Time Series Clustering via NMF in Networks

Guowang Du[1], Lihua Zhou[1*], Yuan Fang[1], Ming Yang[2]
School of Information Science and Engineering Yunnan University[1] , Yunnan University of TCM[2]
Kunming, China
18468178560@l63.com, lhzhou@ynu.edu.cn, fy1990825@163.com, 702392882@qq.com

***Abstract*—Time series data mining has attracted a lot of attention in the last decade, especially the research on the clustering of time series data. Network-based clustering technology, transforming data of time series into a network and then used community detection methods of network to cluster time series, is a new approach to cluster time series data. This approach takes the advantage that a network can describe the relationship between any pair or any group of data samples, but the effectiveness of clustering heavily dependent on the performance of algorithms of community detection. In this paper, we cluster time series by transforming them into network and detecting communities by non-negative matrix factorization (NMF). Experimental evaluations illustrate the superiority of our approach compared with the state-of-the-arts such as Multilevel.**

Keywords—Time Series Clustering； Network; Non-negative matrix factorization (NMF)

I. Introduction

Time series, collections of numerical values obtained from sequential measurements over time, are generally used to describe the current status and future variations of objects over time lags[1]. With the rapid growth of digital sources of information, enormous amounts of time series data, such as health trajectories of individuals[2], climate data[3], are being continually generated and collected. Mining these data is helpful to discover the hidden knowledge and information [4][5][6], such as temporal associations [7], community behavior patterns[8], thus it is receiving increasing attention from researchers in the recent years.

Clustering time series, i.e. dividing a set of time series into groups such that similar ones are put in the same group [9], is a fundamental task of time series data mining (TSDM). This task has been extensively studied and a great number of approaches have been proposed. Ferreiraa & Zhao [10] thought that time series clustering requires not only local information, but also global knowledge to capture the pattern formation of given time series, but in general only the local relationship amongst neighbor data samples can be easily identified, while long distance global relationship remains unknown in the original form of time series. To use the global knowledge for clustering time series, Ferreiraa & Zhao [10] proposed a network-based approach. This approach uses distance functions to transform a set of time series into a network (represented by a graph), where each time series is represented by a node and the most similar ones are connected, and then apply community detection algorithms to identify groups of strongly connected nodes and, consequently, identifies time series clusters. This network-based clustering technique can capture arbitrary cluster shapes, because the ability of network characterizing both local and global relationship among nodes as well as the ability of community detection algorithms identifying connectivity patterns (such patterns can be any shape in the Euclidean space) can be utilized by transforming time series from time-space domain to topological domain. However, this method depends on the effects of network construction and community detection algorithms to a certain extent.

NMF [11] is a powerful tool for data analysis with enhanced interpretability. It aims at learning the representative parts of the original data by approximating the target matrix into the product of two low-rank matrices. In the real world problems, the non-negative constraint of the input matrix is required, but this feature is directly explained by extra ordinary matrix factorization methods, and NMF is fast for convergence. In the last decades, non-negative matrix factorization has shown very good results in clustering and has been used to clustering many types of data, including documents [12], images [13], micro array data[1] and community discovery [15].

In this paper, we use network-based methods for clustering time series. We first transform time series data into network, then we apply to NMF and extend NMF to discovery communities, and each community is equivalent to a cluster.

The remaining sections of this paper are organized as follows: a brief overview of related work about clustering time series and NMF is given in Section II. Section III introduces our proposed method for time series clustering. Experiments and results are presented in Section IV. Finally, we conclude the paper in Section V.

II. RELATED WORK

A. Time Series Clustering

Clustering time series is one of the most important tasks in data mining. The purpose of clustering is to find similar time

* Corresponding Author

978-1-5386-5886-4/18/$31.00 ©2018 IEEE
SERA 2018, June 13-15, 2018, Kunming, China

series and divides them into a cluster. Zakaria et al. [16] learned shapelet (local patterns of a time series) from unlabeled time series and used the k-means algorithm to cluster shapelet. This method can cluster time series with noise and different lengths because it uses only some local patterns and deliberately ignoring the rest of the data. Huang et al. [17] proposed a new k-means type smooth subspace clustering algorithm (TSkmeans) for time series data, where the smooth subspaces is represented by weighted time stamps which indicate the relative discriminative power of these time stamps for clustering objects. To use the global knowledge for clustering time series, Ferreiraa & Zhao [10] transformed the time series into the form of a network and then clustered it through a community detection algorithm to obtain satisfactory results. Yang [18] proposed a HMM-based partitioning ensemble based on hierarchical clustering refinement to solve the problems of initialization and model selection for temporal data clustering. Bicego [19] proposed hidden Markov models to cluster time series. Hidden Markov models are trained by each time series, and then the similarity of the two models is measured by the output probability of HMM, and the distance between time series is measured by the similarity.

