MULTI-VIEW K-MEANS WITH LAPLACIAN EMBEDDING

Zhezheng Hao[1,2] *Zhoumin Lu*[2,3] *Feiping Nie*[2,3*] *Rong Wang*[2] *Xuelong Li*[2]

[1] School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.
[2] School of Artificial Intelligence, Optics and Electronics (iOPEN),
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.
[3] School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.

ABSTRACT

Most of the existing multi-view clustering algorithms are performed in the original feature space, and their performance in heavily reliant on the quality of the raw data. Besides, some two-stage strategies cannot achieve ideal results due to the absence of capturing the correlation between views. In view of this, we propose Multi-View K-means with Laplacian Embedding (MVKLE), which is capable of clustering multi-view data in the learned embedding space. Specifically, we employ local structure-preserving dimensionality reduction to obtain the underlying representation of each view, and obtain the clustering results directly through an effective optimization formulation. Experiments on several common multi-view datasets demonstrate the superiority of the proposed method.

***Index Terms*—** Multi-view clustering, Nonnegative matrix factorization, Laplacian eigenmaps, Graph Constraints

1. INTRODUCTION

Data mining and analytics are applied in many practical tasks. Multi-view data may exhibit heterogeneous characteristics while retaining immanent associations. Now multi-view learning has emerged as a promising direction of machine learning. Compared with traditional single-view clustering, multi-view clustering takes advantage of valuable feature knowledge from diverse views to improve the performance of clustering. In the context of multi-view clustering, one inherent issue that all types of algorithms seek to deal with is finding a methodology to maximize the quality and complementarity of clusters in each view, while considering the consistency of clustering across views [1]. The naive approach either concatenates the features of all views together into one macro view, or performs clustering on each view independently. Nevertheless, since the features of each view have specific statistical significance, both approaches fail to exploit the complementary information of multiple views and are prone to overfitting [2].

*Corresponding author. This work was supported in part by the National Natural Science Foundation of China under Grant 62276212 and Grant 62176212.

The concept of multiple views was first applied to clustering tasks by Bickel et al. [3]. They extended the EM algorithm to multiple views and applied it to text data. Models based on various principles have been proposed these years, including co-training algorithms [4] [5], multi-view graph clustering [6] [7], multi-view subspace clustering [8] [9], etc. Although these multi-view clustering algorithms effectively utilize multi-view information, they still suffer from some drawbacks. On the one hand, the existing algorithms usually rely on the original features of each view and are less capable of mining a unified feature representation of multi-view data. On the other hand, two-stage multi-view clustering methods lack the ability to refine components in one coherent optimization framework. In addition, these methods are susceptible to poor quality raw data, leading to uninspiring results.

To address the above issues, we propose Multi-View K-means with Laplacian Embedding (MVKLE). The main contributions of this paper can be summarized as follows:

- We propose a succinct multi-view clustering method MVKLE, which learns the underlying representation of each view separately and can effectively retain the local structure of the original data. MVKLE can be applied not only to feature data, but also to graph data.
- We put Laplace embedding and k-means into a unified optimization framework that takes the complementarity and consistency between views into account. Clustering in embedding space effectively reduces the reliance on the original space.
- Suitable optimization methods are employed to obtain the indicator matrix directly without post-processing. The experimental results also verify the superiority of the proposed algorithm.

Notations: Matrices are written in boldface uppercase throughout the paper. Data matrix of the v-th view is denoted as $\mathbf{X}_v \in \mathbb{R}^{d_v \times n}$, where n is the number of samples and d_v represents the dimension of features. The matrix is called an indicator matrix, if each row of it has one and only one element equal to 1 indicating the cluster membership and the rest elements are all 0. $\mathbf{\Phi}^{n \times k}$ denotes the set of all indicator

matrices. The transpose, the i-th row, the j-th column, the (i,j)-th entry, the trace and the Frobenius norm of $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}_{i\cdot}$, $\mathbf{A}_{\cdot j}$, $\mathbf{A}_{ij}$, $Tr(\mathbf{A})$,$\|\mathbf{A}\|_F$, respectively.

