Application of KITSAT-3 Images: Automated Generation of Fuzzy Rules and Membership Functions for Land-cover Classification of KITSAT-3 Images

WONKYU PARK, SOONDAL CHOI
Satellite Technology Research Center, KAIST
373-1 Yusong-gu, Kusong-dong, Taejon, 305-701
E-mail: wpark@satrec.kaist.ac.kr
Korea

KEY WORDS: KITSAT-3, Fuzzy Logic, Land-Cover, Classification, Membership Function, Rules

ABSTRACT: The paper presents an automated method for generating fuzzy rules and fuzzy membership functions for pattern classification from training sets of examples and an application to the land-cover classification. Initially, fuzzy subspaces are created from the partitions formed by the minimum and maximum of individual feature values of each class. The initial membership functions are determined according to the generated fuzzy partitions. The fuzzy subspaces are further iteratively partitioned if the user-specified classification performance has not been archived on the training set. Our classifier was trained and tested on patterns consisting of the DN of each band, (XS1, XS2, XS3), extracted from KITSAT-3 multispectral scene. The result represents that our classification method has higher generalization power.

1. INTRODUCTION

KITSAT-3 has been launched in May 26, 1999. One of the key missions of KITSAT-3 is earth observation, and it is successfully transmitting the earth observation images. KITSAT-3 has multi-spectral camera with 13.5m (nadir) resolution and the image is adequate for land-cover classification in the sense of resolution and multispectral characteristics. However, so far, classifications using the conventional methods such as maximum likelihood classifier was erroneous, because it assumes the normal distribution for each class patterns, which is not true in the satellite images.

Fuzzy rule-based methods have been used in a wide range of engineering applications (Park *et al.* 1996, Cox 1994) In most of the cases, fuzzy rules were generated from human experts' knowledge about the objects and the corresponding membership functions were intuitively adjusted. Recently, there have been several methods proposed for constructing fuzzy rules and calculating membership functions based on learning schemes of neural networks, and/or genetic algorithms from training set (Abe and Lan 1995, Simpson 1992, Ishibuchi *et al.* 1992, Ishbuchi *et al.* 1995, Homifar 1995). However, only a few of the reported techniques were developed targeting pattern classification problems. In Abe and Lan 1995, Simpson 1992), expandable hyperboxes were used and each box defined a region of a multidimensional pattern space containing patterns with full class membership. The membership functions were defined with respect to the hyperbox min-max points by the learning method of a three-layer neural network Hence, the classification system was called fuzzy min-max neural network (Simpson 1992).

Ishibuchi *et al.* 1995 proposed a method for selecting the superior fuzzy rules and membership functions for pattern classification problems based on a training set using genetic algorithm. However, due to the use of genetic algorithm optimization, the method requires significant training time, especially for larger dimensional pattern spaces and larger numbers of target classes.

Keeping the inherent properties of fuzzy logic pattern classifiers, the aim of our study is to design a trainable system with the following properties.

- The system offers high performance fuzzy if-then rule-based classification.
- Training is controlled by a single tuning parameter that is easily understood by users.
- Computational cost is reduced by creating finer partitions where needed.
- System training time is short compared to other fuzzy learning algorithms.

To achieve the stated behavior, our method presented below adaptively partitions the feature space using *a priori* information. In this paper, the generation method of fuzzy rules and membership functions is summarized and its application to satellite image is introduced.

2. Adaptive Multi-Scale Feature Space Partitioning and Rule Generating

Our adaptive fuzzy partitioning method starts with an initial partitioning of the pattern space. Then, it iteratively partitions the subspaces in which the classification performance is low. Training of our method can be outlined as follows:

Step 1. Partition the pattern space using the adaptive grids defined by the minimum and maximum of individual feature values in each class.

Step 2. Select the fuzzy subspace in which the classification performance is the lowest and perform its additional partitioning.

Step 3. Stop if the classification error in the training set is lower than the pre-specified training error. Go to Step 2, otherwise.

These steps are now described in more detail.

