**Eighth Int. Conf. on IFSs, Varna, 20-21 June 2004**
**NIFS Vol. 10 (2004), 4, 61-69**

# Medical Diagnostic Reasoning Using a Similarity Measure for Intuitionistic Fuzzy Sets

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 similarity measure for intuitionistic fuzzy sets (cf. Atanassov [1, 2]) and show its usefulness in medical diagnostic reasoning. We point out advantages of this new concept over the method proposed previously (cf. Szmidt and Kacprzyk [23]) where distances were used instead of the proposed similarity measure.
**Keywords:** intuitionistic fuzzy sets, similarity measure, medical disagnostic reasoning

## 1 Introduction

Intuitionistic fuzzy sets (Atanassov [1], [2]), due to an additional degree of freedom in comparison with fuzzy sets (Zadeh [26]), can be viewed as their generalization. The additional degree of freedom let us better model imperfect information which is omnipresent in any conscious decision making. We will present here intuitionistic fuzzy sets as a tool for a more human consistent reasoning under imperfectly defined facts and imprecise knowledge.

An example of medical diagnosis will be presented assuming that there is a database, i.e. a 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. The problem description uses the concept of an intuitionistic fuzzy set that makes it possible to render two important facts. First, values of symptoms change for each patient as, e.g., temperature goes up and down, pain increases and decreases, etc. Second, in a medical database describing illnesses for different patients it should be taken into account that for different patients suffering from the same illness, values of the same symptom can be different.

The proposed method of diagnosis involves a new measure of similarity for intuitionistic fuzzy sets. For each patient the similarity measures for his particular set of symptoms and a set of symptoms that are characteristic for each diagnosis are calculated. The lowest obtained value points out a proper diagnosis.

The material in the article is organized as follows. In Section 2 we briefly overview intuitionistic fuzzy sets. In Section 3 we propose the new measure of similarity for intuitionistic fuzzy sets. In Section 4 we use the proposed similarity measure to single out the diagnosis for the considered patients. We compare the obtained solution with the final diagnosis pointed out by looking for the smallest distance between symptoms characteristic for a patient and symptoms decsribing considered illnesses (see Szmidt and Kacprzyk [23]]). Finally, we finish with some conclusions in Section 5.

## 2 Brief introduction to intuitionistic fuzzy sets

As opposed to a fuzzy set in $X$(Zadeh [26]) , given by

$$A' = \{< x, \mu_{A'}(x) > | x \in X\} \tag{1}$$

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

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\} \tag{2}$$

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

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{3}$$

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

Obviously, each fuzzy set may be represented by the following intuitionistic fuzzy set

$$A = \{< x, \mu_{A'}(x), 1 - \mu_{A'}(x) > | x \in X\} \tag{4}$$

For each intuitionistic fuzzy set in $X$, we will call

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{5}$$

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

In our further considerations we will use the notion of the complement elements, which definition is a simple consequence of a complement set $A^C$

$$A^C = \{< x, \nu_A(x), \mu_A(x) > | x \in X\} \tag{6}$$

The application of intuitionistic fuzzy sets instead of fuzzy sets means the introduction of another degree of freedom into a set description. Such a generalization of fuzzy sets gives us an additional possibility to represent imperfect knowledge what leads to describing many real problems in a more adequate way.

Applications of intiutionistic fuzzy sets to group decision making, negotiations and other situations are presented in Szmidt [9], Szmidt and Kacprzyk [11], [12], [14], [22], [23].

The formula (7) can also be stated as

$$Sim(X,F) = \frac{l_{IFS}(X,F)}{l_{IFS}(X,F^C)} = \frac{l_{IFS}(X^C,F^C)}{l_{IFS}(X,F^C)} = \\ = \frac{l_{IFS}(X,F)}{l_{IFS}(X^C,F)} = \frac{l_{IFS}(X^C,F^C)}{l_{IFS}(X^C,F)} \tag{9}$$

It is worth noticing that

- $Sim(X,F) = 0$ means the identity of $X$ and $F$.
- $Sim(X,F) = 1$ means that $X$ is to the same extent similar to $F$ and $F^C$ (i.e., values bigger than 1 mean in fact a closer similarity of $X$ and $F^C$ to $X$ and $F$).
- When $X = F^C$ (or $X^C = F$), i.e. $l_{IFS}(X,F^C)$=$l_{IFS}(X^C,F)$= 0 means the complete dissimilarity of $X$ and $F$ (or in other words, the identity of $X$ and $F^C$), and then $Sim(X,F) \to \infty$.
- When $X = F = F^C$ means the highest possible entropy (see [20]) for both elements $F$ and $X$ i.e. the highest "fuzziness" – not too constructive a case when looking for compatibility (both similarity and dissimilarity).