B. NMF

Non-negative Matrix Factorization (NMF) was originally proposed as a method for finding matrix factors with parts-of-whole interpretations. Later NMF has emerged as a powerful tool for data analysis with enhanced interpretability. In recent years, it has been successfully applied to environmetrics [20], chemometrics [22], bioinformatics [23], community discovery [15] etc. Wang et al [15] proposed symmetric NMF, a symmetric NMF and joint NMF to detect communities on undirected, directed and compound networks. Lin et al [21] proposed FacetNet to detect communities and their evolution in dynamic temporal networks.

Because NMF does not require the derived latent space to be orthogonal and it guarantees that non-negative value is taken in all the latent directions, NMF is superior to the tradition matrix factorization for solving the clustering problem in most real applications[17]. In order to achieve wider application, NMF has been extended in recent years. Hoyer [25] combines the NMF algorithm and sparse coding to form a non-negative sparse coding method, so that the decomposed coefficients have better sparseness, and the original information is represented by a smaller number of elements, which can save storage space and improve the operation efficiency. Kuang [27] proposed a symmetric non-negative matrix method to cluster symmetrical graphs and achieved good results. Cai [28] proposed a non-negative matrix method based on graph regularization to cluster symmetric graphs, and added graph regular constraints to traditional non-negative matrices. Zhang [23] proposed Non-negative Matrix Factorization on Kernels to extract more useful features hide in the original data. [28]proposed robust graph regularized nonnegative matrix factorization (RGNMF) to cluster graphs, and got a good effect.

III. The Method for Clustering Time Series via NMF in Networks

A time series X_i is an ordered sequence of t real values $X_i=(x_{i1}, x_{i2}, \ldots x_{it})$, $x_{ij} \in R$, $j \in \{1, 2, \ldots, t\}$. Let $X = \{X_1, X_2, \ldots, X_N\}$ represent a collection of N univariate time series.

The method proposed in this paper can be roughly classified into the following two steps: (1) network construction; (2) Clustering via Non-negative Matrix Factorization.

A. The Network Construction

A network is often represented as a graph $G = (X, E)$ composed by a set of nodes $X = \{X_1, \ldots, X_N\}$ and a set of edges $E = \{(X_i, X_j) \mid X_i, X_j \in X\}$, where the nodes represent objects and edges indicate interactions amongst objects. To construct a network, let each series correspond to a node, and compute the distance between any two time series X_i and X_j . If the distance between series X_i and X_j is smaller than a threshold ε, there is an edge between X_i and X_j, and the weight of the edge is assigned as the similarity between X_i and X_j. The adjacency matrix of the final graph G is represented by matrix V.

B. Clustering via Non-negative Matrix Factorization

Given a m x n nonnegative matrix V where all the elements are nonnegative, NMF decomposes V into two nonnegative matrices $W_{m \times c}$ and $H_{c \times n}$ such that $V = WH$,*s.t.* $W \geqslant 0, H \geqslant 0$. In general, c is much smaller than $\min\{m, n\}$. W is referred to as the base matrix and H is called the relation matrix. The largest coefficient in the i-th column of W indicates the cluster that object belongs to. W and H together constitute non-negative linear relationship of matrix V. The optimal W and H are learned by minimizing a particular loss function, such as Euclidean distance :

$$D(W,H) = \frac{1}{2} \| V - WH \|_F^2 \quad (1)$$

where $\|\bullet\|_F$ denotes the Frobenius norm of the matrix. Lee[11] gives the following multiplicative iteration rules:

$$h_{ij}^{t+1} \leftarrow h_{ij}^{t} \frac{(W^T V)}{(WV^T V)} \tag{2}$$

$$\mathrm{w_{ij}}^{t+1} \leftarrow \mathrm{w_{ij}}^{t} \frac{(\mathrm{VH^T})}{(\mathrm{WHH^T})} \tag{3}$$

NMF has been extended as semi definite non-negative matrix (Semi-NMF), kernel non-negative matrix factorization(KNMF), Sparse non-negative matrix factorization(SNMF), robust graph regularized nonnegative matrix factorization (RGNMF) and Symmetric non-negative matrix factorization(Sym-NMF) to improve the performance of NMF. In the following, we briefly introduce them.