2. MULTI-VIEW K-MEANS WITH LAPLACIAN EMBEDDING

Formulation of multi-view clustering problems is typically accomplished by means of Nonnegative Matrix Factorization (NMF). A brief revisit of it is introduced before our model formulation. Our model is proposed in this section followed by optimization and analysis.

2.1. Multi-view Clustering Revisit

The equivalence of NMF with k-means has been demonstrated in the literature [10]. A number of researchers have extended it to multi-view clustering, a general framework of which is

$$\min_{\mathbf{A}_v,\mathbf{B}} \|\mathbf{X}_v - \mathbf{A}_v\mathbf{B}^T\|_F^2 \quad \text{s.t. } \mathbf{A}_v \geq 0, \mathbf{B} \in \mathbf{\Phi}^{n\times k}. \tag{1}$$

where $\mathbf{A}_v \in \mathbb{R}^{d_v\times k}$ denote the basis matrix of the v-th viewpoint, and $\mathbf{B} \in \mathbb{R}^{n\times k}$ denotes the shared indicator matrix. Liu et al. [11] introduced consistency between different views at NMF, i.e., a penalty term $\sum_{v=1}^{M}\|\mathbf{B}_v - \mathbf{B}^*\|_F^2$ is incorporated to measure the difference in the indicator matrix. Cai et al. [12] proposed graph regularization NMF to preserve the local geometric structure, i.e., by adding the smoothness function $J = tr(\mathbf{BLB}^T)$. Zhu et al. [13] proposed one-step multi-view spectral clustering, and yet the redundant parameter settings limit its application.

2.2. The Proposed Method

Inspired by [8], our approach explores the underlying complementary information from multiple views while finding the respective underlying representations. We employ Laplacian Eigenmaps (LE), a nonlinear dimensionality reduction approach, to learn the underlying representation of each view. LE uses a local perspective to construct relationships between data. According to the literature [14], the similarity matrix $\mathbf{W}$ of the single-view data can be readily obtained in several ways, such as through the Gaussian kernel function $\mathbf{W}_{ij} = exp(-\|\mathbf{X}_{\cdot i} - \mathbf{X}_{\cdot j}\|_2^2/(2\sigma^2))$. In turn, the diagonal degree matrix $\mathbf{D}_{ii} = \sum_{j=1}^{n}\mathbf{W}_{ij}$ and the regularized Laplace matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$ can be derived. The main idea is that if a pair of data is similar, it should also be close in its reduced dimensional subspace. LE can be formulated as:

$$\min_{\mathbf{Y}} \sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{Y}_{\cdot i} - \mathbf{Y}_{\cdot j}\|_2^2 W_{ij}, \tag{2}$$

where $\mathbf{Y} \in \mathbb{R}^{d'\times n}$ is the dimensionality-reduced data matrix. Further a concise form of dimensionality reduction (2) can be derived:

$$\begin{aligned}(2) &\Leftrightarrow \min_{\mathbf{Y}} \sum_{i=1}^{n}\mathbf{D}_{ii}\mathbf{Y}_{\cdot i}^T\mathbf{Y}_{\cdot i} - \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{W}_{ij}\mathbf{Y}_{\cdot i}^T\mathbf{Y}_{\cdot i} \\ &\Leftrightarrow \min_{\mathbf{Y}} Tr(\mathbf{YDY}^T) - Tr(\mathbf{Y}^T\mathbf{YW}) \\ &\Leftrightarrow \min_{\mathbf{Y}} Tr(\mathbf{YLY}^T).\end{aligned} \tag{3}$$

Since the solution can degenerate to a single point and can have an arbitrary scale, adding constraints to this optimization problem $\mathbf{YDY}^T = \mathbf{I}$, which meets the normalized cut in spectral clustering. LE is thus converted into a generalized eigenvalue problem.

