Neurocomputing 138 (2014) 120–130

Contents lists available at ScienceDirect

Neurocomputing

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

Dual-graph regularized concept factorization for clustering

Jun Ye [a,b,*], Zhong Jin [a]

[a] School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing 210094, China
[b] School of Natural Sciences, Nanjing University of Posts & Telecommunications, Nanjing 210003, China

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form
26 December 2013
Accepted 13 February 2014
Communicated by Deng Cai
Available online 13 April 2014

Keywords:
NMF
Concept factorization
Dual-graph regularized
Document clustering
Image clustering

ABSTRACT

In past decades, tremendous growths in the amount of text documents and images have become omnipresent, and it is very important to group them into clusters upon desired. Recently, matrix factorization based techniques, such as Non-negative Matrix Factorization (NMF) and Concept Factorization (CF), have yielded impressive results for clustering. However, both of them effectively see only the global Euclidean geometry, whereas the local manifold geometry is not fully considered. Recent research has shown that not only the observed data are found to lie on a nonlinear low dimensional manifold, namely data manifold, but also the features lie on a manifold, namely feature manifold. In this paper, we propose a novel algorithm, called dual-graph regularized concept factorization for clustering (GCF), which simultaneously considers the geometric structures of both the data manifold and the feature manifold. As an extension of GCF, we extend that our proposed method can also be apply to the negative dataset. Moreover, we develop the iterative updating optimization schemes for GCF, and provide the convergence proof of our optimization scheme. Experimental results on TDT2 and Reuters document datasets, COIL20 and PIE image datasets demonstrate the effectiveness of our proposed method.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Clustering is one of the most important research topics in both machine learning and data mining communities. It arms at partitioning the data into groups of similar objects. An enormous number and variety of methods have been proposed over the past several decades to solve clustering problems [1]. Generally, clustering methods can be categorized as agglomerative and partitional. Agglomerative clustering methods group the data points into a hierarchical tree structure using bottom-up approaches. The procedure starts by placing each data point into a distinct cluster and then iteratively merges the two most similar clusters into one parent cluster. On the other hand, data partitioning methods decompose the data set into a given number of disjoint clusters which are usually optimal in terms of some predefined criterion functions [2]. Both of them have been well studied and investigated in previous literatures [3,4].

In the last decade, matrix factorization based approaches have attracted considerable attention for clustering. With regard to these methods, each text document or image in the corpus is often treated as a data point in the high dimensional linear space. Clustering analysis aims to look for similar data points and ensure them within the same cluster in maximum degree. Intuitively, similar samples are more likely to be grouped together than different ones, and this could be attributed to the fact that characteristics shared by similar ones in original data spaces are inherited by new representations in lower dimensional spaces, which makes the clustering more easily. There are particularly two popular matrix factorization methods widely applied to clustering analysis, i.e., Nonnegative Matrix Factorization (NMF) [5] and Concept Factorization (CF) [2]. CF mainly strives to address the limitations and meanwhile inherits all the strengths of NMF, such as better semantic interpretation and easily derived clustering results. In CF, each concept or component is modeled as a linear combination of the data points while each data point consists of a linear combination of the concepts. In general, CF is more advantageous than NMF, because of its merits that it can be applied to any data points taking both positive and negative values. However, regardless of NMF or CF, they only consider using the global Euclidean geometry to find new basis vectors, according to how the new data representation is generated [6]. However, many previous studies have shown human generated text data is probably sampled from a submanifold of the ambient Euclidean space [7–10]. In fact, the human generated text documents cannot possibly "fill up" the high dimensional Euclidean space uniformly.

* Corresponding author at: School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing 210094, China.
Tel.: +86 1360 1587 306.
E-mail address: yj8422092@163.com (J. Ye).

http://dx.doi.org/10.1016/j.neucom.2014.02.029
0925-2312/© 2014 Elsevier B.V. All rights reserved.

Therefore, the intrinsic manifold structure needs to be considered while learning new data representations [11]. Inspired by this, Li et al. [12] proposed discriminative orthogonal nonnegative matrix factorization (DON), in order to obtain a good data representation that preserves both the local geometrical structure and the global discriminating information. And also in order to preserve the intrinsically geometrical structure and use the prior knowledge, Li et al. [13] proposed locally constrained a-optimal nonnegative projection (LCA). They are all NMF-based methods.

