ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 1, pp: (82-90), Month: January - March 2016, Available at: www.researchpublish.com

Decision Making In Fuzzy Logic Environment Using ID3 Decision Tree Algorithm

Nirlipta Pallabee Samapika[1], Aparimita Swain [2]

[1]Lecturer, [1] Department of Computer Science Engineering, [1] Hi-Tech College of Engineering, Bhubaneswar, India
[2]M.Tech Scholar, [2]School of Computer Engineering, KIIT University, Bhubaneswar, India

Abstract: **This paper presents the classification technique of data mining to identify the class of an attribute with classical decision tree approach (ID3) and then to add fuzzification to improve the result of ID3. This ID3 algorithm has been implemented on weather dataset due to its easiness to use and effectiveness. ID3 algorithm is based on information gain theory and entropy values of each individual attribute. Fuzzy set theory is implemented to represent the dataset with linguistic variable and combines tree growing and cropping to determine the structure of the tree. ID3 shows better accuracy in terms of discrete value. Fuzzy logic resolves this issue for better classification on decision making. Finally fuzzification results showed more accurate classification as compared to classical decision approach.**

Keywords: **ID3 (Iterative Dichotomizer), Decision Tree, Fuzzy logic.**

1. INTRODUCTION

1.1. DECISION TREE:

A decision tree is a workable model and classification scheme which generates a tree and a set of rules, that will predict the value of a target variable based on the set of input variables. The set of records available for developing classification methods is generally divided into two disjoint subsets-a *training set* and a *test set* [7]. Training sets are used for deriving the classifier, while test set is used for measuring the accuracy of the classifier. The accuracy of the classifier is determined by the percentage of the test examples that are correctly classified [2]. The attributes of the records are categorized into two different types. Attributes whose domain is numerical are called the *numerical attributes*, and the attributes whose domain is not numerical are called the *categorical attributes* [3]. There is one distinguished attribute called the class label. The major strengths of the decision tree methods are they are able to generate understandable rules having both numerical and categorical attributes. Decision trees also provide a clear indication of which fields are most important for prediction or classification. [1]

1.2 FUZZY RULE-BASED SYSTEMS (FRBS):

FRBS comprises of two main components Knowledge Base (KB) and Inference Engine (IE) [11].Knowledge can be represented through various ways. The most common way of representing the human knowledge is in the form of natural language expression. The KB consists of knowledge specific to the domain of application. It represents the knowledge about the problem being solved in the form of fuzzy linguistic IF-THEN rules. The IE uses the knowledge in the KB for performing suitable reasoning for user queries. The IE is needed to obtain an output from FRBS when an input is being specified [10]. This can be expressed in an expression in the form of IF-THEN rule based form like IF premise (Antecedent), THEN conclusion (Consequent) parameters [4]. The block diagram of an FRBS is shown in Fig. 1.

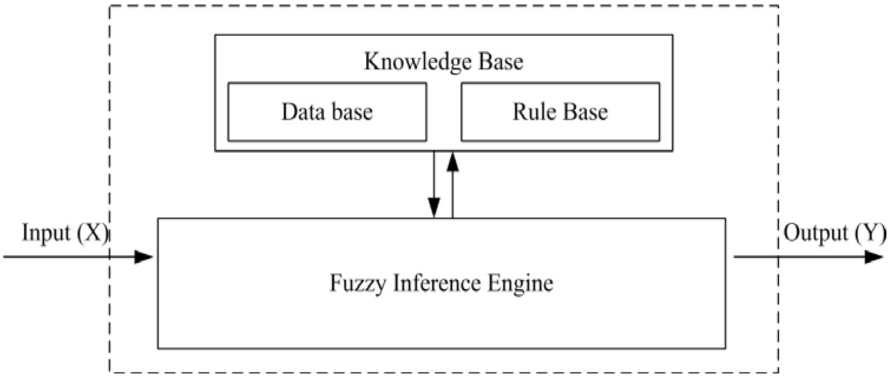


Fig.1. Block diagram of FRBS

A FRBS consists of three functional blocks, namely fuzzification ,inference, and defuzzification [4]. Fuzzification is the process, in which the input parameters or the crisp quantities are converted into appropriate fuzzy quantities to express measurement of uncertainty [14]. The fuzzified measurements are then used by inference engine to evaluate the control rules that are stored in the fuzzy rule base and finally fuzzified output is determined. The defuzzification unit converts the output into single crisp value [9].

