An Efficient Algorithm for Inducing Fuzzy Rules from Numerical Data

J.N. Wu and K.C. Cheung

Department of Mechanical Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong
jnwu@hkusub.hku.hk or kccheung@hkucc.hku.hk

Abstract

This paper proposes a modified but more powerful algorithm for inducing fuzzy if-then rules from numerical data. Data mining is performed before the process of fuzzy inference in view of the presence of noise in data. Therefore, the minimum number of learning attributes can be used to induce fuzzy rules automatically from training examples. The results obtained by this modified method demonstrate that it is more efficient and effective than relevant works in aspects of time complexity and space complexity.

Introduction

In many classification problems, classification rules are often obtained by automatic knowledge acquisition methods. The process of acquiring knowledge from experts is still one of the difficult problems in expert systems development. Thus, it is expected that inductive inference techniques which have been used to build expert systems could ease this process.

Recently, many methods have been developed for inducing if-then rules from training examples. However, these methods are often inadequate or improper in expressing and handling the vagueness and ambiguity associated with human thinking and perception (Yuan and Shaw 1995). It is quite important that fuzzy set theory should be introduced in expert systems to manage uncertainty and noise because of its simplicity and similarity to human reasoning. Up to now, various fuzzy machine learning methods have been used in classification problems for automatically inducing fuzzy if-then rules. These include the methods proposed by Hong and Lee (Hong and Lee 1996), Ishibuchi, et al. (Ishibuchi, et al. 1995), Nozaki et al. (Nozaki et al. 1997), Yuan and Shaw (Yuan and Shaw 1995). Among them the inductive method proposed by Hong and Lee is very simple and yet powerful. It can automatically derive membership functions and fuzzy if-then rules from a given set of training classification examples and can significantly reduce the time and effort needed to develop a fuzzy expert system.

However, the weakness of Hong and Lee's method is that it employs data sets directly to generate membership functions and induce fuzzy rules. If the original training data contains noise, the speed of inductive learning will be adversely affected.

On the other hand, the algorithm named Chi2 (Liu and Setiono 1995) can automatically select features according to the characteristics of the original data. Feature selection can eliminate some irrelevant attributes. Using only relevant attributes in inductive learning can, in general, shorten the learning process.

In order to keep the advantage of Chi2 method and overcome the weakness of Hong and Lee's algorithm, we present a method which combines their ideas to deal with classification problems with numerical attributes before deriving fuzzy if-then rules. A modified algorithm for fuzzy inductive learning is designed so that the effectiveness of the learning process improves on a large scale both in time and space requirements. After data mining, the minimum number of learning attributes are used in the process of inducing fuzzy if-then rules by the modified fuzzy inductive algorithm.

Fuzzy Classification Problem and Fuzzy Classification Rules

Fuzzy Classification Problem

A typical classification problem consists of a particular object described in terms of the values of all relevant attributes together with the classification given. A fuzzy classification problem means that either the object or the class is fuzzy. In general, each object takes one of the mutually exclusive fuzzy values for each attribute and each object is classified into only one of the mutually exclusive class represented in fuzzy terms (Quinlan 1986). An n-dimensional input-output pair given as training data for constructing a fuzzy system can be described as (Ishibuchi et al. 1995):

$$\{(x_p\,;y_p)\quad p=1,2,\ldots,n\} \tag{1}$$

where $x_p=(x_{p1}, x_{p2}, \ldots, x_{pm})$ is the input vector of the pth input-output pair and y_p is the corresponding output. The input is often referred to as a set of attribute values and the output as the class values.

Fuzzy If-Then Rules

In a fuzzy classification system, a fuzzy classification rule defines a fuzzy relation from the condition fuzzy set (attribute set) to the conclusion fuzzy set (class set). The process of inference, which derives conclusions when conditions are satisfied, is based on fuzzy logic. The derived rules in this paper are assumed in the following form:

$$\textbf{if } x_{p1} \wedge x_{p2} \wedge \ldots \wedge x_{pm} \textbf{ then } y_p \quad (2)$$

where the symbol "$\wedge$" means AND.

For example, in the Golf Playing Problem (Quinlan 1988), the pth training example can have four attributes:

x_p=(Outlook, Temperature, Humidity, Windy),

and each attribute has fuzzy values

Outlook(x_{p1})=(Rainy, Overcast, Sunny),
Temperature(x_{p2})=($\leq$69, >69, $\leq$75, >75),
Humidity(%)(x_{p3})=(>80, $\leq$80),
Windy(x_{p4})=(True, False).

