Fuzzy Classification Using EM Algorithm

Sang-Hoon Lee

Department of Industrial Engineering, Kyungwon University

ABSTRACT - This study proposes a fuzzy classification using EM algorithm. For cluster validation, this approach iteratively estimates the class-parameters in the fuzzy training for the sample classes and continuously computes the log-likelihood ratio of two consecutive class-numbers. The maximum ratio rule is applied to determine the optimal number of classes.

Key Words: Mixed Distribution, EM Estimation, Fuzzy Classification, Cluster Validation.

1. Introduction

The proposed method is in two stages. First, it finds the best partition of the image field using the spatial region growing segmentation (Lee, 2004), which is a hierarchical clustering operation (Anderberg, 1973) of step-by-step merging of smaller clusters into lager ones with the restriction that pixels in a cluster should be spatially contiguous. The segmentation algorithm is based on the local mutually closest neighbors (MCN) and multi-window operation using a pyramid-like structure to increase computational efficiency. Any two regions, which are adjacent each other in the partition, is supposed to be non-uniform on a given statistical criterion. The resultant partition from the segmentation is dependent on the criterion determining the level corresponding to the best partition in the hierarchy of the clustering algorithm. For the adequate criterion, the segmentation procedure then segregates inner regions, which correctly define the classes, from the uncertain regions in the boundaries between different classes. In the second stage, the fuzzy classification, which conducts iterative identification of expected maximum-likelihood parameters of the class (Liang *et al.*, 1992), is applied for the segmented regions in order to determine the sample classes of ground truth. This scheme iteratively estimates the parameters of the sample classes by decreasing the number of classes assumed for the scene of observed area, and continuously computes the ratio of log-likelihoods of two consecutive class-numbers. The optimal number of classes is determined at the point with the maximum ratio.

2. CN-chain Spatial Region Growing

One essential structural characteristic involves hierarchy of scene information. Under the constraint of the hierarchical structure, it is then possible to determine natural image segments by combining hierarchical clustering with spatial region growing. Hierarchical clustering (Anderberg, 1973) is an approach for step-by-step merging of small clusters into larger ones. Clustering algorithm utilize a similarity/dissimilarity measure that is computed between all pairs of candidates being considered for merging, a rule for selecting the pairs to be merged, and a rule for "cutting" the hierarchical tree. The computational efficiency of hierarchical clustering segmentation is mainly dependent on how to find the best pair to be merged. The closest neighbor of region j is defined as

$$\mathrm{CN}(j) = \arg\min_{k \in \mathbf{R}_j} d(j,k)$$

where $d(j,k)$ is the dissimilarity measure between regions j and k, and $\mathbf{R}_j$ is the index set of regions considered to be merged with region j. The pair of regions is then defined as MCN iff $k = \mathrm{CN}(j)$ and $j = \mathrm{CN}(k)$. It is easily shown that the best pair is one of the MCN's. Thus, the search is limited in the set of MCN's in the hierarchical clustering procedure.

For the region growing segmentation, the clustering procedure successively merges a pair among them of two regions which neighbor in image space, that is, $\mathbf{R}_j$ is the collection of the regions which are spatially adjacent with region. If all the pixels in the sample image are initially considered to be individual clusters, the algorithm requires exorbitant memory for the values associated with the merging process for large multi-channel imagery. According to the merger, the neighbor configuration and the set of MCN's must be updated, and the computational time for the update of the configurations and the search of the best pairs exponentially increases as the number of clusters in the initial state increases. To alleviate the memory problem and improve the computational performance of the algorithm, a multi-window strategy of boundary blocking operation can be used by constructing a pyramid-like hierarchy system.

