Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 158 (2007) 284–294
FUZZY sets and systems
www.elsevier.com/locate/fss

A weighted fuzzy classifier and its application to image processing tasks

Tomoharu Nakashima[a], Gerald Schaefer[b,*], Yasuyuki Yokota[a], Hisao Ishibuchi[a]

[a] *College of Engineering, Osaka Prefecture University, Japan*
[b] *School of Engineering and Applied Science, Aston University, UK*

Available online 20 October 2006

Abstract

Many image processing applications involve a pattern classification stage. In this paper we propose a classifier based on fuzzy if–then rules that allows the incorporation of weighted training patterns which can be used to adjust the sensitivity of the classification with respect to certain classes. The antecedent part of fuzzy if–then rules are specified by partitioning each attributes into fuzzy sets while the consequent class and the degree of certainty are determined from the compatibility and weights of training patterns. We also introduce a learning method which adjusts the degree of certainty in order to provide improved classification performance and reduced costs. Experimental results on several image processing tasks demonstrate the efficacy of the proposed method.

Keywords: Fuzzy classifier; Fuzzy if–then rules; Pattern classification; Image processing

1. Introduction

While in the past fuzzy rule-based systems have been mainly applied to control problems [12,9], recently they have also been used in pattern classification tasks. Various methods have been proposed for the automatic generation of fuzzy if–then rules from numerical data for pattern classification [1–6,8,10].

Many image processing applications involve, after some pre-processing and feature selection and extraction, a pattern classification stage. Typical examples include face recognition where each image is assigned to one person, or the medical diagnosis of patient imagery. In some of these certain classes can be judged as bearing higher importance than others. In such cases the misclassification/rejection of a particular input pattern will cause extra costs. For example, in medical diagnosis of cancer, diagnosing people with cancer as not having the disease could be penalised more than diagnosing healthy individuals as cancer candidates.

In this paper we introduce a pattern classification algorithm based on fuzzy if–then rules that allows more emphasis to be put on one or more classes. We regard pattern classification as a cost minimisation problem and employ the concept of weights for training patterns. The weight of an input pattern can be viewed as the cost of misclassification/rejection

* Corresponding author. Tel.: +44 121 2043470; fax: +44 121 2043681.
E-mail address: G.Schaefer@aston.ac.uk (G. Schaefer).

0165-0114/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2006.10.011

of it. Fuzzy if–then rules are generated by considering the weights as well as the compatibility of training patterns. We also propose a learning method, based on incremental learning principles, for the generated rules which adjusts the grades of certainty in order to improve the classification performance and reduce the overall cost.

We apply our classification algorithm to several image processing applications such as the diagnosis of breast cancer from images of fine needle aspirates of breast mass, and the classification of satellite images. The experimental results confirm the efficacy of our proposed method as compared to standard fuzzy classification approaches.

The rest of the paper is organised as follows. Section 2 describes the concept of pattern classifiers based on fuzzy if–then rules. Section 3 introduces our fuzzy classifier that employs weighted training patterns. Section 4 then details our proposed learning algorithm while Sections 5 and 6 provide experimental results on several imaging data sets. Section 7 concludes the paper.

2. Fuzzy classification

Let us assume that our pattern classification problem is an n-dimensional problem with M classes and m given training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$. Without loss of generality, we assume each attribute of the given training patterns to be normalised into the unit interval [0, 1]; that is, the pattern space is an n-dimensional unit hypercube $[0, 1]^n$. In this study we use fuzzy if–then rules of the following type as a base of our fuzzy rule-based classification systems:

$$\begin{aligned}&\text{Rule } R_j\text{: If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn}\\ &\text{then Class } C_j \text{ with } CF_j, \quad j = 1, 2, \ldots, N,\end{aligned} \tag{1}$$

where R_j is the label of the jth rule, $A_{j1}, \ldots, A_{jn}$ are antecedent fuzzy sets on the unit interval [0, 1], C_j is the consequent class (i.e. one of the M given classes), and CF_j is the grade of certainty of the fuzzy if–then rule R_j. As antecedent fuzzy sets we use triangular fuzzy sets as in Fig. 1 where we show a partition of the unit interval into three fuzzy sets.

