Local Factor Analysis with Automatic Model Selection: A Comparative Study and Digits Recognition Application

Lei Shi and Lei Xu

Chinese University of Hong Kong, Shatin, NT, Hong Kong
{shil, lxu}@cse.cuhk.edu.hk

Abstract. A further investigation is made on an adaptive local factor analysis algorithm from Bayesian Ying-Yang (BYY) harmony learning, which makes parameter learning with automatic determination of both the component number and the factor number in each component. A comparative study has been conducted on simulated data sets and several real problem data sets. The algorithm has been compared with not only a recent approach called Incremental Mixture of Factor Analysers (IMoFA) but also the conventional two-stage implementation of maximum likelihood (ML) plus model selection, namely, using the EM algorithm for parameter learning on a series candidate models, and selecting one best candidate by AIC, CAIC, and BIC. Experiments have shown that IMoFA and ML-BIC outperform ML-AIC or ML-CAIC while the BYY harmony learning considerably outperforms IMoFA and ML-BIC. Furthermore, this BYY learning algorithm has been applied to the popular MNIST database for digits recognition with a promising performance.

1 Introduction

Clustering and dimension reduction have been considered as two of the fundamental problems in the literature of unsupervised learning. It is well known that Gaussian mixture model (GMM) with full covariance matrices requires sufficient training data to guarantee the reliability of the estimated model parameters, while GMM with diagonal covariance matrices requires a relatively large number of Gaussians to provide high recognition performance. Local factor analysis (LFA) (also called mixture of factor analyzers (MFA)) combines the widely-used GMM model with one well known dimension reduction approach, namely factor analysis (FA). Via local structure analysis, LFA is able to reduce the freedom degree of covariance matrices to achieve a good generalization. Several efforts have been made on such a topic of local dimensionality reduction.

In the literature of LFA research, the conventional method performs the maximum likelihood (ML) learning in help of one of typical statistical criteria to select both component number and local dimensions of local factor analysis. However, it suffers a huge computing cost. Bayesian Ying-Yang (BYY) learning was proposed as a unified statistical learning framework firstly in 1994 and systematically developed in the past decade. Providing a general learning framework,
BYY harmony learning consists of a general BYY system and a fundamental harmony learning principle as a unified guide for developing new regularization techniques, a new class of criteria for model selection, and a new family of algorithms that perform parameter learning with automatic model selection. Readers are referred to [15,17] for a recent systematic review. Applying the BYY harmony learning to local factor analysis, an adaptive learning algorithm has been developed that performs local factor analysis with both the local dimensions of each component and the number of components automatically determined during parameter learning [14,16].

This paper investigates the automatic BYY harmony learning based LFA, in comparison with the ML learning via criteria of AIC, CAIC, BIC, as well as a recently proposed approach called Incremental Mixture of Factor Analyzers (IMoFA) [11] that makes an increasing model selection during learning. A comparative study is conducted via experiments on not only simulated data but also several real problem data sets, as well as a popular digit recognition database, respectively. The rest of this paper is organized as follows. In Section 2, we review FA and LFA, together with typical statistical criteria and the recent proposed algorithm IMoFA. Section 3 will further introduce the BYY harmony learning based LFA. After a series of comparative experiments in Section 4, we apply the BYY-LFA to the popular MNIST database of digit recognition in Section 5. Finally, we conclude in Section 6 and make further discussion in Section 7.

2 FA and LFA

2.1 Factor Analysis Model

Factor analysis (FA) is a classical dimension reduction technique aiming to find the hidden causes and sources [8]. Provided a \(d\)-dimensional vector of observable variables \(\mathbf{x}\), the FA model is given by \(\mathbf{x} = \mathbf{Ay} + \mathbf{e}\), where \(\mathbf{A}\) is a \(d \times k\) loading matrix, \(\mathbf{y}\) is a \(m\)-dimensional unobservable latent vector assumed from Gaussian \(G(\mathbf{y}|\mathbf{0}, \mathbf{I}_k)\) with \(m < d\) generally, \(\mathbf{e}\) is a \(d\)-dimensional random noise vector assumed from Gaussian \(G(\mathbf{e}|\mathbf{0}, \Psi)\) with \(\Psi\) being a diagonal matrix. Moreover, \(\mathbf{y}\) and \(\mathbf{e}\) are mutually independent. Therefore, \(\mathbf{x}\) is distributed with zero mean and covariance \(\mathbf{AA}^T + \Psi\). The goal of FA is to find \(\theta = \{\mathbf{A}, \Psi\}\) that best models the structure of \(\mathbf{x}\). One widely used method to estimate \(\theta\) is the maximum likelihood (ML) learning that maximizes the log-likelihood function, usually implemented by the expectation-maximization (EM) algorithm [18].

2.2 Local Factor Analysis

Local factor analysis (LFA) (or also called mixture of factor analyzers (MFA)), is a useful multivariate analysis tool to explore not only clusters but also local subspaces with wide applications including pattern recognition, bioinformatics, and financial engineering [14,10]. LFA performs clustering analysis and dimension reduction in each cluster (component) simultaneously. Provided \(\mathbf{x}\) as a \(d\)-dimensional random vector of observable variables, the mixture model assumes