3. Fuzzy Classification

Consider a problem classifying M regions in K classes. The data set of region m, $\mathbf{X}_m = \{\mathbf{x}_j, j \in J_m\}$, where J_m is the index set of pixels in region m and $\mathbf{x}_j$ is the data vector of pixel j, is associated with an unobserved image class k, which is to be estimated. This association between $\mathbf{X}_m$ and class k can be specified completely with an unobserved indicator vectors, $\mathbf{s}_m = \{s_{km}, k = 1, \cdots, K\}$. In ideal situation, the kth element of $\mathbf{s}_m$ has unit value and all the other elements are zero if region m belongs to class k. The mixture probability distribution of the complete data set $\mathbf{Z} = \{\mathbf{X}_m, \mathbf{s}_m\}$ is then expressed as

$$F(\mathbf{Z} \mid \mathbf{W}, \Theta) = \prod_m \prod_k w_k^{s_{km}} f_k^{s_{km}}(\mathbf{X}_m \mid \boldsymbol{\theta}_k)$$

where $\mathbf{W} = \{w_k\}$ represents the weights of the components $\{f_k\}$ in the mixture distribution, $\sum_k w_k = 1$, and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}$ is the set of parameters that define the classes. The fuzzy procedure calculates the indicator variables $\{s_{km}\}$ as fuzzy vectors in the E-step, and the likelihood of $\mathbf{W}$ and $\boldsymbol{\Theta}$ is maximized in the M-step using $\{s_{km}\}$ estimated in the E-step (Ling *et al.*, 1992). For the assumption of additive Gaussian image model, EM iterative approach to compute the fuzzy vector is summarized in Fig. 1.

Next, using the training samples and reducing one by one the number of classes, the fuzzy training is successively performed for cluster validation, that is, to determine the best number of classes. In this study, the best number of classes, K^* is selected at the level with the maximum ratio of log-likelihood differences in successive steps, that is,

$$K^* = \underset{K_{\min} \le K \le K_{\max}}{\arg\max} \left\{ \frac{LL(K-1) - LL(k)}{LL(K) - LL(K+1)} \right\}$$

where $LL(K)$ is the log-likelihood of K classes. Fig. 2 outlines the Fuzzy classification including cluster validation.

Finally, the unlabelled regions are assigned into one of the K^* classes by MLC using the class parameters estimated in the fuzzy training.

4. Conclusions

In most of remote sensing applications, there exists geophysical connectedness in the observed scene, the proposed method is based on the segmentation using a spatial region-growing clustering which is advantageous relative to the pixel-by-pixel one for the analysis of the patterns with spatial contiguity.

The fuzzy vector is supposed to have unit value for the class that the sample belongs to and zero for the others. The proposed fuzzy approach classifies the regions based on the fuzzy vectors estimated by EM algorithm.

References

Anderberg, M.R, 1973. *Cluster Analysis for Application*, Academic Press, New York.

Lee, S-H., 2004. Unsupervised image classification using region-growing segmentation based on CN-chain, 20(3): 235-248 (in Korean).

Liang, Z, R. J. Jaszczak, and R. E. Coleman, 1992. Parameter estimation of finite mixture using the EM algorithm and information criteria with application to medical image processing, *IEEE Trans. Nucl. Sci.*, 39: 1126-1133.

• E-step - Calculateing Indicator Vectors

$$s_{km}^{(h)} = \frac{w_k^{(h)} f_k(\mathbf{X}_m \mid \boldsymbol{\theta}_k^{(h)})}{\sum_k w_k^{(h)} f_k(\mathbf{X}_m \mid \boldsymbol{\theta}_k^{(h)})}$$

$$f_k(\mathbf{X}_m \mid \boldsymbol{\theta}_k^{(h)}) \propto \left|\Sigma_k^{(h)}\right|^{-\frac{N_m}{2}} \exp\left\{-\frac{1}{2} \sum_{j \in J_m} (\mathbf{x}_j - \boldsymbol{\mu}_k^{(h)})' \Sigma_k^{(h)-1} (\mathbf{x}_j - \boldsymbol{\mu}_k^{(h)})\right\}$$

Conditional Probabilities of Data Set X_m belonging to Class k at hth Iteration

$$\sum_k s_{km}^{(h)} = 1$$

• M-step - Computing Maximum Likelihood Estimates of W, Θ

$$w_k^{(h+1)} = \frac{1}{N} \sum_m N_m s_{km}^{(h)}$$

$$\boldsymbol{\mu}_k^{(h+1)} = \frac{1}{N w_k^{(h+1)}} \sum_m s_{km}^{(h)} \sum_{j \in J_m} \mathbf{x}_j$$

$$\Sigma_k^{(h+1)} = \frac{1}{N w_k^{(h+1)}} \sum_m s_{km}^{(h)} \sum_{j \in J_m} (\mathbf{x}_j - \boldsymbol{\mu}_k^{(h+1)})' (\mathbf{x}_j - \boldsymbol{\mu}_k^{(h+1)})$$

$$N = \sum_m N_m$$

Fig. 1. EM steps to generate Fuzzy Vectors.

