Multiview Subspace Clustering for Multi-kernel Low-Redundancy Representation Learning

Zhuo Wang[1](✉), Ao Li[1], Jie Li[2], and Yangwei Wang[2]

[1] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China
wz1997sapphire@163.com

[2] Shandong Provincial Innovation and Practice Base for Postdoctors, Weihaizhenyu Intelligence Technology Co., Ltd., Weihai, China

Abstract. The purpose of the multiview subspace clustering algorithm is to construct a consensus subspace representation matrix by looking for complementary information among multiple views. However, most of the existing algorithms only learn the common information shared between multiple views and ignore the different information among multiple views, which will also have a positive impact on the clustering effect. To solve this problem, we integrate the subspace representation matrix of all views, introduce tensor analysis, and learn to obtain the low-rank tensor subspace representation matrix to capture the high-order correlation between multiple views. Comprehensive experiments have been conducted on three data sets, and the experimental results show that the proposed algorithm is much better than the comparison algorithms in recent literature, and the superiority of the proposed algorithm is verified.

Keywords: Multiview subspace clustering · Multi-kernel · Low-redundancy representation

1 Introduction

In real life, with the development of advanced technology, an object can often be fully described by multiple views. For example, using different languages to describe the same thing; shoot the same object from different angles. These views can be various measurements, such as real-time images, text records in different languages, etc. Each of these views generates a description of the same object, and different views typically describe the same and different information about the object. Learning better representations by analyzing all views is a challenging problem [1].

Considering that a single view is not enough to describe the information of data, subspace clustering has been extended to the case of multiple views. Different from the single-view subspace clustering method, Xu [2] proposed that the multiview learning method should fully use the principles of consensus and complementarity to ensure success. In order to realize the consensus principle, a classical strategy is to obtain a

S. Shi et al. (Eds.): 6GN 2021, LNICST 439, pp. 285–294, 2022.
https://doi.org/10.1007/978-3-031-04245-4_25

potential subspace that can be shared by multiple views through subspace learning [3]. At the same time, some scholars proposed many algorithms to explore the complementary information between different views, and achieved good performance. Gao [4] proposed a algorithm, which performed subspace clustering on each view of the data and used a common class indicator matrix to ensure the consistency of the data. Although the effect of this algorithm is better than that of single-view clustering, it does not restrict the self-representation coefficient of the data. It cannot well explore the internal structure of the data. Zhang [5] presented a method to learn the potential representation based on the features of multiple views and generate a common subspace representation to explore the potential complementary information among multiple views. Inspired by Sparse Subspace Clustering (SSC) [6] and Low-Rank Representation (LRR) [7], Zhang [8] proposed a algorithm, which extended LRR-based subspace clustering to multiview learning using rank-constrained subspace tensors of different modal expansion. It is a good way to explore complementary information from multiple sources, and greatly improve the subspace clustering.

Although the above multiview subspace clustering method has good performance, it still has some limitations. These clustering methods only use the linear subspace of the data in the original feature space, which is not enough to capture the complex correlation between actual data. In this regard, Abavisani [9] proposed a multi-mode extension method based on SSC and LRR subspace clustering algorithms, and used kernel learning to make the multimodal subspace clustering method nonlinear. Li [10] proposed to capture higher-order, nonlinear relationships between different views by introducing the Hilbert Schmidt Independent Standard (HSIC).

The above multiview subspace clustering method is proved to be effective in many scenarios. However, most of the above models use the original data and corresponding kernel matrix as the model's input, and both the original data and corresponding kernel matrix contain a lot of redundant information, which will affect the clustering effect. To solve this problem, we present a multiview subspace clustering method with multi-kernel and low redundancy representation learning. Raw data through learning method to get the corresponding kernel matrix, at the same time to avoid a single-kernel method is heavily dependent on the choice of kernel function problems, we select multiple kernel function to map, respectively mapped to the corresponding high-dimensional reproducing kernel Hilbert space. Then, the kernel matrix is removed by feature decomposition, and the low redundancy data self representation matrix is obtained, and then the subspace representation matrix is constructed on each view, and then the subspace representation matrix of each view is collected to construct the tensor to capture the consistency and difference information between different graphs, to obtain better clustering results. The contributions are summarized as follows.

