Concept Factorization with Optimal Graph Learning for Data Representation

Zhenqiu Shu[1,2], Xiao-jun Wu[2(✉)], Honghui Fan[1], Congzhe You[1], Zhen Liu[2], and Jie Zhang[1]

[1] School of Computer Engineering, Jiangsu University of Technology,
Changzhou 231001, China
shuzhenqiu@163.com

[2] Jiangsu Provincial Engineering Laboratory of Pattern Recognition and
Computational Intelligence, Jiangnan University, Wuxi 214122, China
wu_xiaojun@jiangnan.edu.cn

Abstract. In recent years, concept factorization methods become a popular data representation technique in many real applications. However, conventional concept factorization methods cannot capture the intrinsic geometric structure embedded in data using the fixed nearest neighbor graph. To overcome this problem, we propose a novel method, called Concept Factorization with Optimal Graph Learning (CF_OGL), for data representation. In CF_OGL, a novel rank constraint is imposed on the Laplacian matrix of the initial graph model, which encourages the learned graph with exactly c connected components for the data with c clusters. Then the learned optimal graph regularizer is integrated into the model of concept factorization. Therefore, this learned structure is benefit to the clustering analysis. In addition, we develop an efficient and effective iterative optimization algorithm to solve our proposed model. Extensive experimental results on three benchmark datasets have demonstrated that our proposed method can effectively improve the performance of clustering.

Keywords: Concept factorization · Data representation · Geometric structure · Rank constraint · Laplacian matrix · Regularizer

1 Introduction

Data representation methods have been widely applied to various fields in pattern recognition and machine learning [1–3]. Over the past few decades, matrix factorization methods have become one of the most popular data representation

techniques due to its efficiency and effectiveness. Many classical matrix factorization techniques, such as Singular Value Decomposition (SVD) [4], Principal Component Analysis (PCA) [5], Nonnegative Matrix Factorization (NMF) [6] and Concept Factorization (CF) [7], have shown the encouraging performances in image classification, object tracking, document clustering, etc. [8,9].

NMF has aroused increasing interests due to its physical and theoretical interpretations. NMF naturally leads to a part-based representation of data by imposing the nonnegative constraint on both coefficient and basis matrices. The basic idea behind NMF is to seek two nonnegative matrices to approximate the original data matrix. However, NMF cannot deal with the data matrix containing with some negative elements due to the noise or outlier. Therefore, Xu et al. [7] proposed a variation of NMF, called Concept Factorization (CF), for document clustering. Different from the NMF methods, CF can deal with the data matrix mixed with nonnegative elements. In order to discover the local geometric structure of data, Cai et al. [10] proposed a Locally Consistent Concept Factorization (LCCF) method for data representation. It models the manifold structure of data using the graph regularizer. Shu et al. [11] proposed a Local Learning Concept Factorization (LLCF) method to learn the discriminant structure and the local geometric structure, simultaneously, by adding the local learning regularization term into the model of CF. Motivated by the deep learning, Li et al. [12] proposed a multilayer concept factorization method to discover the structure information hidden in data using the multilayer framework. Pei et al. [13] developed a CF with adaptive neighbors method for clustering. The idea of this proposed method is to integrate an ANs regularizer into the CF decomposition. However, the aforementioned methods cannot update dynamically the optimal graph model, which is used to explore the intrinsic geometric manifold structure of data in matrix decomposition.

To solve this issue, we propose a novel method named as Concept Factorization with Optimal Graph Learning (CF_OGL) in this paper. Specifically, we impose a rank constraint on the Laplacian matrix of the initially given graph, and then iteratively update it. Therefore, the learned graph has exactly c connected components, whose structure is beneficial to the clustering applications. Then the learned graph regularizer is used to constrain the model of the concept factorization method, and thus the geometric structure of data can be better preserved in low dimensional feature space. Extensive experimental results on three datasets demonstrate that our proposed CF_OGL method outperforms other state-of-the-art methods in clustering.

This paper is organized as follows: We briefly describe both CF and LCCF algorithms in Sect. 2. In Sect. 3, we introduce our proposed CF_OGL algorithm and then derive its updating rules. In Sect. 4, we carry out some experiments to investigate the proposed CF_OGL algorithm. Finally, conclusions are drawn in Sect. 5.

2   The Relative Work

In this section, the models of both CF and LCCF are briefly presented.

