Similarity measure between fuzzy set, fuzzy numbers and fuzzy rules using T Fuzzy Assessment Methodology

M.Kalpana, A.V Senthil Kumar

Department of Social Sciences, Anbil Dharmalingam Agricultural College and Research Institute, Tamil Nadu Agricultural University, Trichy, Tamil Nadu, India

Abstract: This paper expresses the prominent futures of fuzzy expert system by applying the algorithm T Fuzzy Assessment Methodology. Fuzzy expert system consists of the following elements such as fuzzification interface, T Fuzzy Assessment Methodology, and defuzzification. T Fuzzy Assessment Methodology uses the K Ratio to find overlapping between membership function and T Fuzzy similarity measure the similarity between fuzzy set, fuzzy number and fuzzy rule. Similar fuzzy sets are merged to form a common set; a new methodology was framed to identify the similarity between fuzzy rules with fuzzy numbers. Similar fuzzy numbers are merged to reduce the number of rules. The efficiency of the proposed algorithm was implemented using MATLAB Fuzzy Logic tool box to construct fuzzy expert system to diagnosis diabetes.

Key words: T Fuzzy Assessment Methodology (TFAM), K Ratio, T Fuzzy similarity measure, Diabetes application.

INTRODUCTION

Diabetes Mellitus is not only a hereditary disease but a heterogeneous group of diseases, which lead to high blood glucose levels due to defects in either insulin secretion or insulin action. Fuzzy Expert System is very important to diagnosis the patient suffering from diabetes. Chang and Lilly (2004) a fuzzy classification system was derived directly from the data, at the beginning fuzzy classifier is empty with no rules and no membership function. Then rules and membership function are created in this process. Polat and Gunes (2007) used principal component analysis to diabetes disease dataset, has 8 features which is reduced to 4 features and adaptive neuro-fuzzy inference is conducted to diagnosis diabetes. The American Diabetes Association (2007) categorizes diabetes into two types as type-1 and type-2 diabetes. Type-1 is most common for children and young adults and type-2 diabetes is common form of diabetes that the body does not produce adequate insulin. Kahramanli and Allahverdi (2007) developed a hybrid neural network that includes artificial neural network and fuzzy neural network. Method used in hybrid neural network is classification which increases the reliability of the result for heart and diabetes data.

Chang-Shing Lee (2011) designed fuzzy expert system using the algorithm fuzzy decision making mechanism to diagnosis diabetes. Fuzzy ontology is applied to diabetes data. With the fuzzy expert system and fuzzy ontology medical staff decides the patient is affected by diabetes or not. M. Kalpana and A. V Senthilkumar (2011a) developed a fuzzy expert system using the algorithm fuzzy verdict mechanism to diagnosis the diabetes. The proposed fuzzy expert system uses the concept of fuzzification and defuzzification. A. V Senthilkumar and M. Kalpana (2011) designed intensified fuzzy verdict mechanism which consists of fuzzy inference, implication and aggregation. M. Kalpana and A. V Senthilkumar (2012a) proposed a fuzzy expert system using Correlation fuzzy logic to find the relationship between two membership function. M. Kalpana and A. V Senthilkumar (2012b) Design and implemented Fuzzy Expert System using Fuzzy Assessment Methodology to diagnosis the diabetes. Vincent C. Yen(1999) suggested a measure of nearness between rules for fuzzy expert system. Rami Zwick et al. (1987) gave a similarity measure for fuzzy concept using geometric and set theoretic. Nineteen similarity measures were reviewed compared in behavioral experiments. Detlef D. Nauck (2003) suggested an index to measure the interpretability of fuzzy rule for classification problems and demonstrated using bench mark and real world data sets. Zsolt Csaba Jahanyak et al. (2005) introduced distance based similarity measures. The rule antecedence of fuzzy rule was not fully covered in input universe. Therefore similarity measures were used to distinguish the similarity of non overlapping fuzzy sets. Julian Luengo et al. (2009) the behavior of fuzzy rule based classification system selected proposal called Positive Definite Fuzzy Classifier which is a fuzzy system that uses Support Vector Machines to get accurate results and low number of rules. Wlodzislaw Duch et al. (2004) opens a new way to generate similarity between fuzzy rules based on individual features. Raouf ketata et al. (2007) introduced a new approach for fuzzy rule reduction with new fuzzy set. This methodology was applied to truck backer upper
control and liver trauma diagnostic. M.J Gacto et al. (2011) proposed interpretability measures for fuzzy rule based systems. A taxonomy based system proposed is “Complexity versus Semantic interpretability”. Wen-Liang Hung et al. (2008) proposed similarity measures between intuitionistic fuzzy sets. A new measure is used to evaluate students answer scripts. Samia Nefti et al. (2008) proposed a merging parameter fuzzy set based on clustering. H Bunke et al. (2001) proposed a new similarity measure with numeric and symbolic features with Euclidean distance function. For medical expert system the proposed similarity measure is applied. Magne et al. (1998) used a similarity measure to rule simplification that reduces fuzzy sets. Fuzzy sets are merged to create a common fuzzy set to replace in the rule.

