Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)

doi:10.3233/FAIA220020

Multi-View Subspace Clustering by Combining $\ell_{2,p}$- Norm and Multi-Rank Minimization of Tensors

Zhaoxian Deng[a] and Zhiqiang Zeng[a,1]
[a]*College of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China*

Abstract. In this article, based on the self-represented multi-view subspace clustering framework, we propose a new clustering model. Based on the assumption that different features can be linearly represented by data mapped to different subspaces, multiview subspace learning methods take advantage of the complementary and consensus informations between various kind of views of the data can boost the clustering performance. We search for the tensor with the lowest rank and then extract the frontal slice of it to establish a well-structured affinity matrix. Based on the tensor singular value decomposition (t-SVD), our low-rank constraint can be achieved. We impose the $\ell_{2,p}$-norm to flexibly control the sparsity of the error matrix, making it more robust to noise, which will enhance the robustness of our clustering model. With combining $\ell_{2,p}$-norm and tensor multi-rank minimization, the proposed Multi-view Subspace Clustering(MVSC) model can effectively perform clustering with multiple data resources. We test our model on one real-world spoon dataset and several publicly availabe datasets. Extensive evaluation methods have proved that our model is effective and efficient.

Keywords. Multi-view features, Subspace clustering, $\ell_{2,p}$-norm, t-SVD

1. Introduction

In daily life, for a specific object, we can usually recognize it from multiple perspectives. For example, we can identify a person from his appearance, voice, fingerprints, iris etc.[1]; a picture can be represented by different features, such as harris, lpb and gabor; We can describe a cat together with pictures and text descriptions. Generally, multi-view data has two characteristics: consistency and complementarity. The consistency means that data from different perspectives are used to represent the same object; the complementarity means that data from a specific perspective has characteristics that other perspectives do not have. Therefore, when improving clustering performance, complementary information from multiple views provides more contribution than single-view data.

Xu et al. believe that collaborative training, graph-based learning methods, and subspace learning constitute the three main categories of multi-perspective learning[1]. Multi-view data provides more information for clustering tasks. In this article, we mainly study the problem of MVSC, which aims to divide unlabeled multi-view data into various

[1] Corresponding Author. E-mail: zqzeng@xmut.edu.cn

realistic clusters [2], and ensure that the distances of samples within the same cluster are close to each other, while the samples in different clusters are far from each other. Subspace clustering assumes that different clusters have different low-dimensional subspaces, and high-dimensional data can be reconstructed from data existing in subspaces. In other words, subspace clustering can reduce the dimensionality of high-dimensional data and obtain clustering results at the same time. Sparse subspace clustering (SSC) successfully uses sparse subspace representation to solve the coefficient matrix, and the clustering results will be outputted when the affinity matrix was inputted into the spectral clustering[3]. Low-rank representation (LRR) effectively enforces a lowest rank on the representation matrix so that the subspace to which the data belongs will be recovered and the noise of the data will be removed[4]. Although the above two single-view clustering algorithms have achieved good clustering effect, data usually come from different perspectives or fields in practical application. Therefore, based on self-representation, many researchers proposed MVSC algorithm. Cao et al. [5] use Hilbert-Schmidt Independence criterion to obtain the enhanced complementary information extracted form multiple views of the data and improve the clustering performance. Zhang et al. [6] stack the subspace representation matrix into a tensor and impose low-rank constraints on the tensor to explore the high-order correlation between multi-perspective data. However, the tensor nuclear norm(TNN) cannot satisfy the requirements of Tucker rank and ℓ_1-norm for compact convex relaxation[2], which is a simple rank sum norm, lacking clear physical meaning [7]. Lu et al. [8] extended matrix singular value decomposition to higher-order tensors and proposed tensor nuclear norm which can be effectively solve by t-SVD. Xie et al. [7] project the tensors into the low-dimensional matrix space and then obtain the minimized nuclear norm of the tensors through the t-product operation. A well-structured subspace representation matrix is obtained through the constraints of low-rank tensors, which can form a robust affinity matrix for spectral clustering[7].

