WAC 2012 1569538657

Incremental Learning of Fuzzy Rule-Based Classifiers for Large Data Sets

Tomoharu Nakashima*, Takeshi Sumitani*, and Andrzej Bargiela†

*Department of Engineering, Osaka Prefecture University
Gakuen-cho 1-1, Naka-ku, Sakai, Osaka 599-8531, Japan
Email: nakashi@cs.osakafu-u.ac.jp, takeshi.sumitani@ci.cs.osakafu-u.ac.jp

†Department of Computer Science, The University of Nottingham
Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, U.K.
Email: andrzej.bargiela@nottingham.ac.uk

***Abstract*—Incremental construction of fuzzy rule-based classifiers is studied in this paper. It is assumed that not all training patterns are given a priori for training classifiers, but are gradually made available over time. It is also assumed the previously available training patterns can not be used in the following time steps. Thus fuzzy rule-based classifiers should be constructed by updating already constructed classifiers using the available training patterns at each time step. Two methods are proposed for the incremental construction of fuzzy rule-based classifiers. The first method updates the fuzzy if-then rules by considering individual training patterns separately while in the second method all available training patterns are used together in the update procedure of the fuzzy if-then rules. A series of computational experiments are conducted in order to examine the performance of the proposed incremental construction methods of fuzzy rule-based classifiers using a simple artificial pattern classification problem.**

I. Introduction

Fuzzy systems based on fuzzy if-then rules have been researched in various fields such as control [?], classification and modeling [?]. A fuzzy rule-based classifier is composed of a set of fuzzy if-then rules. Fuzzy if-then rules are generated from a set of given training patterns. Advantages of fuzzy classifiers are mainly two-folds: First, the classification behavior can be easily understood by human users. This can be done by carefully checking the fuzzy if-then rules in the fuzzy classifier because fuzzy if-then rules are inherently expressed in linguistic forms. Another advantage is nonlinearity in classification. It is well known that non-fuzzy rule-based classifiers are difficult to perform non-linear classification because classification boundaries are always parallel to attribute axes in most cases. The nonlinearity of fuzzy classification leads to high generalization ability of fuzzy rule-based classifiers while its classification behavior is linguistically understood.

A fuzzy rule-based classifier in this paper consists of a set of fuzzy if-then rules. The number of fuzzy if-then rules is determined by the dimensionality of the classification problem and the number of fuzzy partitions used for each attribute. A fuzzy if-then rules is generated by calculating the compatibility of training patterns with its antecedent part for each class. The calculated compatibilities are summed up to finally determine the consequent class of the corresponding fuzzy if-then rule. An unseen pattern is classified by the fuzzy rule-based classifier (i.e. a set of generated fuzzy if-then rules) using a fuzzy inference process.

In general, as the amount of information keeps growing due to the development of high-performance computers and high-capacity memories, it is difficult for any information systems to efficiently and effectively process a huge amount of data at a time. This is because it takes intractably long time to retrieve whole data and it is not possible to handle the intractably huge amount of data by just one information system. Also, it is possible that training patterns are generated over time and the designers of information systems have to handle the dynamically available patterns in a manner of streaming process. This paper focuses on the latter case in the construction process of fuzzy rule-based classifiers.

In order to tackle with such streaming data, fuzzy rule-based classifiers need to adapt themselves to newly available training patterns. In this paper, two methods are proposed for incrementally constructing fuzzy rule-based classifiers. In the first method, the fuzzy if-then rules that constitute a fuzzy rule-based classifier are updated by considering the compatibility of each of the available training patterns with them while in the second method all available training patterns are considered together in the updating procedure of the fuzzy if-then rules. A series of computational experiments are conducted in order to examine the generalization ability of the constructed fuzzy rule-based classifiers by the proposed methods for a two-dimensional incremental pattern classification problem.

II. Pattern Classification Problems

A. Conventional Pattern Classification

The standard type of pattern classification is explained in this subsection. Let us assume that a set of training patterns is given before constructing a classifier. A training pattern consists of a real-valued input vector and its corresponding target class. Consider, for example, an n-dimensional C-class pattern classification problem. It is also assumed that m training patterns $\vec{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$ are given a priori. The task then is to construct a classifier that correctly classify an unseen pattern using the given set of the training patterns.

B. Incremental Pattern Classification

The incremental pattern classification problem in this paper is defined as the classification task that involves an incremental process of obtaining training patterns. That is, the full set of training patterns cannot be obtained beforehand. Instead, a small number of training pattens are gradually made available as the time step proceeds. It is also assumed that classification of new patterns should be made during the course of the incremental process of obtaining training patterns. Thus, a classifier should be constructed using the training patterns that have been made available thus far.

