IEICE TRANS. INF. & SYST., VOL.E107–D, NO.1 JANUARY 2024

LETTER

Shared Latent Embedding Learning for Multi-View Subspace Clustering*

Zhaohu LIU[†], ***Nonmember***, **Peng SONG**[†a)], ***Member***, **Jinshuai MU**[†], ***and*** **Wenming ZHENG**[††], ***Nonmembers***

SUMMARY Most existing multi-view subspace clustering approaches only capture the inter-view similarities between different views and ignore the optimal local geometric structure of the original data. To this end, in this letter, we put forward a novel method named shared latent embedding learning for multi-view subspace clustering (SLE-MSC), which can efficiently capture a better latent space. To be specific, we introduce a pseudo-label constraint to capture the intra-view similarities within each view. Meanwhile, we utilize a novel optimal graph Laplacian to learn the consistent latent representation, in which the common manifold is considered as the optimal manifold to obtain a more reasonable local geometric structure. Comprehensive experimental results indicate the superiority and effectiveness of the proposed method.

key words: *multi-view learning, subspace clustering, latent representation, graph Laplacian*

1. Introduction

With the advent of information technologies, a large amount of multi-view data from different areas or multiple sources have rapidly emerged. Multi-view learning has attracted extensive attention in the field of machine learning [1]. In practice, an object is often pictured from distinct aspects [2]. For instance, given a picture, we might consider different features such as pixel, color, saturation, and brightness, which are different aspects of the picture, etc. In these situations, it is a big challenge to obtain accurate clustering results for multi-view clustering.

To tackle the above issue, over the past decade, a variety of multi-view subspace clustering methods have been presented. For instance, in [3], Zhang et al. present a latent multi-view subspace clustering (LMSC) approach, which aims to seek the latent representation to perform data reconstruction. In [4], Zhou et al. propose a dual learning method to mine the shared information and simultaneously maintain the specific property of each view. In [5], Niu et al. develop a one-step multi-view subspace clustering with incomplete views (OMVSC-IV) framework, which simultaneously performs the non-negative embedding and spectral embedding subspace clustering. In [6], Si et al. develop an exclusivity constraint to explore the diversity of different views. However, there exist some shortcomings in the above methods. They do not fully take into account the intra-view similarities of each view. Meanwhile, they do not pay attention to the optimal local geometric structure of the original data space.

To address the above-mentioned issues, in this letter, we present a new multi-view subspace clustering method, called shared latent embedding learning for multi-view subspace clustering (SLE-MSC). Specifically, we perform multi-view subspace clustering based on matrix factorization, and utilize the pseudo-labels to fully explore the intra-view similarities within each view. Moreover, we apply an optimal graph regularization strategy to preserve the local geometry of the original data space. The overall framework of SLE-MSC is demonstrated in Fig. 1.

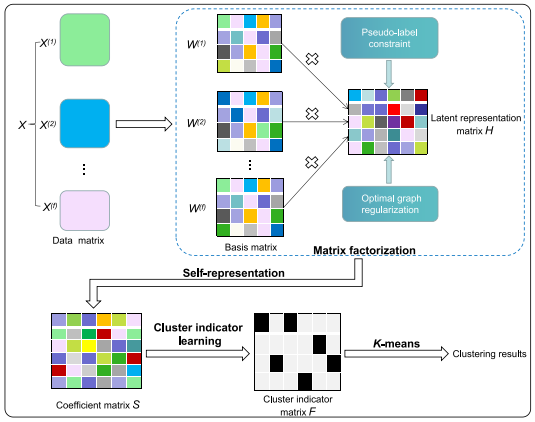


Fig. 1 The framework of SLE-MSC.

2. Proposed Method

Denote $X = [X^1; X^2; \ldots; X^m] \in \mathbb{R}^{\sum_{f=1}^{m} d^f \times n}$ as a set of multi-view data with m different views, where $X^f \in \mathbb{R}^{d^f \times n}$ represents the data matrix of the f-th view. In this work, we assume that there exists a consistent latent representation for multiple views, which can well describe the data in essence and mine the shared latent structure. Note that the kernel norm cannot be easily solved and thus we use the Frobenius norm instead of the kernel norm. The objective function is formulated as follows:

Manuscript received July 24, 2023.
Manuscript revised September 7, 2023.
Manuscript publicized October 17, 2023.

