Analyzing Consequences of Diabetes Mellitus Using Intuitionistic Fuzzy Set

Sujit Das¹, Samarjit Kar²
¹Dept. of Comp. Sc. & Enng., Dr. B.C. Roy Engg. College, Durgapur, India
²Dept. of Mathematics, National Institute of Technology Durgapur, India
sujit_cse@yahoo.com, kar_s_k@yahoo.com

Abstract - Diabetes is a chronic illness which may lead to several major consequences in future like Vision Loss, Heart Failure, Stroke, Nerve Damage and Foot Ulcer. There is no exact cure of diabetes, as a result prevention and self-education is the only way to deal with it. In this article we propose to analyze the consequences of diabetes mellitus using intuitionistic fuzzy set. Using hamming distance and similarity measurements this article has investigated the consequences which may appear for a patient. We have proposed an approach using entropy to analyze the result and ensured that the results are correct. Also using entropy measurement it is scrutinized that some consequences have almost no chances to appear.

Keywords - Diabetes mellitus, entropy, hamming distance, intuitionistic fuzzy set, similarity measurements

I. INTRODUCTION

Atanassov ([1], [2], [3]) introduced the concept of intuitionistic fuzzy set which is a generalization of fuzzy set [4]. Over the last decades, intuitionistic fuzzy set theory has been applied to many different fields, such as decision making [5], logic programming [6], topology [7], medical diagnosis ([8], [9]) pattern recognition [10], machine learning and market prediction [11] etc. Often it is found that in many real life situations where the fuzzy linguistic variable is expressed by the membership function only is not enough to generate a feasible solution. Decision making problem like medical analysis claims a fair chance of existing the non-null hesitation part in the prospect of reasoning the imperfect facts and imprecise knowledge for better diagnostic prediction purposes.

Diabetes has become a significant public health problem and is growing rapidly worldwide. Statistics revealed that expected incident in Australia to increase from 4% to 10% by 2010 [18]. Situation is worse in India, an estimated 40 million Indians suffer from diabetes, and the problem seems to be growing at a frightening rate. It has been predicted that by 2020, the number is expected to be double and would to reach epidemic proportions, but the alarming part is that even half of the numbers of diabetics in India remain undiagnosed. Diabetes has debilitating consequences on many of the body’s vital organs if remained unchecked and uncontrolled.

Since the last decade many of the researchers and academicians are involved in diagnosing the various aspects of diabetes mellitus. Lee and Wang [18] in 2010 proposed ontology based fuzzy expert system using fuzzy linguistic concept for diabetes decision support problem. Sapna and Tamilasari [19] in 2009 have suggested Fuzzy Relational Equation (FRE) based procedure in preventing diabetic heart attack which is basically like a fuzzy binary relationship concept defined on a set. Yasini et al. [20] in 2008 focused their work on the Type 1 diabetes patients and developed a closed-loop control algorithm for blood glucose regulation. Lee et al. [21] in 2008 had concentrated on the fact that for diabetes patients, appropriate diabetes diet is very crucial. For this purpose authors have proposed an intelligent agent, called the personal food recommendation agent, based on the ontology model for diabetic food recommendation. Hans and Nehls [22] in 2000, proposed one of the earliest model to simulate human glucose metabolism using fuzzy arithmetic to represent the uncertainty involved in the physiological model. The authors have concentrated on the Type 1 diabetes only.

In this article we have worked to predict a proper diagnosis depending upon a set of related symptoms using the knowledge of intuitionistic fuzzy set parameters: membership value, non membership value and hesitation margin. After the diagnostic prediction we have ensured that those predictions are quite strong and also assured that some diagnosis will never appear to some patients using entropy estimation. Diabetes (Diabetes Mellitus) is a lifelong disease, categorized as a metabolism disorder. Consequence of Diabetes Mellitus may be Vision Loss, Heart Failure, Stroke, Nerve Damage and Foot Ulcer. Our aim is to find out the maximum chance’s consequence and ensuring that some consequences will never emerge using entropy analysis depending on membership value, non membership value and hesitation index of the various parameters normally used for checking the Diabetes Mellitus. The various attributes for Diabetes used by Pima Indians Diabetes Database (PIDD) [12] is depicted in the following chart (Table 1).

