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

Non-negative Sparse Laplacian regularized Latent Multi-view Subspace Clustering

Cong-Zhe You, Hong-Hui Fan, Zhen-Qiu Shu
Jiangsu University of Technology
Changzhou, China
youcongzhe@126.com, fanhonghui@jsut.edu.cn, shuzhenqiu@163.com

***Abstract*—Recently, in the area of artificial intelligence and machine learning, subspace clustering of multi-view data is a research hotspot. The goal is to divide data samples from different sources into different groups. We proposed a new subspace clustering method for multi-view data which termed as Non-negative Sparse Laplacian regularized Latent Multi-view Subspace Clustering (NSL2MSC) in this paper. The method proposed in this paper learns the latent space representation of multi view data samples, and performs the data reconstruction on the latent space. The algorithm can cluster data in the latent representation space and use the relationship of different views. However, the traditional representation-based method does not consider the non-linear geometry inside the data, and may lose the local and similar information between the data in the learning process. By using the graph regularization method, we can not only capture the global low dimensional structural features of data, but also fully capture the nonlinear geometric structure information of data. The experimental results show that the proposed method is effective and its performance is better than most of the existing alternatives.**

Keywords-Multi-View Representation, Subspace Clustering, Latent Representation, Graph Regularization, Laplacian Matrix.

I. Introduction

Clustering analysis is a commonly used processing and analysis tool in artificial intelligence and machine learning fields. At present, researchers have proposed a large amount of clustering analysis methods, and have been widely used. However, due to the continuous progress of data science, the rapid development of data sources and access methods, the data obtained is more and more complicated. Among them, multi-view data is a wide-ranging phenomenon. However, traditional clustering methods are mostly proposed and developed for single-view data, in the face of multi-view data clustering, the performance is not ideal. Therefore, how to design a clustering analysis algorithm for multi-view datasets is a huge challenge. In order to effectively solve the problem of multi view data clustering, researchers have proposed many multi view cluster analysis methods from many aspects.

Numerous studies have shown that high-dimensional data can usually be embedded in the low-dimensional subspaces, so the subspace learning algorithm for multi-view data based on spectral clustering has been widely applied and achieved good performance. Representative algorithms include subspace clustering algorithm for multi-view data (MSC)[1], latent subspace clustering algorithm for multi-view data (LMSC) [3] and diversity-induced subspace clustering algorithm for multi-view data (DiMSC)[2].

In this paper, a Laplacian regularized multi-view clustering method based on latent representation is proposed. At the same time, we further introduce sparse and non-negative constraints in the algorithm model to improve the performance and rationality of the algorithm.

II. Related Works

Multi-view datasets contain a wealth of information from multiple sources, which is very useful for clustering analysis. Therefore, studying clustering methods for multi-view data is very important. Based on the single-view subspace clustering algorithm, LT-MSC [4] is proposed. The algorithm model can be expressed as follows:

$$\begin{gathered} min_{Z^v,E^v}\|Z\|_* + \lambda\|E\|_{2,1} \\ \text{s.t. } X^v = X^v Z^v + E^v,\ v = 1,\cdots,m \end{gathered} \tag{1}$$

In the above optimization function, Z is the coefficient of the data representation, $Z = \emptyset(Z^1,\cdots,Z^m)$, and E is the reconstruction error, $E = [E^1,\cdots,E^m]$, $\emptyset(\cdot)$ is the construction method of a tensor Z with a dimension of $n \times n \times m$ by combining the representation coefficient matrix $Z^v\ (v = 1,\cdots,m)$ of different views into one 3-order tensor. Finally, each view's representation is merged through

$$Z = \frac{1}{m}\sum_{v=1}^{m}(|Z^v| + |Z^v|^T) \tag{2}$$

However, the fusion method is too simple to fully and accurately explore the complementary information between different view data samples

At present, there are also many studies that use weight-and rules to fuse multi-view representations. However, it still cannot solve the fuse error problem. In the single-view subspace clustering analysis algorithm, the Latent Space Sparse Subspace Clustering (LS3C) [5] algorithm can simultaneously reduce the dimensionality and sparse representation. Inspired by this idea, the work in [3] proposed a LMSC method, which treats different views as projections from the same latent space and performs subspace aggregation on this latent space. The LMSC model can be expressed:

$$\begin{gathered} min_{Z,P,H,E_h,E_r}\|E\|_{2,1} + \lambda\|Z\|_* \\ \text{s.t. } E = [E_h;\ E_r],\ H = HZ + E_r, \\ X = PH + E_h,\ PP^T = I \end{gathered} \tag{3}$$

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00062

In the above optimization problem, $P = [(P^1)^T, \cdots, (P^m)^T]^T$ is a projection matrix of the multi-view data and the data set is represented as $X = [(X^1)^T, \cdots, (X^m)^T]^T$.

