ROBUST RANK CONSTRAINED SPARSE LEARNING: A GRAPH-BASED METHOD FOR CLUSTERING

Ran Liu, Mulin Chen, Qi Wang[*]*, Xuelong Li*

School of Computer Science and Center for OPTical IMagery Analysis and Learning(OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China

ABSTRACT

Graph-based clustering is an advanced clustering techniuqe, which partitions the data according to an affinity graph. However, the graph quality affects the clustering results to a large extent, and it is difficult to construct a graph with high quality, especially for data with noises and outliers. To solve this problem, a robust rank constrained sparse learning method is proposed in this paper. The $L_{2,1}$-norm objective function of sparse representation is introduced to learn the optimal graph with robustness. To preserve the data structure, the graph is searched within the neighborhood of the initial graph. By incorporating a rank constraint, the learned graph can be directly used as the cluster indicator and the final results is obtained without additional post-processing. Plenty of experiments on real-world data sets have proved the superiority and the robustness of the proposed approach.

Index Terms— Unsupervised Learning, Clustering, Graph-Based Clustering, Sparse Representation, ALM

1. INTRODUCTION

With the development of information technology, the large number of unlabeled data have greatly promoted the development of an important research topic in machine learning, clustering [1]. Among numerous clustering techniques, graph-based clustering has achieved dominant performance and can mine the internal structure of data, so it is widely used in practice, such as document clustering and image segmentation [2]. Graph-based clustering simulates data as spatial points, that is, constructs a data similarity graph based on the data relationship, and then accomplishes the clustering task by graph-theoretic optimization. The quality of similarity graphs largely determines the quality of clustering results. However, it is hard to produce a high quality graph for different datasets, especially when the input data contains noises. To solve this problem, we propose a Robust Rank Constrained Sparse Learning (RRCSL) method. The sparse representation and $L_{2,1}$-norm are combined to learn the desired graph and reduce the impact of data noise and outliers [3]. In addition, the optimal graph is searched within the neighbor of the predefined graph, which ensures that the similarity graph reflects the relationship between the data accurately. The data similarity graphs by the traditional graph-based clustering method require an additional post-processing (such as K-means) to complete the clustering. The rank constrainted makes the learned graph have a clear structure that can be classified directly [4].

In summary, the main contributions of this paper are as follows:

- The sparse representation is combined with $L_{2,1}$-norm to upgrade the construction quality of the graph and reduce the impact of data noise and outliers.
- To ensure the similarity graph reflects the relationship between that data accurately, we construct an initial graph and explore the optimal similarity graph in its neighborhood.
- Rank constraint is added into the objective function, which can avoid additional post-processing.

2. THE PROPOSED METHOD

In this section, on the basis of introducing sparse representation, we will propose a Robust Rank Constrained Sparse Learning (RRCSL) method. Then we will address the issue of the corresponding optimization method.

2.1. Sparse Representation

Suppose the data matrix is $X = [x_1, ..., x_n] \in \mathbb{R}^{d \times n}$, where each column corresponds to one sample, and each row corresponds to one feature. In most cases, X is dense with lots of redundant and irrelevant data. It is necessary to convert these dense data into appropriate sparse forms and this process is sparse representation.

Suppose $y \in \mathbb{R}^{d \times 1}$ is a new observation of X and $\beta \in \mathbb{R}^{n \times 1}$ is a weight vector used to approximate y. The purpose of sparse representation learning is to solve the follow-

[*]Qi Wang is the corresponding author. This work was supported by the National Key R&D Program of China under Grant 2018YFB1107403, National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, 61871470 and 61761130079.