The proposed T Fuzzy Assessment Methodology (TFAM) finds the similarity between the fuzzy set, fuzzy number and fuzzy rule. The methods proposed compares the three fuzzy set at the time and the sets are reduced. With the reduced set the similarity between the rules are achieved and the rules are reduced to improve the accuracy of the fuzzy expert system. This paper is organized as follows: the first parts deals with the Design of fuzzy expert system. The experimental results, implemented in MATLAB fuzzy logic toolbox are presented and experimental results indicate that the proposed method are compared with other methods (Chang, X., et al., 2004; Polat, K., et al., 2007; Kahramanli, H., et al., 2007; Chang-Shing, et al., 2011; Lee, C.S., et al., 2007).

Design of Fuzzy Expert System:

The fuzzy expert system includes Fuzzification interface, T Fuzzy Assessment Methodology (TFAM) and Defuzzification interface for diabetes represented in Fig. 1.

Pima Indians Diabetes Database:

The Pima Indians Diabetes Database (Lee, C.S., et al., 2007) is used to test the proposed algorithm T Fuzzy Assessment Methodology.

Modeling Fuzzy Expert Systems:

Fuzzy expert system can be designed using the following steps.
1. Fuzzification interface
2. T Fuzzy Assessment Methodology
3. Defuzzification interface

Fuzzy set and fuzzy numbers are listed in Table 1.

Fuzzification Interface:

The transformation of crisp inputs into fuzzy values and the fuzzy values taken as input for the T Fuzzy Assessment Methodology. Membership function adopted is triangular function with the parameter set {a,b,c} as shown in Eq. (1). The parameter is fixed with Minimum value, Mean, Standard Deviation, Maximum value for each variables (Kalpana, M., et al., 2011b). Then the membership function $\mu(x)$ of the triangular fuzzy numbers (William Siler, 2005) is given by

$$
\mu(x) = \begin{cases} 
0, & x \leq a \\
(x - a) / (b - a), & a < x \leq b \\
(c - x) / (c - b), & b < x < c \\
0, & x > c 
\end{cases}
$$

- (1)

Fig. 1: Diagram of the Fuzzy Expert System for diabetes
Table 1: Representation of Fuzzy variables and numbers

<table>
<thead>
<tr>
<th>Fuzzy Variables</th>
<th>Representation of Fuzzy Variables</th>
<th>Fuzzy Numbers</th>
<th>Representation of fuzzy numbers</th>
<th>Fuzzy triangular numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose D1</td>
<td>low d11 [71, 94.41, 121.27]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium d12 [94.41, 121.27, 148.12]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high d13 [121.27, 148.12, 196]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INS D2</td>
<td>low d21 [0.15, 16.89]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium d22 [15.16, 89.82, 194.81]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high d23 [89.82, 194.81, 198.9]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI D3</td>
<td>low d31 [0.24, 46.33, 32.24]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium d32 [24.46, 33.24, 24.2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high d33 [3.24, 42.03, 17.67]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPF D4</td>
<td>low d41 [0.13, 0.21, 0.44]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium d42 [0.21, 0.44, 0.67]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high d43 [0.44, 0.67, 0.96]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age D5</td>
<td>young d51 [21, 21, 22]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium d52 [21, 22, 24]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>old d53 [22, 24, 25]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DM O</td>
<td>verylow O1 [0.0, 0.1, 0.2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>low O2 [0.15, 0.25, 0.33]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>medium O3 [0.287, 0.327, 0.3997]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high O4 [0.329, 0.623, 0.762]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>veryhigh O5 [0.731, 0.831, 1]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**T Fuzzy Assessment Methodology:**