[†]The authors are with the School of Computer and Control Engineering, Yantai University, Yantai 264005, China.
[††]The author is with the Key Laboratory of Child Development and Learning Science, Ministry of Education (Southeast University), Southeast University, Nanjing 210096, China.

*This study was partially funded by the Natural Science Foundation of Shandong Province under Grants ZR2023MF063, the Fundamental Research Funds for the Central Universities, and the Graduate Innovation Foundation of Yantai University (GIFYTU).

a) E-mail: pengsong@ytu.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2023EDL8044

$$\min_{W,H,S} \|X - WH\|_F^2 + \lambda\|H - HS\|_F^2 + \gamma\|S\|_F^2 \quad (1)$$
$$\text{s.t. } \|W_{:,j}\|_2^2 \le 1,\ S \ge 0,\ diag(S) = 0,\ rank(L_s) = n - c$$

where $\|\cdot\|_F$ is the Frobenius norm, $H \in \mathbb{R}^{d\times n}$ is the consistent latent representation, and $W = [W^1; W^2; \ldots; W^m] \in \mathbb{R}^{\sum_{f=1}^m d^f \times d}$ is the multi-view basis matrix, in which $W^f \in \mathbb{R}^{d^f\times d}$ and $S \in \mathbb{R}^{n\times n}$ are the basis matrix of the f-th view and the self-representation matrix of H, respectively, and d^f and d are the corresponding dimensionalities, $diag(\cdot)$ represents the main diagonal elements of a matrix, and λ and γ are balancing parameters. $\|W_{:,j}\|_2^2 \le 1$ is used to prevent H from being too small when scaling, and $\|S\|_F^2$ is a regularization term to avoid overfitting. $rank(L_s) = n - c$ makes that the number of clusters c equals the number of the connected components in S. $L_s = D - \frac{S^T+S}{2}$ is the Laplace matrix of S, and $D \in \mathbb{R}^{n\times n}$ denotes a diagonal matrix with $D_{ii} = \sum_j^n \frac{S_{ij}+S_{ji}}{2}$.

Note that Eq. (1) aims to obtain the consistent latent representation of all views, the intra-view similarities within each view are often neglected. To cope with the above issue, we introduce a linear projection matrix of each view to project the consistent latent representation to the discriminative pseudo-label space. Thus, we can capture the intra-view similarities of the original data space. Equation (1) can be rewritten as follows:

$$\begin{aligned}\min_{W,H,S,Q^f} &\|X - WH\|_F^2 + \alpha\sum_{f=1}^m \|Y^f - Q^f H\|_F^2 \\ &+ \lambda\|H - HS\|_F^2 + \gamma\Psi(Q^f, S)\end{aligned} \quad (2)$$
$$\text{s.t. } \|W_{:,j}\|_2^2 \le 1,\ S \ge 0,\ diag(S) = 0,\ rank(L_s) = n - c$$

where $\Psi(Q^f, S)$=$\|Q^f\|_F^2 + \|S\|_F^2$, in which $\|\cdot\|_F^2$ is used to avoid overfitting, $Y^f \in \mathbb{R}^{c\times n}$ and $Q^f \in \mathbb{R}^{c\times d}$ are the pseudo-label matrix and the linear projection matrix of the f-th view, respectively, γ is a balancing parameter, and $\|Q^f\|_F^2$ is a regularization term to avoid overfitting. Note that Y^f can be computed by using clustering approaches on the data matrix of the f-th view.

To discover a more reasonable local geometric structure of data space, we introduce an optimal graph regularization, which treats a common manifold as the optimal manifold. The optimal graph regularization is formulated as follows:

$$\min_H Tr(HL_*H^T) \quad (3)$$

where $L_* = D_* - M_*$ denotes the optimal manifold. $D_* = \left\{d_{ii}^*\right\}^{n\times n}$ represents a diagonal matrix with $d_{ii}^* = \sum_{j=1}^n m_{ij}^*$, and M_* is the optimal similarity matrix, which can be defined by taking the minimum similarity value of all views as follows:

$$M_* = \left\{m_{ij}^*\right\}^{n\times n}, \quad m_{ij}^* = \min\left\{m_{ij}^f\right\}_{f=1..m} \quad (4)$$

