ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

FUZZY RULE EXTRACTION FOR FRUIT DATA CLASSIFICATION

S.V.S. Ganga Devi

Professor & Head, Department of MCA, K.S.R.M. College of Engineering, Kadapa – 516 003 (A.P.)
Email: svsgangadevi_06@yahoo.co.in

ABSTRACT: Decision Tree algorithms provide one of the most popular methodologies for symbolic knowledge acquisition. The resulting knowledge, a symbolic Decision Tree along with a simple inference mechanism, has been praised for comprehensibility. The most comprehensible Decision Trees have been designed and then rules are extracted for perfect symbolic data. Over the years, additional methodologies have been investigated and proposed to deal with continuous or multi-valued data and with missing or noisy features. Recently, with the growing popularity of fuzzy representation in Decision Trees are introduced to deal with similar situations. Fuzzy representation bridges the gap between symbolic and non-symbolic data by linking qualitative linguistic terms with quantitative data. In this paper first Fuzzy Decision Tree for Fruit data classification is constructed and then the fuzzy classification rules are extracted.

Keywords: Classification, Decision Tree, Fuzzy ID3, Fuzzy rules.

1. INTRODUCTION

With the increasing amount of data readily available, automatic knowledge acquisition capabilities are becoming more and more important. When all data elements pre-classified, no extensive domain knowledge is available, and the objective is to acquire knowledge in the form of classification rules describing those classes (simply to classify future data), the knowledge acquisition process is called Supervised Learning from examples. Decision Trees are one of the most popular methods for learning and reasoning for feature based examples. [7]. However, they have been criticized for their persistent over-reliance on near perfect data and for the resulting degradation in the presence of imperfect data. Data imperfection might have been the result of noise, imprecise measurements, subjective evaluations, inadequate descriptive language or simply missing data. Additional problems arise from a continuous or simply large nominal attributes, all such domains have to be partitioned. Quinlan has proposed some methods for dealing missing features both in trained data and in the examples to be classified [8]. Continuous domains have been addressed by CART [1] and subsequently by C4.5 [6] algorithms, with the usual approach being to use threshold values to split a domain. However, the resulting knowledge generally exhibits lower comprehensibility and overspecialization. These problems were in turn addressed by True pruning techniques [6]. To overcome the problem, some researchers, suggested Fuzzy Decision Tree from data [2] [3] [4] [5] [9] [10] [11] [12]. It is capable of processing a mixture of symbolic, numeric and fuzzy termed data. However, all domains must be partitioned into fuzzy sets defined by the user.

This paper is organized as follows. Section 2 introduces the Fuzzy Decision Tree, Section 3 deals with Fuzzy ID3 algorithm to construct Fuzzy Decision Tree. Section 4 deals with Experiment results. Finally, Section 5 draws the conclusion.

2. FUZZY DECISION TREE

In classification the goal in to build or find a model in order to predict the category of data based on some predictor variables. A Decision Tree is a representation of the classification rules. Fuzzy Decision Trees provide a more robust way to avoid misclassification. A Fuzzy Decision Tree consists of nodes for testing attributes, edges for branching by values of fuzzy sets defined by user and leaves for deciding class names with certainties. In Decision Tree, an instance can be classified into an affirmative class, but in Fuzzy Decision Tree an instance may be classified into different classes with different membership grades.

3. FUZZY ID3

In the beginning, Fuzzy ID3 is only an extension of the ID3 algorithm achieved by applying a fuzzy set of data (several data with membership grades). It generates a Fuzzy Decision Tree using fuzzy sets defined by a user for all attributes.

Assume that we have a set of data D, where each data has l numerical values for attributes A_1, A_2, …… A_l and one class variable C = {C_1, C_2, …..C_n} and fuzzy sets F_{i1}, F_{i2},……..F_{im} for the attribute A_i (the values of m varies for every attribute). Let D^{C_k} to be a fuzzy subset in D whose class is C_k. Let |D| denote the sum of membership values in a fuzzy set of data D. The information gain $G(A_i, D)$ for the attribute A_i by a fuzzy set of data D is defined by,

$$G(A_i, D) = I(D) - E(A_i, D) \quad \ldots\ldots \quad (1)$$

Where

$$I(D) = -\sum_{k=1}^{n}\left(p_k . \log_2 p_k\right) \quad \ldots\ldots. \quad (2)$$

$$E(A_i, D) = \sum_{j=1}^{m}\left(p_{ij} . I\left(D_{F_{ij}}\right)\right) \quad \ldots\ldots. \quad (3)$$

$$p_k = \frac{\left|D^{C_k}\right|}{|D|} \quad \ldots\ldots.. \quad (4)$$

$$p_{ij} = \frac{\left|D_{Fij}\right|}{\sum_{j=1}^{m}\left|D_{F_{ij}}\right|} \quad \ldots\ldots.. \quad (5)$$

3.1 ALGORITHM

Fuzzy Decision Tree is generated by using the following steps.

