Multi-Layer Concept Factorization Algorithm Based on Dual Graph Regularization (Postprint)

Authors: Zhang Xian, Ye Jun

Date: 2018-05-18T00:00:00+00:00

Abstract

Abstract: To further mine hidden information among data, within the framework of the Multi-layer Concept Factorization (MCF) algorithm and considering the data manifold and feature manifold under each layer of decomposition, we propose a Dual Graph regularized Multi-layer Concept Factorization (DGMCF) algorithm. This algorithm learns hierarchically through layer-by-layer data decomposition, constructs Laplacian graphs for both the data space and feature attribute space at each layer to reflect the multiple geometric structure information of the data manifold and feature manifold, thereby enabling more effective feature extraction from complex data. The objective function is solved using an alternating iterative method, and the convergence of the algorithm is proven. Experiments on three real-world databases (TDT2, PIE, COIL20) demonstrate that the proposed method outperforms other methods in terms of clustering representation effectiveness.

Full Text

Dual-Graph Regularized Multilayer Concept Factorization

Zhang Xian, Ye Jun
(School of Natural Sciences, Nanjing University of Posts & Telecommunications, Nanjing 210023, China)

Abstract: To further excavate hidden information among data, this paper proposes a novel dual-graph regularized multilayer concept factorization (DGMCF) algorithm within the framework of multilayer concept factorization (MCF). The algorithm performs learning in a hierarchical manner through layer-by-layer decomposition, constructing Laplacian graphs in both data space and feature attribute space at each decomposition layer to reflect the multivariate geometric structure information of the data manifold and feature manifold. This enables more effective feature extraction from complex data. An alternating iterative method is employed to optimize the objective function, and the convergence of the algorithm is proven. Experiments on three real-world databases (TDT2, PIE, COIL20) demonstrate that the proposed method outperforms other approaches in terms of clustering representation effectiveness.

Keywords: concept factorization; multilayer decomposition; dual regularization; manifold learning; clustering

0 Introduction

How to mine hidden information and effective data from massive high-dimensional data has become a research hotspot in machine learning, data mining, and social network analysis [1–3]. Appropriate data representation can uncover potential structures within data, facilitating further processing. Currently, matrix factorization methods have attracted considerable attention as an effective data processing approach. Commonly used matrix factorization algorithms include singular value decomposition (SVD), nonnegative matrix factorization (NMF) [4], and concept factorization (CF) [5].

Under the frameworks of NMF and CF, numerous extensions have been proposed. By incorporating label information, Liu et al. [6] proposed semi-supervised matrix factorization methods, namely constrained nonnegative matrix factorization (CNMF) and constrained concept factorization (CCF) [7]. Both methods embed label information as hard constraints within the NMF and CF frameworks. By integrating manifold learning principles [8–11] into NMF and CF, Cai et al. [12] proposed graph regularized nonnegative matrix factorization (GNMF), and Cai et al. [13] proposed locally consistent concept factorization (LCCF). Both methods have achieved promising results in text clustering and face recognition applications.

However, the aforementioned methods perform only single-layer decomposition of data. Existing literature [14,15] indicates that single-layer decomposition struggles to capture hierarchical information hidden in original data, thereby limiting data representation capability. Consequently, X. Li et al. [16] proposed multilayer nonnegative matrix factorization (MNMF), which can better separate original signals from mixed signals. Within the CF framework, Li et al. [16] proposed multilayer concept factorization (MCF), which can obtain hierarchical information hidden in high-dimensional data through a multilayer model. Considering geometric structure information of data, Li et al. [17] recently incorporated manifold learning into each layer of MCF, proposing graph regularized multilayer concept factorization (GMCF). GMCF not only utilizes multilayer structures but also considers the manifold geometric structure of each data layer, which indeed helps excavate the most essential information.

GNMF, LCCF, and GMCF methods all exploit spatial structure information of data, effectively improving learning quality. However, these works only consider the distribution structure of data space without utilizing structural information from the feature attribute space. Recently, Shang and Ye et al. [18,19] proposed graph dual regularization nonnegative matrix factorization (DNMF) and graph dual regularization concept factorization (GCF), respectively, which simultaneously consider the geometric structures of both data manifold and feature manifold, achieving excellent results. This further demonstrates that not only can geometric structure information of data space effectively improve learning quality, but geometric structure information of feature attribute space can also assist learning.

Therefore, to further mine hidden information among data while leveraging geometric structure information, this paper proposes a dual-graph regularized multilayer concept factorization (DGMCF) algorithm under the MCF framework, considering both data manifold and feature manifold at each decomposition layer.

