Decision Tree Based Fuzzy Reasoning

Thirumalaimuthu Thirumalaiappan Ramanathan

Faculty of Education, Science, Technology and Mathematics
University of Canberra, ACT, Australia
thirumalaimuthu.thirumalaiappanramanathan@canberra.edu.au

Dharmendra Sharma

Faculty of Education, Science, Technology and Mathematics
University of Canberra, ACT, Australia
dharmendra.sharma@canberra.edu.au

Abstract—*Fuzzy logic techniques are efficient in solving complex, ill-defined problems that are characterized by uncertainty of environment and fuzziness of information. Fuzzy logic allows handling uncertain and imprecise knowledge and provides a powerful framework for reasoning. Fuzzy reasoning models are relevant to a wide variety of subject areas such as engineering, economics, psychology, sociology, finance, and education. For most of these applications, fuzzy system or hybrid fuzzy system is developed to deal with complex data. This paper presents a new hybrid fuzzy system model which is named as decision tree based fuzzy expert model (DFEM). Decision tree learning is used to extract patterns from dataset and the extracted patterns are used in terms of rules in the fuzzy rule base. The proposed model is tested on a car evaluation dataset and it shows best results.*

Decision tree learning; fuzzy logic (keywords)

I. INTRODUCTION

This paper presents a hybrid model, DFEM where the decision tree [1] is used to design the rules in the fuzzy rule base. Fig. 1 show DFEM architecture. In a fuzzy expert system [2], the inputs are fuzzified into fuzzy sets. The Fuzzy sets are reasoned through the rules in the fuzzy rule base. Then the fuzzy output are converted into crisp output using defuzzification technique. DFEM architecture hybrids the fuzzy expert system design by using the decision tree learning for the rule base. The decision tree is trained using a dataset. For all possible combination of inputs, the trained decision tree is tested and the results are in the fuzzy rule base. The idea of DFEM is because the decision tree classification is directly based on the dataset used for its training. For any complex data like incomplete information, the classification performance could be improved if the decision is used in the inference engine of fuzzy system as the fuzzy system is able to deal with complex data through fuzzy reasoning.

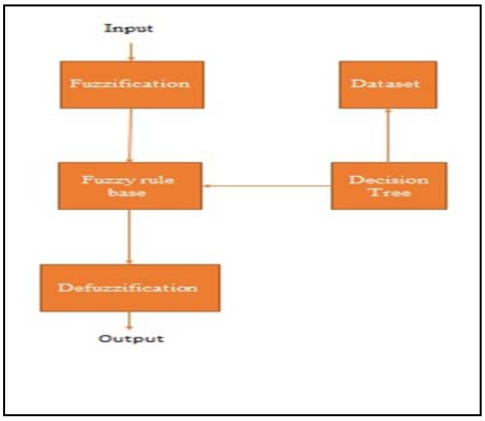


Figure 1. DFEM Architecture

II. Decision Tree

Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. In the decision tree structure, leas represent classifications and branches represent conjunction of features that lead to those classifications. The machine learning technique for inducing a decision tree classifier from data (training objects) is called decision tree learning. The main goal of classification is to build a model that can be used for prediction. In a classification problem, we are given a data set of training objects, each object having several attributes. There is one distinguished attribute called a decision class; it is a dependent attribute, whose value should be determined using the induced decision tree. The remaining attribute are used to determine the value of the decision class. Classification is thus a process of mapping instances represented by attribute-value vectors to decision classes [3]. In the DFEM, the decision tree is used to extract patterns from the dataset and its classification results are used as rules in the fuzzy rule base.

III. FUZZY LOGIC

Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth: truth values between “completely true” and “completely false”. It was proposed by [4] to deal with the uncertainty of natural language. Fuzzy set theory can be used to model through approximation the vaguely defined classes or categories. DFEM combines fuzzy reasoning and decision tree learning approach for classifying the input data.

IV. DATASET

Car evaluation dataset used for the classification is a benchmark dataset from [5]. It contains 1728 instances. The input attributes are buying price, maintenance price, number of doors, capacity in terms of persons to carry, size of luggage boot, estimated safety of the car. The output classes are unacceptable, acceptable, good, very-good. For the DFEM, only three inputs: buying price, maintenance price and estimated safety are used. The four output classes are made into two output classes: unacceptable or acceptable. The outputs acceptable, good, very-good are considered as acceptable output class. The decision tree can group the dataset only into two groups, so the four output classes are made into two output classes.