The majority of the above MVSC algorithms use $\ell_{2,1}$ regularizers to constrain the noise matrix of each view. However, as indicated in recent studies [6,7], the $\ell_{2,1}$- norm is sample-specific, which may assume data is represented in Laplacian distribution [6]. The $\ell_{2,1}$- norm is an extension of ℓ_1-norm, andℓ_1-norm is a convex relaxation ofℓ_0-norm. Although convex relaxation problems are easy to obtain optimal solutions, such solutions may be suboptimal. To cope with this problem, inspired by this paper Xie et al. [7], we introduce a more robust and sparse solution with the $\ell_{2,p}$-norm for clustering [10].By controlling the sparsity of subspace representation matrix with a flexibale parameter, our model can flexibly deal with data in different distributions for various applications.

2. Related Work

Subspace clustering algorithms, both single-view and multi-view, will be briefly reviewed in this section. Some necessary notations are also introduced for better understanding .of our method.

2.1. Notations

X is indicated in bold uppercase, which is defined as a matrix in this article. As a vectors, **x** is represented by bold lowercase. Tensor are written as bold calligraphy letters,

e.g., $\mathcal{X}$. The formulation of the $\ell_{2,P}$-norm is to effectively constraint the out of samples of $X \in R^{n\times m}$:

$$\|X\|_{2,p} = \left(\sum_{i=1}^{n}\left(\sum_{j=1}^{m} m_{ij}^2\right)^{\frac{2}{p}}\right)^{\frac{1}{p}} \tag{1}$$

where p is a variable value, and the $\ell_{2,p}$-norm with different p values can constrain different types of noise.

2.2. Single-view subspace clustering

Affinity matrix or similarity matrix can well represent the relationship between data. Clustering results can be obtained by applying clustering method on affinity matrix. Therefore, the construction of affinity matrix directly affects the clustering performance. Self-representation based subspace learning believes that high-dimensional original data can be reconstructed from data existing in low-dimensional subspaces. Some studies [12,13,14,15] use affinity matrix, which can reveal the similarities between data, to achieved advanced clustering performance. The general clustering model of single view subspace based on self-representation is as follows:

$$\min_{Z,E} \|E\|_{\ell} + \lambda\Omega(Z) \qquad \text{s.t. X=XZ+E} \tag{2}$$

Where $X \in R^{d\times n}$ is a data matrix in which the n data samples are represented by each column. The subspace representation matrix is represented by **Z**, **E** is used to constrain sample noise; $\|\cdot\|_{\ell}$ is a proper norm, $\Omega(Z)$ is a regularizer, λis a parameter to balance error losses and regularization terms. In order to obtain the local relationship between the data, SSC[3] uses ℓ_1-norm constraint **Z**, while LRR[4] uses $\ell_{2,1}$-norm constraint **E** to remove the specific noise of samples. Although these two single-view subspace clustering algorithms have achieved promising performance, the clustering performance cannot be further improved due to the inability to effectively utilize the complementary information among multi-view data.

2.3. Multi-view subspace clustering

Each view of data from multiple perspectives has information specific to other views, which can not be ignored for the improvement of clustering performance. Making full use of these information is one of the characteristics of MVSC algorithm. Most MVSC algorithms can be represented by the following model:

$$\min_{Z^{(v)},E^{(v)}} \sum_{v=1}^{m}\left\|E^{(v)}\right\|_{\ell} + \lambda\Omega\left(Z^{(v)}\right) \qquad \text{s.t. } X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}. \tag{3}$$

Here $\mathbf{X}^{(v)}$ represents the data matrix of the vth view. $\mathbf{Z}^{(v)}$ and $\mathbf{E}^{(v)}$ are the corresponding representation matrix and error matrix of $\mathbf{X}^{(v)}$, respectively. m represents

the number of views. Applying different constraints on the first and second terms in Formula (3) will result in different multi-view clustering methods. Belhumeur et al. [16] obtain better complementary information by applying the Hilbert-Schmidt criterion on different $\mathbf{Z}^{(v)}$. By imposing the nuclear norm on the unfolding tensor and forcing the error matrix with the $\ell_{2,1}$-norm, the high-order relationship between different views has been obtained by Zhang et al. [6]. However, the requirements of Tucker rank and ℓ_1-norm for compact convex relaxation are not satisfied by this unfolding-based rank constraint, namely the tensor nuclear norm. To make the above problem effectively solved, By constraining the tensor nuclear norm based on t-SVD, Xie[7] et al ensure consistent information between various kind of views of the data..

3. The Proposed Method

Inspired by t-SVD-MSC, we impose the $\ell_{2,p}$-norm on the subspace representation matrix to reduce the influence of noise and increase the model's flexibility.

