Research on the Algorithm of Image Classification Based on Gaussian Mixture Model

Z. Meng

Institute of Information Engineering
China University of Geosciences
Beijing, China, 100083

G.Q. Yao

Institute of Information Engineering
China University of Geosciences
Beijing, China, 100083

*Abstract-*Remote sensing image is according to a certain proportion of objective record and reflect electromagnetic radiation of earth surface intensity information, is a form of expression we obtain the remote sensing information. With the development of remote sensing technology, remote sensing image classification is more and more important, and in recent years also appeared some new methods for the classification of remote sensing image. This paper introduces and realizes an unsupervised learning method, an algorithm of image classification based on Gaussian mixture model. When solving this model, find that the function is too complex to direct derivation, so we introduce the EM algorithm to solve it. This paper gives the detailed process of EM algorithm, explains the geometric significance of EM algorithm, and proves its convergence, and realizes the EM algorithm for Gaussian mixture model. The image of this experiment is nearby Beijing airport Results show the algorithms 'applicability and effectiveness.

Keywords-GMM; EM; image classification

I. INTRODUCTION

In recent years, remote sensing technology has become a kind of powerful technology for the current human study earth's resources and environment [1]. In the study of remote sensing technology, the classification is an important aspect of remote sensing image, is also the focus and hotspot in the research of the scholars. Remote sensing image classification can be more intuitive and vividly to reflect the remote sensing image information. The remote sensing image classification method is more and more, each method has its own characteristics.

Supervised classification need a training sample in advance, and the training sample should be chosen good, should have certain representativeness, should have enough. Such as maximum likelihood method, parallel polyhedron classification method, minimum distance classification, decision tree classification method,and so on.Instead, unsupervised classification have no prior knowledge, its commonly used methods in the image classification are ISODATA(Iterative Self Organizing Data Analysis Techniques Algorithm) and K-Means method. In recent years, remote sensing image classification introduced some new methods. Such as artificial neural network, SVM (support vector machine), the method of expert system knowledge, and so on.

II. IMAGE CLASSIFICATION BASED ON GAUSSIAN MIXTURE MODEL

At present more mature remote sensing image classification method based on probability and statistics [2] are mostly supervised classification method, such as maximum likelihood method, etc. But some unsupervised classification methods, such as Gaussian mixture model, is not common in remote sensing image classification method. Gaussian mixture model is the use of the Gaussian probability density function (Gaussian distribution curve) accurately quantify things, will be a thing is decomposed into a number of model based on Gaussian probability density function is formed.

In this article, M denotes the number of models in the Gaussian mixture model (the classification number), j denotes the jth Gaussian model, N denotes the number of samples(pixel), i denotes the ith pixel, dim denotes the number of dimensions, d denotes the d th dimension, the top right corner with $*$ mark denotes the latest values of the parameters.

A. Gaussian Mixture Model

The Normal distribution is given by

$$X \sim N(\mu, \sigma^2) \tag{1}$$

The parameter μ in this definition is the mean or expectation of the distribution (and also it's median). The parameter σ is its standard deviation; its variance is therefore σ^2.The Probability Density Function (PDF) for Equation (1) is given by

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2}$$

And its high dimensional PDF [3] is given by

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{(2\pi)^d \sigma^2}} \exp\left(-\frac{1}{2}(x-\mu)^T (\sigma^2)^{-1}(x-\mu)\right) \tag{3}$$

Gaussian Mixture Model (GMM) is the extension of SGM. Because of the GMM can be divided into several categories and also can better reflects the actual circumstances of the actual object, it used more widely in recent years, especially used in the directions of speech and image, also got good results. Its mathematical function model [4] is given by

$$p(x_i|\alpha, \mu, \sigma) = \sum_{j=1}^{M} \alpha_j f_j(x_i; \mu_j, \sigma_j) \tag{4}$$

Where $0 \leq \alpha_j \leq 1$, $\sum_{j=1}^{M} \alpha_j = 1$, and α_j denotes the coefficient of the jth Gaussian model.

