GENERATING FUZZY DECISION TREES BASED ON A NEW HEURISTIC

Shi Benguang[1]
Mathematics Group, Zhengzhou Grain College, Zhengzhou 450005, Henan, P.R.China

and

Wang Xizhao
Department of Mathematics, Hebei University, Baoding 071002, Hebei, P.R.China

The induction of fuzzy decision trees is an important way of acquiring imprecise knowledge automatically. Fuzzy ID3 and its variants are popular and efficient methods of making fuzzy decision trees from a group of training examples. This paper presents a new algorithm for generating fuzzy decision trees. Distinguished from Fuzzy ID3, this new algorithm takes the maximum variance as the heuristic.

The induction of decision trees is an efficient way of learning from examples([1]). Many methods have been developed for constructing decision trees([2]) and these methods are very useful in building knowledge-based expert systems. The cognitive uncertainty such as vagueness and ambiguity has been incorporated into the knowledge induction process by using fuzzy decision trees([3]). Fuzzy ID3 algorithm and its variants([3,4,5,6,7]) are popular and efficient methods of making fuzzy decision trees. The fuzzy ID3 can generate fuzzy decision trees without much computation. It has the great matching speed and is especially suitable for large-scale learning problems.

Consider a directed tree where each edge of the tree links two nodes, the initial node and the terminal node. The former is called the fathernode of the latter while the latter is said to be the sonnode of the former. The node having not its fathernode is the root whereas the nodes having not any sonnodes are called leaves.

A decision tree is a kind of directed tree. The induction of decision trees is an important way of inductive learning. It has two key aspects, training and matching. The former is a process of constructing trees from a training set which is a collection of objects whose classes are known, while the latter is a process of judging classification for unknown cases. ID3 is a typical algorithm for generating decision trees([1]).

---

[1] Corresponding author

A fuzzy decision tree is a generalization of the crisp case. Before giving the heuristic, we explain some symbols. Throughout this paper, X is supposed to be a given finite set, F(X) is the family of all fuzzy subsets defined on X , and the cardinality measure of a fuzzy subset, A, is defined by M(A)= $\sum_{x\in X} A(x)$ (The membership of a fuzzy subset is denoted by itself).

There are many methods of representing fuzzy decision trees. In essential, the representation in this paper is as the same as in the article[3] where the fuzzy decision tree is regarded as fuzzy partitioning. The training strategy is to use a heuristic to accomplish its search. The general matching strategy of fuzzy decision trees is described as follows.

(a) Matching starts from the root and ends in a leaf along the branch of the maximum membership. (b) If the maximum membership at the node is not unique, matching proceeds along several branches. (c) The decision with maximum degree of truth is assigned to the matching result.

There are two key points in the process of constructing fuzzy decision trees. One is the selection of expanded attributes. They are such attributes that according to values of attributes (which are fuzzy subsets) trees are expanded at the nodes considered *i.e.* sonnodes of the nodes are generated. The other is the judgment on leaves. Nodes are usually regarded as leaves if the relative frequency of one class is greater than or equal to a given threshold value.

Learning algorithms typically use heuristic to guide their search. A general learning algorithm for generating fuzzy decision trees can described as follows.

Consider the whole training set which is regarded as the first candidate node.
WHILE there exist candidate nodes
DO select one using the search strategy; if the selected one is not a leaf, then generate its sonnodes by selecting the expanded attribute using a heuristic. These sonnodes are regarded as new candidate nodes.

From existing references, two powerful heuristics guiding the selection of expanded attributes in the fuzzy decision tree generation can be found. One, called Fuzzy-ID3, is based on minimum fuzzy entropy([4,5,7]) whereas the other is based on the reduction of classification ambiguity([3]). The latter is a variant of the former. These two heuristics are briefly described as follows.

Let there be N training examples and n attributes $A^{(1)},\cdots,A^{(n)}$. For each $k(1\le k\le n)$, the attribute $A^{(k)}$ takes m_k values of fuzzy subsets, $A_1^{(k)},\cdots,A_{m_k}^{(k)}$. For simplicity, the fuzzy classification is considered to be two fuzzy subsets, denoted by P and N respectively. For each attribute value (fuzzy subset), $A_i^{(k)}(1\le k\le n, 1\le i\le m_k)$, its relative frequencies concerning P and N are

$$p_i^{(k)} = M(A_i^{(k)}\cap P)/M(A_i^{(k)}) \text{ and } q_i^{(k)} = M(A_i^{(k)}\cap N)/M(A_i^{(k)})$$

respectively. The relative frequency is regarded as the degree of truth of a fuzzy rule in [3].

Heuristic (1). Fuzzy ID3 based on the minimum fuzzy entropy.

Select such an integer k_0 (the k_0-th attribute) that $E_{k_0} = Min_{1\le k\le n} E_k$

where $E_k = \sum_{i=1}^{m_k}\left(M(A_i^{(k)}) \Big/ \sum_{j=1}^{m_k} M(A_j^{(k)}) \right) Entr_i^{(k)}$, $k = 1,2,\cdots,n$;

$Entr_i^{(k)} = -p_i^{(k)}\log_2 p_i^{(k)} - q_i^{(k)}\log_2 q_i^{(k)}$ denotes the fuzzy entropy of classification.