For multi-view clustering problems, many two-step approaches do not balance complementarity and consistency. We use a one-step approach for clustering data from multiple views. Consider mapping the original data from different views to their respective feature spaces by Laplacian Eigenmaps, i.e., $\mathcal{L}_v : \mathbf{X}_v \to \mathbf{Y}_v$, where $\mathbf{Y}_v \in \mathbb{R}^{k_v\times n}$. Further, since the low-dimensional representations of each view are distinct, the basis matrices of k-means for each view are therefore distinct. Then the objective function can be written as:

$$\begin{aligned}&\min_{\mathbf{Y}_v,\mathbf{A}_v,\mathbf{B},\alpha} \sum_{v=1}^{M}\alpha_v^p\left(Tr\left(\mathbf{Y}_v\mathbf{L}_v{\mathbf{Y}_v}^T\right) + \lambda\|\mathbf{Y}_v - \mathbf{A}_v\mathbf{B}^T\|_F^2\right) \\ &\text{s.t. } \mathbf{Y}_v\mathbf{D}_v{\mathbf{Y}_v}^T = \mathbf{I}, \mathbf{B} \in \mathbf{\Phi}^{n\times k}, \alpha^T\mathbf{1} = \mathbf{1}, \alpha \geq 0,\end{aligned} \tag{4}$$

where λ is a trade-off parameter which controls the balance betweeen the two items. The base matrix $\mathbf{A}_v$ varies from view to view, while the final indicator matrix $\mathbf{B}$ is uniform. α_v^p is the exponential decay form weight of each view, where $p > 1$ is a scalar to keep the smooth distribution.

2.3. Optimization

In this section, we employ the alternating optimization approach to solve Eq. (4).

Step 1 Update $\mathbf{A}_v$ with $\mathbf{Y}_v$, $\mathbf{B}$, α fixed. The optimization problem becomes simple:

$$\min_{\mathbf{A}_v} \sum_{v=1}^{M}\alpha_v^p\|\mathbf{Y}_v - \mathbf{A}_v\mathbf{B}^T\|_F^2. \tag{5}$$

Since the objective function is convex with respect to $\mathbf{A}_v$, we force its derivative with respect to $\mathbf{A}_v$ to be 0. Ignoring irrelevant terms, we can obtain:

$$\mathbf{A}_v = \mathbf{Y}_v\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1} \tag{6}$$

It can be further found that $\mathbf{A}_v$ represent the centroids in the v-th view. This is consistent with the idea of k-means. $\mathbf{B}^T\mathbf{B}$ is a diagonal matrix and its inverse straightforward to obtain.

Step 2 Update $\mathbf{B}$ with $\mathbf{Y}_v$, $\mathbf{A}_v$, α fixed. The optimization problem can be reduced to

$$\begin{aligned}&\min_{\mathbf{B}\in\mathbf{\Phi}^{n\times k}} \sum_{v=1}^{M}\alpha_v^p\|\mathbf{Y}_v - \mathbf{A}_v\mathbf{B}^T\|_F^2 \\ \Leftrightarrow &\min_{\mathbf{B}\in\mathbf{\Phi}^{n\times k}} \sum_{i=1}^{n}\sum_{v=1}^{M}\alpha_v^p\|(\mathbf{Y}_v)_{\cdot i} - \mathbf{A}_v\mathbf{B}_{i\cdot}^T\|_2^2.\end{aligned} \tag{7}$$

Since only one element of each row of $\mathbf{B}$ is equal to 1 and the rest are equal to 0, the optimization process of the i-th row of $\mathbf{B}$ is

$$\mathbf{B}_{ij}^{(t)} = \begin{cases}1, & j = \arg\min_j \sum_{v=1}^{M}\alpha_v^p\|(\mathbf{Y}_v)_{\cdot i} - (\mathbf{A}_v)_{\cdot j}\|_2^2, \\ 0, & \text{otherwise.}\end{cases} \tag{8}$$