Let us denote the available training patterns at the time step t as $\vec{x}_p^t$, $p = 1, 2, \ldots, m^t$, where m^t is the number of training patterns that became available at time t. The task at time T is to construct a classifier using $\sum_{t=1}^{T} m^t$ training patterns $\vec{x}_p^t$, $p = 1, 2, \ldots, m^t$, $t = 1, 2, \ldots, T$.

III. Fuzzy Rule-Based Classifier

In this paper, a fuzzy rule-based classifier proposed in [**?**] is used. It should be noted that the idea of the classification confidence can be applied to any forms of fuzzy classifiers if they are rule-based systems. An overview of the system in [**?**] is given below.

A. Fuzzy If-Then Rule

In a pattern classification problem with n dimensionality and M classes, we suppose that m labeled patterns, $\vec{x}_p = \{x_{p1}, x_{p2}, \cdots, x_{pn}\}$, $p = 1, 2, \cdots, m$, are given as training patterns. We also assume that without loss of generality, each attribute of $\vec{x}_p$ is normalized to a unit interval $[0, 1]$. From the training patterns we generate fuzzy if-then rules of the following type:

$$R_q\text{: If } x_1 \text{ is } A_{q1} \text{ and } \cdots \text{ and } x_n \text{ is } A_{qn} \text{then Class } C_q \text{ with } CF_q, \quad q = 1, 2, \cdots, N, \tag{1}$$

where R_q is the label of the q-th fuzzy if-then rule, $\vec{A}_q = (A_{q1}, \cdots, A_{qn})$ represents a set of antecedent fuzzy sets, C_q a the consequent class, CF_q is the confidence of the rule R_q, and N is the total number of generated fuzzy if-then rules.

We use triangular membership functions as antecedent fuzzy sets. Figure 1 shows triangular membership functions which divide the attribute axis into five fuzzy sets. Suppose that an attribute axis is divided into L fuzzy sets. The membership function of the k-th fuzzy set is defined as follows:

$$\mu_k(x) = \max\left\{1 - \frac{|x - x_k|}{v}, 0\right\}, k = 1, \cdots, L, \tag{2}$$

where

$$x_k = \frac{k-1}{L-1},\ k = 1, \cdots, L, \tag{3}$$

and

$$v = \frac{1}{L-1}. \tag{4}$$

Let us denote the compatibility of a training pattern $\vec{x}_p$ with a fuzzy if-then rule R_q as $\mu_{\vec{A}_q}(\vec{x}_p)$. The compatibility $\mu_{\vec{A}_q}(\vec{x}_p)$ is calculated as follows:

$$\mu_{\vec{A}_q}(\vec{x}_p) = \prod_{i=1}^{n} \mu_{A_{qi}}(x_{pi}),\ q = 1, 2, \cdots, N, \tag{5}$$

where $\mu_{A_{qi}}(x_{pi})$ is the compatibility of x_{pi} with the fuzzy set A_{qi} and x_{pi} is the i-th attribute value of $\vec{x}_p$. Note that $\mu_{A_{qi}}(x_{pi})$ is calculated by (2).

The number of fuzzy rules to be generated is L^n. That is, the number of rules increases exponentially for the division number and the dimensionality.

B. Generating Fuzzy If-Then Rules

A fuzzy classification system consists of a set of fuzzy if-then rules. The fuzzy if-then rules are generated from the training patterns $\vec{x}_p$, $p = 1, 2, \ldots, m$. The number of generated fuzzy if-then rules is determined by the number of fuzzy partitions for each axis (i.e., L in (2) $\sim$ (4)). That is, the number of generated fuzzy if-then rules is the number of combinations of fuzzy sets that are used for attribute axes. Although different numbers of fuzzy partitions can be used for different axes, in this paper we assume that it is the same for all axes. In this case, the number of fuzzy if-then rules is calculated as $N = L^n$ where n is the dimensionality of the pattern classification problem at hand. In this paper, it is supposed that all attributes are divided in the same way (i.e., the same fuzzy partition). An illustrative example is shown in Fig. 2. In Fig. 2, a two-dimensional pattern space is divided into $3^2 = 9$ fuzzy subspaces as each attribute is divided into three fuzzy sets. Each subspace is labeled with a rule label ($R_1 \sim R_9$). For example, the antecedent part of Rule R_6 has the fuzzy set A_3 for attribute x_1 and A_2 for attribute x_2. In this way, the total number of generated fuzzy if-then rules and the antecedent part of each fuzzy if-then rule are automatically

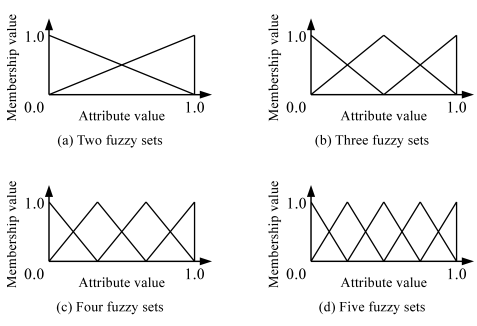

Fig. 1. Triangular fuzzy sets.