Our fuzzy rule-based classification system consists of N fuzzy if–then rules each of which has a form as in Eq. (1). There are two steps in the generation of fuzzy if–then rules: specification of antecedent part, and determination of consequent class C_j and grade of certainty CF_j. The antecedent part of the rules is initialised manually. Then the consequent part (i.e. consequent class and grade of certainty) is determined from the given training patterns [4]. In [7] it is shown that the use of the grade of certainty in fuzzy if–then rules allows us to generate comprehensible classification systems with high classification performance.

2.1. Fuzzy rule generation

Let us assume that m training patterns $\mathbf{x}_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, \ldots, m$, are given for an n-dimensional C-class pattern classification problem. The consequent class C_j and the grade of certainty CF_j of a fuzzy if–then rule are

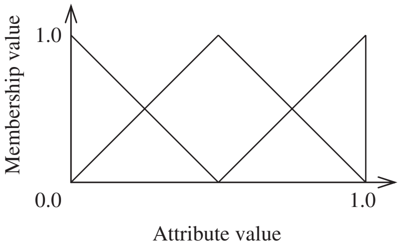

Fig. 1. Triangular fuzzy membership function.

determined in the following two steps:

1. Calculate $\beta_{\text{Class } h}(j)$ for Class h as

$$\beta_{\text{Class } h}(j) = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_j(\mathbf{x}_p), \tag{2}$$

where

$$\mu_j(\mathbf{x}_p) = \mu_{j1}(x_{p1}) \cdot \ldots \cdot \mu_{jn}(x_{pn}), \tag{3}$$

and $\mu_{jn}(\cdot)$ is the membership function of the fuzzy set A_{jn}. In this paper we use triangular fuzzy sets as in Fig. 1.

2. Find Class $\hat{h}$ that has the maximum value of $\beta_{\text{Class } h}(j)$:

$$\beta_{\text{Class } \hat{h}}(j) = \max_{1 \leqslant k \leqslant C} \{\beta_{\text{Class } k}(j)\}. \tag{4}$$

If two or more classes take the maximum value, the consequent class C_j of the rule R_j cannot be determined uniquely. In this case, C_j is specified as $C_j = \phi$. Thus each fuzzy if–then rule has only a single consequent class and cannot have multiple consequent classes. If a single class $\hat{h}$ takes the maximum value, let C_j be Class $\hat{h}$. The grade of certainty CF_j is determined as

$$CF_j = \frac{\beta_{\text{Class } \hat{h}}(j) - \bar{\beta}}{\sum_h \beta_{\text{Class } h}(j)} \tag{5}$$

with

$$\bar{\beta} = \frac{\sum_{h \neq \hat{h}} \beta_{\text{Class } h}(j)}{c - 1}. \tag{6}$$

2.2. Fuzzy reasoning

Using the rule generation procedure outlined above we can generate N fuzzy if–then rules as in Eq. (1). After both the consequent class C_j and the grade of certainty CF_j are determined for all N rules, a new pattern $\mathbf{x} = (x_1, \ldots, x_n)$ can be classified by the following procedure:

1. Calculate $\alpha_{\text{Class } h}(\mathbf{x})$ for Class h, $j = 1, \ldots, C$, as

$$\alpha_{\text{Class } h}(\mathbf{x}) = \max\{\mu_j(\mathbf{x}) \cdot CF_j | C_j = h\}. \tag{7}$$

2. Find Class h' that has the maximum value of $\alpha_{\text{Class } h}(\mathbf{x})$:

$$\alpha_{\text{Class } h'}(\mathbf{x}) = \max_{1 \leqslant k \leqslant C} \{\alpha_{\text{Class } k}(\mathbf{x})\}. \tag{8}$$

If two or more classes take the maximum value, then the classification of **x** is rejected (i.e. **x** is left as an unclassifiable pattern), otherwise **x** is assigned to Class h'.

2.3. A numerical example (1)

In this subsection we show a simple numerical example of the fuzzy classification system. Let us suppose that 20 training patterns are given in a two-dimensional pattern space as shown in Fig. 2. Ten of these are of Class 1 and the other 10 of Class 2.