ing problem to learn β:

$$\min_{\beta} \|y - X\beta\|_2^2 + \omega_0 \|\beta\|_1 . \tag{1}$$

The first term of Eq.(1) makes β refactor X, and then ensures that y is approximately equal to $X\beta$. The second term makes β as sparse as possible. Based on this, Nie et al. [5] proposed the Simplex Sparse Representation (SSR) [5]. The similarity $\alpha_i \in \mathbb{R}^{(n-1)\times 1}$ between the i-th sample and other samples can be expressed by the following objective equation:

$$\min_{\alpha_i \geq 0} \|x_i - X_{-i}\alpha_i\|_2^2 + \omega_0 \|\alpha_i\|_1 , \tag{2}$$

where $X_{-i} = [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n] \in \mathbb{R}^{d\times(n-1)}$ represents the data matrix that removes column i. To make the graph to be invariant to rotation, it is necessary to constrain $\alpha_i^T \mathbf{1} = 1$. Meanwhile $\|\alpha_i\|_1$ becomes constant, so Eq.(2) can be rewritten as:

$$\min_{\alpha_i \geq 0, \alpha_i^T \mathbf{1}=1} \|x_i - X_{-i}\alpha_i\|_2^2 . \tag{3}$$

Eq.(3) is the objective equation of the SSR method. This method assumes that a data point can be approximated by a linear combination of the other points, so that a sparse data similarity graph can be learned without prior parameter. However, it is obvious that Eq.(3) squares the residue errors, so this method is sensitive to outliers and noise. Furthermore, for the same set of x_i and X_{-i}, there may be multiple linear combinations of α_i. For the same sample, minimizing Eq.(3) leads to multiple solutions. Therefore the learned graph may not reflect the correct relationship of samples, which further produces incorrect clustering results.

2.2. Robust Rank Constrained Sparse Learning method

In this section, based on the sparse representation, we will propose Robust Rank Constrained Sparse Learning (RRCSL) method.

In order to facilitate subsequent optimization, we rewrite Eq.(3) into the matrix form:

$$\begin{aligned} &\min_{S} \|X - XS\|_F^2 , \\ &s.t. \quad s_{ij} \geq 0, \sum_j s_{ij} = 1, \end{aligned} \tag{4}$$

where $S \in \mathbb{R}^{n\times n}$ is the desired graph and $\|\bullet\|_F$ is Frobenius norm.

However, in Eq.(4), the problem of poor robustness due to squared errors has not been solved, so we consider introduce the $L_{2,1}$-norm objective function:

$$\begin{aligned} &\min_{S} \|X - XS\|_{2,1} , \\ &s.t. \quad s_{ij} \geq 0, \sum_j s_{ij} = 1. \end{aligned} \tag{5}$$

The $L_{2,1}$-norm of the matrix M is defined as:

$$\|M\|_{2,1} = \sum_i \sqrt{\sum_j |m_{ij}|^2} = \sum_i \|m_j\|_2 . \tag{6}$$

Compared to $\|\bullet\|_F^2$ in Eq.(4), the $\|\bullet\|_{2,1}$ has better robustness to data noise and outliers. The $L_{2,1}$-norm can also make the model insensitive to feature rotation. In addition, rather than pursuing the flat sparsity, the $L_{2,1}$-norm ensures the structural sparsity within the learned graph.

It should be noted that there may be multiple linear combinations of S for the same X. This results in that the similarity graph of data does not accurately reflect the relationship between the data. Therefore we construct an initial graph and find the optimal similarity graph within the neighborhood of the initial graph:

$$\begin{aligned} &\min_{S} \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 , \\ &s.t. \quad s_{ij} \geq 0, \sum_j s_{ij} = 1, \end{aligned} \tag{7}$$

where B is the initial similarity graph of the input which can be generated by the Nie et al. [4], and $\alpha > 0$ is the controlling parameter. Eq.(7) can make the learned optimal similarity graph closer to the initial similarity graph, thus insuring optimal similarity graph is well represented and making up for the shortcomings of sparse representation learning that cannot grasp the local relationship.

The additional post-processing (such as K-means) is required to complete the clustering task, which reduces the quality and efficiency of clustering. To deal with this problem, we constrained in the process of learning data similarity graph, ensuring the learned graph have a clear structure that can be classified directly [6]. This requires the following theorem:

Theorem 1 [7]: *The multiplicity k of the eigenvalue zero of the Laplacian matrix L_s is equal to the number of connected components in the graph associated with S.*

The Laplacian matrix is $L_s = D_s - \left(S^T + S\right)/2$, where D_s is the diagonal matrix and i-th diagonal element is $\sum_j \left(s_{ij} + s_{ji}\right)/2$. Given the cluster number k, we can constrain the rank of L_s to be $n - k$, so that the learned graph contains exact k connected components. In this way, the clustering result can be directly obtained according to the connection relationship of the data points in the similarity graph [4].