2.1   CF

Concept factorization is a popular matrix factorization technique to deal with high dimensional data. Given a data matrix $X = [x_1, x_2, ..., x_n] \in R^{m \times n}$, x_i denotes a m-dimensional vector. In CF, the entire data points are used to lineally represent each underlying concept, and all the concepts seek to lineally approximate to each data point, simultaneously. Therefore, we can give the objective function of CF as

$$X = XUV^T \tag{1}$$

where $U \in R^{n \times k}$ and $V \in R^{n \times k}$. Using the Euclidean distance metric to measure the reconstruction error, its minimization problem can be given as follows:

$$\min_{U,V} \left\| X - XUV^T \right\|_F^2 \tag{2}$$
$$s.t. U \geq 0, V \geq 0$$

where $\|\cdot\|_F$ denotes the matrix $Frobenius\ norm$. Using the multiplicative updating algorithm, we derive the updating rules of Eq. (2) as follows:

$$u_{ij}^{t+1} \leftarrow u_{ij}^t \frac{(KV)_{ij}}{(KUV^TV)_{ij}}$$
$$v_{ij}^{t+1} \leftarrow v_{ij}^t \frac{(KW)_{ij}}{(VW^TKW)_{ij}} \tag{3}$$

where $K = X^T X$. To deal with the nonlinear data, CF is easily kernelized using kernel trick.

2.2   LCCF

Traditional CF method fails to consider the manifold structure information of data. To solve this issue, Cai et al. [10] proposed the LCCF method, which models the manifold structure embedded in data using the fixed graph model. Therefore, the objective function of LCCF can be given as follow:

$$\min_{U,V} \left\| X - XUV^T \right\|_F^2 + \lambda tr(V^TLV) \tag{4}$$
$$s.t. U \geq 0, V \geq 0$$

where λ stands for a balance parameter, and $tr(.)$ denotes the trace of a matrix. D is a diagonal matrix, $D_{ii} = \sum_S W_{ij}$, $L = D - W$. Similarly, we derive the updating rules of Eq. (4) as follows:

$$u_{ij}^{t+1} \leftarrow u_{ij}^t \frac{(KV)_{ij}}{(KUV^TV)_{ij}}$$
$$v_{ij}^{t+1} \leftarrow v_{ij}^t \frac{(KW+\lambda WV)_{ij}}{(VW^TKW+\lambda DV)_{ij}} \tag{5}$$

According to the rules (5), we can achieve a local minimum of Eq. (4).

3   The Proposed Method

3.1   Motivation

Traditional CF methods cannot effectively explore the intrinsic geometric manifold structure embedded in high dimensional data using the fixed graph model. By addint the rank constraint into the Laplacian matrix of the initially given graph, we learn an optimal graph model with exactly c connected components. In CF_AGL, the learned graph regularizer is further constructed, and then imposed on the model of CF. Therefore, our proposed method explores the semantic information hidden in high dimensional data effectively.

3.2   Constrained Laplacian Rank (CLR)

A graph learning method, called Constrained Laplacian Rank (CLR), was proposed to explore the intrinsic geometric structure of data, whose goal is to learn an optimal graph model [14]. Therefore, the CLR method is formulated by the following optimization problem:

$$J_{CLR} = \min_{\sum_j a_{ij}=1, a_{ij}\geq 0, rank(L_Q)=n-k} \|Q - W\|_F^2 \tag{6}$$

where L_Q stands for the Laplacian matrix of the matrix Q. Denote $\sigma_i(L_Q)$ as the i-th smallest eigenvalue of L_A. It is worth noting that $\sigma_i(L_Q) > 0$ because of its positive semidefinition. Therefore, Eq. (6) can be reformulated as the following problem for a large enough value of σ_i:

$$J_{CLR} = \min_{\sum_j q_{ij}=1, q_{ij}\geq 0} \|Q - W\|_F^2 + 2\lambda \sum_{i=1}^{k} \sigma_i(L_Q) \tag{7}$$

According to the Ky Fans Theorem, we have the following equivalent definition as

$$\sum_{i=1}^{k} \sigma_i(L_Q) = \min_{F \in n \times k, F^T F=I} Tr(F^T L_Q F) \tag{8}$$