Next, given a data matrix of the f-th view $X^f = [X_1^f, \ldots, X_n^f] \in \mathbb{R}^{d^f\times n}$, the adjacency matrix $M_f \in \mathbb{R}^{n\times n}$ is obtained by using the k-nearest neighbors (KNN) algorithm. Then, the similarity matrix $M_f = \left\{m_{ij}^f\right\}^{n\times n}$ is defined as follows:

$$m_{ij}^f = \begin{cases} sim_{ij}, & \text{if } X_i^f \in \mathcal{N}^k(X_j^f) \text{ or } X_j^f \in \mathcal{N}^k(X_i^f) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $\mathcal{N}^k(X_i^f)$ is a set of the k nearest neighbors of X_i^f, and sim_{ij} represents the similarity measurement between X_i^f and X_j^f. The similarity can be measured by different methods such as heat kernel, binary, cosine and dot-product [7]. For simplification, here we choose the binary strategy.

Combining Eqs. (2) and (3), the overall objective is written as

$$\begin{aligned}\min_{W,H,S,Q^f} &\|X - WH\|_F^2 + \alpha\sum_{f=1}^m \|Y^f - Q^f H\|_F^2 \\ &+ \beta Tr(HL_*H^T) + \lambda\|H - HS\|_F^2 + \gamma\Psi(Q^f, S) \\ \text{s.t. } &\|W_{:,j}\|_2^2 \le 1,\ S \ge 0,\ diag(S) = 0,\ rank(L_s) = n - c\end{aligned} \quad (6)$$

where α and β are balancing parameters controlling the pseudo-label constraint and the optimal graph regularization, respectively. According to the Ky Fan's Theorem [8], the final objective function of the proposed SLE-MSC is formulated as

$$\begin{aligned}\min_{W,H,S,Q^f,F} &\|X - WH\|_F^2 + \alpha\sum_{f=1}^m \|Y^f - Q^f H\|_F^2 \\ &+ \beta Tr(HL_*H^T) + \lambda\|H - HS\|_F^2 + \gamma\Psi(Q^f, S) + \theta Tr(F^T L_s F) \\ \text{s.t. } &\|W_{:,j}\|_2^2 \le 1,\ S \ge 0,\ diag(S) = 0,\ F^T F = I\end{aligned} \quad (7)$$

where $F \in \mathbb{R}^{n\times c}$ represents an indicator matrix.

To solve Eq. (7), we propose an iterative optimization scheme. The problem is divided into several sub-problems, and the optimization steps are listed as follows:

(1) Fix W, H, F and S, and update Q^f: the problem in Eq. (7) can be written as

$$\pounds^f = \min_{Q^f} \alpha\|Y^f - Q^f H\|_F^2 + \gamma\|Q^f\|_F^2 \quad (8)$$

Setting the partial derivative $\frac{\partial \pounds^f}{\partial Q^f} = 0$, we can obtain

$$Q^f = \alpha Y^f H^T(\alpha HH^T + \gamma I)^{-1} \quad (9)$$

(2) Fix Q^f, H, F and S, and update W: By introducing an auxiliary variable $T \in \mathbb{R}^{\sum_{f=1}^m d^f\times d}$, we can get the following objective function:

$$\min_{W,T} \|X - WH\|_F^2 \quad \text{s.t. } W = T,\ \|T_{:,j}\|_2^2 \le 1 \quad (10)$$

To address the above problem, we utilize the alternating direction method of multipliers (ADMM) algorithm [9], and

can get

$$\begin{cases} W^{r+1} = \arg\min_{W} \|X - WH\|_F^2 + \mu\|W - T^r + J^r\|_F^2 \\ T^{r+1} = \arg\min_{T} \mu\|W^{r+1} - T + J^r\|_F^2 \text{ s.t. } \|T_{:,j}\|_2^2 \leq 1 \\ J^{r+1} = J^r + W^{r+1} - T^{r+1}, \text{ update } \mu \text{ if appropriate} \end{cases} \tag{11}$$

where r is the number of iterations, and $J \in \mathbb{R}^{\sum_{f=1}^{m} d^f \times d}$ denotes an intermediate variable. As the ADMM algorithm has good convergence performance, W can converge rapidly in each optimization step.