The constraint $rank\left(L_s\right) = n - k$ is a complex nonlinear constraint, but it can be equivalently converted into an easy-to-solve form by an effective method. Assuming $\sigma_i\left(Ls\right)$ is its i-th smallest eigenvalue, when λ is large enough, Eq.(7) with rank constrained is equivalent to:

$$\begin{aligned} &\min_{S} \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 + 2\lambda \sum_{i=1}^{k} \sigma_i(L_s), \\ &s.t. \quad s_{ij} \geq 0, \sum_j s_{ij} = 1. \end{aligned} \tag{8}$$

According to Ky Fan et al. [8]:

$$\begin{aligned} &\sum_{i=1}^{k} \sigma_i(L_s) = \min_{F} Tr(F^T L_s F), \\ &s.t. \quad F \in \mathbb{R}^{n\times\ k}, F^T F = I, \end{aligned} \tag{9}$$

where $F \in \mathbb{R}^{n\times k}$ is an indicator matrix. Therefore, Eq.(8) is

equivalent to:

$$\begin{aligned} &\min_{S,F} \|X - XS\|_{2,1} + \alpha \|S - B\|_F^2 + \lambda Tr(F^T L_s F), \\ &s.t. \quad s_{ij} \geq 0, \sum_j s_{ij} = 1, F^T F = I. \end{aligned} \tag{10}$$

The above formula is the objective equation of RRCSL method for single-view clustering. By solving Eq.(10), we get a robust and sparse data similarity graph with the clear cluster structure. In the following part, we will present an optimization algorithm to solve Eq.(10).

2.3. Optimization Algorithm

Since $\|X - XS\|_{2,1}$, $\|S - B\|_F^2$ and L_s all depend on S, Eq.(10) is a complex optimization problem. An efficient approach called Augmented Lagrangian Multiplier (ALM) [9] can be used for tackling Eq.(10).

Let $E = X - XZ$, $Z = S$, then Eq.(10) is transformed into the following ALM problem:

$$\begin{aligned} &\min_S \|E\|_{2,1} + \alpha \|Z - B\|_F^2 + \lambda Tr(F^T L_s F) \\ &+ \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2, \\ &s.t. \quad s_{ij} \geq 0, \sum_j s_{ij} = 1, F^T F = I, \end{aligned} \tag{11}$$

where $\mu \in \mathbb{R}^{1\times 1}$ is a regularity coefficient, $\Lambda_1 \in \mathbb{R}^{d\times n}$ and $\Lambda_2 \in \mathbb{R}^{d\times n}$ are penalty parameters.

Update E:

When updating E, we fix Z, S and F. Denoting $C = X - XZ - \frac{\Lambda_1}{\mu}$, thus Eq.(11) becomes:

$$\min_E \|E\|_{2,1} + \frac{\mu}{2} \|E - C\|_F^2 . \tag{12}$$

According to Nie et al. [10] [11], let $\delta = \frac{1}{\mu}$, and the optimal solution of E is:

$$E(:,i) = \begin{cases} C(:,i) - \frac{C(:,i)}{\mu \|C(:,i)\|}, \text{ if } \|C(:,i)\|_2 > \frac{1}{\mu} \\ 0, \quad \text{otherwise} , \end{cases} \tag{13}$$

where $E(:,i)$ and $C(:,i)$ are i-th columns of E and C respectively.