2. ID3 ALGORITHM

It constructs an unpruned decision tree and only deals with nominal attribute without missing values. The basic idea employed in this algorithm is to construct a top-down, greedy search decision tree through the given set of test attributes [2]. The algorithm makes use of entropy which gives the information about the degree of doubt. The algorithm selects the attribute for classifying by comparing the information gain. [1] Three number of sets used in ID3, these are Learning sets(S), Attribute sets (A) and Attribute values (V). ID3 uses information gain to help it decide which attribute goes into a decision node. To define information gain we have to define entropy [3]. To describe ID3 algorithm, we use strategies (attribute) and decision to 'PLAY TENNIS'. The symbolic attribute description.

Table 1: Symbolic attributes representation

ATTRIBUTE	POSSIBLE VALUES
OUTLOOK	SUNNY, OVERCAST, RAIN
TEMPERATURE	HOT, MILD, COOL
HUMIDITY	HIGH, NORMAL
WINDY	TRUE, FALSE
DECISION	N(NEGATIVE), P(POSITIVE)

Table 2: Input-Output learning set (play tennis)

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	DECISION
SUNNY	HOT	HIGH	FALSE	N
SUNNY	HOT	HIGH	TRUE	N
OVERCAST	HOT	HIGH	FALSE	P
RAIN	MILD	HIGH	FALSE	P
RAIN	COOL	NORMAL	FALSE	P
RAIN	COOL	NORMAL	TRUE	N
OVERCAST	COOL	NORMAL	TRUE	P
SUNNY	MILD	HIGH	FALSE	N
SUNNY	COOL	NORMAL	FALSE	P
RAIN	MILD	NORMAL	FALSE	P
SUNNY	MILD	NORMAL	TRUE	P
OVERCAST	MILD	HIGH	TRUE	P
OVERCAST	HOT	NORMAL	FALSE	P
RAIN	MILD	HIGH	TRUE	N

3. METHODOLOGY

We all are affected by weather, having uncertainty and imprecision. It is completely different from the classical system. Fuzzy Logic can aim at modelling the imprecise models of reasoning and deal with the approximate rather than precise models [4]. Therefore fuzzy logic is chosen as a suitable method for weather nominal dataset. It was implemented in MATLAB 13. The weather parameters outlook, temperature, humidity, wind are used to predict whether with the above conditions play will be possible or not.

3.1 ENTROPY:

It is used to measure as how informative the node is. In ID3, entropy is calculated for each remaining attribute. The attribute with smallest entropy is used to split S on this iteration [1, 2, 3].The higher the entropy, the higher the potential to improve the classification.

$Entropy(S) =- P(positive)\log_2 P(positive)-P(negative)\log_2 P(negative)$ [1]

P(positive): proportion of positive examples in S

P(negative): proportion of negative examples in S

3.2 GAIN:

It is the measure of the difference in entropy from before to after, the set S is split on an attribute. The attribute with largest information gain is used to split the set S on iteration. The information gain, Gain(S,A) of an attribute A,

$Gain(S,A)= Entropy(S)$ - Sum for v from 1 to n of $(|S_v|/|S|) * Entropy(S_v)$ [2]

4. COMPUTATION

Entropy(Root Node Subset) = -(9/14)log2(9/14) -(5/14)log2(5/14)=0.940 Gain(S,Windy) = Entropy(S)-(8/14)Entropy (S_{false}) -(6/14)Entropy (S_{true})

Table 3: Decision value for windy input variable

WINDY	P(POSITIVE)	N(NEGATIVE)
FALSE	6	2
TRUE	3	3

$Entropy_{windy}(S)$=8/14*I(6,2)+6/14*I(3,3) I(6,2)=-6/8log2(6/8)-2/8log2(2/8)=0.811