1. Generate the root note that has a set of all data i.e., a fuzzy set of all data with membership value 1.
2. If a node t with a fuzzy set of data D satisfies the following conditions:
   (a) The proportion of a data set of a class C_k is greater than or equal to a threshold θ_r, that is $\frac{\left|D^{C_k}\right|}{|D|} \geq \theta_r$,
   (b) The number of a data set is less than a threshold θ_n that is $|D| < \theta_n$,
   (c) There are no attributes for more classification.
   Then it is a leaf node and assigned by the class name with possibilities.
3. If a node does not satisfy the above conditions, it is not a leaf node and the test node is generated as follows:
   (a) For A_i's (i=1, 2, …. l), calculate the information gains $G(A_i, D)$ and select the test attribute A_{max} which has the highest information gain.
   (b) Divide D into a fuzzy subsets D1, D2, …….Dm according to A_{max}, where the membership value of the data in D_j is the product of the membership value in D and the value of $F_{max,\ j}$ of the value of A_{max} in D.
   (c) Generate new nodes t_1, t_2,……t_m for fuzzy subsets D1, D2, …. Dm and label the fuzzy sets $F_{max,\ j}$ to edges that connect between the nodes t_j and t.
   (d) Replace D by Dj (j=1, 2,…..m) and repeat from 2 recursively.

4. EXPERIMENTAL RESULTS

In Fruit data, the attribute variables Xj (j=1,…,10) are considered as follows. X1 as number of days after flowering, X2 as weight of whole fruit in grams, X3 as weight of fruit after storage, X4 as penetrometer, X5 as solids, X6 as brix, X7 as glucose, X8 as fructose, X9 as sucrose and X10 as flavor. The class variable as acceptability (Not_acceptable,acceptable or Excellent).

A collection of learning patterns U = {1,…n} can be described by input attribute vector X=[x_1, x_2,……x_p]. x_j^i (i=1,….n) represents i^{th} pattern of attribute x_j. Each attribute xj has been fuzzified into 3 Fuzzy Sets Low, Medium, and High by using Fuzzy C-means Algorithm. To fuzzify input attributes, we select Gaussian membership functions out of many alternatives due to its differentiable nature. We calculate the membership degree of the i^{th} value of attribute x_j on the Fuzzy Set F_{jk} (j=1…..p, k=1,2,3) is given by,

$$\mu = \exp\left(- \frac{\left(x_j^i - C_{jk}\right)^2}{2\sigma_{jk}^2} \right)$$

Where C_{jk}, σ_{jk} are center and width of Gaussian membership function of j^{th} attribute on the k^{th} cluster. Then we get the fuzzy representation for the given input data. Now we run the fuzzy ID3 algorithm.

MATLAB implementation is done to construct the Fuzzy Decision Tree for Fruit data classification by using Fuzzy ID3 algorithm.

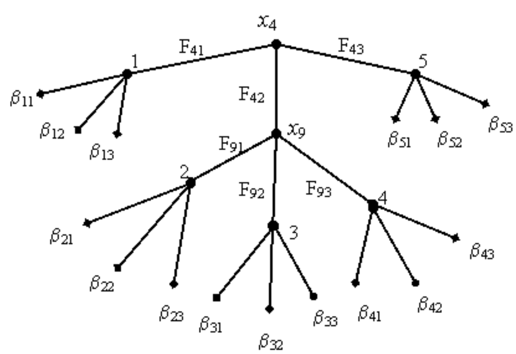

Fig.1 Fuzzy Decision Tree

Fig.1 shows the Fuzzy Decision Tree(FDT) for Fruit data. It consists of 2 internal nodes and five leaf nodes. Certainty factors for the three classes associated with all the five leaf nodes are

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \\ \beta_{41} & \beta_{42} & \beta_{43} \\ \beta_{51} & \beta_{52} & \beta_{53} \end{bmatrix} = \begin{bmatrix} 0.83 & 0.03 & 0.14 \\ 0.01 & 0.97 & 0.02 \\ 0.84 & 0.06 & 0.10 \\ 0.01 & 0.96 & 0.03 \\ 0.97 & 0.02 & 0.01 \end{bmatrix}$$

Classification accuracy of this FDT on Trained Fruit data is 81% and on Test data, it is 79.1%. Size of FDT is 7. Number of leaf nodes are 5. Time taken to build the model is 0.4 secs.

