Fuzzy Classification System by Self Generated Membership Function Using Clustering Technique

Shruti S. Jamsandekar[1] and **Ravindra R. Mudholkar**[2]

Submitted in August, 2013; Accepted in February, 2014

Abstract - In this paper we have proposed fuzzy classification system by generating the membership functions using semi supervised learning method. k-means clustering technique is used to form clusters and to obtain membership centers, each cluster boundary values. These boundary values are approximated to vertex values of membership functions for overlapping the membership functions. The work is extended with auto creation of fuzzy inference system for classification to perform the classification on the given input data.

Index Terms – fuzzy classification, clustering, k means, membership function

NOMENCLATURE
FIS – Fuzzy Inference System

1.0 INTRODUCTION

Fuzzy classification is one of the applications of fuzzy logic which is used to deal with classification problems. In development of a fuzzy classification system, the important task is to construct membership functions and to find a set of suitable fuzzy rules in the fuzzy classification system. Fuzzy membership functions and fuzzy rules can be formulated based on expert knowledge approach and other alternative is use data driven approach, these approaches correspond to either manual or automatically through a machine learning process based on training instances respectively, Mostly expert knowledge is used to formulate membership function range and if then rules for inference and it has an advantage as it has link with domain knowledge but it can be very subjective with different experts generating different membership functions and rules for same application. The other approach of generation of fuzzy membership functions based on the input data will convert crisp data into linguistic terms. Most previous attempts in automatic generation of fuzzy membership functions have used machine learning and statistical methods [16, 3, 4, 7, 11]. In this paper, we present a new method for constructing membership functions using k means clustering technique, based on cluster centers and approximation of boundary values of each cluster obtained. The paper has 2 major contributions, first approach data driven membership function generation by clustering algorithm and second contribution is automatic creation of *if then* fuzzy classification system to deal with student semester examination data. The automatic creation of the fuzzy classification system from the membership function generated reduces the hassle of every time creating a new fuzzy Inference system to perform classification on change of data input.The rest of this paper is organized as follows. In Section 2 work in related field is reviewed. In section 3, we briefly review basic concepts of fuzzy sets. The proposed method to construct membership functions based on clustering is explained in section 4. The details and steps involved in automatic creation of fuzzy inference system for classification are in section 5 In Section 6 we discuss the experimental results of the proposed method. The conclusions and future work are discussed in Section7.

[1]*Department of Computer Studies, SIBER, Kolhapur. (MS), India-416004. shrutijams@gmail.com*
[2]*Department of Electronics Shivaji University, Kolhapur, (MS), India-416004.*
E-mail: [1]shrutijams@gmail.com and
[2]rrm_eln@unishivaji.ac.in

2.0 RELATED WORK

In [1] the authors have proposed a fuzzy model for calculating the group maturity rating for different groups within a software organization based on the defect density, residual defect density and the review effectiveness of the historical projects. In [2] authors have discussed fuzzy multi-objective optimization model approach for selecting the optimal COTS software product among alternatives for each module in the development of modular software system they have used fuzzy methodology for the estimation of reliability and cost. In[3] authors discuss an improvement in performance in Fuzzy Rule Based Classification Systems by using interval valued fuzzy set and cooperative tuning of methodology .Oyelade, O. J et.al [15] proposed a method using Kmeans clustering algorithm for analyzing students' results based on cluster analysis. In [7] Sushmita Mitra et.al discussed a new methodology for encoding connection weights of Fuzzy MLP and a technique generating an architecture of the fuzzy MLP in terms of hidden nodes and links.

M.Holena [5] proposed an algorithm to extract rules for fuzzy system using MLP with continuous activation function. In [16] the authors discussed framework for generation of fuzzy membership functions from training data using genetic algorithm. In [17] authors propose a method to generate rules and learning for knowledge base using genetic algorithm.

Roberto R.F. Mendes et.al [4] proposes a system for fuzzy classification by using genetic algorithm to generate fuzzy rules and an evolutionary algorithm to construct membership functions.