Based on these given training patterns, we generate fuzzy if–then rules following the procedure described in Section 2.1. Nine fuzzy if–then rules are generated to form the classification system as each attribute is partitioned into three fuzzy sets (see Fig. 1); the generated rules are depicted in Fig. 3. The classification boundary generated by the classifier is shown in Fig. 4 from where we can see that two training patterns are misclassified by the system.

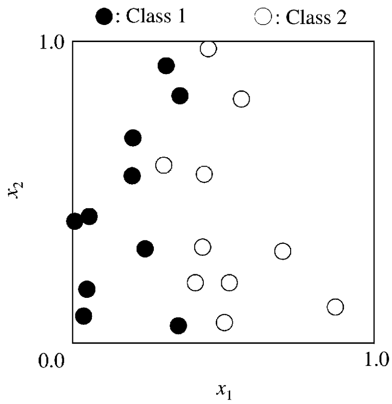

Fig. 2. A two-dimensional example.

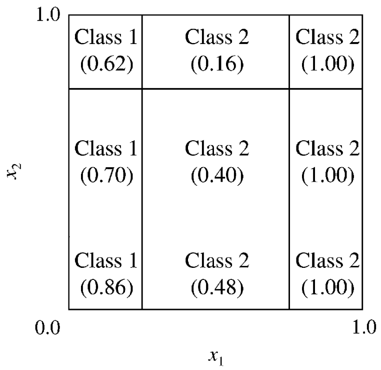

Fig. 3. Fuzzy if–then rules generated from 20 training patterns.

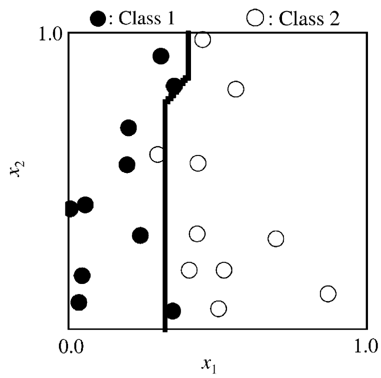

Fig. 4. Classification boundary by a fuzzy classification system.

3. Weighted fuzzy classification

In this paper we extend the principle of fuzzy classification to accommodate weighted training patterns. The idea is based on an understanding that in certain cases misclassification of a particular input pattern will cause extra costs. For example in diagnosis of cancer, diagnosing people with cancer as not having the disease could be penalised more than diagnosing healthy individuals as cancer candidates.

The pattern classification problem is re-formulated as a cost minimisation problem. The concept of weight is introduced for each training pattern in order to handle this situation. The weight of an input pattern can be viewed as the cost of misclassification/rejection of it. Fuzzy if–then rules are generated by considering the weights as well as the compatibility of training patterns.

In order to incorporate the concept of weight, Eq. (2) of the fuzzy rule generation is modified to

$$\beta_{\text{Class } h}(j) = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_j(\mathbf{x}_p) \cdot \omega_p, \tag{9}$$

where ω_p is the weight associated with training pattern p.

We note that this fuzzy rule generation method can also be applied to the standard pattern classification problem where there are no pattern weights. In this case, the class and the grade of certainty are determined from training patterns by specifying a pattern weight as $\omega_p = 1$ for $p = 1, \ldots, m$.

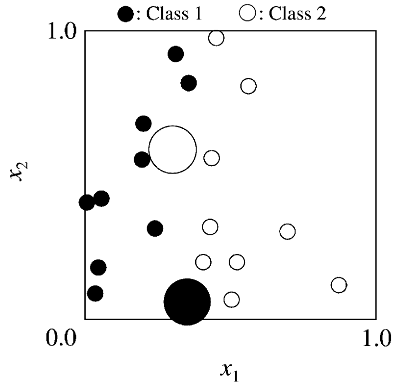

Fig. 5. Weighted training patterns.