1) *SNMF.* Hoyer [25] combined the NMF algorithm and sparse coding to form a non-negative sparse coding, so that the decomposition has good sparsity, and the original information is expressed with fewer elements, which can save storage space and improve the operation efficiency. However, Hoyer's algorithm was slow. In order to improve the interpretability of the basis vectors and speed up the algorithm, Li [24] implemented the following model:

$$\min_{V,H} \frac{1}{2} \|V - WH\|_F^2 + \lambda \sum |h_i|_1 \tag{4}$$

subject $V,W,H \geqslant 0$ $h_i \in H$, $i \in \{1,.....,c\}$, λ represents sparse coefficient .

2) *Semi-NMF.* The origin NMF only works for non-negative data, so Ding [26] extended it to semi definite non-negative matrix (Semi-NMF) which removes the non-negative constraints on the matrix V and basis matrix W. Semi-NMF can be applied to the matrix of mixed signs, therefore it expands NMF to many fields. Semi-NMF minimizes the following objective functions:

$$\min_{V,W} \frac{1}{2} \|V - WH\|_F^2 \tag{5}$$

Subject to $H > 0$.

3) *KNMF.* The origin NMF only processes the linear structure data and cannot make full use of the data information. Zhang [23] extends the original non-negative matrix factorization to kernel non-negative matrix factorization (KNMF). The advantages of KNMF over NMF are: 1) it could extract more useful features hidden in the original data through some kernel-induced nonlinear mappings; 2) it can deal with data where only relationships (similarities or dissimilarities) between objects are known; 3) it can process data with negative values by using some specific kernel functions (e.g. Gaussian). Thus, KNMF is more general than NMF.

4) *Sym-NMF.* Although origin NMF can cluster data, it cannot make full use of the data information when data has nonlinear structure. Kuang [20] proposed a symmetric non-negative matrix method to cluster symmetrical graph. Sym-NMF can get information about nonlinear structure and can get better clustering effects in some situations. Sym-NMF minimizes the following objective functions:

$$\min_{0 \leq H} \|V - HH^T\|_F^2 \tag{6}$$

where H is a non-negative matrix of size , and k is the number of clusters requested.

5) *RGNMF.* Peng [29] proposed robust graph regularized nonnegative matrix factorization (RGNMF), which also takes advantage of the nonlinear structures of the data for clustering in a unified framework. In addition, this method accounts for within-sample outliers and impulsive noise. In order to optimize RGNMF, Peng proposed two variant formulations: RGNMF-ALM and RGNMF-Multi. RGNMF-Multi got very good effect. RGNMF-Multi minimizes the following objective functions:

$$\min_{W,H,S} \|X - WH - S\|_F^2 + \alpha \|S\|_0 + \varphi Tr(HLH^T) \tag{7}$$

Subject to $W,H \geqslant 0$. Where S represents the sparse matrix which can be regarded as outliers, φ is the penalty term to control L which is Laplace matrix and α is a tradeoff parameter to control the sparsity of S.

Next, we outline our algorithm steps.

Algorithm 1: NMF-Based Clustering

Input: Adjacent matrix V, Cluster number r, Relative Error Threshold β, Maximum iteration K

Output: W,H, Cluster CS

1. Initialize: Randomly initial W, H, CS
2. m,n=size(V)
3. $last_{err}$=1+err
4. For iter=1 to K
5. Compute $V_c = W * H$
6. err= $\frac{1}{2}\|V - V_c\|$
7. If $last_{err}$-err< β
8. Break

```
9.    Else
10.      Last_err=err
11.      Fixed W, update H
12.      Fixed H, update W
13.    End If
14. End For
15. For i=1 to m
16.    index = argMax_j(W_ij), j ∈ {1,...,r}
17.    i → CS{index}
18. End For
```

Algorithm 1 shows non-negative matrix decomposition clustering methods. For different factorization methods, Step 11 and Step 12 are different. Because different decomposition methods have different constraints for non-negative matrix factorization. There are two termination conditions: number of iterations K, relative error threshold β.