Recently, Cai et al. [11] proposed locally consistent concept factorization (LCCF) based on CF to extract the underlying concepts which are consistent with the low dimensional manifold structure. The obtained concepts can well capture the intrinsic geometrical structure and the documents associated with similar concepts can be well clustered. However, the method mentioned above focuses on one-sided clustering, i.e., clustering the data based on the similarities along the feature. Considering the duality between data points and features, several co-clustering algorithms have been proposed and shown to be superior to traditional one-sided clustering [14–20]. Gu et al. [19] proposed a Dual Regularized Co-Clustering (DRCC) method based on semi-nonnegative matrix tri-factorization. In order to discover an appropriate intrinsic manifold, Li et al. [20] proposed realtional multimanifold co-clustering based on symmetric nonnegative matrix tri-factorization. Based on NMF, Shang et al. [17] proposed Graph Dual Regularization Non-negative Matrix Factorization (DNMF) for co-clustering, which achieves an encouraging performance.

Motivated by recent progress in dual regularization [17–20] and concept factorization [2,11], we propose a novel algorithm called dual-graph regularized concept factorization for clustering (GCF), which simultaneously considers the geometric structures of the data manifold as well as the feature manifold. We encode the geometric structure information of data and feature spaces by constructing two nearest neighbor graphs, respectively. Our proposed algorithm GCF is based on the CF, it can be optimized by iterative multiplicative updating schemes, and their convergence proof is been provided. To summarize, the main contributions of this work include:

1. We propose a novel dual-graph regularized concept factorization (GCF) algorithm which simultaneously considers the geometric structure information contained in data points as well as features.
2. We develop iterative multiplicative updating optimization schemes to solve our proposed algorithm GCF, and provide the convergence proof of the optimization scheme.

The remainder of this paper is organized as follows: Section 2 presents a brief overview of some related works. A novel GCF algorithm is proposed in Section 3. As an extension of GCF, the algorithm for negative data is described in Section 4. Experimental results on many real-world datasets are presented in Section 5. Section 6 is conclusions.

2. Related works

In this section, we briefly review some related works to our research work.

2.1. NMF

Consider a data matrix $\mathbf{X}=[\mathbf{x}_1,\dots,\mathbf{x}_N]\in\mathbf{R}^{M\times N}$, each column of $\mathbf{X}$ is a sample vector. NMF aims to decompose $\mathbf{X}$ into two low rank nonnegative matrices, basis matrix $\mathbf{U}=[u_{ik}]\in\mathbf{R}^{M\times K}$ and feature matrix $\mathbf{V}=[v_{jk}]\in\mathbf{R}^{N\times K}$, such that $\mathbf{X}\approx\mathbf{U}\mathbf{V}^T$, where $K\ll\min\{M,N\}$. Therefore, the objective optimization problem of NMF can be concluded as follows:

$$\min_{U,V}:\mathbf{J}_{\mathbf{NMF}}=||\mathbf{X}-\mathbf{U}\mathbf{V}^T||_F^2\quad s.t.\quad U,V\geq 0 \tag{1}$$

Several methods have been proposed to find a solution to this nonlinear optimization problem. The multiplicative updates rules were first investigated by Lee and Seung [21] as follows:

$$u_{ik}^{t+1}=u_{ik}^t\frac{(\mathbf{XV})_{ik}}{(\mathbf{U}\mathbf{V}^T\mathbf{V})_{ik}};\qquad v_{jk}^{t+1}=v_{jk}^t\frac{(\mathbf{X}^T\mathbf{U})_{jk}}{(\mathbf{V}\mathbf{U}^T\mathbf{U})_{jk}} \tag{2}$$

Theorem 1. *[21] for $\mathbf{X},\mathbf{U},\mathbf{V}\geq 0$, the objective function $\mathbf{J}_{\mathbf{NMF}}$ in Eq. (1) is nonincreasing under each of the above multiplicative updating rules stated in Eq. (2).*

The nonnegative constraints on $\mathbf{U}$ and $\mathbf{V}$ require the combination coefficients among different basis can only be positive. This is the most significant difference between NMF and other matrix factorization methods, e.g., SVD. Unlike SVD, no subtractions can occur in NMF. For this reason, it is believed that NMF can learn a parts-based representation have been observed in many real world problems such as face analysis, document clustering.

