Pattern Recognition 48 (2015) 1011–1022

Contents lists available at ScienceDirect

Pattern Recognition

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

Low-rank matrix factorization with multiple Hypergraph regularizer

Taisong Jin [a,b], Jun Yu [c,*], Jane You [d], Kun Zeng [a], Cuihua Li [a], Zhengtao Yu [e]

[a] Computer Science Department, School of Information Science and Engineering, Xiamen University, Xiamen, China
[b] Science and Technology on Electro-optic Control Laboratory, Luoyang, 471009, China
[c] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 10018, China
[d] Department of Computing, The Hong Kong polytechnic University, Kowloon, Hong Kong
[e] School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, 650500, China

ARTICLE INFO

Article history:
Received 11 January 2014
Received in revised form
23 July 2014
Accepted 2 September 2014
Available online 16 September 2014

Keywords:
Hypergraph
Matrix factorization
Manifold
Alternating optimization

ABSTRACT

This paper presents a novel low-rank matrix factorization method, named MultiHMMF, which incorporates multiple Hypergraph manifold regularization to the low-rank matrix factorization. In order to effectively exploit high order information among the data samples, the Hypergraph is introduced to model the local structure of the intrinsic manifold. Specifically, multiple Hypergraph regularization terms are separately constructed to consider the local invariance; the optimal intrinsic manifold is constructed by linearly combining multiple Hypergraph manifolds. Then, the regularization term is incorporated into a truncated singular value decomposition framework resulting in a unified objective function so that matrix factorization is changed into an optimization problem. Alternating optimization is used to solve the optimization problem, with the result that the low dimensional representation of data space is obtained. The experimental results of image clustering demonstrate that the proposed method outperforms state-of-the-art data representation methods.

1. Introduction

In recent years, the development of data acquisition technology has resulted in a rapid accumulation of a large amount of high-dimensional data in many fields. Selecting suitable data representation is usually the first step to design an effective data analysis algorithm. Many data representation methods have been presented, among which matrix factorization [1–3] is important for simple and easy implementation such as the following: QR, LU decomposition, Truncated singular value decomposition (TSVD), Non-negative matrix factorization (NMF), etc. TSVD and NMF are two common methods, which can be implemented iteratively via matrix–vector products and multiplicative updating rules, respectively. Because of the current popularity of NMF, most work focuses on introducing a new regularization term to the original NMF framework. These regularization terms make the data representation more suitable for classification, clustering and retrieval. Many NMF variants [4–7] were proposed by adding a regularization term such as sparse constraint to the original NMF framework. The geometrical structure of data space is also considered as a manifold regularization term, and it usually assumes that the data is sampled from the intrinsic low-dimensional manifold. Furthermore, manifold learning [40–41] uses a so-called locally invariance idea, namely nearby points are likely to be similarly embedded to discover the underlying geometrical structure.

Manifold learning [8–11,36–39] usually uses a p-nearest neighbor graph to model the intrinsic manifold. This graph considers only the pair-wise relationship between the two data samples. Due to manifold underlay from the sampled data set, selecting the optimal manifold for the task of analyzing specific data is usually difficult and time consuming. To address this problem, a multiple graph regularization term [12] is added to the original NMF framework. A multiple graph regularized NMF method is proposed where graph selection and matrix factorization are automatically implemented by alternating optimization. However, another problem, ignored by most researchers, is the balance between matrix factorization and graph regularization. If the same approximate rank is taken, NMF has a larger approximation error in the Frobenius norm than the traditional TSVD method. A good data representation method should have as small an approximation error as possible and yet preserve the nonlinearity of data space. Based on this idea, a novel, low-rank matrix factorization [13], named MMF, is proposed where the manifold regularization term is added to the TSVD framework to leverage the regularization term and matrix factorization. MMF more effectively preserves the nonlinear structure of data space.

Based on recent efforts in both matrix factorization and manifold learning, this paper proposes a novel low-rank matrix factorization method with a multiple Hypergraph regularizer. It uses Hypergraph

* Corresponding author.
E-mail address: yujun@hdu.edu.cn (J. You).

http://dx.doi.org/10.1016/j.patcog.2014.09.002
0031-3203/

to model the high-order relationship among the data samples. In order to select the optimal intrinsic manifold for a low-rank matrix factorization task, the proposed method fuses multiple Hypergraphs to approximate the intrinsic manifold inspired by an ensemble manifold regularizer [14]. Furthermore, to enhance the regularization effectiveness, a novel multiple Hypergraph regularization term is formulated and incorporated into the traditional TSVD framework.

Multiple graph regularized NMF [12] (MGrNMF) is closely relevant to our proposed method. MGrNMF performs better on real world datasets. However, it has two disadvantages: (1) the high-order relationship among samples are ignored in graph-based learning methods, and MGrNMF only considers the pair-wise relationship between two samples; (2) the balance between linear factorization and manifold regularization is ignored, i.e., the non-negative constraint in MGrNMF leads to the larger approximation error of linear factorization, which further affects the regularization term. The proposed regularization in MHGrMMF adopts Hypergraph instead of simple graph to model the intrinsic manifold. Hypergraph can effectively model high-order relationship among data samples. Therefore, the first disadvantage of MGrNMF is addressed. In addition, MHGrMMF incorporates Hypergraph regularization term into the TSVD framework, which relaxes the non-negativity constraint in MGrNMF to the orthogonal constraint. Due to the fact that TSVD framework has a smaller approximation error in Frobenius norm than NMF framework, MHGrMMF can effectively leverage the linearity factorization and the Hypergraph manifold regularization. Thus, the second disadvantage of MGrNMF is addressed.