T Fuzzy Assessment Methodology (T FAM) three triangular membership functions (MFs) are used for each input variable (D1, D2, D3, D4, D5) and four triangular MFs for the output variable (O) using eqn. (1) with parameters D1 {low [Min, Mean-SD, Mean], medium [Mean-SD, Mean, Mean+SD], high[Mean, Mean+SD, Max]}, D2 {low [Min, Mean-SD, Mean], medium [Mean-SD, Mean, Mean+SD], high[Mean, Mean, Max+SD]}, D3 {low [Min, Mean-SD, Mean], medium [Mean-SD, Mean, Mean+SD], high[Mean, Mean, Max+SD]}, D4 {low [Min, Mean-SD, Mean], medium [Mean-SD, Mean, Mean+SD], high[Mean, Mean, Max+SD]} and D5 {low [Min, Mean-SD, Mean], medium [Mean-SD, Mean, Mean+SD], high[Mean, Mean, Max+SD]} listed in Table 1(Kalpana, M., et al., 2011b). In TFMAM T-norm operator used is algebraic product and T-conorm operator used is algebraic sum(Peter, H Sydenham., et al., 2005).

**K ratio:**

For the input variable the numbers of membership functions are designed. The membership functions overlap between each other. The intersection of membership function starts with two points. Let the first point at the beginning be Lower Member (LM) and the second point next to Lower Member be Upper Member (UM). The Fuzzy Mid Value (FMV) is the midmost point between the UM and LM. A straight line is drawn above the FMV. The intersection of the line from the first membership function is P2. Fuzzy Start Value (FSV) is calculated by FSV=FMV-0.5. A straight line is drawn above the FSV. The intersection of the line from the second membership function is P1. Hence K ratio is calculated as

\[ K = \frac{P1 + P2}{LM - UM} \]  

K ratio lies between 0 to 1. If the K ratio is greater than 1 then the membership function are fixed to the limit as LM=0.5 points are moved after LM UMK=0.5 points are moved before UM

To calculate the K ratio for the input variable Diabetes Pedigree Function (D4) for the membership function(low and medium) d41 and d42 given as P1=0.75, P2=0.5, and UM=0.44, LM=0.2 as shown in Fig. 2 and K ratio is calculated using the Eq.(2). The value for K ratio =5.208333. K ratio is greater than 1. To fit the membership function calculate FMV=0.32. So the LM can be changed to LMK and UM can be changed to UMK.

**Rules for Fuzzy Expert System:**

1. If (D1 is d11) or (D2 is d21/d22) or (D3 is d31) or (D4 is d41) or (D5 is d51 or d52) then (O is O1 or O2).
2. If (D1 is d11) or (D2 is d21/d22) or (D3 is d33) or (D4 is d41) or (D5 is d51 or d52) then (O is O1 or O2).
3. If (D1 is d12) or (D2 is d23) or (D3 is d33) or (D4 is d42) or (D5 is d51 or d52) then (O is O3).
4. If (D1 is d13) or (D2 is d21ord22) or (D3 is d33) or (D4 is d43) or (D5 is d51or d52) then (O is O4orO5).
5. If (D1 is d11) or (D2 is d21ord22) or (D3 is d32) or (D4 is d41) or (D5 is d51or d52) then (O is O1orO2).
6. If (D1 is d12) or (D2 is d21ord22) or (D3 is d33) or (D4 is d42) or (D5 is d51or d52) then (O is O4orO5).
7. If (D1 is d13) or (D2 is d23) or (D3 is d32) or (D4 is d41) or (D5 is d51or d52) then (O is O4orO5).
8. If (D1 is d11) or (D2 is d21ord22) or (D3 is d32) or (D4 is d41) or (D5 is d51or d52) then (O is O1orO2).
9. If (D1 is d11) or (D2 is d21ord22) or (D3 is d31) or (D4 is d41) or (D5 is d53) then (O is O1orO2).

**T Fuzzy Similarity Measure Between Fuzzy Set And Fuzzy Numbers:**

To get the adequate rule base, membership functions and sufficient number of fuzzy set are very essential (Raouf Ketata, et al., 2007). Fuzzy set similarity always considers the two fuzzy set, the proposed T Fuzzy similarity measure considered three fuzzy set A, B, and C. Consider the fuzzy variable glucose with its fuzzy number low, medium and high.

