DEMPSTER-SHAFER THEORY IN MEDICAL DIAGNOSIS SUPPORT

Ewa Straszecka*, Joanna Straszecka**

*Div. of Biomedical Engineering, Inst. of Electronics, Silesian Technical University
16 Akademicka St., 44-100 Gliwice, Poland
phone (48)(32)2371725, fax (48)(32)2372225, e-mail ewa@biomed.iele.polsl.gliwice.pl

**Department of Internal Diseases and Rheumatology
Medical University of Silesia
102 Edukacji St., 43-100 Tychy, Poland
phone (48)(32)2276071-215, fax (48)(32)2277723

ABSTRACT: The paper presents possible application of Dempster-Shafer theory of evidence in medical diagnosis support. An interpretation of diagnostic rules and determination of a basic probability assignment are proposed.
A method of illustration of a diagnosis uncertainty when some symptoms are unknown is described. The theory can be particularly useful in rare diseases diagnosis support as well as whenever frequency of symptoms occurrence significantly depends on a population.

1. INTRODUCTION

Pure probability theory can hardly be used in medical diagnosis support. It is difficult to find whether disease symptoms are independent, if symptoms observed for one living organism can be treated as autonomous at all. Estimation of *a priori* probability of a disease, which is necessary for Bayes conditional probability calculations, is also troublesome. An interpretation of a situation when some symptoms (often many of them) are unknown is almost impossible, or at least complicated. However probability - based measures cannot be neglected. They play an important role in many reasoning processes and their usefulness was proved in practical applications (for instance in the famous expert system MYCIN). A theory that takes much of probability properties and softens its rigid assumptions is the Dempster-Shafer theory (DST) of evidence.

In DST the knowledge is represented by a link from a predicate which is a focal element (a symptom or several symptoms) to a hypothesis (diagnosis in the discussed problem). Knowledge about the predicate consists in a basic probability assignment (bpa). It is a measure m defined for every focal element [1]:

$$m(f) = 0$$
$$\sum_{a \in T} m(a) = 1 \qquad (1)$$

where: f - false predicate, T - set of focal elements.

Such an unconstraint definition allows omitting most of the probability drawbacks. Nevertheless, the bpa, which is the fundamental part of DST, must be created very carefully. Indeed the bpa carries all the knowledge about a diagnosis, as diagnostic rules only link focal elements (a) with a hypothesis (b) thus making possible to calculate a belief measure [1]:

$$\text{Bel}\ (a) = \sum_{(b \Rightarrow a) = t} m\ (b) \qquad (2)$$

where t stands for truth.

In the paper an interpretation of diagnostic rules and a determination of basic probability assignment are proposed.
A procedure of belief calculation when several symptoms are unknown is suggested. Eventually, an idea of combining bpas for expert knowledge and statistical data is given.

2. AN INTERPREATION OF MEDICAL DIAGNOSIS IN TERMS OF DST

Rules of medical diagnosis can be interpreted in DST as links that join symptoms to a hypothesis. The hypothesis can be a disease or health. Symptoms are focal elements. The links are evaluated by a bpa, which must be determined for every focal element, i.e. for every symptom or closed collection of symptoms. The belief in the hypothesis is calculated as a sum of focal element assignments, which are of its favor, according to the formula (2). Let's assume that we have a collection of symptoms of a disease (d) and health (h) denoted by $\{s_{d1}, \ldots, s_{dk}\}$ and $\{s_{h1}, \ldots, s_{hn}\}$, respectively. They can be considered as sets with

The study is a part of the grant No.8T11E01215 of the State Committee for Scientific Research

an empty conjunction: $\{s_{d1}, \dots, s_{dk}\}$ $\{s_{h1}, \dots, s_{hn}\}=\emptyset$. Obviously, it happens that a symptom of a disease occur with a patient who is not ill, but certainly it is not a symptom of health.