When solving Equation(4) of the three parameters, introducing the maximum likelihood function, which is a very commonly used parameter estimation, by maximizing the likelihood function to get the parameters. Make $\theta_j = (\alpha_j, \mu_j, \sigma_j)$, then Equation (4) can be expressed as $p(x_i|\theta)$. The likelihood function of Equation (4) is given by

$$P(X|\theta) = \prod_{i=1}^{N} p(x_i|\theta) \tag{5}$$

Whose value maybe vary small, so the log likelihood for θ is then given by

$$L(X|\theta) = \sum_{i=1}^{N} \ln\left(\sum_{j=1}^{M} \alpha_j f_j\left(x_i;\ \mu_j, \sigma_j\right)\right) \tag{6}$$

It is difficult to maximize Equation (6) by getting its derivative. Therefore we solve it by EM algorithm [5].

B. EM Algorithm

EM algorithm is a kind of iterative algorithm. In each iteration is divided into two steps, Expectation step (E-step) and Maximization step(M-step).Denote by y the x belongs to the Gaussian distribution, then the observation set is $\{(x_1, y_1), (x_2, y_2), (x_3, y_3) \cdots\}$, instead of $\{x_1, x_2, x_3 \cdots\}$. But the reality is that we don't know y. According to Equation (6)

$$\partial \log L(\theta) / \partial \theta = 0 \tag{7}$$

The EM algorithm assumes that Y is on the basis of the previous estimates of the parameters, to maximizing Equation (6).

The Q function is defined as the conditional expectation of the complete-data log likelihood given X and the current estimates θ, which is given by

$$Q(\theta^*, \theta) = E_Y[L(X, Y|\theta^*)| X, \theta] \tag{8}$$

then

$$Q(\theta^*, \theta) = \sum_{j=1}^{M} \sum_{i=1}^{N} \log(\alpha_j)\, p(j|x_i, \theta)$$
$$+ \sum_{j=1}^{M} \sum_{i=1}^{N} \log\left(f_j\left(x_i;\ \mu_j, \sigma_j\right)\right) p(j|x_i, \theta) \tag{9}$$

make $\beta(i, j) = p(j|x_i, \theta)$ then

$$\beta(i, j) = \frac{\alpha_j f_j(x_i; \mu_j, \sigma_j)}{p(x_i|\alpha, \mu, \sigma)} \tag{10}$$

To Equation (9) using the Lagrange multiplier method, we get

$$\alpha_j^* = \frac{1}{N} \sum_{i=1}^{N} \beta(i, j) \tag{11}$$

$$u_j^* = \frac{\sum_{i=1}^{N} \beta(i,j) x_i}{\sum_{i=1}^{N} \beta(i,j)} \tag{12}$$

$$\sigma_j^* = \frac{\sum_{i=1}^{N} \beta(i,j)(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{N} \beta(i,j)} \tag{13}$$

In conclusion, in the EM algorithm E-step gets Function $Q(\theta^*, \theta)$, the M-step updates parameters.

We already know that the EM algorithm is gradually approximation of maximizing $\log L(\theta)$ through iterative. Hypothesis θ^* is got by iteration θ, so we hope the $L(\theta)$ value increasing,which is $\log L(\theta^*) > \log L(\theta)$. Make:

$$B(\theta^*, \theta) = \log L(\theta^*) - \log L(\theta) \tag{14}$$

Based on Jensen inequality, we get the lower bound

$$B(\theta^*, \theta) = \sum_Y P(Y|X, \theta) \log\left(\frac{P(X|Y,\theta^*)P(Y|\theta^*)}{P(Y|X,\theta)\log P(X|\theta)}\right) \tag{15}$$

The $L(\theta) \geq B(\theta^*, \theta)$. So $B(\theta^*, \theta)$ is the lower bound of $L(\theta)$, and by Equetion(15)

$$L(\theta) = B(\theta, \theta) \tag{16}$$

So that the θ can make $B(\theta^*, \theta)$ increasing, also can make $L(\theta)$ increasing. The geometric meaning is shown in Figure 1.
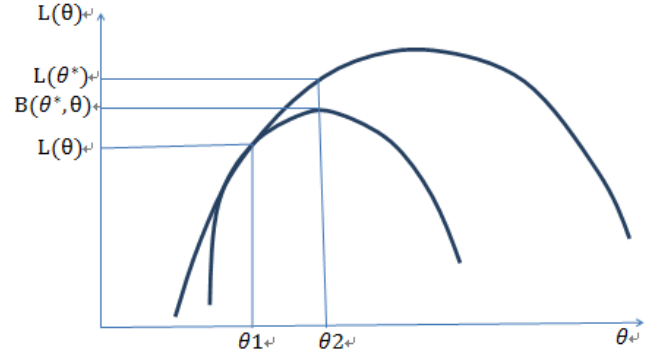

FIGURE I.   THE GEOMETRIC MEANING OF EM ALGORITHM.