3.1. Formulation

The formulation of our proposed method is:

$$\begin{aligned}&\min_{E^{(v)},Z^{(v)}} \lambda\|E\|_{2,p} + \|\mathcal{Z}\|_{\circledast}\\ &\text{s.t. } X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)},\ \text{v=1,...,V},\\ &\mathcal{Z} \quad \phi\left(\mathrm{Z}^{(\mathrm{v})},\mathrm{Z}^{(\mathrm{v})},\ldots,\mathrm{Z}^{(\mathrm{v})}\right),\\ &\mathrm{E}=[\mathrm{E}^{(1)};\mathrm{E}^{(2)};\ldots;\mathrm{E}^{(\mathrm{V})}],\end{aligned} \tag{4}$$

where $\phi(\cdot)$ represents the stacking operation. A 3-order tensor $\mathcal{Z}$ will be stacked by diverse $\mathbf{Z}^{(v)}$. $\|\quad\|_{\circledast}$ is the nuclear norm of $\mathcal{Z}$, which is optimized based on the t-SVD. For a comprehensive review of the nuclear norm and t-SVD, please refer to Xie et al.[7]. The concatenated error matrix can be effectively represented by the $E=[E^{(1)};E^{(2)};...,E^{(V)}]$. In every view, E connects vertically along the $E^{(v)}$ column[4]. We can flexibly control the sparsity of $\mathbf{E}$ and obtain the best representation $\mathcal{Z}$ by optimizing the above objective function.

3.2. Optimization

To optimize the formulation (4), inspired by the variable-splitting technique, we make the tensor $\mathcal{Z}$ separable by introducing an auxiliary tensor variable $\mathcal{Q}$ **T**hen, we can rewrite our model as follows:

$$\begin{aligned}&\mathcal{L}\left(\mathrm{Z}^{(\mathrm{v})},\ldots,\mathrm{Z}^{(\mathrm{V})};\mathrm{E}^{(\mathrm{v})},\ldots,\mathrm{E}^{(\mathrm{V})};\mathcal{Q}\right)\\ &=\lambda\|\mathrm{E}\|_{2,\mathrm{p}} + \left\|\mathcal{Q}\right\|_{\circledast} + \sum_{\mathrm{v}=1}^{\mathrm{V}}\left(\langle \mathrm{D}_{\mathrm{v}}, \mathrm{X}^{(\mathrm{v})} - \mathrm{X}^{(\mathrm{v})}\mathrm{Z}^{(\mathrm{v})} - \mathrm{E}^{(\mathrm{v})}\rangle\right.\\ &\left.+\frac{\mu}{2}\left\|\mathrm{X}^{(\mathrm{v})} - \mathrm{X}^{(\mathrm{v})}\mathrm{Z}^{(\mathrm{v})} - \mathrm{E}^{(\mathrm{v})}\right\|_{\mathrm{F}}^{2}\right) \quad +\langle\mathcal{M},\mathcal{Z}-\mathcal{Q}\rangle\end{aligned}$$

$$+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{Q}\|_F^2, \tag{5}$$

here the two Lagrange multipliers are effectively represented by $\mathbf{D}_v$, $\mathcal{M}$, respectively, and the two penalty parameters are represented by μ and ρ.

Problem (5) can be converted to solving the following three sub-problems:

1) $Z^{(v)}$-Subproblem (with the other two variables are fixed):

$$\begin{aligned}&\min_{Z^{(v)}}\langle D_v, X^{(v)}-X^{(v)}Z^{(v)}-E^{(v)}\rangle+\frac{\mu}{2}\left\|X^{(v)}-X^{(v)}Z^{(v)}-E^{(v)}\right\|_F^2\\&+\langle M^{(v)}, Z^{(v)}-Q^{(v)}\rangle+\frac{\rho}{2}\left\|Z^{(v)}-Q^{(v)}\right\|_F^2\end{aligned} \tag{6}$$

Let the derivative of formulation (6) be zero, then the optimal solution of $\mathbf{Z}^{(v)}$ will be obtained:

$$\begin{aligned}Z^{(v)^*}=&\left(I+\frac{\mu}{\rho}X^{(v)^T}X^{(v)}\right)^{-1}\left(\left(X^{(v)^T}D_v+\mu X^{(v)^T}X^{(v)}\right.\right.\\&\left.\left.-\mu X^{(v)^T}E^{(v)}-M^{(v)}\right)/\rho+Q^{(v)}\right)\end{aligned} \tag{7}$$

