Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Neurocomputing 55 (2003) 739–745

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Letters

Topological local principal component analysis☆

Zhi-Yong Liu*, Lei Xu

Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, People's Republic of China

Abstract

In help of the Kohonen's self-organizing maps we present a topological local principal component analysis model which is capable of exploiting both the global topological structure and each local cluster structure. A newly proposed self-organizing strategy that can enhance the learning speed is introduced to train the model.

Keywords: Local PCA; SOM; Self-organizing strategy

1. Introduction

Principal component analysis (PCA) [6] is frequently adopted for dimensionality reduction as the mean square error (MSE) upon reconstruction is minimized. However, when the data is from multi-modes, performing PCA can be far from satisfactory and thus its local extension, local PCA, has been suggested [2,3,7–9]. Local PCA can be implemented in several ways. The approach studied in [4] estimates a gaussian mixture

$$p(x|\theta) = \sum_{i=1}^{k} \alpha_i G(x|\mu_i, \Sigma_i), \tag{1}$$

where $\alpha_i > 0$, $\sum_{i=1}^{k} \alpha_i = 1$, and $G(x|\mu_i, \Sigma_i)$ denotes a gaussian probability density function (PDF) with mean vector μ_i and covariance matrix Σ_i, in the first step and then performs PCA on each cluster or gaussian in the second step. A better way is to perform the two steps coordinately such that the first step and the second step are

☆ The work described in this paper was fully supported by a grant from the Research Grant Council of the Hong Kong SAR (Project No: CUHK 2050263).

* Corresponding author. Tel.: +852-26098443; fax: +852-26035302.
E-mail address: zyliu@cse.cuhk.edu.hk (Z.-Y. Liu).

0925-2312/$ - see front matter
doi:10.1016/S0925-2312(03)00414-4

performed alternatively [2,7–9]. Moreover, instead of using the full covariance matrix, each covariance matrix is represented either in a constrained form in terms of the principal component and the noise variance [2,7] or an equivalent way that directly considers each principal component and the noise variance in a reconstruction cost [8]. They are more preferred especially in the case of small sample size, since the free parameters to be estimated are considerably reduced. Typically, we consider the following orthonormal decomposition for the covariance matrix Σ [9],

$$\Sigma = \varsigma I + W\Psi W^{\mathrm{T}} \tag{2}$$

where $\Sigma \in \mathbb{R}^{d\times d}, W = [\phi_1, \ldots, \phi_m] \in \mathbb{R}^{d\times m}, \Psi = \mathrm{diag}(\sigma_1, \ldots, \sigma_m) \in \mathbb{R}^{m\times m}, m \leqslant d$, and d denotes the data dimensionality with the constraints: $W^{\mathrm{T}}W = \mathrm{I}, \varsigma > 0, \sigma_j > 0, (j = 1, 2, \ldots, m)$. It follows that $\phi_1, \ldots, \phi_m$ are the first m principal components of Σ. Thus, we can study the local PCA by estimating the gaussian mixture model with the covariance matrix given by Eq. (2). All the local PCA models studied so far, however, only consider each local cluster structure, with little emphasis being placed on the global topological structure of data. In this paper, in help of the Kohonen's SOM we present a topological local PCA to make up for the deficiency. Moreover, we adopt a newly proposed self-organizing strategy [9] to accelerate the self-organizing process. In the rest of this paper Section 2 is devoted to the topological local PCA and a comparison between two self-organizing strategies and Section 3 presents an experiment, followed by the conclusions in Section 4.

2. Topological local PCA

In this section we introduce the topological local PCA model based on the Kohonen's SOM. The SOM, as an unique data analysis tool that can explore the underlying data topological structure [4], usually consists of two layers with the lower one as the input associated with the samples, and the upper one being an array of nodes (neurons) that are usually laid out in a two-dimensional rectangular or hexagonal lattice. The SOM is usually trained by the competitive learning, with the neighbors of the winner node also learned in a certain extent. When viewed as a clustering approach, the conventional SOM actually makes an assumption that all clusters possess spherical shape.

As shown in Fig. 1, we generalize the conventional SOM to the topological local PCA model, with each neuron on the top layer associated with a weight vector $\{\alpha_i, \mu_i, \varsigma_i, W_i, \Psi_i\}$ which parameterizes a local subspace structure. Compared with the local PCA model, the topological local PCA model further considers the global topological relationship among the clusters, via the linkages between the neurons. On the other hand, compared with the conventional SOM which assumes all clusters are with spherical shape, the topological local PCA further considers the cluster structure, via each neuron's weight vector.