1.1 Concept Factorization (CF) Algorithm

Given a nonnegative matrix $X \in \mathbb{R}^{M\times N}$ where each column represents a sample, the CF algorithm seeks two nonnegative matrices $W \in \mathbb{R}^{K\times M}$ and $V \in \mathbb{R}^{K\times N}$ such that $X \approx XWV$. The objective function of CF can be expressed as:

$$J_{CF}(W,V) = \|X - XWV\|_F^2 \quad \text{s.t.} \quad W, V \geq 0$$

where V represents the low-dimensional data representation to be learned. The corresponding multiplicative update rules have been detailed in reference [5].

1.2 Multilayer Concept Factorization (MCF) Algorithm

For complex data, particularly ill-posed data, Li et al. [16] adopted a hierarchical approach in concept factorization, performing concept factorization layer by layer, and proposed the multilayer concept factorization (MCF) algorithm. The MCF concept is: let $X_1 = X$ be the first layer decomposition, use X_1 as initial data for the second layer decomposition to obtain $X_2 = X_1W_1V_1$, and so on, obtaining the l-th layer concept decomposition $X_{l+1} = X_lW_lV_l$. The objective function of multilayer concept factorization can be expressed as:

$$J_{MCF}(W,V) = \sum_{l=1}^{L} \|X_l - X_lW_lV_l\|_F^2 \quad \text{s.t.} \quad W_l, V_l \geq 0$$

The corresponding multiplicative update rules have been detailed in reference [5].

1.3 Dual-Graph Regularized Concept Factorization (GCF) Algorithm

Ye et al. [19] proposed a dual-graph regularized concept factorization algorithm based on CF, simultaneously considering geometric structure information of both data and feature attributes. Its objective function can be expressed as:

$$J_{GCF}(W,V) = \|X - XWV\|_F^2 + \alpha \mathrm{Tr}(V^T L_v V) + \beta \mathrm{Tr}(W^T L_w W) \quad \text{s.t.} \quad W,V \geq 0$$

where $\alpha, \beta \geq 0$ are regularization parameters. The multiplicative update rules for equation (3) have been detailed in reference [16].

2 Dual-Graph Regularized Multilayer Concept Factorization (DGMCF)

Recent studies show that not only do observed data lie on a low-dimensional submanifold called the data manifold, but also data features lie on a low-dimensional submanifold called the feature manifold [18]. During concept factorization, since each layer of data decomposition involves obtaining geometric structure information of both data manifold and feature manifold, two graphs are used at each layer to characterize the geometric structures of data manifold and feature manifold, namely the data graph and feature graph.

2.1 Constructing Dual Graph Regularization Terms

Data Graph Construction at Layer l:
Let the vertex set of the graph at layer l be the dataset $\{x_1^{(l)}, \dots, x_{N_l}^{(l)}\}$. If the i-th sample and the j-th sample are mutual nearest neighbors, there exists an edge between them with weight $s_{ij}^{(l)}$, and the corresponding neighborhood matrix is $S_l^V \in \mathbb{R}^{N_l \times N_l}$. The Laplacian matrix of the data graph at layer l can be expressed as $L_l^V = D_l^V - S_l^V$, where D_l^V is a diagonal matrix with elements $d_{ii}^V = \sum_j s_{ij}^{(l)}$.

The smoothness of the representation can be expressed as:

$$\frac{1}{2} \sum_{i,j=1}^{N_l} s_{ij}^{(l)} \|v_i^{(l)} - v_j^{(l)}\|^2 = \mathrm{Tr}(V_l^T L_l^V V_l)$$

Feature Graph Construction at Layer l:
Similarly, at layer l, the vertex set of the graph is defined as the feature set

$\{x_1^{(l)T}, \ldots, x_{M_l}^{(l)T}\}$. The corresponding neighborhood matrix $S_l^W \in \mathbb{R}^{M_l \times M_l}$ is defined as:

$$s_{ij}^{(l)} = \begin{cases} 1 & \text{if } x_i^{(l)} \text{ and } x_j^{(l)} \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

The Laplacian matrix of the feature graph at layer l can be expressed as $L_l^W = D_l^W - S_l^W$, where D_l^W is a diagonal matrix. The smoothness of the representation is:

$$\frac{1}{2}\sum_{i,j=1}^{M_l} s_{ij}^{(l)} \|w_i^{(l)} - w_j^{(l)}\|^2 = \mathrm{Tr}(W_l^T L_l^W W_l)$$

2.2 DGMCF Objective Function

To simultaneously consider geometric structure information of both sample data manifold and feature manifold, we add regularization terms based on data graph and feature graph to the MCF objective function, obtaining the DGMCF objective function:

$$J_{DGMCF}(W,V) = \sum_{l=1}^{L} \left(\|X_l - X_l W_l V_l\|_F^2 + \alpha \mathrm{Tr}(V_l^T L_l^V V_l) + \beta \mathrm{Tr}(W_l^T L_l^W W_l) \right) \quad \text{s.t.} \quad W_l, V_l \geq 0$$

where $\alpha, \beta \geq 0$ are regularization parameters.

