2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Dual Graph regularized NMF with Sinkhorn Distance

Yunmeng Zhang, Zhenqiu Shu*, Jie Zhang*, Congzhe You, Zonghui Weng, Honghui Fan, Feiyue Ye
School of Computer Engineering, Jiangsu University of Technology,
Changzhou, 231001, China
Corresponding authors: shuzhenqiu@126.com, zhangjie@jsut.edu.cn

***Abstract*- Many researchers have paid more attention to the application of non-negative matrix factorization (NMF) in data representation. Recently, some regularization methods can improve the performances by utilizing the data and feature manifold, simultaneously. In this work, a new method, named dual graph regularized NMF with Sinkhorn distance (DSDNMF) is presented. It not only synchronously takes the data structure and feature structure into consideration, but also measures the reconstruction error by adopting the Earth Mover's Distance (EMD) to make full use of the feature correlation. Therefore, DSDNMF can effectively explore the semantic structure information of data in contrast to traditional methods. Besides, we introduce an efficient strategy to optimize our proposed model. Comprehensive experiments on the COIL20 and PIE datasets manifest the superiority of DSDNMF.**

Keywords-*NMF; data presentation; data manifold; feature manifold; EMD; semantic structure.*

I. INTRODUCTION

Data representation plays a vital role in some fields, such as data mining and pattern recognition [1]. Through suitable representation, the clustering performance of high-dimensional data can be significantly improved. Unfortunately, traditional pattern recognition methods are not only computationally expensive in dealing with high-dimensional data, but also easily lead to so-called "dimensionality curses". The goal of data representation methods is to effectively discover the latent semantic information hidden in data. In the past few decades, various data representation methods were developed to achieve this goal.

Over the past decades, NMF is a particularly attractive method owing to its strong psychological and physiological interpretation [2]. Different from other methods, NMF imposes the non-negative constraint, and thus results in the part-based representation that may be represented in the same way as data in the human brain. NMF has shown excellence results in dealing with real high-dimensional data [3]. One drawback of the original NMF is that it neglects the latent manifold in the data, which can raise the representation capability in many tasks. To solve this issue, Cai *et al.* [4] discover the manifold structure embedded in data by constructing a nearest neighbor graph. Many studies proposed that the data space and feature space lie in a low- dimensional submanifold structure [5]. Shang *et al.* [7] further proposed to construct two dual graph regularizers to exploit the geometric structure in data and feature space. Tong *et al*. [7] introduced a dual graph regularized NMF method for hyperspectral unmixing by constructing two regularizers in both spectral space and spatial space. The above-mentioned methods use the l_2 norm to measure the reconstruction cost and the distance between the samples. Therefore, the feature correlation of original data cannot be fully utilized in low-dimensional representation space. To alleviate this problem, Roman Sandler *et al.* [8] proposed to quantify the cost by using EMD instead of Frobenius norm. EMD aims to reflect the minimal amount by moving the mass between two distributions, and thus has been applied to some real problems in recent years. Since EMD is insensitive to the relationship between different dimensional features, EMD based NMF can achieve more robust performance than traditional l_2 norm based NMF [9]. Qian *et al*. [10] further proposed a nonnegative matrix factorization with Sinkhorn distance (SDNMF) method. It uses a graph regularizer to explore the geometric manifold structure effectively, and the correlation between features is fully considered by adopting an approximation of EMD as the metric.

In this article, a new NMF algorithm, namely dual graph regularized NMF with Sinkhorn distance (DSDNMF), is proposed, which takes both data and feature structure and the feature correlation information into consideration. We employ Sinkhorn distance to model the relationship between different dimensional features. Meanwhile, the dual graph regularizer is adopted to preserve the local geometric structures in dual space. In addition, we propose an effective multiplicative updating algorithm to optimize the model. The experimental results manifest that our DSDNMF algorithm is better than other competitors.

The framework of the paper is organized as follows. We briefly present the related works in Section 2. Details of our proposed DSDNMF algorithms are introduced in Section 3. In Section 4, the experimental results and their analysis are presented. Finally, we summarize the work in Section 5.

