SEMI-SUPERVISED CLASSIFICATION BASED ON GAUSS MIXTURE MODEL FOR REMOTE IMAGERY

XIONG Biao [a, *], ZHANG Xiaojun [a,] JIANG Wanshou [a]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

Commission VI, WG VI/4, WG VII/5

KEY WORDS: Virtual Globe; Remote Sensing Image; Thematic Information; Semi-Supervised Classification; Gauss Mixture Model; EM algorithms

ABSTRACT:

Virtual globes support users with remote images from multiple sources, and support data analysis, information extraction and even knowledge discovery. But when extracting thematic information, those remote images are so complex that we should provide a large amount of label data, which is much expensive and difficult for manual collection, to get sufficient classification result. Semi-Supervised Classification, which utilizes few labeled data assigned with unlabeled data to determine classification borders, has great advantages in extracting classification information from mass data. We find Gauss Mixture can excellently fit the remote sensing image's spectral feature space, propose a novel thought in which each class's feature space is described by one Gauss Mixture Model, and then apply the thought in Semi-Supervised Classification. A large number of experiences shows by using a small amount of label samples, the method proposed in this paper can achieve as good classification accuracy as other supervised classification methods(such as Support Vector Machine Classification, Object Oriented Classification), which need large amount of label samples, and so has a strong application value.

1. Instructions

Today the wide popularity of virtual globe software such as Google Earth, Microsoft Virtual Earth can provide normal users with rapidly increasingly num of remote imagery with high solution and from multiple sources. However, the development of data processing method fails to go up with the pace of image acquirement technology, how to extract information from the image data quickly has become a problem. Supervised classification used in extracting classification information commonly need sufficient and appropriate sample point, not only requires staff have extensive experience in sampling, but also time-consuming and labor-intensive. So it is difficult to meet the request of rapidly image processing.

Semi-Supervised Learning uses a small amount of label data and unlabeled data to get accuracy boundary of classification [1]. Semi-supervised classification recognizes that the label data is limited and continues to learn with the new emerging data, which is much according with the brain's learning mechanisms. Therefore, in the past ten years, the Semi-Supervised Learning is rapid developed, and has been quickly applied in the network label, image indexing, voice recognition and other aspects [1] [2], some scholars have introduced it into the remote sensing image classification [2][3].

The study of Gaussian Mixture Model began in 1894, initially it was used for voice signal processing, image segmentation, video background modeling, moving object detection, etc. but few used for image classification [4]. As the Gaussian mixture model theory is much mature, and could also well fit spectral feature space of remote sensing images, Gaussian Mixture Model in the field of remote sensing image classification should also have good application.

Normally, people use a Gaussian mixture model to describe the entire data set, one certain Gaussian component corresponding to one class. In the remote sensing image, a few types of features' histograms have more than one peak, and some even have no significant peak. It is difficult to describe these features by Gaussian Probability density function, but Gaussian mixture model fit better. In this paper, each category were described by one Gaussian mixture model and classified with Bayesian classification rules, and the results of last classification are used as the next training sample, iterative processing. A large amount of experience shows that the method proposed in the paper only needs much less labels than other supervised classification methods but could achieve as good classification accuracy as them.

2. Gaussian mixture model and Maximum Likelihood Classification

Bayesian classifier is still the most widely used classification algorithm, but generally it assumes that the training data obey Gaussian distribution, which brings a lot of restrictions on the practical application. In this paper, each category of the remote sensing data is described by one Gaussian mixture model, experiments show that the probability function of each category can be fully expressed by Gaussian mixture model with only 3-5 components, which make Gaussian mixture model has strong practical value.