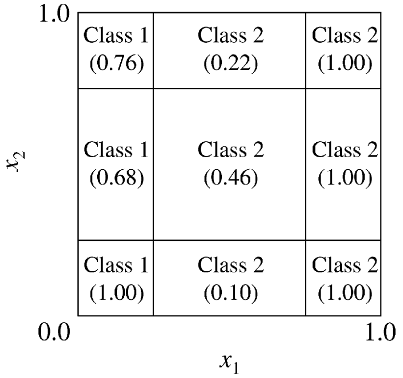

Fig. 6. Generated fuzzy if–then rules from the weighted training patterns.

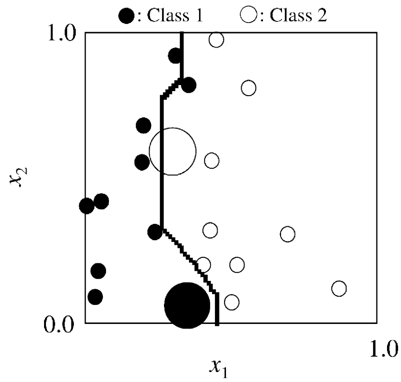

Fig. 7. Classification boundary by the weighted fuzzy classification system.

3.1. A numerical example (2)

Let us consider again the two-dimensional numerical example from Fig. 2 but in contrast we now assume that each training pattern has a weight assigned to it. We illustrate this situation in Fig. 5 where the size of the circles indicates the assigned weights (the assigned weights are 3 and 1, respectively, for the large and small circles).

We generate a weighted fuzzy classification system as described above from these weighted training patterns. The generated fuzzy if–then rules are shown in Fig. 6 and the resulting classification boundary is given in Fig. 7. We can see that although one training pattern is misclassified those important patterns with a larger weight (i.e. 3) are correctly classified while they were misclassified by the conventional fuzzy classification system in Section 2.3. The total cost (i.e. the sum of weights from misclassified patterns) by the weighted fuzzy classification is 2.0 ($= 1.0 \times 2$) which is significantly smaller than the 6.0 ($= 3.0 \times 2$) achieved by the conventional fuzzy classification system.

4. Learning fuzzy if–then rules for weighted training patterns

A learning method that adjusts the grades of certainty CF_j can be employed to achieve improved classification performance. It is based on an incremental learning approach where the adjustment occurs whenever classification of training patterns is performed. When a training pattern is correctly classified we reinforce the grade of certainty of the fuzzy if–then rule that is used for the classification. On the other hand, we decrease the grade of certainty of a fuzzy if–then rule if a training pattern is not successfully classified.

Let us assume that we have generated fuzzy if–then rules by the rule-generation procedure detailed in Section 2.1. We also assume that a fuzzy if–then rule R_j is used for the classification of a training pattern $\mathbf{x}_p$. That is, R_j has the

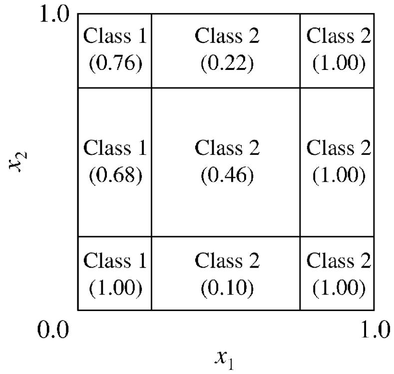

Fig. 8. Fuzzy if–then rules after the learning.

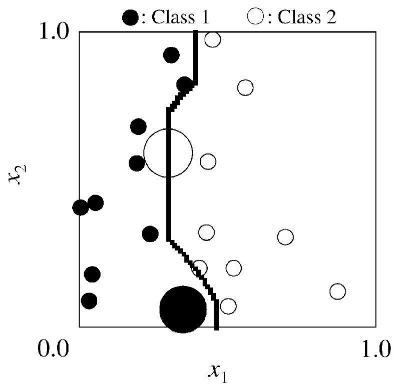

Fig. 9. Classification boundary after the learning of fuzzy if–then rules.

maximum product of the compatibility and the grade of certainty (see Eq. (7)). The proposed learning method adjusts the grades of certainty of R_j as follows:

$$CF_j^{\text{new}} = CF_j^{\text{old}} - \eta \cdot \omega_p \cdot CF_j^{\text{old}} \quad \text{if } \mathbf{x}_p \text{ is misclassified} \tag{10}$$

and

$$CF_j^{\text{new}} = CF_j^{\text{old}} + \eta \cdot \omega_p \cdot (1 - CF_j^{\text{old}}) \quad \text{if } \mathbf{x}_p \text{ is correctly classified,} \tag{11}$$

where ω_p is the weight of the training pattern $\mathbf{x}_p$, and η (the learning rate) is a positive constant value in the interval [0; 1].