Therefore, we can further rewrite the problem (7) as follows:

$$\min_{Q} \|Q - W\|_F^2 + \lambda Tr(F L_Q F^T)$$
$$s.t. FF^T = I, Q1 = 1, Q \geq 0, Q \in R^{n \times n} \tag{9}$$

3.3   Our Proposed Method

By integrating the learned graph regularization term into the model of CF, the objective function of our proposed CF_OGL method can given as follows:

$$\min_{Q,U,V} (\|Q - W\|_F^2 + \beta Tr(F L_Q F^T) + \lambda Tr(V L_Q V^T)$$
$$+ \mu \|X - XUV^T\|)$$
$$s.t. FF^T = I, V \geq 0, U \geq 0, Q1 = 1, Q \geq 0, Q \in R^{n \times n} \tag{10}$$

It is impractical to find the global optimal solution of problem (10) because it is not a convex problem in U, V and A together. Fortunately, we can achieve a local solution by optimizing the variables alternatively. Therefore, the optimization scheme of our proposed CF_OGL method mainly consists of two parts:

Fixing Q and F, Update U and V. By fixing the variables A and F, the Eq. (10) can be rewritten as the following problem:

$$\min_{U,V}(\lambda Tr(VL_QV^T) + \mu \|X - XUV^T\|)$$
$$s.t.VV^T = I, Q \geq 0, U \geq 0 \tag{11}$$

Similarly, it is easy to derive the updating rules of problem (10) as follows:

$$u_{ij}^{t+1} \leftarrow u_{ij}^t \frac{(KV)_{ij}}{(KUV^TV)_{ij}} \tag{12}$$

$$v_{ij}^{t+1} \leftarrow v_{ij}^t \frac{(KW+\lambda WV)_{ij}}{(VW^TKW+\lambda DV)_{ij}} \tag{13}$$

Fixing U and V, Update Q and F. By fixing U and V, we can rewrite the Eq. (10) as the following optimization problem:

$$\min_{Q,U,V}(\|Q - W\|_F^2 + \beta Tr(FL_QF^T))$$
$$s.t.FF^T = I, Q1 = 1, Q \geq 0, Q \in R^{n \times n} \tag{14}$$

(A) When Q is fixed, the Eq. (14) becomes

$$\min_{FF^T=I}(\beta Tr(FL_QF^T)) \tag{15}$$

It is easy to know that the solution scheme of F can be converted into solving the k smallest eigenvalues problem of L_Q.

(B) When F is fixed, the problem (14) becomes the following optimization problem:

$$\min \sum_{i,j=1} (q_{i,j} - w_{i,j})^2 + \frac{\beta}{2} \sum_{i,j=1} \|f_i - f_j\|_2^2 q_{i,j}$$
$$s.t. \sum_j q_{ij} = 1, q_{ij} \geq 0 \tag{16}$$

For each row w_i, we have the vector form as

$$\min_{q_i \geq 0, q_i 1=1} \left\| q_i - (w_i - \frac{\beta}{2}d_i^T) \right\|_2^2 \tag{17}$$

where $d_{ij} = \|f_i - f_j\|_2^2$. The problem (17) can be solved by the optimization algorithm proposed in [14].

4   Experimental Results

In this section, we carry out some experiments to investigate the proposed CF_OGL method on the Yale, ORL and FERET datasets. To demonstrate its effectiveness, the proposed CF_OGL method is compared with several state-of-the-art methods, such as K-means, PCA, NMF, CF and LCCF. Two well accepted measurements, such as accuracy (AC) and normalized mutual information (NMI), are used as metrics to quantify the performance of data representation in clustering.

Algorithm 1. Summary of the proposed CF_OGL method

Input: Initial graph W, parameter β, λ, μ, Data matrix X.
Output: The coefficient matrix V;
Initialization: $W \leftarrow W_0$, $Q \leftarrow Q_0$, $L_Q = Laplace(Q)$
Repeat
Fixing Q and F:
  Update U by Eq.(12);
  Update V by Eq.(13);
Fixing U and V:
  Update F by Eq.(15);
  Update Q by Eq.(17);
  Compute L_Q by $L_Q = D_Q - \frac{Q+Q^T}{2}$;
Until convergence

4.1   Yale Face Dataset

The Yale face dataset includes a total of 165 face images from 15 individuals. In each experiment, the P categories images were randomly sampled from the Yale dataset to evaluate the performances of all methods. We run all methods ten times for each value of P, and recorded their average results. The results of all methods on the Yale dataset are shown in Table 1. We can clearly see that our proposed CF_OGL method outperforms other state-of-the-art methods regardless of the choices of P. Specifically, the average AC and NMI of the proposed CF_OGL method are 3.7% and 5.1% higher than those of LCCF, respectively. The main reason is that our propose CF_OGL method can learn an optimal graph structure, which can significantly improve the clustering performance than LCCF.