4.1 FUZZY CLASSIFICATION RULES

From Fuzzy Decision Tree, human interpretable fuzzy Classification rules are extracted in the form "If $path_m$ then $leaf_m$", where m is the number of leaf nodes. These rules are clear and easy to understand for all the Agricultural people. As a result they know which attributes are important.

Five fuzzy rules can be extracted from the Fig1.

- If (x_4 is F_{41}) then y = Not_acceptable (β_{11}) , y = Acceptable (β_{12}) and y = Excellent(β_{13}).
- If (x_4 is F_{42} ^ x_9 is F_{91}) then y = Not_acceptable (β_{21}) , y = Acceptable (β_{22}) and y = Excellent(β_{23})
- If (x_4 is F_{42} ^ x_9 is F_{92}) then y = Not_acceptable (β_{31}) , y = Acceptable (β_{32}) and y = (β_{33}).
- If (x_4 is F_{42} ^ x_9 is F_{93}) then y = Not_acceptable (β_{41}) , y = Acceptable (β_{42}) and y = Excellent (β_{43}).
- If (x_4 is F_{43}) then y = Not_acceptable (β_{51}) , y = Acceptable(β_{52}) and y = (β_{53}).

Where y is the decision attribute to classify each pattern into a single class and β_{ml} is the certainty factor.

5. CONCLUSION

Fuzzy Decision Tree uses clustering to fuzzify continuous valued attribute data into linguistic terms and matches unseen instances by fuzzy reasoning. Such cognitive certainties included in classification problems are explicitly represented, measured and incorporated into the knowledge induction process. The generalization capability of Fuzzy Decision Tree is high when compared to the Crisp Decision Tree. In this paper, Fuzzy Decision Tree for Fruit data classification is constructed and then the more human interpretable rules extracted for Fruit data .

BIBLIOGRAPHY

[1] Breiman L., Friedman J.H., Olshen R.A., and Stone C.J., "Classification and Regression Trees", *Wadsworth*, 1984.

[2] Chengming Qi, "A new partition criterion for Fuzzy Decision Tree algorithm", *Workshop on Intelligent Information Technology Application (IITA 2007)*, pp 43-46, 2007.

[3] Janikow C.Z, "Fuzzy Decision Trees, Issues and methods", *IEEE Transactions on Systems, Man and Cybernetics*, Vol.28 (1), pp 1-14, 1998.

[4] Koen-Myung Lee, Kyung-Mi Lee, Jee-Hyong Lee, Hyung Lee-Kwang, "A Fuzzy Decision Tree induction method for Fuzzy data", *IEEE International Fuzzy Systems Conference proceedings*, Seoul, Korea, 1999.

[5] Olaru C and Wehenkil L, "A Complete Fuzzy Decision Tree technique", *Fuzzy Sets and Systems*, Vol.138 (2), 2003.

[6] Quinlan J.R., "*C4.5: Programs for Machine Learning*", Morgan Kaufmann, San Mateo, CA, 1993.

[7] Quinlan J.R., "Induction of Decision Trees", *Machine Learning*, Vol.1, pp.81-106, 1986.

[8] Quinlan J.R., "Unknown attribute values in Induction", *Proceedings of the Sixth International workshop on Machine Learning*, pp.164-168, 1989.

[9] Sushmita Mitra, Kishori M. Konwar and Sankar K. Pal, "Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-based network: Generation and Evaluation", *IEEE Transactions on Systems, Man and Cybernetics,* Vol.32, No.4, 2002.

[10] Wang LX and Mandel J.M, "Generating Fuzzy rules by learning from Examples", *IEEE Transactions on Systems, Man and Cybernetics*, Vol.22 (6), pp 1414-1427, 1992.

[11] Wang X.Z, Yeung D.S and Tsang, E.C.C., "A comparative study on heuristic algorithms for generating Fuzzy Decision Trees", *IEEE Transactions on SMC-B31,* pp.215-226, 2001.

[12] Yuan, Y and Shaw M.J, "Induction of Fuzzy Decision Trees", *Fuzzy Sets and Systems*, Vol.69, pp.125-139, 1995.