Where the horizontal axis is θ, in the longitudinal axis the above curve is $L(\theta)$ and the below curve is $B(\theta^*, \theta)$, $B(\theta^*, \theta)$ is the lower bound of log likelihood function of $L(\theta)$. By Equation (16), the two functions are equal at point $\theta 1$. Find $\theta 2$ make B maximization, at this moment L function also increases. Recalculate the Q function with the new $\theta 2$ as the next iteration. In this process, the log likelihood function L is increasing, so the EM algorithm must be convergence [6].

C. GMM and EM Algorithm for Remote Sensing Image Classification

About remote sensing image classification, using the EM algorithm for GMM is as follows

---

Algorithm 1 EM algorithm for solving GMM
Input: the number of samples N, the number of
     dimensions dim, the number of models, the sample
     matrix data[dim, N]
Output: the vector classification results result
alpha0[1, M] = {1./M};
miu0[dim, M] = rand();
digma0[dim, M] = {1};
while($\sum \theta - \sum \theta 0) > threshold$) {
    // E-step:
    bate[i, j] = $\frac{\alpha_j f_j(x_i; \mu_j, \sigma_j)}{p(x_i|\alpha, \mu, \sigma)}$
    // M-step:
    alpha[j] = $\frac{1}{N} \sum_{i=1}^{N} \beta(i, j)$
    miu[:, j] = $\frac{\sum_{i=1}^{N} \beta(i,j) x_i}{\sum_{i=1}^{N} \beta(i,j)}$
    sigma[d, j] = $\frac{\sum_{i=1}^{N} \beta(i,j)(x_i - \mu_i)(x_i - \mu_i)^T}{\sum_{i=1}^{N} \beta(i,j)}$
    alpha0 = alpha
    miu0 = miu

```
    sigma0 = sigma
}
result[i] =  max   α_j f_j(x_i; μ_j, σ_j)
           1≤j≤M
```

Firstly initialization parameter $\theta_j = (\alpha_j, \mu_j, \sigma_j)$. The alpha, which is α, is $1 * M$ vector, initialized to $1/M$; miu, which is μ, is dim $* M$ matrix, initialized to the random number; sigma, which is σ, is dim $* M$ matrix, initialized to unit matrix. Then do the EM algorithm iteration until its convergence.

Judging convergence conditions generally have two methods:

1) the loglikelihood function convergence, that is $|L(X|\theta^*) - L(X|\theta)| < \varepsilon$

2) the parameters convergence, that is $\sum_\theta |\theta^* - \theta| < \varepsilon$

Through experiment, the effect of these two methods are close, but obviously method 2 has smaller computational cost.

III. EXPERIMENTAL RESULT

The image of this experiment is nearby Beijing airport, taken by NASA's Landsat Landsat-8 on July 31, 2013. The classification results is shown in Figure 2 to Figure 5.
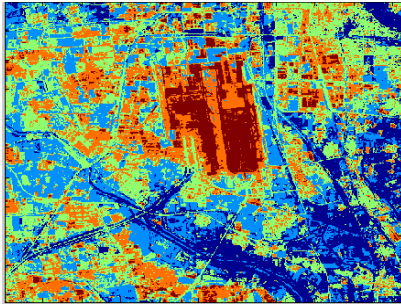

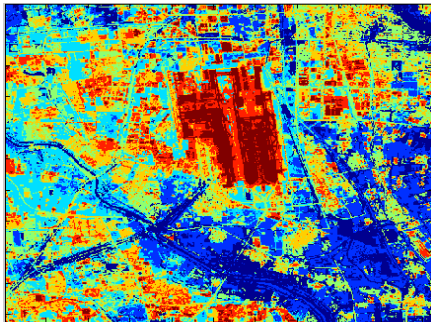FIGURE II. 5 CATEGORIES.