2.1 Fuzzy Partitioning

When we select the appropriate features, the patterns in the pattern space for each class should look like a cluster or several scattered clusters. Then, training patterns within a fuzzy subspace formed by the maximum and minimum values of an individual feature axis in each class are likely to belong to a single class (Fig.1). However, subspaces containing patterns from different classes may present. Such subspaces would contribute to classification errors. Fig 1 shows the initially determined partitions. Values $\tilde{x}_1$, $\tilde{x}_4$ correspond to the minimum and maximum values of the class 0 patterns (marked by •) with respect to the x axis. Similarly, $\tilde{x}_2$, $\tilde{x}_3$ values represent the minimum and maximum values of the feature x in class 1 (marked by +)

As can be seen in Fig. 1 each fuzzy subspace tends to contain patterns from a specific class.

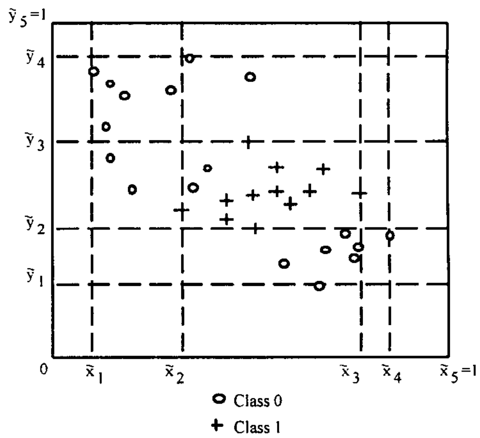

Fig 1. Partitions formed after the initial partition

2.2 Determination of Membership Functions and Rules

After the initial partitioning is performed, fuzzy membership functions and fuzzy rules are generated for each partition. The membership functions are shown in Fig. 2 The membership functions are affected by the width of the adjacent partitions. That is, it has ability to incorporate with information of the neighboring patterns.

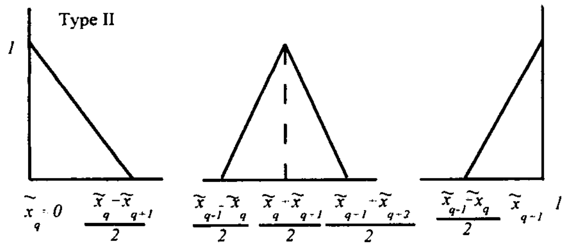

Fig 2. The membership functions.

For N-class classification problems in the pattern space $[0, 1]\times[0, 1]$, the fuzzy if-then rule corresponding to a *q*-*r* subspace created during the *w*-th iteration can be represented as

$$\begin{aligned} &Rule: R_{qr}^{w}: \textit{If } x \textit{ is } FS_q \textit{ and } y \textit{ is } FS_r, \textit{then} \\ &\qquad \bar{p} \textit{ belongs to a class } C_{qr} \\ &\qquad \textit{with } CCF = CCF_{qr} \\ &\qquad q = 0,1,\cdots 2N; r = 0,1,\cdots,2N \end{aligned}$$

where $\bar{p} = (x, y)$ is a pattern vector, R_{qr}^{w} is the name of the fuzzy if-then rule, and FS_q and FS_r are the names of *q*-th and *r*-th fuzzy sets which are represented by the fuzzy membership functions μ_q and μ_r

The parameters C_{qr} and CCF_{qr} can be calculated as suggested in the Ishibuchi's rule generation framework (Ishibuchi *et al.* 1995):

Step 1: Calculate β_L for each class $L = 0,1,\cdots,N-1$ as

$$\beta_L = \sum_{\bar{p}s' \text{ that belong to } L} \mu_q(x)\mu_r(y),$$

where β_L is the sum of the compatibility of $\bar{p}s'$ *that belong to* L to the fuzzy if-then rule R_{qr} (Ishibuchi *et al.* 1995).

Step 2: Find class *K* such that

$$\beta_K = \max\{\beta_0, \beta_1, \cdots, \beta_{N-1}\}$$

Step 3: If a single class *K* takes the maximum value in

$$CCF_{qr} = \frac{\beta_K - \beta}{\sum_{L=1}^{N-1} \beta_L},$$

where $\beta = \sum_{L=0, L\neq K}^{N-1} \frac{\beta_L}{N-1}$ and the consequent C_{qr} is *K*.