A symptom is defined as a health parameter with a relation to its norm:

$$symptom \equiv parameter < relation > norm$$

For instance T_3 laboratory test is a symptom of hyperthyroidism if its result is above the upper threshold of the norm and is a symptom of health if its result falls into norm limits. Thus the norm plays an important role in the division of symptoms for these which imply a disease and those which indicate health. In the diagnosis of hyperthyroidism the symptoms of the disease and health are most often adequate: the same *j-th* health parameter and norm with different relations are s_{dj} and s_{hj}. Naturally other symptoms can be used too, but usually they are in minority. The adequacy of symptoms set off the role of a norm. An automatic change of a symptom meaning and importance is possible if a user of a diagnosis support system changes its norm. This property becomes essential whenever norms are established individually by laboratories, as it is in case of hyperthyroidism tests.

A symptom can be as well defined as a complex one:

$$symptom_1 \wedge symptom_2 \equiv symptom$$

For instance symptoms related to laboratory tests T_3 and FT_3 are interpreted together as one (complex) symptom.

According to (2) two bpas must be defined for hyperthyroidism diagnosis: m_d and m_h, specifying measures for every symptom of the disease and health. The bpas can be found in different ways:

- flat bpa, for which

$$m_d(s_{di}) = \frac{1}{k}, \quad i = 1, \dots, k; \quad m_d(s_{hj}) = \frac{1}{n}, \quad j = 1, \dots, n;$$

- bpa which is described in medical handbooks or expresses an expert knowledge - an importance of a symptom influences values of $m_d(s_{di})$ and $m_h(s_{hj})$; score test interpretation can be the basis of such an assignment;
- bpa defined on probability (frequency of occurrence).

The main advantage of the third method is a possibility to represent empirical distribution. The bpa can express which symptoms imply a disease most often in a population, and thus allow distinguishing symptoms, which are most frequent. To this end, frequencies of occurrence for a disease symptom and a health symptom are calculated. On the basis of frequencies, a new bpa might be built according to (1) and the conditional probability definition. However, a problem is that in many cases probability *a priori* of a disease, i.e. $P(d)$ cannot be estimated. In that case a sensitivity of a symptom can be used to adjust its influence on a diagnosis.

$$m_d(s_{di}) = [\mathrm{P}(s_{di}/d) - \mathrm{P}(s_{di}/\neg d)] \quad i = 1, \dots, k;$$

where brackets denote normalisation so that (1) holds true. Sometimes it is justified to use importance of a symptom in the diagnosis. In that case bpa can be determined as:

$$m_d(s_{di}) = [\mathrm{P}(s_{di}/d) - \mathrm{P}(\neg s_{di}/d)] \quad i = 1, \dots, k;$$

where $\neg s_{di}$ denotes the absence of the symptom.

In similar way m_h can be built.

A difference among frequencies of a symptom occurrence observed for ill and healthy patients represents the significance of each symptom in a diagnosis. The difference can be considered as $m_d(s_{di})$ value or, when it is negative, it can show that a predicate, though generally true, imply a contradictory conclusion in a population. Thus, the belief measure can be built on a probability assignment in the following way:

$$\mathrm{Bel}(d) = \sum_{(s_{di} \Rightarrow d) = \mathrm{t}} m(s_{di}) = \sum_{(s_{di} \Rightarrow d) = \mathrm{t}} [\mathrm{P}(s_{di}/d) - \mathrm{P}(s_{di}/\neg d)] = \quad i = 1, \dots k;$$

If s_{id} and s_{ih} are a couple with the same health parameter, then sensitivity as well as importance calculation is straightforward. A sample of such calculations for the bpa estimation in hyperthyroidism is given in table 1. Note that the assumption of couple of symptoms allows: $\mathrm{P}(s_{di}/\neg d) = \mathrm{P}(s_{hi}/h)$. The negative value for S2 symptom shows that the symptom is not relevant to the disease for the small population under consideration. Its elimination from calculations results in higher certainty and better differentiation among diagnoses.

Table 1.