Based on the following general distance metric between the sample $x(t)$ and cluster i [9–11],

$$d_{it} = -\ln[\alpha_i G(x(t)|\mu_i, \Sigma_i)], \tag{3}$$

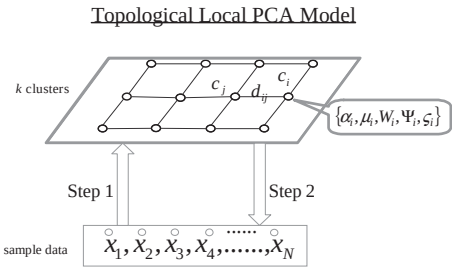

Fig. 1. Sketch of the topological local PCA model

where each covariance matrix Σ_i is given by Eq. (2), in help of the conventional self-organizing strategy the topological local PCA model can be learned via the following iterative algorithm,

Step 1: For sample $x(t)$, determine the winner cluster c:

$$c = \arg\min_i \{-\ln[G(x(t)|\mu_i, \Sigma_i)\alpha_i]\} \tag{4}$$

Step 2: Update the parameters $\theta_i \in \{\alpha_i, \mu_i, W_i, \Psi_i, \varsigma_i\}$:

$$\theta_i(t+1) = \theta_i(t) + h_{ci}(t)\eta(t)\nabla_{\theta_i}\ln[\alpha_i G(x(t)|\mu_i, \varsigma_i \mathrm{I} + W_i \Psi_i W_i^{\mathrm{T}})], \tag{5}$$

where h_{ci} denotes the neighborhood kernel around the winner neuron c, for which a typical choice is as $h_{ci}(t)=\exp(-d_{ci}^2/2\sigma^2(t))$ [4] with d_{ci} denoting the distance between neuron i and c, and $\sigma(t)$ monotonically decreasing with time t. The detailed updating forms are referred to [5].

The conventional self-organizing strategy has been the major approach for the self-organizing process since its inception. However, it is rather slow as it is time consuming to regularly organize the neurons starting from a random initialization.

A new self-organizing strategy described as "*stronger persons gain and then form a team*" was proposed in [10] (Section 5). By this strategy, a group of winning neurons are all updated and are relocated to become topologically close, in help of a learning process [10] in which the second, third ... winner neurons are forced to exchange with the neighboring neurons around the winner one. As discussed in [1], the two self-organizing are analogous in some sense to the way human live. In the conventional strategy, every neuron should adapt itself to his surrounding environment in order to get along with their neighbors; On the other hand, in the new strategy the neurons migrate and team up directly with the those who possess similar characteristic. Therefore, they need not go through the process of adapting themselves slowly. Specifically, for the purpose of implementing the topological local PCA model the three steps in [10] are adopted as follows:

Step 1: For a sample $x(t)$, find the first n winner neurons, denoted by N_c, according to (4).

Step 2: Update the weights of all of the neurons in N_c according to (5) with $h_{ci}(t)$ being removed.

Step 3: The second to the last one in N_c are exchanged with the topologically neighboring neurons of the first winner one.

As suggested in [10], we use it to replace the conventional strategy in the early phase of learning to accelerate the self-organizing process.

3. Experimental illustration

We present an experiment to illustrate the topological local PCA model which can explore both each local structure and the global topological structure, as well as compare the convergence speed of the two self-organizing strategies. The 900 20-dimensional data samples used in this experiment are equally generated via 30 gaussian's, whose covariance matrices are all with "line" shapes, i.e., the largest eigenvalue is much bigger than the remainder 19 ones. The 30 gaussian's roughly belong to 5 *super* clusters, as shown in Fig. 2, where the data is projected on the plane spanned by the first two dimensions.

We adopt a 5×6 rectangular grid for the topological local PCA to map the 30 clusters, and adopt the difference d of the average distance between two iterations as the convergence criteria, i.e., $d(t) = (1/n)\sum_{i=1}^{n} |d_i(t) - d_i(t-1)|$, where d_i denotes the distance between two neighboring neurons, and n is the total number of linkages,