2.2 Iterative Generation of Fuzzy Membership Functions and Rules

In Fig. 1, the fuzzy subspace $[\tilde{x}_2, \tilde{x}_3]\times[\tilde{y}_2, \tilde{y}_3]$ contains a mixture of patterns from class 0 and class 1. Hence, the fuzzy subspaces inside of which there are patterns from more than one class need to be partitioned more finely. Therefore, the fuzzy subspace with the lowest classification performance is further partitioned as discussed earlier (see Fig. 3).

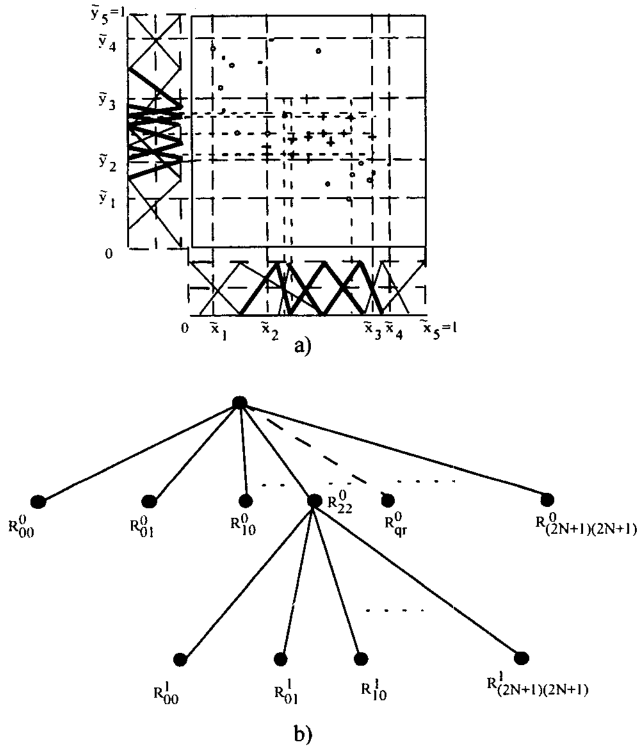

Fig 3. The membership functions and partitions after the final iteration

Hence, our method generates rules based on finer partitions in a hierarchical fashion. Consequently, the decision boundary can exhibit more substantial non-linearity.

2.3 Classification Using Generated Rules and Membership Functions

When a final rule set is generated following the above steps, a new pattern $\bar{p}$ is classified using the simplified defuzzification methods.

Step 1: Calculate α_T for each class T ($T = 0,1,\cdots,N-1$) as

$$\alpha_T = \max\{\mu_q(x)\cdot\mu_r(y)\cdot CCF_{qr} \mid C_{qr} = T \,\&\, R_{qr} \in S\}$$

Step 2: Find Class K such that

$\alpha_K = \max\{\alpha_0,\alpha_1,\cdots,\alpha_{N-1}\}$

Then, $\bar{p}$ belongs to class K.

3. Application to Satellite Image Classification

Land-use classification on satellite images is old but important problem in remote sensing society. Satellite image classification is difficult especially because of atmospheric scattering, topography, the sun and view angles, class mixture, and within-class reflectance variability. Hence, the degree of non-linearity becomes more serious so that a sophisticated classifier is necessary.

Most of image analysts tend to select training sites from regions where their classes are obvious. In this case, the generalization issue arises which is a common problem in pattern classification field. In order to achieve the generalization, a classifier must be able to capture entire characteristics of every class from small training sites. In the following section, the performance of our method is assessed and also compared with that of quadratic Gaussian classifier, 1 nearest neighbor classifier.

3.1 Data

Two kinds of satellite image were prepared. SPOT multi-spectral scene was used for accuracy assessment. We tested our classifier on KITSAT-3 image. The SPOT scene was taken in Sept. 24, 1997 over Kwang-Ju and Cheonra-BookDo, Korea. The KITSAT-3 scene was taken in Oct. 4, 1999 over Seosan and Dangjin area, Korea. The image telemetry was received and processed at Satellite Technology Research Center.