In other words, when applying measure (7) to analyse the similarity of two objects, one should be interested in the values $0 \leq Sim(X,F) < 1$.

The proposed measure (7) was constructed for selecting objects which are more similar than dissimilar [and well-defined in the sense of possessing (or not) attributes we are interested in]. For further discussion concerning the proposed measure (including its name, range of possible values, connections with the Jaccard's index, and the literature) we refere an interested reader to Szmidt and Kacprzyk [24].

Now we will show that a measure of similarity defined as mentioned above, (7), between $X(\mu_X, \nu_X, \pi_X)$ and $F(\mu_F, \nu_F, \pi_F)$ is more powerful then a simple distance between them. Medical diagnostic reasoning will help us to show the fact.

# 4 Medical diagnostic reasoning

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 involves elements of intuitionistic fuzzy sets.

We consider the same data as those of De, Biswas and Roy [3]. 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 1 – each symptom is described by three numbers: membership $\mu$, non-membership $\nu$, hesition margin $\pi$. For example, for malaria, we have: the temperature is high ($\mu = 0.7$, $\nu = 0$, $\pi = 0.3$), whereas for the chest problem, we have: temperature is low ($\mu = 0.1$, $\nu = 0.8$, $\pi = 0.1$).

The set of patients considered is $P$ = {*Al, Bob, Joe, Ted*}. The symptoms characteristic for the patients are given in Table 2 – as before, we need all three parameters ($\mu$,$\nu$,$\pi$)

Table 1: Symptoms characteristic for the diagnoses considered

| | *Viral fever* | *Malaria* | *Typhoid* | *Stomach problem* | *Chest problem* |
|---|---|---|---|---|---|
| *Temperature* | (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) |
| *Headache* | (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) |
| *Stomach pain* | (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) |
| *Cough* | (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) |
| *Chest pain* | (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 2: Symptoms characteristic for the patients considered

| | *Temperature* | *Headache* | *Stomach pain* | *Cough* | *Chest pain* |
|---|---|---|---|---|---|
| *Al* | (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) |
| *Bob* | (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) |
| *Joe* | (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) |
| *Ted* | (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) |

to describe each symptom (see Szmidt and Kacprzyk [23]). Our task is to derive a proper diagnosis for each patient $p_i$, $i = 1, \ldots, 4$. In our previous article (Szmidt and Kacprzyk [23]) we proposed to solve the problem in the following way

- to calculate for each patient $p_i$ a distance (we used the normalised Hamming distance) of his symptoms (Table 2) from a set of symptoms $s_j$, $j = 1, \ldots, 5$ characteristic for each diagnosis $d_k$, $k = 1, \ldots, 5$ (Table 1),

- to single out the lowest obtained distance which points out a proper diagnosis.

The normalised Hamming distance for all the symptoms of the $i$-th patient from the $k$-th diagnosis is equal to

$$l(s(p_i), d_k) = \frac{1}{10}\sum_{j=1}^{5}(|\mu_j(p_i) - \mu_j(d_k)| + |\nu_j(p_i) - \nu_j(d_k)| + |\pi_j(p_i) - \pi_j(d_k)|) \quad (10)$$

The distances (10) for each patient from the considered set of possible diagnoses are given in Table 3. The lowest distance points out a proper diagnosis: Al suffers from malaria, Bob from stomach problem, Joe from typhoid, whereas Ted from fever.

Now we will solve the same task - deriving a proper diagnosis for each patient $p_i$, $i = 1, \ldots, 4$ using the proposed similarity measure (7) To do so, we propose

- to calculate for each patient $p_i$ a similarity measure (7) between his symptoms (Table 2) and symptoms $s_j$, $j = 1, \ldots, 5$ characteristic for each diagnosis $d_k$, $k = 1, \ldots, 5$ (Table 1),

- to single out the lowest value from the obtained similarity measures which points out a proper diagnosis.

Table 3: The normalized Hamming distances for each patient from the considered set of possible diagnoses

| | *Viral fever* | *Malaria* | *Typhoid* | *Stomach problem* | *Chest problem* |
|---|---|---|---|---|---|
| *Al* | 0.28 | 0.24 | 0.28 | 0.54 | 0.56 |
| *Bob* | 0.40 | 0.50 | 0.31 | 0.14 | 0.42 |
| *Joe* | 0.38 | 0.44 | 0.32 | 0.50 | 0.55 |
| *Ted* | 0.28 | 0.30 | 0.38 | 0.44 | 0.54 |

