IJCSE International Journal of Computer Science and Engineering Open Access
Research Paper Volume-2, Issue-1 E-ISSN: 2347-2693

Prediction Of A Class Variable In Classification Problem Using Fuzzy Inference Method

S.V.S. Ganga Devi

Professor in MCA, K.S.R.M. College of Engineering, Kadapa – 516003, A.P., India,
Email: svsgangadevi_06@yahoo.co.in

www.ijcaonline.org

Received: 22/12/2013 Revised: 10/01/204 Accepted: 18/01/204 Published: 31/01/204

ABSTRACT- A popular and particularly efficient method for making a decision tree for classification from symbolic data is ID3 algorithm. Revised algorithms for numerical data have been proposed, some of which divide a numerical range into several intervals or fuzzy intervals. Their decision trees, however, are not easy to understand. A new version of ID3 algorithm to generate a understandable fuzzy decision tree using fuzzy sets defined by a user. In this paper, first the fuzzy decision tree is constructed for the given data and then fuzzy reasoning is applied in order to predict the class variable.

Key Words- Fuzzy Technique

I. INTRODUCTION

Knowledge acquisition from data is very important in knowledge engineering. A popular and efficient method is ID3 algorithm proposed by J.R. Quinlan [1], [2] in 1979, which makes a decision tree for classification from symbolic data.

The decision tree consists of nodes for testing attributes, edges for branching by values of symbols and leaves for deciding class names to be classified. ID3 algorithm applies to a set of data and generates a decision tree which minimizes the expected value of the number of tests for classifying the data.

For numerical data, its revised algorithms have been proposed, which divide a numerical range of attribute into several intervals. To make a decision tree flexible, some algorithms are proposed to fuzzify the interval [3], [4]. Their decision trees, however, are not easy to understand because

(1) we cannot know how a range of attribute is divided into intervals,

(2) a range of attribute may be divided into different intervals on different test nodes,

(3) one attribute may appear more than one time in one sequences of tests.

Moreover, we need a long sequence of tests since the decision tree is binary.

As for numerical data, many tuning methods for fuzzy rules are also proposed in the research of fuzzy control, which is called neuro fuzzy technique [5]. Since it generates rules that contain all combinations of all fuzzy sets in attributes, it has several difficulties as follows:

(1) so many fuzzy rules are generated,

(2) fuzzy sets in the rules are not understandable since they are tuned for fitting the training data,

(3) the more the number of attributes are, the less convergent the error is.

II. FUZZY ID3 ALGORITHM

ID3 algorithm [1], [2] applies to a set of data and generates a decision tree for classifying the data, algorithm, called fuzzy ID3 algorithm, is extended to apply to a fuzzy set of data 9several data with membership grades) and generates a fuzzy decision tree using fuzzy sets defined by a user for all attributes. A fuzzy decision tree consists of nodes for testing attributes, edges for branching by test values of fuzzy sets defined by a user and leaves for deciding class names with certainties. An example of fuzzy decision trees is shown in Figure 1.

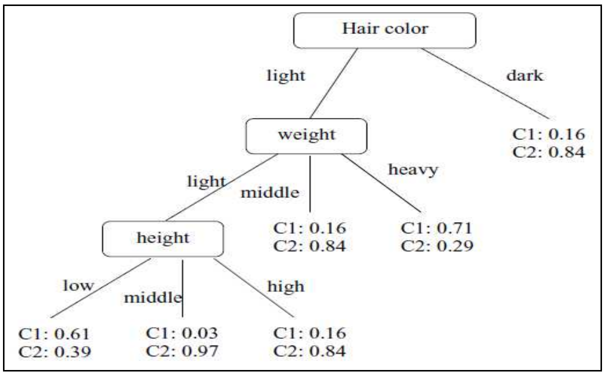

Figure 1: Fuzzy decision tree

It is very similar to ID3, except ID3 selects the test attribute based on the information gain which is

Corresponding Author: *S.V.S. Ganga Devi*

computed by the probability of ordinary data but this is the probability of membership values for data.

Fuzzy ID3Algorithm:

Create a root node that has a set of fuzzy data with membership value 1.