Step 3 Update Y_v with $\mathbf{A}_v$, $\mathbf{B}$, α fixed. Converting the Frobenius norm form in problem Eq. (4) into trace form, the optimization problems can be rewritten as:

$$\begin{aligned}\min_{\mathbf{Y}_v} \sum_{v=1}^{M} \alpha_v^p \Big(& Tr\left(\mathbf{Y}_v \mathbf{L}_v {\mathbf{Y}_v}^T\right) + \lambda Tr\left(\mathbf{Y}_v \mathbf{Y}_v^T\right) \\ & -\lambda Tr\left(\mathbf{Y}_v \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T {\mathbf{Y}_v}^T\right)\Big) \\ \text{s.t. } & \mathbf{Y}_v \mathbf{D}_v {\mathbf{Y}_v}^T = \mathbf{I},\end{aligned} \tag{9}$$

where $\mathbf{A}_v$ is substituted by the solution in Eq. (6). Let $\tilde{\mathbf{Y}}_v = \mathbf{Y}_v \mathbf{D}_v^{\frac{1}{2}}$. Since the optimization problem is independent for each view, Eq (9) can thereby be translated into a concise form for each of the M views:

$$\begin{aligned}\min_{\tilde{\mathbf{Y}}_v} & Tr(\tilde{\mathbf{Y}}_v \mathbf{H}_v \tilde{\mathbf{Y}}_v^T) \\ \text{s.t. } & \tilde{\mathbf{Y}}_v \tilde{\mathbf{Y}}_v^T = \mathbf{I},\end{aligned} \tag{10}$$

where $\mathbf{H}_v = \mathbf{D}_v^{\frac{1}{2}} \left(\mathbf{L}_v + \lambda\mathbf{I} - \lambda\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\right)\mathbf{D}_v^{\frac{1}{2}}$ denoting a constant matrix for the v-th view. Obviously, this problem is an eigenvector problem, which is derived by the Lagrange multiplier method. The minimum of (10) is the sum of the k_v smallest eigenvalues of $\mathbf{H}_v$, while the row vectors of $\tilde{\mathbf{Y}}_v$ are the corresponding orthogonal eigenvectors. In experiments, $\mathbf{H}_v$ is not only positive semi-definite but often sparse, which facilitates the solution.

Step 4 Update α with $\mathbf{Y}_v$, $\mathbf{A}_v$, $\mathbf{B}$ fixed. The optimization problem becomes a simple power programming problem:

$$\min_{\alpha} \sum_{v=1}^{M} \alpha_v^p \theta_v \quad \text{s.t. } \alpha^T\mathbf{1} = \mathbf{1}, \alpha \geq 0, \tag{11}$$

where $\theta_v = Tr\left(\mathbf{Y}_v \mathbf{L}_v {\mathbf{Y}_v}^T\right) + \lambda\|\mathbf{Y}_v - \mathbf{A}_v\mathbf{B}^T\|_F^2$ denoting a non-negative constant for the v-th view. To solve this optimization problem with constraints, the Lagrangian function is written as follows:

$$l(\alpha, \mu, \xi) = \sum_{v=1}^{M} \alpha_v^p \theta_v - \mu(\alpha^T\mathbf{1} - \mathbf{1}) - \xi^T\alpha, \tag{12}$$

where $\mu \in \mathbb{R}$ and $\xi \in \mathbb{R}^M, \xi \geq 0$ are Lagrangian multipliers. According to the complementarity conditions of the KKT conditions, simplify to get $p\alpha_v^{p-1}\theta_v - \mu = 0$. The closed solution of α obtained by eliminating μ with $\sum_{v=1}^{M} \alpha_v = 1$,

$$\alpha_v = \frac{\theta_v^{\frac{1}{1-p}}}{\sum_{v=1}^{M} \theta_v^{\frac{1}{1-p}}}. \tag{13}$$

The detailed procedure of the proposed MVKLE is outlined in Algorithm 1.

2.4. Theoretical Analysis

Theorem 1 *Optimization problem Eq. (10) is convex for* $\tilde{\mathbf{Y}}_v$.