2.3 Solving the DGMCF Objective Function

The objective function J_{DGMCF} is a non-convex function of two variables W and V, making global optimization impractical. We employ an alternating iteration method to obtain a local optimum. Let $J_l = \|X_l - X_l W_l V_l\|_F^2 + \alpha \mathrm{Tr}(V_l^T L_l^V V_l) + \beta \mathrm{Tr}(W_l^T L_l^W W_l)$. The Lagrangian function of equation (9) is:

$$L(W_l, V_l) = J_l + \mathrm{Tr}(\Psi_l W_l^T) + \mathrm{Tr}(\Phi_l V_l^T)$$

where Ψ_l and Φ_l are Lagrange multipliers. Taking partial derivatives of the function with respect to W_l and V_l and applying KKT optimality conditions yields the DGMCF update formulas:

$$w_{jk}^{(l)} \leftarrow w_{jk}^{(l)} \frac{(X_l^T X_l V_l^T + \beta X_l^T X_l S_l^W W_l)_{jk}}{(X_l^T X_l W_l V_l V_l^T + \beta X_l^T X_l D_l^W W_l)_{jk}}$$

$$v_{jk}^{(l)} \leftarrow v_{jk}^{(l)} \frac{(X_l^T X_l W_l + \alpha S_l^V V_l)_{jk}}{(V_l W_l^T X_l^T X_l W_l + \alpha D_l^V V_l)_{jk}}$$

The DGMCF method stops iterating when the maximum number of iterations is reached or when the stopping criterion $\|J_{DGMCF}^{(t)} - J_{DGMCF}^{(t-1)}\| < \varepsilon$ is satisfied.

2.4 Convergence Proof of DGMCF Algorithm

Definition 1 A function $G(w, w')$ is an auxiliary function for $F(w)$ if it satisfies: 1. $G(w, w') \geq F(w)$ for any w 2. $G(w, w) = F(w)$

Lemma 1 If G is an auxiliary function for F, then F is non-increasing under the update:

$$w^{(t+1)} = \arg\min_{w} G(w, w^{(t)})$$

Lemma 2 The function $G(w_{ab}, w_{ab}^{(t)})$ defined below is an auxiliary function for $F_{ab}(w)$:

$$G(w_{ab}, w_{ab}^{(t)}) = F_{ab}(w_{ab}^{(t)}) + F'_{ab}(w_{ab}^{(t)})(w_{ab} - w_{ab}^{(t)}) + \frac{(X_l^T X_l W_l V_l V_l^T + \beta X_l^T X_l D_l^W W_l)_{ab}}{w_{ab}^{(t)}}(w_{ab} - w_{ab}^{(t)})^2$$

Proof: By Definition 1, we need to show $G(w_{ab}, w_{ab}^{(t)}) \geq F_{ab}(w_{ab})$. The Taylor expansion of $F_{ab}(w_{ab})$ is:

$$F_{ab}(w_{ab}) = F_{ab}(w_{ab}^{(t)}) + F'_{ab}(w_{ab}^{(t)})(w_{ab} - w_{ab}^{(t)}) + \frac{1}{2}F''_{ab}(w_{ab}^{(t)})(w_{ab} - w_{ab}^{(t)})^2$$

Thus, proving $G(w_{ab}, w_{ab}^{(t)}) \geq F_{ab}(w_{ab})$ is equivalent to proving:

$$\frac{(X_l^T X_l W_l V_l V_l^T + \beta X_l^T X_l D_l^W W_l)_{ab}}{w_{ab}^{(t)}} \geq \frac{1}{2}F''_{ab}(w_{ab}^{(t)})$$

Since $F''_{ab}(w_{ab}^{(t)}) = 2(X_l^T X_l)_{aa}(V_l V_l^T)_{bb} + 2\beta(X_l^T X_l)_{aa}(D_l^W)_{bb}$, and using the inequality $\frac{w_{ab}^{(t)}}{w_{ab}^{(t)}} \geq \frac{w_{ab}^{(t)}(X_l^T X_l W_l V_l V_l^T + \beta X_l^T X_l D_l^W W_l)_{ab}}{(X_l^T X_l W_l V_l V_l^T + \beta X_l^T X_l D_l^W W_l)_{ab}}$, we can show that equation (15) holds, i.e., $G(w_{ab}, w_{ab}^{(t)}) \geq F_{ab}(w_{ab})$.