However, there is a problem of this method is that the complementary information between different data views is ignored, and no structural constraints are used on the representation coefficient matrix Z. In addition, this kind of method needs to use the representation coefficient matrix Z to do spectral clustering to get the final clustering partition result. Therefore, the accuracy of the coefficient matrix Z of the data representation affects the accuracy of the final result.

III. NON-NEGATIVE SPARSE LAPLACIAN REGULARIZED LATENT MULTI-VIEW SUBSPACE CLUSTERING

In this paper, we do some research on the subspace clustering problem of multi-view latent representations. Given a multi-view data set $\left\{[x_i^{(1)}; \cdots; x_i^{(V)}]\right\}_{i=1}^{N}$ with V different views, the number of samples is N. Many studies of the researchers have shown that different data views can be represented by the same latent representation space. Therefore, the goal of this method is to find such a common latent representation space h for each data sample point.

A. Objective Function

The objective function of the proposed Non-negative Sparse Laplacian regularized Latent Multi-view Subspace Clustering (NSL2MSC) is represented as

$$\begin{gathered} min_{P,H,Z,E_h,E_r} \|Z\|_* + \lambda_1 \|Z\|_1 + \\ \lambda_2 \sum_{ij} \|z_i - z_j\|^2 W_{ij} + \lambda_3 \|E_h\|_{2,1} + \lambda_4 \|E_r\|_{2,1} \\ \text{s.t. } H = HZ + E_r, X = PH + E_h, \\ PP^T = I, Z \geq 0 \end{gathered} \tag{4}$$

where $P\ and\ X$ and are reconstruction models aligned and the multi-view observations, respectively. E_r and E_h represent the errors because of the subspace representation and the latent representation respectively.

In the objective function, the third term is the constraint term added based on the manifold assumption. Under the manifold assumption, the relationship between the two samples can be expressed as

$$min \sum_{ij} \|z_i - z_j\|^2 W_{ij} \tag{5}$$

Where, z_i and z_j are respectively the coefficients matrix of sample points x_i and x_j under some transformations.

Manifold constraints are important for the construction of many algorithm models, such as dimension reduction algorithm, clustering algorithm and semi-supervised learning algorithm. Where D is a diagonal matrix, which represents the degree of the matrix. D_{ii} is the i-th diagonal element, and its value is the sum of all the similarity relationship associated with y, i.e. $D_{ii} = \sum_j W_{ij}$, Therefore, we can express the objective function of graph Laplacian as follows [6].

$$L = D - W$$

In addition, the objective function of graph embedding (5) can easily be expressed as follows

$$min\ tr(ZLZ^T)$$

Therefore, through these algebraic transformations, the model of the proposed method can be represented as the following matrix form

$$\begin{gathered} min_{P,H,Z,E_h,E_r} \|Z\|_* + \lambda_1 \|Z\|_1 + \\ \lambda_2 tr(ZLZ^T) + \lambda_3 \|E\|_{2,1} \\ \text{s.t. } E = [E_h; E_r], H = HZ + E_r, \\ X = PH + E_h,\ PP^T = I, Z \geq 0 \end{gathered} \tag{6}$$

B. Model Optimization

The objective function in Eq. (6) proposed in this paper can get a latent representation of the data and get meaningful similarity matrix from the latent representation which learns from multiple views. Although for the variables in the algorithm model, such as P, H, Z, E_h and E_r the proposed method cannot guarantee joint convexity. For each variable we can perform iterative optimization by fixing the other variables separately. For example, LADMAP [7] algorithm is an effective optimization method to solve this problem.