Proof. We first proof that $\mathbf{H}_v$ is an $n \times n$ semi-positive definite matrix, which is equivalent to the point where $\left(\mathbf{L}_v + \lambda\mathbf{I} - \lambda\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\right)$ is a semi-positive definite matrix. The semi-positive property of the Laplace matrix $\mathbf{L}_v$ here is well known. For an arbitrary nonzero vector x,

$$\begin{aligned} & x^T\left(\mathbf{I} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\right)x \\ = & \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} \frac{1}{|C_t|} x_i \sum_{x_i, x_j \in C_t} x_j \\ = & \sum_{t=1}^{k} \frac{1}{|C_t|} \sum_{x_i \in C_t} \left(|C_t| x_i^2 - x_i \sum_{x_j \in C_t} x_j\right) \\ = & \sum_{t=1}^{k} \sum_{x_i, x_j \in C_t} \frac{(x_i - x_j)^2}{|C_t|} \geq 0. \end{aligned} \tag{14}$$

C_t represents the t-th cluster, satisfying $\sum_{t=1}^{k} |C_t| = n$. In the above equation, for each pair x_i, x_j in the same cluster, there is one and only one quadratic component $(x_i - x_j)^2$. That is why the above equation holds. Thus $\mathbf{H}_v$ is semi-positive definite with $\lambda > 0$. By matrix differentiation, the Hessian matrix of $Tr(\tilde{\mathbf{Y}}_v \mathbf{H}_v \tilde{\mathbf{Y}}_v^T)$ with respect to $\tilde{\mathbf{Y}}_v^T$ is $2\mathbf{I} \otimes \mathbf{H}_v$, which is also semi-positive definite. The convexity of the optimization problem Eq. (10) is consequently proved.

Algorithm 1 Multi-view K-means with Laplacian Embedding

Input: Data of M views, $\mathbf{X}_v \in \mathbb{R}^{d_v \times n}, v = 1, \cdots, M$, the number of clustering categories k, the power of weight p, the trade-off parameter λ.

Output: Indicator matrix $\mathbf{B} \in \mathbf{\Phi}^{n \times k}$.

Initialization: $\mathbf{Y}_v$ is initialized by LE respectively. The initial $\mathbf{B}$ is the result of k-means in a random view. $\alpha = (\frac{1}{M}, \frac{1}{M}, \cdots, \frac{1}{M})$.

repeat

1. Update $\mathbf{A}_v$ according to Eq. (6).
2. Update $\mathbf{B}$ according to Eq. (8).
3. Update $\mathbf{Y}_v$ according to Eq. (10).
4. Update α according to Eq. (13).

until $\mathbf{Y}_v$, $\mathbf{A}_v$, $\mathbf{B}$, and α converge.

Convergence Analysis The optimization procedure of Eq. (4) is divided into four subproblems. The update for $\mathbf{A}_v$ and $\mathbf{Y}_v$ are convex optimization, where the optimal solution is obtained by matrix derivation. The optimization for $\mathbf{B}$ is n independent problems, each of which is solved directly for the minimum. The closed solution of α is also optimal. Therefore, by solving the subproblems alternatively, our proposed algorithm ensures to find the optimal solution to each subproblem and finally, the overall objective value will non-increasingly converge to local optimum.

Complexity Analysis The time complexity of MVKLE mainly comes from the update procedure. For ease of representation, we approximate all k_v to k, which has a negligible impact on complexity. The time complexity of single update $\mathbf{A}_v$, $\mathbf{B}$, $\mathbf{Y}_v$, α is $\mathcal{O}(Mk^2n)$, $\mathcal{O}(Mk^2n)$, $\mathcal{O}(Mkn^2)$, $\mathcal{O}(Mk^2n + Mkn^2)$ respectively. Moreover, the com-

plexity of the construction of the initial Laplacian matrix is $\mathcal{O}(Mdn^2)$. Therefore the overall time complexity of $\mathcal{O}\left(Mdn^2 + (Mk^2n + Mkn^2)t\right)$, where the number of iterations t is experimentally proven to be usually small. Since M, k, t is usually very small in practice, its complexity is comparable to the previous similarity-based work.