The dataset is grouped into two datasets. One datasets contains 1678 data which is used for training the decision tree. The other dataset contains 50 data and is used to test the DFEM.

TABLE 1 SAMPLE DATASET [5]

Buying Price	Maintenance Price	Estimated Safety	Output
Low	Med	High	Acceptable
Med	High	Low	Unacceptable
Vhigh	Vhigh	Med	Unacceptable

Table 1 shows sample data from the car evaluation dataset. The inputs buying price and maintenance price have four possible data: low, medium, high and very high. The input estimated safety has three possible data: low, medium and high. The output is either acceptable or unacceptable.

Using this dataset, below are the questions that are asked from the model.

- *Can a hybrid decision tree-fuzzy system be designed to deal with uncertain data?*
- *Can such a hybrid system show good accuracy on classifying the dataset?*

V. DFEM ALGORITHM

Below shows the algorithm of DFEM.

Get Inputs ();

Create Fuzzy Sets ()

    Trapezoidal membership function ();

    Triangular membership function ();

Decision Tree ()

    Load Dataset ();

    Train Decision Tree ();

Inference engine()

For all possible combination of input fuzzy sets()

Decision Tree-Classification ();

Add Decision Tree-Results to Rule base ();

Fuzzy reasoning ();

Defuzzification ()

Centre of Gravity ();

Print Output ();

Mamdani's fuzzy inference method [6] is used to implement the DFEM. MATLAB is used for implementation. As per the above DFEM, the inputs are converted into fuzzy sets using trapezoidal and triangular membership function [7]. Decision is trained using the dataset. Then the trained decision tree is used to classify all possible combination of the input fuzzy sets. The decision tree results with the corresponding inputs are added to the rule base. The fuzzy rule base reasons the fuzzy sets based on these rules. The fuzzy output is then defuzzified using centre of gravity (COG) [8] method.

A. Fuzzy Sets

The triangular and trapezoidal membership functions are used to design the input and output fuzzy sets. The triangular and trapezoidal membership functions are described below [9]:

The trapezoidal curve is a function of a vector, x, and depends on four scalar parameters *a*, *b*, *c*, and *d*, as given by

$$f(x;a,b,c,d) = \max\left(\min\left(\frac{x-a}{b-a},1,\frac{d-x}{d-c}\right),o\right)$$

The triangular curve is a function of a vector, x, and depends on three scalar parameters *a*, *b*, and *c*, as given by

$$f(x;a,b,c) = \max\left(\min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right),o\right)$$