(1) The low redundant data representation is used to replace the original data as the model input, thus reducing the influence of the redundant information in the original data on the clustering effect.
(2) Feature decomposition into multiview subspace clustering algorithm provides low redundancy data representation. At the same time, the multiview subspace clustering algorithm leads the eigendecomposition to produce the low redundancy data representation, which is more suitable for clustering.

(3) Integrate the subspace representation matrix of all views to construct a tensor to explore the high-order correlation among multiple views and obtain a better clustering effect.

2 Related Work

This paper mainly explores the multiview subspace clustering method for multi-kernel low-redundant representation learning. The following will introduce some basic related work and analysis involved in this method.

2.1 Subspace Clustering

Given a dataset $\mathbf{X} \in \mathbb{R}^{\mathrm{d}\times \mathrm{n}}$ of n data from k class clusters. The general formula of subspace clustering algorithm can be expressed as:

$$\min_{\mathbf{Z}} \ \mathcal{L}(\mathbf{X}, \mathbf{XZ}) + \lambda\Omega(\mathbf{Z}) \quad \text{s.t.} \ \mathbf{Z} \in \mathbb{R}^{n\times n} \tag{1}$$

where $\mathcal{L}(\cdot)$ and $\Omega(\cdot)$ represent the regularization terms. Considering the noise contained in the original data, an error matrix $\mathbf{E}$ was adopted, and the least square regression was proposed to express the target formula:

$$\begin{aligned} &\min_{\mathbf{Z}} \ \|\mathbf{E}\|_F + \lambda\|\mathbf{Z}\|_F \\ &\text{s.t.} \ \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \quad \mathrm{diag}(\mathbf{Z}) = \mathbf{0} \end{aligned} \tag{2}$$

2.2 Multiview Subspace Clustering

Given data containing V views, where $\mathbf{X}_v \in \mathbb{R}^{\mathrm{d_v}\times \mathrm{n}}$ represents the original data, d_v means the characteristic dimension of the v view, n represents the number of samples of the original data, and $\mathbf{Z}_v$ is the subspace representation matrix of the v view. The multiview subspace clustering algorithm can be expressed as:

$$\min_{\{\mathbf{Z}_v\}_{i=1}^{v}, \mathbf{Z}} \ \mathcal{L}(\{\mathbf{X}_\mathrm{v}, \ \mathbf{X}_\mathrm{v}\mathbf{Z}_\mathrm{v}\}_{\mathrm{v}=1}^{\mathrm{V}}) + \lambda\Omega(\{\mathbf{Z}_\mathrm{v}\}_{\mathrm{v}=1}^{\mathrm{V}}, \mathbf{Z}) \tag{3}$$

We extend the least square regression algorithm in (2) to the framework that follows the multiview subspace clustering algorithm in (3), and get the objective function of the least square regression algorithm on multiview subspace clustering as follows:

$$\begin{aligned} &\min_{\mathbf{Z}_v, \beta} \ \lambda\sum_{v=1}^{V}\|\mathbf{Z}_v\|_F^2 + \sum_{v=1}^{V}\beta_v\|\mathbf{X}_v - \mathbf{X}_v\mathbf{Z}_v\|_F^2 \\ &\text{s.t.} \ \mathrm{diag}(\mathbf{Z}_v) = \mathbf{0} \end{aligned} \tag{4}$$

where λ represented the trade-off parameter, and β_v represents the weight coefficients of different views.

3 Model

3.1 Objective

We first define several kernel mappings $\{\ \phi_s(\cdot)\}\ _{s=1}^{S}$. For the v - th view, the kernel matrix is calculated as follows:

$$\mathbf{K}_s^{(v)}(i,j) = \phi_s\left(\mathbf{x}_i^{(v)}\right)^{\mathrm{T}}\phi_s\left(\mathbf{x}_j^{(v)}\right) \tag{5}$$

where $i, j \in \{1, 2, \cdots, n\}$ represents the instance index, $x_i^{(v)}$ represents the i - th column vector of the v - th view, $\mathbf{K}_s^{(v)}$ represents the s - th kernel matrix of the i - th view. According to $m = S{*}\mathrm{V}$, it can be seen that when there are V views and S kings of kernel mappings, there will be m corresponding kernel matrices. The set of its kernel matrix is $\{\ \mathbf{K}_q\}\ _{q=1}^{m}$.

