Neurocomputing 72 (2009) 1817–1825

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Linear manifold topographic map formation based on an energy function with on-line adaptation rules

Peyman Adibi, Reza Safabakhsh *

Computational Vision/Intelligence Laboratory, Computer Engineering Department, Amirkabir University of Technology, Tehran 15914, Iran

ARTICLE INFO

Article history:
Received 4 December 2007
Accepted 3 July 2008
Communicated by E.W. Lang
Available online 29 July 2008

Keywords:
Linear manifold
Topographic map
Subspace
Self-organizing map
Energy function
On-line learning

ABSTRACT

The lack of an energy function is an important problem in many topographic map formation methods. This paper describes formation of a map, called linear manifold topographic map, based on minimization of an energy function. Using multiple low-dimensional linear manifolds as data representation elements, the data distributions of many problems with high-dimensional data spaces can be simply and parsimoniously modeled. Two sets of on-line adaptation rules are obtained based on stochastic gradient descent on the energy functions devised for a *soft* and a *hard* data assignment. Experimental results show good performance of the map in comparison to other relevant techniques.

1. Introduction

Learning low-dimensional manifolds for representing high-dimensional data is a common and useful technique for data representation with many applications in pattern recognition and signal processing domains. In general, three learning approaches can be considered for this type of data representation: (1) finding a single subspace or linear manifold representing the data, (2) obtaining multiple linear manifolds or subspaces from the data, and (3) learning nonlinear manifolds of the data. The second approach can be considered as a compromise between the other two. Typically, the methods of the first approach have limited representation abilities because of the global linearity of the learnt manifold. On the other hand, the methods in the third approach are usually too complex, due to dealing with nonlinear manifolds. In addition, some of them do not show a good performance on the large-scale real-world problems or special data manifolds [8,43]. Multiple linear manifold learning techniques, i.e. the methods in the second approach, use a piecewise linear representation, which avoids the global linearity limitation of the first approach and the complexity and other problems of the third approach at the same time.

Principal component analysis (PCA) [9] and independent component analysis (ICA) [4] can be mentioned as popular examples of the first approach. Examples of the third approach include principal curves and surfaces [12,13], nonlinear PCA [24,41], kernel PCA [36], ISOMAP [37], local linear embedding (LLE) [32], and Laplacian eigenmap [3]. A variety of efforts can be found which belong to the second approach. Local PCA methods such as [10,16,18] (among others) try to make PCA in partitions of data space. The probabilistic PCA (PPCA) mixture model, proposed by Tipping and Bishop [39], uses the maximum likelihood estimation with the expectation-maximization (EM) algorithm to find the linear manifolds. Its advantage over the previously proposed local PCA approaches is in defining a probability density for the data.

The adaptive subspace self-organizing map (ASSOM) proposed by Kohonen [19] is another model of the second approach mentioned above, which learns the basis vectors of multiple lower-dimensional subspaces using a competitive on-line learning method in a topographic map. Emergence of the transformation invariant feature filters in the network was demonstrated by Kohonen et al. [20,23]. Application of the model in texture segmentation is studied by Ruiz-del-Solar et al. [33–35]. ASSOM is a variant of the self-organizing map (SOM) [21], which takes advantage of SOM's key features such as topology preservation and on-line (or incremental) learning ability. These features are the main advantages of the ASSOM in comparison to the PPCA mixture model. In addition, there are some biological interpretations for ASSOM (e.g. see [28,22]). Some problems are also reported with ASSOM, such as slow convergence and instability [44,26], and limited representation power due to the origin intersection property of the subspaces [25].

Zhang et al. [44] tried to improve the stability of ASSOM using local nonlinear PCA. As a result, the simplicity advantage of subspaces due to linearity is lost in their work. Liu [25] added a

* Corresponding author.
E-mail addresses: adibip@aut.ac.ir (P. Adibi), safa@aut.ac.ir (R. Safabakhsh).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.07.003

simple SOM learning process to the ASSOM algorithm in order to remove its origin intersection problem. He called his model adaptive manifold SOM (AMSOM), since the subspaces are replaced by linear manifolds. In the PCASOM network, proposed by Lopez-Rubio et al. [26], the local mean and local covariance matrices are iteratively approximated in a topographic map. The model needs to solve the eigenvalue–eigenvector problem periodically during the learning process to obtain the basis vectors of each linear manifold. This makes the computational cost of the algorithm higher than the original ASSOM which directly learns the basis vectors.

Lopez-Rubio et al. [27] also proposed a similar model called PCA competitive learning (PCACL), but without topology preservation. Another improved model, called AOSSOM, is proposed by Zheng et al. [45], which, as in [25,26], learns the mean vectors in addition to the basis vectors of the manifolds. Recently, the information maximization approaches in the kernel-based linear manifold topographic maps have also been proposed by the authors [1,2], which prevent the above-mentioned problems of ASSOM, while maintaining its simplicity, low computations, topology preservation, and on-line learning advantages.

As noted by Bishop et al. [6] and Heskes [14], there are several theoretical issues with the standard SOM. One of its most important problems is the lack of an objective or energy function which could be optimized by the algorithm. This is the result of heuristically inserting the neighborhood function in the stochastic approximation-based learning rules of the SOM. This problem is also inherited by the ASSOM and all of its modifications mentioned above. To remove this problem, an energy function for the linear manifold topographic map formation is proposed in this paper. Learning rules are found which perform gradient descent on the energy function to learn the mean and basis vectors of multiple linear manifolds in a topographic map. Inspired by the work of Heskes [14], the energy function is defined to include the neighborhood functions. This way, topology preservation of the map will be more reliable. An important application of this property is in data visualization tasks. Topology preservation is also a key factor when SOM is considered as a *piecewise constant* approximator of data principal curves or surfaces [29]. In the linear manifold topographic map, topology preservation, instead, gives us a *piecewise linear* approximation of the data principal manifold.

