Neurocomputing 158 (2015) 1–12

Contents lists available at ScienceDirect

Neurocomputing

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

Local regularization concept factorization and its semi-supervised extension for image representation

Zhenqiu Shu*, Chunxia Zhao, Pu Huang

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 15 September 2014
Received in revised form
5 January 2015
Accepted 7 February 2015
Communicated by: Haowei Liu
Available online 16 February 2015

Keywords:
Local learning
Predictor
Label
Semi-supervised
Constraints
Clustering

ABSTRACT

Matrix factorization methods have been widely applied for data representation. Traditional concept factorization, however, fails to utilize the discriminative structure information and the geometric structure information that can improve the performance in clustering. In this paper, we propose a novel matrix factorization method, called *Local Regularization Concept Factorization* (*LRCF*), for image representation and clustering tasks. In LRCF, according to local learning assumption, the label of each sample can be predicted by the samples in its neighborhoods. The new representation of our proposed LRCF can encode the intrinsic geometric structure and discriminative structure of the high-dimensional data. Furthermore, in order to utilize the label information of labeled data, we propose a semi-supervised version of LRCF, namely *Local Regularization Constrained Concept Factorization* (*LRCCF*), which incorporates the label information as additional constraints. Moreover, we develop the corresponding optimization schemes for our proposed methods, and provide the convergence proofs of the optimization schemes. Various experiments on real databases show that our proposed LRCF and LRCCF are able to capture the intrinsic latent structure of data and achieve the state-of-the-art performance.

1. Introduction

In the past years, data representation has become a fundamental problem in various research fields like computer vision and machine learning [1–8]. Thus, we often need to deal with the high dimensional data. As we all know, it poses a serious challenge for data analysis and processing. Consequently, many efforts have been devoted to seeking a suitable low-dimensional representation for the high-dimensional data. Matrix factorization is a popular technique for data representation, which aims to find two or more low rank matrices to approximate the original data. Up till now, many researchers have proposed a set of matrix factorization methods [9–16] to represent the high-dimensional data based on different purposes.

Among matrix factorization methods, PCA [9] and LDA [10] are two well-known methods for data representation and feature extraction, which have been widely used to learn a low-dimensional representation. PCA and LDA effectively see only the global geometric structure, but they fail to discover the underlying manifold structure. To solve this issue, a variety of manifold-based learning algorithms, such as isometric feature mapping (ISOMAP) [11], locally linear embedding (LLE) [12] and Laplacian Eigenmap (LE) [13], have been developed for dimensionality reduction and feature extraction. One limitation of these manifold learning methods is that they cannot map new coming samples. In order to solve this issue, He et al. proposed Locality Preserving Projections (LPP) [14] and Neighborhood Preserving Embedding (NPE) [15], which are the linearization of LE and LLE, respectively. In addition, kernel-based techniques [16] are employed to deal with the nonlinear distribution data. The basic idea is to implicitly map the original data into high-dimensional Reproducing Kernel Hilbert Space (RKHS) via kernel trick. Such methods make it possible for the nonlinear structure of data in original input space to become linear in RKHS space, and thus the linear techniques in pattern recognition can be applied to handle the data in kernel space.

Unlike the methods mentioned previously, NMF [17] has received considerable attention due to its psychological and physiological interpretation in the human brain. NMF, as a parts-based representation method, tries to find two low-rank non-negative matrices whose product is a good approximation to the original high-dimensional data. However, in practical applications, it is difficult that all elements of the original data are required to be non-negative due to noise or outlier. In addition, NMF fails to be performed on negative data owing to the non-negative limitation. Besides, it is difficult that NMF can be kernelized to improve the performance in clustering. To overcome these disadvantages, a variation of NMF, namely Concept Factorization (CF) [18], is proposed for data representation. In CF, each cluster can be linearly represented by all the samples, and each

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.02.014
0925-2312/

sample can be linearly represented by the cluster centers. Compared with NMF, CF can not only be kernelized, but also can be performed in any case.

Recently, Cai et al. [19] developed a graph regularization non-negative matrix factorization (GNMF) algorithm to capture nonlinear manifold structure of the original data set with resort to the Laplacian graph regularization technique. This approach can preserve the intrinsic geometrical manifold structure properties from the original high-dimensional space to new representation space. Unfortunately, during matrix factorization processing, GNMF only considers the favorite similarities of the data pairs, and neglects the dissimilarities for the data pairs. To solve this issue, a general framework, called non-negative graph embedding (NGE) [20], is proposed for non-negative data decomposition by integrating the characteristics of both intrinsic and penalty graphs. In NGE, the intrinsic graph and the penalty graph are used to characterize the intra-class compactness and the variance information of the data, respectively. GNMF and NGE, however, take no consideration of multi-geometry information of the data. To overcome this drawback, Zeng et al. [20] proposed a hyper-graph regularized nonnegative matrix factorization (HGNMF) algorithm for data representation. HGNMF extracts the multi-geometry information of samples by constructing a weighted hyper-graph regularization term. Furthermore, Cai et al. [21] developed a locally consistent concept factorization (LCCF) algorithm for documents clustering. It assumes that the nearby data points are likely to be in the same cluster, which is called local consistency assumption. GNMF and LCCF are able to discover the underlying local geometrical structure, but the high dimensional data points may not always satisfy the locality conditions. Since the local points share the greatest similarity, it would be more natural to represent the basis vectors by using a few nearby anchor points, which may lead to a more efficient representation of the data. To solve this issue, Chen et al. [22] proposed a non-negative local coordinate factorization (NLCF) method by adding a local coordinate constraint that can ensure the locality of the low dimensional representation. Then Liu et al. [23] proposed a locality-constrained concept factorization (LCF) algorithm to impose a locality constraint on the objective function of concept factorization. In reality, locality constraint cannot well reveal the intrinsic structure since it only requires the concept to be as close to the original data points as possible. To address this problem, Li et al. [24] introduced a graph-based local concept coordinate factorization (GLCF) method based on LCF. GLCF not only respects the intrinsic structure of the data via manifold kernel learning, but also considers the locality constraints in revealing the underlying concepts. Additional, some further studies of the matrix factorization have also been developed [25–31] based on different constraints during the last few years.