Literature [11] pointed out that the redundant information contained in the original data could not be removed by simple kernel learning. Thus, we use the eigendecomposition method to obtain the low redundancy data representation, i.e.

$$\begin{aligned} &\underset{\mathbf{U}_q}{\arg\max}\quad \mathrm{Tr}(\mathbf{UKU}^{\mathrm{T}}) \\ &\text{s.t.}\ \ \mathbf{U} \in \mathbb{R}^{\mathrm{c}\times\mathrm{n}} \end{aligned} \tag{6}$$

We use the low redundancy data obtained in (6) to represent, i.e. $\{\mathbf{U}_q\}_{q=1}^{m}$, replace the input of Eq. (4), i.e. $\{\mathbf{X}_q\}_{q=1}^{m}$. By integrating these two processes into one framework, we can get:

$$\begin{aligned} &\min_{\{\mathbf{U}_q\}_{q=1}^{m},\ \mathbf{Z}_q,\ \beta,\ \gamma} \quad \lambda\sum_{q=1}^{m}\left\|\mathbf{Z}_q\right\|_F^2 + \sum_{q=1}^{m}\beta_q\left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Z}_q\right\|_F^2 - \sum_{q=1}^{m}\gamma_q\mathrm{Tr}(\mathbf{U}_q\mathbf{K}_q\mathbf{U}_q^{\mathrm{T}}) \\ &\text{s.t.}\ \ \mathrm{diag}(\mathbf{Z}_q) = \mathbf{0} \end{aligned} \tag{7}$$

where γ_q represents the weight coefficients of different views.

Meanwhile, we propose to use the tensor kernel norm to capture the higher-order correlation between different views. We construct a low-rank tensor subspace representation matrix $\mathcal{Z}$ of order three by integrating all subspace representation matrix $\mathbf{Z}_q$, and obtain the primary objective function of our model:

$$\begin{aligned} &\min_{\{\mathbf{U}_q\}_{q=1}^{m},\ \{\mathbf{Z}_q\}_{q=1}^{m},\ \mathcal{Z},\ \beta,\ \gamma} \quad \lambda\|\mathcal{Z}\|_{\circledast} + \sum_{q=1}^{m}\beta_q\left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Z}_q\right\|_F^2 - \sum_{q=1}^{m}\gamma_q\mathrm{Tr}\left(\mathrm{U}_q\mathrm{K}_q\mathrm{U}_q^{\mathrm{T}}\right) \\ &\text{s.t.}\ \ \mathcal{Z} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}\times\mathrm{q}},\ \mathbf{Z}_{\mathrm{q}} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}},\ \mathrm{U}_{\mathrm{q}}\mathbf{U}_{\mathrm{q}}^{\mathrm{T}} = \mathbf{I},\ \mathbf{U}_{\mathrm{q}} \in \mathbb{R}^{\mathrm{c}\times\mathrm{n}} \\ &\beta^{\frac{1}{2}\mathrm{T}}\mathbf{1} = 1,\ \beta \in \mathbb{R}_+^{\mathrm{m}},\ \gamma^{\mathrm{T}}\gamma = 1,\ \gamma \in \mathbb{R}_+^{\mathrm{m}} \end{aligned} \tag{8}$$

3.2 Optimization

An alternate optimization strategy is adopted to minimize the solution of each variable iteratively under the condition that other variables remain unchanged.

Due to the high correlation between variables $\mathcal{Z}$ and $\mathbf{Z}_q$, it is tough to solve Eq. (8). We introduce auxiliary variable $\mathcal{Q}$ to make the variables separable, so optimization the Eq. (8) becomes:

$$\begin{aligned} \min_{\{\mathbf{U}_q\}_{q=1}^m,\ \{\mathbf{Z}_q\}_{q=1}^m,\ \mathcal{Q},\ \mathcal{Z},\ \beta,\ \gamma} \quad & \lambda\|\mathcal{Q}\|_{\circledast} + \|\mathcal{Q}-\mathcal{Z}\|_F^2 + \sum_{q=1}^m \beta_q \left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Z}_q\right\|_F^2 \\ & - \sum_{q=1}^m \mathrm{Tr}\Big(\mathbf{U}_q\mathbf{K}_q\mathbf{U}_q^{\mathrm{T}}\Big) \\ \text{s.t.} \quad & \mathcal{Q} = \mathcal{Z} \end{aligned} \tag{9}$$