I(3,3) =-3/6log2(3/6) - 3/6log2(3/6)=1

$Entropy_{windy}(S)$=8/14*0.811 +6/14*1=0.892

Gain(windy)= Entropy(S)- $Entropy_{windy}(S)$ =0.940-0.892=0.048

Similarly –

Gain(S, Humidity) =0.151

Gain(S, Temperature)=0.029 Gain(S, Outlook) = 0.246

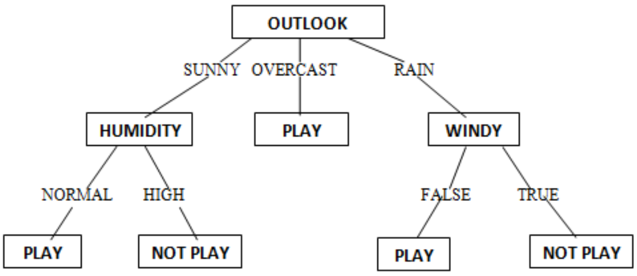


Fig .2. Decision based on ID3

5. WORKING PRINCIPLE OF FLC USING MAMDANI APPROACH

FLC consists of fuzzy control rules using linguistic variables. The fuzzy control rules are presented in the form of IF (a set of conditions are satisfied) THEN (a set of consequences can be prepared). Both the antecedents and consequents of the IF-THEN rules are represented using some linguistic terms. The inputs of FLC should be given by fuzzy sets, and therefore, the crisp inputs should be fuzzified. The output of an FLC is always fuzzy set and to get the corresponding crisp value the defuzzification method is used.

The fuzzification of input variables involves the following steps [11]:

- Measure all the input variables.

- Fixed and uniform mapping of the input variables are done so that the ranges of the input variable are transferred into corresponding universe of discourse.

- Fuzzification process is performed that converts the input data into corresponding linguistic values, which may be viewed as label of fuzzy sets.

Fuzzy linguistic approach provides a systematic way of representing linguistic variables in a natural evaluation procedure [16]. A fuzzy linguistic label can be represented by a fuzzy number, which is represented by a fuzzy set [11].Sets capture the ability to handle uncertainty by approximation methods [16].For the inputs and output, triangular membership functions were used in order to keep the design of the FLCs simple. A degree of overlapping of two was used, as shown in Figure. Furthermore, a universe of discourse normalized to the range of [0.0, 1.0] was utilized.

Table 4: Parameter ranges of variable Outlook

Linguistic terms	Membership function	Range of the parameter
Rain	Trimf	[0.0, 0.4]
Overcast	Trimf	[0.3, 0.7]
Sunny	Trimf	[0.6, 1.0]

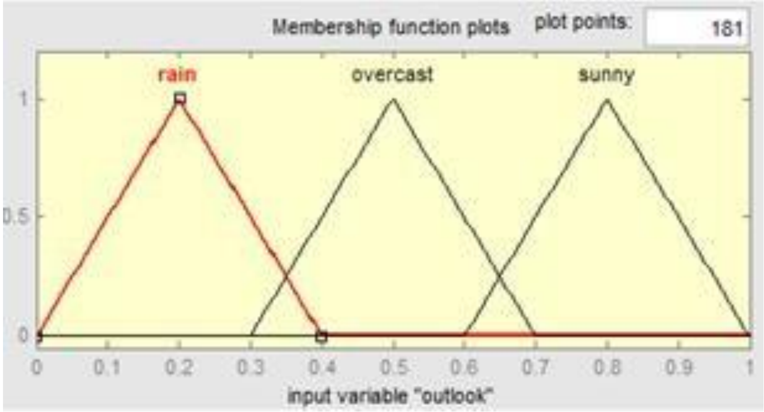


Fig. 3.Membership function distributions for OUTLOOK variable: V1 = {rain},V2={overcast},V3 = {sunny}

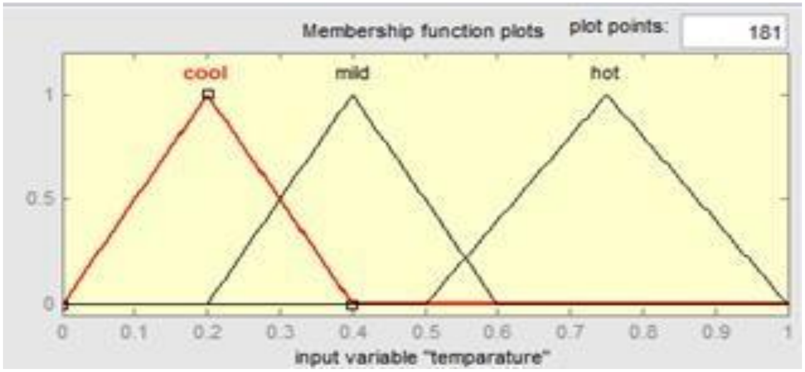


Fig.4. Membership function distributions for

TEMPERATURE variable: V_1={Cool},V_2={Mild},V_3={Hot}.