* **Corresponding author:** XIONG Biao, PhD candidate, majors in remote sensing image process, and intelligent information extraction. **Email:** xiongbiaostar@163.com

2.1. Gaussian mixture model and maximum likelihood estimation

Set $X = [X_1, \cdots, X_d]^T$ is a d-dimensional random variable, $x = [x_1, \cdots, x_d]^T$ is an example of that. If its' probability density function can be expressed as the weighted average of k-component distribution:

$$P(x \mid \theta) = \sum_{m=1}^{k} \alpha_m p(x \mid \theta_m) \tag{1}$$

Then we can accept that X subject to the limited mixing distribution, and the corresponding model is a finite mixture model. In the equation, $\alpha_1, \ldots, \alpha_k$ is the distribution probability of the various elements; θ_m is the first m-component distribution parameters; $\{\alpha_1, \cdots, \alpha_k; \theta_1, \cdots, \theta_k\}$ is the collection of all the parameters; at the same time, α_m must meet the following conditions: $\alpha_m \geq 0$, m = 1 , , and

$$\sum_{m=1}^{k} \alpha_m = 1 \tag{2}$$

If we assume that the distributions of the components are all Gaussian distribution, then the corresponding model is Gaussian mixture model. The d-dimensional Gaussian mixture model parameters θ in fact is determined by two parameters: mean vector μ and covariance matrix Σ [6].

With the constraints of Equation (2), Equation (1)'s parameter analytical solution is very complex, so generally we use the iterative method [7]. That is to first establish a sample maximum likelihood equation, and then use EM algorithm to estimate the parameters and mixing parameters of each class.

The basic assumption of Maximum Likelihood Estimation is that all N samples set X= {x(1) , …,x(N) } is independent, then its likelihood function can be defined as follows:

$$l(\theta) = p(X \mid \theta) = \prod_{n=1}^{N} p(x_n \mid \theta) \tag{3}$$

Further definition of the log-likelihood function is as follows:

$$H(\theta) = \ln[l(\theta)] = \sum_{n=1}^{N} \ln p(x_n \mid \theta) \tag{4}$$

Substitute equation (1) into equation (4), then:

$$H(\theta) = \sum_{n=1}^{N} \ln \sum_{m=1}^{k} \alpha_m p(x \mid \theta_m) \tag{5}$$

The maximum likelihood estimation is to find ways of making the largest estimate of the value of θ in Equation (4),

$$H(\hat{\theta}) = \max \sum_{n=1}^{N} \ln \sum_{m=1}^{k} \alpha_m p(x \mid \theta_m), \tag{6}$$

2.2. EM algorithm

EM parameter estimation algorithm is proposed by Dempster, etc. [6], which is divided into E (Expectation) step and the M (Maximization) step. E-step calculate the expectations of the likelihood function --- Q function, M-step make selection of the largest parameters, and then choose the parameters substituted into the E-step, computing expectations, and so forth. EM algorithm will eventually converge to the optimal solution in maximum likelihood sense. Its' advantage is that it does not need to know analytic solution, and the speed of calculation is much fast. For the Gaussian mixture model using EM algorithm parameter estimation process can be described as follows:

E-step: First initialize parameters: μ_m , Σ_m and α_m ,then calculate the posterior probability of samples n belongs to the class m:

$$Q_{mn} = \alpha_m \, p(x \mid \theta_m) \tag{7}$$

Normalized as follows:

$$R_{mn} = \frac{Q_{mn}}{Q_n} = \frac{\alpha_m \, p(x \mid \theta_m)}{\sum_{m=1}^{k} \alpha_m \, p(x \mid \theta_m)} \tag{8}$$

In which

$$p(x \mid \theta_m) =$$

$$2\pi^{-d/2} \mid \Sigma_m \mid^{-1/2} \exp\{-\frac{1}{2}(x - \mu_m)^T \Sigma_m^{-1} (x - \mu_m)\}$$

Obey Gaussian distribution.

