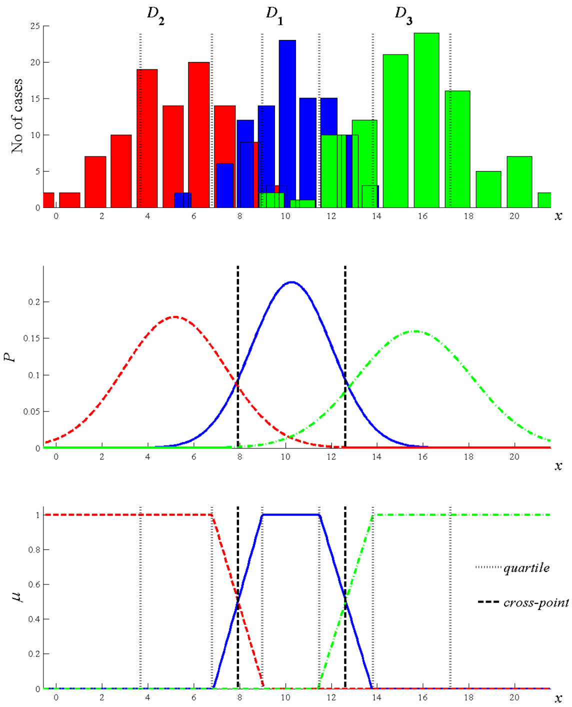

Fig.1. Membership function design.

membership function determination is a novel approach. It is assumed, that values of membership functions in the cross-points equal 0.5 and values in the quartiles equal 1 (fig.1). The points that are originally set at quartiles, can be moved during research toward and away from the mean value. In this way an investigation of the most appropriate shape of the membership function can be done. The shape is gradually changed until boundary functions have been obtained. The latter are: a triangular function and a function with slopes of 0.99 gradient (close to the characteristic function). The shapes have been tested for the three kinds of data. The first database is available on the Internet: ftp.ics.uci.edu/pub/machine-learning-data-bases/thyroid-disease, files new-thyr.*. The database is related to the problem of thyroid gland diseases and the reference [1] provides its comprehensive statistical analysis. The data are called the “Internet data” from now on. Number of training/test cases for the data have been as follows: hyperthyroidism 75/75, euthyroidism 15/20,and hypothyroidism 15/15. The data have been used to check whether a change of the membership function for real data can improve quality of diagnosis support. Next, data have been simulated with comparable statistic parameters to the Internet data. Each diagnostic group has included 100/100 cases. In such a way sets for 50 simulation runs of the method have been prepared. The simulated data have been used to find out which shape of membership functions is the most efficient. The third database has been the well-known Iris Plants Database from http://www.ics.uci.edu. The data have been divided into 25/25 train/test cases for the three defined categories. The database is a benchmark in research works and so it has been used to verify correctness of the proposed method.

5. RESULTS

Verification of the proposed methods has consisted in an error investigation. The error is the percentage of wrong diagnoses indicated for the test data. Wrong diagnoses have been outputs different from that specified in the test data as well as undetermined diagnoses (because of belief equality). Three diagnostic categories have been considered. Authors of previous investigations [1] have concluded that several statistical methods have been acceptable for selection of two categories (e.g. “hyperthyroidism” and “the rest”), but a classification for the three categories has always failed. In case of the proposed method the error found for the Internet data has been reduced to 2.67% when the optimum threshold for the imprecision level has been chosen. Generally, good results have been obtained for the threshold in the [0.1, 0.9] interval. Thus, the choice of the threshold during a diagnosis do not limit variety of considered symptoms, it only excludes unclear observations (for which $\eta_j < 0.1$) and accept for investigation quite reliable, thought not entirely precise symptoms. The upper limit of the advantageous η_T values, which is lower than 1, indicates that an exclusive consideration of the most precise symptoms makes the diagnosis worse. The error has been significantly smaller in comparison to the classical methods [1]. Moreover, it has occurred only in case of euthyroidisim, so patient's health would not be jeopardized. Yet, a modification of membership functions makes results even better.