The pattern vectors consisted of the DNs of each band (i.e., [XS1, XS2, XS3]) and were linearly scaled to [0, 1].

3.2 Study protocol

We chose the popular 6 classes for SPOT images: water, urban, forest, agricultural area, bare land, and shadow.

Two image analysts participated for preparation of training sites and testing sites. One of them selected regions for each class by visual analysis of image with a help of maps (Set I), the other analyst performed the same task without knowing the previous image analyst's sites (Set II). The classification accuracy was assessed in two way; We performed leave-one-out test (Sonka *et al.* 1993) on Set I and Set II. Also, in order to test the generalization ability of our classifier and other classifier, Set I was used for training and Set II was for testing and vice versa. The classification performance is reported as the ratio of correctly classified patterns to the total number of test patterns.

3.3 Result

The performances of each classifier on training set, training set using leave-one-out test, and independent test set are summarized in Table 1, 2, and 3, respectively.

As shown in leaving-one-out test (Table 2), Set I may be close to linear patterns (even if Set I has non-linearity) because the 1-nearest neighbor classifier showed higher performance while Set II is a set of nonlinear patterns because the quadratic Gaussian classifier showed better result on Set II. However, in the cross validation procedure (Table 3), the classification performance using the quadratic Gaussian classifier and the 1-nearest classifier became worse.

Table 1. Training and testing on the same set from SPOT

	Classification Accuracy	
	Set I	Set II
Proposed method	93.18	94.18
Quadratic Gaussian	96.13	95.92
1-nearest neighbor	98.92	95.89

Table 2. Leaving-one-out test from SPOT

	Classification Accuracy	
	Set I	Set II
Proposed method	93.09	94.12
Quadratic Gaussian	96.12	95.92
1-nearest neighbor	96.60	92.18

Table 3. The performance of each classifier on the independent test set from SPOT

	Classification Accuracy	
	Training on Set I & Testing on Set II	Training on Set II & Testing on Set I
Proposed method	98.95	92.72
Quadratic Gaussian	94.73	87.89
1-nearest neighbor	94.21	89.96

Hence, we may conclude as follows.

- The quadratic Gaussian classifier over-fitted on nonlinear pattern set, Set II, so that the testing on linear Set I became worst.
- The 1-nearest classifier could not model the nonlinear decision boundary.
- Our proposed method could generalize no matter how we chose the training set. That is, our method has higher generalization power than other methods.

The SPOT image classified into 6 classes using our method trained on Set I is shown in Fig. 4

The classifier was tested on KITSAT-3 images. The training set was generated from KITSAT-3 images and the classification was performed on KITSAT-3 images. By visual inspection, each class was successfully distinguished except some pixels that belong to tree class were classified as shadows (See Fig 5). That's due to the transmission errors from the satellite. However, those transmission errors can be eliminated by a smart median filter that is remaining as a future work. Also, the brightness of urban area and bare land area is very similar so that missclassification occurred. However, even if the spectral response is similar, we can successfully distinguish two classes using texture analysis because the texture pattern is different.

4. Discussion

As seen from Tables 3, our method is better than the Gaussian classification algorithm and the 1-nearest neighbor classifier when we used independent test sites in this study. It means that our method has more generalization power that is most important in practical situation where collecting ground truth for training set is limited. However, the two conventional classification algorithm showed higher performance on training set itself and in leave-one-out test as shown in Table 1, 2. The results reflect that the conventional classifiers overfit the training data and cannot resolve the conflicts in the training data due to the imaging artifacts. Also, the conventional classifiers may overfit the noise training examples. This contributes to classification error. In contrary, our proposed method generates small size partitions (see Fig. 3) for highly non-linear decision boundary or for noise training patterns without changing the entire shape of the decision boundary. That is because our method generates the fuzzy partitions only where the partition is needed and our method tries to keep the partition as large as possible. Hence, our fuzzy classifier has a higher generalization power.