Update Z:

When update Z, we fix E, S and F, thus Eq.(11) becomes:

$$\begin{aligned} \min_Z \alpha \|Z - B\|_F + \frac{\mu}{2} \left\| E - X + XZ + \frac{\Lambda_1}{\mu} \right\|_F^2 \\ + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2 . \end{aligned} \tag{14}$$

Taking the derivative of Eq.(14) with respect to Z and let it be zero, then we get the optimal solution for Z as follows:

$$Z = \frac{\left[\frac{2\alpha}{\mu} B - 2X^T \left(E - X + \frac{\Lambda_1}{\mu} \right) + S - \frac{\Lambda_2}{\mu} \right]}{\left(\frac{2\alpha}{\mu} I + X^T X + I \right)}. \tag{15}$$

Update S:

When update S, we fix E, Z and F, thus Eq.(11) becomes:

$$\min_S \lambda Tr(F^T L_s F) + \frac{\mu}{2} \left\| Z - S + \frac{\Lambda_2}{\mu} \right\|_F^2 . \tag{16}$$

Let s_{ij} denote the element in S. According to the properties of the Laplace matrix, we have the following equation for any F:

$$Tr(F^T L_s F) = \frac{1}{2} \sum_{ij} (\|F(i,:) - F(j,:)\|_2^2 s_{ij}). \tag{17}$$

Denoting $H = Z + \frac{\Lambda_2}{\mu}$ and $d_{ij} = \sum_{ij} \|F(i,:) - F(j,:)\|_2^2$, Eq.(16) becomes:

$$\min_S \frac{\lambda}{2} \sum_{ij} d_{ij} s_{ij} + \frac{\mu}{2} \|H - S\|_F^2 . \tag{18}$$

Let h_{ij} denote the element in H. Eq.(18) becomes:

$$\min_S \frac{\lambda}{2} \sum_{ij} d_{ij} s_{ij} + \frac{\mu}{2} \sum_{ij} (h_{ij} - s_{ij})^2. \tag{19}$$

Note that the solution to Eq.(19) is independent for each row, so we can optimize each row separately to solve the above problem. Eq.(19) can be written in the following form further:

$$\min_S \left\| s_j^2 - (h_j - \frac{\lambda}{2\mu} d_j) \right\|_2^2 . \tag{20}$$

which can be solved with an efficient iterative algorithm [5].

Update F:

When update F, we fix E, Z and S, thus Eq.(11) becomes:

$$\begin{aligned} &\min_F Tr(F^T L_s F), \\ &s.t. \quad F \in \mathbb{R}^{n\times k}, F^T F = I. \end{aligned} \tag{21}$$

The number of clusters is k. Calculating the first k minimum eigenvalues of L_s, and the optimal solution of F consists of k eigenvectors corresponding to these eigenvalues.

Update Λ_1,Λ_2 and μ:

In each iteration, the ALM parameters are updated as follows [10]:

$$\begin{aligned} \Lambda_1 &= \Lambda_1 + \mu (E - X + XZ), \\ \Lambda_2 &= \Lambda_2 + \mu (Z - S), \\ \mu &= \rho \mu, \end{aligned} \tag{22}$$

where $\rho > 1$ is the update rate, and a larger ρ brings faster convergence speed, but accompanied by poorer results.

3. EXPERIMENTS

The robust rank constrained sparse learning (RRCSL) method is evaluated on nine commonly used real-world benchmark datasets. The clustering experimental results will be compared with six state-of-the-art clustering methods. In addition, the robustness of RRCSL method is also proved.

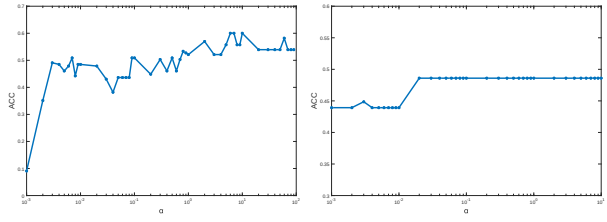
3.1. Experiments Setup

Nine real-world benchmark datasets are employed to demonstrate the good performance of the proposed RRCSL

Table 1. Clustering experimental results in terms of Accuracy/NMI.