Lemma 3 The function $G(v_{ab}, v_{ab}^{(t)})$ is an auxiliary function for $F_{ab}(v)$. (The proof follows similarly to Lemma 2 and is omitted due to space limitations.)

Theorem 1 For given data X and arbitrary $W_l, V_l \geq 0$, the alternating iterative update rules (11) and (12) monotonically decrease the objective function $J_{DGMCF}(W, V)$.

Proof: From Lemmas 2 and 3, substituting equations (15) and (17) into equation (13) yields:

$$J_{DGMCF}(W^{(t+1)}, V^{(t+1)}) \leq J_{DGMCF}(W^{(t)}, V^{(t)})$$

Since equations (15) and (17) are auxiliary functions for $F_{ab}(w)$ and $F_{ab}(v)$ respectively, the function values are monotonically decreasing under iterative updates (11) and (12).

2.5 DGMCF Algorithm Steps

Input: Dataset matrix X, number of layers L, decomposition dimension K, regularization parameters α, β, nearest neighbor parameter p, maximum iterations $IterMax$, and error limit ε.

Procedure: 1. Initialize parameters: Set $X_1 = X$, randomly generate nonnegative matrices W_l and V_l for each layer. 2. For $l = 1$ to L: a. Construct data graph and feature graph adjacency matrices S_l^V and S_l^W using p nearest neighbors. b. Compute diagonal matrices D_l^V and D_l^W, then calculate Laplacians L_l^V and L_l^W. c. For $t = 1$ to $IterMax$: - Fix V_l, update W_l using equation (10) to obtain $W_l^{(t+1)}$. - Fix W_l, update V_l using equation (11) to obtain $V_l^{(t+1)}$. - Repeat until $\|J_{DGMCF}^{(t)} - J_{DGMCF}^{(t-1)}\| < \varepsilon$. d. Set $X_{l+1} = X_l W_l V_l$. 3. End for

Output: Factorized matrices $\{W_l, V_l\}_{l=1}^{L}$.

2.6 Complexity Analysis of DGMCF Algorithm

To compare computational complexity with other algorithms, we analyze CF, GCF, GMCF, and DGMCF. After t iterations:

- CF and GCF have complexity $O(tNKM + NKM + NK^2p + MK^2p)$.
- GMCF and DGMCF both perform L-layer decomposition. GMCF's graph construction costs $O(LNK^2p)$, while DGMCF' s dual graph construction costs $O(LNK^2p + LMK^2p)$.
- Total complexities: GMCF is $O(LtNKM + LNK^2p + LMK^2p)$, DGMCF is $O(LtNKM + LNK^2p + LMK^2p)$.

3.1 Clustering Experiments

To verify the effectiveness of the proposed algorithm, we conducted clustering comparison experiments on the text database TDT2 and image databases PIE and COIL20, comparing DGMCF with CF, LCCF, MCF, GMCF, and GCF.

Datasets: - TDT2 text database: Contains 10,021 documents from 56 categories. - **PIE face database:** Selected 11,554 images of 68 individuals (164 images for person 38, 170 images for others). Images are 32×32 grayscale. - **COIL20 object image database:** Contains 1,440 images of 20 objects, each 32×32 grayscale.

Experimental Setup: For MCF, GMCF, and DGMCF, we set $L = 3$ and maximum iterations per layer to 500. For LCCF, GMCF, GCF, and DGMCF, the number of nearest neighbors p is set to 5. For different cluster numbers

$k \in \{2, \ldots, 10\}$, the average clustering results over 20 runs are shown in Tables 1–3.

Results: From Tables –, we observe: a) On TDT2, DGMCF improves ACC and NMI by 11.23% and 11.67% over CF, 1.94% and 2.65% over GCF, and 2.28% and 2.33% over GMCF. On PIE, improvements are 13.70% and 12.18% over CF, 3% and 2.45% over GCF, and 3.49% and 2.90% over GMCF. On COIL20, improvements are 10.93% and 13.18% over CF, 3.36% and 4.02% over GCF, and 4.02% and 7.17% over GMCF.

b) LCCF utilizes geometric structure information of data distribution and achieves better clustering than CF, demonstrating the effectiveness of data geometry, especially for image datasets with underlying manifold structures. GCF outperforms LCCF by exploiting both data and feature manifold structures. GMCF surpasses CF and LCCF through hierarchical decomposition with data manifold information at each layer.

c) DGMCF achieves the best clustering quality by leveraging multilayer decomposition with both data and feature manifold structures at each layer, particularly effective for image datasets.