M Step: Maximize Equation (8), get new parameters $\tilde{\alpha}_m, \tilde{\mu}_m, \tilde{\Sigma}_m$. The specific Equation is as follows:

$$\tilde{\alpha}_m = \frac{\sum_{n=1}^{N} R_{mn}}{N} \tag{9}$$

$$\tilde{\mu}_m = \frac{\sum_{n=1}^{N} R_{mn} x_n}{N \tilde{\alpha}_m} \tag{10}$$

$$\tilde{\Sigma}_m = \frac{\sum_{n=1}^{N} R_{mn} (x_n - \tilde{\mu}_m)(x_n - \tilde{\mu}_m)^T}{N \tilde{\alpha}_m} \tag{11}$$

With the Equation (8) (9) (10) (11), the result could convergence after several iterations, then the posterior probability of samples n belongs to the class m can be the acquired.

2.3. Maximum Likelihood Classification based on GMM

Suppose each category in remote sensing image data can be represented as a Gaussian mixture model, the probability model of class l can be written as follows:

$$P(x \mid \theta_l) = \sum_{m=1}^{k_l} \alpha_{lm} p(x \mid \theta_{lm}) \tag{12}$$

In which $\theta_l = \{\alpha_{l1}, \cdots, \alpha_{lk_l}; \theta_{l1}, \cdots, \theta_{lk_l}\}$ is the parameter set, k_l , determined by the spectral distribution of the characteristics of selected features, is the best number of Gaussian components, the probability distribution $p(x \mid \theta_{lm})$ follows Gaussian distribution.

If L class data sets are known to each category of the probability distribution function, we can apply Bayesian classifier to estimate probability of a data point x belongs to each category, and then the data points are divided into the greatest probability class. According to probability theory Beyes Equation, the posterior probability of unknown data points x belong to the class as follows:

$$P(w_l \mid x) = \frac{P(x \mid w_l)P(w_l)}{P(x)} \tag{13}$$

In which, $P(w_l)$ is the priori probability, witch is the probability of class w_l appears in the image; $P(x \mid w_l)$ is the likelihood probability, which indicated that probability of class w_l contains point x which can be calculated by Equation (12).

Since $P(x) = \sum_{l=1}^{L} P(x \mid w_l)P(w_l)$ has nothing to do with the class w_l, which is a common factor for different types, and does not work when comparing size, can be removed when we determine categories, then the largest Maximum Likelihood rule becomes:

$x \in w_l$, If and only if

$$P(x \mid w_l)P(w_l) \geq P(x \mid w_j)P(w_j) \tag{14}$$

All l and j are from 1, 2, 3 ,... , L possible classes

3. Semi-Supervised classification based on GMM

It is generally believed that the study of semi-supervised learning began from B. Shahshahani and D. Landgrebe [5]. Semi-supervised learning accepts that the label data is relatively small, and could not fully represent classification space, during the classification process assigned with the label data and unlabeled data, we can establish a reasonable link between unlabeled data's distribution and learning objectives, and then improve the performance of classifier [8]. The existing semi-supervised learning algorithm can be divided into three categories [6]: (1) Generative model-based classifier; (2) algorithm based on graph regularization framework for semi-supervised learning; (3) cooperative training (co-training) algorithm.

Generative model for the general semi-supervised classification assumes that the probability distribution of entire data set obey a Gaussian mixture model, each category can be represented by a Gaussian function [3][6]. However, according above discussion, we can know that the spectral features distribution in remote sensing image is complex, features of each type is difficult to expressed by Gaussian PDF, while it is found that GMM with four or so components can well describe the spectral features space of each type. So we utilize this model in the Generate model-based classification. Then we can use a large number of unlabeled data to estimate accurate model parameters to improve classifier' generalization ability.