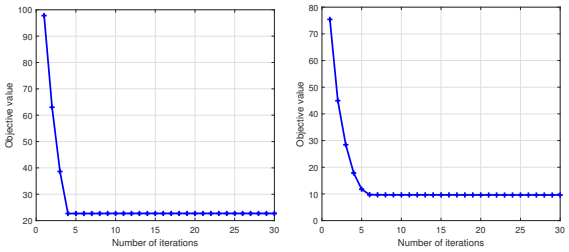
	SRBCT	Iris	Yale	Wine	Glass	UMIST	Coil20	Yeast	Semeion
K-means	0.42/0.14	0.79/0.63	0.47/0.54	0.96/0.86	0.48/0.33	0.43/0.63	0.55/0.72	0.40/0.26	0.57/0.52
NMF	0.43/0.17	0.63/0.40	0.33/0.41	0.72/0.44	0.44/0.30	0.30/0.48	0.47/0.60	0.13/0.01	0.37/0.32
CAN	0.50/0.27	0.57/0.42	0.52/0.54	0.92/0.78	0.45/0.28	0.67/0.81	0.83/0.91	0.40/0.15	0.61/0.61
CLR	0.39/0.18	0.84/0.72	0.50/0.58	0.91/0.73	0.41/0.24	0.61/0.79	0.82/0.90	0.49/0.29	0.41/0.40
SSR	0.42/0.16	0.68/0.58	0.59/0.58	0.95/0.84	0.38/0.30	0.56/0.71	0.62/0.80	0.41/0.23	0.61/0.61
KMM	0.42/0.24	0.81/0.68	0.44/0.50	0.54/0.36	0.46/0.32	0.38/0.59	0.43/0.64	0.42/0.26	0.58/0.53
RRCSL	**0.59/0.41**	**0.87/0.75**	**0.60/0.63**	**0.97/0.88**	**0.48/0.34**	**0.81/0.90**	**0.88/0.93**	**0.50/0.31**	**0.63/0.62**

in single-view clustering, including SRBCT [12], iris [13], Yale [14], wine [15], glass [16], umist [17], coil20 [18], yeast [19] and Semeion [20]. Two widely used clustering performance measures are adopted to evaluate the clustering results, including Accuracy (ACC) and normalized mutual information (NMI). We compared our clustering methods with 7 clustering methods, including K-means, NMF [21] methods, Clustering with Adaptive Neighbors (CAN) [22], Constrained Laplacian Rank (CLR) [4], Simplex Sparse Representation (SSR) [5] and K-Multiple-Means (KMM) [23]. As for the parameter λ, a heuristic way is employed to determine the valua of λ: at the beginning, let λ be a small value like $1e-4$, then we calculate whether the number of zero eigenvalues of L_s is equal to k in each iteration. The only variable parameter is controlling parameter α.



(a) YALE (b) GLASS

Fig. 1. Convergence analysis on datasets.



(a) YALE (b) GLASS

Fig. 2. Convergence analysis on datasets.

3.2. Performance

After the same normalization processing of the data of the nine datasets, the RRCSL method and six comparison algorithms are applied respectively to carry out clustering experiments. The performances are shown in Table 1 (bold indicates the best). We can see that the proposed method RRCSL outperform other methods in all the experiments.

Taking the two datasets of Yale and Glass as examples, we let the variable parameter α fluctuate within a large range of 0.0001 to 10. It can be observed that the performance of RRCSL method is very stable as shown in Figure 1. Meanwhile, we prove the convergence of the proposed algorithm on the same datasets as shown in Figure 2.

3.3. The Robustness Experiments

The robustness of RRCSL is verified by taking Yale face dataset as an example. On each 16-by-16-pixel image, a part is randomly selected for occlusion processing. Table 2 summarizes the results for all methods on the occluded Yale dataset. Obviously, the RRCSL method has the best robustness compared to others.

Table 2. The Accuracy/NMI of the robustness experiments.

Occlusion	2×2	3×3	4×4
K-means	0.46/0.53	0.45/0.51	0.42/0.48
NMF	0.34/0.42	0.33/0.40	0.32/0.39
CAN	0.51/0.56	0.44/0.45	0.43/0.49
CLR	0.47/0.56	0.51/0.55	0.43/0.46
SSR	0.55/0.56	0.53/0.53	0.48/0.51
KMM	0.36/0.41	0.34/0.38	0.33/0.40
RRCSL	**0.59/0.63**	**0.57/0.60**	**0.50/0.58**

4. CONCLUSION

In this paper, we propose a new graph-based robust rank constrained sparse learning (RRCSL) method for clustering. In our method, the sparse representation with a rank constraint is adopted to learn the optimal data graph, and $L_{2,1}$-norm is combined to improve the construction quality of the graph and reduce the impact of data noise and outliers. We futher construct an initial graph and explore the optimal similarity graph in the neighborhood of it. Extensive experiments on real-world datasets have proved the superiority and robustness of the proposed method.