2.2. DRCC

Gu et al. [19] proposed a dual regularized co-clustering (DRCC) method based on graph regularized (semi-)NMF, which imposes graph regularization on both the data points and features cluster assignment matrices. The objective optimization problem can be concluded as follows:

$$\begin{aligned}\min_{U,S,\mathbf{V}}:\mathbf{J}_{\mathbf{DRCC}}&=||\mathbf{X}-\mathbf{USV}^T||_F^2+\lambda\mathrm{Tr}(\mathbf{V}^T\mathbf{L}_{\mathbf{V}}\mathbf{V})+\mu\mathrm{Tr}(\mathbf{U}^T\mathbf{L}_{\mathbf{U}}\mathbf{U})\\ &s.t.\quad \mathbf{U},\mathbf{V}\geq 0\end{aligned} \tag{3}$$

where $\lambda,\mu\geq 0$ are the regularization parameters, and $\mathbf{S}$ is a matrix whose entries can take any signs. $\mathbf{L}_{\mathbf{V}}=\mathbf{D}^{\mathbf{V}}-\mathbf{W}^{\mathbf{V}}$ is the graph Laplacian of the data graph which reflects the label smoothness of the data points, where $\mathbf{W}^{\mathbf{V}}$ is the weight matrix and $\mathbf{D}^{\mathbf{V}}$ is a diagonal matrix whose entries are column sums of $\mathbf{W}^{\mathbf{V}}$. $\mathbf{L}_{\mathbf{U}}=\mathbf{D}^{\mathbf{U}}-\mathbf{W}^{\mathbf{U}}$ is the graph Laplacian of the feature graph which reflects the label smoothness of the feature. The multiplicative updating rules minimizing Eq. (3) are given as [19].

$$\mathbf{S}=(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{X}\mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1},$$

$$v_{jk}\leftarrow v_{jk}\sqrt{\frac{[\lambda\mathbf{W}^V\mathbf{V}+\mathbf{A}^++\mathbf{V}\mathbf{B}^-]_{jk}}{[\lambda\mathbf{D}^V\mathbf{V}+\mathbf{A}^-+\mathbf{V}\mathbf{B}^+]_{jk}}},$$

$$u_{ik}\leftarrow u_{ik}\sqrt{\frac{[\mu\mathbf{W}^U\mathbf{U}+\mathbf{P}^++\mathbf{U}\mathbf{Q}^-]_{ik}}{[\mu\mathbf{D}^U\mathbf{U}+\mathbf{P}^-+\mathbf{U}\mathbf{Q}^+]_{ik}}} \tag{4}$$

where $\mathbf{A}=\mathbf{X}^T\mathbf{US}=\mathbf{A}^+-\mathbf{A}^-$, $\mathbf{B}=\mathbf{S}^T\mathbf{U}^T\mathbf{US}=\mathbf{B}^+-\mathbf{B}^-$, $\mathbf{P}=\mathbf{XVS}^T=\mathbf{P}^+-\mathbf{P}^-$ and $\mathbf{Q}=\mathbf{SV}^T\mathbf{VS}^T=\mathbf{Q}^+-\mathbf{Q}^-$, where $\mathbf{A}_{ij}^+=(|\mathbf{A}_{ij}|+\mathbf{A}_{ij})/2$, $\mathbf{A}_{ij}^-=(|\mathbf{A}_{ij}|-\mathbf{A}_{ij})/2$

Theorem 2. *[19] For $\mathbf{U},\mathbf{V}\geq 0$, the objective function $\mathbf{J}_{\mathbf{DRCC}}$ in Eq. (3) is non-increasing under each of the above updating rules stated in Eq. (4).*

Gu et al. [19] have proved that the iterative multiplicative updating scheme stated in Eq. (4) will find local minima of the objective function $\mathbf{J}_{\mathbf{DRCC}}$.

Download English Version:

https://daneshyari.com/en/article/406518

Download Persian Version:

https://daneshyari.com/article/406518

Daneshyari.com