From Definition 1, similarity measure (7) for $p_i$ patient - between his/her symptoms and the symptoms characterisctic for diagnosis $d_k$, is

$$\begin{aligned} Sim(s(p_i), d_k) &= \frac{1}{5}\sum_{j=1}^{5}[(|\mu_j(p_i) - \mu_j(d_k)| + |\nu_j(p_i) - \nu_j(d_k)| + \\ &+ |\pi_j(p_i) - \pi_j(d_k)|)]/[(|\mu_j(p_i) - \nu_j(d_k)| + \\ &+ |\nu_j(p_i) - \mu_j(d_k)| + |\pi_j(p_i) - \pi_j(d_k)|)] \end{aligned} \quad (11)$$

For example, for Al, similarity measures for all his symptoms and respective symptoms of a choosen diagnosis - Chest problem $Ch$ are

- for temperature $T$

$$\begin{aligned} Sim_T(Al, Ch) &= [|0.8 - 0.1| + |0.1 - 0.8| + |0.1 - 0.1|]/ \\ &/ [|0.8 - 0.8| + |0.1 - 0.1| + |0.1 - 0.1|] \to \infty \end{aligned} \quad (12)$$

- for headache $H$

$$\begin{aligned} Sim_H(Al, Ch) &= [|0.6 - 0| + |0.1 - 0.8| + |0.3 - 0.2|]/ \\ &/ [|0.6 - 0.8| + |0.1 - 0| + |0.3 - 0.2|] = 3.5 \end{aligned} \quad (13)$$

- for stomach pain $SP$

$$\begin{aligned} Sim_{SP}(Al, Ch) &= [|0.2 - 0.2| + |0.8 - 0.8| + |0 - 0|]/ \\ &/ [|0.2 - 0.8| + |0.8 - 0.2| + |0 - 0|] = 0 \end{aligned} \quad (14)$$

- for cough $C$

$$\begin{aligned} Sim_C(Al, Ch) &= [|0.6 - 0.2| + |0.1 - 0.8| + |0.3 - 0|]/ \\ &/ [|0.6 - 0.8| + |0.1 - 0.2| + |0.3 - 0|] = 2.33 \end{aligned} \quad (15)$$

- for chest pain $ChP$

$$\begin{aligned} Sim_{ChP}(Al, Ch) &= [|0.1 - 0.8| + |0.6 - 0.1| + |0.3 - 0.1|]/ \\ &/ [|0.1 - 0.1| + |0.6 - 0.8| + |0.3 - 0.1|] = 3.5 \end{aligned} \quad (16)$$

Table 4: Similarities of symptoms for each patient to the considered set of possible diagnoses

| $R$ | $Viral$ $fever$ | $Malaria$ | $Typhoid$ | $Stomach$ $problem$ | $Chest$ $problem$ |
|---|---|---|---|---|---|
| $Al$ | 0.75 | 1.19 | 1.31 | 3.27 | $\infty$ |
| $Bob$ | 2.1 | 3.73 | 1.1 | 0.35 | $\infty$ |
| $Joe$ | 0.87 | 1.52 | 0.46 | 2.61 | $\infty$ |
| $Ted$ | 0.95 | 0.77 | 1.67 | $\infty$ | 2.56 |

Similarity measure $Sim(Al, Ch)$ taking into account all his symptoms (i.e., (12)-(16)) is

$$Sim(Al, Ch) = \frac{1}{5}(\infty + 3.5 + 0 + 2.33 + 3.5) = \infty$$

what means that at least one of Al's symptoms is quaite opposite as specified for $ChestProblem$. It is an important clue which does not occur at all when we consider just distances between symptoms instead of the proposed similarity measure.

All the results for the considered patients are in Table 4. The obtained results (Table 4) are different as they were when we considered instead of similatity measure (7) just the distances (Table 3). As previously, Bob suffers from stomach problems, Joe from typhoid, but Al from fever (not from malaria), and Ted suffers from malaria (not from fever). These differences are because the similarity measure (10) can be small but at the same time the distance between the symptom of the patient and the complementary symptom characteristic for the examined illness can be smaller (even equal to 0 as it was for (12)). See also Szmidt and Kacprzyk [24].

# 5 Conclusions

By employing intuitionistic fuzzy sets in databases we can express a hesitation concerning objects under consideration. The method proposed in this article, performing diagnosis on the basis of the calculation of a new similarity measure for intuitionistic fuzzy sets, makes it possible to avoid drawing conclusions about strong similarity between intuitionistic fuzzy sets on the basis of the small distances between these sets.

It is also worth stressing that the proposed method takes into account the values of all symptoms. As a result, our approach makes it possible to introduce weights for all symptoms (for instance, for some illnesses some symptoms can be more important). This weighting scheme is impossible in the method proposed by De, Biswas and Roy [3] because the max-min-max rule “neglects” in fact most values except for the 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] 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.