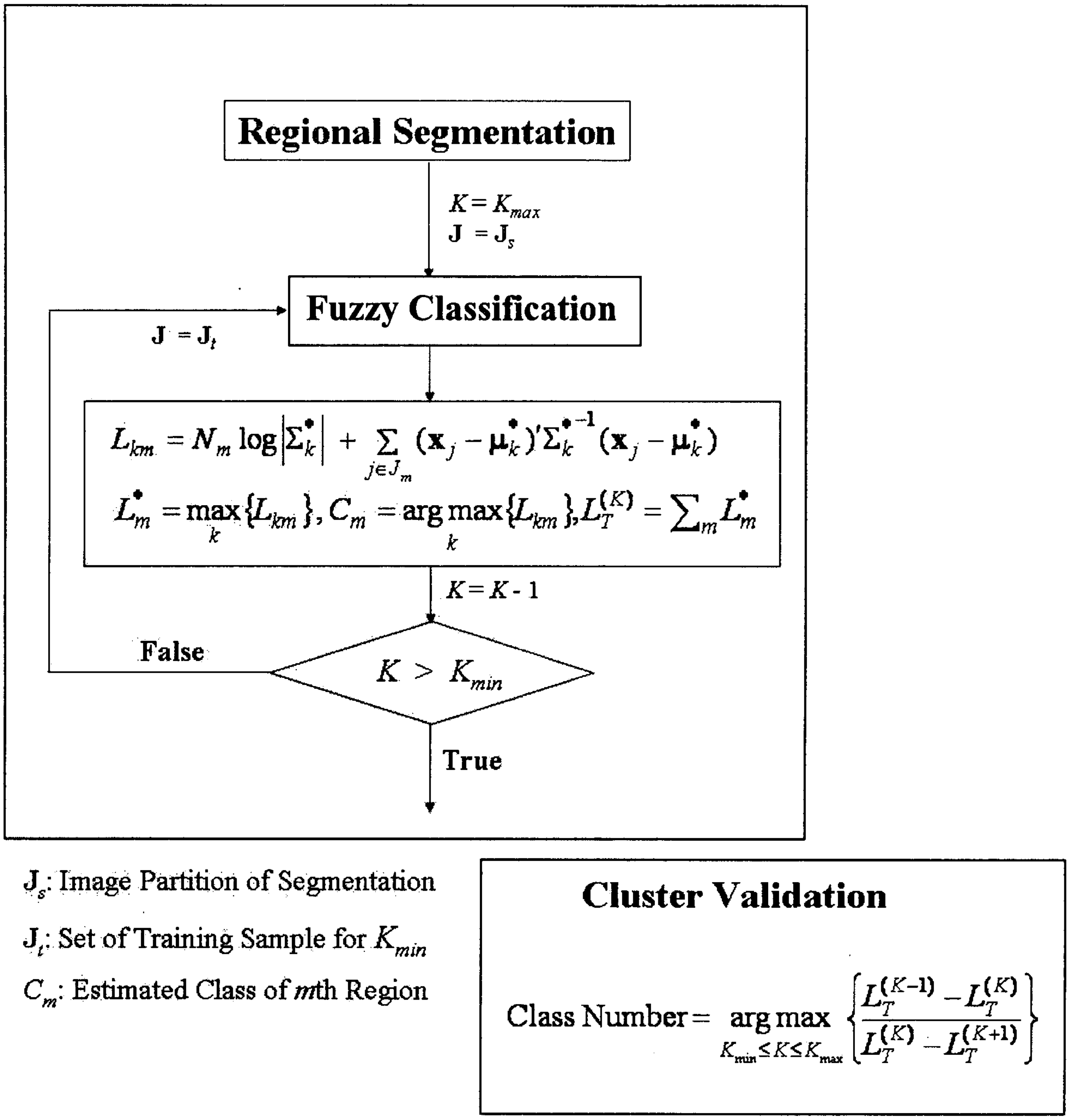

Fig. 2. Cluster Validation