[6] Proposed a clustering technique by using genetic algorithm which searches the cluster center in the cluster space and optimizing the similarity metrics. Authors in [8] proposed a design for fuzzy classification system based on labeled data using genetic algorithm, they have also reduced the rule set for the system. In [9] authors have discussed different clustering algorithms and used fuzzy logic based clustering in controlling multi compressor system. [10] M.A. Hogo presented an evaluation method of E-learners using Fuzzy C-means and Kernelized Fuzzy C-Means to predict the student profile and group them into different learning categories. Author in [11] have experimented apriori association rule mining using Weka tool to predict the student performance and find the correlation among attributes.
In [12] authors have proposed a new method of constructing fuzzy membership functions using α-cuts of fuzzy sets from training data set. *C Vialardi et.al in* [13] Proposed a recommendation system in higher education using data mining techniques, for student to choose appropriate course enrollment depending upon their academic performance. Hisao Ishibuch et.al [14] proposed an approach for automatic construction of fuzzy classification system using minimize set of rules generated by from genetic algorithm

3.0 INTRODUCTION TO FUZZY LOGIC

A fuzzy set A in a universe of discourse X is defined as the following set pairs [18]

$$A = \{\mu A(x) : x \in X\} \qquad (1)$$

Where, $\mu_A(x)$: $X \to [0, 1]$ is a mapping called the membership function of fuzzy set *A* and $\mu_A(x)$ is called
the degree of belongingness or membership value or degree of membership of $x \in X$ in the fuzzy set *A.* we write (1) in the following form:

$$A = \{\mu A(x)/x : x \in X\} \qquad (2)$$

For brevity, however, we often equate fuzzy sets with their membership functions i.e. instead of a fuzzy set *A*
characterized by $\mu_A(x)$ we will often say fuzzy sets μ_A [4, 8].

Example**:** Suppose $X = \{6, 2, 0, 4\}$. A fuzzy set of X may be given by $A = \{0.2/6, 1/2, 0.8/0, 0.1/4\}$.
Construction of membership function is based on the system design data and choice of the suitable shape. There are many shapes of membership functions. However, the application context dictates the choice of the suitable shape.
For the problem domain addressed in this study, system components have maximum and minimum value that cannot be exceeded. Therefore, any candidate membership function shape should have two extreme bounds with zero and hundred as range values. Triangular and trapezoidal shapes are the simplest MF shapes that meet this requirement.

The membership function of the triangular fuzzy set A can be represented by a triple (b; c; a), where "c" is called the center of the triangular fuzzy set A; "b" is called the left vertex of the triangular fuzzy set A; "a" is called the right vertex of the triangular fuzzy set A.

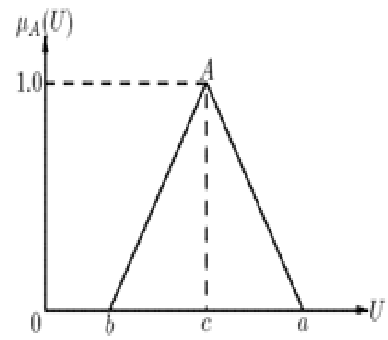

Figure 1: Triangular Membership Function

$$triangular(x; a, b, c) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ \dfrac{c-x}{c-b} & b \leq x \leq c \\ 0 & x \geq c \end{cases} \qquad (3)$$

The membership function of the trapezoidal fuzzy set can be represented by a function of a vector, x, and depends on four scalar parameters a, b, c, and d, as given by

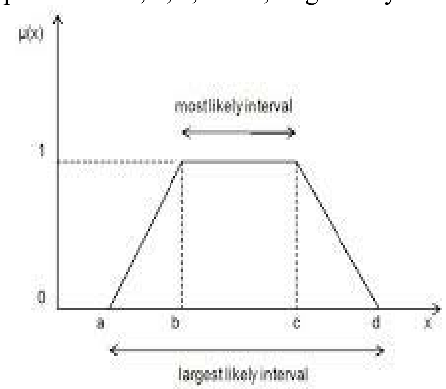

Figure 2: Trapezoidal Membership Function

$$Trapezoidal(x; a, b, c, d) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & b \leq x \leq c \\ \dfrac{d-x}{d-c} & c \leq x \leq d \\ 0 & d \leq x \end{cases} \qquad (4)$$