- Glucose low $d_{11}$ (A)
- medium $d_{12}$ (B)
- high $d_{13}$ (C)

\[
S(A, B, C) = \left( \frac{\max(A, B) - \min(A, B)}{3, \max(B, C) - \min(B, C)} \right)
\]  

**Glucos e(A, B, C) = \left( \frac{148.12 - 71}{3}, \frac{196 - 94.41}{3} \right) = (25.7, 33.9) \]  

With Eq.(3) similarity for the variable Glucose is calculated. The values are not similar, so the fuzzy set cannot be merged. Consider BMI with its fuzzy numbers low, medium and high and the similarity are measured using the Eq. (3)

- BMI low $d_{31}$ (A)
- medium $d_{32}$ (B)
- high $d_{33}$ (C)

\[
BMI(A, B, C) = \left( \frac{43.03 - 0}{3}, \frac{67 - 24}{3} \right) = (14.1, 14.3)
\]
The values of BMI low $d_{31}$ (A), medium $d_{32}$ (B) and high $d_{33}$ (C) are similar, the three sets can be merged into two sets as BMI low and BMI high. Rule 5 has BMI medium. It can be merged with Rule 1, all the parameters in Rule 1 are same except BMI medium. Rule 8 has BMI medium. It can be merged with Rule 2, all the parameters in Rule 2 are same except BMI medium. Finally Rule 5 and Rule 8 are merged with Rule 1 and Rule 2.

**T Fuzzy Similarity Measure Between Fuzzy Rules:**

The rule base consists of nine if-then rules with antecedent (Vincent, C Yen., 1999). The Antecedent of rule are $d_{11}, d_{12}, d_{13}, d_{21}, d_{22}, d_{23}, d_{31}, d_{32}, d_{33}, d_{41}, d_{42}, d_{43}, d_{51}, d_{52}, d_{53}$. The antecedent part of the rule evaluated with OR operator. Uncertainty are managed between the rules using Fact Values (Kalpana, M., et al., 2013) using measure of credulity and measure of incredulity. Compute the degree of similarity between all rules in the order.

Consider two rules

Rule 3: If $(D1$ is $d_{12}$) or $(D2$ is $d_{23})$ or $(D3$ is $d_{33})$ or $(D4$ is $d_{42})$ or $(D5$ is $d_{51}$ or $d_{52})$ then $(O$ is $O3)$.

Rule 4: If $(D1$ is $d_{13})$ or $(D2$ is $d_{21}$ or $d_{22})$ or $(D3$ is $d_{33})$ or $(D4$ is $d_{43})$ or $(D5$ is $d_{51}$ or $d_{52})$ then $(O$ is $O4$ or $O5)$.

Degree of Similarity (DS) = Total number of similar parameter between the rule

\[
\frac{\text{Total number of input and output parameter}}{\text{Total number of input and output parameter}} = \frac{2}{6} = 33\%
\]

DS = 33% < 50%

Rule 3 and Rule 4 are dissimilar

Consider another two rules

Rule 4: If $(D1$ is $d_{13})$ or $(D2$ is $d_{21}$ or $d_{22})$ or $(D3$ is $d_{33})$ or $(D4$ is $d_{43})$ or $(D5$ is $d_{51}$ or $d_{52})$ then $(O$ is $O4$ or $O5)$.

Rule 6: If $(D1$ is $d_{12})$ or $(D2$ is $d_{21}$ or $d_{22})$ or $(D3$ is $d_{33})$ or $(D4$ is $d_{42})$ or $(D5$ is $d_{51}$ or $d_{52})$ then $(O$ is $O4$ or $O5)$.

Degree of Similarity (DS) = Total number of similar parameter between the rule

\[
\frac{\text{Total number of input and output parameter}}{\text{Total number of input and output parameter}} = \frac{4}{6} = 66\%
\]

DS = 66% > 50%

Rule 4 and Rule 6 are similar. Dissimilar parameters in the rule are Glucose($d_{12}$ and $d_{13}$) and DPF($d_{42}$ and $d_{43}$). Let us consider the dissimilar parameter Glucose($d_{12}$ and $d_{13}$)

\[
\text{Glucose}(\text{cvalued}_{12}) = \frac{94.41 + 121.27 + 148.12}{3} = 121.3
\]

\[
\text{Glucose}(\text{cvalued}_{13}) = \frac{121.27 + 148.12 + 196}{3} = 155.13
\]

\[
\text{Glucose}(\text{kcalvalue}) = \text{Glucose}(\text{cvalued}_{13}) - \text{Glucose}(\text{cvalued}_{12}) = 155.13 - 121.3 = 33.83
\]

\[
\text{Glucose}(\text{kbase}) = \text{First triangular number of } d_{13} \cdot \text{First triangular number of } d_{12} = 121.27 - 94.41 = 26.86
\]

Glucose(Kcalvalue) > Glucose(Kbase) ie, $d_{12}$ and $d_{13}$ are reduced. Minimum value $d_{12}$ is deleted, Maximum value $d_{13}$ is considered in rule.