If a node t with a fuzzy set of data D satisfies the following conditions, then it is a leaf node and assigned by the class name

The proportion of a class C_k is greater than or equal to

$$\theta_r, \frac{|D^{Ct}|}{|D|} \geq \theta_r$$

The number of a data set is less than θ_n

There are no attributes for more classifications

If a node D does no satisfy the above conditions, then it is not a leaf-node. And an new sub-node is generated as follow:

- For $A_i's(i=1,....L)$ calculate the information gain G and select the test attribute $A_{\max}$ that maximizes them.
- Divided D into fuzzy subset $D_1,.....,D_m$ 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.
- Generate new nodest $t_1,.....,t_m$ for fuzzy sets $D_1,.....,D_m$ and label the fuzzy sets $F_{\max,j}$ to edges that connect between the nodes t_j and t
- Replace D by $D_j\,(j=1,2,....m)$ and repeat from 2 recursively.

III. REASONING WITH FUZZY ID3

We must start reasoning from the top node (Root) of the fuzzy decision tree. Repeat testing the attribute at the node, branching an edge by its value of the membership function (μ) and multiplying these values until the leaf node is reached. After that we multiply the result with the proportions of the classes in the leaf node and get the certainties of the classes at these leaf nodes. Repeat this action until all the leaf nodes are reached and all the certainties are calculated. Sum up the certainties of the each class respectively and choose the class with highest certainty.

We present the sample calculation by means of Figure 2. Each of tree node and leaf node represent the value of the membership function of the attribute at the node and the proportion of each class at the node respectively. By using method called X-X-+, we have the result that the sample belongs to C_1 and C_2 with probabilities 0.355 and 0.645 respectively. These probabilities are complement.

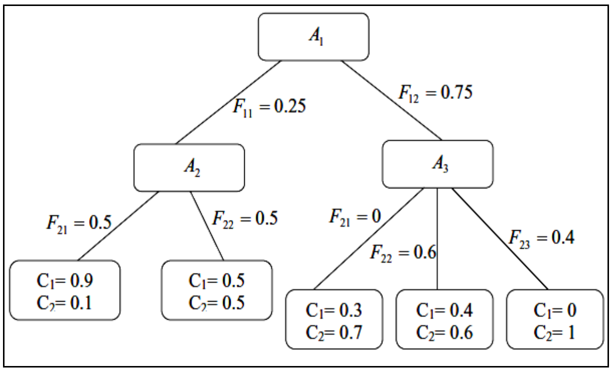

Figure 2: Representing of fuzzy reasoning

C_1=0.25*0.5*0.9+0.25*0.5*0.5*+0.75*0*0.3+0.75*0.6*0.4+0.75*0.4*0 = 0.355

C_2 =0.25*0.5*0.1+0.25*0.5*0.5+0.75*0*0.7+0.75*0.6*0.6+0.75*0.4*1=0.645

$$C_1 + C_2 = 1$$

IV. CONCLUSION

First Fuzzy Decision tree was constructed from numerical data using fuzzy sets defined by a user. Then the given sample belongs to which class is predicted by an inference method used by FDT. This can be considered as a method to generate fuzzy rules from a set of numerical data.

REFERENCES

[1] J.R. Quinlan (1979}: “Discovering Rules by Induction from large collections of Examples”, in d.Michie (ed.): Expert Systems in the Micro Electronics Age, Edinburgh University Press.

[2] J.R. Quinlan (1986): “Induction of Decision Trees”, Machine Learning, Vol.1, pp.81-106.

[3] T. Tani and M. Sakoda (1991): “Fuzzy Oriented Expert System to Determine Heater Outlet Temperature Applying Machine Learning”, 7th Fuzzy System Symposium (Japan Society for Fuzzy Theory and Systems), pp.659-662 (in Japanese).

[4] S. Sakurai and D. Araki (1992): “Application of Fuzzy Theory to Knowledge Acquisition”, 15th Intelligent System Symposium (Society of Instrument and Control Engineers), pp.169-174 (in Japanese).

[5] H. Ichihashi (1993): “Tuning Fuzzy Rules by Neuro-Like Approach”, Journal of Japan Society for Fuzzy Theory and Systems, Vol.5, No.2, pp.191-203 (in Japanese).

[6] F. Kawachi and T. matsuura (1990): “Development of Expert System for Diagnosis by Gas in Oil and Its Evaluation in Practice Usage”, Technical Meeting on electrical Insulation Material (The Institute of Electrical Engineers of Japan), EIM-90-40 (In Japanese).