5. Conclusions

This paper focused on two major topics, the automated determination of fuzzy membership functions and rules for pattern classification, and its application to design an accurate and automated method for satellite image classification. The automated determination of fuzzy membership functions and rules was developed and successfully applied to the satellite image classification.

The key issue to accurate classification is "inductive" ability, meaning that from a limited number of training examples, the classifier should induce the entire characteristics of each class. In this point of view, our method showed better inductive power than other classification methods. The classification of KITSAT-3 images is successful in visual inspection, and the quantitative accuracy assessment is remaining for the future work.

References

W Park, E A Hoffman, and M Sonka. Fuzzy logic approach to extraction of intrathoracic airway trees from three-dimensional CT images. In *Image Processing, Proceedings SPIE Vol. 2710*, Bellingham, WA, 1996. SPIE.

E Cox. *The Fuzzy Systems Handbook.* AP Professional, Cambridge, 1994.

T T Phan and G Chen. A fuzzy logic approach to object recognition, In *Proceedings SPIE, vol. 2061, Applications of Fuzzy Systems*, pages 547-556, New York, 1994. SPIE.

S Abe and M Lan. A method for fuzzy rules extraction directly from numerical data and its application to pattern classification. *IEEE Trans. on Fuzzy Systems*, 3(2):129-139, 1995

P K Simpson. Fuzzy min-max neural networks - Part 1: Classification. *IEEE Trans. on Fuzzy Systems*, 3(2):129-139, 1992.

H Ishibuchi, K Nozaki, and H Tanaka. Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Sets and Syst.*, 52:21-32, 1992.

H Ishibuchi, K Nozaki, N Yamamoto, and H Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Trans. on Fuzzy Systems*, 3(2):260-270, 1995.

A Homifar and E McCormick. Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. *IEEE Trans. on Fuzzy Systems*, 3(2):129-139, 1995.

M Sonka, V Hlavac, and R Boyle. *Image Processing, Analysis, and Machine Vision.* Chapman and Hall, London, New York, 1993.

Fig 4. SPOT images and the classified image

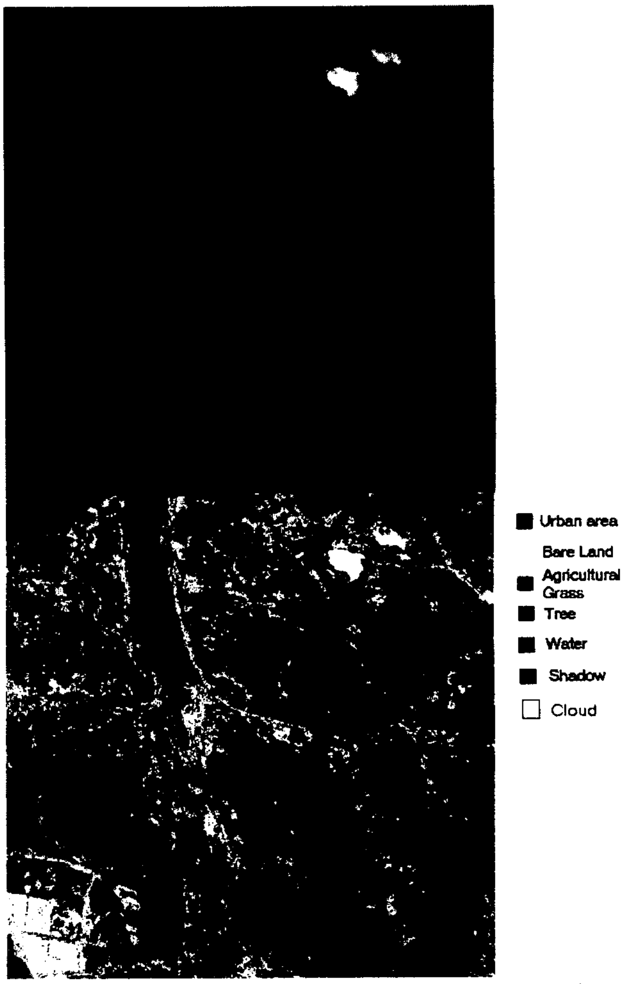

Fig 5 KITSAT-3 Images (above) and the classification result (below)