Let us consider the dissimilar parameter DPF($d_{42}$ and $d_{43}$)
DPF(cvalued \_d_{43}) = \frac{0.21 + 0.44 + 0.67}{3} = 0.44 \\
DPF(cvalued \_d_{45}) = \frac{0.44 + 0.67 + 0.96}{3} = 0.69 \\
DPF(kcalvalue) = DPF(cvalued \_d_{43}) - DPF(cvalued \_d_{42}) = 0.69 - 0.44 = 0.25 \\
DPF(kbase) = \text{First triangular number of } d_{43} - \text{First triangular number of } d_{42} \\
= 0.44 - 0.21 = 0.23 \\

DPF (Kcalvalue) > DPF (Kbase) ie, d_{42} and d_{43} are reduced. Minimum value d_{42} is deleted, Maximum value d_{43} is considered in rule.

So the rules Rule 4 and Rule 6 can be reduced into one Rule as
If (D1 is d_{13}) or (D2 is d_{21} or d_{22}) or (D3 is d_{33}) or (D4 is d_{43}) or (D5 is d_{51} or d_{52}) then (O is O4 or O5).

By using the T Fuzzy similarity measure between fuzzy set, fuzzy number and rules nine if-then rules are reduced to six rules to get accurate result and to reduce the time in the construction of rules.

1. If (D1 is d_{11}) or (D2 is d_{21} or d_{22}) or (D3 is d_{31}) or (D4 is d_{41}) or (D5 is d_{51} or d_{52}) then (O is O1 or O2).
2. If (D1 is d_{11}) or (D2 is d_{21} or d_{22}) or (D3 is d_{33}) or (D4 is d_{41}) or (D5 is d_{51} or d_{52}) then (O is O1 or O2).
3. If (D1 is d_{12}) or (D2 is d_{23}) or (D3 is d_{33}) or (D4 is d_{42}) or (D5 is d_{51} or d_{52}) then (O is O3).
4. If (D1 is d_{13}) or (D2 is d_{21} or d_{22}) or (D3 is d_{33}) or (D4 is d_{43}) or (D5 is d_{51} or d_{52}) then (O is O4 or O5).
5. If (D1 is d_{13}) or (D2 is d_{21} or d_{23}) or (D3 is d_{31}) or (D4 is d_{41}) or (D5 is d_{53}) then (O is O4 or O5).
6. If (D1 is d_{11}) or (D2 is d_{21} or d_{22}) or (D3 is d_{31}) or (D4 is d_{41}) or (D5 is d_{53}) then (O is O1 or O2).

MIN and SUM operator:
Antecedent part of the rule gives a single number for implication process. The antecedent part of the rule are \(d_{11}, d_{12}, d_{13}, d_{21}, d_{22}, d_{23}, d_{31}, d_{32}, d_{33}, d_{41}, d_{42}, d_{43}, d_{51}, d_{52}, d_{53}\). The process of mapping result of fuzzification from antecedent part into consequence is termed as implication. To fire more than one fuzzy rule at same time, MIN operation is used by the system. The output of each rule is combined into single fuzzy set by aggregation process using SUM operation.

Defuzzification Interface:
The result obtained from the aggregation is fuzzy value. To convert the fuzzy value obtained from TFAM into crisp value defuzzification process is conducted. Centroid method is used for defuzzification process.

T Fuzzy Assessment Methodology analyzes the personal physical data, converts the results into knowledge and the patterns of statement for output descriptions. The pattern of the statement helps the medical practitioner to diagnosis the patient from diabetes.