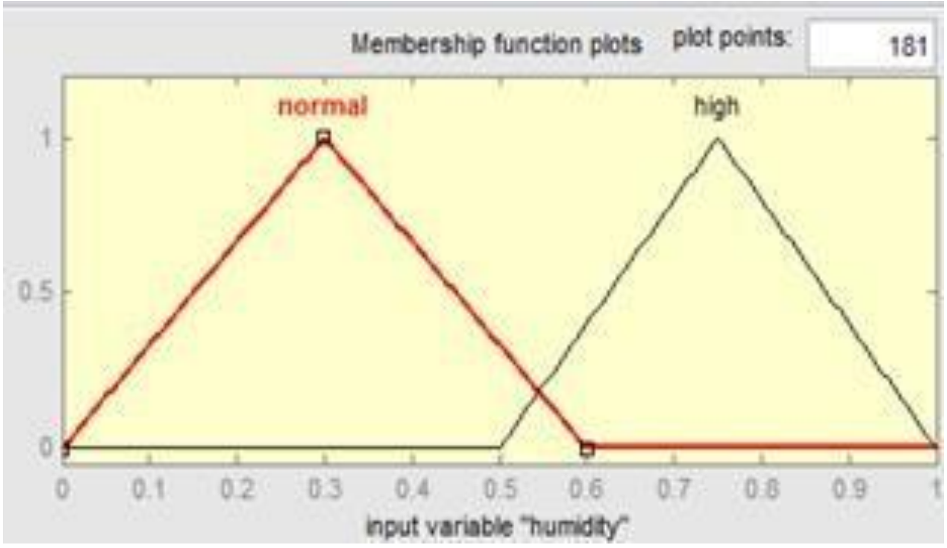


Fig.5. Membership function distributions for HUMIDITY variable: *V*1 = {normal}, *V*2 = {high}

Table 5 : Parameter ranges of variable Temperature

Linguistic terms	Membership function	Range of the parameter
Cool	Trimf	[0.0, 0.4]
Mild	Trimf	[0.2, 0.6]
Hot	Trimf	[0.5, 1.0]

Table 6: Parameter ranges of variable Humidity

Linguistic terms	Membership function	Range of the parameter
Normal	Trimf	[0.0, 0.6]
High	Trimf	[0.5, 1.0]

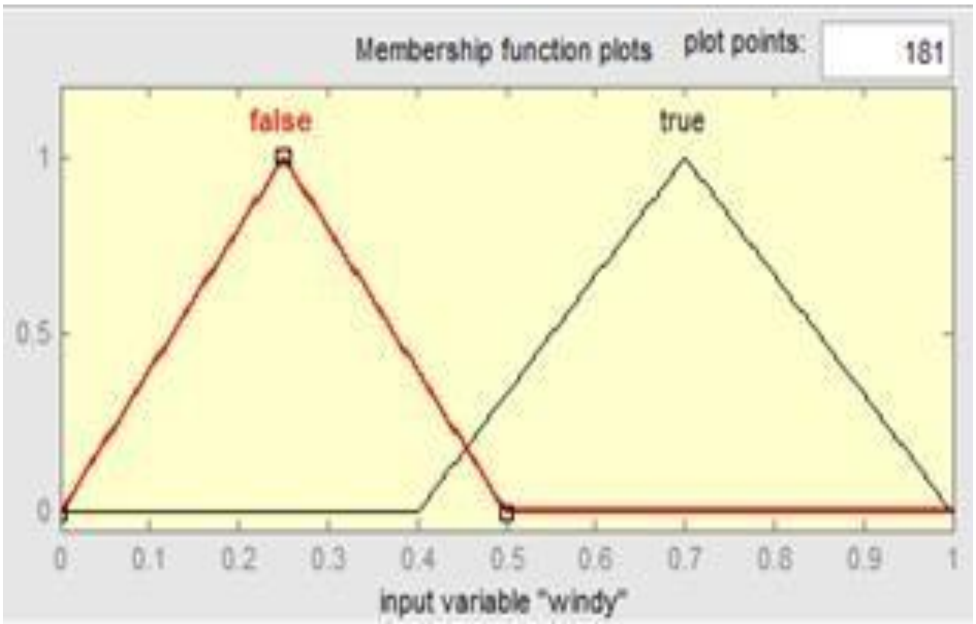


Fig.6. Membership function distributions for WINDY variable: *V*1 = {false}, *V*2 = {true}

