Listy Biometryczne – Biometrical Letters
Vol. 30 (1993), No. 1, 3-12

On modelling membership function values in diagnostic decisions

Tadeusz Gerstenkorn and Elżbieta Rakus

Institute of Mathematics, University of Łódź,
S.Banacha 22, PL 90-238 Łódź, Poland,
Laboratory of Sports Medicine, Medical Academy of Łódź,
Al.Politechniki 4, PL 93-590 Łódź, Poland

Summary

The paper presents some methods of Fuzzy Set Theory which can act as a helpful apparatus in determining medical diagnosis on the basis of symptoms. A max-min composite fuzzy relation equation is a mathematical key to finding the solution of this problem. The relationship between patients and symptoms and diagnoses is considered as a fuzzy relation. Extending the very well-known conceptions it is demonstrated that one can compute a degree of the membership functions which is characteristic of the patient-symptom relation. The S-function class has been applied to symptoms, which take continuous values, as a method of estimating the membership function. In the case of symptoms, which have been characterized by taking discrete values, concept of the linguistic variable is introduced to obtain the same result. The final appreciation of the patient-symptom relation has been done by means of the concrete mathematical solution and it may be helpful to physicians in confirming their intuitive assumption.

1. Introduction

Fuzzy set theory has recently brought attempts to establish mathematical conceptions assisting the medical diagnosing process. In these inquiries the mathematical apparatus differs many a time considerably, but all the methods are connected by the common assumption that medical diagnoses are established hierarchically on the basis of the symptoms that have been found in the set of patients being consulted. Let us introduce three non-fuzzy sets: a set of symptoms $S=\{S_1, S_2,\ldots,S_m\}$, a set of diagnoses $D=\{D_1, D_2,\ldots,D_n\}$ and a set of patients

Key words: fuzzy sets, composite fuzzy relation equation, diagnostic decisions, modelling membership function values, patient-symptoms-diagnosis relation

$P = \{P_1, P_2, \ldots, P_k\}$. The symptoms adopted in the set S are connected with the diagnoses from the set D and it is assumed that the data concerning their occurrence in the population of patients P are available. The method of medical diagnosing we show here has arisen on the basis of the compositional rule of inference, represented by the relation equation

$$PD = PS \circ SD \tag{1}$$

where PD, PS and SD are fuzzy relations, and the sign "∘" denotes a composition of the fuzzy relations in the max-min sense. This problem was discussed in papers by Adlassnig (1980 and 1986), Rakus (1991) and Sanchez (1979). The present report includes new solutions and complements which, may be, will improve the quality of the method and allow one to take decisions in doubtful cases, especially connected with PS. Let us describe the fuzzy relations occurring in equation (1), starting their characterization with the relation PS which is a mathematical equivalent of the relationship: patient-symptom. This relation is composed of pairs (P_r, S_i), $r = 1,\ldots,k$, $i = 1,\ldots,m$, and the membership degree of each pair describes the intensity of the symptom found in the patient.

2. Description of the method

Every symptom is a fuzzy set and can be represented in a different form, according to the kind of attribute. We shall distinguish three types of symptoms characterized by a qualitative attribute wearing a simple complexion, a qualitative attribute wearing a more complicated complexion called here a discrete qualitative attribute and a quantitative attribute. We shall next show the procedure of forming the fuzzy sets corresponding to these attributes, and also the way of calculating the membership degrees.

1) S_i is qualitative attribute – the fuzzy set characterizing this type of symptom is defined in the space $X = \{0,1\}$ with the membership degree 0 for the element $x = 0$ corresponding to the lack of the symptom and the membership degree 1 in the case of the presence of the symptom, which is symbolically assigned to the element $x = 1$. This set can be formally accepted as

$$S_i = 0/0 + 1/1\ . \tag{2}$$

There are biological qualitative parameters whose occurrence can be described more widely by, for instance, introducing a kind of questionnaire including alternative answers to the questions concerning the given attribute. Let us call this attribute a symptom of discrete type and assign to it the fuzzy set corresponding to this tendency.