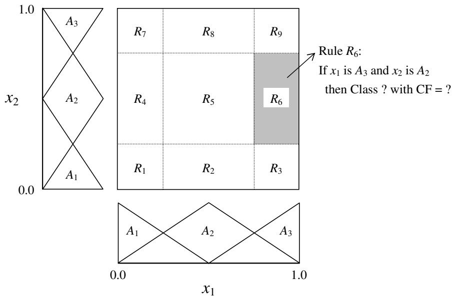

Fig. 2. Two-dimensional illustrative example of specifying the antecedent part of a fuzzy if-then rule (three fuzzy sets for both the two attributes).

determined after the number of fuzzy sets for each attribute is determined.

The consequent part of fuzzy if-then rules (i.e., C_q and CF_q in (1)) is determined from the given training patterns once the antecedent part is specified. The consequent class C_q of the fuzzy if-then rule R_q is determined as follows:

$$C_q = \arg \max_{h=1,\ldots,M} \beta_h^q, \tag{6}$$

where

$$\beta_h^q = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p). \tag{7}$$

That is, the most matching class with the fuzzy if-then rule is selected considering the given training patterns. If there is not any training pattern that is covered by the fuzzy if-then rules, the consequent class is set as empty. Also, in the case where multiple classes have the maximum value in (6), the consequent class is set as empty. The confidence CF_q is determined as follows:

$$CF_q = \frac{\beta_{C_q} - \bar{\beta}}{\sum_{h=1}^{m} \beta_h^q}, \tag{8}$$

where

$$\bar{\beta} = \frac{1}{M-1} \sum_{h \neq C_q} \beta_h^q. \tag{9}$$

There are other formulations for determining the confidence. Interested readers are referred to [**?**] for the discussion on the confidence calculation and the performance evaluation.

C. Classification of Unseen Patterns

Generated fuzzy if-then rules in the previous subsections are used to assign a class label to an unseen pattern which is not included in the set of training patterns. Let us denote an n-dimensional unseen pattern as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. The fuzzy inference is employed to classify unseen patterns in the fuzzy classification system in this paper. The class of an unseen pattern $\mathbf{x}$ is classified as Class C that is determined by the following equation:

$$C = \arg \max_{h=1,\ldots,M} \{\alpha_h\}, \tag{10}$$

where

$$\alpha_h = \max_{\substack{q=1,\ldots,N \\ C_q = h}} \{\mu_{\mathbf{A}_q(\mathbf{x})} \cdot CF_q\}. \tag{11}$$

In the above equations, M is the number of classes and N is the number of fuzzy if-then rules. In (10), if there are multiple classes that have the same maximum value of α_h, the classification of the unseen pattern is rejected.

IV. Incremental Construction of Fuzzy Rule-Based Classifiers

Since not all training patterns are available at a time but are available over time, it is necessary for already generated fuzzy if-then rules to adapt themselves to the training patterns that are newly made available. In this paper two methods for incrementally constructing fuzzy rule-based classifiers are proposed. Both the two methods update the summed compatibilities that are calculated in (7), but in different manners.

Incremental method A: The summed compatibilities are updated so that the new training patterns are considered equally as the previously available training patterns.

Incremental method B: The summed compatibilities are updated so that the higher weights are put for the new training patterns than the previously available training patterns.

The following subsections explain the above incremental construction methods of fuzzy rule-based classifiers.

A. Incremental Method A

Let us assume that at time T a fuzzy classifier has been already constructed. It is also assumed that at time $T+1$ new training patterns $\vec{x}_p^{T+1}$, $p = 1, 2, \ldots, m^{T+1}$ are made available. As in Subsection III-B, each fuzzy if-then rule has a summed compatibility β for each class.

The update procedure of the summed compatibilities is written as follows:

For each of new training patterns $\vec{x}_p^{T+1}$, $p = 1, 2, \ldots, m^{T+1}$, do the following steps:

Step 1: Calculate the compatibility $\mu_q(\vec{x}_p^{T+1})$ of $\vec{x}_p^{T+1}$ with the j-th fuzzy if-then rule R_q, $q = 1, 2, \ldots, N$, where N is the total number of generated fuzzy if-then rules.