5. REFERENCES

[1] Wei Zhu, Feiping Nie, and Xuelong Li, "Fast spectral clustering with efficient large graph construction," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2492–2496.

[2] Qi Wang, Zequn Qin, Feiping Nie, and Xuelong Li, "Spectral embedded adaptive neighbors clustering," *IEEE Transactions on Neural Networks and Learning Systems (T-NNLS)*, vol. 30, no. 4, pp. 1265–1271, 2018.

[3] Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding, "Efficient and robust feature selection via joint l2, 1-norms minimization," in *Advances in neural information processing systems*, 2010, pp. 1813–1821.

[4] Feiping Nie, Xiaoqian Wang, Michael I Jordan, and Heng Huang, "The constrained laplacian rank algorithm for graph-based clustering.," in *AAAI*, 2016, pp. 1969–1976.

[5] Jin Huang, Feiping Nie, and Heng Huang, "A new simplex sparse learning model to measure data similarity for clustering.," in *IJCAI*, 2015, pp. 3569–3575.

[6] Qi Wang, Mulin Chen, Feiping Nie, and Xuelong Li, "Detecting coherent groups in crowd scenes by multiview clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, vol. 42, no. 1, pp. 46–58, 2020.

[7] Bojan Mohar, Y Alavi, G Chartrand, and OR Oellermann, "The laplacian spectrum of graphs," *Graph theory, combinatorics, and applications*, vol. 2, no. 871-898, pp. 12, 1991.

[8] Ky Fan, "On a theorem of weyl concerning eigenvalues of linear transformations i," *Proceedings of the National Academy of Sciences*, vol. 35, no. 11, pp. 652–655, 1949.

[9] Michael JD Powell, "A method for nonlinear constraints in minimization problems," *Optimization*, pp. 283–298, 1969.

[10] Jin Huang, Feiping Nie, Heng Huang, and Chris Ding, "Robust manifold nonnegative matrix factorization," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 8, no. 3, pp. 11, 2014.

[11] Ming Yuan and Yi Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[12] Javed Khan, Jun S Wei, Markus Ringner, Lao H Saal, Marc Ladanyi, Frank Westermann, Frank Berthold, Manfred Schwab, Cristina R Antonescu, Carsten Peterson, et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature medicine*, vol. 7, no. 6, pp. 673, 2001.

[13] Ronald A Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[14] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[15] Stefan Aeberhard, Danny Coomans, and Olivier De Vel, "Improvements to the classification performance of rda," *Journal of chemometrics*, vol. 7, no. 2, pp. 99–115, 1993.

[16] Ping Zhong and Masao Fukushima, "Regularized nonsmooth newton method for multi-class support vector machines," *Optimisation Methods and Software*, vol. 22, no. 1, pp. 225–236, 2007.

[17] Harry Wechsler, Jonathon P Phillips, Vicki Bruce, Francoise Fogelman Soulie, and Thomas S Huang, *Face recognition: From theory to applications*, vol. 163, Springer Science & Business Media, 2012.

[18] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al., "Columbia object image library (coil-20)," 1996.

[19] Paul Horton and Kenta Nakai, "A probabilistic classification system for predicting the cellular localization sites of proteins.," in *Ismb*, 1996, vol. 4, pp. 109–115.

[20] Massimo Buscema, "Metanet*: The theory of independent judges," *Substance use & misuse*, vol. 33, no. 2, pp. 439–461, 1998.

[21] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[22] Feiping Nie, Xiaoqian Wang, and Heng Huang, "Clustering and projected clustering with adaptive neighbors," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 977–986.

[23] Feiping Nie, Cheng-Long Wang, and Xuelong Li, "K-multiple-means: A multiple-means clustering method with specified k clusters," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, pp. 959–967.