Table 7: Parameter ranges of variable Windy

Linguistic terms	Membership function	Range of the parameter
False	Trimf	[0.0, 0.5]
True	Trimf	[0.4, 1.0]

Two linguistic terms, namely play and not play were used to represent the output variable. The Mamdani min-operator was utilized for aggregation and defuzzification was done using the center of sums (COS) method [17].

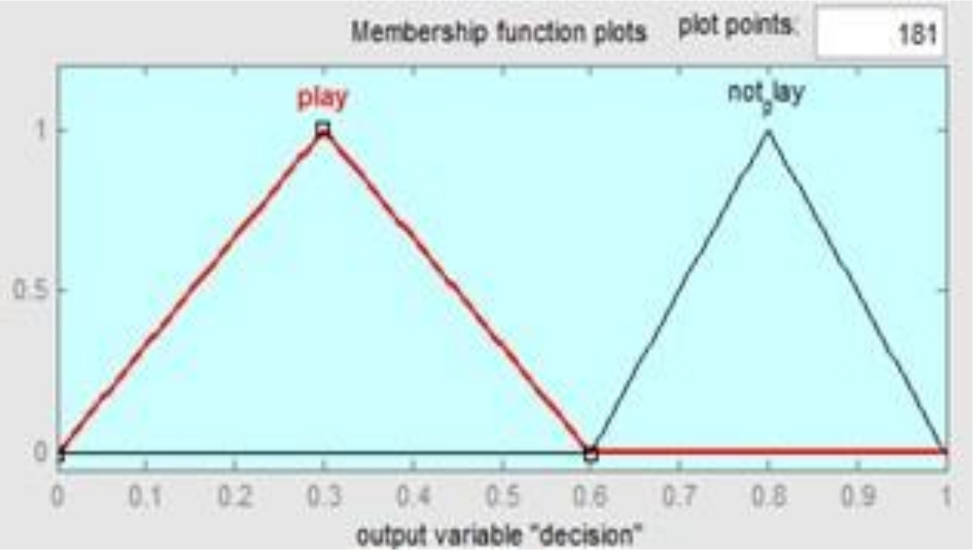


Fig.7. Membership function distributions for DECISION (output) fuzzy variable: *V1* = {play}, *V2* = {not play}

Table 8: Parameter ranges of variable Decision

Linguistic terms	Membership function	Range of the parameter
Play	Trimf	[0.0, 0.6]
Not play	Trimf	[0.6, 1.0]

5.1 DETERMINING FUZZY RULE BASE FROM INPUT AND OUTPUT VARIABLES:

Rules are the fundamental to FRBS. Rules establishes a relation between the input and output [4]. In the present problem, four input variables were considered and they were represented by a total of ten linguistic terms. Thus, there could be a maximum of rules in the FRBS. Considering the above example we can construct 36 rules by taking OUTLOOK, TEMPERATURE, HUMIDITY, WINDY as input and DECISION as output.

These RULES are as follows:

RULE1: IF outlook is *sunny* AND temperature is *hot* AND humidity is *high* AND windy is *false* THEN decision is *not play*

RULE2: IF outlook is *overcast* AND temperature is *hot* AND humidity is *high* AND windy is *true* THEN decision is *play*

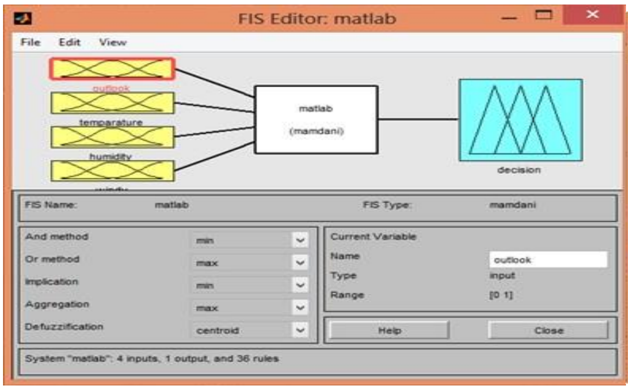
.