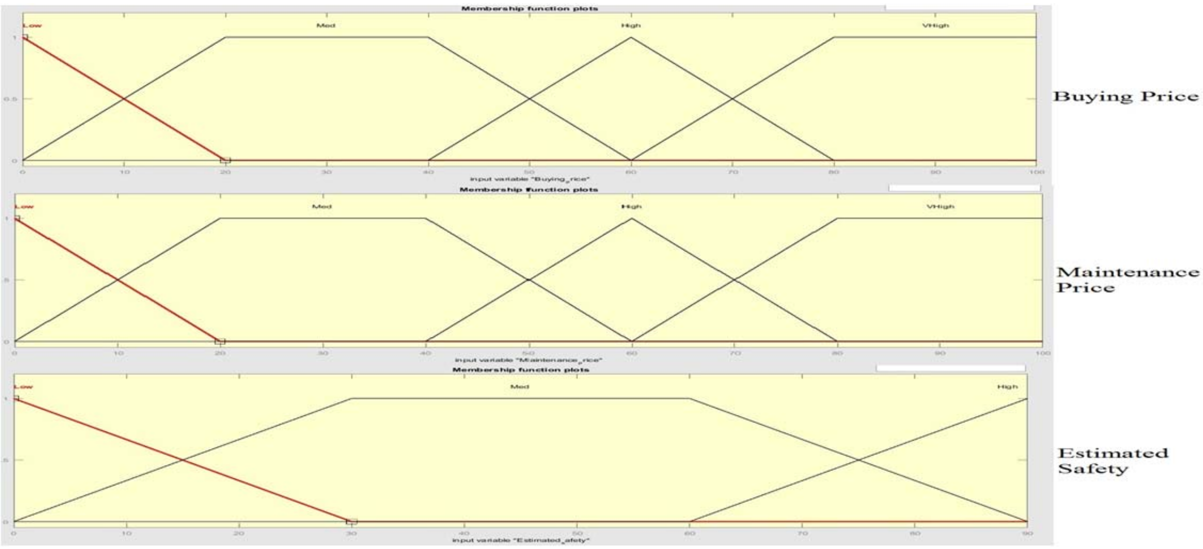
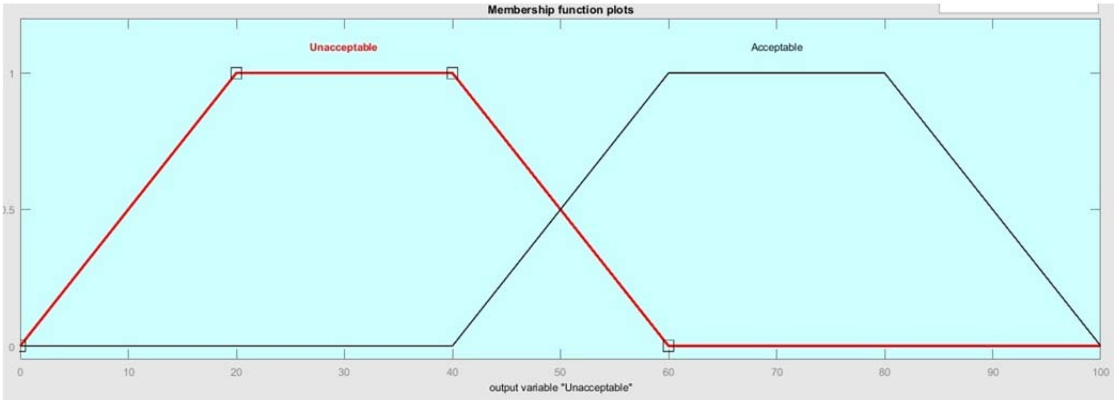


Figure 2. Input Fuzzy Sets

Fig. 2 shows the fuzzy sets of inputs: buying price, maintenance price and estimated safety of the car. The buying price and maintenance price inputs have four fuzzy sets: Low, Medium, High, VHigh. The estimated safety input has three fuzzy sets: Low, Medium, High. The ranges for the input is chosen randomly and equally divided for the fuzzy sets. For example, the input range for buying price input is chosen randomly as [0 100] and equally divided for the four fuzzy sets Low, Medium, High and VHigh using triangular and trapezoidal membership function.



Output

Figure 3. Output Fuzzy Sets

Fig. 3 shows the output fuzzy set. The output has two fuzzy sets: unacceptable, acceptable, which is designed using trapezoidal membership function. Similar to the input fuzzy sets, the output range is chosen randomly and equally divided for the two output fuzzy sets.

B. Fuzzy Rules

In the fuzzy rule base, as discussed in the previous sections, decision tree used to the design the rules. Fig. 4 shows graph of decision tree learning trained using the dataset.