Step 2: Update the summed compatibilities β_h^q of the fuzzy if-then rule R_q for Class h as follows:

$$\beta_h^q \leftarrow \begin{cases} \dfrac{n_h^q \cdot \beta_h^q + \mu_q(\vec{x}_p^{T+1})}{n_h^q + 1}, \\ \quad \text{if } \vec{x}_p^{T+1} \in \text{Class } h \\ \quad \text{and } \mu_q(\vec{x}_p^{T+1}) > 0.0, \\ \beta_h^q, \text{ otherwise.} \end{cases} \tag{12}$$

The above procedure calculates the new summed compatibility as the weighted average of the previous summed compatibility and the compatibility with the new pattern. The weight assigned for the summed compatibility is the number of training patterns that were previously available. Thus the new summed compatibility can be seen as the summed compatibility for $\vec{x}_p$, $p = 1, 2, \ldots, m^t$, $t = 1, 2, \ldots, T+1$ that is calculated by the standard fuzzy rule-generation procedure in Subsection III-B.

B. Incremental Method B

The second method of incrementally constructing fuzzy rule-based classifiers does not take into consideration of the number of training patterns that were used to calculate the summed compatibilities. Instead, they are modified so that they approach the compatibility of the new training patterns. The following steps explain the procedure:

For each of new training patterns $\vec{x}_p^{T+1}$, $p = 1, 2, \ldots, m^{T+1}$, do the following steps:

Step 1: Calculate the compatibility $\mu_q(\vec{x}_p^{T+1})$ of $\vec{x}_p^{T+1}$ with the j-th fuzzy if-then rule R_q, $q = 1, 2, \ldots, N$, where N is the total number of generated fuzzy if-then rules.

Step 2: Update the summed compatibilities β_h^q of the fuzzy if-then rule R_q for Class h as follows:

$$\beta_h^q \leftarrow \begin{cases} \beta_h^q + \gamma \cdot \delta_h^q & \text{, if } \vec{x}_p^{T+1} \in \text{Class } h \\ & \text{and } \mu_q(\vec{x}_p^{T+1}) > 0.0, \\ \beta_h^q & \text{, otherwise,} \end{cases} \quad (13)$$

where

$$\delta_h^q = \sum_{\vec{x}_p \in \text{Class } h} \mu_{\vec{A}_q}(\vec{x}_p) - \beta_h^q, \quad (14)$$

and γ is a positive constant in the closed interval $[0.0, 1.0]$.

V. Computational Experiments

In this section, the performance of fuzzy rule-based classifiers are examined. The proposed methods in the last section are used in the computational experiments. Two modes of incremental pattern classification problems are used in the experiments. One is a static incremental pattern classification problem where the classification boundaries between different classes remain unchanged during the course of the experiments. The second mode of the incremental pattern classification is dynamic in that the classification boundaries changes over time. For both the two modes, only a small number of training patterns are made available at each time step.

A. Static Pattern Classification Problem

In this section, the performance of the proposed methods were examined for a static incremental pattern classification problem. During the course of the experiments, the classification boundaries remain unchanged. Figure 3 shows the problem that is used in this subsection. A training pattern in this classification problem constitutes of a two-dimensional input vector in the domain space $[0.0, 1.0]^2$ and a target output which is one of four classes. The two diagonal lines in the two-dimensional pattern space are the classification boundaries between the four classes.

A fuzzy rule-based classifier is constructed from training patterns that are temporarily available. In the computational experiments, the number of temporarily available training patterns at a single time unit is specified as 1, 5, 10, and 100. The new training patterns are generated by first randomly sampling a two-dimensional input vector according to the uniform distribution, then specify the target output as the class in which the sampled vector falls. The procedure of generating new training patterns was iterated for 100 times. Thus the total number of new training patterns in a single run is 100, 500, 1000, and 10000 depending on the number of generated training patterns at a single time step. For the purpose of examining the generalization ability of the constructed fuzzy if-then rules, 10000 input patterns are uniformly generated in the domain space $[0, 1]^2$. The target class of a test pattern is determined by the area in which the pattern falls. For Incremental method B, the value of the positive constant γ in (13) is specified as $0.0, 0.1, \ldots, 0.9$, and 1.0. The number of fuzzy sets for each attribute is specified as two, three, four, and five (see Fig. 1). For each parameter specification, the computational experiments were conducted ten times to obtain the average performance of the proposed methods.