4.0 PROPOSED METHOD TO GENERATE MEMBERSHIP FUNCTIONS BASED ON K MEANS CLUSTERING TECHNIQUE

In the proposed method we have created a Fuzzy Inference System for two input variables and one output variable. A combination of two membership functions triangular and trapezoidal is used. The figure of triangular is shown in Figure 1and that of trapezoidal is in Figure 2. For input as well as output variables of the system the extreme membership functions are trapezoidal and membership function in between are triangular. In this paper we have used KMeans clustering algorithm to form clusters and constructed the triangular membership functions. Taking the cluster center as center c of triangular set referring to Figure 1. The extreme trapezoidal fuzzy sets are then approximated .

4.1 Membership Functions for Input Variables Created as Follows

I. Kmeans Clustering algorithm is used to form clusters on student examination data for each subject marks to form 3 cluster. The three cluster center found forms the center of 3 inner triangular fuzzy membership functions

II. The two vertexes of each of the triangular fuzzy membership function referring to Figure 1. b and a are calculated by calculating the maximum and minimum value of each cluster

III. The maximum value obtained for each cluster is increased by 10% which formed vertex b of each of the triangular membership function (Figure 1)

IV. The minimum value obtained for each cluster is decreased by 10% which formed the vertex a of each of the triangular membership function(Figure 1)

V. The trapezoidal membership functions for the input variables are constructed as follows

a) Left extreme trapezoidal membership function according to figure 2 a = 0, b = 0 and c value is calculated by 10 % difference of Minimum of first triangular membership function cluster with itself and d is calculated by 15 % difference of Minimum of first triangular membership function cluster with itself

b) Right extreme trapezoidal membership function according to figure 2 c = 100, d = 100 value and a is calculated as 10% difference of maximum obtained for third triangular membership cluster with itself and b is calculated as 5 % difference of maximum obtained for third triangular membership cluster with itself

4.2 Membership Functions for Output Variables Created as Follows

I. For output variable the inner triangular membership function is calculated as follows

II. Input variable corresponding first triangular membership function centers average forms the center of first triangular membership of output variable and same procedure is repeated for second and third triangular membership function centers

III. The two vertexes of 3 triangular membership for output variable referring to Figure 1

a) a vertex of each of the 3 triangular membership function is Average of minimum of first cluster, second cluster and third cluster for input variable(calculated for input in 4.1) respectively

b) b vertex of each of the 3 triangular membership function is Average of maximum of first cluster, second cluster and third cluster for input variable calculated for input in 4.1) respectively.

IV. The trapezoidal membership functions for the output variables are constructed as follows

a) Left extreme trapezoidal membership function according to figure 2 a, b with zero value and c is 20 and d is 35. Here 35 is taken since the domain data is student examination marks and passing is 35 out of 100

b) Right extreme trapezoidal membership function according to figure 2 c, d is assigned value 100, a is calculated as 10% difference of maximum obtained for third triangular membership cluster with itself and b is calculated as 5 % difference of maximum obtained for third triangular membership cluster with itself

5.0 AUTOMATIC CREATION OF FUZZY INFERENCE SYSTEM (FIS) FOR CLASSIFICATION

Matlab fuzzy tool was used to create the proposed fuzzy system. The fuzzy inference system is mostly constructed manually; instead in this experiment we tried the approach of automatic creation of Fuzzy inference System (FIS). In this paper we generated FIS given the input data for classification. The generation of FIS is enumerated in steps below.