Tests of the Internet and simulated data have revealed a considerable influence of the membership function shape on the performance of the method. The distance between the quartile from training data and the cross-point of theoretical distributions $d=|q-x_{cros}|$ has been the base for the shape modification. The original quartile points have been moved for the distance $[-\delta d, \delta d]$, where δ has been changed with 0.05 step from δ=0, to δ=1, until the border function (almost a characteristic function or a triangular function) has been reached. Lower errors have been observed for the membership functions with $\mu(x)=1$ for $x\in[q_{low}-\delta^*d,\ q_{up}+\delta^*d]$ where $\delta^*\in[0.25, 0.75]$. The errorless classification has occurred for δ=0.65 and δ=0.7 for the Internet data. The minimum of mean errors for simulations appears for $\delta\in[0.65, 0.75]$. An analysis of the calculation results shows that the trapezoidal membership function should have maximum membership of 1 for a wide range of its domain and then should descent to 0 with a big gradient. Thus, the shape of the membership function differs from that used in classical fuzzy rules. The latter are often triangular or trapezoidal with slopes of small gradient, as they are meant to correspond normal distribution. Still, the membership functions for fuzzy focal elements should not have extremely steep slopes as for the shapes close to the characteristic functions again the error reaches 2.7% for the Internet data and increases beginning from δ=0.8 for simulations. Hence, membership functions should allow for better performance of the diagnosis support algorithm than strict norms.

In case of the Iris Plant Databases the error has depended on the training set. Calculations have resulted in the global error of 5.33% (2 cases wrongly classified in the groups of versico and virgini) when cases were divided at random. This result has been obtained for the similar shapes of membership functions, i.e. the similar $[q_{low}-\delta^*d,\ q_{up}+\delta^*d]$ interval. It has been suspected that sound training data may improve performance. Therefore, wrongly classified cases have been attached to the training data until the error has stabilized. Thus, all difficult cases were gathered in one set.

Then the learning and test sets have been exchanged. In this way a division of cases for sound training data and difficult test data have been made. Under such conditions, the classification of the test data has been errorless. It can be concluded that though the error is not big for random training data, it is better to split the data into two sub-sets and to select suitable training cases.

6. DISCUSSION AND CONCLUSIONS

An adequate representation of medical knowledge in diagnosis support systems is crucial for understanding diagnostic hints by physicians and the right performance of the systems. Fuzzy sets have been suggested for modeling symptoms difficult for two-valued logic description. However, it is important to create an opportunity to define not only symptoms, but also evidence by means of membership functions. Thus, arises the problem of a uniform representation of all kinds of input information and heuristic rules as well as matching them to get a diagnostic conclusion. The paper proposes a solution that is based on the extended Dempster-Shafer theory and fuzzy sets. Symptoms that are not related to a real-number scale, for instance pain, can be defined on the [0, 100] interval, as examples from medical practice suggest. Fuzzy evidence can be used to model ambiguous opinions of patients, for example: “the symptom occurred about 3 weeks ago”, “mild pain” or “slightly better heat tolerance”. A patient can choose a suitable linguistic expression and afterwards a fuzzy set may be assigned to it during an inference. The method described in this paper makes it possible to interpret all cases of symptom-evidence matching.

The present research resulted in indications for an efficient method of designing membership functions. Usually, one membership function is needed for each diagnostic hypothesis. All possible diagnostic hypotheses have to be considered simultaneously during design of membership functions. The functions have to cover all the domain of a medical parameter. It is convenient when all membership functions of a medical parameter cross at the same level, e.g. 0.5. Then, a threshold can be chosen that excludes the most unreliable symptoms from inference, yet permits other imprecise findings to influence the diagnosis. Trapezoidal membership functions work generally better than triangular. The proposed method of the membership function designing and tuning is convenient for interpretation of all kinds of imprecise medical parameters, regardless their domains. The interpretation is intuitively clear for physicians. Taking into account the above mentioned remarks the proposed methods can be considered as a convenient model of knowledge representation for the purpose of the diagnosis support.

BIBLIOGRAPHY

[1] COOMANS D., BROECKAERT I., JONCKHEER M., MASSART D. L., Comparison of multivariate discrimination techniques for clinical data-application to the thyroid functional state, Meth. Inform. Med., Vol. 22, pp. 93-101, 1983.

[2] HO K., SPENCE J., MURPHY M.F., Review of pain-measurement tools, Annals of Emergency Medicine Vol. 27:4, pp. 427-432, 1996.

[3] KACPRZYK J., FEDRIZZI M. (Eds.), Advances in Dempster-Shafer Theory of Evidence. J. Wiley, New York. (1994)

[4] STRASZECKA E., An interpretation of focal elements as fuzzy sets, Int. J. of Intel. Systems Vol.18, pp.821-835, 2003.

[5] SCHUERZ M., HIPF G., GRABNER G., An assessment of different approaches to defining fuzzy membership functions semi-automatically, Proc. ERUDIT-Workshop, pp.129-137, Vienna, Austria, 2000.