One epoch of the proposed learning method involves examining all given training patterns. Thus, there will be m adjustments of fuzzy if–then rules after all m training patterns are examined. The learning process is summarised as follows:

1. Generate fuzzy if–then rules from m given training patterns by the procedure in Section 2.1.
2. Set K as $K = 1$.
3. Set p as $p = 1$.
4. Classify $\mathbf{x}_p$ by using the fuzzy if–then rules generated in Step 1.
5. After $\mathbf{x}_p$ is classified, adjust the grades of certainty using Eqs. (10) or (11).
6. If $p < m$ let $p := p + 1$ and go to Step 4. Otherwise go to Step 7.
7. If K reaches a pre-specified value, stop the learning procedure. Otherwise let $K := K + 1$ and go to Step 3.

Note that K in the above learning procedure corresponds to the number of epochs.

4.1. A numerical example (3)

We return to the numerical example of 20 training patters from above and apply the learning procedure with $\eta = 0.02$ for 100 epochs. The rules generated by the weighted fuzzy classification system after learning are shown in Fig. 8. The resulting classification boundary is given in Fig. 9 from where we can see that now all 20 training samples are correctly classified and hence the total misclassification cost is reduced to 0.0.

5. Experimental results for weighted classifier

In this section we examine the performance of the proposed weighted classifier described in Section 3 (that is *without* the learning algorithm introduced in Section 4) in various imaging applications. Under the assumption that a weight is assigned to each training pattern which can be viewed as the relative importance of the patterns, we use the concept of classification/rejection cost to construct a weighted fuzzy classification with integrated learning as outlined above.

We define a cost function $\texttt{Cost}(S)$ of a fuzzy classification system S as follows:

$$\texttt{Cost}(S) = \sum_{p=1}^{m} \omega_p \cdot z_p(S), \tag{12}$$

where m is the number of training patterns, ω_p is the weight of the training pattern $\mathbf{x}_p$, and $z_p(S)$ is a binary variable set according to the classification result of the training pattern $\mathbf{x}_p$ by S: $z_p(S) = 0$ if $\mathbf{x}_p$ is correctly classified by S, and $z_p(S) = 1$ otherwise (i.e. $\mathbf{x}_p$ is misclassified or rejected). We use this cost function as well as the classification rate as performance measures.

In order to construct a fuzzy classification system it should be determined how to partition each attribute variable into fuzzy sets. In this paper we divide each axis into three fuzzy sets as shown in Fig. 1. The number of generated fuzzy if–then rules in a fuzzy classification system depends on the partition of attributes and the dimensionality of the pattern classification problem. Since there are three fuzzy sets for each attribute, the possible number of combinations of antecedent fuzzy sets is $N = 3^n$ where n is the number of attributes.

We use two weight assigning methods for determining the weights of patterns. In the first method we assume it is important to correctly classify a certain class. Thus the weights of training patterns of this focussed class are specified as $\omega_p = 1.0$. On the other hand the weights of the other training patterns are specified as $\omega_p = 0.5$. That is, the cost of misclassifying/rejecting a training pattern from the focussed class is twice as large as that from the other classes. The second weight assigning method considers the distribution of classes in a data set. The weight of a class specified by this method is large if the proportion of the class is small. Thus it is assumed that classification of minor classes with a small number of patterns is more important than major classes with a large number of training patterns, a situation that is often the case for e.g. medical data sets. The weight of a training pattern $\mathbf{x}_p$ from Class c is specified by the inverse of the proportion of the class over the given training patterns as follows:

$$\omega_p = \omega_c = \frac{1}{Z} \cdot \frac{m}{N_c}, \quad p = 1, \ldots, m, \quad c = 1, \ldots, C, \tag{13}$$

where ω_p is the weight of the training pattern $\mathbf{x}_p$ that is from Class c, w_c is the weight of Class c patterns, m is the number of given training patterns, N_c is the number of Class c patterns, and Z is a normalisation factor that makes the maximum value of Class weights a unit value (i.e. $\max_c \omega_c = 1$).