5.1. Create a MATLAB® workspace variable FIS name.

FIS = NEWFIS (name of FIS, type of FIS, andMethod, or Method, implementation method, aggregation method, defuzzication method)

Example:
a= newfis('clustclassFIS','mamdani','min','max', 'min','max','centroid');

5.2. Create input variable and add it to FIS .

Addvar method adds an a variable to an FIS.[19]
The parameters of the function are as follows
First – FIS name
Second – the type of variable input or output
Third – the name of the variable
Fourth – the vector with range values of the variable

a=addvar(a,'input','sub1_marks',[0 100]);

5.3 Create membership function add it to FIS

ADDMF method adds a new membership function to an FIS
a=addmf(a,varType,varIndex,mfName, mfType,mfParams)[19]

The parameters of the function are in this order:
First - a matlab variable name of a FIS structure in the workspace
Second- the type of variable to be add the membership function (input or output)
Third – the index value of the membership function
Fourth- string representing the label of the membership function eg. Low, medium etc.
Fifth - A string representing the type of membership function eg. Triangular, trapezoidal etc.
Sixth- the vector specifying the membership function range values

In this experiment the vector of parameters passed is as per membership range calculated in section 4.1

5.4 Similarly create and add all input variables and output variable with their membership functions to the FIS.

In this experiment the vector of parameters passed to the output variable membership function is as per membership range calculated in section 4.2

5.5 Form the Rule base with rules defined and add it to FIS (to carry out inference mechanism)

a) Create rule list matrix for the FIS

The format of the rule list matrix is as follows. If there are m inputs to a system and n outputs, there must be exactly $m + n + 2$ columns to the rule list matrix.

The first m columns refer to the inputs of the FIS system. The next n columns refer to the outputs of the FIS system. Each column for inputs and outputs contains a number that refers to the index of the membership function for that variable. The $m + n + 1$ column contains the weight that is to be applied to the rule. It can ale value between o and 1.The $m + n + 2$ column contains the fuzzy operator for the rule's antecedent to be used, the value used is 1 if AND fuzzy operator and value is 2 if OR fuzzy operator.[19]

b) addrule(a,ruleList)

addrule method is used to add rule list to the FIS
It has two parameters as follows .
First- is the MATLAB workspace variable FIS name.
Second- is a matrix which represents a given rule.
Example

```
ruleList=[    1   3   2   1   1
              1   4   3   1   1
              5   2   2   1   1];
```

a = addrule(a,ruleList);
If the system a has two inputs and one output, the first rule can be interpreted as:"If Input 1 is MF 1 and Input 2 is MF 1, then Output 1 is MF 1."

6.0 EXPERIMENTAL RESULTS

Comparison of the experimental results of the proposed method with the existing methods
To illustrate the method proposed

i. Accept the student examination data. Student data set for two subject marks in semester exam is shown in , Appendix A
ii. Clustering the student data using kmeans algorithm
a) The cluster center for subject 1 data are center of three triangular membership functions of input variable 1

55.26 39.50 71.66

b) The cluster center for subject 2 data are center of three triangular membership functions of input variable 2
60.95 4 8.47 28.66

iii.. Calculate the left and right vertex points of the 3 triangular membership functions
a) Compute maximum and minimum for subject 1 and subject 2 which is input 1 and input 2 respectively.

Input 1 Cluster number	Maximum value	Minimum value
Cluster_ 1	46	29
Cluster_2	63	48
Cluster_3	81	65

Table 1: Max and Min of input 1

Input 1 Cluster number	Maximum value	Maximum value
Cluster_ 1	81	56
Cluster_2	53	43
Cluster_3	36	20

Table 2: Max and Min of input 2

b) Approximate values calculated for overlapping of membership of triangular membership functions of input_1

Left and right vertex of the 3 triangular membership function for input 1.

Membership Function	Right Vertex	Left Vertex
low	39.6	18
medium	58.3	38.7
high	89.1	50.4

Table 3: Left and Right vertex points of input 1

Left and right vertex of the 3 triangular membership function for input 2.

Membership Function	Right Vertex	Left Vertex
low	50.6	26.1
medium	69.3	43.2
high	89.1	58.5

Table 4: Left and Right vertex point of input 2

The following is the membership function generated of input1, input2 variables for subject1 and subject 2 respectively using steps of section 3.1 and section 4.