The rest of this paper is organized as follows: a brief introduction about intuitionistic fuzzy set, various distance measurements, similarity measurements and entropy is defined in Section II. Medical diagnosis of diabetes using intuitionistic fuzzy set is described in

| Table 1 |

| PIMA INDIAN DIABETES DATABASE (PIDD) |

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregnancy</td>
<td>Number of times pregnant</td>
<td>-</td>
</tr>
<tr>
<td>Glucose</td>
<td>Plasma glucose concentration in a 2-hour OGTT</td>
<td>mg/dl</td>
</tr>
<tr>
<td>DBP</td>
<td>Diastolic blood pressure</td>
<td>mmHg</td>
</tr>
<tr>
<td>TSFT</td>
<td>Triceps skin fold thickness</td>
<td>mm</td>
</tr>
<tr>
<td>INS</td>
<td>2-hour serum insulin</td>
<td>U/ml</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
<td>kg/m²</td>
</tr>
<tr>
<td>DPF</td>
<td>Diabetes pedigree function</td>
<td>-</td>
</tr>
<tr>
<td>Age</td>
<td>Age</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE 1
Section III. Section IV includes the analysis and experimental results using distances, similarity measurements and our proposed entropy investigation followed by conclusion in Section V.

II. BRIEF INTRODUCTION ON INTUITIONISTIC FUZZY SET

A. Intuitionistic Fuzzy Set

In this section we have introduced the basic characteristics of intuitionistic fuzzy set. Let $X$ be a universe of discourse, then a fuzzy set

$$A = \{ x, \mu_A(x) | x \in X \}$$

(1)
defined by Zadeh [4] is characterized by a membership function $\mu_A : X \rightarrow [0,1]$, where $\mu_A(x)$ represents the degree of membership of the element $x$ to the set $A$. Atanassov ([1], [2]) introduced a more general version of fuzzy set called intuitionistic fuzzy set, shown as follows:

$$\hat{A} = \{ x, \mu_A(x), \nu_A(x) | x \in X \}$$

(2)

which is characterized by a membership function $\mu_A : X \rightarrow [0,1]$ and a non-membership function $\nu_A : X \rightarrow [0,1]$ with the condition

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

(3)

for all $x \in X$ where the numbers $\mu_A(x)$ and $\nu_A(x)$ denotes the degree of membership and the degree of non-membership respectively of the element $x$ to the set $\hat{A}$. For each intuitionistic fuzzy set $\hat{A}$ in $X$, the hesitation margin or intuitionistic fuzzy index which expresses a lack of knowledge of whether $x$ belongs to $A$ or not is defined as

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

(4)

where $x \in X$ and it is hesitation degree of whether $x \in \hat{A}$ or $x \notin \hat{A}$. Also for each $x \in X$,

$$0 \leq \pi_A(x) \leq 1$$

(5)

Note that for an IFS $\hat{A}$, if $\mu_A(x) = 0$ then $\nu_A(x) + \pi_A(x) = 1$ and if $\mu_A(x) = 1$ then $\nu_A(x) = 0$ and $\pi_A(x) = 0$. For each fuzzy set $A$ in $X$ we can say

$$\pi_A(x) = 1 - \mu_A(x) - [1 - \mu_A(x)] = 0, \text{ for every } x \in X.$$  

(6)

B. Distances between Intuitionistic Fuzzy Sets

This experiment has used one of the most popular distances between intuitionistic fuzzy sets ([13], [14]) A, B in $X = \{ x_1, x_2, x_3, \ldots, x_n \}$ i.e. normalized hamming distance.

