

Intuitionistic Fuzzy Sets in Intelligent Data Analysis for Medical Diagnosis

Eulalia Szmidt and Janusz Kacprzyk

Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
E-mail: {szmidt, kacprzyk}@ibspan.waw.pl

Abstract. We propose a new approach for medical diagnosis by employing intuitionistic fuzzy sets [cf. Atanassov [1,2]]. Solution is obtained by looking for the smallest distance [cf. Szmidt and Kacprzyk [7,8]] between symptoms that are characteristic for a patient and symptoms describing illnesses considered. We point out advantages of this new concept over the method proposed by De, Biswas and Roy [4] where intuitionistic fuzzy sets were also applied, but the max-min-max rule was used instead of taking into account all, unchanged symptom values as proposed in this article.

1 Introduction

For many real world problems, imperfect, imprecise information is a vital part of the problem itself, and a continuing reasoning without proper modelling tools may lead to generating inaccurate inferences.

Traditional methods of analysis are oriented toward the use of numerical techniques. By contrast, much of human reasoning involves the use of variables whose values are not numerical. This observation is a basis for the concept of a linguistic variable, that is, a variable whose values are words rather than numbers, in turn represented by fuzzy sets.

The use of linguistic variables represents a significant paradigm shift in system analysis. More specifically, in the linguistic approach the focus of attention in the representation of dependencies shifts from difference and differential equations to fuzzy *IF – THEN* rules in the form *IF X is A THEN Y is B*, where *X* and *Y* are linguistic variables and *A* and *B* are their linguistic values, e.g. *IF Pressure is high THEN Volume is low*.

Description of system behaviour in the language of fuzzy rules lowers the need for precision in data gathering and data manipulation, and in effect may be viewed as a form of data compression.

But there are situations when description by a (fuzzy) linguistic variable, given in terms of a membership function only, seems too rough. For example, in decision making problems, particularly in the case of medical diagnosis, sales analysis, new product marketing, financial services, etc. there is a fair chance of the existence of a non-null hesitation part at each moment of evaluation of an unknown object.

Intuitionistic fuzzy sets (Atanassov [1,2]) can be viewed in this context as a proper tool for representing hesitancy concerning both the membership and

non-membership of an element to a set. To be more precise, a basic assumption of fuzzy set theory that if we specify the degree of membership of an element in a fuzzy set as a real number from $[0, 1]$, say a , then the degree of its non-membership is automatically determined as $1 - a$, need not hold for intuitionistic fuzzy sets. In intuitionistic fuzzy set theory it is assumed that the non-membership should not be more than $1 - a$. The difference let us express a lack of knowledge (hesitancy concerning both the membership and non-membership of an element to a set). In this way we can better model imperfect information – for example, we can express the fact that the temperature of a patient changes, and other symptoms are not quite clear.

In this article we will present intuitionistic fuzzy sets as a tool for reasoning in the presence of imperfect facts and imprecise knowledge. An example of medical diagnosis will be presented assuming there is a database, i.e. description of a set of symptoms S , and a set of diagnoses D . We will describe a state of a patient knowing results of his/her medical tests. Description of the problem uses the notion of an intuitionistic fuzzy set. The proposed method of diagnosis involves intuitionistic fuzzy distances as introduced in (Szmidt and Kacprzyk [7,8]). Advantages of such an approach are pointed out in comparison with the method presented in (De, Biswas and Roy [4]) in which the max-min-max composition rule was applied.

The material in the article is organized as follows. In Section 2 we briefly overview intuitionistic fuzzy sets. Section 3 presents De, Biswas and Roy's [4] approach for medical diagnosis via intuitionistic fuzzy relations, or – to be more precise – via the max-min-max composition. In Section 4 we propose a new approach for solving the same problem – we also use intuitionistic fuzzy sets but the final diagnosis is pointed out by the smallest distance between symptoms characteristic for a patient and symptoms describing considered illnesses. Finally, we finish with some conclusions in Section 5.