The fuzzy class values (y_p) can be either Don't Play or Play.

A fuzzy classification rule can be expressed as:

Rule: **if** rainy and windy **then** don't play.

The Procedures of Fuzzy Inductive Learning

In this section, a more efficient fuzzy inductive method is proposed to automatically derive fuzzy if-then rules and membership functions from numerical data sets. There are three main stages in our fuzzy inductive learning algorithm: selecting features from numerical attributes, deriving membership functions and fuzzy if-then rules, and defuzzifying output data to non-fuzzy values.

Selecting Features from Numeric Attributes

The Chi2 algorithm which is based on the χ^2 statistic is a useful and reliable tool for selecting features from numerical attributes (Liu and Setiono 1995). We employed this method here as an approach for training data preprocessing. The process of feature selection is shown as follows:

Step 1: Set a high significance level (SigLevel) for all numeric attributes for discretization.
Step 2: Sort each numerical attribute.
Step 3: Calculate the χ^2 values for every pair of adjacent intervals:

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{k}\frac{(A_{ij}-E_{ij})^2}{E_{ij}} \quad (3)$$

where k is the number of classes, A_{ij} is the number of patterns in the ith interval and jth class, $R_i=\sum_{j=1}^{k}A_{ij}$ is the number of patterns in the ith interval, $C_j=\sum_{i=1}^{2}A_{ij}$ is the number of patterns in the jth class, $N=\sum_{i=1}^{2}R_i$ is the total number of patterns, and $E_{ij}=R_i*C_j/N$ is the expected frequency of A_{ij}.

Step 4: Merge the pair of adjacent intervals with the lowest χ^2 value until all χ^2 values of adjacent pairs of intervals exceed the parameter determined by SigLevel.
Step 5: If inconsistency rate (δ) is exceeded in the discretized data then stop; else set SigLevel0=SigLevel, SigLevel=decreSigLevel, and go to Step 2.
Step 6: Sort each numerical attribute i which is associated with a SigLevel[i]=SigLevel0 again.
Step 7: Merge each attribute one by one.
Step 8: If the inconsistency rate is not exceeded then SigLevel[i] is decremented for attribute i's next round of merging; else attribute i will not be involved in further merging.
Step 9: If no attribute values can be merged then stop; else go to Step 6.

At the end of the feature selection process, if an attribute is merged into only one value then it is not relevant in representing the original data set. That means feature selection is accomplished.

The Fuzzy Inference Process

In 1996, Hong and Lee (Hong and Lee 1996) proposed a very simple and useful method for automatically generating reasonable membership functions and appropriate decision rules from training data. The method is adapted as the following 8 steps.

Inducing fuzzy rules process:

Step 1: Cluster and fuzzify the class values if they are numerical.

(1) Sort $\{y_p, p=1, 2, \ldots, n-1\}$ to get $\{y_p \mid y_p \leq y_{p+1}, p=1, 2, \ldots, n-1\}$.

(2) Calculate the value of similarity s_p between y_p and y_{p+1}:

$$s_p=\begin{cases} 1-\dfrac{diff_p}{C\times\sigma_s} & \text{for } diff_p \leq C\times\sigma_s \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $diff_p=y_{p+1}-y_p$, σ_s is the standard derivation of $diff_p$'s, and C is a control parameter deciding the shape of the membership functions of similarity.

(3) Cluster the class values by using the α-cut of similarity: if $s_p<\alpha$ then divide the two adjacent data into different groups else put them into the same group. After that, the result in the form of (y_p, R_j) is obtained, meaning that the class value of the pth training example will be clustered into the jth produced region $R_j=(a_j, c_j]$.

(4) Calculate a triad (a_j, b_j, c_j) of a triangle defined as the membership function of the output data by the following formulas, where b_j is the central-vertex-point. If y_i, y_{i+1}, ..., y_k belong to the jth region R_j, then.

$$b_j=\frac{y_{ps}\times s_p+y_{p+1s}\times\frac{s_p+s_{p+1}}{2}+y_{p+2s}\times\frac{s_{p+1}+s_{p+2}}{2}+\cdots+y_{i-1s}\times\frac{s_{k-2}+s_{k-1}}{2}+y_{is}\times s_{k-1}}{s_p+\frac{s_p+s_{p+1}}{2}+\frac{s_{p+1}+s_{p+2}}{2}+\cdots+\frac{s_{k-2}+s_{k-1}}{2}+s_{k-1}} \tag{5}$$

$$a_j=b_j-\frac{b_i-y_{ps}}{1-\mu_j(y_{ps})} \tag{6}$$

$$c_i=b_j+\frac{y_{ks}-b_i}{1-\mu_j(y_{ks})} \tag{7}$$

where $\mu_j(y_{ps})=\mu_j(y_{ks})=\min(s_p, s_{p+1}, \ldots, s_{k-1})$ represents the membership of belonging to the jth region.