(3) Fix Q^f, W, F and S, and update H: The optimization for H in Eq. (7) can be equivalently rewritten as

$$\begin{aligned} \min_{H} \|X - WH\|_F^2 + \alpha \sum_{f=1}^{m} \|Y^f - Q^f H\|_F^2 \\ + \beta tr(HL_* H^T) + \lambda\|H - HS\|_F^2 \end{aligned} \tag{12}$$

The above equation is a smooth and convex optimization problem that can be efficiently solved. By differentiating Eq. (12) with respect to H and setting it to zero, we can get the following optimal solution H^*:

$$P_1 * H^* + H^* * P_2 = P_3 \tag{13}$$

where $P_1 = W^T W + \alpha \sum_{f=1}^{m} Q^{f^T} Q^f$, $P_2 = \lambda(I-S)(I-S)^T + \beta L_*$, and $P_3 = W^T X + \alpha \sum_{f=1}^{m} Q^{f^T} Y^f$. As discussed in [10], the above problem in Eq. (13) has been successfully solved.

(4) Fix Q^f, W, F and H, and update S: We can get the following optimization problem:

$$\begin{aligned} &\min_{S} \|H - HS\|_F^2 + \frac{\gamma}{\lambda}\|S\|_F^2 + \frac{\theta}{\lambda} Tr(F^T L_s F) \\ &\text{s.t. } S \geq 0, \; diag(S) = 0 \end{aligned} \tag{14}$$

By introducing an intermediate variable $K \in \mathbb{R}^{n \times n} = H^T H$ and an auxiliary variable $Z \in \mathbb{R}^{n \times n}$, the problem can be equivalently rewritten as follows:

$$\begin{aligned} &\min_{S} Tr(K - 2KS + S^T KS) + \frac{\gamma}{\lambda}\|Z\|_F^2 + \frac{\theta}{\lambda} Tr(F^T L_s F) \\ &\text{s.t. } S \geq 0, \; Z = S, \; diag(S) = 0 \end{aligned} \tag{15}$$

By using the augmented Lagrange multiplier (ALM) algorithm [9], we can obtain the following augmented Lagrange function:

$$\begin{aligned} \mathfrak{R}(Z, S, \Phi) = Tr(K - 2KS + S^T KS) + \frac{\gamma}{\lambda}\|Z\|_F^2 \\ + \frac{\theta}{\lambda} Tr(F^T L_s F) + \frac{\delta}{2}\|Z - S + \frac{\Phi}{\delta}\|_F^2 \end{aligned} \tag{16}$$

where $\Phi \in \mathbb{R}^{n \times n}$ is a Lagrange multiplier, and $\delta > 0$ is a penalty parameter. We can fix other variables to minimize the above problem with respect to Z, S and Φ alternatively.

- By setting $P = S - \frac{\Phi}{\delta}$, we can update Z element-wisely as follows:

$$Z_{ij} = \max(|P_{ij}| - \frac{\gamma}{\lambda\delta}, 0) \cdot sign(P_{ij}) \tag{17}$$

- By setting $E = Z + \frac{\Phi}{\delta}$, we can update S element-wisely as follows:

$$\min_{S_i} S_i^T (\frac{\delta}{2} I + K) S_i + (\frac{\theta}{2\lambda} a_i^T - \delta E_i^T - 2k_{i,:}) S_i \tag{18}$$

  where $a_i \in \mathbb{R}^{n \times 1}$ represents a column vector with the j-th element $a_{ij} = \|F_{i,:} - F_{j,:}\|^2$. By setting the derivative of Eq. (18) with respect to S_i to zero, we can obtain S.

- We can update the variable Φ as follows:

$$\Phi = \Phi + \delta(Z - S) \tag{19}$$

(5) Fix Q^f, W, S and H, and update F: The optimization for F in Eq. (7) can be equivalently rewritten as

$$\min_{F} Tr(F^T L_s F) \quad \text{s.t. } F^T F = I \tag{20}$$

By utilizing the c eigenvectors of L_s corresponding to the c smallest eigenvalues, we can get the optimal solution of F, and update Q^f, W, H, S and F until convergence. Finally, we take the optimal F as the input and perform K-means to get the final clustering results.

3. Experiments

3.1 Experimental Setup

In our experiments, we utilize four multi-view datasets, including Yale [1], ORL [2], 3-Source [11], and HW [12]. Several conventional and state-of-the-art clustering methods are selected for comparison, including K-means, LMSC [3], AWNLRR [13], GMC [11], JSMC [14], and AGLMF [2]. To ensure fairness, we download all codes of the compared methods from the authors' homepages, and adjust the parameters according to the recommendations. For our method, we choose the values of α and β by searching the range of $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$.