5.1. Breast cancer diagnosis

The first application we evaluate our proposed classifier on is the diagnosis of breast cancer from digitised images of fine needle aspirates of breast mass. Fluid samples were extracted using a fine needle from the patient's breast mass, placed on a glass slide and stained to highlight the nuclei of constituent cells [11]. From the captured images a number of features were derived which were then used as the input patterns for classification. First, curve-fitting techniques were applied to extract the boundaries of the nuclei. For each nucleus 10 features were extracted, namely radius, standard deviation of grey-scale values, perimeter, area, smoothness (local variation in radius lengths), compactness, concavity, number of concave parts, symmetry and fractal dimension. The mean, standard deviation and maximum values of these over all nuclei in the image were then calculated to provide a feature vector with 30 values [11].

In total the data set comprises 569 samples of which 357 are known to constitute benign and the remaining 212 malignant cases. Since there are two classes (i.e. benign and malignant), we examined the performance of fuzzy classifiers with three weight assignment: benign focussed (1.0 for benign patterns and 0.5 for malignant patterns), malignant focussed (1.0 for malignant patterns and 0.5 for benign patterns), and class-proportional. Ten-fold cross validation was performed where the given data set is divided into 10 subsets and each subset is used as test data set while the other nine data sets are used as training patterns. The experimental results, expressed in terms of classification rate and total cost are listed in Tables 1 and 2 which also provide results for a conventional fuzzy rule-based classifier as described in Section 2. We note that the performance of the conventional method is constant as it does not consider the weight of training patterns. Table 1 shows the performance for the training patterns and Table 2 that of the test data set. From there we can see that in two of the three cases there is a clear improvement, both in terms of overall cost (the main aim of the proposed classifier) as well as in classification performance. For both the benign-focussed and class-proportional cases the cost is more than halved as compared to a standard fuzzy classifier. In turn the classification rate is improved from 89.98 to 92.97 and 94.38, respectively.

Table 1
Experimental results for training patterns on breast cancer diagnosis

	Classification rate (%)		Cost	
	Proposed	Conventional	Proposed	Conventional
Benign focussed	92.26	89.28	**2.74**	6.04
Malignant focussed	72.37	89.28	7.89	**3.12**
Class-proportional	93.86	89.28	**2.88**	6.05

Table 2
Experimental results for test patterns on breast cancer diagnosis

	Classification rate (%)		Cost	
	Proposed	Conventional	Proposed	Conventional
Benign focussed	92.97	89.98	**24**	57
Malignant focussed	72.58	89.98	78	**28.5**
Class-proportional	94.38	89.98	**26.31**	57

Table 3
Weight specification for satellite image classification.

	Class					
	Red	Cotton	Grey	Damp	Veg. stubble	Very damp
Red focussed	1.0	0.5	0.5	0.5	0.5	0.5
Veg. stubble focussed	0.5	0.5	0.5	0.5	1.0	0.5
Very damp focussed	0.5	0.5	0.5	0.5	0.5	1.0
Class-proportional	0.387	0.866	0.432	1.000	0.883	0.400

5.2. Satellite image classification

The second application is concerned with the classification of satellite image data. A database of a number of satellite images obtained from the Landsat multi-spectral scanner were obtained [13]. Each image consists of four spectral bands, two in the visible part of the spectrum (corresponding roughly to green and red sensitivities) and the remaining two in the infrared. Parts of these images were segmented and, using a moving window of 3×3 pixels at each position, a feature vector of $3 * 3 * 4 = 36$ pixel values extracted [13].