2) Let, at the first stage of our considerations, S_i be a linguistic variable (Gerstenkorn, Rakus, 1990; Kacprzyk, 1986; Saitta, Torasso, 1981) whose set of terms is formed indirectly by answers to the questions $q_p^{(i)}$, $i = 1,\ldots,m$, $p = 1,\ldots,Q^{(i)}$, where $Q^{(i)}$ constitutes the number of questions describing the symptom. Each question is connected with the name of the fuzzy variable defined in the discrete reference set U whose elements are codes of answers to the questions $q_p^{(i)}$, denoted by $s_{p,t}^{(i)}$, $p = 1,\ldots,Q^{(i)}$, $t = 1,\ldots,t^{(p)}$, where $t^{(p)}$ expresses the number of alternative answers to the question posed. In the further procedure, one should assign weights $w_{p,t}^{(i)}$ for the encoded answers $s_{p,t}^{(i)}$. These weights are numbers belonging to the interval [–1,1]. It is assumed that negative numbers are connected with negative answers to the question posed, that is, not confirming the occurrence of the symptom, while a positive value of the weight gives the answer a positive character, i.e. witnesses the presence of a symptom in the patient. The weight equalling 0 or close to this value is reserved for the case of a lack of answer or a statement bringing no information. The introduction of fuzzy restrictions is not essential at this stage of reasoning.

Let us treat S_i as a fuzzy set defined in the space X where elements $x \in$ X are the numbers

$$x = \sum_{p=1}^{Q^{(i)}} w_{p,t_p}^{(i)} \tag{3}$$

which have arisen as a sum of weights of answers to the questions connected with this attribute.

The values x constitute a support of the fuzzy set, bounded below by a number α which is calculated as

$$\alpha = \sum_{p=1}^{Q^{(i)}} \min_{1 \le t \le t^{(p)}} w_{p,t_p}^{(i)} \tag{4}$$

and, respectively, bounded above by a number γ expressed by the formula

$$\gamma = \sum_{p=1}^{Q^{(i)}} \max_{1 \le t \le t^{(p)}} w_{p,t_p}^{(i)} \; . \tag{5}$$

Consequently, a symptom of discrete type is a fuzzy set with values from the interval $[\alpha,\gamma]$, whose membership function is proposed to be adopted as follows:

$$\mu_{S_i}(x) = S(x,\alpha,\beta,\gamma) \; , \tag{6}$$

where S is a standard function of the class S, described by the formula (see: Adlassnig 1980, Czogała and Pedrycz 1985)

$$S(x,\alpha,\beta,\gamma) = \begin{cases} 0 & \text{for } x \le \alpha, \\ 2\left(\dfrac{x-\alpha}{\gamma-\alpha}\right)^2 & \text{for } \alpha < x \le \beta, \\ 1-2\left(\dfrac{x-\gamma}{\gamma-\alpha}\right)^2 & \text{for } \beta < x \le \gamma, \\ 1 & \text{for } x > \gamma. \end{cases} \qquad (7)$$

If the value of the weight assigned to an extremely negative answer is fixed to be −1, and the weight of its most positive variant is equal to +1, and this rule is preserved for each question, then the graph of the function S will take the form of Figure 1.

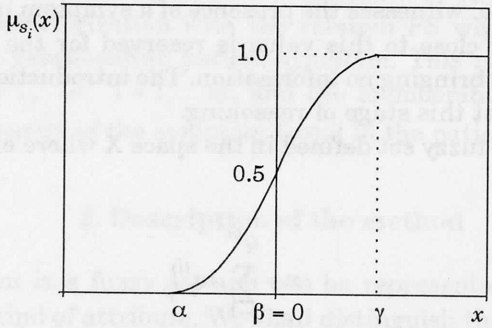

Figure 1. The membership function of the fuzzy set S_i representing a symptom of the discrete type

Analyzing the graph, let us notice that to the number α which is a sum of minimal weights, i.e. those connected with answers negating the occurrence of the symptom, corresponds the membership degree equal to 0, always being an exponent of health and assigned to the patient in whom the symptom S_i did not get pathological. The number γ is a sum of maximal "positive" weights confirming the revealed symptom. Therefore, it is a logical consequence that the membership degree assigned to it takes the value 1. A vague piece of information or its lack, which in the numerical interpretation is expressed by value β, has its representation in the form of the membership degree equalling 0.5.