.

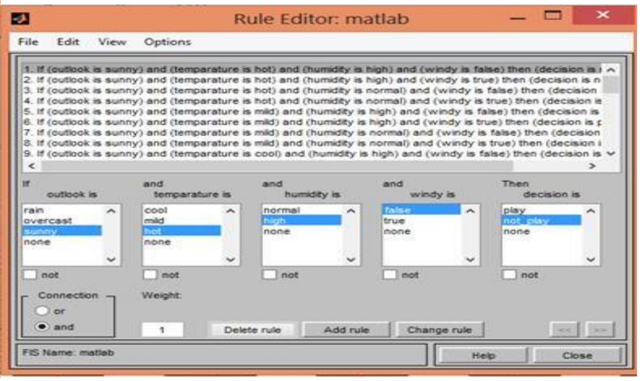
.

RULE36: IF outlook is *rain* AND temperature is *cool* AND humidity is *normal* AND windy is *true* THEN decision is *not play*

After writing all the rules the diagrammatic representation of all the input and output parameters are shown as below-



FIS Name: matlab
FIS Type: mamdani
And method: min
Or method: max
Implication: min
Aggregation: max
Defuzzification: centroid
Current Variable
Name: outlook
Type: input
Range: [0 1]
System "matlab": 4 inputs, 1 output, and 36 rules

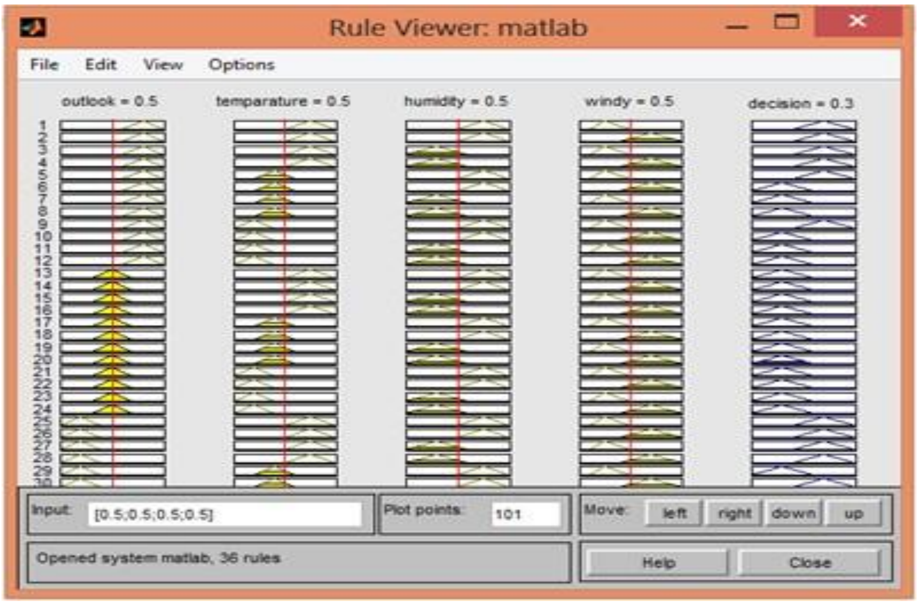


1. If (outlook is sunny) and (temparature is hot) and (humidity is high) and (windy is false) then (decision is
2. If (outlook is sunny) and (temparature is hot) and (humidity is high) and (windy is true) then (decision is n
3. If (outlook is sunny) and (temparature is hot) and (humidity is normal) and (windy is false) then (decision
4. If (outlook is sunny) and (temparature is hot) and (humidity is normal) and (windy is true) then (decision is
5. If (outlook is sunny) and (temparature is mild) and (humidity is high) and (windy is false) then (decision is
6. If (outlook is sunny) and (temparature is mild) and (humidity is high) and (windy is true) then (decision is p
7. If (outlook is sunny) and (temparature is mild) and (humidity is normal) and (windy is false) then (decision
8. If (outlook is sunny) and (temparature is mild) and (humidity is normal) and (windy is true) then (decision i
9. If (outlook is sunny) and (temparature is cool) and (humidity is high) and (windy is false) then (decision is