IV Experiments and Results

In this section, we first use Dynamic Time Warping (DTW) to measure the distance between time series and construct ε-nearest neighbour networks (ε-NN) by using the method proposed by Ferreiraa & Zhao[10], because their experimental results demonstrated that ε-NN based on DTW can lead to better performance. Then, we use NMF, Semi-NMF, KNMF, SNMF, RGNMF and Sym-NMF respectively to decompose adjacent matrix of network. Next, we compare our results with those obtained by Multilevel[10], so as to test the effectiveness of NMF to detect communities. In addition, we explore the effects of ε and β.

A. DataSet

In our experiment, we use 11 distinct time series datasets of the UCR time series repository[30]. Table I describes the attributes of the datasets.

Table I Data Set Description

Number	Name	Number of clusters	Number of series	Length of each time series
1	beef	5	30	470
2	cinc_ecg_torso	4	40	1639
3	coffee	2	28	286
4	diatom_size_reduction	4	16	345
5	ecg_five_days	2	23	136
6	mote_strain	2	20	84
7	oliveoil	4	30	570
8	sony_AIBO_Robot_surface	2	27	65
9	sony_AIBO_Robot_surface_ii	2	20	70
10	symbols	6	25	398
11	two_lead_ecg	2	23	82

B. Evaluation

We use Rand Index (*RI*) to evaluate the clustering performance by comparing the experimental results with the correct cluster labels provided by UCR. The Rand index is calculated as follows:

$$RI = \frac{TP + TN}{N(N-1)/2} \tag{8}$$

where TP (true positive) is the numbers of pairs of time series that are correctly put in the same cluster, TN (true negative) is the number of pairs that are correctly put in different clusters and N is the total number of series. The range of *RI* is 0 to 1. The larger the *RI*, the higher the accuracy of the clustering results.

C. Experiment Result and Analysis

Table II shows the comparison between NMF-based algorithm and Multilevel [10], where the distance threshold ε is set to 1e-4, the relative error threshold β of the NMF, Sym-NMF, SNMF, KNMF is 1e-5, and the relative error threshold β of the Semi-NMF is 1e-4, and the relative error threshold β of RGNMF algorithm is 1e-2. In addition, for SNMF the sparsity of the sparse coefficient λ is set to 0.1 and for RGNMF α and φ are set to 1,10 respectively.

It can be seen from table II that the average Rand Index for NMF, SNMF, Sym-NMF, Semi-NMF, KNMF, RGNMF and Multilevel is 0.799, 0.801, 0.832, 0.835, 0.815, 0.838,0.790 respectively. It indicates that NMF methods are superior to Multilevel method in most data sets, especially in the case of RGNMF and Semi-NMF. They not only have higher average values than that of the Multilevel algorithm, but also have higher median values than that of the Multilevel algorithm. But NMF-based methods show poor effect on the beef and cinc_ecg_torso datasets.

In Figure 1 and 2, the horizontal axis number i indicates the i-th dataset and the vertical axis indicates the Rand Index.

Next we test the effect of similarity threshold ε. For different ε values, the transformation of the network may be very different, and even affect the effectiveness of the clustering algorithm. Figure 1 shows the Rand Index of RGNMF for the network under different threshold ε. From Figure 1, we can see that the smaller the threshold ε, the better the clustering result. Especially when the threshold ε is 1e-4, the clustering result is best.

Table II Rand index for NMF、KNMF、Semi-NMF、SNMF、Sym-NMF and Multilevel

	Sym-NMF	Semi-NMF	RGNMF	KNMF	SNMF	NMF	Multilevel
beef	0.786	0.800	0.813	0.783	0.782	0.786	**0.83**
cinc_ecg_torso	0.671	0.692	0.702	0.664	0.667	0.654	**0.75**
Coffee	0.801	0.928	**1**	0.862	0.929	0.802	0.6
diatom_size_reduction	1	0.958	**1**	0.933	0.933	0.958	0.97
ecg_five_days	0.700	0.699	**0.762**	0.699	0.644	0.644	0.63
mote_strain	0.900	**0.90**	0.81	0.732	0.811	0.811	0.78
oliveoil	0.894	0.871	**0.894**	0.853	0.876	0.874	0.88
sony_AIBO_Robot_surface	0.811	0.811	**1**	1	0.900	0.811	0.83
sony_AIBO_Robot_surface_ii	0.858	0.858	**0.857**	0.795	0.858	0.858	0.85
Symbols	0.970	**0.973**	0.890	0.943	0.88	0.897	0.97
two_lead_ecg	0.763	0.700	**0.833**	0.700	0.700	0.700	0.61
Average	0.832	0.836	**0.869**	0.815	0.816	0.799	0.791
Median	0.811	0.858	**0.874**	0.795	0.858	0.811	0.83