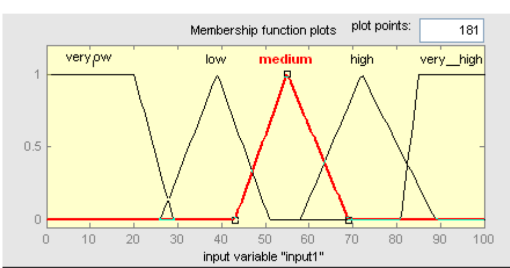

Figure 3: Generated membership function plot for input 1

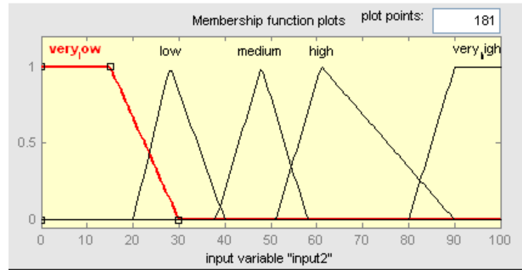

Figure 4: Generated membership function plot for input 2

The following is the membership function generated of output variables of fuzzy inference system for The following is the membership function generated of output variables of fuzzy inference system for classification using steps of section 3.1 and section 4.

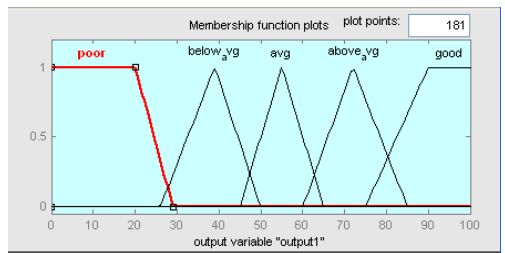

Figure 5: Generated membership function plot for output

c) The rule base for Inference created

ruleList= [1 1 1 1 1
1 2 2 1 1
1 3 2 1 1
1 4 3 1 1
5 2 2 1 1
5 3 3 1 1
2 5 4 1 1
3 1 2 1 1
3 2 3 1 1
3 3 3 1 1
4 2 2 1 1
4 3 3 1 1
4 4 4 1 1
4 5 5 1 1
1 5 3 1 1
2 1 1 1 1
2 2 2 1 1
2 3 2 1 1
2 4 3 1 1
3 4 4 1 1
3 4 4 1 1
4 1 2 1 1
5 1 2 1 1
5 4 4 1 1
5 5 5 1 1]

Snapshot of results obtained by Fuzzy Classification based on Membership function generated by Clustering (FCMFC) .

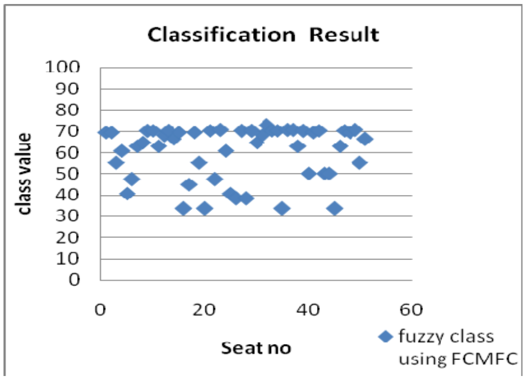

Figure 6: Classification result of Auto generated Fuzzy Inference System

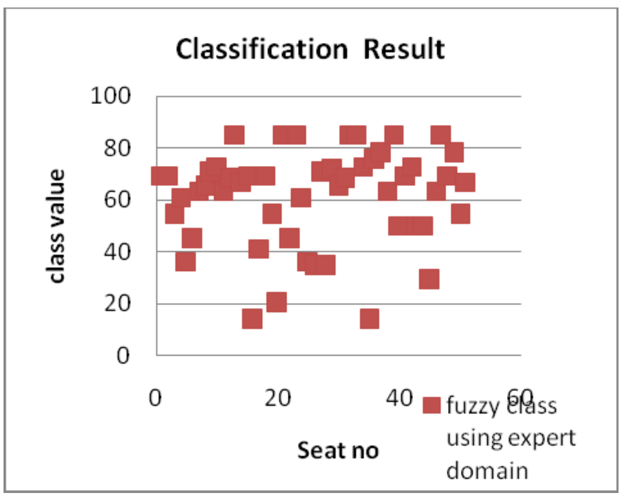

Figure 7: Classification result of Fuzzy Inference System using expert domain knowledge