In order to optimize the objective function, we first introduce an auxiliary variable J into the algorithm model to separate the objective function. In this way, we describe the optimization problem:

$$\begin{gathered} min_{P,H,Z,E_h,E_r,J} \|Z\|_* + \lambda_1 \|J\|_1 + \\ \lambda_2 tr(ZLZ^T) + \lambda_3 \|E\|_{2,1} \\ \text{s.t. } E = [E_h; E_r], H = HZ + E_r, \\ X = PH + E_h,\ PP^T = I, Z = J, J \geq 0 \end{gathered} \tag{7}$$

The augmented Lagrange function of the above problem is

$$\begin{aligned} & L(P, H, Z, E_h, E_r, J) \\ & = \|Z\|_* + \lambda_1 \|J\|_1 + \lambda_2 tr(ZLZ^T) \\ & + \lambda_3 \|E\|_{2,1} + \Phi(M_1, H - HZ - E_r) \\ & + \Phi(M_2, Z - J) \\ & + \Phi(M_3, P - PH - E_h) \end{aligned}$$

In addition, in order to facilitate the subsequent calculation and representation, we give the definitions of Φ: $\Phi(C, D) = \frac{\mu}{2} \|D\|_F^2 + \langle C, D \rangle$,where μ is a positive penalty scalar and $\langle \cdot,\cdot \rangle$ defines the matrix inner product. According to LADMAP, by fixing the other variables, the variables P, H, Z, E, J can be updated by solving the following optimization problems iteratively.

(1) ***P***-subproblem:

$$\begin{gathered} \mathrm{P}^* = \arg\min \Phi(\mathrm{M}_3, \mathrm{X} - \mathrm{PH} - \mathrm{E_h}) \\ \text{s.t. } PP^T = I \end{gathered} \tag{8}$$

(2) ***H***-subproblem:

$$\begin{aligned} H^* = arg\ min\ & \Phi(M_3, X - PH - E_h) + \\ & \Phi(M_1, H - HZ - E_r) \end{aligned} \tag{9}$$

(3) ***Z***-subproblem:

$$Z^* = \arg min_Z \|Z\|_* + \langle \nabla_Z q(Z_k), Z - Z_k \rangle + \frac{\eta}{2}\|Z - Z_k\|_F^2 \tag{10}$$

where

$$q(Z, E_k, J_k, M_1^K, M_2^K) = \lambda_2 tr(ZLZ^T) + \frac{\mu}{2}\left\|Z - J_k + \frac{1}{\mu}M_2^k\right\|_F^2 + \frac{\mu}{2}\left\|H - HZ - E_h^k + \frac{1}{\mu}M_1^k\right\|_F^2$$

(4) ***E***-subproblem:

$$\begin{aligned} E^* &= \arg min_E \|E\|_{2,1} + \Phi(M_3, X - PH - E_h) + \Phi(M_1, H - HZ - E_r) \\ &= arg\ min_E \frac{1}{\mu}\|E\|_{2,1} + \frac{1}{2}\|E - G\|_F^2 \end{aligned} \tag{11}$$

Where the matrix G in the above formula is obtained by connecting the two matrices $(X - PH + M_3/\mu)$ and $(H - HZ + M_1/\mu)$ vertically.

(5) J-subproblem:

$$\begin{aligned} J^* &= arg\ min_{J\geq 0}\ \lambda_1\|J\|_1 + \Phi(M_2, Z - J) \\ &= arg\ min_{J\geq 0}\ \lambda_1\|J\|_1 + \frac{\mu}{2}\|J - (Z + M_2/\mu)\|_F^2 \end{aligned} \tag{12}$$

(6) Updating Multipliers:

$$\begin{cases} M_1 = M_1 + \mu(H - HZ - E_r) \\ M_2 = M_2 + \mu(Z - J) \\ M_3 = M_3 + \mu(X - PH - E_h) \end{cases}$$

C. Complexity and Convergence

The method proposed is optimized through six subproblems. Where, the time complexity of solving the P-subproblem is $O(k^2d + d^3)$, n , k and d respectively represent the number of data samples, the latent space dimensions and the dimensions of multi-view data features. To solve the J-subproblem, the time complexity is $O(n^3)$. To solve the H-subproblem, Bartels Stewart algorithm was used to solve Sylvester equation, and the time complexity was $O(k^3)$. In the process of solving the Z-subproblem optimally, the main task is to compute the inverse of the matrix, and the time complexity is $O(n^3)$ .For E subproblems and Lagrange multiplier calculations, the main task is to calculate matrix multiplication, with the time complexity $O(dkn + kn^2)$. Thus, the total time complexity of each iteration is $O(k^2d + d^3 + k^3 + n^3 + dkn + kn^2)$. The total time complexity can be expressed as $O(d^3 + n^3)$ because of $k \ll d$. The convergence of the algorithm is difficult to prove, but the effect of the algorithm on the real dataset shows that the proposed method has strong convergence and stability even if H is randomly initialized.

IV. Experiments

In the experiment, we chose four real datasets (MSRCV1 [8], extended YaleB, Still DB [9], BBCSport [10]) as test data to verify the effectiveness of the proposed method. And some the state-of-the-art algorithms including SPC, LRR[11],Co-Reg SPC[12],RMSC[10] and LMSC[3] are adopted as the compared methods.