4.2   ORL Face Dataset

The ORL face dataset contrains 400 face images from 40 distinct subjects. For some subjects, the face images were taken at different times, varying the lighting and facial expressions. In this experiment, we adopted the above similar experimental scheme to investigate the effectiveness of our proposed CF_OGL method. Table 2 provides the clustering results of six methods on the ORL face dataset. It can be observed that CF_OGL can achieve the best performance among all the compared methods. The main reason is that our proposed CF_OGL method can learn the optimal graph and thus effectively preserve the intrinsic geometric structure of data. Therefore, it outperforms other state-of-the-art methods on this dataset (Figs. 1 and 2).

Fig. 1. Some samples from the Yale dataset

Table 1. The clustering performances on the Yale face dataset

P	AC						NMI					
	K-means	PCA	NMF	CF	LCCF	CF_OGL	K-means	PCA	NMF	CF	LCCF	CF_OGL
8	0.536	0.520	0.511	0.545	0.556	**0.575**	0.602	0.590	0.589	0.561	0.592	**0.628**
9	0.480	0.458	0.477	0.438	0.454	**0.505**	0.560	0.540	0.558	0.516	0.542	**0.575**
10	0.489	0.469	0.465	0.490	0.460	**0.514**	0.589	0.575	0.577	0.551	0.533	**0.604**
11	0.462	0.480	0.461	0.480	0.480	**0.497**	0.587	0.596	0.577	0.559	0.569	**0.604**
12	0.440	0.451	0.431	0.437	0.428	**0.456**	0.564	0.566	0.564	0.530	0.536	**0.575**
13	0.439	0.427	0.423	0.400	0.422	**0.449**	0.573	0.566	0.564	0.513	0.534	**0.585**
14	0.448	0.454	0.422	0.432	0.431	**0.474**	0.589	0.602	0.581	0.542	0.546	**0.610**
15	0.427	0.423	0.408	0.396	0.386	**0.429**	0.583	0.582	0.568	0.519	0.518	**0.596**
AVG	0.468	0.461	0.451	0.456	0.453	**0.490**	0.581	0.577	0.572	0.536	0.546	**0.597**

Fig. 2. Some samples from the ORL dataset

Table 2. The clustering performances on the ORL face dataset

P	AC						NMI					
	K-means	PCA	NMF	CF	LCCF	CF_OGL	K-means	PCA	NMF	CF	LCCF	CF_OGL
20	0.386	0.384	0.373	0.244	0.377	**0.394**	0.582	0.578	0.573	0.367	0.571	**0.584**
22	0.375	0.382	0.364	0.230	0.370	**0.391**	0.596	0.592	0.575	0.371	0.591	**0.596**
24	0.368	0.374	0.357	0.239	0.373	**0.406**	0.592	0.598	0.579	0.388	0.585	**0.613**
26	0.356	0.364	0.342	0.228	0.344	**0.374**	0.585	0.595	0.569	0.384	0.577	**0.596**
28	0.341	0.350	0.328	0.209	0.339	**0.360**	0.586	0.588	0.568	0.376	0.575	**0.591**
30	0.334	0.324	0.332	0.206	0.328	**0.355**	0.587	0.582	0.580	0.382	0.578	**0.589**
32	0.326	0.339	0.330	0.200	0.338	**0.361**	0.587	0.582	0.580	0.382	0.578	**0.589**
34	0.332	0.332	0.334	0.206	0.330	**0.345**	0.588	0.597	0.578	0.385	0.589	**0.607**
36	0.328	0.324	0.312	0.204	0.336	**0.345**	**0.600**	0.595	0.585	0.392	0.584	0.597
38	0.320	0.315	0.315	0.185	0.320	**0.322**	0.596	0.600	0.583	0.383	0.587	**0.604**
40	0.3075	0.315	0.302	0.192	0.315	**0.318**	0.589	0.597	0.581	0.400	0.5953	**0.598**
AVG	0.343	0.346	0.335	0.213	0.343	**0.361**	0.591	0.593	0.577	0.385	0.585	**0.598**