2 Brief Introduction to Intuitionistic Fuzzy Sets

As opposed to a fuzzy set (Zadeh [9]) in $X = x$, given by

$$A' = \{ \langle x, \mu_{A'}(x) \rangle \mid x \in X \} \quad (1)$$

where $\mu_{A'} : X \rightarrow [0, 1]$ is the membership function of the fuzzy set $A' : \mu_{A'}(x) \in [0, 1]$; is the membership of $x \in X$ in A' , an intuitionistic fuzzy set (Atanassov [2]) $A \in X$ is given by

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \} \quad (2)$$

where: $\mu_A : X \rightarrow [0, 1]$ and $\nu_A : X \rightarrow [0, 1]$ such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \quad (3)$$

and $\mu_A(x), \nu_A(x) \in [0, 1]$ denote the degree of membership and non-membership of $x \in A$, respectively.

Obviously, each fuzzy set corresponds to the following intuitionistic fuzzy set

$$A = \{ \langle x, \mu_{A'}(x), 1 - \mu_{A'}(x) \rangle \mid x \in X \} \quad (4)$$

For each intuitionistic fuzzy set in X , we will call

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \quad (5)$$

a *hesitation margin* (or *intuitionistic fuzzy index*) of $x \in A$, and it is a hesitation degree of whether x belongs to A or not [cf. Atanassov [2]]. It is obvious that $0 \leq \pi_A(x) \leq 1$, for each $x \in X$.

On the other hand, for each fuzzy set A' in X , we evidently have

$$\pi_{A'}(x) = 1 - \mu_{A'}(x) - [1 - \mu_{A'}(x)] = 0, \text{ for each } x \in X \quad (6)$$

Therefore, we can state that if we want to fully describe an intuitionistic fuzzy set, we must use any two functions from the triplet:

- membership function,
- non-membership function, and
- hesitation margin.

In other words, the application of intuitionistic fuzzy sets instead of fuzzy sets means the introduction of another degree of freedom into a set description (i.e. in addition to μ_A we also have ν_A or π_A).

3 An Intuitionistic Fuzzy Sets Approach to Medical Diagnosis Due to De, Biswas and Roy [4]

By following the reasoning of De, Biswas and Roy [4] (which is an extension of Sanchez's approach [5,6]), we will now consecutively recall their approach to medical diagnosis via intuitionistic fuzzy sets, or to be more precise – via intuitionistic fuzzy relations that in effect boils down to applying the max-min-max composition [3].

The approach presented by De, Biswas and Roy [4] involves the following three steps:

- determination of symptoms,
- formulation of medical knowledge based on intuitionistic fuzzy relations, and
- determination of diagnosis on the basis of composition of intuitionistic fuzzy relations.

A set of n patients is considered. For each patient $p_i, i = 1, \dots, n$, a set of symptoms S is given. As a result, an intuitionistic fuzzy relation Q is given from the set of patients to the set of symptoms S .

Next, it is assumed that another intuitionistic fuzzy relation R is given – from a set of symptoms S to the set of diagnoses D . The composition T of intuitionistic fuzzy relations R and Q describes the state of a patient given in terms of a membership function, $\mu_T(p_i, d_k)$, and a non-membership function, $\nu_T(p_i, d_k)$, for each patient p_i and each diagnosis d_k .

The functions are calculated in the following way [4]:

$$\mu_T(p_i, d_k) = \bigvee_{s \in S} [\mu_Q(p_i, s) \wedge \mu_R(s, d_k)] \tag{7}$$

and

$$\nu_T(p_i, d_k) = \bigwedge_{s \in S} [\nu_Q(p_i, s) \vee \nu_R(s, d_k)] \tag{8}$$

where $\bigvee = \max$ and $\bigwedge = \min$.

Example 1. [4] Let there be four patients: Al, Bob, Joe and Ted, i.e. $P = \{Al, Bob, Joe, Ted\}$. The set of symptoms considered is $S = \{temperature, headache, stomach\ pain, cough, chest\ pain\}$. The intuitionistic fuzzy relation $Q(P \rightarrow S)$ is given in Table 1.

Table 1.

Q	<i>Temperature</i>	<i>Headache</i>	<i>Stomach pain</i>	<i>Cough</i>	<i>Chest pain</i>
<i>Al</i>	(0.8, 0.1)	(0.6, 0.1)	(0.2, 0.8)	(0.6, 0.1)	(0.1, 0.6)
<i>Bob</i>	(0.0, 0.8)	(0.4, 0.4)	(0.6, 0.1)	(0.1, 0.7)	(0.1, 0.8)
<i>Joe</i>	(0.8, 0.1)	(0.8, 0.1)	(0.0, 0.6)	(0.2, 0.7)	(0.0, 0.5)
<i>Ted</i>	(0.6, 0.1)	(0.5, 0.4)	(0.3, 0.4)	(0.7, 0.2)	(0.3, 0.4)

Let the set of diagnoses be $D = \{Viral\ fever, Malaria, Typhoid, Stomach\ problem, Chest\ problem\}$. The intuitionistic fuzzy relation $R(S \rightarrow D)$ is given in Table 2.