A. Performance Comparison

To evaluate the characteristics of each algorithm, we adopt four evaluation indexes, NMI, ACC, F-measure and RI. For the 5 compared algorithms in the experiment, we adjust all parameters to the best performance in the corresponding paper. In the experiment, for all experimental data sets, the dimension of latent space of the proposed algorithm is set to $K = 500$, and the value range of three parameters $\lambda_1, \lambda_2, \lambda_3$ is from $\{1e^{-3}, 1e^{-2}, 1e^{-1}, 1, 1e^1, 1e^2, 1e^3\}$. Each experiment is run independently for 30 times and the results are recorded in Table 1-4.

TABLE I. Clustering Results on the MSRCV1

Method	NMI		ACC		F-Measure		RI	
	Average	Std	Average	Std	Average	Std	Average	Std
SPC	0.5961	0.0267	0.6831	0.0441	0.5566	0.0414	0.8698	0.0089
LRR	0.5235	0.0153	0.6063	0.0025	0.4743	0.0056	0.8612	0.0007
Co-Reg SPC	0.6056	0.0134	0.6993	0.0123	0.5581	0.0203	0.9025	0.0028
RMSC	0.5987	0.0059	0.7161	0.0072	0.6092	0.0151	0.8886	0.0023
LMSC	0.6724	0.0117	0.8121	0.0115	0.6713	0.0159	0.9027	0.0017
NSL2MSC	0.7206	0.0121	0.8469	0.0101	0.7037	0.0126	0.9258	0.0026

TABLE II. Clustering Results on the Extended Yale B

Method	NMI		ACC		F-Measure		RI	
	Average	Std	Average	Std	Average	Std	Average	Std
SPC	0.3807	0.0115	0.3815	0.0349	0.3049	0.0117	0.2347	0.0139
LRR	0.6345	0.0048	0.6286	0.0145	0.5263	0.0069	0.4511	0.0019
Co-Reg SPC	0.1819	0.0054	0.2073	0.0076	0.1703	0.0008	0.0882	0.0006
RMSC	0.1627	0.0119	0.1868	0.0136	0.1672	0.0107	0.0718	0.0123
LMSC	0.7231	0.0111	0.7328	0.0152	0.6259	0.0082	0.5825	0.0121
NSL2MSC	0.8026	0.0117	0.7818	0.0120	0.7406	0.0038	0.7121	0.0058

TABLE III. Clustering Results on the Still DB

Method	NMI		ACC		F-Measure		RI	
	Average	Std	Average	Std	Average	Std	Average	Std
SPC	0.1056	0.0075	0.2928	0.0065	0.2222	0.0073	0.7319	0.0063
LRR	0.1125	0.0026	0.3146	0.0041	0.2473	0.0051	0.7353	0.0008
Co-Reg SPC	0.1055	0.0018	0.2701	0.0028	0.2317	0.0039	0.7414	0.0004
RMSC	0.1188	0.0063	0.2917	0.0197	0.2383	0.0204	0.7359	0.0051
LMSC	0.1418	0.0036	0.3289	0.0031	0.2709	0.0058	0.7426	0.0002
NSL2MSC	0.1521	0.0047	0.3442	0.0058	0.2838	0.0062	0.7569	0.0006

TABLE IV. Clustering Results on the BBCSport

Method	NMI		ACC		F-Measure		RI	
	Average	Std	Average	Std	Average	Std	Average	Std
SPC	0.7084	0.0058	0.8005	0.0332	0.7633	0.0035	0.8919	0.0008

LRR	0.6983	0.0022	0.7913	0.0035	0.7715	0.0028	0.8816	0.0015
Co-Reg SPC	0.7242	0.0008	0.7556	0.0057	0.7722	0.0013	0.8930	0.0006
RMSC	0.8133	0.0102	0.8428	0.0139	0.8772	0.0090	0.9265	0.0027
LMSC	0.8263	0.0071	0.9031	0.0052	0.8871	0.0071	0.9465	0.0006
NSL2MSC	0.8536	0.0063	0.9436	0.0067	0.9048	0.0066	0.9617	0.0011

Table 1 shows the clustering results on the MSRCV1 data set. On the MSRCV1 dataset, we can see that the proposed method is much better than the compared method. The reason behind this is that the latent representation space learned can make better use of multiple views of data.