2) $\mathbf{E}^{(v)}$-Subproblem (with variables $\mathcal{Z}$ and $\mathcal{Q}$ are fixed):

$$\begin{aligned}\mathrm{E}^*=&\ \mathrm{argmin}_{\mathrm{E}}\lambda\|\mathrm{E}\|_{2,\mathrm{p}}+\sum_{\mathrm{v}=1}^{\mathrm{V}}\left(\langle \mathrm{D_v}, \mathrm{X^{(v)}}-\mathrm{X^{(v)}Z^{(v)}}-\mathrm{E^{(v)}}\rangle\right.\\&\left.+\frac{\mu}{2}\left\|\mathrm{X^{(v)}}-\mathrm{X^{(v)}Z^{(v)}}-\mathrm{E^{(v)}}\right\|_{\mathrm{F}}^2\right)\\=&\ \mathrm{argmin}_{\mathrm{E}}\frac{\lambda}{\mu}\|\mathrm{E}\|_{2,\mathrm{p}}+\frac{1}{2}\|\mathrm{E}-\mathrm{B}\|_{\mathrm{F}}^2\\=&\ \mathrm{arg}min_E\sum_{j=1}^{n}\left(\frac{\lambda}{\mu}\left\|e_j\right\|_2^p+\frac{1}{2}\left\|e_j-b_j\right\|_2^2\right)\end{aligned} \tag{8}$$

where $\mathbf{B}$ is obtained by connecting $\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}+(1/\mu)\mathbf{D}^{(v)}$ end to end.

Therefore, equation (8) can be alternatively optimized by solving the following n independent subproblems :

$$\begin{aligned}&\left\|e_j^*\right\|_2=arg\min_{e_j}\lambda\left\|e_j\right\|_2^p+\frac{1}{2}\left\|e_j-b_j\right\|_2^2,\\&j=1,2,\ldots,n.\end{aligned} \tag{9}$$

The optimal solution of problem (9):

$$e_j^*=\left|e_j\right|_2^*\left(\frac{b_j}{\|b_j\|_2}\right) \tag{10}$$

Where $\left|e_j\right|_2^*$ is the optimal solution of the following formulation:

$$\left|e_j\right|_2 = arg\min_{\left|e_j\right|_2} \lambda\left|e_j\right|_2^p + \frac{1}{2}\left(\left|e_j\right|_2 - \left|b_j\right|_2\right)^2 \tag{11}$$

Subproblem (11) is also a standard ℓ_p problem,which can be solved by using the methods in [17].

3) Q-Subproblem (with variables $\mathcal{Z}$ and **E** are fixed):

$$\mathcal{Q}^* = argmin_{\mathcal{Q}}\|\mathcal{Q}\|_{\circledast} + \frac{\rho}{2}\left\|\mathcal{Q} - \left(\mathcal{Z} + \frac{1}{\rho}\mathcal{M}\right)\right\|_F^2 \tag{12}$$

The optimization procedure of this problem can refer to the Algorithm 2 of Xie et al.[7].

The two Lagrange multipliers $\mathbf{D}_v$ and $\mathcal{M}$can be updated as follow:

$$D_v^* = D_v + \mu\left(X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\right), \tag{13}$$

$$\mathcal{M}^* = \mathcal{M} + \rho(\mathcal{Z} - \mathcal{Q}) \tag{14}$$

Finally, we described the optimization procedure of the $\ell_{2,p}$-norm based MVSC method in Algorithm 1.

Algorithm 1: the optimization algorithm to solve problem (4)

Input:K is the cluster number, Mutil-view data matrices:$\{X^{(v)}\}$ and λ

Initialized $\mathbf{Z^{(v)}} = 0, \mathbf{E^{(v)}} = 0, \mathbf{D_v} = 0, \mathbf{i} = 1, \ldots, V;$ $\quad = \quad = 0;$

While not converge **do**

  for v = 1 : V **do**

  Update $Z^{(v)}$ by using (7);

  end

  Update E by using (8);

  for v = 1 : V **do**

  Update D_v by using (13);

  end

  Obtain $\mathcal{Z} = \phi\left(Z^{(1)}, Z^{(2)}, \ldots, Z^{(V)}\right)$;

  Update $\mathcal{Q}$ by using (12);

  Update $\mathcal{M}$ by using (14);

  Update parameters μ and ρ; $\mu = \min(\eta\mu, \mu_{max}), \rho = \min(\eta\rho, \rho_{max})$;

  $\left(Q^{(1)}, \ldots, Q^{(v)}\right) = \phi^{-1}(\mathcal{Q})$;

  Check the convergence conditions:

  $\left\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\right\|_\infty < \varepsilon$ and $\left\|Z^{(v)} - Q^{(v)}\right\|_\infty < \varepsilon$;

end

Obtain the affinity matrix by

  $L = \frac{1}{V}\sum_{v=1}^{V}\left|Z^{(v)}\right| + \left|Z^{(v)^T}\right|$;

Output: Spectral clustering algorithm is run on L to obtain clustering result C .