Finally, we test the effect of relative error thresholds β. Figure 2 shows Rand Index of RGNMF for the network under different β. In most cases, we can find that for different relative error thresholds, the RGNMF achieves similar results. However, RGNMF achieves better results on the 3-th(coffee) dataset when the relative error threshold β is 1e-2 and RGNMF achieves better results on the and 10-th(Symbols) when the relative error threshold β is 1e-4. Overall, RGNMF produces better results when the relative error threshold β is 1e-2.

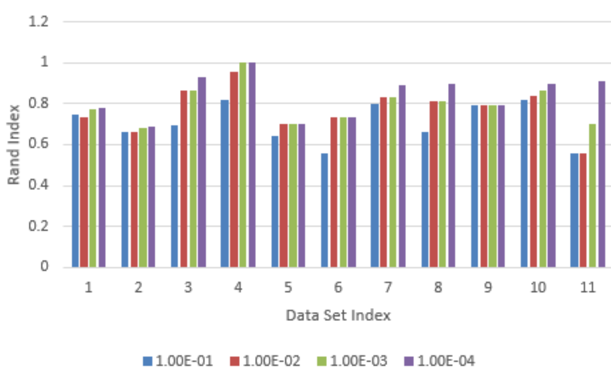

Figure 1 Rand Index for the network under different threshold ε

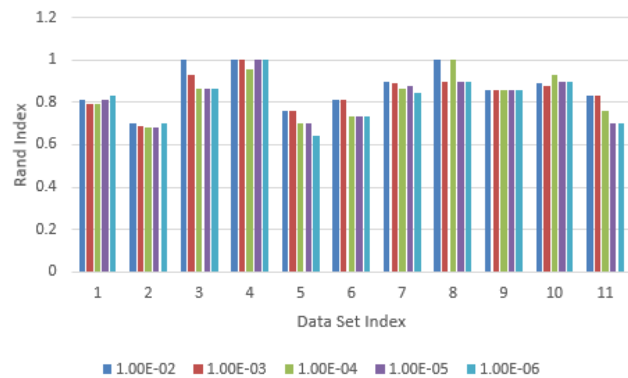

Figure 2 Rand Index for network graph of different β

V CONCLUSION

In this paper, we cluster time series by transforming them into network and detecting communities by matrix factorization. We use DTW to measure similarity between time series and construct ε-nearest neighbor networks and use NMF, Semi-NMF, KNMF, SNMF , RGNMF and Sym-NMF respectively to decompose adjacent matrix of network. The experimental results of 11 datasets of UCR time series repository indicates that clustering based on matrix factorization is effective, especially for RGNMF.

VI ACKNOWLEDGEMENT

The authors would like to sincerely thank anonymous reviewers for their valuable comments. This research was supported by the National Natural Science Foundation of China (61762090, 61262069, 61472346, 61662086), The Natural Science Foundation of Yunnan Province (2016FA026, 2015FB114), the Project of Innovative Research Team of Yunnan Province, Program for Innovation Research Team (in Science and Technology) in University of Yunnan Province (IRTSTYN).

REFERENCES

[1] W. Deng, G. Wang, J. Xu. Piecewise two-dimensional normal cloud representation for time-series data mining[J]. Information Sciences, 2016, 374: 32-50.

[2] S. Ghassempour, F. Girosi, A.Maeder. Clustering multivariate time series using hidden Markov models[J]. International journal of environmental research and public health, 2014, 11(3): 2741-2763.

[3] Y. Liu. Scalable Multivariate Time-Series Models for Climate Informatics[J]. Computing in Science & Engineering, 2015, 17(6): 19-26.

[4] D. Tao, J. Cheng, M Song, X. Lin. Manifold Ranking-Based Matrix Factorization for Saliency Detection[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 27(6):1122-1134.

[14] H. Kim, H. Park. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis[J]. Bioinformatics, 2007, 23(12): 1495-1502.