3.2 Parameter Selection

DGMCF has four main parameters: nearest neighbors p, regularization parameters α and β, and number of layers L. The discussion of L is detailed in [17]; we set $L = 3$. Here we examine parameter stability.

When analyzing α and β, we set $p = 5$. When analyzing p, we set $\alpha = \beta = 100$. For comparison, LCCF, GMCF, and GCF use the same parameter values. Experiments on TDT2, PIE, and COIL20 are shown in Figures [Figure 1: see original paper] and [Figure 2: see original paper].

Observations: a) DGMCF is very stable for α and β in the range [1, 1000], consistently achieving good clustering results and outperforming other algorithms.

b) Clustering accuracy decreases when p is too large, as oversized neighborhoods fail to accurately reflect the intrinsic geometric structure.

---

4 Conclusion

This paper proposes a dual-graph regularized multilayer concept factorization algorithm that leverages multilayer decomposition to simultaneously exploit geometric structure information from both data and feature manifolds at each layer, enabling deeper mining of the most essential data features. We derive the objective function, iterative update rules, and provide convergence proof.

Extensive experiments demonstrate the algorithm' s superior performance in data representation compared to existing methods.

References

[1] Yi Yugen, Shi Yanjiao, Zhang Huijie, et al. Label propagation based semi-supervised nonnegative matrix factorization for feature extraction [J]. Neurocomputing, 2015, 149: 1021-1037.

[2] Wang Mengmeng, Zuo Wanli, Wang Ying, et al. A weighted nonnegative matrix factorization based algorithm for multi-dimensional user personality trait recognition [J]. Chinese Journal of Computers, 2016, 39(12): 2562-2577.

[3] He Chaobo, Tang Yong, Yang Asong, et al. Large-scale topic community mining based on distributed nonnegative matrix factorization [J]. Scientia Sinica Informationis, 2016, 46(6): 714-728.

[4] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization [J]. Nature, 1999, 401: 788-791.

[5] Xu Wei, Gong Yihong. Document clustering by concept factorization [C]//Proc of International Conference on Research and Development in Information Retrieval. 2004: 202-209.

[6] Liu Haifeng, Wu Zhaohui, Li Xuelong, et al. Constrained nonnegative matrix factorization for image representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1299-1311.

[7] Liu Haifeng, Yang Genmao, Wu Zhaohui, et al. Constrained concept factorization for image representation [J]. IEEE Trans on Cybernetics, 2014, 44(7): 1214-1224.

[8] Tenenbaum J B, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(5500): 2319–2323.

[9] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323-2326.

[10] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering [C]//Advances in Neural Information Processing Systems. 2002: 585-591.

[11] Guan Naiyang, Tao Dacheng, Luo Zhigang, et al. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent [J]. IEEE Trans on Image Processing, 2011, 20(7): 2030-2048.

[12] Cai Deng, He Xiaofei, Han Jiawei, et al. Graph regularized nonnegative matrix factorization for data representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1548-1560.

[13] Cai Deng, He Xiaofei, Han Jiawei. Locally consistent concept factorization for document clustering [J]. IEEE Trans on Knowledge and Data Engineering,

2011, 23(6): 902-913.

[14] Cichocki A, Zdunek R. Multilayer NMF using projected gradient approaches [J]. International Journal of Neural Systems, 2007, 17(6): 431-446.

[15] Rajabi R, Ghassemian H. Spectral unmixing of hyperspectral imagery using multilayer NMF [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(1): 38-42.

[16] Li Xue, Zhao C, Shu Z, et al. Multilayer concept factorization for data representation [C]//Proc of the 10th International Computer Science & Education. 2015: 486-491.

[17] Li Xue, Shen Xiaobo, Shu Zhenqiu, et al. Graph regularized multilayer concept factorization for data representation [J]. Neurocomputing, 2017, 238: 139-151.

[18] Shang Fanhua, Jiao Lichen, Wang Fei. Graph dual regularization nonnegative matrix factorization for co-clustering [J]. Pattern Recognition, 2012, 45(6): 2237-2250.

[19] Ye Jun, Jin Zhong. Dual-graph regularized concept factorization for clustering [J]. Neurocomputing, 2014, 138(3): 120-130.

Note: Figure translations are in progress. See original paper for figures.