[4] Cross V. and Sudkamp T. (2002) Similarity and Compatibility in Fuzzy Set Theory. Physica-Verlag.

[5] Dengfeng L. and Chuntian C. (2002) New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. Pattern Recognition Letters, Vol. 23, pp. 221-225.

[6] Dubois D. and Prade H. (1982) New results about properties and semantics of fuzzy sets operators. In P.Wang, S. Chang (eds.) Fuzzy Sets and Possibility Theory Recent Developments, pages 3-13, Pergamon Press.

[7] 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.

[8] Sanchez E. (1976) Resolution of composition fuzzy relation equations. Inform. Control, Vol. 30, pp. 38–48.

[9] Szmidt E. (2000): Applications of Intuitionistic Fuzzy sets in Decision Making. (D.Sc. dissertation) Techn. Univ., Sofia, 2000.

[10] Szmidt E. and Baldwin J. (2003) New Similarity Measure for Intuitionistic Fuzzy Set Theory and Mass Assignment Theory. Proc. of the 7-th Int. Conf. on IFSs. Sofia, 2003.

[11] Szmidt E. and Kacprzyk J. (1996a) Intuitionistic fuzzy sets in group decision making, *Notes on IFS*, 2, pp. 15–32.

[12] Szmidt E. and Kacprzyk J. (1996b) Remarks on some applications of intuitionistic fuzzy sets in decision making, *Notes on IFS*, Vol.2, No.3, pp. 22–31.

[13] Szmidt E. and Kacprzyk J. (1997) On measuring distances between intuitionistic fuzzy sets, Notes on IFS, Vol.3, No.4, pp.1 - 13.

[14] Szmidt E. and Kacprzyk J. (1998a) Group Decision Making under Intuitionistic Fuzzy Preference Relations. 7th Int. Conf. IPMU'98 (Paris, La Sorbonne, July 6-10), pp. 172-178.

[15] Szmidt E. and Kacprzyk J. (1998b) Applications of Intuitionistic Fuzzy Sets in Decision Making. 8th Congr. EUSFLAT'99 (Pamplona, September 8-10), pp. 150–158.

[16] Szmidt E. and J. Kacprzyk J. (2000) Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, Vol.114, No.3, pp.505 - 518.

[17] Szmidt E. and Kacprzyk J. (2000) On Measures on Consensus Under Intuitionistic Fuzzy Relations. Proc. of Eight Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems IPMU 2000, Madrid, July 3-7, 2000, pp. 1454-1461.

[18] Szmidt E. and Kacprzyk J. (2001) Distance from consensus under Intuitionistic Fuzzy Preferences. In: Proc. on Eurofuse Workshop on Preference Modelling and Applications, Granada, Spain, April 25-27, 2001, pp. 73-77.

[19] Szmidt E. and Kacprzyk J. (2001) Distance from Consensus Under Intuitionistic Fuzzy Preferences. Proc. EUROFUSE Workshop on Preference Modelling and Applications, Granada, April 25–27, pp. 73–78.

[20] Szmidt E., Kacprzyk J. (2001) Entropy for intuitionistic fuzzy sets. Fuzzy Sets and Systems, vol. 118, No. 3, pp. 467–477.

[21] Szmidt E. and Kacprzyk J. (2001) Analysis of Consensus under Intuitionistic Fuzzy Preferences. Proc. Int. Conf. in Fuzzy Logic and Technology. De Montfort Univ. Leicester, UK, September 5-7, pp. 79–82.

[22] Szmidt E. and Kacprzyk J. (2002) Analysis of Agreement in a Group of Experts via Distances Between Intuitionistic Fuzzy Preferences. Proc. 9th Int. Conf. IPMU 2002, Annecy, France, July 1–5, pp. 1859–1865.

[23] Szmidt E. and Kacprzyk J. (2002) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L.Jain, J. Kacprzyk (Eds.): Recent Advances in Intelligent Paradigms and and Applications. Springer-Verlag, pp. 57–70.

[24] Szmidt E. and Kacprzyk J. Similarity of Intuitionistic Fuzzy Sets and Jaccard Coefficient. (accepted for IPMU'04)

[25] Tversky A. (1977) Features of similarity. Psychol. Rev. Vol. 84, pp. 327–352.

[26] Zadeh L.A. (1965) Fuzzy sets. Information and Control 8, pp. 338 - 353.

[27] Zadeh L.A. and Kacprzyk J. (1999) (Eds.) Computing with words in information/intelligent systems. 1. Foundations. 2. Applications. Physica-Verlag. Heidelberg, New York.