Table 2 shows the clustering performance on the Extended YaleB dataset. The clustering performance of most algorithms on this dataset is poor, the main reason is that the illumination changes greatly in the dataset, which seriously affects the clustering performance. However, the NSL2MSC algorithm proposed in this paper has achieved better results. It is 7.95%, 4.9%, 11.47% and 12.96% higher than LMSC algorithm in NMI, ACC, F-measure and RI indexes, respectively.

Table 3 shows the clustering results on the still DB dataset. The clustering performance of each algorithm is not very good. However, from the four indicators, the NSL2MSC algorithm proposed in this paper has achieved relatively promising clustering results.

Table 4 shows the clustering results on BBCSport dataset. The clustering performance of NSL2MSC proposed in this paper is at least 2% higher than that of RMSC. Compared with LMSC method, our NSL2MSC method still achieves better result.

B. Parameters Effect

In the NSL2MSC algorithm model proposed in this paper, there are several regularization parameters. Next, the Extended YaleB data set is taken as an example to test the effent of the four parameters $\lambda_1, \lambda_2, \lambda_3$ and K on ACC and NMI. We change one parameter while fixing the other. Figure 1 shows the ACC and NMI results for different parameter settings on the Extended Yale B dataset. It calearly can be seen that the results show that in a large range of parameters, the NSL2MSC algorithm is superior to other algorithms.

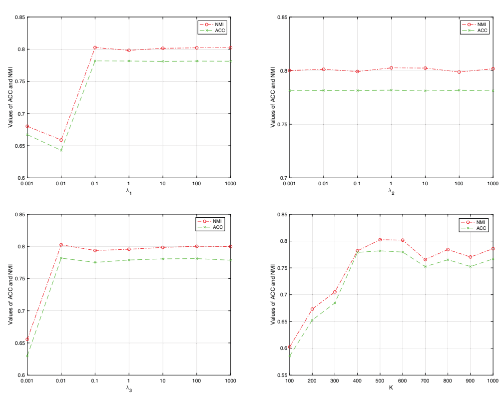

Figure 1. The parameter effect of $\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \boldsymbol{\lambda_3}$ and $\mathbf{K}$ on Extended Yale-B dataset.

V. Conclusion

In this paper, a new multi view subspace clustering method was proposed, which is called NSL2MSC. The main innovation of this algorithm lies in the use of graph regularization to make full use of the complementary information among views in the process of learning multi view latent subspace representation. In this way, this method can not only represent the global information of the data, but also can capture the local geometric structure information of the data. A large number of image clustering experiments show the effectiveness of this method.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 61902160, 61806088), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No.19KJB520006) and the foundation of Changzhou Science and Technology Plan (Applied Basic Research) (Grant No. CJ20190076).

References

[1] Gao, H., Nie, F., Li, X., & Huang, H. (2015). Multi-view Subspace Clustering. *international conference on computer vision*.

[2] Cao, X., Zhang, C., Fu, H., Liu, S., & Zhang, H. (2015). Diversity-induced Multi-view Subspace Clustering. *computer vision and pattern recognition*.

[3] Zhang, C., Hu, Q., Fu, H., Zhu, P., & Cao, X. (2017). Latent Multi-view Subspace Clustering. *computer vision and pattern recognition*.

[4] Zhang, C., Fu, H., Liu, S., Liu, G., & Cao, X. (2015). Low-Rank Tensor Constrained Multiview Subspace Clustering. *international conference on computer vision*.

[5] Patel, V. M., Van Nguyen, H., & Vidal, R. (2013). Latent Space Sparse Subspace Clustering. *international conference on computer vision*.

[6] Chung F R K, Graham F C. Spectral graph theory(1997). American Mathematical Soc..

[7] Lin, Z., Liu, R., & Su, Z. (2011). Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation. *neural information processing systems*.

[8] Xu, J., Han, J., & Nie, F. (2016). Discriminatively Embedded K-Means for Multi-view Clustering. *computer vision and pattern recognition*.

[9] Ikizler, N., Cinbis, R. G., Pehlivan, S., & Duygulu, P. (2008). Recognizing actions from still images. *international conference on pattern recognition*.

[10] Xia, R., Pan, Y., Du, L., & Yin, J. (2014). Robust multi-view spectral clustering via low-rank and sparse decomposition. *national conference on artificial intelligence*.

[11] Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., & Ma, Y. (2013). Robust Recovery of Subspace Structures by Low-Rank Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 171-184.

[12] Kumar, A., Rai, P., & Daume, H. (2011). Co-regularized Multi-view Spectral Clustering. *neural information processing systems*.