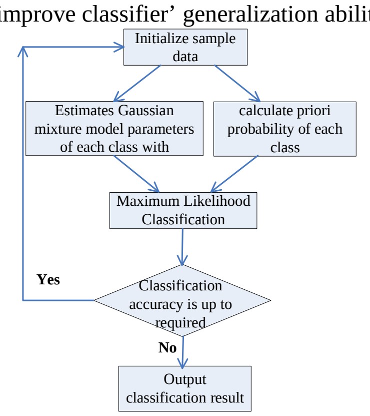

Figure 1. Process of Semi-Supervised Classification based on Gauss Mixture Model

Algorithm flow shown in Figure 1,we first use the label data to estimate model parameters, and set the ratio between number of label data of a given class and the total number of label data as the priori probability. Then we use Equation (13) to calculate the posterior probability and classify the entire data set classification with Bayes rule, then use classification results as label data for the next train, so loops until the classification accuracy meet requirements, or the number of iterations exceeds a certain given value. A large number of experiment shows that the method proposed in this paper only need a small amount of label data can achieve required precision.

4. Experiments

We intercept an image with high spatial resolution from Google Earth for experiments to verify the validity of the algorithm. As shown in Figure 2, the region is rich of features, including small trees, water, roads and housing. As roads and housing are very close in color features and this paper only use three-band images (R, G, B) color information for classification, it is difficult to separate them, and we divide them into the same class, impermeable layer.

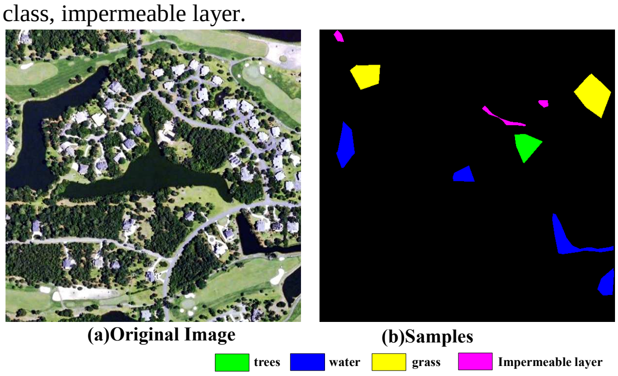

Figure 2. Experiment data and label samples

Figure 3 shows the validity of Gaussian mixture model in describing the spectral features of remote image. We can find, from Histogram curves, that the natural features in the high spatial resolution image is complex, only the characteristics of trees show Gaussian distribution, while most of other parts have multiple peaks, such as the water, grassland, however, the feature of impermeable layer distributed as a long strip in addition to a apparent peak. Such a complex feature space could be difficult to be described effectively with Gaussian

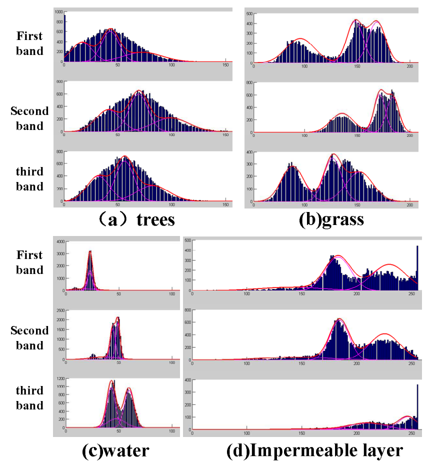

Figure 3. Histograms of Samples and Curves of GMMs

function which only has one peak. The relationship between component number of Gaussian mixture model and its fitting error is shown in Figure 4, in which fitting error $Err = -H(\hat{\theta})$, $H(\hat{\theta})$ is maximum likelihood value estimated according to equation (6). In order to facilitatedisplay and comparison, the figure uses the relative fitting error, which is the ratio between the fitting error and the maximum fitting errors. From Figure 4 we can see that the probability density function fitting the spectral characteristics of thedistribution of four kinds of surface features ,when compared with the Gaussian function, hybrid model in three components when the