4. Experiment

4.1. Experiment settings

In this section, by conducting experiments on four image datasets. We verify the effectiveness of our model. We describe these datasets in Table1.

Table 1. The real world datasets.

Dataset	Instances	Views	Clusters
ORL	400	3	40
YaleB	650	3	10
Yale	165	3	15
Spoon	240	3	3

There are 40 individuals in ORL dataset, and each individual contains 10 pictures taken from different conditions. For example, facial expressions, facial details, times and lighting.

There are 38 distinct subjects in YaleB dataset. With different illumination, 64 near frontal images has been captured in each subject. In our experiment, we take the first 10 individuals(650 images).

The Yale data set consists of 165 different grayscale images, each of which is a group of 15 images. Each group of photos represents a different facial expression or shape.

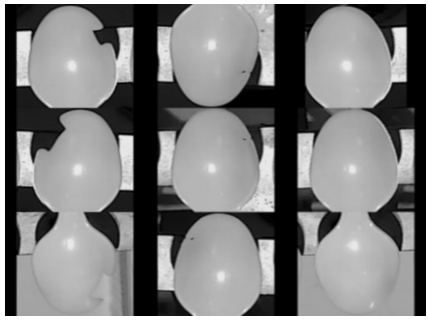
The Spoon data set is collected by ourselves. There are a total of 20 spoons with varying degrees of defect, and each 80 is a group. Some samples are shown in Fig. 1.

For the first three face dataset,we extract the intensity, LBP[19] and Gabor[20] feature. As for the Spoon dataset, Harris ,LBP and Sift feature are extracted.

We choose the t-SVD-MSC algorithms as baselines: the competitor solve the TNN by using t-SVD so that the low-rank tensor can be obtained,which ensure the consensus between different views of the data.

Two evaluation metrics will be used to assess the clustering performances[18]: Normalized Mutual Information(NMI), Accuracy(ACC). For these metrics, higher value ind-icates better performance.

On ORL dataset,we set λ=0.1 and p=0.2 in our proposed method. We set λ=0.1 and p=1.9 on YaleB dataset. On Yale dataset, λ=1.1 and p=0.8. As for the spoon dataset, we set λ=0.1 and p=0.7 in our method. For the t-SVD-MSC, the settings of the parameters will follow the experiments in its papers[7]. All experiments were implemented in Matlab 2020b on an lntel(R) Core(TM) i5-8300H CPU @ 2.30GHz with 16G memory and NVIDIA GeForce GTX 1050 Ti GPU.



(a) (b) (c)

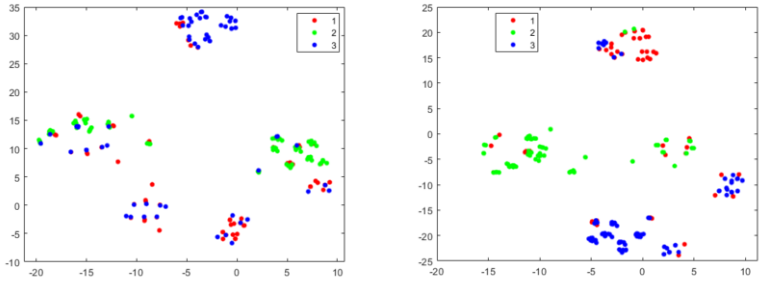
Figure 1. Samples from Spoon dataset. (a) has a significant gap, (b) with small defects and (c) is the complete one.

Table 2. Clustering results of compared algorithms (ACC%).

	ORL	YaleB	Yale	Spoon
t-SVD-MSC	96.6±0.30	55.6±0.10	87.2±1.30	61.7±0.00
our work	96.7±0.40	55.6±0.50	88.1±1.00	62.5±0.00

Table 3. Clustering results of compared algorithms (NMI%).