Focal element	$m(s_{di}) = [\mathrm{P}(s_{di}/d) - \mathrm{P}(s_{di}/\neg d)]$
S1	0.106311
S2	-0.10968
S3	0.025061
S3∧S4	0.145353

It's worth noticing that in the table there is a complex symptom (S3∧S4) and that it is not independent from the other

focal element (S3).
Generally the bpa can be defined as follows:

$$m_h(s_i) = \begin{cases} \dfrac{P(s_i / h) - P(s_i / \neg h)}{\sum\limits_{(s_i \Rightarrow h)=t} P(s_i / h) - P(s_i / \neg h)} & \text{if } P(s_i / h) - P(s_i / \neg h) > 0 \\ 0 & \text{otherwise} \end{cases}$$

or

$$m_h(s_i) = \begin{cases} \dfrac{P(s_i / h) - P(\neg s_i / h)}{\sum\limits_{(s_i \Rightarrow h)=t} P(s_i / h) - P(\neg s_i / h)} & \text{if } P(s_i / h) - P(\neg s_i / h) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Where *h* stands for a hypothesis this time (which can be health or a disease).
Such a definition of bpa allows expressing dependencies existing in a small population without medical rules violation.

3. CERTAINTY OF A DIAGNOSIS

Symptoms are focal elements in DST. However we are free to define one more focal element which represents a collection of unknown symptoms. The focal element denoted by *è* represents a level of ignorance in case of some symptoms are not known. The *m*(*è*) is equal to the sum of bpa values for unknown symptoms. In this way (1) holds true and the interpretation of the situation can be: we define individual bpa for every possible collection of unknown symptoms.

Let's assume that it is a uniform bpa. If all symptoms are examined then $m_d(s_{di})$=0.1 i=1,...,10 and m_d(*è*) = 0. If some symptoms, for instance four, are unknown , m_d(*è*) is equal to a sum of probability assignments for the not stated symptoms, i.e. m_d(*è*) = 0.4. Let's say that the other 6 symptoms indicate the disease. In this case Bel(*d*)= 0.6 and Cert (*d*)= 0.4. At first glance Cert seems to be a useless measure as if for a patient every known symptom is in favour of the disease, i.e. $\forall_{s_{di}} (s_{di} \Rightarrow d) = t$ then Bel(*d*) +Cert(*è*)=1. However that is not the case when:

- some present patient's symptoms point to alternative diagnosis (health)
- there are symptoms about which a physician is not fully convinced; often a doctor signs "+/-" (not clear) mark in a questionnaire instead of "+" (present) or "-" (absent). The "+/-" mark cannot be considered as an unknown symptom.

Every experienced researcher will admit that the above mentioned cases cover over 90% of diagnoses.
Hence, Cert as a measure of ignorance is very useful. In the present problem it is sufficient to define the certainty of a diagnosis as Cert= 1-*m*(*è*), in spite of a number of other, more complicated definitions given in references [5]. The *Cert* in more details is discussed in [3], [4].
The measures Bel and Cert make it possible to represent both what is the possibility that a patient is ill and how sure is the diagnosis. The calculations are very simple, but give a good representation of information used in reasoning during the diagnostic process.

4. DEFINING BASIC PROBABILITY ASSIGNMENTS AS A COMBINATION OF INFORMATION

Two sources of information - for instance tables from medical handbooks and statistical data of a hospital information system can be joined in the Dempster-Shafer theory. The both sources of basic probability assignments are combined in the following way [1]:

$$\forall_{a \neq f} \; m(a) = \frac{\sum_{b \wedge c = a} m_1(b) \cdot m_2(c)}{\sum_{b \wedge c \neq f} m_1(b) \cdot m_2(c)} \qquad (5)$$