4.3   FERET Face Dataset

The FERET face database contains 200 different individuals with about 7 face samples for each individual. Here, we randomly chose P categories samples from the FERET dataset, and mixed them as the experimental subset for clustering. All methods were run ten times, and then their average performances were recorded as the final results. The clustering performances for each method on the FERET dataset are summarized in Table 3. It is easy to find that the average performance of our proposed CF_OGL method has certain advantage compared with other methods in clustering (Fig. 3).
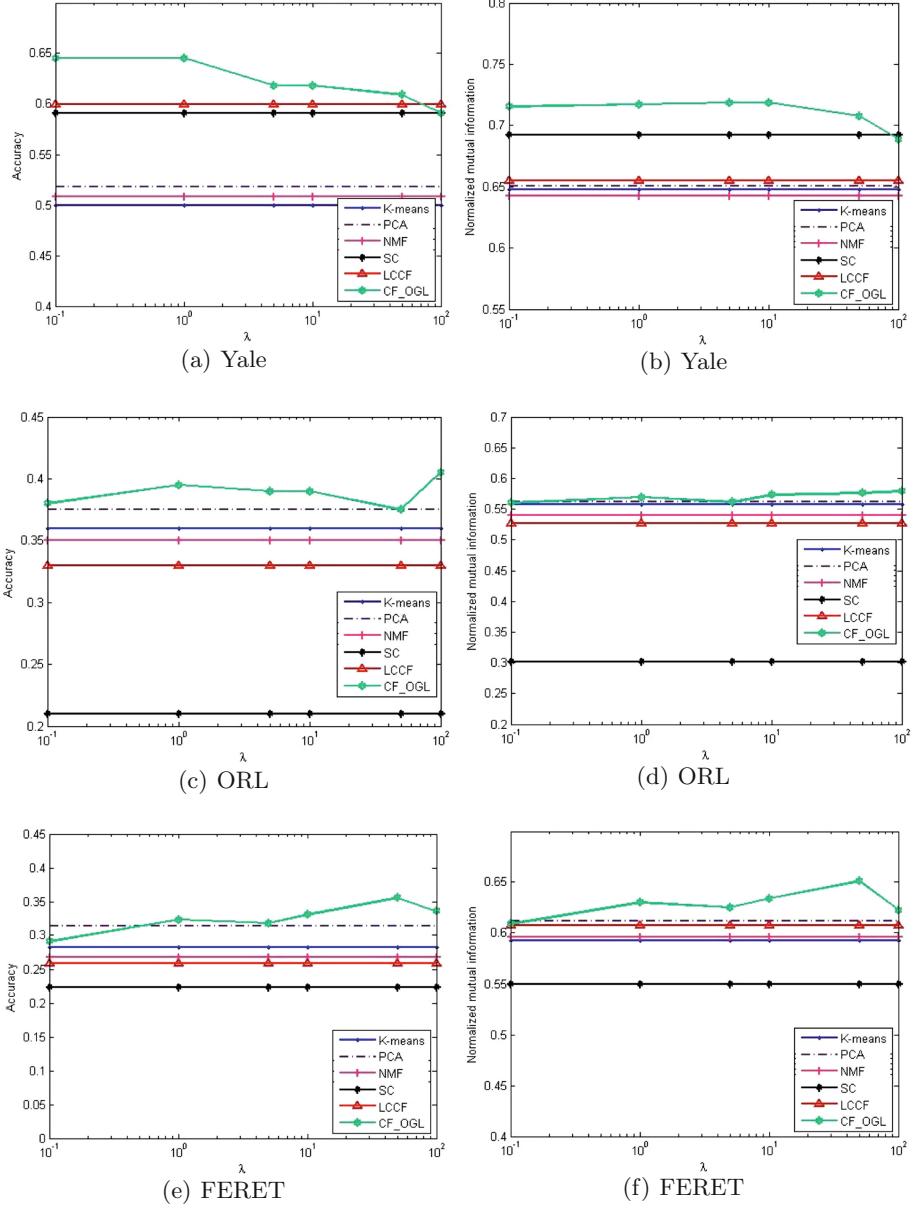
Fig. 3. Some samples from the FERET dataset