The paper is organized as follows. Some preliminary definitions are given in Section 2. The on-line learning rules are obtained for a soft and a hard version of winner assignment in Section 3. Then, the performances of the rules are evaluated through experiments in Section 4, and Section 5 concludes the paper.

2. Preliminary definitions

The n-dimensional data points are assumed to approximately locate on multiple *regional* d-dimensional linear manifolds, with $d<n$. There are N processing units or neurons arranged in a p-dimensional fixed-topology lattice $\mathbf{L}$. Usually, the lattice dimensionality p is considered to be equal to 1 or 2, especially for visualization purposes, but in general p has an arbitrary value. To each neuron i of the map, a linear manifold, determined by a mean vector $\mathbf{m}_i$ and d orthonormal basis vectors $(\mathbf{b}_i^1,\ldots,\mathbf{b}_i^d)$, is assigned. The learning subjects thus will be the set of mean vectors and basis vectors of all the map neurons, denoted by M and B, respectively. Also, a neighborhood function h_{ij} is defined between each of the two neurons i and j of the map, which is usually a decreasing function of the lattice distance between the two neurons.

For each input vector $\mathbf{x}$ and each neuron i, a vector $\boldsymbol{\varphi}_i$ is defined as $\boldsymbol{\varphi}_i=\mathbf{x}-\mathbf{m}_i$. The vector $\boldsymbol{\varphi}_i$ can be decomposed into two orthogonal vectors $\hat{\boldsymbol{\varphi}}_i$ and $\tilde{\boldsymbol{\varphi}}_i=\boldsymbol{\varphi}_i-\hat{\boldsymbol{\varphi}}_i$ by projecting it onto the linear manifold as

$$\hat{\boldsymbol{\varphi}}_i=\sum_{k=1}^{d}(\boldsymbol{\varphi}_i^{\mathrm{T}}\mathbf{b}_i^k)\mathbf{b}_i^k,\quad \tilde{\boldsymbol{\varphi}}_i=\boldsymbol{\varphi}_i-\sum_{k=1}^{d}(\boldsymbol{\varphi}_i^{\mathrm{T}}\mathbf{b}_i^k)\mathbf{b}_i^k \tag{1}$$

Two distances, $\hat{r}_i=\|\hat{\boldsymbol{\varphi}}_i\|$ and $\tilde{r}_i=\|\tilde{\boldsymbol{\varphi}}_i\|$, are then defined as *on-manifold* and *off-manifold* distances of $\mathbf{x}$ with respect to the manifold of neuron i, where $\|\cdot\|$ is the Euclidean norm of a vector. Fig. 1 shows the projection of an input vector in the two-dimensional (2-D) input space on a one-dimensional (1-D) linear manifold.

In the original ASSOM, only the off-manifold distances $\tilde{r}_i$ contribute in the error function, and thus the algorithm is insensitive to the on-manifold distances. It is clear that clustering the data based on the proximity to the subspaces alone (as done in the original ASSOM) will not necessarily result in localized partitions [17]. But, when we consider adaptable mean vectors for linear manifolds, we implicitly assume a regional nature for the manifolds. In other words, we want a neuron to represent the data which are not only close to its manifold, but also moderately close to its mean vector as well. It means that the lower on-manifold distances as well as off-manifold ones should result in lower error function values. Thus, for neuron i and input $\mathbf{x}$, the *regional error* is defined as

$$e_i(\mathbf{x})=\frac{1}{2}\sum_j h_{ij}(\alpha_j^2\hat{r}_j^2+\tilde{r}_j^2),\quad 0\leqslant\alpha_j\leqslant 1 \tag{2}$$

We see that a combination of on- and off-manifold distances is used in the above error term. Also, the distances of all the map neurons j contribute in this error proportionally to the neighborhood function h_{ij}, to be able to define the energy function for the topographic map. The parameter α adjusts the mutual importance of $\hat{r}$ and $\tilde{r}$ in the error term. Its value cannot be greater than one in order to maintain the manifold nature of the representation elements. Note that if $\alpha=0$, the on-manifold distance is removed from the error and we will have the global linear manifolds similar to the ASSOM. This will result in problems in data representation and, as mentioned before, it is not consistent with the local PCA approaches. On the other hand, if $\alpha=1$, the effect of the basis vectors is removed and the model will be similar to a type of SOM presented in [14]. Thus, typically we must have $0<\alpha<1$. The value of α in this range indicates that the role of the off-manifold distance, as expected from a manifold learning method, is more important than the role of the on-manifold distance in the regional error term.

In a probabilistic framework, we consider the probability of assigning current input vector $\mathbf{x}$ to neuron i as p_i, with the constraint $\sum_i p_i=1$. Now, for the current input vector, the

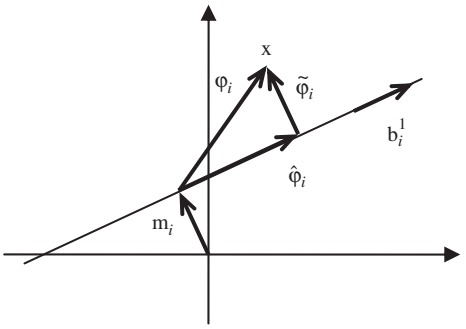

Fig. 1. The projection of a 2-D input data onto a 1-D linear manifold with an assigned mean vector.