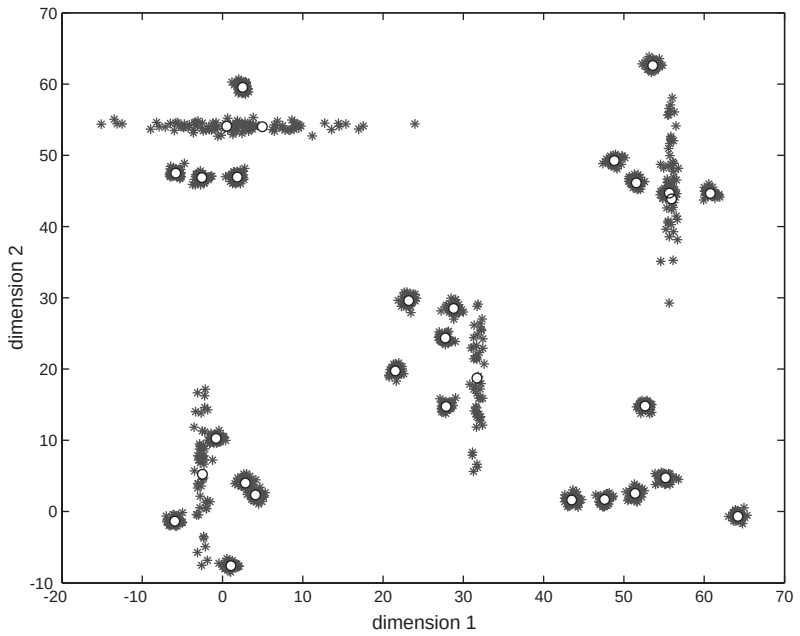

Fig. 2. The original 20-dimensional data samples, for viewable, projected on the plane spanned by the first two dimensions, where the 30 clusters are all with line shape.

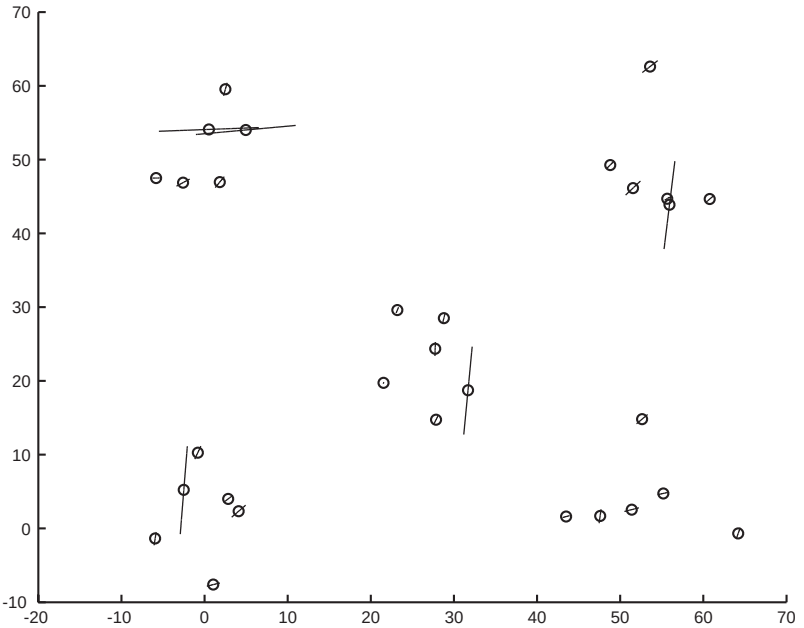

Fig. 3. Experiment results projected on the plane spanned by the first two dimensions, where each "○" represents one cluster, and the line represents its first principal component.

which is 49 for the 5×6 rectangle lattice. The algorithm is regarded as converged when d becomes smaller than a preset threshold.

The experimental results are shown in Figs. 3 and 4 and the convergence comparison is shown in Fig. 5. Fig. 3 shows the learned local cluster structure which, for viewable, is projected onto the plane spanned by the first two dimensions. The topological relationships among the clusters are illustrated by Fig. 4, where the gray value, normalized by the largest one, of each neuron represents the average distance between itself and its neighboring neurons. We can notice that Fig. 4 can be roughly partitioned into five "regions" that reflect the topological relationship among the five *super* clusters of the original data samples.

From Fig. 5 we can notice that the new strategy is much faster than the conventional one, especially at the beginning dozens of steps. Specifically, while the new strategy takes about 350 steps ($\sim$ 27.3 s) to converge, the conventional one takes about 500 steps ($\sim$ 42.4 s).

4. Conclusions

In this paper we studied the topological local PCA model with the help of Kohonen's SOM. We also adopted a newly proposed self-organizing strategy that can accelerate the conventional self-organizing speed to train the model.