II. THE RELATED WORK

Given a data matrix $X = [x_1, x_2, ..., x_n] \in \mathbb{R}^{M\times N}$, each sample denotes a column of elements in matrix *X*. NMF tries to seek the matrices $U \in \mathbb{R}^{M\times D}$ and $V \in \mathbb{R}^{N\times D}$, and their product is to approximate the original matrix X. The model of NMF can be written as

$$J_{NMF} = \sum_{i=1}^{N}\sum_{j=1}^{M}(X_{ij} - (UV^T)_{ij})^2 = \left\| X - UV^T \right\|_F^2, \quad (1)$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. Lee and Seung [3] proposed the multiplicative iterative algorithm to optimize the problem (1), and derived the following the updating rules:

$$U_{ik} \leftarrow U_{ik}\frac{(XV)_{ik}}{(UV^TV)_{ik}}, \quad V_{jk} \leftarrow V_{jk}\frac{(X^TU)_{jk}}{(VU^TU)_{jk}}. \quad (2)$$

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00046

III. THE PROPOSED METHOD

A. Earth Mover's Distance

Given two histograms x and y, the definition of their EMD is given as follows:

$$d_M(x,y)=\min_{T_{pq}\geq 0}\sum_{p,q=1}^{m}M_{pq}T_{pq}\quad s.t.\ \sum_{q=1}^{m}T_{pq}=x_p,\sum_{p=1}^{m}T_{pq}=y_q,\forall p,q \quad . \tag{3}$$

We call Eq. (3) as the transportation problem. T_{pq} stands for the amount of traffic from p to q. M_{pq} is usually defined by L_1 or L_2 distance, and denotes the ground distance between two samples.

However, the computational complexity of EMD is very expensive. Cuturi *et al.* [11] proposed an efficient optimization algorithm. They use the entropy regularization term to smooth classical optimal transport problem and show that the Sinkhorn-Knopps matrix extension algorithm can be used to calculate the optimal distance. We call the new distance as Sinkhorn distance, whose mathematical form is given as follows:

$$d_M^{\lambda}(x,y)=\min_{T_{pq}\geq 0}(\sum_{p,q}M\odot T+\frac{1}{\lambda}H(T)) \ , \tag{4}$$

where $H(T)=-\sum_{p,q}T_{pq}\log T_{pq}$ denotes the entropy of T.

B. Data and feature graph regularizations

Many studies have proposed that the structure information among samples in both data and feature space lies on a low-dimensional manifold [6]. Therefore, we can derive a dual graph regularizer to preserve the geometric structures in dual space.

In feature space, we can use the coefficient matrix V to effectively exploit the geometric manifold structure. Therefore, we formulate the data graph regularization term as follows:

$$\min_V R_V=\sum_{j,s}^{n}W_{1js}\left\|v_j-v_s\right\|^2 \ , \tag{5}$$

where W_1 is the affinity matrix. Sinkhorn distance is used as the metric to construct the graph model.

Similarly, we can explore the data manifold by using the basis matrix U. In data space, we give the data graph regularization term as follows:

$$\min_U R_U=\sum_{i,s}^{m}W_{2is}\left\|u_i-u_s\right\|^2 \ , \tag{6}$$

where W_2 is the affinity matrix. Similarly, Sinkhorn distance is used as the metric to construct the graph model.

C. Objective function of DSDNMF

Since the approximation error between X and Y are usually non-Gaussian distribution, we can improve the representation performance by adopting Sinkhorn Distance as the metric. Therefore, the model of SNMF can be written as follows:

$$O=\sum_{j=1}^{n}d_M^{\lambda}(x_j,u_k v_{jk}) \ . \tag{7}$$

By integrating the graph regularizers (5) and (6) into formula (7), the loss function of DSDNMF is given as follows:

$$O=\sum_{j=1}^{n}d_M^{\lambda}(x_j,u_k v_{jk})+\frac{\xi}{4}R_V+\frac{\sigma}{4}R_U \ , \tag{8}$$

where δ and σ are the positive regularization parameters. It is obvious that the problem (8) is nonconvex in both U and V together, and thus it is only achieved local optimization solution. Similarly, we adopt the multiplicative iterative algorithm to solve (8). Therefore, we can present the rules of the problem (8) as

$$u_{ik}\leftarrow u_{ik}\frac{\sum_s v_{sk}\frac{\sum_t T_s^{*it}}{y_{is}}+\sigma\sum_{s\neq j}W_{is}u_{is}}{\sum_s v_{sk}+\sigma u_{ik}\sum_{s\neq j}W_{is}} \ , \tag{9}$$

$$v_{jk}\leftarrow u_{jk}\frac{\sum_s u_{sk}\frac{\sum_t T_j^{*st}}{y_{sj}}+\xi\sum_{s\neq j}W_{js}v_{js}}{\sum_s u_{sk}+\xi v_{jk}\sum_{s\neq j}W_{js}} \ , \tag{10}$$

where T_s^{*it} is the (i,t)-entry of the optimal transportation matrix between x_s and y_s.