3. EXPERIMENTS

3.1. Experiment Setup

In this section, our experiments are conducted on four real-world datasets, including Yale, SCENE [15], 3Sources [11], BBCSports [16]. Yale is a widely-used face image dataset consisting of 165 gray-scale images of 15 subjects, where 3 views are used in our experiments. SCENE contains 2688 outdoor images with 8 categories, and consists of 4 views. 3Sources text dataset includes 169 stories, each with a theme and 3 views, which are selcted from news articles [17]. BBCSport dataset is from bbcsport news corpora, which consists of 544 documents with 5 themes.

In order to verify the capability of the proposed algorithm, we select four prominent algorithms to conduct comparison experiments with it on the above dataset, including Robust Multi-view K-means Clustering (RMKMC) [18], Reweighted Discriminatively Embedded K-Means (RDEKM) [19], Multi-view Clustering in Latent Embedding Space (MCLES) [9], Mixed Embedding Approximation (MEA) [20]. For comparison, three commonly employed evaluation matrices are selected to evaluate the clustering performance, which are Accuracy (ACC), Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) [21].

To avoid that the diagonal elements of $\mathbf{B}^T\mathbf{B}$ to have 0, i.e., the corresponding cluster is empty, all $(\mathbf{B}^T\mathbf{B})^{-1}$ is replaced with $(\mathbf{I}+\mathbf{B}^T\mathbf{B})^{-1}$ in the experiments, which has negligible effect on the results. To facilitate the process of tuning parameters and to conform to the sense of embedding, we set all k_v uniformly to k. There are two hyperparameters p and λ in our proposed method. A shortcut to determine the hyperparameters is to select p in the range $[2, 3, \cdots, 10]$ and λ in the range $[10^{-5}, 10^{-4}, \cdots, 10^{5}]$. Similar grid parameter-search manners are adopted in other baseline methods.

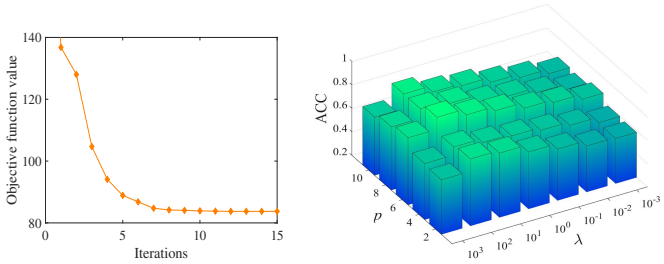
3.2. Performance Analysis

The experimental results of MVKLE with four comparison algorithms are given in Table 1, where each result is the average of 10 runs after tuning the parameters. The best results for each metric are marked in bold. Our algorithm achieves optimal results in most cases and is applicable to multiple types of data sets (text, speech, image, etc.). In summary, MVKLE demonstrates superior performance on real-world datasets.

Fig. 1 (a) shows the convergence curve of Algorithm 1 on BBCSport dataset. In extensive experiments, convergence speed is quick and the number of iteration steps usually not exceeding 15. In addition, we experimented with random initialization for $\mathbf{B}$ instead of k-means and considered the consequences on convergence, for which the convergence rate can still be guaranteed for suitable λ and p. Due to the fast convergence rate, the practical running time of MVKLE is not significantly slower than those of the linear complexity comparison algorithms. Fig. 1 (b) presents the 3-D bar chart of parameter sensitivity, which depicts the influence of different settings of the two major parameters p, λ on the clustering results. When p is too large or too small, the weight of each view will be over- or under-emphasized. Similarly, if λ is not set appropriately, the two terms in the objective function are not sufficiently constrained to each other during the update and lead to poor performance. It can be seen that the algorithm is effective for a large range of parameters.

Table 1: Comparison on real-world datasets.

Dataset	Metric	RMKMC	RDEKM	MCLES	MEA	MVKLE
Yale	ACC	0.494	0.642	0.662	0.587	**0.679**
	NMI	0.543	0.659	0.677	0.597	**0.684**
	ARI	0.498	0.636	0.641	0.565	**0.659**
SCENE	ACC	0.423	0.555	0.622	0.603	**0.647**
	NMI	0.347	0.380	0.506	0.504	**0.523**
	ARI	0.219	0.329	0.392	0.391	**0.411**
3Sources	ACC	0.546	0.588	0.674	0.574	**0.686**
	NMI	0.381	0.529	0.650	0.418	**0.663**
	ARI	0.347	0.520	0.583	0.435	**0.586**
BBCSport	ACC	0.705	0.844	0.869	0.748	**0.885**
	NMI	0.584	0.751	0.784	0.557	**0.814**
	ARI	0.460	0.718	0.770	0.395	**0.792**



(a) Convergence Curve (b) Parameter Sensitivity

Fig. 1: Convergence curve and parameter sensitivity on BBCSport.