$\mathcal{Z}$-subproblem: It's clear that each of these subspaces represents the matrix $\{\mathbf{z}_q\}_{q=1}^m$ that is independent, so fixed $\{\mathbf{Z}_t\}_{t=1,t\neq q}^m$, $\mathcal{Z}$, $\mathcal{Q}$, $\{\mathbf{U}_q\}_{q=1}^m$, β and γ. The optimization Eq. (9) becomes:

$$\begin{aligned} \min_{\mathbf{Z}_q} \quad & \left\|\mathbf{Z}_q - \mathbf{Q}_q\right\|_F^2 + \beta_q\left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Z}_q\right\|_F^2 \\ \text{s.t.} \quad & \mathbf{Z}_q \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}} \end{aligned} \tag{10}$$

The Eq. (11) is optimized by using the formula $\mathbf{Z}_q - \mathbf{Q}_q = \mathbf{Z}'_q$:

$$\begin{aligned} \min_{\mathbf{Z}'_q} \quad & \left\|\mathbf{Z}'_q\right\|_F^2 + \beta_q\left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Q}_q - \mathbf{U}_q\mathbf{Z}'_q\right\|_F^2 \\ \text{s.t.} \quad & \mathrm{diag}(\mathbf{Z}'_q) = 0 \end{aligned} \tag{11}$$

The Eq. (11) is optimized by using the formula $\mathbf{U}'_q = \mathbf{U}_q - \mathbf{U}_q\mathbf{Q}_q$:

$$\min_{\mathbf{Z}'_q} \quad \left\|\mathbf{Z}'_q\right\|_F^2 + \beta_q\left\|\mathbf{U}'_q - \mathbf{U}_q\mathbf{Z}'_q\right\|_F^2 \tag{12}$$

Since the diagonal of $\mathbf{Z}'_q$ is forced to be 0, we remove the i - th column of $\mathbf{U}'_q = \{\mathrm{u}_q^{'(i)}\}_{i=1}^n \in \mathbb{R}^{\mathrm{c}\times\mathrm{n}}$ to get $\mathbf{F}_q^{(i)} = \{\mathrm{u}_q^{'(1)}, \ldots, \mathrm{u}_q^{'(i-1)}, \mathrm{u}_q^{'(i+1)}, \ldots, \mathrm{u}_q^{'(n)}\} \in \mathbb{R}^{\mathrm{c}\times\mathrm{n}}$, and optimize each column of $\mathbf{Z}'_q$ to get:

$$\begin{aligned} \min_{\mathbf{z}_q^{'(i)}} \quad & \left\|\mathbf{z}_q^{'(i)}\right\|_F^2 + \beta_q\left\|\mathrm{u}_q^{'(i)} - \mathbf{F}_q^{(i)}\mathbf{z}_q^{'(i)}\right\|_F^2 \\ \text{s.t.} \quad & \mathbf{z}_q^{'(i)} \in \mathbb{R}^{\mathrm{n}-1} \end{aligned} \tag{13}$$

Then optimization $\mathbf{Z}_q^{'(i)*}-$ is:

$$\begin{aligned} & \mathrm{z}_q^{'(i)*} = (\mathbf{E}_q^{(i)})^{-1}\beta_q\mathbf{u}_q^{'(i)}\mathbf{F}_q^{(i)\mathrm{T}} \\ \text{s.t.} \quad & \mathbf{E}_q^{(i)} = \mathbf{I} + \beta_q\mathbf{F}_q^{(i)\mathrm{T}}\mathbf{F}_q^{(i)} \end{aligned} \tag{14}$$

$\mathcal{Q}$-subproblem: fixed $\mathcal{Z}$, $\mathbf{Z}_q$, $\{\mathbf{U}_q\}_{q=1}^m$, β and γ. Optimization Eq. (9) becomes:

$$\begin{aligned} &\min_{\mathcal{Q}} \ \lambda\|\mathcal{Q}\|_{\circledast} + \|\mathcal{Q} - \mathcal{Z}\|_F^2 \\ &\text{s.t.} \ \ \mathcal{Q} \in \mathbb{R}^{\mathrm{n \times n \times q}} \end{aligned} \tag{15}$$

According to the tensor kernel norm minimization algorithm based on T-SVD in reference [12], it can be solved.