IV. EXPERIMENTS

We conduct some clustering experiments on COIL20 and PIE datasets to evaluate our proposed DSDNMF method. Besides, some popular traditional methods, such as *k*-means (KM), NMF, GNMF, EMDNMF, SDNMF, and DGNMF, are compared with our proposed DSDNMF method.

A. COIL20 dataset

There are 1440 samples of 20 objects in the COIL20 database. It captured 72 images from different angles to each object. The size of all sample images is 32×32. Figure 1 shows the examples from the COIL20 dataset.

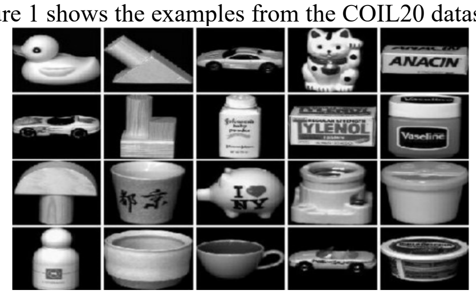

Figure 1. Some images in the COIL20 dataset

In each time, P categories samples were randomly selected to investigate the performance of the proposed method. To be fair, all experiments were independently run ten times, and the average performances are recorded. Tables 1 and 2 report the final results of all algorithms with different P values on the COIL20 dataset. From these tables, we can find that the capability of the proposed DSDNMF method is powerful than other competitors regardless of the value of P. This is because DSDNMF not only uses dual graph regularizers to explore the structure in dual space, but also uses Sinkhorn Distance as the metric to take advantage of the feature correlation of data.

Table 1 Accuracy on COIL20 database

P	*KM*	*NMF*	*GNMF*	*EMDNMF*	*SDNMF*	*DGNMF*	*DSDNMF*
10	0.647	0.598	0.727	0.650	0.752	0.733	**0.790**
12	0.635	0.538	0.732	0.655	0.759	0.745	**0.770**
14	0.631	0.620	0.720	0.629	0.726	0.731	**0.754**
16	0.619	0.649	0.768	0.578	0.810	0.801	**0.816**
18	0.580	0.655	0.746	0.577	0.785	0.762	**0.807**
20	0.589	0.620	0.724	0.587	0.706	0.732	**0.752**
avg	0.617	0.613	0.736	0.612	0.756	0.750	**0.781**

Table 2 Normalized mutual information on COIL20 database

P	*KM*	*NMF*	*GNMF*	*EMDNMF*	*SDNMF*	*DGNMF*	*DSDNMF*
10	0.717	0.736	0.821	0.728	0.822	0.821	**0.837**
12	0.671	0.626	0.805	0.682	0.817	0.810	**0.833**
14	0.726	0.721	0.831	0.699	0.834	0.841	**0.844**
16	0.720	0.722	0.840	0.649	0.860	0.856	**0.871**
18	0.701	0.721	0.868	0.675	0.870	0.849	**0.893**
20	0.711	0.702	0.839	0.691	0.856	0.844	**0.872**
avg	0.707	0.704	0.834	0.687	0.843	0.837	**0.858**

B. PIE dataset

There are 41,368 samples of 68 individuals on PIE database. The size of each image is 32×32. For convenience, we randomly picked out 50 samples from each category for clustering. Figure 2 shows some examples of the PIE face database.

Figure 2. Some images in PIE face dataset

In the second experiment, we also adopted the experimental scheme as above. Table 3 and Table 4 display the performances of seven algorithms on the PIE dataset. We can see clearly that DSDNMF outperforms all competitors. The main reason is that the dual graph regularizers can utilize the latent structure embedded in dual space effectively, and the Sinkhorn Distance is adopted to model the relationship between different dimensional features.