a) Normalized Hamming Distance

$$l_{\text{gs}}(A,B) = \frac{1}{2n} \sum_{i=1}^{n} \left( |\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)| \right)$$

(7)

with the condition

$$0 \leq l_{\text{gs}}(A,B) \leq 1.$$  

(8)

b) Normalized Euclidean Distance

$$e_{\text{gs}}(A,B) = \frac{1}{2n} \sum_{i=1}^{n} \left( (\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2 \right)$$

(9)

with the condition

$$0 \leq e_{\text{gs}}(A,B) \leq 1.$$  

(10)

C. Similarity Measurements

The similarity measures between two IFSs are used to estimate the degree of similarity between those two sets. Szmidt and Kacprzyk [15] defined a similarity measure using a distance measure which involves both similarity and dissimilarity.

The similarity measurements between $A, B \in IFS(X)$ is defined by

$$S(A,B) = \frac{1}{n} \sum_{i=1}^{n} \frac{1 - \min(\mu_i, \nu_i)}{1 + \max(\mu_i, \nu_i)}$$

(11)

where

$$\mu_i = |\mu_A(x_i) - \mu_B(x_i)| \text{ and } \nu_i = |\nu_A(x_i) - \nu_B(x_i)|.$$  

(12)

D. Entropy

Entropy measures the degree of fuzziness in a fuzzy set. Simply it describes how fuzzy is a fuzzy set. Theoretically the entropy measures the whole missing information which may be necessary to have no doubts when classifying a point to the area of consideration.

For an IFS $\hat{A} = \{ x, \mu_A(x), \nu_A(x) | x \in X \}$, Szmidt et al. [16] defined two kinds of cardinalities for $\hat{A}$ by the following way.

a) the least cardinality or min-sigma-count of $\hat{A}$ given by

$$\min \sum \text{count}(\hat{A}) = \sum_{i=1}^{n} \mu_A(x) \text{.}$$

(13)

b) the biggest cardinality or max-sigma-count of $\hat{A}$ given by

$$\max \sum \text{count}(\hat{A}) = \sum_{i=1}^{n} (\mu_A(x) + \pi_A(x)) \text{.}$$

(14)

Using these two cardinalities, Szmidt et al. [16] defined the entropy measure $E_{\text{sk}}(\hat{A})$ for $\hat{A}$ as

$$E_{\text{sk}}(\hat{A}) = \frac{1}{n} \sum_{i=1}^{n} \max \text{count}(\hat{A} \cap \hat{A}^c)$$

(15)

Where

$$\hat{A} \cap \hat{A}^c = \{ (x, \min(\mu_A(x), \nu_A(x)), \max(\nu_A(x), \mu_A(x))) \}$$

(16)
A \cap \hat{A} = \{(x, \max \{\mu_x(x), \nu_x(x)\}, \min \{\nu_x(x), \mu_x(x)\}) \}

(17)

For an IFS \( \hat{A} \), Wang et al. [17] presented a different entropy formula by

\[
E_{\text{tr}}(\hat{A}) = \frac{1}{n} \sum_{x=1}^{n} \min \{\mu_x(x), \nu_x(x)\} + \pi_x(x)
\]

\[
\max \{\nu_x(x), \mu_x(x)\} + \pi_x(x)
\]

(18)

III. MEDICAL DIAGNOSIS USING INTUITIONISTIC FUZZY SET

We have maintained a sample medical knowledge base consists of the values of tested symptoms S and proper diagnosis D. This knowledge base is prepared by arranging a thorough discussion with a group of related experts from a local Govt. hospital. The knowledgebase is formulated in terms of intuitionistic fuzzy sets. For experimentation we have taken the set of diagnosis be D = \{Vision loss, Heart failure, Stroke, Foot ulcer, Nerve damage\} and the set of symptoms be S = \{Pregnant, Glucose, DBP, TSFT, INS, BMI, DPF, Age\} which are commonly observed in a diabetic patient. The details of data are given in Table 2. According to the basis of intuitionistic fuzzy set (IFS), this article has described each symptom by three numbers: membership value \( \mu_x(x) \), non-membership value \( \nu_x(x) \) and hesitation margin \( \pi_x(x) \). Table 3 represents the considered set of patients \( P = \{\text{Jac, Karl, Rob, Smith, Riki}\} \) and their symptoms S with an intuitionistic fuzzy set.