In our experiments, three metrics in terms of accuracy (ACC), normalized mutual information (NMI), and Purity, are adopted, and the average clustering results and standard deviations are recorded by running 10 times.

3.2 Results and Discussions

The experimental results are shown in Table 1. From the table, we can have the observations as follows.

First, our algorithm achieves better clustering results than the baseline algorithms on all datasets, which manifests the superiority of our algorithm.

Second, multi-view clustering methods are more advantageous in comparison with single-view clustering methods on multi-view datasets. This indicates that the multi-view algorithms are more suitable for multi-view data in comparison with the single-view algorithms.

Last, compared with multi-view subspace approaches based on the latent representation, i.e., LMSC, the SLE-MSC algorithm gains much better clustering performance. The main reason is that LMSC only considers the inter-view similarities between different views to learn the consistency latent matrix. By contrast, our method fully considers the intra-view similarities within each view while considering inter-view similarities. Meanwhile, our method fully takes into account the local geometric structure of the original data space, which can help learn better consistent latent representation.

Moreover, we also conduct ablation experiments. We remove the influence of the pseudo-label constraint term, the local geometric structure of the original data space, and the matrix regularization term by setting $\alpha = 0$, $\beta = 0$, and $\gamma = 0$ in Eq. (7), and can get three corresponding variants of SLE-MSC, called SLE-MSC$_1$, SLE-MSC$_2$, and SLE-MSC$_3$. Figure 2 illustrates the comparison results. From the figure, we can observe that each regularization term plays a vital role in the proposed model.

Table 1 Comparison results of all methods on Yale, ORL, 3-Sources, and HW datasets (mean±standard deviation).

Datasets	Methods	ACC	NMI	Purity
Yale	K-means	59.33±1.23	63.14±1.37	61.33±1.75
	LMSC	62.48±3.55	64.70±2.80	62.90±3.38
	AWNLRR	63.33±0.90	64.45±0.49	64.42±0.54
	GMC	65.45±0.00	67.36±0.00	66.06±0.00
	JSMC	68.48±0.00	70.07±0.00	68.48±0.00
	AGLMF	70.30±0.00	71.77±0.00	70.30±0.00
	SLE-MSC	**73.94±0.00**	**76.53±0.00**	**73.94±0.00**
ORL	K-means	62.25±1.50	78.37±1.26	65.55±0.90
	LMSC	77.00±4.01	89.58±2.05	80.55±4.02
	AWNLRR	66.92±0.16	78.40±0.15	67.82±0.16
	GMC	65.25±0.00	81.66±0.00	72.50±0.00
	JSMC	77.50±0.00	88.03±0.00	80.00±0.00
	AGLMF	72.00±0.00	87.00±0.00	77.75±0.00
	SLE-MSC	**84.00±0.00**	**91.88±0.00**	**86.75±0.00**
3-Sources	K-means	53.85±3.55	27.52±5.14	59.47±4.44
	LMSC	56.92±6.45	49.44±2.76	71.77±2.82
	AWNLRR	75.14±0.26	59.82±0.18	75.14±0.26
	GMC	69.23±0.00	54.80±0.00	74.56±0.00
	JSMC	77.51±0.00	69.52±0.00	82.24±0.00
	AGLMF	69.82±0.00	64.14±0.00	78.10±0.00
	SLE-MSC	**83.43±0.00**	**76.84±0.00**	**86.98±0.00**
HW	K-means	75.18±5.15	70.73±1.86	76.32±3.38
	LMSC	83.72±4.19	76.61±3.14	83.75±4.16
	AWNLRR	61.10±0.05	62.04±0.04	64.85±0.05
	GMC	82.95±0.00	85.86±0.00	85.50±0.00
	JSMC	55.35±0.00	52.69±0.00	58.95±0.00
	AGLMF	83.75±0.00	86.74±0.00	86.40±0.00
	SLE-MSC	**92.35±0.00**	**87.10±0.00**	**92.35±0.00**

3.3 Parameter Sensitivity Analysis

In this section, we also conduct experiments to analyze the parameter sensitivity of the proposed method, i.e., α, β, λ, and γ. The experimental results are shown in Fig. 3. As shown in the figure, for the parameters α and β, we set the parameter values in the range of $10^{-3} \sim 10^{3}$. For the parameters λ and γ, we set different ranges on different

