BYY HARMONY ENFORCING REGULARIZATION FOR GAUSSIAN MIXTURE LEARNING

Hongyan Wang, Jinwen Ma

Department of Information Science, School of Mathematical Sciences
Peking University, Beijing, 100871, China

ABSTRACT

In this paper, a Bayesian Ying-Yang (BYY) harmony enforcing regularization (BYY-HER) algorithm is proposed for Gaussian mixture learning with a sample dataset on both parameter estimation and model selection, i.e., selecting an appropriate number of Gaussians in the mixture, through a regularization process from the BYY harmony learning to the maximum likelihood learning. It has been demonstrated by experiments on synthetical and real sample datasets that our proposed BYY-HER algorithm can not only select the correct number of actual Gaussians in a dataset, but also obtain good parameter estimations for the parameters in the true mixture.

Keywords: Gaussian mixture; BYY Harmony learning; Automated model selection; Regularization; Maximum likelihood

1. INTRODUCTION

As a powerful tool for data clustering, Gaussian mixture model has been extensively studied in the literature for either data modeling or clustering analysis on a sample dataset. Although there have been various statistical or unsupervised competitive learning methods to do such a task, e.g. the EM algorithm [1] for Maximum Likelihood (ML), k-means algorithm [2] for the least Mean Square Error (MSE), it is usually assumed that the number of Gaussians, or clusters, in the dataset is pre-known. However, in many instances this key information is not available and then the selection of an appropriate number of Gaussians must be made before or during the estimation of the parameters in the mixture, which is a rather complicated and difficult task [3].

As the number k of Gaussians is just a scale of the Gaussian mixture model, its determination is actually referred to as model selection. Thus, the general Gaussian mixture modeling is actually a compound problem of estimation and model selection. In fact, this compound problem has been investigated by many researchers from different directions. The traditional method was to choose an optimal number of Gaussians via certain selection criterion. Among these criteria, Akaike's information criterion (AIC) [4] is well known. But the validating process is computationally consumptive because we need to repeat the entire parameter learning process at a large number of possible values of k.

Recently, the Bayesian Ying-Yang (BYY) harmony learning system [5-7] has developed a new learning mechanism that makes model selection automatically during parameter learning. In fact, the BYY harmony learning has already implemented on the Guassian mixture modeling for the parameter learning with automated model selection. In order to do so, a bidirectional architecture (BI-architecture) and a backward architecture (B-architecture) of the BYY learning system were constructed for a finite mixture such that the Gaussian mixture modeling can be transformed into a BYY harmony learning problem on them. Actually, some efficient BYY harmony learning algorithms have been al-

Jinwen Ma, the corresponding author, Telephone:86-10-62760609, Email: jwma@math.pku.edu.cn.

tor shifting from 0 to 1, the BYY-HER algorithm turns the BYY harmony learning into the ML learning finally. Since the BYY harmony learning has the ability of automated model selection and the ML estimate is consistent, the BYY-HER algorithm will lead to a good estimate of the parameters with correct model selection as the regularization proceeds properly, which is actually demonstrated by the simulation and practical experiments.

In the sequel, the BYY-HER algorithm will be derived in Section 2. Some typical simulations and practical experiment are conducted in Section 3. Finally, we conclude briefly in Section 4.

2. BYY HARMONY ENFORCING REGULARIZATION ALGORITHM

According to the BYY harmony theory [7,11], we can get the following harmony function:

$$J(\Theta_k) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} p(j|x_i)\ln[\alpha_j q(x_i|m_j, \Sigma_j)], \qquad (1)$$

on the B-architecture with the parameters $\Theta_k = \{\alpha_j, m_j, \Sigma_j, p(j|x)\}$. By certain transformations, $J(\Theta_k)$ can be divided into two parts as follows:

$$J(\Theta_k) = L(\Theta_k) - O_N(p(y|x)),$$

where the first part4490/F81-isstfirstpelihoodst

$L(\Theta k$

cal maximum solution. The experiment results on the two synthetic datasets in the case of $k = 8$ are shown in Figs 3 & 4, respectively. It can be observed that four estimated Gaussians can match the actual ones in each dataset accurately, with the extra Gaussians being canceled automatically. Moreover, it can be further found that the estimations of the parameters are as good as the ML estimators.