The data set comprises 2000 training vectors and 4435 test samples. Each sample is a member of one of the following six classes: red soil, cotton crop, grey soil, damp grey soil, vegetation stubble, and very damp grey soil. As examples we designed weighted fuzzy classifiers emphasising the importance of the red soil, vegetable stubble, and very damp grey soil classes, respectively, as well as a classifier following the class-proportional method from Eq. (13) (the distribution of weights for these cases is given in Table 3). Tables 4 and 5 show the classification rates and total cost for the training and test patterns, respectively. We see that in the red soil and vegetable stubble focussed cases our proposed classifier provides both a lower cost and a better classification performance compared to a conventional classifier. On the other hand, for the other two cases we get results which are worse than those of a standard fuzzy classification system. This suggests that the weighted classifier alone, though providing good performance, cannot perfectly adapt and hence illustrates the need for an adaptive learning process as the one proposed in Section 4.

5.3. Image segmentation

The third data set we investigated is an image segmentation data set provided by the Vision Group at the University of Massachusetts [13]. Several outdoor images were manually segmented into areas corresponding to seven classes, namely brickface, sky, foliage, cement, window, path, and grass to provide a ground truth for the data set. The images were then divided into 3×3 regions and various features extracted to serve as input patterns for the classification stage.

Table 4
Experimental results for training patterns on satellite image classification

	Classification rate (%)		Cost	
	Proposed	Conventional	Proposed	Conventional
Red focussed	60.85	57.65	**394.0**	582.5
Veg. stubble focussed	59.65	57.65	**489.5**	542.0
Very damp focussed	51.25	57.65	494.5	**459.0**
Class-proportional	52.65	57.65	598.79	**578.25**

Table 5
Experimental results for test patterns on satellite image classification

	Classification rate (%)		Cost	
	Proposed	Conventional	Proposed	Conventional
Red focussed	62.66	62.86	**837.0**	1315.0
Veg. stubble focussed	63.99	62.86	**951.5**	1058.0
Very damp focussed	55.51	62.86	1072.0	**877.0**
Class-proportional	55.51	62.86	1214.90	**1132.56**

Table 6
Experimental results for training patterns on the image segmentation data set

	Classification rate (%)		Cost	
	Proposed	Conventional	Proposed	Conventional
Brickface focussed	63.84	63.9	**3.89**	4.19
Sky focussed	59.12	63.9	4.3	**3.8**
Cement focussed	65.58	63.9	**3.84**	5.29
Class proportional	65.52	63.9	**6.72**	7.06

Table 7
Experimental results for test patterns on the image segmentation data set

	Classification rate (%)		Cost	
	Proposed	Conventional	Proposed	Conventional
Brickface focussed	68.57	69.05	**33.5**	65
Sky focussed	64.29	69.05	37.5	**35.5**
Cement focussed	72.86	69.05	**29.05**	33
Class proportional	69.05	69.05	**65**	**65**

The extracted features comprise the results of edge and line detection algorithms and colour and intensity information for each region [13].

The data set contains 210 instances of which 30 patterns are given from each class. Each pattern consists of 19 attributes. Similar to Section 5.2 we designed classifiers focussing on some of the segmentation classes, namely brickface, sky and cement, respectively, as well as a class-proportional classifier. Ten-fold cross validation was then performed on the data set. The results for training and test patterns are given in Tables 6 and 7. Compared to a conventional fuzzy rule-based classifier our proposed method reduces the cost in two of the cases (for the brickface focussed classifier by about half) while maintaining the same performance for the class-proportional case and giving slightly higher costs for the sky focussed experiment.

6. Experimental results of classifier with learning

The results shown and discussed in the previous section confirm that the proposed weighted fuzzy classification system provides, in the majority of cases, reduced overall cost (and often also improved classification). However, in

Table 8
Classification results on breast cancer data set after learning

	Classification rate (%)		Cost	
	With learning	Conventional	With learning	Conventional
Benign focussed	**89.99**	89.98	**45**	57
Malignant focussed	88.06	**89.98**	37	**28.5**
Class-proportional	**91.74**	89.98	**39.65**	57

Table 9
Classification results on satellite image data set after learning

	Classification rate (%)		Cost	
	With learning	Conventional	With learning	Conventional
Red focussed	**67.85**	62.86	**897.5**	1315.0
Veg. stubble focussed	61.20	**62.86**	**955.5**	1058.0
Very damp focussed	**63.47**	62.68	**852**	877.0
Class-proportional	**63.00**	62.68	**910.44**	1132.56