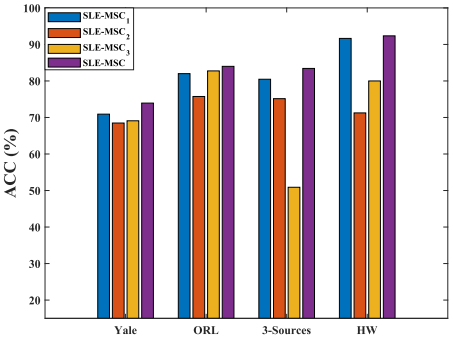
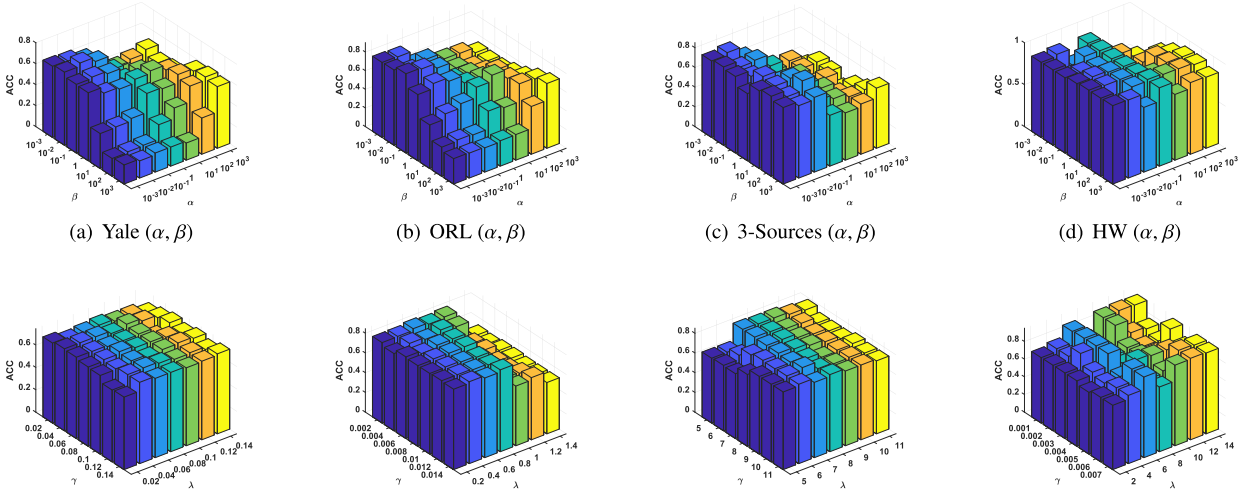


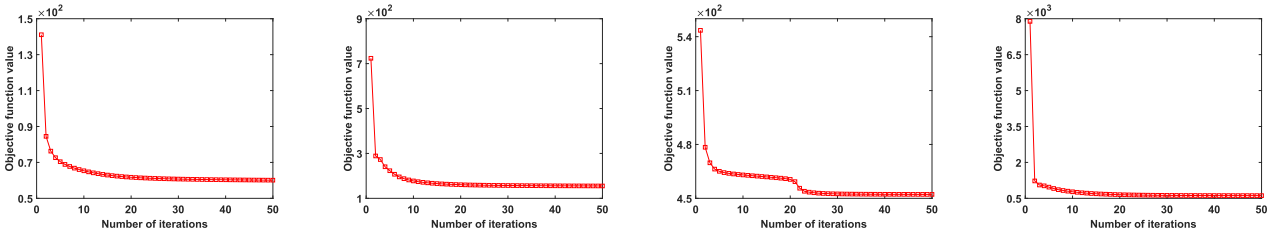
Fig. 2 Clustering results of SLE-MSC and its special cases.



(a) Yale (α, β) (b) ORL (α, β) (c) 3-Sources (α, β) (d) HW (α, β)

(e) Yale (λ, γ) (f) ORL (λ, γ) (g) 3-Sources (λ, γ) (h) HW (λ, γ)

Fig. 3 Parameter sensitivity analysis of the proposed method in terms of ACC results, i.e., α, β, λ and γ.



(a) Yale (b) ORL (c) 3-Sources (d) HW

Fig. 4 The convergence analysis of SLE-MSC on all datasets.

datasets. For example, λ and γ are both set as $0.02 \sim 0.14$ on the Yale dataset. On the ORL dataset, λ and γ are set as $0.2 \sim 1.4$ and $0.002 \sim 0.014$, respectively. On the 3-Sources dataset, λ and γ are both set to $5 \sim 11$. On the HW dataset, λ and γ are set to $2 \sim 14$ and $0.001 \sim 0.007$, respectively. From these results, we can conclude that the α, β are less sensitive to different values, and λ and γ are relatively sensitive to different values.

