An Improved Algorithm for Image Classification

Asma Lak¹*, Naser Parhizgar² and Mahbobeh Lak¹

¹Department of Electrical Engineering, College of Engineering, Geneva branch, Islamic Azad University, Geneva, Iran
²Department of Electrical Engineering, College of Engineering, Shiraz branch, Islamic Azad University, Shiraz, Iran

Abstract

In this paper a new method has been proposed to less time (by reducing the repetition rate) and more accuracy in EM algorithm running. In this paper MATLAB software has been used to simulation. The image which has been used to simulation taken by AVIRIS in June 1992 in Indiana with 145*145 pixels. EM algorithm is an iterative algorithm and uses from the results of previous step to re-estimate. This repetition is continued till the stop condition is satisfied. The proposed method uses from the probability of belonging of each pixel to a class and compares it with estimated probability of previous stage. After running proposed algorithm the time of running and the accuracy have been improved. The time has been reduced from 1370 to 1581 seconds because the number of iterations has been reduced from 14 to 12. The accuracy increased from 66.757 to 69.979. This algorithm is useable to classify the images which need to more accuracy.

1. Introduction

Traditional machine learning use limited labeled data to classify the images which are expensive and difficult. The solution is semi-supervised learning. This method is using from labeled and unlabeled data to better accuracy and better classifier to classification and because of it, it is an interested theory¹,².

In this research it is tried to semi-supervised graph-based present a method to improve image classification which tries to reduce the time of classification and increase the accuracy. There are sixteen classes with Gaussian mixture model for each class. When the huge unlabeled data are used, the Expectation-Maximization (EM) algorithm can use. Also there is identified weight matrix to detect outliers and lead to improve the accuracy.

2. Harmonic Mixture and Em Algorithm

There are two important subjects to consider the graph-based semi-supervised learning.

- How the un-supervised points in graph can be considered?
- How the calculation for unlabeled data matrix has been reduced?

It is done by mixing the graph method and mixture model. For instance for semi-supervised learning the Gaussian Mixture Model (GMM)³,⁴,⁵ is used. Learning is done by EM algorithm witch use of unsupervised points and is a parameter model with a few parameters. Mixture models and graph-based semi-supervised methods lead

*Author for correspondence
An Improved Algorithm for Image Classification

An Improved Algorithm for Image Classification

to different hypothesis for relationship between labeled and unlabeled data. From the graph method point of view, sequent model is a backbone graph that is smaller with small calculation. From the mixture model point of view, this method is inductive and uses the new points. For example in a typical mixture model for generative process first of all select a class y, then chooses a mixture component .

\[ m \in \{1, ..., M\} \] with \( p(m | y) \) and finally a point \( x \) according to \( p(x | m) \) is generated.

\[ p(x, y) = \sum_{m=1}^{M} p(m)p(y | m)p(x | m) \quad (1) \]

For all y: \( p(y | m) > 0 \).

\( q_i(m | i) \) is introduced on mixture membership, one for each i . By Jensen’s inequality:

\[
L(\Theta) = \sum_{i \in L} \sum_{m=1}^{M} q_i(m | x_i, y_i) \frac{p(m)p(y | m)p(x | m)}{q_i(m | x_i)} \\
+ \sum_{i \in U} \sum_{m=1}^{M} q_i(m | x_i) \frac{p(m)p(x | m)}{q_i(m | x_i)} \\
\geq \sum_{i \in L} \sum_{m=1}^{M} q_i(m | x_i, y_i) \log \frac{p(m)p(y | m)p(x | m)}{q_i(m | x_i)} \\
+ \sum_{i \in U} \sum_{m=1}^{M} q_i(m | x_i) \log \frac{p(m)p(x | m)}{q_i(m | x_i)} \geq F(q, \Theta) \quad (2)
\]

The EM algorithm works by iterating coordinate-wise ascend on \( q \) and \( \Theta \) to maximize \( F(q, \Theta) \). The E step fixes \( \Theta \) and finds the \( q \) that maximizes \( F(q, \Theta) \). After EM converges, with some summarization, the classification of a new point \( x \) is done by

\[
p(y = i | x) = \frac{\sum_{m=1}^{M} p(y = i | m)p(m | x)}{\sum_{m=1}^{M} p(x | m)p(m)} \]

\[ (3) \]

For more information the reference 5.