Coming back to the relation PS, the pair (P_r,S_i), $r = 1,\ldots,k$, $i = 1,\ldots,m$, has the membership degree calculated on account of (7) and on the basis of the value x defined by (3) with the parameters α,γ given by formulae (4), (5). These operations are made in the case of the physician's decision that the symptom should be characterized by means of a special questionnaire.

3) The last (third) type of symptom to be considered is a quantitative attribute, i.e. one that takes values continuously from a known interval determined by the

physician. These values are calculated by means of laboratorial tests, special measurements, and the like. Similarly as before, the symptom S_i is a fuzzy set whose support is the interval containing all values of the attribute under consideration. Let us write down this interval as [VMIN, VMAX] with notations VMIN – the minimal value the examined parameter can take, and VMAX – respectively, the maximal value.

Denote by the symbols VN_1, VN_2 ends of the interval in which quantities characteristic of a healthy man occur. It should be noticed that, outside the interval (VN_1, VN_2), both the deficiency and the excess of a biological indicator are signs of a disease, most often connected with different diagnoses. In view of the above, it seems purposeful to divide the originally fixed interval [VMIN, VMAX] into three subintervals, namely,

$$[\mathrm{VMIN}, \mathrm{VMAX}] = [\mathrm{VMIN}, \mathrm{VN}_1] \cup (\mathrm{VN}_1, \mathrm{VN}_2) \cup [\mathrm{VN}_2, \mathrm{VMAX}] \quad . \qquad (8)$$

In the case of symptoms whose uniform growth of values reflects a progressive morbid state, we propose to adopt the following membership function:

$$\mu_{S_i}(x) = \begin{cases} S(x, \mathrm{VN}_2, \dfrac{\mathrm{VN}_2 + \mathrm{VMAX}}{2}, \mathrm{VMAX}) & \text{for } x \in [\mathrm{VN}_2, \mathrm{VMAX}], \\ 0 & \text{for } x \in (\mathrm{VN}_1, \mathrm{VN}_2), \\ 1 - S(x, \mathrm{VMIN}, \dfrac{\mathrm{VMIN} + \mathrm{VN}_1}{2}, \mathrm{VN}_1) & \text{for } x \in [\mathrm{VMIN}, \mathrm{VN}_1] \end{cases} \qquad (9)$$

whose graph is shown in Figure 2.

In the case when the physician finds that the growth of a value characterizing a symptom does not matter essentially till some moment, and only greater values are connected with a violent deterioration of the health condition, it would be purposeful to apply a concentration operation for the fuzzy set S_i, that is, to use the membership function of the type:

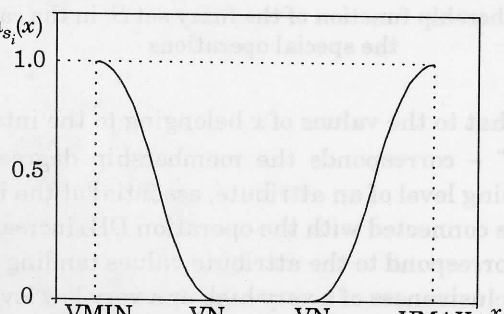

Figure 2. The membership function of the quantitative S_i characteristic of the progressive morbid state

$$\mu_{CON(S_i)}(x) = \{\mu_{S_i}(x)\}^2 . \tag{10}$$

Conversely, if the physician decides that what constitutes the greatest danger for one's health is the growth of the symptom value at its first stage, then it is advisable to introduce a dilution operation for the symptom S_i which changes the membership function in the following manner:

$$\mu_{DIL(S_i)}(x) = \{\mu_{S_i}(x)\}^{1/2} . \tag{11}$$

The "branches" of the membership function, modified in such a manner, are shown in Figure 3. In the description of this figure, the following notations are adopted:

$$\mu_{S_i'}(x) = S(x, VN_2, \frac{VN_2 + VMAX}{2}, VMAX) \tag{12}$$

$$\mu_{S_i''}(x) = 1 - S(x, VMIN, \frac{VMIN + VN_1}{2}, VN_1) \tag{13}$$

which is compatible with formula (9).