The property is particularly useful in case of medical reasoning [2] as it makes possible to interpret general knowledge base in terms of a specific database. Let's assume that two bpas: m_{d1} and m_{d2} are defined. For every disease symptom a joined bpa (m_d) is equal:

$$\forall_{s_{di}\neq f}\ m_d(s_{di}) = \frac{\sum_{s_{di}\wedge s_{di}=s_{di}} m_{d1}(s_{di})\cdot m_{d2}(s_{di})}{\sum_{s_{dj}\wedge s_{dj}\neq f\ j=1,\ldots,k} m_{d1}(s_{dj})\cdot m_{d2}(s_{dj})} = \frac{m_{d1}(s_{di})\cdot m_{d2}(s_{di})}{\sum_{s_{dj}\neq f\ j=1,\ldots,k} m_{d1}(s_{dj})\cdot m_{d2}(s_{dj})} \qquad (6)$$

On the analogy m_h can be obtained.
If m_{d1} is a flat (uniform) i.e. $m_{d1}(s_{di})= m_c$=const, i=1,...k, and all disease symptoms are known and present then the joined bpa:

$$\forall_{s_{di}\neq f}\ m_d(s_{di}) = \frac{m_c\cdot m_{d2}(s_{di})}{\sum_{s_{dj}\neq f\ j=1,\ldots,k} m_c\cdot m_{d2}(s_{dj})} = \frac{m_c\cdot m_{d2}(s_{di})}{m_c \sum_{s_{dj}\neq f\ j=1,\ldots,k} m_{d2}(s_{dj})} = m_{d2}(s_{di})$$

is equal to m_{d2} . It means that if the general knowledge does not distinguish any symptom, symptoms' importance is given by the information from database. Of course, if m_{d1} carry some information, it is included in m_d too. Naturally the formula (6) can be used for symptoms of disease and health as well as for unknown symptoms.
If m_{d1} is an assignment built on the basis of medical handbooks and m_{d2} is an empirical distribution the combination (6) helps to ‘see’ general knowledge ‘in the light’ of a population under consideration.
The possibility to combine bpas enables flexible interpretation of a support system knowledge base and therefore is of major importance for diagnosis support systems.

5. CONCLUSIONS

The Dempster-Shafer theory of evidence (DST) is a suitable tool for diagnosis supporting. It makes it possible to combine knowledge from medical handbooks and databases. However, its essential element – basic probability assignment must be carefully designed. First disease symptoms must be precisely determined. Next a sort of bpa (uniform, sensitivity- or importance-based) must be chosen. A flat (uniform) bpa can be successfully combined with other information, therefore it is very easy to make an option of symptom weight choice in a diagnosis support system. It is important as combining similar assignments can blow up weights of some symptoms.
It is possible to interpret certainty of a diagnosis using the same bpa as for belief measure calculation. Belief and certainty factors make a diagnosis more complete than in case of one imprecision parameter.
Authors’ experiences of application of DST for hyperthyroidism diagnosis support show that the theory is easy for implementing in every software environment.
Patients' data are often incomplete, for rare diseases they are sometimes limited to a collection of several instances. Statistical interpretation of such data is doubtful or impossible. DST can help to use the data without carelessness of their insufficient quality.
DST could find an extensive field of application in information description in medical diagnosis support.

REFERENCES

[1] J. Gordon, E.H. Shortliffe - The Dempster-Shafer Theory of Evidence, Rule -Based Expert Systems, Buchanan B.G., Shortliffe E.H. edts, Addison Wesley 1984, pp.272-292

[2] E.H. Shortliffe, B.G. Buchanan - A Model of Inexact Reasoning in Medicine, Mathematical Biosciences 23, 1975, pp.351-379

[3] Straszecka E., Straszecka J. - Dempster-Shafer theory of evidence in application to medical diagnosis support, Proc. Int. Workshop BEMI’97 Sept. 97, Gliwice, Poland pp 40-44

[4] Straszecka E., Straszecka J. - Dempster-Shafer theory and basic probability assignment in diagnosis support, Proc. Joint EUROFUSE-SIC’99, in print

[5] R.R. Yager, M. Fedrizzi, J. Kacprzyk - Advances in the Dempster-Shafer Theory of Evidence, John Wiley & Sons Inc., 1994