### 3. Improvement of Em Algorithm

EM algorithm is an iterative algorithm. In any stage of algorithm for each unlabeled pixel, probability of each class are estimated and with respect to the most estimated probability for that pixel, the label is determined and finally the classification are done. In next step of the algorithm, the results of previous step are used to re-estimate. This repetition is continued till the stop condition is satisfied.

In this paper, in the each step of algorithm, the probability of belonging of each pixel to a class is compared with estimated probability of previous stage. If this value is lower, the value of probability of previous stage is selected and otherwise, the probability of current stage is considered and according to this value of probability the classification is done.

The image which is used taken by AVIRIS in June 1992 in Indiana. This image has 145*145 pixels and radiometric accuracy is 8 bits. Based on cultivated plants and conditions of ground each area has 16 various classes5,6,7.

### 4. Results

In Table 1, the total classification accuracy in the various stages, for proposed EM algorithm and previous8 method has been shown.

As it can be seen the number of iteration of EM has been reduced and also total classification accuracy has been reach to 70% from 62.4% in the final stage of algorithm. These values in Figure 1. have been compared.

![Figure 1. Results of comparison for previous and proposed method.](image)

**Table 1.** Comparison between previous method of image classification

| Preparation | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   |
|-------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Previous method | 62.406 | 63.631 | 64.355 | 64.615 | 64.779 | 64.943 | 65.213 | 65.599 | 65.802 | 66.081 | 66.197 | 66.573 | 66.737 | 66.757 |
| Proposed method  | 62.406 | 65.059 | 66.381 | 67.528 | 68.310 | 68.840 | 69.419 | 69.641 | 69.699 | 69.834 | 69.940 | 69.979 |    |    |
The classified image in the first stage and final stage of proposed algorithm has been shown in Figure 2 and 3.

![First step of classified image](image1)

**Figure 2.** First step of classified image

![Final step of classified image](image2)

**Figure 3.** Final step of classified image

### 5. Conclusion

In this paper we proposed a method to improve the EM algorithm to work better by reducing the time of running and increasing the accuracy. In this method in each step of algorithm, the probability of belonging of each pixel to a class is compared with estimated probability of previous stage. If this value is lower, the value of probability of previous stage is selected and otherwise, the probability of current stage is considered and according to this value of probability the classification is done. By the used method the total time of running of proposed algorithm is 1581 seconds. The iterations have been reduced 2 stages (from 14 to 12). Also the time of running of previous algorithm has been reduced to 1712 seconds. The accuracy increased from 66.757 to 69.979. These results show that our proposed method is very good and useful to image classification.

### 6. References

1. Jimenez L, Landgrebe DA. Hyperspectral data analysis and feature reduction via projection pursuit, IEEE Trans Geosci Remote Sens. 1999 Nov; 37(6):2653–67.
2. Lak A, Parhizgar N, Lak M. Improvement of outliers detection in image classification. Journal of Applied Sciences. 2015. p. 1–5. Available from: 10.3923/jas
3. Nigam K, Ghani R. Analyzing the effectiveness and applicability of co-training. Ninth International Conference on Information and Knowledge Management; McLean USA. 2000 Nov; 6(11):86–93.
4. Castelli V, Cover T. The relative value of labeled and unlabeled samples in pattern recognition with an unknown mixing parameter. IEEE Transactions on Information Theory. 1996; 42(6): 2101–17.
5. Lak M, Keshavarz A, Pourghassem H. Graph-based hyperspectral image classification using outliers detection based on spatial information and estimating of the number of gmm mixtures. IEEE International Conference on Communication Systems and Network Technologies; Washington, DC: USA. 2013 Apr. p. 196–200.
6. Rosset S, Zhu J, Zou H, Hastie T. A method for inferring label sampling mechanisms in semi-supervised learning. In: Saul LK, Weiss Y, Bottou L, editors. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press; 2008. p. 1161–8.
7. Luxburg VU, Bousquet O, Belkin M. Limits of spectral clustering. In: Saul LK, Weiss Y, Bottou L, editors. Advances in neural information processing systems. Cambridge, MA: MIT Press; 2005.