Therefore, the composition T (7)–(8) is given in Table 3.

But as the max-min-max composition was used when looking for T , "dominating" symptoms were in fact only taken into account. So, in the next step an improved version of R is calculated for which the following holds [4]:

- $S_R = \mu_R - \nu_R \pi_R$ is the greatest, and
- equations (7)–(8) are retained.

Effects of the presented improvements [4] are given in Table 4.

□

Table 2.

<i>R</i>	<i>Viral fever</i>	<i>Malaria</i>	<i>Typhoid</i>	<i>Stomach problem</i>	<i>Chest problem</i>
<i>Temperature</i>	(0.4, 0.0)	(0.7, 0.0)	(0.3, 0.3)	(0.1, 0.7)	(0.1, 0.8)
<i>Headache</i>	(0.3, 0.5)	(0.2, 0.6)	(0.6, 0.1)	(0.2, 0.4)	(0.0, 0.8)
<i>Stomach pain</i>	(0.1, 0.7)	(0.0, 0.9)	(0.2, 0.7)	(0.8, 0.0)	(0.2, 0.8)
<i>Cough</i>	(0.4, 0.3)	(0.7, 0.0)	(0.2, 0.6)	(0.2, 0.7)	(0.2, 0.8)
<i>Chest pain</i>	(0.1, 0.7)	(0.1, 0.8)	(0.1, 0.9)	(0.2, 0.7)	(0.8, 0.1)

Table 3.

<i>T</i>	<i>Viral fever</i>	<i>Malaria</i>	<i>Typhoid</i>	<i>Stomach problem</i>	<i>Chest problem</i>
<i>Al</i>	(0.4, 0.1)	(0.7, 0.1)	(0.6, 0.1)	(0.2, 0.4)	(0.2, 0.6)
<i>Bob</i>	(0.3, 0.5)	(0.2, 0.6)	(0.4, 0.4)	(0.6, 0.1)	(0.1, 0.7)
<i>Joe</i>	(0.4, 0.1)	(0.7, 0.1)	(0.6, 0.1)	(0.2, 0.4)	(0.2, 0.5)
<i>Ted</i>	(0.4, 0.1)	(0.7, 0.1)	(0.5, 0.3)	(0.3, 0.4)	(0.3, 0.4)

Table 4.

	<i>Viral fever</i>	<i>Malaria</i>	<i>Typhoid</i>	<i>Stomach problem</i>	<i>Chest problem</i>
<i>Al</i>	0.35	0.68	0.57	0.04	0.08
<i>Bob</i>	0.20	0.08	0.32	0.57	0.04
<i>Joe</i>	0.35	0.68	0.57	0.04	0.05
<i>Ted</i>	0.32	0.68	0.44	0.18	0.18

It seems that the approach proposed in [4] has some drawbacks. First, the max-min-max rule alone (Table 3) does not give a solution. To obtain a solution, the authors [4] propose some changes in medical knowledge $R(S \rightarrow D)$. But it is difficult to justify the proposed changes as medical knowledge $R(S \rightarrow D)$ is (at least, should be without a doubt) based on many cases, and knowledge of physicians, so it is difficult to understand sudden, arbitral changes in it.

Next, the type of changes: $S_R = \mu_R - \nu_R \pi_R$ means that the membership function describing relation R (medical knowledge) is weakened. But, in the idea of intuitionistic fuzzy sets there is nowhere assumed that the membership function can decrease because of the hesitation margin or the non-membership function. The hesitation margin (or part of it) can be split between the membership and non-membership functions (so, in fact, it can be added to, not subtracted from the membership function). Summing up, the proposed improvements, although leading to some solutions, are difficult to understand because of arbitral (both from practical (doctors' knowledge) and theoretical (theory of intuitionistic fuzzy sets)) points of view.

4 Medical Diagnosis via Distances for Intuitionistic Fuzzy Sets

To solve the same problem as in [4], but without manipulations in medical knowledge base, and with taking into account all the symptoms characteristic for each patient, we propose a new method based on calculating distances between diagnoses and patient tests.