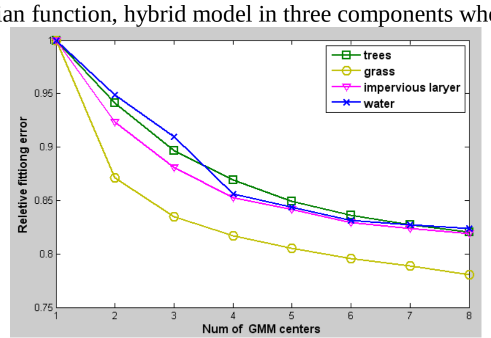

Figure 3. Relationship between Fitting error and centers of GMM

fitting error can be reduced by about 10%, while in the five components can be reduced by 15% when, if more up to increase the number of components has been difficult to significantly improve the fitting precision, while the increase the number of components would increase the calculating cost. Therefore, normal applications only need Gaussian mixture model with 3-5 components, and the following experiment use three components.

Semi-supervised classification uses a small amount of label data together with the unlabeled data to get refined classification surfaces according to the whole data set. As shown in Table 1, after 5 iterations this classification method get a relatively stable result, the classification accuracy has been significantly enhanced. Figure 5 shows the distribution of various types of surface features for the feature space through several iterations, water surrounded by trees, iterative, after continuing to move the center of the distribution of trees, while at the beginning the grass only a small corner, after iterative slow expansion of the annexation of trees, especially a lot of impermeable layer location. Figure 6 is the result of the classification after several iterations. What we should note is that although our method can take advantage unlabeled data to improve the classifier's generalization ability, when collecting label data, we still need to capture the typical surface features, and otherwise the classification results will not be significantly optimized.

Iterative	1	2	3	4	5	6	7	8
Kappa coefficient	0.785	0.837	0.845	0.849	0.850	0.851	0.850	0.851

Table.1 Classification Accuracy

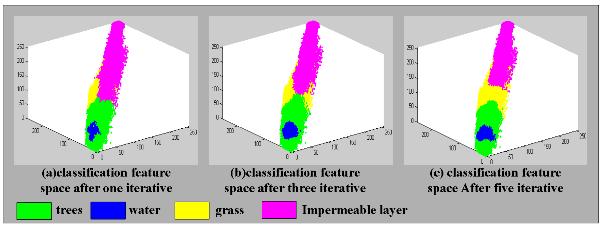

Figure 5. Feature Spaces of Classification after Several Iterative

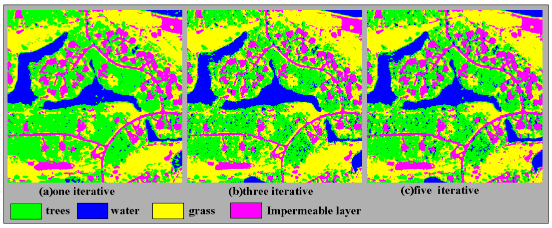

Figure 6. Classification Results after Several Iterative

In Figure 7, our method is compared with other classical methods. The sample data used in Maximum likelihood method, Object-Oriented classification and Support Vector Machines (SVM) is shown in Figure 2.b, which is more than twice of the number of samples used in our method. From Figure 5 we can see, water are completely surrounded by trees in the feature space, general methods are hard to get good classification results, area 1 in Figure 7 shows the maximum likelihood method and support vector machines divided a lot of water regions into trees, while the object-oriented and our method can get excellent results. Object-oriented method is good at extracting surface features, but often break strip features divided them into other objects, such as roads in No. 2 region in Figure 7, while other methods, including our method can get better results. As shown in table 2, the overall classification accuracy kappa coefficient shows that our method is slightly higher than the maximum likelihood method and Support Vector Machine method, less than the object-oriented method, but visual result of our method is better than other classification methods.