Table 3 Accuracy on PIE face database

P	*KM*	*NMF*	*GNMF*	*EMDNMF*	*SDNMF*	*DGNMF*	*DSDNMF*
36	0.164	0.271	0.461	0.245	0.468	0.455	**0.475**
40	0.177	0.297	0.443	0.273	0.437	0.421	**0.463**
44	0.184	0.263	0.454	0.261	0.452	0.410	**0.462**
48	0.193	0.299	0.514	0.278	0.460	0.425	**0.476**
52	0.191	0.278	0.441	0.254	0.430	0.413	**0.470**
56	0.193	0.285	0.488	0.287	**0.502**	0.456	0.495
60	0.177	0.289	0.458	0.246	0.459	0.398	**0.476**
avg	0.183	0.283	0.465	0.263	0.458	0.425	**0.473**

Table 4 Normalized mutual information on PIE face database

P	*KM*	*NMF*	*GNMF*	*EMDNMF*	*SDNMF*	*DGNMF*	*DSDNMF*
36	0.289	0.465	0.590	0.409	0.584	0.569	**0.611**
40	0.325	0.501	0.587	0.469	0.583	0.574	**0.605**
44	0.336	0.470	0.618	0.466	0.610	0.598	**0.625**
48	0.370	0.522	0.685	0.486	**0.687**	0.604	0.659
52	0.360	0.506	0.574	0.488	0.598	0.561	**0.618**
56	0.387	0.522	0.654	0.503	0.654	0.614	**0.661**
60	0.390	0.528	0.641	0.469	0.625	0.590	**0.647**
avg	0.351	0.502	0.621	0.470	0.620	0.587	**0.632**

C. Parameters selection

Since the different parameter settings have a certain influence on the performances, we conducted some experiments to analysis the sensitivity of the parameters. We display the results of DSDNMF by varying the parameters λ, ξ and σ. For simplicity, we set the same values of the parameters ξ and σ. Specifically, their values are set as [0.001, 0.01, 0.1, 1, 10, 100], respectively and the parameter λ is set as [1, 10, 50, 100, 500, 1000]. Figures 3 and 4 show the clustering result of DSNMF by varying values of the parameters ξ and σ on COIlL20 and PIE image datasets. Figures 5 and 6 show the performances of DSDNMF varied with parameter λ on the COIlL20 and PIE datasets. It is obvious that DSDNMF can achieve a stable performance when these parameters vary in a wide range.

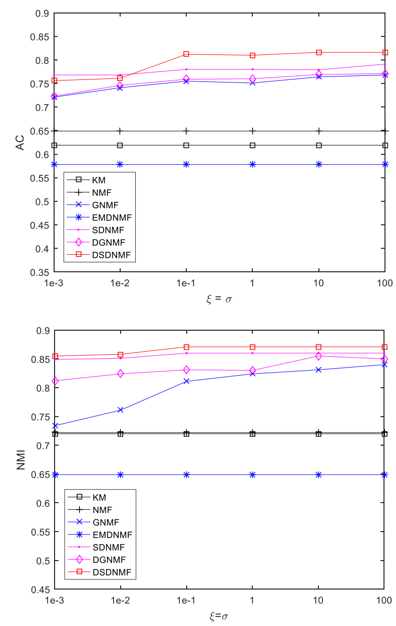

Figure3: Clustering performance on the COIlL20 with varying the parameters $\xi=\sigma$.

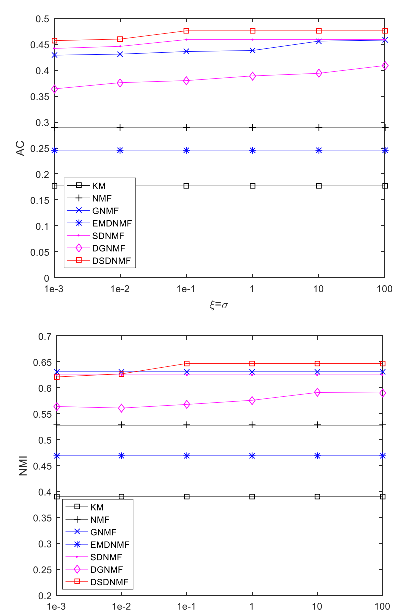

Figure4: Clustering performance on the PIE with varying the parameters $\xi = \sigma$.