The detail result in shown in Appendix B

7.0 CONCLUSION

The paper describes an approach for classification of students using fuzzy logic with membership function generated using semi supervised k means clustering and creation of Fuzzy inference system for reasoning automatically given the input data The approach was carried out on student examination data ,after obtaining the results difference is seen between the classes that are created based on the fuzzy membership function created using expert knowledge and proposed Fuzzy Logic classification method . The class value for the FCMFC are in range of 30 to 75 which means classification of students is below average to above average referring to figure 5, comparing it to expert domain method the class vale is in range of 10 an above 80.

REFERENCES

[1]. A. K. Verma, Anil R and Dr. Om Prakash Jain ,Fuzzy Logic Based Revised Defect Rating for Software Lifecycle Performance" Prediction Using GMR,BIJIT - BVICAM's International Journal of Information Technology, Vol. 1 No. 1; January – June, 2009

[2]. P. C. Jha, Shivani Bali and P. K. Kapur Fuzzy Approach for Selecting Optimal COTS Based Software ProductsUnder Consensus Recovery Block Scheme BIJIT - BVICAM's International Journal of Information Technology, January – June, 2011; Vol. 3 No.1

[3]. J. Sanz, et.al, "A Genetic Algorithm for Tuning Fuzzy Rule-Based Classification Systems with Interval- Valued Fuzzy Sets" WCCI 2010 IEEE World Congress on Computational Intelligence July, 18-23, 2010 - CCIB, Barcelona, Spain

[4]. Roberto R.F. Mendes et.al " Discovering Fuzzy Classification Rules with Genetic Programming and Co-evolution ", LNAI 2168, pp. 314-325, Springer-Verlag Berlin Heidelberg 2001

[5]. Martin Holena "Extraction Of Fuzzy Logic Rules From Data By Means Of Artificial Neural Networks"Kybernetika — Volume 41, Number 3, Pages 297-314 (2005)

[6]. Ujjwal Maulik and Sanghamitra Bandyopadhyay" Genetic algorithm-based clustering technique", Journal of Pattern Recognition Society ,published by Elsevier Vol 33 pp 1455-1465, (2000)

[7]. Sushmita Mitra, Rajat K. De, and Sankar K. Pal, Fellow, IEEE "Knowledge-Based Fuzzy MLP for Classification and Rule Generation " IEEE Transactions On Neural Networks", Vol. 8, No. 6,pp 1338-1350 November 1997

[8]. Johannes A. Roubos a, Magne Setnes b, and Janos Abonyi, "Learning fuzzy classification rules from labeled data " International Journal of Information Sciences published by Elsevier vol. 150 pp 77–93 (2003)

[9]. Gursewak S. Brar, Yadwinder S. Brar and Yaduvir Singh "An Efficient Data Clustering Algorithm Using Fuzzy Logic For Control Of A Multi Compressor System", IETECH Journal of Electrical Analysis, Vol: 1, No: 2,pp 130-136, 2007

[10]. Mofreh A. Hogo, "Evaluation of E-Learners Behavior using Different Fuzzy Clustering Models: A Comparative Study", (IJSIS)International Journal of Computer Science and Information Security Vol.7 No.2 pp: 131-140 (2010)

[11]. Suchita Borkar and K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining" IJCSMC International Journal of Computer Science and Mobile Computing, ISSN 2320–088X , Vol. 2, Issue. 7, pg.273 – 279 (2013)

[12]. Tzu-Ping Wu and Shyi-Ming Chen, "A new method for constructing membership functions and fuzzy rules from training examples", Systems, Man, and Cybernetics, Part