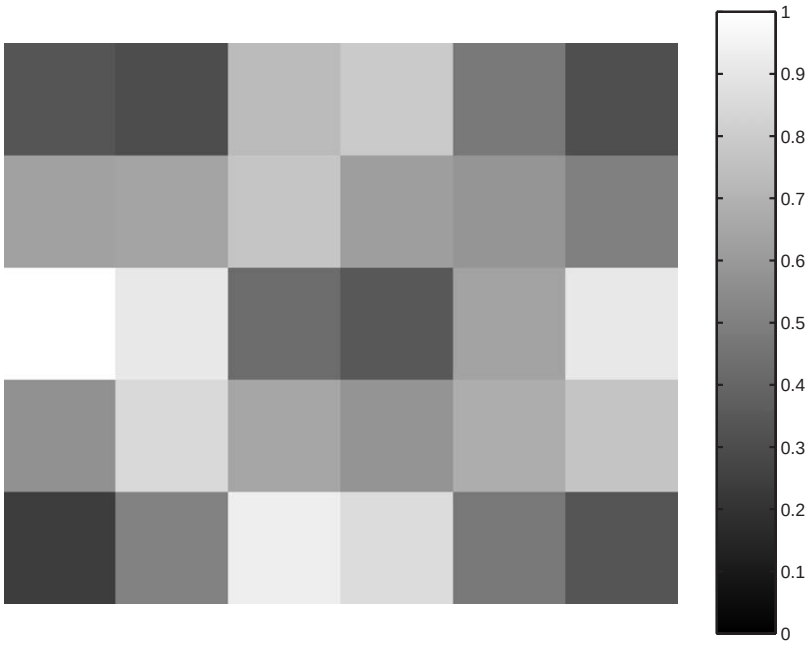

Fig. 4. The obtained global topological relationships.

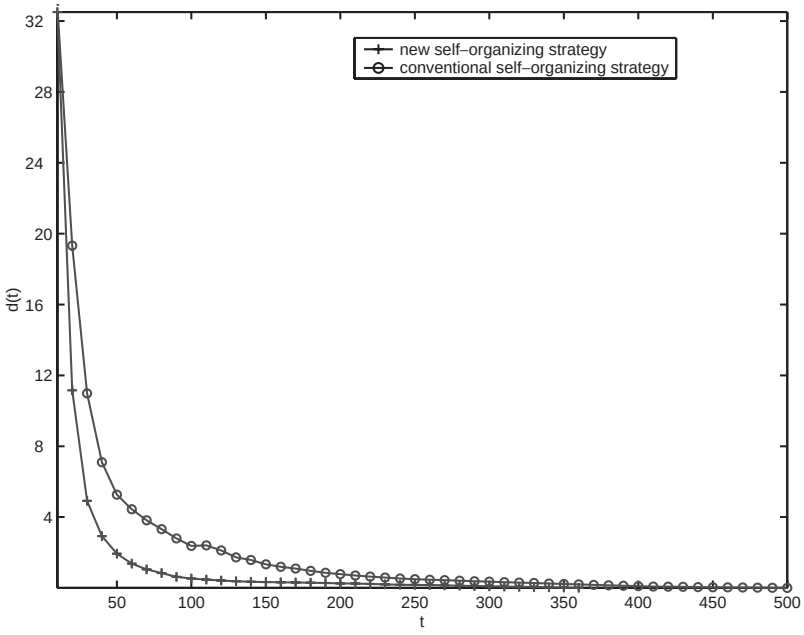

Fig. 5. The comparative convergence curves.

References

[1] K.Y. Chan, W.S. Chu, L. Xu, Empirical comparison between two computational strategies for topological self-organization, Proceedings of the Fourth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'03), to appear.
[2] G.E. Hinton, M. Revow, P. Dayan, Recognizing handwritten digits using mixtures of linear models, Adv. Neural Inform. Process. System 7 (1995) 1015–1022.
[3] N. Kambhatla, T. Leen, Dimension reduction by local principal component analysis, Neural Comput. 9 (7) (1997) 1493–1516.
[4] T. Kohonen, Self-Organizing Maps, Springer, Berlin, Heidelberg, 1995.
[5] Z.Y. Liu, K.C. Chiu, L. Xu, Improved system for object detection and star/galaxy classification via local subspace analysis, Neural Networks 16 (3,4) (2003) 437–451.
[6] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 1986.
[7] M. Tipping, C. Bishop, Mixtures of probabilistic principal component analysis, Neural Comput. 11 (2) (1999) 443–482.
[8] L. Xu, Multisets modeling learning: an unified theory for supervised and unsupervised learning, Proceedings of 1994 IEEE International Conference on Neural Networks I (ICNN'94, Orlando, FL), 1994, pp. 315–320.
[9] L. Xu, An overview on unsupervised learning from data mining perspective, Advances in Self-Organizing Maps: Proceedings of the Workshop on Self-Organizing Maps (WSOM'01), 2001, pp. 181–210.
[10] L. Xu, Byy harmony learning, structural rpcl, and topological self-organizing on unsupervised and supervised mixture models, Neural Networks 15 (8–9) (2002) 1125–1151.
[11] L. Xu, Data smoothing regularization, multi-sets-learning, and problem solving strategies, Neural Networks 3 (2003), in press.