U-subproblem: Because the data with low redundancy means that $\{\mathbf{U}_q\}_{q=1}^m$ is independent of each other, fixed $\{\mathbf{U}_t\}_{t=1,t\neq q}^m$, $\mathcal{Z}$, $\mathcal{Q}$, $\{\mathbf{Z}_q\}_{q=1}^m$, β and γ. The optimization Eq. (9) becomes:

$$\min_{\mathbf{U}_q} \ \beta_q\left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Z}_q\right\|_F^2 - \gamma_q \mathrm{Tr}(\mathbf{U}_q\mathbf{K}_q\mathbf{U}_q^{\mathrm{T}}) \tag{16}$$

Equation (16) can be translated into:

$$\begin{aligned} &\max_{\mathbf{U}_q} \ \ \mathrm{Tr}(\mathbf{U}_q\mathbf{N}_q\mathbf{U}_q^{\mathrm{T}}) \\ &\text{s.t.} \ \ \mathrm{N}_q = 2\beta_q\mathbf{Z}_q^{\mathrm{T}} - \beta_q\mathbf{Z}_q\mathbf{Z}_q^{\mathrm{T}} + \gamma_q\mathbf{K}_q \end{aligned} \tag{17}$$

Equation (17) can be efficiently solved by eigendecomposition.

β-subproblem: fixed $\mathcal{Z}$, $\mathcal{Q}$, $\{\mathbf{Z}_q\}_{q=1}^m$, $\{\mathbf{U}_q\}_{q=1}^m$ and γ. Optimization Eq. (9) becomes:

$$\begin{aligned} &\min_{\beta} \ \ \beta^{\mathrm{T}}v \\ &\text{s.t.} \ \ v_q = \left\|\mathbf{U}_q - \mathbf{U}_q\mathbf{Z}_q\right\|_F^2 \end{aligned} \tag{18}$$

According to the Cauchy-Schwartz inequality, we optimize the Eq. (18) as:

$$\beta_q^* = 1/\left(v_q\sum_{t=1}^m \frac{1}{v_t}\right)^2 \tag{19}$$

γ-subproblem: fixed $\mathcal{Z}$, $\mathcal{Q}$, $\{\mathbf{Z}_q\}_{q=1}^m$, $\{\mathbf{U}_q\}_{q=1}^m$ and β. Optimization Eq. (9) becomes:

$$\begin{aligned} &\max_{\gamma} \ \ \gamma^{\mathrm{T}}v \\ &\text{s.t.} \ \ v_q = \mathrm{Tr}(\mathbf{U}_q\mathbf{K}_q\mathbf{U}_q^{\mathrm{T}}) \end{aligned} \tag{20}$$

According to the Cauchy-Schwartz inequality, we optimize the Eq. (20) as:

$$\gamma_q^* = v_q/\left(\sum_{t=1}^m v_t^2\right)^{\frac{1}{2}} \tag{21}$$

To sum up, we summarize the solving steps of the objective function of the proposed model in Algorithm 1.

Using the low-rank tensor subspace representation matrix $\mathcal{Z}$, $\mathbf{J} = \frac{1}{m}\sum_{q=1}^{m}\left(\left|\mathcal{Z}^{(q)} + \mathcal{Z}^{(q)\mathrm{T}}\right|\right)$ is used to calculate the fusion subspace representation matrix $\mathbf{J}$ ($\mathcal{Z}^{(q)}$ represents the q slice of $\mathcal{Z}$ along the angle of view), and then using the fusion subspace representation matrix, $\mathbf{T} = \frac{1}{2}(\mathbf{J} + \mathbf{J}^{\mathrm{T}})$ is used to calculate the affinity subspace representation matrix $\mathbf{T}$, which is sent into the spectral clustering algorithm to calculate the clustering results.

Algorithm 1

Input：the original data $\{\mathbf{X}_v\}_{v=1}^{V}$, dimensions represented by low redundancy data c and parameter λ.

Output：the low-rank tensor subspace representation matrix $\mathcal{Z}$.