If outlook is: rain, overcast, sunny (selected), none — □ not
and temparature is: cool, mild, hot (selected), none — □ not
and humidity is: normal, high (selected), none — □ not
and windy is: false (selected), true, none — □ not
Then decision is: play, not_play (selected), none — □ not

Connection: ○ or ● and
Weight: 1
FIS Name: matlab

Fig.8. 4input 1 output model using FLC **Fig.9. Firing fuzzy rules**

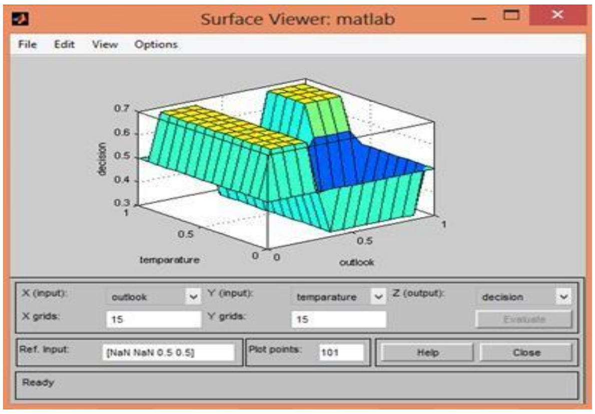


Input: [0.5;0.5;0.5;0.5]
Plot points: 101
Opened system matlab, 36 rules

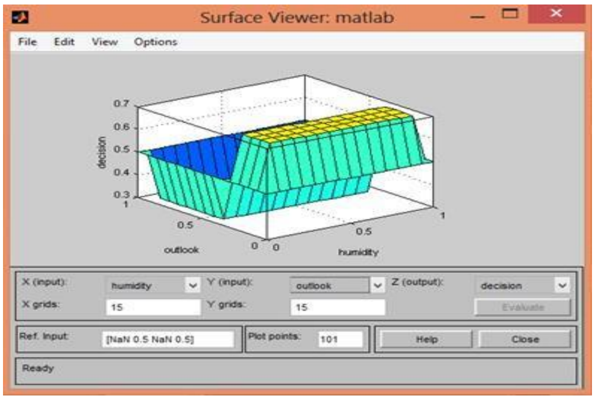
Fig.10.Input output relationship on firing the rules in rule base

5.2 Results and Discussion:

The performance of traditional fuzzy logic based on Mamdani approach showed better accuracy on classification. The results are shown in graphically in Fig 11 and Fig 12.When all the parameters are at 0.5, then decision value is at 0.3 i.e. not *play*. By keeping 3 parameters constant if any one parameter will change then also decision value will not be affected. On variations of temperature and outlook, keeping humidity and windy as constant value, then decision will remain constant. When windy and outlook are changed, keeping humidity and temperature as constant, then decision will remain constant. It is also observed that during this study when temperature and humidity are high then decision is not play.

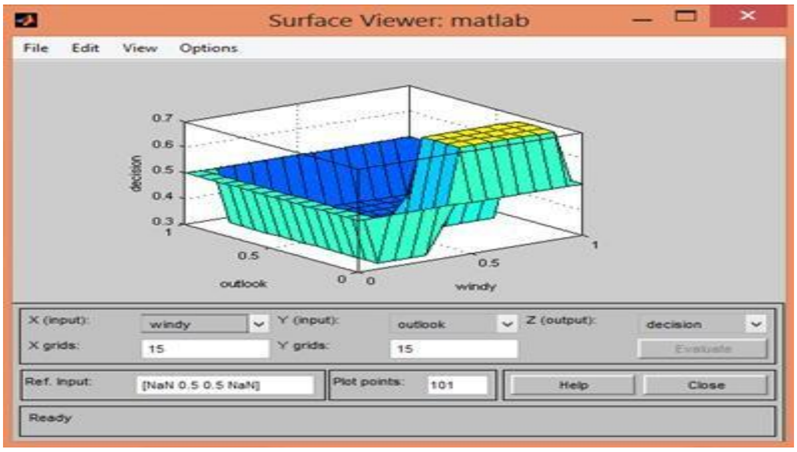


X (input): outlook
Y (input): temparature
Z (output): decision
X grids: 15
Y grids: 15
Ref. Input: [NaN NaN 0.5 0.5]
Plot points: 101
Ready