[15] F. Wang, T. Li, X.Wang, et al. Community discovery using non-negative matrix factorization[J]. Data Mining and Knowledge Discovery, 2011, 22(3): 493-521.

[16] J. Zakaria, A. Mueen, E. Keogh. Clustering Time Series Using Unsupervised-Shapelets[C]// IEEE, International Conference on Data Mining. IEEE Computer Society, 2012:785-794.

[17] X. Huang, Y. Ye, H. Guo, et al. DSKmeans: a new kmeans-type approach to discriminative subspace clustering[J]. Knowledge-Based Systems, 2014, 70: 293-300.

[18] Yun. Yang, Jianmin. Jiang. HMM-based hybrid meta-clustering ensemble for temporal data. Knowl.-Based Syst. 56: 299-310 (2014)

[19] M. Bicego , V. Murino , MAT. Figueiredo. Similarity-based clustering of sequences using hidden Markov models[C]//International Workshop on Machine Learning and Data Mining in Pattern Recognition. Springer, Berlin, Heidelberg, 2003: 86-95.

[20] G. Palla, I. Derényi, I. Farkas, T. Vicsek, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. nature, 2005, 435(7043): 814.

[21] Y. R. Lin, Y. Chi, S. Zhu, H. Sundaram. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks[C]//Proceedings of the 17th international conference on World Wide Web. ACM, 2008: 685-694

[22] Y. L. Xie, PK. Hopke, P. Paatero. Positive matrix factorization applied to a curve resolution problem[J]. Journal of Chemometrics, 1998, 12(6): 357-364.

[5] L. Wang, P. Wu, H. Chen. Finding Probabilistic Prevalent Co-Locations in Spatially Uncertain Data Sets[J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 25(4):790-804.

[6] K. Yue, Q. Fang, X. Wang, et al. A Parallel and Incremental Approach for Data-Intensive Learning of Bayesian Networks[J]. IEEE Trans Cybern, 2017, 45(12):2890-2904.

[7] D. E. H. Zhuang, G. C. L. Li, A. K. C. Wong. Discovery of temporal associations in multivariate time series[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(12): 2969-2982.

[8] N. İlhan, ŞG. Öğüdücü. Predicting community evolution based on time series modeling[C]// Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. ACM, 2015: 1509-1516.

[9] P. Esling, C. Agon. Time-series data mining[J]. ACM Computing Surveys (CSUR), 2012, 45(1): 12.

[10] L. N. Ferreira, L. Zhao. Time series clustering via community detection in networks[J]. Information Sciences, 2015, 326(C):227-242.

[11] D. Lee, H. S. Seung. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788.

[12] W. Xu, X. Liu, Y. Gong. Document clustering based on non-negative matrix factorization[C]//Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003: 267-273.

[13] D. Guillamet, J. Vitrià. Non-negative Matrix Factorization for Face Recognition[J]. 2002, 35:336-344.

[23] D. Zhang, Z. H. Zhou, S. Chen. Non-negative matrix factorization on kernels[C]//Pacific Rim International Conference on Artificial Intelligence. Springer, Berlin, Heidelberg, 2006: 404-412.

[24] Y. Li, A. Ngom. The non-negative matrix factorization toolbox for biological data mining[J]. Source code for biology and medicine, 2013, 8(1): 10.

[25] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints[J]. Journal of machine learning research, 2004, 5(Nov): 1457-1469.

[26] C. H. Q. Ding, T. Li, M. I. Jordan. Convex and Semi-Nonnegative Matrix Factorizations[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(1):45-55.

[27] D. Kuang, C. Ding, H. Park. Symmetric non-negative matrix factorization for graph clustering[C]//Proceedings of the 2012 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, 2012: 106-117.

[28] D. Cai, X. He, X. Wu, J. Han. Non-negative matrix factorization on manifold[C]//Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008: 63-72.

[29] P. Chong, K. Zhao et al. Robust Graph Regularized Nonnegative Matrix Factorization for Clustering[J]. Acm Transactions on Knowledge Discovery from Data, 2017, 11(3):33.

[30] E. Keogh, Q. Zhu, B. Hu, Y. Hao., X. Xi, L. Wei, C. A. Ratanamahatana, The-UCR-time-series-dataset,http://www.cs.ucr.edu/~eamonn/time_series_data/, [Online; accessed Sep-2014] (2008)