4. CONCLUSION

In this paper, we propose a Multi-View K-means with Laplacian Embedding to overcome the shortcomings of the existing algorithms. The proposed method clusters low-dimensional features under a unified optimization schema, and the final results are obtained directly by a decent optimization method. Compared with the existing multi-view clustering algorithms, it has the distinct advantages of fewer parameters, easy parameter-tuning, stable and fast convergence. MVKLE is applicable to data with only graphs without features, which accommodates the proliferation of graph data. Experiments have been conducted on four benchmark datasets, and the results demonstrated the effectiveness of our model, compared to the four excellent clustering methods.

5. REFERENCES

[1] Kamalika Chaudhuri, Sham M Kakade, Karen Livescu, and Karthik Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proceedings of the International Conference on Machine Learning*, 2009, pp. 129–136.

[2] Chang Xu, Dacheng Tao, and Chao Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.

[3] Steffen Bickel and Tobias Scheffer, "Multi-view clustering," in *Proceedings of the IEEE International Conference on Data Mining*, 2004, pp. 19–26.

[4] Abhishek Kumar and Hal Daumé, "A co-training approach for multi-view spectral clustering," in *Proceedings of the International Conference on Machine Learning*, 2011, pp. 393–400.

[5] Annalisa Appice and Donato Malerba, "A co-training strategy for multiple view clustering in process mining," *IEEE Transactions on Services Computing*, vol. 9, no. 6, pp. 832–845, 2015.

[6] Feiping Nie, Jing Li, Xuelong Li, et al., "Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification.," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2016, pp. 1881–1887.

[7] Chenping Hou, Feiping Nie, Hong Tao, and Dongyun Yi, "Multi-view unsupervised feature selection with adaptive similarity and view weight," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1998–2011, 2017.

[8] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu, "Generalized latent multi-view subspace clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 86–99, 2018.

[9] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, and Dong Huang, "Multi-view clustering in latent embedding space," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, pp. 3513–3520.

[10] Chris HQ Ding, Tao Li, and Michael I Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2008.

[11] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the SIAM International Conference on Data Mining*, 2013, pp. 252–260.

[12] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2010.

[13] Xiaofeng Zhu, Shichao Zhang, Wei He, Rongyao Hu, Cong Lei, and Pengfei Zhu, "One-step multi-view spectral clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 10, pp. 2022–2034, 2018.

[14] Andrew Ng, Michael Jordan, and Yair Weiss, "On spectral clustering: Analysis and an algorithm," in *Proceedings of the Advances in Neural Information Processing Systems*, 2001, vol. 14, pp. 849–856.

[15] A Monadjemi, BT Thomas, and M Mirmehdi, "Experiments on high resolution images towards outdoor scene classification," in *Proceedings of the Computer Vision Winter Workshop*, 2002, pp. 325 – 334.

[16] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014, vol. 28, pp. 2149–2155.

[17] Derek Greene and Pádraig Cunningham, "A matrix factorization approach for integrating multiple data views," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2009, pp. 423–438.

[18] Xiao Cai, Feiping Nie, and Heng Huang, "Multi-view k-means clustering on big data," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2013, pp. 2598–2604.

[19] Jinglin Xu, Junwei Han, Feiping Nie, and Xuelong Li, "Re-weighted discriminatively embedded k-means for multi-view clustering," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 3016–3027, 2017.

[20] Danyang Wu, Feiping Nie, Rong Wang, and Xuelong Li, "Multi-view clustering via mixed embedding approximation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 3977–3981.

[21] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan, *Introduction to Information Retrieval*, vol. 39, 2008.