IV. ANALYSIS OF MEDICAL DIAGNOSIS VIA THREE INDICATORS: DISTANCES, SIMILARITY MEASUREMENTS AND ENTROPY

To analyze the problem we have initially considered hamming distance between diagnoses and symptoms of each patient suffering from diabetes mellitus. To prescribe proper diagnosis we have calculated the hamming distance for each patient \( p_i \) \((i=1,2,\ldots,5)\) from the set of symptoms \( s_j \), \((j=1,2,3,\ldots,8)\) (Table 3) to the set of diagnosis \( d_k \), \((k=1,2,\ldots,5)\) (Table 2). After measuring the distance, the lowest obtained distance points out a proper diagnosis. We know that it is necessary to include all the three parameters such as membership function, non membership function and hesitation margin to calculate the distance for intuitionistic fuzzy sets ([13], [14]). The normalized hamming distance can be calculated by the expression described as (7). The distances (7) for each patient from a considered set of possible diagnosis are given in Table 4. Proper diagnosis is pointed out by the minimum distance. According to the desired result, Jac may suffer from vision loss, Carl may suffer from foot ulcer, Rob from stroke, Smith from vision loss and Riki also from vision loss.

As proved by Szmidt and Kacprzyk ([13], [14]) the final outcome of diagnosis problem by using normalized Euclidean distance will be the same as the result drawn from Hamming distance measurements. The result of similarity measurements using the expression (11) for each patient \( p_i \) \((i=1,2,\ldots,5)\) from the set of symptoms \( s_j \), \((j=1,2,3,\ldots,8)\) (Table 3) to the set of diagnosis \( d_k \), \((k=1,2,\ldots,5)\) (Table 2) is depicted in Table 5. Then the proper diagnosis \( d_k \) for the \( i \)th patient \( p_i \) is derived according to the biggest numerical value from the obtained similarity measures. From Table 5, we can see Jac may suffer from vision loss, Carl may suffer from foot ulcer, Rob from stroke, Smith from vision loss and Riki also from vision loss. These results obtained here are similar with the ones obtained from Table 4.