	ORL	YaleB	Yale	Spoon
t-SVD-MSC	99.3±0.10	59.8±0.20	89.8±1.40	30.9±0.00
our work	99.3±0.10	59.4±0.60	90.2±1.60	31.9±0.00



(a) **(b)**

Figure 2. Clustering results of our method. (a) Original data. (b) Clustering results on Spoon dataset.

4.2. Experiment results

Table 2 show the clustering results on the four datasets, which are obtain by using different methods. Our method has achieved a higher metrics than t-SVD-MSC on almost all datasets, which shows that our model is more robust to noise by controlling the value of p. It is worth noting that on the ORL data set, our method only improved a little bit over t-SVD-MSC on ACC, while other metrics were the same. The visual clustering results on the spoon dataset are shown in Figure 2, which show that our method can effectively cluster the real-world spoon defect dataset.

5. Conclusion

We construct a new model to effectively solve the multi-view clustering problem by imposing the nuclear norm and $\ell_{2,p}$-norm on the representation tensor and the vertical concatenation error matrix respectively. The$\ell_{2,p}$-norm regularizer can fit the variety of sparsity requirements via a flexiable parameter. At the same time, the introduction of tensor nuclear norm allows us to better probe the high-order relationship between multiple views of data. An efficient algorithm has been proposed to optimize our model. Extensive evaluation methods have proved that our model is effective and efficient on one real-world spoon defect dataset and three publicly available datasets.

Acknowledgments

This paper was supported by National Natural Science Foundation of China (Grant Nos.61871464), National Natural Science Foundation of Fujian Province (Grant Nos.2020J01266, 2021J011186), the "Climbing" Program of XMUT (Grant No.XPDKT20031), Scientific Research Fund of Fujian Provincial Education Department (Grant No. JAT200486), Program of XMUT for high-Level talents introduction plan (Grant No.YKJ19003R).

References

[1] Xu, Chang, Dacheng Tao, and Chao Xu. A survey on multi-view learning. arXiv preprint arXiv:1304.5634 (2013).
[2] Xia, Wei, et al. Multiview Subspace Clustering by an Enhanced Tensor Nuclear Norm. IEEE Transactions on Cybernetics (2021).
[3] Elhamifar, Ehsan, and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. IEEE transactions on pattern analysis and machine intelligence 35.11 (2013): 2765-2781.
[4] Liu, Guangcan, et al. Robust recovery of subspace structures by low-rank representation. IEEE transactions on pattern analysis and machine intelligence 35.1 (2012): 171-184.
[5] Cao, Xiaochun, et al. Diversity-induced multi-view subspace clustering. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[6] Zhang, Changqing, et al. Low-rank tensor constrained multiview subspace clustering. Proceedings of the IEEE international conference on computer vision. 2015.
[7] Xie, Yuan, et al. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. International Journal of Computer Vision 126.11 (2018): 1157-1179.
[8] Lu, Canyi, et al. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[9] Xie, Yuan, et al. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. International Journal of Computer Vision 126.11 (2018): 1157-1179.
[10] Tao, Hong, et al. Effective discriminative feature selection with nontrivial solution. IEEE transactions on neural networks and learning systems 27.4 (2015): 796-808.
[11] Yan, Yan, et al. GLocal tells you more: coupling GLocal structural for feature selection with sparsity for image and video classification. Computer Vision and Image Understanding 124 (2014): 99-109.
[12] Wang, Shusen, et al. Efficient subspace segmentation via quadratic programming. Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011.
[13] Luo, Dijun, et al. Multi-subspace representation and discovery. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2011.
[14] Lu, Canyi, et al. Correlation adaptive subspace segmentation by trace lasso. Proceedings of the IEEE international conference on computer vision. 2013.
[15] Lu, Can-Yi, et al. Robust and efficient subspace segmentation via least squares regression. European conference on computer vision. Springer, Berlin, Heidelberg, 2012.
[16] Belhumeur, Peter N., Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on pattern analysis and machine intelligence 19.7 (1997): 711-720.
[17] Nie, Feiping, et al. Robust matrix completion via joint schatten p-norm and lp-norm minimization. 2012 IEEE 12th International Conference on Data Mining. IEEE, 2012.
[18] Schütze, Hinrich, Christopher D. Manning, and Prabhakar Raghavan. Introduction to information retrieval. Vol. 39. Cambridge: Cambridge University Press, 2008.
[19] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24.7 (2002): 971-987.
[20] Lades, Martin, et al. Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on computers 42.3 (1993): 300-311.