3.2. On the Iris data

We further apply the BYY-HER algorithm to the classification of the Iris data (from http://archive.ics.uci.edu/ml/). This dataset consists of 150 samples of three classes which are Versicolor, Iris Virginica and Iris Setosa. Each class contains 50 samples and each sample or datum is 4-dimensional with measures of the plants morphology. Because the BYY-HER algorithm is in an unsupervised learning mode, we ignore the indexes of these samples. As the learning process has been accomplished, each sample is classified according to its maximum posterior probability $p(j|x_t)$.

By setting $k = 6$, $a = 2.0$ and $b = 200$, $\lambda_0 = 1e-200$, we implement the BYY-HER algorithm on the Iris data set with $0.01 \leq \lambda \leq 0.99$, and t being increased by 0.1 at each time. It has been found by the experiments that the BYY-HER algorithm can detect the three actual categories and the average accuracy is 96.7% (five samples in the second class were misclassified).

4. CONCLUSIONS

We have proposed the BYY harmony enforcing regularization (BYY-HER) algorithm on Gaussian mixture for both model selection and parameter estimation. The BYY-HER algorithm implements strongly the BYY harmony learning for automated model selection at the previous learning stage and gradually transforms to the maximum likelihood (ML) learning for good estimation of the parameters at the final learning stage. It is demonstrated by the simulation and practical experiments that the BYY-HER algorithm can detect the number of actual Gaussians in the sample dataset and obtain accurate estimation of the parameters in the Gaussian mixture.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of China for grant 60771061.

5. REFERENCES

[1] R. A. Render, H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.* vol. 26, no.2, pp. 195-293, 1984.

[2] P. A. Devijver, J.kitter, "A Statistical Approach," *Englewood Cliffs, Prentice-Hall, NJ*, 1982.

[3] J. A. Hartigan, "Distribution problems in clustering," *Classification and clustering*, J. Van Ryzin Eds., pp. 45-72, New York: Academic Press, 1977.

[4] H. Akaike. "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control,* vol. AC-19, pp. 716-723, 1974.

[5] L. Xu, "Bayesian Ying-Yang machine:, clusteing and number of cluters," *Pattern recgnition Lett.,* vol.18, pp. 1167-1178, 1997.

[6] L. Xu, "Best harmony, unified RPCL and automated model selection for unsupervised and supervised learning on Gaussian mixtures, three-layer nets and ME-RBF-SVM models," *Int. J. Neural Syst.,* vol.11, no.1, pp. 43-69, 2001.

[7] L. Xu, "BYY harmony learning, structural RPCL, and topological self-organizing on mixture modes," *Neural Networks,* vol.15, pp. 1231-1237, 2002.

[8] J. Ma, T. Wang, L. Xu, "A gradient BYY harmony learning rule on Gaussian mixture with automated model selection," *Neurocomputing,* vol.56, pp. 481-487, 2004.

[9] J. Ma, L. Wang, "BYY harmony learning on finite mixture: adaptive gradient implementation and a floating RPCL mechanism," *Neural Processing Lett,* vol.24, no.1, pp. 19-40, 2006.

[10] J. Ma, X. He, "A fast fixed-point BYY harmony learning algorithm on Gaussian mixture with automated model selection," *Pattern Recognition Letters*, vol.29, no.6, pp. 701-711, 2008.

[11] J. Ma and J. Liu, "The BYY annealing learning algorithm for Gaussian mixture with automated model selection," *Pattern Recognition,* vol.40, pp. 2029-2037, 2007.

[12] Z. Lu and J. Ma, "A Gradient Entropy Regularized Likelihood Learning Algorithm on Gaussian Mixture with Automatic Model Selection," *Lecture Notes in Computer Science*, vol.3971, pp: 464-469, 2006.

[13] N. Ueda and R. Nakano, "Deterministic annealing EM algorithm," *Neural Networks,* vol.11, pp. 271C282, 1998.

[14] L. Xu, A. Krzyak, E. Oja, "Rival penalized competitive learning for clustering analysis, RBF net, and curve detection," *IEEE Trans. Neural Networks,* vol.4, no.4, pp. 636C649, 1993.