Table 2

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Vision Loss</th>
<th>Heart Failure</th>
<th>Stroke</th>
<th>Foot Ulcer</th>
<th>Nerve Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregnant</td>
<td>(0.4,0.0,0.6)</td>
<td>(0.7,0.0,0.3)</td>
<td>(0.1,0.3,0.4)</td>
<td>(0.1,0.7,0.2)</td>
<td>(0.1,0.8,0.1)</td>
</tr>
<tr>
<td>Glucose</td>
<td>(0.3,0.5,0.2)</td>
<td>(0.2,0.6,0.2)</td>
<td>(0.6,0.1,0.3)</td>
<td>(0.2,0.4,0.4)</td>
<td>(0.0,0.8,0.2)</td>
</tr>
<tr>
<td>DBP</td>
<td>(0.1,0.7,0.2)</td>
<td>(0.0,0.9,0.1)</td>
<td>(0.2,0.7,0.1)</td>
<td>(0.8,0.0,0.2)</td>
<td>(0.2,0.8,0.0)</td>
</tr>
<tr>
<td>TSFT</td>
<td>(0.4,0.3,0.3)</td>
<td>(0.7,0.0,0.5)</td>
<td>(0.2,0.6,0.2)</td>
<td>(0.2,0.7,0.1)</td>
<td>(0.2,0.8,0.0)</td>
</tr>
<tr>
<td>INS</td>
<td>(0.1,0.7,0.2)</td>
<td>(0.1,0.8,0.1)</td>
<td>(0.1,0.9,0.0)</td>
<td>(0.2,0.7,0.1)</td>
<td>(0.8,0.1,0.1)</td>
</tr>
<tr>
<td>BMI</td>
<td>(0.2,0.3,0.5)</td>
<td>(0.7,0.1,0.2)</td>
<td>(0.6,0.3,0.1)</td>
<td>(0.6,0.3,0.1)</td>
<td>(0.2,0.4,0.4)</td>
</tr>
<tr>
<td>DPF</td>
<td>(0.5,0.3,0.2)</td>
<td>(0.6,0.2,0.2)</td>
<td>(0.5,0.3,0.2)</td>
<td>(0.5,0.2,0.3)</td>
<td>(0.1,0.6,0.3)</td>
</tr>
<tr>
<td>Age</td>
<td>(0.4,0.4,0.2)</td>
<td>(0.1,0.2,0.7)</td>
<td>(0.4,0.4,0.2)</td>
<td>(0.8,0.1,0.1)</td>
<td>(0.8,0.1,0.1)</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Patients</th>
<th>Pregnant</th>
<th>Glucose</th>
<th>DBP</th>
<th>TSFT</th>
<th>INS</th>
<th>BMI</th>
<th>DPF</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jac</td>
<td>(0.8,0.1,0.1)</td>
<td>(0.6,0.1,0.3)</td>
<td>(0.2,0.8,0.0)</td>
<td>(0.6,0.1,0.3)</td>
<td>(0.1,0.6,0.3)</td>
<td>(0.2,0.2,0.6)</td>
<td>(0.4,0.5,0.1)</td>
<td>(0.4,0.4,0.2)</td>
</tr>
<tr>
<td>Carl</td>
<td>(0.8,0.0,0.2)</td>
<td>(0.4,0.4,0.2)</td>
<td>(0.6,0.1,0.3)</td>
<td>(0.1,0.7,0.2)</td>
<td>(0.1,0.8,0.1)</td>
<td>(0.7,0.4,0.0)</td>
<td>(0.8,0.1,0.1)</td>
<td>(0.0,0.4,0.6)</td>
</tr>
<tr>
<td>Rob</td>
<td>(0.8,0.1,0.1)</td>
<td>(0.8,0.1,0.1)</td>
<td>(0.0,0.6,0.4)</td>
<td>(0.2,0.7,0.1)</td>
<td>(0.0,0.5,0.5)</td>
<td>(0.7,0.2,0.1)</td>
<td>(0.9,0.1,0.0)</td>
<td>(0.7,0.3,0.0)</td>
</tr>
<tr>
<td>Smith</td>
<td>(0.6,0.1,0.3)</td>
<td>(0.5,0.4,0.1)</td>
<td>(0.3,0.4,0.3)</td>
<td>(0.7,0.2,0.1)</td>
<td>(0.3,0.4,0.3)</td>
<td>(0.8,0.2,0.0)</td>
<td>(0.3,0.3,0.4)</td>
<td>(0.6,0.2,0.2)</td>
</tr>
<tr>
<td>Riki</td>
<td>(0.3,0.3,0.4)</td>
<td>(0.5,0.3,0.2)</td>
<td>(0.6,0.3,0.1)</td>
<td>(0.7,0.2,0.1)</td>
<td>(0.4,0.4,0.2)</td>
<td>(0.3,0.3,0.4)</td>
<td>(0.6,0.4,0.0)</td>
<td>(0.4,0.5,0.1)</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Patients</th>
<th>Vision Loss</th>
<th>Heart Failure</th>
<th>Stroke</th>
<th>Foot Ulcer</th>
<th>Nerve Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jac</td>
<td>0.34</td>
<td>0.50</td>
<td>0.42</td>
<td>0.39</td>
<td>0.81</td>
</tr>
<tr>
<td>Carl</td>
<td>0.65</td>
<td>0.63</td>
<td>0.49</td>
<td>0.39</td>
<td>0.81</td>
</tr>
<tr>
<td>Rob</td>
<td>0.62</td>
<td>0.66</td>
<td>0.48</td>
<td>0.64</td>
<td>0.86</td>
</tr>
<tr>
<td>Smith</td>
<td>0.48</td>
<td>0.50</td>
<td>0.50</td>
<td>0.56</td>
<td>0.76</td>
</tr>
<tr>
<td>Riki</td>
<td>0.40</td>
<td>0.62</td>
<td>0.44</td>
<td>0.56</td>
<td>0.70</td>
</tr>
</tbody>
</table>
It is known that when entropy is zero, the degree of fuzziness is zero i.e. the fuzzy set is no more a fuzzy set, simply it will become a crisp set. But when the entropy is 100% i.e. the set is most fuzzy and then it is quite impossible to predict whether a particular point belongs or does not belong to our set. When a set will be most fuzzy then the hesitation margin will be 1, that’s why we cannot tell if this point belongs or does not belong to our set. In our experiment we have first investigated the entropy measurements of each patient \( p_i \) corresponding to the symptoms \( s_j \) using entropy formulation method defined in expression (18). Then we have explored the entropies for all the diagnosis \( d_j \) corresponding to the symptoms \( s_j \) by using Table 2 and the same method as previous one. Entropies of each patient is depicted by \( E(p_i) \forall i \in \{1,2,...,5\} \) and entropies of each diagnosis is depicted by \( E(d_j) \forall j \in \{1,2,...,5\} \). Now the entropy differences between each patient and the corresponding diagnosis can be formulated by