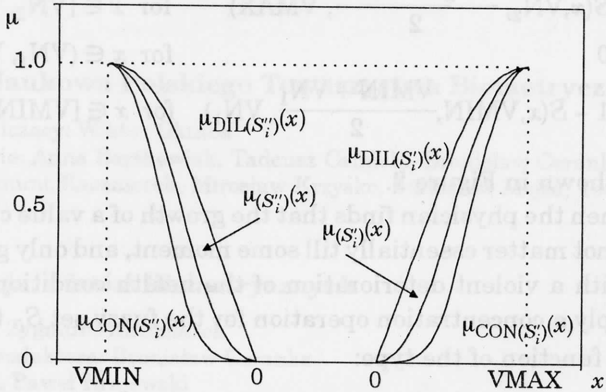

Figure 3. The membership function of the fuzzy set S_i in the case of applying of the special operations

To sum up, note that to the values of x belonging to the interval (VN_1, VN_2) – "the health interval" – corresponds the membership degree equalling 0. An increasing or decreasing level of an attribute, essential at the initial stage of the diagnosing process, is connected with the operation DIL increasing the membership degrees which correspond to the attribute values tending to normal values. In the case of the conclusiveness of a very high or a very low level of the symptom considered, the operation modifies the membership degrees in such a way that the greatest of them should fall to values shifted in the direction of the extremal quantities VMIN, VMAX.

3. Final remarks, example and conclusions

The method of calculating the membership degrees of a quantitative attribute, presented here, ought to be particularly useful in the cases when the physician cannot give precise numerical values of a symptom, recognized as limits of an "increasing level" or "strongly increasing level" if the tendency is known only.

The relation PS, in which the membership degrees have been defined for each pair (P_r, S_i), $r = 1,\ldots,k$, $i = 1,\ldots,m$, by means of one of the methods described above, constitutes the first component of equation (1). The so-called "medical knowledge", i.e. the relation SD expressing a relationship between symptoms and diagnoses, is the second element of the relation equation, and its exact description as well as the final result, i.e. PD, can be found in papers by Adlassnig (1980 and 1986), Rakus (1991) and Sanchez (1979).

As an example of a set of a qualitative type we can present fuzzy sets built on the basis of clinical data collected for women with trichomoniasis where each symptom S_i, $i = 1,2,\ldots,6$, is the mentioned attribute (Gerstenkorn et al. 1991 and Rakus 1991):

S_1 - pruritus of burning sensations in vulva or vagina,
S_2 - pruritus of burning sensation in urethra,
S_3 - pain during uresis,
S_4 - changes of vulva,
S_5 - kolpitis diffusa,
S_6 - kolpitis maculosa.

For the reason of the lack of possibility to show the whole matrix PS constructed for 99 patients (Rakus 1991), we only present here the shortened form of it, i.e.

$$
\mathrm{PS} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ \cdot\cdot \\ P_{99} \end{array}
\begin{array}{c} \begin{array}{cccccc} S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{array} \\
\left[\begin{array}{cccccc}
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 0 & 0 & 0 & 1 & 0
\end{array}\right] \end{array}
$$

The second element of equation (1) is the fuzzy relation concerning the relationship between the symptoms S_i, $i = 1,2,\ldots,6$, and the diagnoses D_j, $j = 1,2,3$.

The above symptoms are characteristic of such diagnoses as:

D_1 - trichomoniasis,

D_2 - mycoses,

D_3 - bacterial infection.

The relationship between S_i and D_j is considered in two aspects – presence S_i in D_j and conclusiveness S_i for D_j.

Two matrices SD_1 and SD_2 constructed by means of methods described in Rakus (1991) correspond with SD which is a component of equation (1).