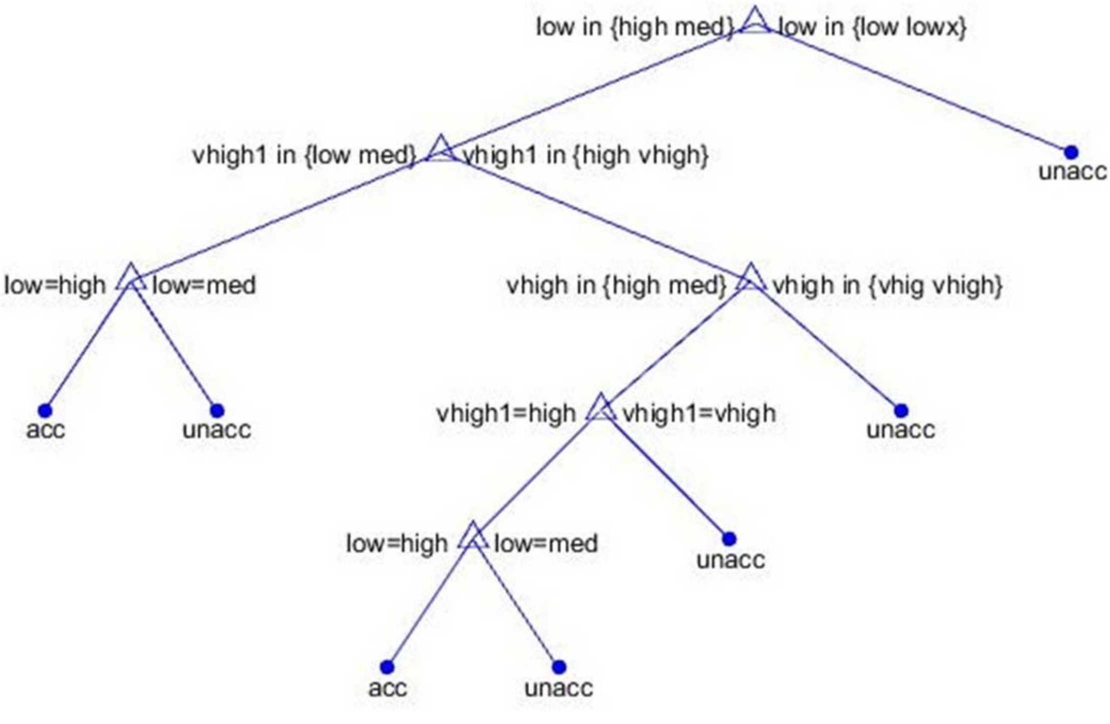


Figure 4. Trained Decision Tree

The trained decision tree is used to test for all possible combination of inputs and the results are added as rules in the fuzzy rule base. There are 48 possible combinations of inputs, and so there are 48 rules.

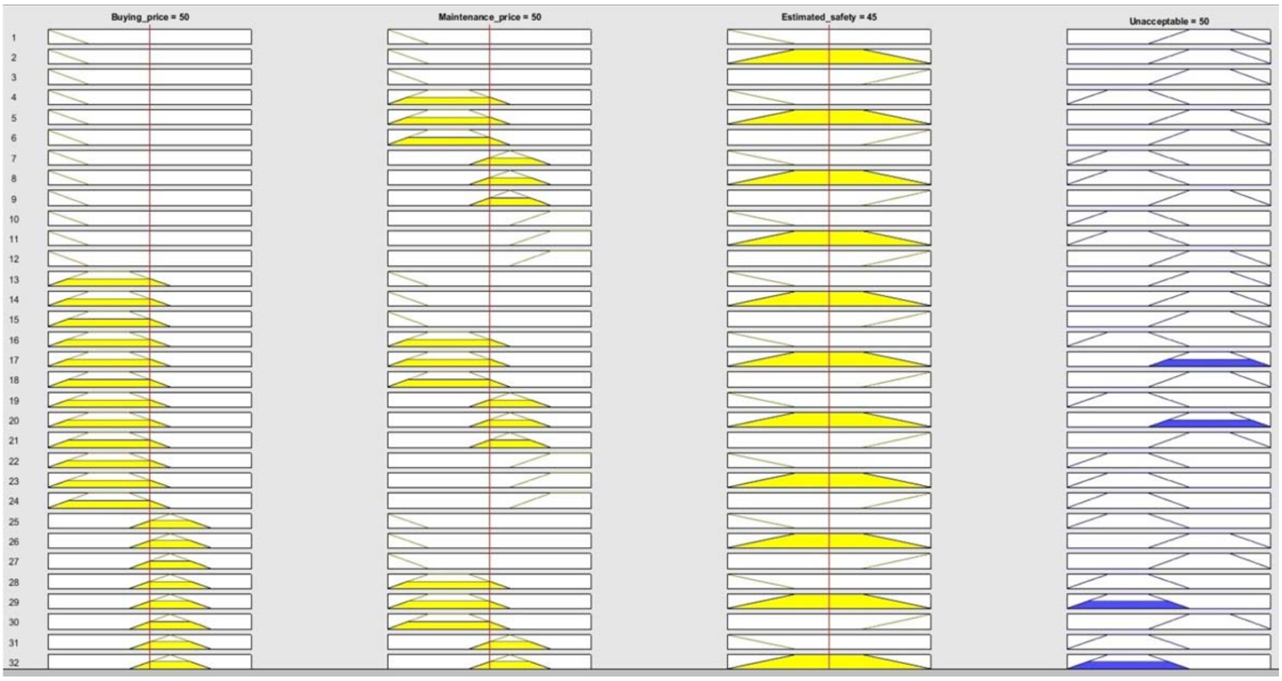


Figure 5. Sample Fuzzy Rule Viewer

Fig. 5 shows sample fuzzy rule base containing rule 1 to 32. In the Fig. 5, the first, second, third columns are inputs buying price, maintenance price, estimated safety respectively and the last column is the output.