X (input): humidity
Y (input): outlook
Z (output): decision
X grids: 15
Y grids: 15
Ref. Input: [NaN 0.5 NaN 0.5]
Plot points: 101
Ready

Fig11. Variation in temperature & outlook **Fig.12. Variation in outlook & humidity**



X (input): windy
Y (input): outlook
Z (output): decision
X grids: 15
Y grids: 15
Ref. Input: [NaN 0.5 0.5 NaN]
Plot points: 101
Ready

Fig.13.Variation in windy & outlook

6. CONCLUSION

In this study, we have focused upon ID3 decision tree learning algorithm with fuzzy logic. Mamdani approach was used in order to determine input–output relationship of this process. We conclude that when fuzzy linguistic variable contaminated with attribute of input parameters it gives better results that of general ID3. However there are some problem in classical decision tree like how to select best criteria to split the training and formulated rules.

REFERENCES

[1] Renuka D. Suryawanshi, D.M. Thakore.(2012), ‘Decision Tree Classification Implementation with Fuzzy Logic’, *IJCSNS International Journal of Computer Science and Network Security*,Vol.12 no.10.pp. 93-97.

[2] Peng Wei, Chen Juhua and Zhou Haiping, ‘An Implementation of ID3 --- Decision Tree Learning Algorithm’,*Project of Comp 9417: Machine Learning* University of New South Wales, School of Computer Science & Engineering, Sydney, NSW 2032, Australia.

[3] HanJ. and Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000.

[4] Neuro Fuzzy and Soft Computing by J.S.R Jang, C.T. Sun, E. Mizutani,PHI, 2009.

[5] E. UtgoffPaul and Brodley Carla E. (1990), 'An Incremental Method for Finding Multivariate Splits for Decision Trees', *Machine Learning:Proceedings of the Seventh International Conference*.pp.58.

[6] Janikow C. Z., ‘Fuzzy Decision Forest.(2003),’ *Proc. of 22nd Int. Conf. of the North American Fuzzy Information Processing Society, Chicago*, pp.480-483.

[7] Quinlan J. R., ‘Decision Trees and Decision making’.(1990), *IEEE Trans. On Systems, Man and Cybernetics*, Vol. 20, Issue 2, pp. 339-346.

[8] Janikow C. Z.(2004), ‘FID4.1: an Overview’, IEEE Annual Meeting of the Fuzzy Information Processing, NAFIPS ’04, Vol. 2, pp. 877-881,

[9] Lee J., Sun J. and Yang L.(2003), ‘An Fuzzy Matching Method of Fuzzy Decision Trees’, *Int. Conf. On Machine Learning and Cybernetics*, Vol. 3, Issue 2, pp. 1569-1573,.

[10] Huang Z. and Shen Q., “Fuzzy Interpolative Reasoning via Scale and Move Transformations,” IEEE Trans. on Fuzzy Systems, Vol. 14, Issue 2, pp. 340-359, 2006.

[11] Mohanty, S.N. Pratihar. D and Suar, D.(2015), ‘Influence of Mood States on Information processing during decision making using fuzzy reasoning tool and neurons-fuzzy system based on Mamdani approach’, *International Journal of Fuzzy Computation and Modeling*, Vol.1, Issue 3, pp.251-268.

[12] Buckley, J.J. and Hayashi, Y. (1994) 'Fuzzy neural networks: a survey', *Fuzzy Sets and Systems*, Vol. 66, No. 1, pp.1–1.

[13] Principles of Soft Computing, Sivanandam, Deepa, Wiley India, 2011.

[14] Mamdani, E.H. and Assilian, S. (1975) 'An experiment in linguistic synthesis with a fuzzy logic controller', *Man-Machine Studies*, Vol. 7, No. 1, pp.1–13.

[15] Nauck, D. and Kruse, R. (1996) *Neuro-fuzzy Systems Research and Application Outside of Japan*, pp.108–134, Soft Computing Series, Asakura Publication, Tokyo.

[16] Pratihar, D.K. (2008) *Soft Computing*, Narosa Publishing House, New Delhi, India.