(a) Yale

(b) Yale

(c) ORL

(d) ORL

(e) FERET

(f) FERET

Fig. 4. Performances of all methods versus different vaules of the parameter λ

(a) Yale

(b) Yale

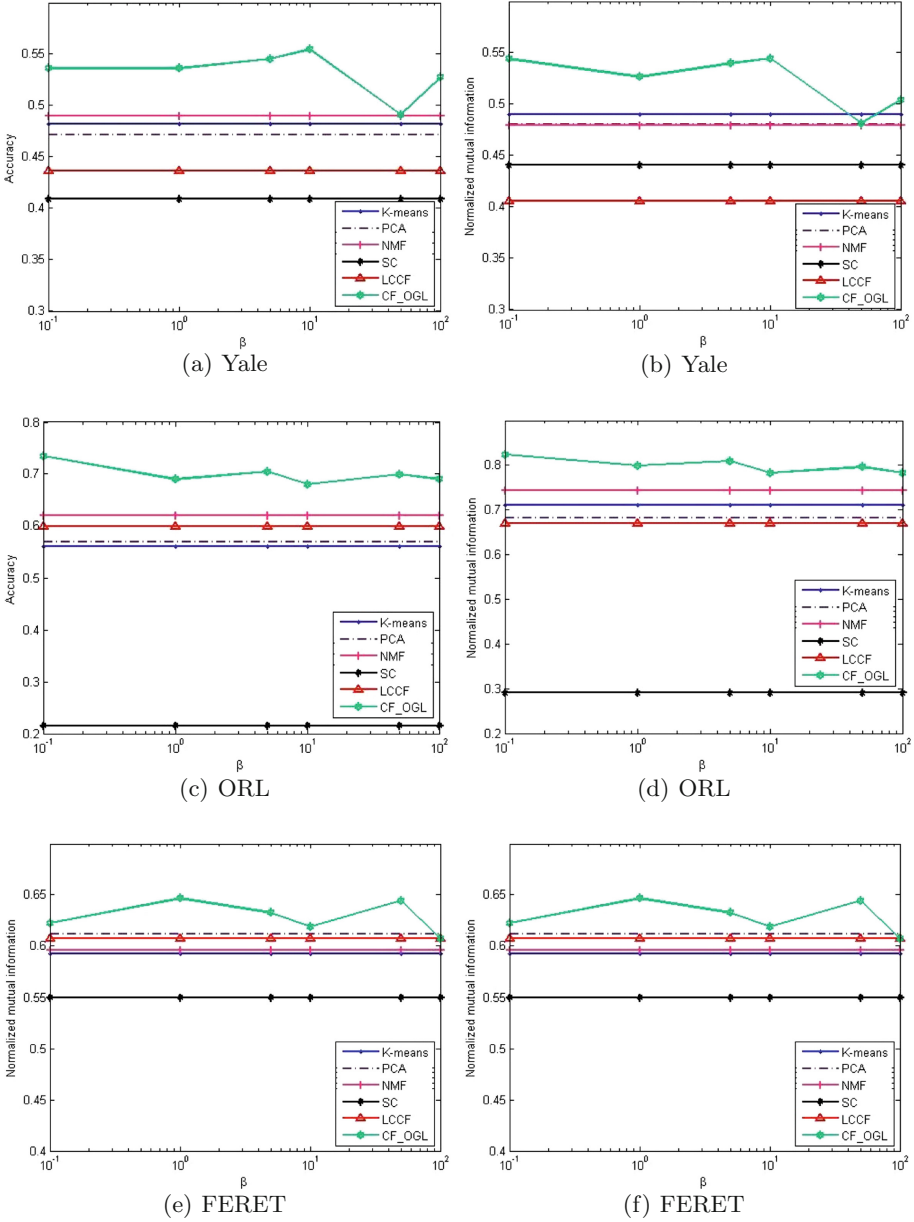(c) ORL

(d) ORL

(e) FERET

(f) FERET

Fig. 5. Performances of all methods versus different vaules of the parameter β

Table 3. The clustering performances on the FERET face dataset

P	AC						NMI					
	K-means	PCA	NMF	CF	LCCF	CF_OGL	K-means	PCA	NMF	CF	LCCF	CF_OGL
80	0.283	0.283	0.270	0.205	0.256	**0.315**	0.596	0.600	0.594	0.527	0.601	**0.617**
100	0.273	0.274	0.255	0.200	0.248	**0.298**	0.606	0.611	0.596	0.546	0.613	**0.622**
120	0.271	0.274	0.253	0.187	0.239	**0.283**	0.617	0.619	0.604	0.543	0.618	**0.629**
140	0.260	0.269	0.243	0.184	0.232	**0.276**	0.619	0.633	0.611	0.545	0.631	**0.642**
160	0.258	0.262	0.233	0.175	0.228	**0.272**	0.627	0.638	0.613	0.537	0.632	**0.643**
180	0.249	0.253	0.231	0.173	0.223	**0.254**	0.632	**0.642**	0.622	0.546	0.639	0.641
200	0.246	**0.257**	0.228	0.171	0.220	0.251	0.634	**0.649**	0.622	0.549	0.645	0.644
AVG	0.263	0.267	0.245	0.185	0.235	**0.278**	0.619	0.627	0.609	0.542	0.626	**0.632**

4.4   The Analysis of the Parameters

In our proposed CF_OGL, the parameters β, λ and μ have an effect on the clustering performance. Specifically, we randomly chose 10, 20, 80 categories samples as the dataset to carry out the experiments. However, the parameter selection of the proposed CF_OGL method is still an open problem. Therefore, we determine the parameters by grid search at first and then change them within certain ranges. Here, we only investigate the parameters β and λ. The Figs. 4 and 5 show the performances of all methods varied with different values of β and λ on three datasets, respectively. It is clear to see that our proposed method can achieve a relative stable performance in a large range.

5   Conclusion

In this paper, a novel matrix factorization technique, called Concept Factorization with adaptive graph learning (CF_OGL), is proposed for data representation. In order to learn an optimal graph, we impose a rank constraint on the Laplacian matrix of the initially given graph. Then the learned graph regularizer is integrated into the model of CF. Therefore, our proposed CF_OGL method effectively exploits the geometric manifold structure embedded in high dimensional data. Experimental results have shown that the proposed CF_OGL algorithm achieves better performance in comparison with other state-of-the-art algorithms.