Proposed Algorithm: T Fuzzy Assessment Methodology (TFAM):

Begin:
1. Input: Terms (D\_1, D\_2, D\_3, D\_4, D\_5) are selected as fuzzy input variables
2. Output: Output term O as fuzzy output variables
3. Input PIDD with N cases
4. Initialize i←1

Method:
Step 1: Create input fuzzy set D\_1(d_{11}, d_{12}, d_{13}), D\_2(d_{21}, d_{22}, d_{23}), D\_3(d_{31}, d_{32}, d_{33}), D\_4(d_{41}, d_{42}, d_{43}), D\_5(d_{51}, d_{52}, d_{53}) and output fuzzy set O(O\_1, O\_2, O\_3, O\_4, O\_5)
Step 2: Calculate the value of min, max mean and standard deviation DO UNTIL (i oglas voor een man. Wat is het verschil in gewicht voor een vrouw? Hoe kan je deze informatie gebruiken bij het bepalen van de optimale pippunne?
LMK = 0.5 points are moved after LM  
UMK= 0.5 points are moved before UM  
Else  
LM and UM

**Step 4**: DO UNTIL(i>N)
If (D1i is d11) or (D2i is d21) or (D3i is d31) or (D4i is d41) or (D5i is d51) then Oi is O3  
END IF  
END DO UNTIL

**Step 5**: Call procedure for T Fuzzy Similarity measure for fuzzy set, fuzzy numbers and rules

**Step 6**: Antecedent part (D1i is d11) or (D2i is d21) or (D3i is d31) or (D4i is d41) or (D5i is d51) into consequent (O is O3) by MIN operator

**Step 7**: Set rules output

---

**Procedure for T Fuzzy Similarity measure for fuzzy set, fuzzy numbers and rules:**

Begin:

**Step 1**: Generate initial fuzzy set, fuzzy numbers and rules for diabetes data.

**Step 2**: In this step, we propose a new similarity measure between three sets A, B and C by the equation

$$S(A, B, C) = \left( \frac{\max(A, B) - \min(A, B)}{3}, \frac{\max(B, C) - \min(B, C)}{3} \right)$$

We conclude that set A, B and C are similar if they have both the values equal otherwise not equal. If the two values in the above equation are equal then merge into two sets A1 and B1.

**Step 3**: Apply the merged set in rules.

**Step 4**: Compute the degree of similarity between all rules in the order.
Consider two rules
Rule 1: if (x1 is A1) or (x2 is B1) or (x3 is C1) or (x4 is D1) or (x5 is E1) then Y is O1
Rule 2: if (x1 is A1) or (x2 is B1) or (x3 is C2) or (x4 is D2) or (x5 is E1) then Y is O1

Degree of Similarity (DS) = Total number of similar parameter between the rule  
-------------------------------------------------------------------------%  
Total number of input and output parameter

DS = 4/6% = 66%

**Step 5**: Constant degree of similarity (CDS) is set to 50%
If (DS > CDS) then  
Goto step 6  
else  
Stop the algorithm

**Step 6**: Calculate cvalue of dissimilar input parameters (C1 and C2) and output parameter (D1 and D2)
cvalue for C1 and C2 are calculated by average of fuzzy numbers.
Kcalvalue = cvalue(C2) – cvalue(C1)
Kbase = First triangular number of C2 – First triangular number of C1
If (Kcalvalue > Kbase) two fuzzy number ie, C1 and C2 are reduced. Minimum value C1 is deleted, Maximum value C2 is considered in rule.
else not reduced.
Similarly cvalue, Kcalvalue, Kbase values of D1 and D2 are calculated and the Rule 1 and Rule 2 are merged into one rule.

End:

**Experimental Results:**
MATLAB Fuzzy Logic toolbox was used to evaluate the performance of the proposed fuzzy expert system. From Pima Indian diabetes dataset knowledge can be analyzed with the proposed T Fuzzy Assessment Methodology. Table 2 indicates the result obtained from TFAM. The acquired result from Table 2 transferred into knowledge and presented in the human understandable form.
Table 2: Final Result for Medical practitioner

<table>
<thead>
<tr>
<th>Data</th>
<th>Glucose (mg/dl)</th>
<th>INS (mu U/ml)</th>
<th>BMI (Kg/m²)</th>
<th>DPF</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement study</td>
<td>177</td>
<td>478</td>
<td>34.6</td>
<td>1.072</td>
<td>21</td>
</tr>
</tbody>
</table>

Assessment Statement:

If (Glucose is Gh) or (INS is INSm) or (BMI is BMIh) or (DPF is DPFh) or (Age is Agey) then (DM is DMh)