The contributions of this paper are listed below:

1) A novel multiple Hypergraph regularization term is proposed where a Hypergraph, instead of a simple graph, is introduced to model the intrinsic manifold; therefore, a high-order relationship among the data samples is considered. A Hyperedge of a Hypergraph connects more than two vertices, which simultaneously capture the locality among the data samples within the same hyperedge. Furthermore, a multiple Hypergraph regularization term is formulated where the intrinsic manifold is approximated by the linear combination of the previously given Hypergraph Laplacians. Minimizing the proposed regularization term can guarantee the smoothness of data representations in low-dimensional space.
2) To enhance regularization efficiency, a low-rank matrix factorization algorithm is proposed, and the multiple Hypergraph regularization term is incorporated into the TSVD framework. Compared with the NMF framework, the proposed algorithm attempts to minimize the approximation error in low-rank factorization; at the same time, it preserves the nonlinear structure of data space. Thus, the proposed algorithm balances the best linear approximation of SVD and nonlinear dimensionality reduction.
3) The clustering performance of the proposed algorithm is analyzed by conducting comprehensive experiments. In addition, a comparison of experimental results with state-of-the-art data representation methods is provided, and a discussion of the details is also presented. In particular, we developed other graph-involved methods and compared them with the proposed method in four real world datasets.

The remainder of the paper is arranged as follows: In Section 2, we introduce related work. In Section 3, we present the unified objective function of the proposed method in detail. In Section 4, we propose an alternate iterative algorithm. In Section 5, we benchmark the proposed method, compare it with state-of-the-art methods, and present a discussion. We conclude the paper in Section 5.

2. Related work

2.1. NMF and TSVD

Matrix factorization is an important low-dimensional data representation method. Assuming N data points with their non-negative feature set $\mathbf{X} = \{\mathbf{X}_i\}$, $i = 1, \ldots, N$, we denote all their features as a non-negative data matrix $\mathbf{X} = [\mathbf{X}_i] \in \mathbb{R}_+^{M \times N}$, where the ith column, $\mathbf{X}_i$, is the feature vector of the ith data point. Matrix factorization aims to locate the two matrices $\mathbf{B} \in \mathbb{R}^{M \times r}$ and $\mathbf{F} \in \mathbb{R}^{r \times N}$, whose product approximates well the original matrix, $\mathbf{X}$, i.e., $\mathbf{X} \approx \mathbf{BF}$. Two common-cost functions to quantify the quality of the approximation are ℓ_2 norm and KL divergence. Recently, the Earth Mover's distance [15], which is more appropriate for the realistic imagery, is also used to quantify the approximation. In practice, we commonly have $r \ll M$ and $r \ll N$. Thus, the low-rank factorization essentially attempts to locate a compressed approximation of original data space where each feature vector $\mathbf{X}_i$ is approximated by a linear combination of the columns of $\mathbf{B}$, weighted by the components of $\mathbf{F}$, as

$$\mathbf{X}_i = \sum_{p=1}^{r} \mathbf{B}_p \mathbf{F}_{pi}. \quad (1)$$

Thus, $\mathbf{B}$ can be regarded as containing a set of basis vectors. Let $\mathbf{F}_i = [\mathbf{F}_{1i}, \ldots, \mathbf{F}_{ri}]$ denote the ith column of $\mathbf{F}$ where $\mathbf{F}_i$ is regarded as a coding vector, or as a new data representation of the ith data point with respect to the new basis, $\mathbf{B}$. Based on the widely used ℓ_2 norm, an objective function of low-rank matrix factorization is formulated as

$$\mathbf{O}^{MF} = ||\mathbf{X} - \mathbf{BF}||_F^2. \quad (2)$$

Various researchers add the constraint, $B > 0$, $F > 0$, to the objective function (2) for the desired properties, which results in many variants of low-rank matrix factorization. Among many variants, NMF and TSVD are two popular methods. NMF adds a non-negative constraint to the factorized matrices $\mathbf{B}$ and $\mathbf{F}$. This leads to part-based data representation, which is useful for the image signal, text, etc, because they are usually considered as low-rank structures and can be represented as an additive combination of few atom signals or sub-texts. Mathematically, the original NMF framework is defined as the minimization problem,

$$\min ||\mathbf{X} - \mathbf{BF}||_F^2 \quad \text{s.t.} \ \mathbf{B} \geq 0, \ \mathbf{F} \geq 0. \quad (3)$$

TSVD amounts to truncating the singular value expansion of $\mathbf{X}$ in such a way that the smallest singular values are discarded. As an alternative to SVD, TSVD has been widely applied to data analysis tasks. For instance, the well known PCA is the TSVD method applied to centered data. TSVD imposes a ortho-normal constraint on the factorized matrix $\mathbf{B}$, whose columns consist of a principal orthogonal basis of data space. The original TSVD is defined as the minimization problem

$$\min ||\mathbf{X} - \mathbf{BF}||_F^2 \quad \text{s.t.} \ \mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_r. \quad (4)$$

For large scale data, iterative methods effectively address the above mentioned two optimization problems. In actual practice, various researchers improve the NMF and TSVD frameworks by adding the regularization term. Sparseness is an important constraint. Among different sparseness constraints, one is selected as the regularization term for incorporation into the NMF framework to determine the better part-based representation of the data. In particular, ℓ_1 [4,5] and $\ell_{1/2}$ norm [6] constraints, respectively, are incorporated into the original NMF framework. Chen et al. [7] proposed a non-negative local coordinate factorization (i.e. a