C. Defuzzification

The COG method is used to convert the fuzzy output into crisp output. In the COG method as described in [10], the fuzzy system first calculates the area under the scaled membership function and within the range of the output variable. The fuzzy logic system then uses the following equation to calculate the geometric center of this area.

$$CoA = \frac{\int_{x_{min}}^{x_{max}} f(x) * x \, dx}{\int_{x_{min}}^{x_{max}} f(x) dx}$$

where CoA is the center of area, x is the value of the linguistic variable, and x_{min} and x_{max} represent the range of the linguistic variable. The COG defuzzification method effectively calculates the best compromise between multiple output linguistic terms.

VI. *PERFORMANCE OF DFEM*

The performance of DFEM is estimated using classification accuracy. The classification accuracy is calculated using true positive (TP), false positive (FP) [11] and using true negative (TN), false negative (FN) [12].

Classification Accuracy = (TP+TN) / (TP+TN+FP+FN).

TP: These are the positive tuples that were correctly labelled by the classifier. If outcome from a prediction is p and the actual value is also p, then it is called TP.

TN: These are the negative tuples that were correctly labelled by the classifier.

FP: These are the negative tuples that were incorrectly labelled by the classifier. However if the actual value is in then it is said to be false positive.

FN: These are the positive tuples that were mislabelled as negative.

The testing dataset is used to test the system. The system outputs were compared with the output class listed in the testing dataset. DFEM showed classification accuracy of 96 %.

VII. CONCLUSION AND FUTURE WORK

The hybrid fuzzy model, DFEM which is a combined approach using fuzzy logic and decision tree learning showed a best performance on classification. DFEM is able to give better accuracy on classification when compared to neuro-fuzzy [13]. The model will be improved to deal with numerical dataset in future.

REFERENCES

[1] Olshen, R.A. and Stone, C.J., 1984. Classification and regression trees. Monterey, CA: Wadsworth & Brooks.Cole Advanced Books & Software. ISBN 978-0-412-04841-8.
[2] Kandel, A., 1991. Fuzzy expert systems. CRC press.
[3] Podgorelec, V. and Zorman, M., 2009. Decision TreesDecision tree (pp. 1826-1845). Springer New York.
[4] Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and systems, 1(1), pp.3-28.
[5] Bohanec, M. and Zupan, B., 1997. UCI Machine Learning Repository.HYPERLINK" https://archive. ics. uci. edu/ml/datasets/Car+ Evaluation" https://archive. ics. uci. edu/ml/datasets/Car+ Evaluation.
[6] Mamdani, E.H. and Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. International journal of man-machine studies,7(1), pp.1-13.
[7] Zadeh, L.A., 1965. Fuzzy sets. Information and control, 8(3), pp.338-353.
[8] Van Leekwijck, W. and Kerre, E.E., 1999. Defuzzification: criteria and classification. Fuzzy sets and systems, 108(2), pp.159-178.
[9] MathWorks, Inc and Wang, W.C., 1998. Fuzzy Logic Toolbox: for Use with MATLAB: User's Guide. Mathworks, Incorporated.
[10] Chinthamani, M.B., Mol, M.P.R., Kamini, M.K. and Sughashini, M.K., 2012. Fuzzy based control using lab view for MISO temperature process.
[11] Sapna Jain, M., Afshar Aalam and Doja, M.N,, 2010. K-Means clustering using Weka Interface. Proceedings of the 4th National Conference; INDIA Com-2010, Computing For Nation Development.
[12] Han, J., Kamber, M. and Pei, J., 2011. Data mining: concepts and techniques. Elsevier.
[13] Serhatlioglu, S., Bozgeyik, Z., Ozkan, Y., Hardalac, F. and Güler, İ., 2003. Neurofuzzy classification of the effect of diabetes mellitus on carotid artery.Journal of medical systems, 27(5), pp.457-464.