References

1. Wu, X., Josef, K., Yang, J., et al.: A new direct LDA (D-LDA) algorithm for feature extraction in face recognition. In: International Conference on Pattern Recognition, vol. 4, pp. 545–548. IEEE Xplore (2004)
2. Shu, Z., Wu, X., Huang, P., et al.: Multiple graph regularized concept factorization with adaptive weights. IEEE Access **6**, 64938–64945 (2018)
3. Zheng, Y., Yang, J., Yang, J., et al.: Nearest neighbour line nonparametric discriminant analysis for feature extraction. Electron. Lett. **42**(12), 679–680 (2006)

4. Zhou, B., Chen, J.: A geometric distortion resilient image watermarking algorithm based on SVD. Chin. J. Image Graph. **9**, 506–512 (2004)
5. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
6. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999)
7. Xu, W., Gong, Y.: Document clustering by concept factorization. In: Proceedings of ACM SIGIR, pp. 202–209 (2004)
8. Shu, Z., Wu, X.J., Hu, C., et al.: Structured discriminative concept factorization for data representation. In: The Third IEEE International Smart Cities Conference (ISC2), pp. 1–4 (2017)
9. Shu, Z., Wu, X., Hu, C.: Structure preserving sparse coding for data representation. Neural Process. Lett. **48**, 1–15 (2018)
10. Cai, D., He, X., Han, J.: Locally consistent concept factorization for document clustering. IEEE Trans. Knowl. Data Eng. **23**(6), 902–913 (2011)
11. Shu, Z., Zhou, J., Huang, P., et al.: Local and global regularized sparse coding for data representation. Neurocomputing **198**(29), 188–197 (2016)
12. Li, X., Shen, X., Shu, Z., Ye, Q., Zhao, C.: Graph regularized multilayer concept factorization for data representation. Neurocomputing **238**, 139–151 (2017)
13. Pei, X., Chen, C., Gong, W.: Concept factorization with adaptive neighbors for document clustering. IEEE Trans. Neural Netw. Learn. Syst. **29**(2), 343–352 (2018)
14. Nie, F., Wang, X., Jordan, M.I., et al.: The constrained Laplacian rank algorithm for graph-based clustering. In: Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press (2016)