B: Cybernetics, IEEE Transactions Vol:29 , Issue: 1 pp 25 – 40, Feb 1999
[13]. Cesar Vialardi, Javier Bravo, Leila Shafti, and Alvaro Ortigosa, "Recommendation in Higher Education Using Data Mining Techniques", Proceedings Educational Data Mining EDM 2009, pg 190-199 , July 2009
[14]. Hisao Ishibuch , Ken Nozaki, Naohisa Yamamoto and Hideo Tanaka, "Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms", Fuzzy Sets and Systems published by Elsevier vol 65 pp 237-253 (1994)
[15]. Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C," Application of k-Means Clustering algorithm for prediction of Students' Academic Performance", (IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 1,pp: 292-295, 2010
[16]. Masoud Makrehchia, Mohamed S. Kamelb, "An information theoretic approach to generating fuzzy hypercubes for if-then classifiers", Journal of Intelligent & Fuzzy Systems vol. no 22 pp: 33–52 (2011)
[17]. Oscar Cordón, Francisco Herrera, and Pedro Villar "Generating the Knowledge Base of a Fuzzy Rule-Based System by the Genetic Learning of the Data Base ", IEEE Transactions On Fuzzy Systems, Vol. 9, No. 4, Pp 667 - 674 ,August 2001
[18]. Shruti S. Jamsandekar and Ravindra R. Mudholkar, "Self generated fuzzy membership function using ANN clustering technique", International Journal of Latest Trends in Engineering and Technology Special issue IDEAS 2013, pp 142-152 (2013)
[19]. MATLAB Fuzzy Logic toolbox 2 user manual

APPENDIX A

seat no	Subject 1 marks	Subject 2 marks
1	58	60
2	58	52
3	51	49
4	53	50
5	46	51
6	49	47
7	54	56
8	55	59
9	59	52
10	61	67
11	54	57
12	57	58
13	70	71
14	56	60
15	58	53
16	29	20
17	48	32
18	58	60
19	51	36
20	34	24
21	73	81
22	49	56
23	76	62
24	53	59
25	46	52
26	45	45
27	59	60
28	45	48
29	60	52
30	55	56
31	57	49
32	81	71
33	72	64
34	61	56
35	30	31
36	63	57
37	65	47
38	54	44
39	70	57
40	50	43
41	58	57
42	61	60
43	50	50
44	50	44
45	41	44
46	54	59
47	73	60
48	58	45
49	65	51
50	51	29
51	56	50

APPENDIX B

seat no	fuzzy class using FCMFC	fuzzy class using expert domain knowledge for membership function
1	69.62347309	69.41016566
2	69.62347309	69.41016566
3	55.18495127	54.1697685
4	61.00549807	61.01374256
5	40.73463047	36.20670391
6	47.38790839	44.875
7	63.05024746	63.32084691
8	64.80762548	65.2206443
9	70.15167421	70.56240231
10	70.38158258	72.85905554
11	63.05024746	63.32084691
12	67.98971257	68.19705015
13	**70.23469398**	**85.08755154**
14	66.41664216	66.78798403
15	69.62347309	69.41016566
16	33.6341888	13.83333333
17	44.98225048	41.21428571
18	69.62347309	69.41016566
19	55.18495127	54.1697685
20	**33.68989073**	**20.1010453**
21	**70.41632333**	**85.1147541**
22	47.38790839	44.875
23	**70.61907508**	**85.1147541**
24	61.00549807	61.01374256
25	40.73463047	36.20670391
26	38.52894557	34.38131313
27	70.15167421	70.56240231
28	38.52894557	34.38131313
29	70.26044891	71.58372067
30	64.80762548	65.2206443
31	67.98971257	68.19705015
32	**72.87273016**	**85.1147541**
33	**70.35233429**	**85.1147541**
34	70.38158258	72.85905554
35	**33.65781565**	**13.83333333**
36	70.66148531	75.40150129
37	70.56933154	78.05734074
38	63.05024746	63.32084691
39	**70.23469398**	**85.08755154**
40	50.23745353	50
41	69.62347309	69.41016566
42	70.38158258	72.85905554
43	50.23745353	50
44	50.23745353	50
45	**33.62465554**	**29.21682243**
46	63.05024746	63.32084691
47	**70.41632333**	**85.1147541**
48	69.62347309	69.41016566
49	70.56933154	78.05734074
50	55.18495127	54.1697685
51	66.41664216	66.78798403