As in [4], to make a proper diagnosis D for a patient with given values of tested symptoms S , a medical knowledge base is necessary. In our case a knowledge base is formulated in terms of intuitionistic fuzzy sets.

To compare the approach proposed in this article with the method of De, Biswas and Roy [4], and described shortly in Section 3, we consider just the same data. Let the set of diagnoses be $D = \{Viral\ fever, Malaria, Typhoid, Stomach\ problem, Chest\ problem\}$. The considered set of symptoms is $S = \{temperature, headache, stomach\ pain, cough, chest-pain\}$.

The data are given in Table 5 – each symptom is described by three numbers: membership μ , non-membership ν , hesitation margin π . For example, for malaria: the temperature is high ($\mu = 0.7, \nu = 0, \pi = 0.3$), whereas for the chest problem: temperature is low ($\mu = 0.1, \nu = 0.8, \pi = 0.1$). In fact data in Table 2 and Table 5 are exactly the same (due to (5)) but by involving in an explicit way the hesitation margin too, we want to stress that the values of all three parameters are necessary in our approach.

The considered set of patients is $P = \{Al, Bob, Joe, Ted\}$. The symptoms characteristic for the patients are given in Table 6 – as before, we need all three parameters (μ, ν, π) describing each symptom but the data are the same (due to (5)) as in Table 1.

Our task is to make a proper diagnosis for each patient $p_i, i = 1, \dots, 4$. To fulfill the task we propose to calculate for each patient p_i a distance of his symptoms (Table 6) from a set of symptoms $s_j, j = 1, \dots, 5$ characteristic for

Table 5.

	<i>Viral fever</i>	<i>Malaria</i>	<i>Typhoid</i>	<i>Stomach problem</i>	<i>Chest problem</i>
<i>Temperature</i>	(0.4, 0.0, 0.6)	(0.7, 0.0, 0.3)	(0.3, 0.3, 0.4)	(0.1, 0.7, 0.2)	(0.1, 0.8, 0.1)
<i>Headache</i>	(0.3, 0.5, 0.2)	(0.2, 0.6, 0.2)	(0.6, 0.1, 0.3)	(0.2, 0.4, 0.4)	(0.0, 0.8, 0.2)
<i>Stomach pain</i>	(0.1, 0.7, 0.2)	(0.0, 0.9, 0.1)	(0.2, 0.7, 0.1)	(0.8, 0.0, 0.2)	(0.2, 0.8, 0.0)
<i>Cough</i>	(0.4, 0.3, 0.3)	(0.7, 0.0, 0.3)	(0.2, 0.6, 0.2)	(0.2, 0.7, 0.1)	(0.2, 0.8, 0.0)
<i>Chest pain</i>	(0.1, 0.7, 0.2)	(0.1, 0.8, 0.1)	(0.1, 0.9, 0.0)	(0.2, 0.7, 0.1)	(0.8, 0.1, 0.1)

Table 6.

	<i>Temperature</i>	<i>Headache</i>	<i>Stomach pain</i>	<i>Cough</i>	<i>Chest pain</i>
<i>Al</i>	(0.8, 0.1, 0.1)	(0.6, 0.1, 0.3)	(0.2, 0.8, 0.0)	(0.6, 0.1, 0.3)	(0.1, 0.6, 0.3)
<i>Bob</i>	(0.0, 0.8, 0.2)	(0.4, 0.4, 0.2)	(0.6, 0.1, 0.3)	(0.1, 0.7, 0.2)	(0.1, 0.8, 0.1)
<i>Joe</i>	(0.8, 0.1, 0.1)	(0.8, 0.1, 0.1)	(0.0, 0.6, 0.4)	(0.2, 0.7, 0.1)	(0.0, 0.5, 0.5)
<i>Ted</i>	(0.6, 0.1, 0.3)	(0.5, 0.4, 0.1)	(0.3, 0.4, 0.3)	(0.7, 0.2, 0.1)	(0.3, 0.4, 0.3)

each diagnosis d_k , $k = 1, \dots, 5$ (Table 5). The lowest obtained distance points out a proper diagnosis.