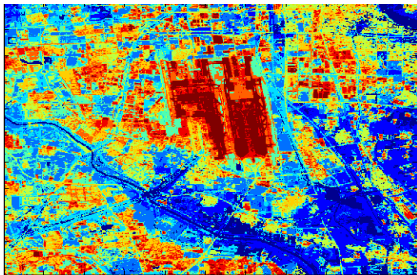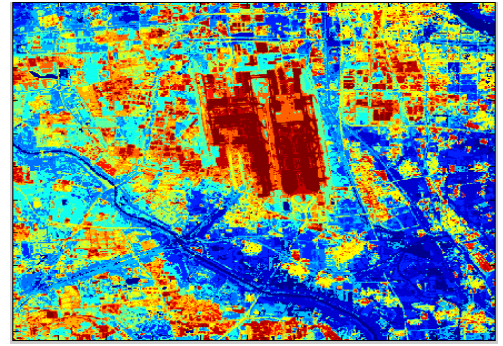FIGURE III. 7 CATEGORIES.

FIGURE IV. 10 CATEGORIES.

FIGURE V. 14 CATEGORIES.

The clustering validity evaluation methods have internal evaluation method, external evaluation method and relative evaluation method [7]. This paper mainly from two aspects of the internal evaluation method, distance of internal clustering and distance between clustering, evaluate the clustering results.

Distance of internal clustering is given by

$$D = \frac{\sum_{i=1}^{M}\left(\frac{\sum_{p \in C_i}|p - m_i|}{|C_i|}\right)}{M} \qquad (17)$$

Where D is the distance of internal clustering, M is the number of categories, C_i is all space objects(sets) in the ith category, $|C_i|$ is the number of space objects in the ith category, p is any space object in the ith category, m_i is the center(average) in the ith category.

Distance between clustering is given by:

$$L = \frac{\sum_{j=1}^{M}|m_j - m|}{M} \qquad (18)$$

Where L is the distance between clustering, M is the number of categories, m_i is the center in the ith category, m is the center of the whole category.

The clustering result evaluation is given in Table 1.

TABLE I.   THE CLUSTERING RESULT EVALUATION.

	Distance of internal clustering	Distance between clustering
5 categories	11.5994	74.9818
7 categories	9.3122	77.9911
10 categories	7.69063	79.7755
14 categories	6.60252	80.8186

In Table 1, it is clearly observed, that two methods(distance of internal clustering and distance between clustering) are both the more categories, the better clustering effect, which is intuitively understandable. If divided into N categories, each object itself is a category, the distance of internal clustering will be 0. Although the more categories, the smaller training error, but the test error will increasing. Therefore in reality it is false that the more categories is the better. The number of categories should judge by reality need.

From the classification results can be seen, compared to the traditional classification methods, the algorithm is a suitable method, reducing the mixing situation.

REFERENCES

[1] Li Shihua,WangJinliang, BiYan, Chen Yao, Zhu Miaoyuan, Yang Shuai, Zhu Jia. Remote Sensing for Land &Resources[J]. Journal of land and resources remote sensing, 2005, (2) : 1-6.

[2] Yang Kai. Principle and method of remote sensing image processing[M]. Beijing: surveying and mapping press, 1988.

[3] Li Hang. Statistical learning methods[M].Beijing: Tsinghua university press, 2012.155-169.

[4] Hedelin P, Skoglund J. Vector quantization based on Gaussian mixture models[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(4) : 385 − 401.

[5] Yang Kai. Remote sensing image processing theories and methods [M]. Beijing: surveying and mapping publishing house, 1988.

[6] Zhu Zhouhua. EM algorithm and its application in the gaussian mixture model[J] . The modern electronic technology, 2003 , 26 (24) : 88 - 90.

[7] Zhang Weijiao, Liu Chunhuang, Li Fangyu. Clustering quality evaluation methods [J]. Computer engineering, 2005, 31 (20) : 10-12.