Inspired by the local learning assumption, we propose a novel concept factorization method, called *Local Regularization Concept Factorization* (*LRCF*), for data representation and clustering tasks. LRCF explicitly considers the local structure of the data by using the local learning technique during the decomposition process. Specifically, for each data point, we can construct a local label predictor to estimate the label of this data point. Besides, in order to consider the available label information, we propose a semi-supervised version of LRCF, called *Local Regularization Constrained Concept Factorization* (*LRCCF*), which takes the limited label information and local structure information into account, simultaneously. Our experimental evaluations on several real data sets show that our proposed LRCF and LRCCF perform better than other state-of-the-art matrix factorization methods.

The major contributions of this paper lie in:

(1) In our proposed LRCF, the label of each data sample can be estimated by the samples in its local neighborhood by adding the local learning regularization. Thus, the low dimensional representation of LRCF can encode both the discriminative information and the instinct geometric structure information of the high-dimensional data. Compared with the conventional CF, we can obtain a better low-dimensional representation of the high-dimensional data.
(2) LRCCF is a semi-supervised learning algorithm, that is, LRCCF makes full use of the prior knowledge of the data. Specifically, it imposes the label information as hard constraints on the basis vectors. Therefore, the original data from the same class can map together in the low dimensional representation space. As a result, LRCCF jointly takes account of both the label information and local structure information. In this way, the new representation of LRCCF can exhibit more discriminative power than the unsupervised learning methods, such as LRCF. Moreover, LRCCF utilizes the label information in a parameter-free way.
(3) We develop the corresponding multiplicative updating optimization schemes with proving their convergence to solve the proposed algorithms. In addition, we give a general scheme to perform on positive as well as negative data.

The rest of this paper is organized as follows. We provide a brief review of NMF, CF and local learning regularization technique in Section 2. We present our proposed approaches as well as the detailed derivations in Section 3. Experimental results are reported in Section 4 with considerable analysis. Finally, we draw a conclusion in Section 5.

2. Related work

In this section, we primarily review some related work to our work.

2.1. NMF

Given a non-negative matrix $\mathbf{X} = [x_1, \cdots, x_n] \in \mathbf{R}^{m\times n}$, where each vector x_i represents a sample. NMF aims to seek two low-rank non-negative matrices $\mathbf{U} \in \mathbf{R}^{m\times k}$ and $\mathbf{V} \in \mathbf{R}^{n\times k}$, where $k \lll \min(m, n)$. Such that the product of $\mathbf{U}$ and $\mathbf{V}$ can well approximate the original data matrix $\mathbf{X}$. Thus, the objective function of NMF can be expressed as follows:

$$O = \|\mathbf{X} - \mathbf{U}\mathbf{V}^T\|_F^2 \quad s.t.\ \ \mathbf{U} > 0, \mathbf{V} > 0 \tag{1}$$

In this paper, $\|\cdot\|_F$ denotes the Frobenius norm. It is easy to check that the objective function of NMF is not convex in both $\mathbf{U}$ and $\mathbf{V}$ together. It is, therefore, unrealistic to find the global solution of the objective function of NMF. The most popular optimization method of NMF is the multiplicative updating algorithm proposed by Lee and Seung [17]. As such, the objective function of NMF in Eq. (1) can be solved by the following updating rules:

$$u_{ij}^{t+1} \leftarrow u_{ij}^{t}\frac{(\mathbf{X}\mathbf{V})_{ij}}{(\mathbf{U}\mathbf{V}^T\mathbf{V})_{ij}}$$

$$v_{ij}^{t+1} \leftarrow v_{ij}^{t}\frac{(\mathbf{X}^T\mathbf{U})_{ij}}{(\mathbf{V}\mathbf{U}^T\mathbf{U})_{ij}}$$

2.2. CF

In CF, each base vector u_j can be represented as a linear combination of the data samples $u_j = \sum_i w_{ij} x_i$, where$w_{ij} \ge 0$. Let $\mathbf{W} = [w_{ij}] \in \mathbf{R}^{N\times K}$, CF aims to seek the following approximation:

$$\mathbf{X} \approx \mathbf{X}\mathbf{W}\mathbf{V}^T \tag{2}$$