(5) Obtain the fuzzy value of each output data from the membership functions produced above. Each training data can then be written as:

$$(x_p;\ (R_1,\mu_1), (R_2,\mu_2), \ldots, (R_k,\mu_k)) \tag{8}$$

Step 2: Construct initial membership functions for input attributes.

Assume each numerical attribute to be a triangle with the smallest predefined unit as an initial membership function.

Step 3: Build a multi-dimensional decision table (each dimension represents an attribute) according to the initial membership functions. Let a *cell* be defined as the contents of a position in the decision table.

Step 4: Simplify the initial decision table according to the following rules:

Rule 1: If cells in two adjacent columns (or rows) are the same, then merge these two columns (or rows) into one.

Rule 2: If two cells are the same or if either of them is empty in two adjacent columns (or rows) and at least one cell in both the columns (or rows) is not empty, then merge these two columns (or rows) into one.

Rule 3: If all cells in a column (or row) are empty and if cells in its two adjacent columns (or rows) are the same, then merge these three columns (or rows) into one.

Rule 4: If all cells in a column (or row) are empty and if cells in its two adjacent columns (or rows) are the same or either of them is empty, then merge these three columns (or rows) into one.

Rule 5: If all cells in a column (or row) are empty and if all the non-empty cells in the column (or row) to its left have the same region, and all the non-empty cells in the column (or row) to its right have the same region, but one different from the previously mentioned region, then merge these three columns (or rows) into two parts.

Step 5: Rebuild membership functions.

The new membership functions for the dimension to be processed are obtained.

Step 6: Derive fuzzy if-then rules from the decision table.

Each cell denoted by $cell_{(d_1,d_2,\ldots,d_n)}=R_j$ in the decision table is used to derive a rule:

$$\textbf{if } x_{p1}=d_1\wedge x_{p2}=d_2, \ldots, \wedge x_{pn}=d_n \textbf{ then } y_p=R_j \tag{9}$$

Step 7: Convert numerical attribute values to linguistic terms according to the membership functions derived.

Step 8: Match the linguistic terms with the fuzzy rules to obtain the fuzzy output regions.

Defuzzify the Output Data to Non-Fuzzy Values

The final class value is converted into non-fuzzy value by averaging the output regions (Zimmermann 1996). Let the membership function for the output region R_i be (a_i, b_i, c_i). The non-fuzzy class value is calculated by the following formula:

$$y_p=\frac{\sum_{j=1}^{K}\mu_{R_j}(x_p)\times b_j}{\sum_{j=1}^{K}\mu_{R_j}(x_p)} \tag{10}$$

where $\mu_{R_j}(x_p)=\min(\mu_{x_{p1}}(x_p),\mu_{x_{p2}}(x_p),\cdots\mu_{x_{pn}}(x_p))$, and K is the number of possible output regions.

Application of Inductive Algorithm

The Iris plant data is used to verify the effectiveness of the proposed method because it is a well known data as the benchmark of classification algorithms. It is a set of data with 150 training examples on flowers proposed by R.A. Fisher in 1936 (Fisher 1936). Each training instance in the set is described in terms of four numerical attributes: Sepal Width, Sepal Length, Petal Width, Petal Length and can be classified into three species: Setosa, Versicolor, and Verginica.

After feature selection, the number of attributes for Iris plant data is reduced from 4 to 2 (only Petal Length and Petal Width are left).

Then, we divide the Iris plant data into two sets, one is the training set and another is the testing set. The fuzzy inductive algorithm was employed in inducing fuzzy classification rules and membership functions by only the two attributes. In the application, half of examples were chosen at random for generating fuzzy if-then rules and the remaining half for testing the derived rules.

The produced membership functions for the two attributes are shown in Figure 1. The derived fuzzy rules and the correct classification ratio are shown in Table 1 and Table 2 respectively.

Discussions

As we know that a rule does not necessarily use all the attributes necessary to describe a particular example, so an ideal rule should use as few attributes as possible. From Table 1, we can see that the rules derived by the modified method consist of only two attributes for the Iris plant data. It can also be seen that these rules possess much higher compactness than the rules obtained by other methods. The high accuracy obtained from the fuzzy inductive method is shown in Table 2. What makes the proposed method effective is just its capability of feature selection before fuzzy inductive learning. As a result, it makes the initial dimension of the decision table small and the time of fuzzy inductive learning short.