Method of classification	Maximum likelihood	Objective -oriented	SVM	Our method
Num of samples	89142	89142	89142	31019
Kappa coefficient	0.795	0.893	0.847	0.850

Table 2. Classification Accuracy in Different Methods

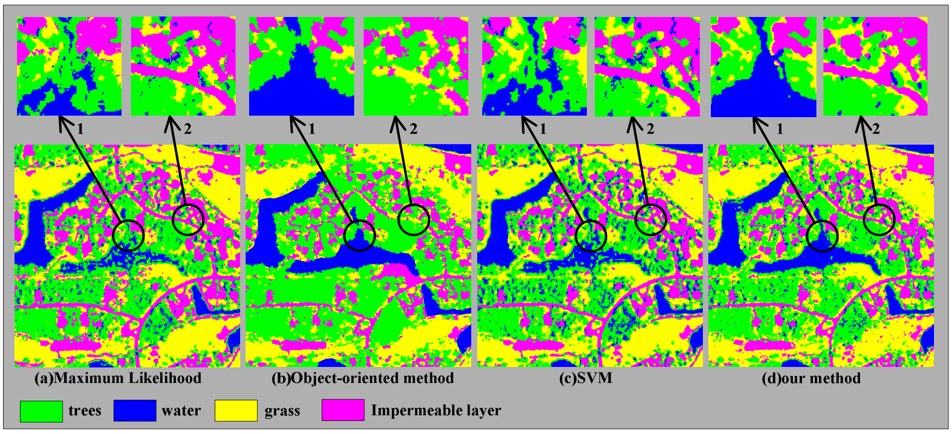

Figure 7. Classification Results in Different Methods

5. Conclusions

Our experiments prove that Gaussian mixture model can excellently describe the remote sensing image feature space, so has great advantages in the remote sensing image classification. At the same time, we only need a small amount of label data and employ unlabeled samples to adjust classifier according the entire image feature space. Experiments show that the method proposed in this paper with a small amount of label data is able to achieve the accuracy of the supervised classification method using a lot of label data, setup basis for rapid processing of remote sensing data, so it has great application value. In the next step we should study the generalization ability of semi-supervised classification applied in a series of images with a sample database, and further reduce the workload of manual collection of label data. We can also study the ability of Gaussian mixture model in describing image local features and texture features to further improve the accuracy of image classification.

6. References

[1] X. Zhu. Semi-supervised learning literature survey [C]. Technical Report 1530, Department of Computer Sciences, University of Wisconsin at Madison, Madison, WI, Apr. 2006.

[2] Chi M, Bruzzone L, Marconcini M, A Novel Transductive SVM for Semi supervised Classification of Remote-Sensing Images [J], IEEE Transactions on Geoscience and Remote Sensing, 2006, 44 (11pt2) : 3363-3373

[3] Luis Gómez-Chova, Gustavo Camps-Valls, Semi supervised Image Classification With Laplacian Support Vector Machines [J], IEEE Transactions on Geoscience and Remote Sensing, 2008, P336-340

[4] Vatsavai, R.R., Shekhar, S., Burk, T.E., A Semi-supervised Learning Method for Remote Sensing Data Mining [C], Tools with Artificial Intelligence, 2005. ICTAI 05.

[5] B. Shahshahani, D. Landgrebe. The effect of unlabeled samples in reducing the small sample size problem and mitigating the hughes phenomenon [J]. IEEE Transactions on Geoscience and Remote Sensing, 1994, 32(5): 1087-1095.

[6] Dempster A, Laird N, Rubin D., Maximum Likehood Estimation from Incomplete Data Via the EM Algorithm [J], J RoyalStatistical Soc B, 1977, 39 :1238

[7] Richards, J. A. and X. Jia, Remote Sensing Digital Image Analysis: An Introduction [M], 1999, Berlin: Springer, 363 p.

[8] McIver D.K. and M. A. Friedl, Using Prior Probabilities in Decision-tree Classification of Remotely Sensed Data [J], Remote Sensing of Environment, 2002, 81:253-256

[9] Wu, W. and G. Shao, 2002, Optimal Combinations of Data, Classifiers, and Sampling Methods for Accurate Characterizations of Deforestation [J], Canadian Journal of Remote Sensing, 28(4):60-609