In (Szmidski and Kacprzyk [7,8]) we proved that the only proper way of calculating the most widely used distances for intuitionistic fuzzy sets is to take into account all three parameters: the membership function, the non-membership function, and the hesitation margin. To be more precise, the normalised Hamming distance for all the symptoms of the i -th patient from the k -th diagnosis is equal to

$$\begin{aligned}
 l(s(p_i), d_k) = & \frac{1}{10} \sum_{j=1}^5 (|\mu_j(p_i) - \mu_j(d_k)| + |\nu_j(p_i) - \nu_j(d_k)| + \\
 & + |\pi_j(p_i) - \pi_j(d_k)|)
 \end{aligned}
 \tag{9}$$

The distances (9) for each patient from the considered set of possible diagnoses are given in Table 7. The lowest distance points out a proper diagnosis: Al suffers from malaria, Bob from stomach problem, Joe from typhoid, whereas Ted from fever.

We obtained the same results, i.e. the same quality diagnosis for each patient when looking for the solution while applying the normalized Euclidean distance [cf. Szmidski and Kacprzyk [7,8]]:

$$\begin{aligned}
 q(s(p_i), d_k) = & \left(\frac{1}{10} \sum_{j=1}^{10} (\mu_j(p_i) - \mu_j(d_k))^2 + (\nu_j(p_i) - \nu_j(d_k))^2 + \right. \\
 & \left. + (\pi_j(p_i) - \pi_j(d_k))^2 \right)^{\frac{1}{2}}
 \end{aligned}
 \tag{10}$$

Table 7.

	<i>Viral fever</i>	<i>Malaria</i>	<i>Typhoid</i>	<i>Stomach problem</i>	<i>Chest problem</i>
<i>Al</i>	0.28	0.24	0.28	0.54	0.56
<i>Bob</i>	0.40	0.50	0.31	0.14	0.42
<i>Joe</i>	0.38	0.44	0.32	0.50	0.55
<i>Ted</i>	0.28	0.30	0.38	0.44	0.54

The results are given in Table 8 – the lowest distance for each patient p_i from possible diagnosis D points out a solution. As before, Al suffers from malaria, Bob from stomach problem, Joe from typhoid, whereas Ted from fever.

Table 8.

	<i>Viral fever</i>	<i>Malaria</i>	<i>Typhoid</i>	<i>Stomach problem</i>	<i>Chest problem</i>
<i>Al</i>	0.29	0.25	0.32	0.53	0.58
<i>Bob</i>	0.43	0.56	0.33	0.14	0.46
<i>Joe</i>	0.36	0.41	0.32	0.52	0.57
<i>Ted</i>	0.25	0.29	0.35	0.43	0.50

5 Conclusions

By employing intuitionistic fuzzy sets in databases we can express a hesitation concerning examined objects. The method proposed in this article, performing diagnosis on the basis of the calculation of distances from a considered case to all considered illnesses, takes into account values of all symptoms. As a result, our approach makes it possible to introduce weights for all symptoms (for some illnesses some symptoms can be more important). Such an approach is impossible in the method described in (De, Biswas and Roy [4]) because the max-min-max rule “neglects” in fact most values except for extreme ones.

References

- [1] Atanassov K. (1986) Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 20 (1986) 87-96.
- [2] Atanassov K. (1999) *Intuitionistic Fuzzy Sets: Theory and Applications*. Physica-Verlag.
- [3] Biswas R. (1997) Intuitionistic fuzzy relations. *Bull. Sous. Ens. Flous.Appl. (BUSEFAL)*, Vol. 70, pp. 22-29.
- [4] De S.K., Biswas R. and Roy A.R. (2001) An application of intuitionistic fuzzy sets in medical diagnosis. *Fuzzy Sets and Systems*, Vol. 117, No.2, pp. 209 – 213.

- [5] Sanchez E. (1977) Solutions in composite fuzzy relation equation. Application to medical diagnosis Brouwerian Logic. In: M.M. Gupta, G.N. Saridis, B.R. Gaines (Eds.), Fuzzy Automata and Decision Process, Elsevier, North-Holland.
- [6] Sanchez E. (1976) Resolution of composition fuzzy relation equations. Inform. Control, Vol. 30, pp. 38-48.
- [7] Szmidt E. and Kacprzyk J. (1997) On measuring distances between intuitionistic fuzzy sets, Notes on IFS, Vol.3, No.4, pp.1 - 13.
- [8] Szmidt E. and J. Kacprzyk J. (2000) Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, Vol.114, No.3, pp.505 - 518.
- [9] Zadeh L.A. (1965) Fuzzy sets. Information and Control 8, pp. 338 - 353.