Justification by Medical Practitioner:

Medical practitioner justification is the person is diabetes

Performance Assessment:

Performance Assessment Statement can be assessed based on the accuracy level. The True Positive (TP) and the True Negative (TN) denote the correct classification. False Positive (FP) is the outcome when the predicted class is yes (or positive) and actual class is no (or negative). Still, a False Negative (FN) is the outcome when the predicted class is no (or negative) and actual class is yes (or positive). Table 3 lists the various outcomes of a two-class prediction (Lee, C.S., et al., 2007). Accuracy is the proportion of the total number of predictions that were correct. The Eq. (4) show the formula for accuracy.

\[
\text{Accuracy} = \frac{TN + TP}{TN + FP + FN + TP} \times 100\% - (4)
\]

Table 3: Different Outcomes of a Two-Class Prediction

<table>
<thead>
<tr>
<th>Actual class</th>
<th>Predicted class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>True positive (TP)</td>
</tr>
<tr>
<td>No</td>
<td>False positive (FN)</td>
</tr>
</tbody>
</table>

The final experiment compares the accuracy of the proposed method with results of studies involving the Pima Indians Diabetes Database (Chang, X., et al., 2004; Polat, K., et al., 2007; Kahramanli, H., et al., 2007; Chang-Shing, et al., 2011; Lee, C.S., et al., 2007). The proposed method achieves the highest accuracy value for “very very young: (AGE: 0-25)” than earlier methods which is indicated in the Table 4.

Table 4: Comparison of accuracy of Proposed Method with Earlier Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (%)</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our study for Very Very Young (AGE: 0-25)</td>
<td>83.15</td>
<td>M. Kalpana and Dr. A. V. Senthil Kumar</td>
</tr>
<tr>
<td>HNFB</td>
<td>78.26</td>
<td>Goncalves et al.</td>
</tr>
<tr>
<td>Logdisc</td>
<td>77.7</td>
<td>Statlog</td>
</tr>
<tr>
<td>IncNet</td>
<td>77.6</td>
<td>Norbert Jankowski</td>
</tr>
<tr>
<td>DIPOL 92</td>
<td>77.6</td>
<td>Statlog</td>
</tr>
<tr>
<td>Linear discr. Anal</td>
<td>77.3-77.7</td>
<td>Statlog, ster and Dobnikar</td>
</tr>
<tr>
<td>A FES for Diabetes Decision very young</td>
<td>77.3</td>
<td>Lee and Wang</td>
</tr>
<tr>
<td>VISIT (Chang-Shing, et al., 2011)</td>
<td>77</td>
<td>Chang and Lilly</td>
</tr>
<tr>
<td>SMART</td>
<td>76.8</td>
<td>Statlog</td>
</tr>
<tr>
<td>GTO DT (5 X CV)</td>
<td>76.8</td>
<td>Bennett and Blue</td>
</tr>
<tr>
<td>ASI</td>
<td>76.6</td>
<td>Ster and Dobnikar</td>
</tr>
<tr>
<td>Fisher discr. Analysis</td>
<td>76.5</td>
<td>Ster and Dobnikar</td>
</tr>
<tr>
<td>MLP+BP</td>
<td>76.4</td>
<td>Ster and Dobnikar</td>
</tr>
<tr>
<td>LVQ(20)</td>
<td>75.8</td>
<td>Ster and Dobnikar</td>
</tr>
<tr>
<td>LFC</td>
<td>75.8</td>
<td>Ster and Dobnikar</td>
</tr>
</tbody>
</table>

Conclusion and Future Research:

This paper presents application of fuzzy expert system for diagnosis of diabetes using T Fuzzy Assessment Methodology. PIDD dataset is processed and the crisp values are converted into fuzzy values in the stage of fuzzification. The T Fuzzy Assessment Methodology uses K ratio to find the overlapping between the membership function, executes rules with fuzzy operator. Similarity between the fuzzy set, fuzzy numbers and fuzzy rules are derived through the proposed T Fuzzy similarity measure between fuzzy sets, fuzzy numbers and fuzzy rules to make a decision on the possibility of individuals suffering from diabetes and to present the knowledge with descriptions. Finally defuzzification is adopted to convert the fuzzy output set to a crisp output. Accuracy achieved through this method is 83.15% which can also improved through future works. Future works includes to modify rules and to add rules to fuzzy expert system to perform similar accuracy.
REFERENCES