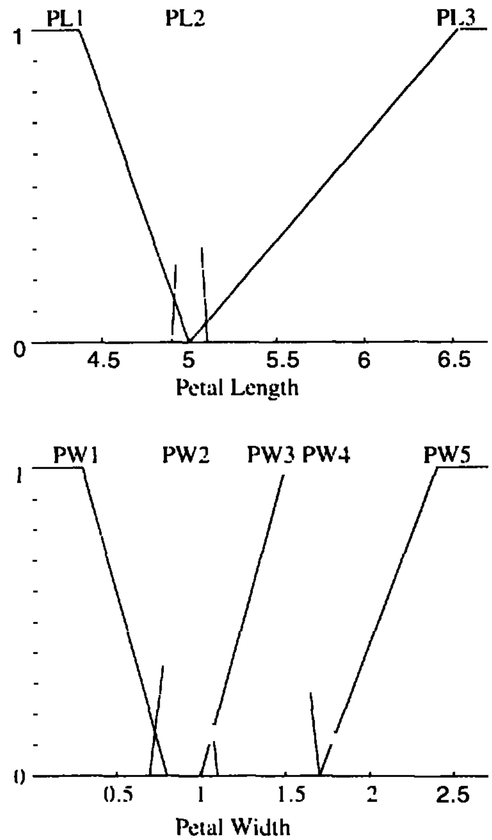

Figure 1: The final membership functions of the two attributes

Rule No.		Petal Length		Petal Width		Class
1	**if**	PL1	and	PW1	**then**	Setosa
2	**if**	PL1	and	PW3	**then**	Versicolor
3	**if**	PL1	and	PW4	**then**	Versicolor
4	**if**	PL1	and	PW5	**then**	Virginica
5	**if**	PL3	and	PW3	**then**	Virginica
6	**if**	PL3	and	PW4	**then**	Versicolor
7	**if**	PL3	and	PW5	**then**	Virginica

Table 1: The produced fuzzy inference rules

No. of runs	Aver. no. of unclassified	Aver. no. of misclassified	Aver. no. of rules	Aver. accuracy
200	0.64	3.8	6.78	94.11

Table 2: Classification accuracy of the fuzzy inductive algorithm for Iris plant data problem

Note that in each run, 75 sets of Iris plant data were chosen randomly for training, and the remaining 75 sets of the data were used for testing.

Conclusion

The method proposed in this study inherits the advantages of data mining and inductive learning. It can automatically derive the membership functions and fuzzy if-then rules from the minimum number of numerical attributes. The empirical results using Iris plant data have shown that the time for inductive learning can be shortened on a large scale by using the modified algorithm. The results also show that the proposed method is a reasonable and efficient tool for feature selection from numerical data and fuzzy rule induction.

References

Yuan, Y.F.; and Shaw, M.J. 1995. Induction of Fuzzy Decision Trees. *Fuzzy Sets and Systems* 69:125-139.

Hong, T.P.; and Lee, C.Y. 1996. Induction of Fuzzy Rules and Membership Functions from Training Examples. *Fuzzy Sets and Systems* 84:33-47.

Ishibuchi, H.; Nozaki, K.; Yamamoto, N.; and Tanaka, H. 1995. Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms. *IEEE Transactions on Fuzzy Systems* 3(3):260-270.

Nozaki, K.; Ishibuchi, H.; and Tanaka, H. 1997. A Simple but Powerful Heuristic Method for Generating Fuzzy Rules from Numerical Data. *Fuzzy Sets and Systems* 86:251-270.

Liu, H.; and Setiono, R. 1995. Chi2: Feature Selection and Discretization of Numeric Attributes. In Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence, 388-391.

Fisher, R.A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7(2):179-188.

Quinlan, J.R. 1988. Induction, Knowledge and Expert Systems. In Artificial Intelligence Developments and Applications, 253-271. Amsterdam: Elsevier.

Quinlan, J.R. 1986. Induction of Decision Trees. *Machine Learning* 1(1):81-106.

Muggleton, S. 1990. *Inductive Acquisition of Expert Knowledge*. Glasgow: Turing Institute Press.

Zimmermann, H.J. 1996. *Fuzzy Set Theory and Its Applications*. Third Edition. Boston: Kluwer Academic Publisher.