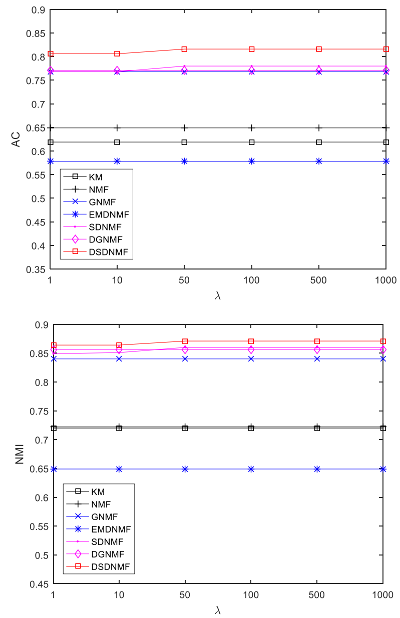

Figure5: Clustering performance on the COIlL20 with varying the parameter λ.

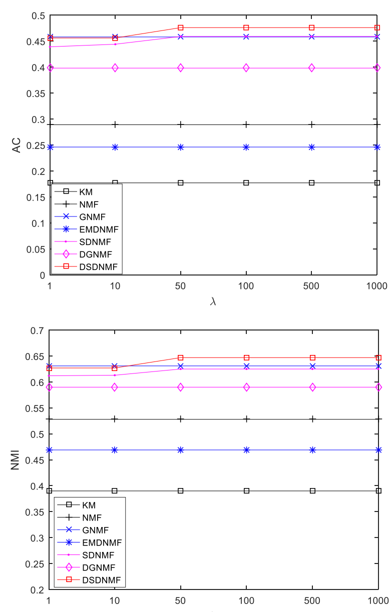

Figure6: Clustering performance on the PIE with varying the parameter λ.

V. CONCLUSIONS

In this work, we present a new algorithm, named DSDNMF, which utilizes the correlation information of features and the dual manifold structure, simultaneously. Specifically, we employ the Sinkhorn distance as the metric to measure the reconstruction errors, and thus can fully utilize the feature correlation. In addition, we adopt a dual graph regularizer to preserve both data and feature manifold in low-dimensional representation space. An efficient multiplicative iteration strategy is used to optimize the proposed model. The experimental results manifest that our DSDNMF method outperforms other competitors in clustering.

ACKNOWLEDGE

This work was supported by the National Natural Science Foundation of China [No. 61603159] and Excellent Key Teachers of QingLan Project in Jiangsu Province.

REFERENCES

[1] Z. Shu, X. Wu, C. You, Z. Liu, P. Li, H. Fan, and F. Ye, “Rank-constrained nonnegative matrix factorization for data representation,” Inf. Sci. vol. 528, 2020, pp. 133-146 .

[2] D. Lee and H. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, 401(6755), 1999, pp. 788-791.

[3] Z. Shu, C. Zhao, and P, Huang, “Local regularization concept factorization and its semi-supervised extension for image representation,” Neurocomputing, 2015, vol. 158, pp. 1-12.

[4] D. Cai, X. He, and J. Han, “Graph regularized nonnegative matrix factorization for data representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, vol. 33(8), pp. 1548-1560.

[5] Q. Gu and J. Zhou, “Co-clustering on manifolds,” The 15th ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2009, pp. 359-368.

[6] F. Shang, L. Jiao, and F. Wang, “Graph dual regularization non-negative matrix factorization for co-clustering,” Pattern Recognition, 2012, vol. 45, pp.2237-2250.

[7] L. Tong, J. Zhou, and X. Bai, “Dual graph regularized NMF for hyperspectral unmixing,” 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2017, pp.1-4.

[8] R. Sandler and M. Lindenbaum, “Nonnegative matrix factorization with earth mover’s distance metric for image analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, vol. 33(8), pp. 1590-1602.

[9] Y. Rubner, C. Tomasi, and Leonidas J Guibas, “The earth mover’s distance as a metric for image retrieval,” International journal of computer vision, 2000, vol. 40(2), pp. 99-121.

[10] Q. Wei, H. Bin, and C. Deng, “Non-negative matrix factorization with Sinkhorn distance,” International Joint Conference on Artificial Intelligence, 2016.

[11] M. Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” In Advances in Neural Information Processing Systems, 2013, pp. 2292- 2300.