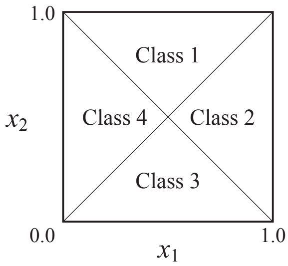

Fig. 3. A static pattern classification problem.

TABLE I

Experimental results by Incremental method A for the static problem. The number in each entry shows the classification performance (%).

# of fuzzy sets	# of patterns 1	2	5	10
2	48.0	49.7	49.5	50.3
3	91.8	92.7	94.4	94.9
4	82.3	82.9	84.3	84.7
5	90.9	92.2	94.1	94.3

The experimental results by Incremental method A is shown in Table I. From this table, it is shown that the classification performance becomes better when the number of available training patterns is larger, which is quite natural as the training classifiers is effective with more training patterns. The classification performance is not very good when the number of fuzzy sets is two and four. This is because the current setting of the triangular fuzzy partitioning is not suitable for the diagonal classification boundaries when the number of fuzzy sets is even (not odd). As Incremental method A eventually considers all training patterns equally at the end of the process. This method should be better than Incremental method B which updates fuzzy if-then rules in a smoothing manner with a positive constant γ.

The results of the experiments by Incremental method B are shown in Fig. 4. Figure 4 shows the classification performance of the constructed fuzzy rule-based classifiers over time with the specified number of available training patterns and a positive constant (i.e., γ). Obviously the classification performance by Incremental method A is better than Incremental method B as the classification problem is static, that is, they remain unchanged during the course of the experiments. As for the specification of the positive constant, the classification performance is the best when $\gamma = 0.1$ among the investigated values for any number of fuzzy partitions except two. As shown in the experiments for Incremental method A, it is difficult to achieve the diagonal classification boundaries by two fuzzy sets for each axis. Thus the classification performance in this parameter specification is rather unstable in the low classification rates.

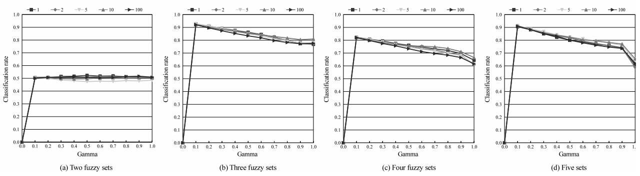

Fig. 4. Experimental results by Incremental method B for the static problem.

B. Dynamic Pattern Classification Problem

In this subsection, a dynamic incremental classification problem is considered where the classification boundaries changes over time. The initial classification boundaries are specified as the same ones as in Fig. 3. However, they rotate clockwise centring on the point $(0.5, 0.5)$ by one degree per time step. A single run of the experiments was conducted for 360 time steps until the classification boundaries get back to their initial position. During the experiments, the classification boundaries generated by the constructed fuzzy rule-based classifiers were investigated.

Only the experimental results by Incremental method B are shown as Incremental method A does not perform well. This is because Incremental method A takes the past training patterns equally into the consideration in the updating procedure of fuzzy if-then rules, leading to poor classification performance. Figure 5 shows the classification boundaries by Incremental method B in a run of the experiments. The parameter specification for this experiment is as follows: The number of new training patterns per time step: 10, $\gamma = 0.1$, and five fuzzy sets for each axis. This figure also shows the true classification boundaries by broken lines. The time step is 290, which means the classification boundaries have rotated 290 degrees from their initial configurations. It is shown that the generated classification boundaries by Incremental method B almost catches the true ones even in the dynamic situation.

VI. Conclusions

Two incremental versions of fuzzy rule-based classifiers was proposed in this paper: Incremental method A and B. The target classification problems are supposed to give streaming

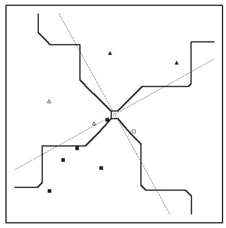

Fig. 5. Experimental results by Incremental method B for the dynamic problem. The true classification boundaries have rotated 290 degrees clockwise from their initial configurations.

training patterns that are made available over time. Fuzzy if-then rules have to be updated according to the new training patterns. Through a series of computational experiments, the performance of the proposed fuzzy rule-based classifiers was examined by two types of classification problems: static and dynamic. The experimental results showed that Incremental method A performs better than Incremental method B for the static problem whereas Incremental method B was able to adapt itself to the dynamic classification problem. This paper is the first attempt to introduce incremental approaches into fuzzy rule-based classifiers for those pattern classification problem where training patterns are available through time. Future works include investigating the performance for dynamic situation quantitatively and improving it. The analysis of computational cost for the incremental learning should be also considered for the implementation of the methods in real-world applications. The application of the proposed method to real-world problems are also considered in our future works.