\[
E(F) = |E(p_i) - E(d_j)| \forall i,j \in \{1,2,...,5\} \tag{19}
\]

The result of the formulation (19) is shown in Table 6. By observing the result we have found that when the entropy difference is maximum for a particular patient then that particular diagnosis had never appeared for that patient. From the derived result set (Table 6) it is clear that patient Jac has maximum entropy difference (0.2597) for Nerve damage. So it can be concluded that patient Jac will never suffer from Nerve damage. Comparing Table 4 and Table 5 we have noted that Jac suffers from Vision loss for which Jac has an entropy difference 0.0986 which is lower relative to other entropy differences. If we take another example then again we get similar kind of results. For the patient Carl, maximum entropy difference is 0.2292 which implies Vision loss. As seen from Table 4 and Table 5, Carl is not suffering from Vision loss. Instead of that Carl is suffering from Foot ulcer which has entropy difference 0.0177, which is again lower. From our experimental view we can agree that when the entropy difference between two set is low, the those two sets are closely related with one another and they could converge but when the entropy difference is larger, then there is very less / no possibility of coming closer or being converged of those two sets.

\section*{V. Conclusion}

In this article we have worked on the consequences of diabetes mellitus problem using intuitionistic fuzzy set by taking into consideration hamming distance, similarity measurements and entropy computation to arrive at a firm conclusion. It is observed that the final result using hamming distance and similarity measurements are same. We have extended the analysis by using entropy measurement which made the final result more strong i.e. which sets are closer to consensus and which are not within reach. The proposed approach shows that when the entropy difference between patient and diagnosis is low, then the diagnosis is in close proximity with the patient and when it is higher, then there is almost no chance of being suffered by that diagnosis. By this analysis one diabetic patient would have sufficient information about the various consequences that he/she has to face in near future and therefore can take some preventive measures or precautions to handle the situation. Future scope of this research work might be to find out the relationship between entropy measurements and the distance measurements in more comprehensive way with a larger dataset by using a suitable algorithmic approach.

\section*{Acknowledgment}

The authors would like to thank the anonymous referees for their valuable comments and suggestions that resulted in a much improved paper.

\section*{References}