Heuristic (2). A variant of Fuzzy ID3 based on the minimum classification ambiguity.

Select such an integer k_0 (the k_0-th attribute) that $G_{k_0} = Min_{1\le k\le n} G_k$

where $G_k = \sum_{i=1}^{m_k}\left(M(A_i^{(k)}) \Big/ \sum_{j=1}^{m_k} M(A_j^{(k)}) \right) Ambig_i^{(k)}$, $k = 1,2,\cdots,n$;

$Ambig_i^{(k)} = Min(p_i^{(k)}, q_i^{(k)}) \,/\, Max(p_i^{(k)}, q_i^{(k)})$ denotes the ambiguity of classification.

The heuristics (1) and (2) can be easily extended to the case of fuzzy classification having more than two fuzzy subsets. The heuristic (2) has an option of significant level which will affect the generation of fuzzy decision trees. Generally, different heuristics will result in different trees. Intuitively, both the weighted average of fuzzy entropies and the weighted average of classification ambiguities will decrease when

$$Max(p_i^{(k)}, q_i^{(k)}) \to 1 \text{ or } Min(p_i^{(k)}, q_i^{(k)}) \to 0.$$

In the process of generating fuzzy decision trees, both heuristic (1) and heuristic (2) attempt to reduce the average depth of trees, *i.e.* on an average, to generate leaves as soon as possible.

Before giving our new heuristic, we explain fuzzy rules using the following definitions.

Definition 1. A fuzzy rule takes a form: IF A THEN B which defines a fuzzy relation from condition fuzzy set A to conclusion fuzzy set B.

A rule IF A THEN B is true means that A implies B, denoted by $A \Rightarrow B$. The implication operator can be interpreted in many ways([3]). As the interpretation of [3], the implication $A \Rightarrow B$ in this paper is understood to be $A \subset B$.

Definition 2. The true degree of a fuzzy rule $A \Rightarrow B$ is defined to be $\sum_{u\in U} \min(\mu_A(u), \mu_B(u)) \big/ \sum_{u\in U} \mu_A(u)$ where A and B are two fuzzy sets defined on the same universe U.

Now, we give a new heuristic algorithm for generating fuzzy decision trees. Suppose the fuzzy classification is considered to be L fuzzy subsets, denoted by C_1, C_2, ... and C_L respectively. For each attribute value (fuzzy subset), $A_i^{(k)} (1 \le k \le n, 1 \le i \le m_k)$, its relative frequencies concerning C_j

$$p_{ij}^{(k)} = M(A_i^{(k)} \cap C_j) \,/\, M(A_i^{(k)}) \quad (j=1, 2, \ldots, L)$$

respectively. From Definition 2., we know that the relative frequency is the true degree of a fuzzy rule. That is, IF $[A^{(k)} = A_i^{(k)}]$ THEN C_j with the true degree $p_{ij}^{(k)}$.

We define the variance by

$$Var_i^{(k)} = \Sigma_{j=1}^{L}(p_{ij}^{(k)} - (\Sigma_{j=1}^{L} p_{ij}^{(k)}) / L)^2$$

The new heuristic based on the minimum error variance is described as follows.

Select such an integer k_0 (the k_0-th attribute) that $V_{k_0} = Max_{1 \le k \le n} V_k$

where $V_k = \sum_{i=1}^{m_k} \left(M(A_i^{(k)}) \Big/ \Sigma_{j=1}^{m_k} M(A_j^{(k)}) \right) Var_i^{(k)}$, $k = 1,2,\cdots,n$;

Compared with existing heuristics for generating fuzzy decision trees, the new algorithm has the following advantages.

(1). It can efficiently describe the statistical characteristics of the training data.

(2). It is especially suitable for the case of generating fuzzy decision trees for continuous-valued attributes.

(3). Experimental results show that, with regard to the test accuracy for unknown cases, the new heuristic is superior to the likes of Fuzzy ID3.

References

[1] J.R.Quinlan, Induction of decision trees, Machine Learning 1(1986)81-106

[2] S.R.Safavian and D.Landgrebe, A survey of decision tree classifier methodology, IEEE Trans. System Man Cybernet. 21(1991)660-674

[3] Y.Yuan and M.J.Shaw, Induction of fuzzy decision trees, Fuzzy Sets and Systems 69(1995)125-139

[4] H.Ichihashi et al., Neuro-Fuzzy ID3, Fuzzy Sets and Systems 81(1996)151-167

[5] K.J.Cios and L.M.Sztandera, Continuous ID3 algorithm with fuzzy entropy measures, Proc. IEEE Internet. Conf. on Fuzzy Systemes(San Diego, CA, March 1992)469-476

[6] R.Weber, Fuzzy-ID3: a class of methods for automatic knowledge acquisition, Proc. 2nd Internet. Conf. on Fuzzy Logic and Neural Networks (Iizuka, Japan, 17-22 July 1992) 265-268

[7] T.Tani and M.Sakoda, Fuzzy Modeling by ID3 algorithm and its application to prediction , Proc. IEEE Internet. Conf. on Fuzzy Systems (San Diego, CA, March 1992)923-930