The matrix SD_1 shows the presence of S_i in D_j by means of the membership degrees of pairs (S_i,D_j) according to Gerstenkorn et al. (1991) and Rakus (1991)

$$SD_1 = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} 0.955 & 0.955 & 0.18 \\ 1.0 & 0.18 & 0.18 \\ 0.82 & 0.005 & 0.005 \\ 0.82 & 0.82 & 0.045 \\ 0.5 & 0.995 & 0.82 \\ 0.5 & 0.005 & 0.005 \end{bmatrix} \end{array}$$

In the same way we can present SD_2 which – as the fuzzy relation – is a representative of the conclusiveness of S_i for D_j (Gerstenkorn et al., 1991 and Rakus, 1991)

$$SD_2 = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} 1.0 & 0.955 & 0.82 \\ 0.955 & 0.5 & 0.5 \\ 0.82 & 0.18 & 0.5 \\ 0.82 & 0.5 & 0.5 \\ 0.5 & 0.82 & 0.5 \\ 0.955 & 0.18 & 0.18 \end{bmatrix} \end{array}$$

The solution of five fuzzy relation equations based on equation (1), i.e.,

$$R_1 = PS \circ SD_1,$$

$$R_2 = PS \circ SD_2,$$

$$R_3 = \min(R_1, R_2),$$

$$R_4 = PS \circ (1 - SD_1),$$

$$R_5 = (1 - PS) \circ SD_1$$

Table 1
Final decisions concerning the accepted diagnoses

P	R_3 D_1	D_2	D_3	R_4 D_1	D_2	D_3	R_5 D_1	D_2	D_3	F
P_1	0.955	0.18	0.18	0.5	0.995	0.995	0.955	0.995	0.82	t
P_2	0.955	0.995	0.82	0.5	0.18	0.955	1.0	0.18	0.18	m
P_3	0.955	0.995	0.82	0.5	0.82	0.82	0.82	0.82	0.05	m
P_4	0.955	0.995	0.82	0.5	0.005	0.18	1.0	0.82	0.18	b
P_5	0.955	0.995	0.82	0.5	0.18	0.955	1.0	0.18	0.18	m
...	....	....	....	....	....	....	....	....	....	...
P_{99}	0.955	0.995	0.82	0.5	0.005	0.18	1.0	0.82	0.18	b

P - Patient
F - Final decision
t - trichomoniasis
m - mycoses

allows to find the final diagnosis by comparing the membership degrees of R_3, R_4 and R_5 as shown in Table 1.

The methods of determining the PS fuzzy relation described here have been applied to practical examinations carried out on the clinical data presented above where the theoretical way of searching allowed us to establish a correct diagnosis (Gerstenkorn et al. 1991). These results were confirmed by laboratorial tests, hence it may be hoped that the primary diagnosing made by means of the mathematical method can be particularly useful especially when the direct access to the apparatus performing special tests is unavailable.

REFERENCES

Adlassnig, K.P. (1980). A fuzzy logical model of computer assisted medical diagnosis. *Meth. Inform. Med.* **19**, 141-148.

Adlassnig, K.P. (1986). Fuzzy modeling and reasoning in a medical diagnostic expert system. *EDV in Medizin und Biologie* **17**, 1/2, 12-20.

Czogała, E. and Pedrycz, W. (1985). *Elementy i metody teorii zbiorów rozmytych.* Warsaw, PWN .

Gerstenkorn, T., Kurnatowska, A. and Rakus, E. (1991). Application of fuzzy set theory to medical diagnosis and treatment of inflammation of genital organs and urinary tract in women (in Polish).*Wiadomości Parazytol.* **36**, 5/6, 251-267.

Gerstenkorn, T. and Rakus, E. (1990). On the utility of the notions of a fuzzy variable and a linguistic variable in natural sciences (in Polish). *Listy Biometryczne – Biometrical Letters* **27**, 1/2, 3-12.

Kacprzyk, J. (1986). *Zbiory rozmyte w analizie systemowej*, Warsaw, PWN.

Rakus, E. (1991). *Zastosowanie teorii zbiorów rozmytych jako metody wspomagającej rozpoznanie i ocenę skuteczności leczenia* (Application of fuzzy set theory as a method assisting medical diagnosis and evaluation of drug action, in Polish). Doctoral thesis, Medical Academy of Łódź.

Sanchez, E. (1979). Medical diagnosis and composite fuzzy relations. In: *Advance in Fuzzy Set Theory and Applications*. Amsterdam, North-Holland, 437-444.

Saitta, L. and Torasso, P. (1981). Fuzzy characterization of coronary disease. *Fuzzy Sets and Systems* **5**, 245-258.

Received 12 October 1992; revised 17 December 1992