3.4 Convergence Analysis

Figure 4 depicts the convergence results of the proposed algorithm. In the figure, the objective value is represented by the vertical axis, and the number of iterations is represented by the horizontal axis. As can be seen from the figure, with the number of iterations increasing, the objective values decline steadily, and can converge within about 25 iterations.

3.5 Complexity Analysis

For the optimization of our method, it can be divided into five parts. For updating Q^f, the time complexity is $O(mt(cnd + nd^2))$. For updating W, the time complexity is $O(t((\sum_{f=1}^{m} d^f)^2 d))$. For updating H, the time complexity is $O(t(d^3))$. For updating S, the time complexity is $O(t(n^3))$. For updating F, the time complexity is $O(t(cn^2))$. Note that we initialize the pseudo-label matrix Y^f by performing K-means on the feature matrix of each view. Thus the time complexity of this process is $O(mcd'nT')$, where d' denotes the maximum dimension of all views, and T' denotes the number of iteration of matrix factorization and clustering. Its time complexity can be ignored compared with the other five parts. We normally set $c < d$ and $c < n$ in this paper. To sum up, the overall time complexity is $O(mtnd^2 + t((\sum_{f=1}^{m} d^f)^2 d + n^3 + d^3))$, where t is the number of iterations.

4. Conclusion

In this letter, we have proposed a novel multi-view clustering framework, called shared latent embedding learning for multi-view subspace clustering (SLE-MSC). Our model integrates the pseudo-label learning and optimal graph learning into the latent representation based multi-view subspace clustering framework, which can simultaneously mine the intra-view similarities of each view and the optimal local geometric structure of the data space. Moreover, a novel iterative optimization scheme is developed to solve SLE-MSC. Extensive experiments on several datasets validate the superiority and effectiveness of our model.

References

[1] M.-S. Chen, J.-Q. Lin, X.-L. Li, B.-Y. Liu, C.-D. Wang, D. Huang, and J.-H. Lai, "Representation learning in multi-view clustering: A literature review," Data Science and Engineering, vol.7, no.3, pp.225–241, 2022.

[2] W. Yang, Y. Wang, C. Tang, H. Tong, A. Wei, and X. Wu, "One step multi-view spectral clustering via joint adaptive graph learning and matrix factorization," Neurocomputing, vol.524, pp.95–105, 2023.

[3] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.4279–4287, 2017.

[4] T. Zhou, C. Zhang, X. Peng, H. Bhaskar, and J. Yang, "Dual shared-specific multiview subspace clustering," IEEE Transactions on Cybernetics, vol.50, no.8, pp.3517–3530, 2020.

[5] G. Niu, Y. Yang, and L. Sun, "One-step multi-view subspace clustering with incomplete views," Neurocomputing, vol.438, pp.290–301, 2021.

[6] X. Si, Q. Yin, X. Zhao, and L. Yao, "Consistent and diverse multi-view subspace clustering with structure constraint," Pattern Recognition, vol.121, p.108196, 2022.

[7] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Neural Computation, vol.15, no.6, pp.1373–1396, 2003.

[8] K. Fan, "On a theorem of Weyl concerning eigenvalues of linear transformations i," Proceedings of the National Academy of Sciences of the United States of America, vol.35, no.11, pp.652–655, 1949.

[9] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," Advances in Neural Information Processing Systems, vol.27, pp.1–9, 2014.

[10] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.586–594, 2015.

[11] H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based multi-view clustering," IEEE Transactions on Knowledge and Data Engineering, vol.32, no.6, pp.1116–1129, 2020.

[12] X. Liu, P. Song, C. Sheng, and W. Zhang, "Robust multi-view nonnegative matrix factorization for clustering," Digital Signal Processing, vol.123, p.103447, 2022.

[13] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted nonnegative low-rank representation," Pattern Recognition, vol.81, pp.326–340, 2018.

[14] X. Cai, D. Huang, G.-Y. Zhang, and C.-D. Wang, "Seeking commonness and inconsistencies: A jointly smoothed approach to multi-view subspace clustering," Information Fusion, vol.91, pp.364–375, 2023.