Table 10
Classification results on image segmentation data set after learning

	Classification rate (%)		Cost	
	With learning	Conventional	With learning	Conventional
Brickface focussed	**70.95**	69.05	**37.5**	65
Sky focussed	**74.29**	69.05	**27**	35.5
Cement focussed	**77.14**	69.05	**29**	33
Class proportional	**73.81**	69.05	**55**	65

certain cases it failed to achieve a better solution. In Section 4 we have introduced a learning algorithm for our classifier which is designed to adjust the grade of certainty so as to provide better classification performance and hence a more cost effective solution. In this section we discuss the experimental results obtained on the same data sets as in Section 5 but using the weighted classifier with integrated learning.

For each data set and each weight assigning method given in Section 5 we applied the learning algorithm to improve the performance of the classification system. For the learning rate η and the number of epochs K we experimented with various parameters ($\eta = 0.1, 0.2, \ldots, 0.5$ $K = 5 \ldots 10$); we show the best results in the following.

The experimental results are shown in Tables 8, 9, and 10 for the breast cancer diagnosis, satellite image and image segmentation tasks, respectively, and again compared to a classical fuzzy rule-based classifier. From the tables it is obvious that after learning much improved results are achieved. Except for the malignant focussed case for the breast cancer data set the total cost of the weighted classifier with learning is now lower in all cases as compared to the standard classifier. Also, comparing these results with those presented in Section 5 we see that a significant improved classification rate is obtained in almost all experiments which demonstrates the efficacy of our approach.

7. Conclusions

In this paper we have introduced a fuzzy classification system that incorporates weighted training patterns. The proposed classifier has been employed in a number of image processing applications and experimental results have demonstrated that it provides improved performance compared to conventional fuzzy classification approaches.

We have also introduced a learning algorithm which adjusts the grade of the certainty of the fuzzy rules and shown that its application greatly improves the performance both in terms of the classification rate and classification cost.

Acknowledgements

The authors would like to acknowledge the support of the Daiwa Foundation.

References

[1] M. Grabisch, The representation of importance and interaction of features by fuzzy measures, Pattern Recognition Lett. 17 (1996) 567–575.
[2] M. Grabisch, F. Dispot, A comparison of some methods of fuzzy classification on real data, in: Second Internat. Conf. on Fuzzy Logic and Neural Networks, 1992, pp. 659–662.
[3] M. Grabisch, J.-M. Nicolas, Classification by fuzzy integral: performance and tests, Fuzzy Sets and Systems 65 (2/3) (1994) 255–271.
[4] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets and Systems 52 (1) (1992) 21–32.
[5] H. Ishibuchi, T. Nakashima, Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes, IEEE Trans. Industrial Electron. 46 (6) (1999) 1057–1068.
[6] H. Ishibuchi, T. Nakashima, Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems, IEEE Trans. Systems, Man Cybernetics—Part B: Cybernet. 29 (1999) 601–618.
[7] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Systems 9 (4) (2001) 506–515.
[8] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic, Prentice-Hall, Englewood Cliffs, NY, 1995.
[9] C.C. Lee, Fuzzy logic in control systems: fuzzy logic controller part I and part II, IEEE Trans. Systems Man Cybernet. 20 (1990) 404–435.
[10] K. Nozaki, H. Ishibuchi, H. Tanaka, Adaptive fuzzy rule-based classification systems, IEEE Trans. Fuzzy Systems 4 (3) (1996) 238–250.
[11] W.N. Street, W.H. Wolberg, O.L. Mangasarian, Nuclear feature extraction for breast tumor diagnosis, SPIE Internat. Symp. on Electronic Imaging: Science and Technology, Vol. 1905, 1993, pp. 861–870.
[12] M. Sugeno, An introductory survey of fuzzy control, Inform. Sci. 30 (1/2) (1985) 59–83.
[13] C.C. Taylor, D. Michie, D.J. Spiegelhalter, Machine Learning, Neural and Statistical Classficiation, 1994.