1. Generate the kernel matrices $\{\mathbf{K}_q\}_{q=1}^{m}$ from $\{\mathbf{X}_v\}_{v=1}^{V}$.
2. Initialize $\mathcal{Q}$, $\{\mathbf{U}_q\}_{q=1}^{m}$, β and γ.
3. **while** $(obj^{t-1} - obj^{t}) / obj^{t} \le \sigma$ **do**
4. Update $\mathbf{Z}_q^{'}$ with Eq. (14).
5. According to formula $\mathbf{Z}_q = \mathbf{Z}_q^{'} + \mathbf{Q}_q$, get the $\mathbf{Z}_q$.
6. Structure tensor $\mathcal{Z} = \phi(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_q)$.
7. Update $\mathcal{Q}$ with Eq. (15).
8. Update $\{\mathbf{U}_q\}_{q=1}^{m}$ with Eq. (17).
9. Update β with Eq. (19).
10. Update γ with Eq. (21).
11. $t = t + 1$.
12. Calculate objective value obj^{t} with Eq. (10).
13. **end while**

4 Experiment

4.1 Dataset

The following three data sets are used to evaluate the effectiveness and superiority.

BBC-Sport It consisting of 737 files from the BBC-Sport website, corresponding to sports news in five subject areas. There are two different views.

ORL It consists of 400 facial images of 40 people of different genders, in which each person has 10 facial images taken from different angles. There are three different views.

UCI-Digits It consists of 2,000 digital images corresponding to 10 categories. Fourier coefficient, pixel average and morphological features are extracted to represent these digital images. There are three different views.

4.2 Comparison of Experimental Settings and Experimental Results

In the experiment, in order to prove the effectiveness of the presented method, this paper will compare the above three data sets with three advanced multiview subspace clustering methods. These three methods are: Latent multi-view subspace clustering (LMSC) in [5], Multimodal sparse and low-rank subspace clustering (MSSC) in literature [9], and Multiview subspace clustering via co-training robust data representation (CoMSC) in [11].

We run each method on each dataset ten times, and take the mean value of the experimental results the 10 times as the result of the clustering evaluation index. For the clustering results, we used three measurement methods of ACC, NMI and purity to evaluate the clustering results. For these indicators, the higher the value, the better the clustering effect.

Table 1. ACC, NMI and Purity comparison of different clustering algorithms on the three public datasets.

	Method	ACC	NMI	Purity
BBC-Sport	LMSC	0.891 ± 0.02	0.796 ± 0.03	0.891 ± 0.02
	MSSC	0.971 ± 0.00	0.898 ± 0.00	0.971 ± 0.00
	CoMSC	0.956 ± 0.02	0.876 ± 0.02	0.956 ± 0.02
	Ours	**0.979 ± 0.02**	**0.933 ± 0.04**	**0.979 ± 0.02**
ORL	LMSC	0.809 ± 0.04	0.909 ± 0.02	0.840 ± 0.03
	MSSC	0.835 ± 0.00	**0.930 ± 0.00**	0.865 ± 0.00
	CoMSC	0.805 ± 0.02	0.903 ± 0.01	0.833 ± 0.01
	Ours	**0.844 ± 0.02**	0.922 ± 0.01	**0.870 ± 0.01**
UCI-Digits	LMSC	0.862 ± 0.03	0.784 ± 0.03	0.862 ± 0.03
	MSSC	0.924 ± 0.00	0.859 ± 0.00	0.924 ± 0.00
	CoMSC	0.903 ± 0.02	0.837 ± 0.03	0.903 ± 0.02
	Ours	**0.986 ± 0.01**	**0.967 ± 0.01**	**0.986 ± 0.01**

The experimental results of the three indexes are shown in Table 1. It can be seen from the experimental results that the method presented in this paper is better than other methods in almost all three indexes. In the there datasets of BBC-Sport, ORL and UCI-Digits, in the ACC index and Purity index, the performance of the presented method was roughly the same as that of the MSSC method in the BBC-Sport and ORL datasets, 6.2% on UCI-Digits; In terms of NMI index, the presented method is slightly lower than the MSSC method only in the ORL dataset and superior to the comparison method in all other datasets.

4.3 Parameter Selection and Convergence Verification

For the method of comparison, we adjusted all the parameters for optimal performance. For our approach, Gaussian kernel mapping, polynomial kernel mapping, linear kernel mapping, symbolic polynomial kernel mapping and inverse polynomial kernel mapping are applied to obtain the kernel matrix corresponding to the original data. We will set the value range of dimension c represented by low redundancy data as $\{k, 2k, \ldots, 20k\}$ and parameter λ as $2.\text{^}\{-10, -8, \ldots, 10\}$, and select the parameter combination with the optimal clustering performance to conducting experiment.

To verify the convergence of the proposed method, the target value of the algorithm is used as the convergence criterion for each iteration. In this paper, the convergence curves of the BBC-Sport dataset at $\lambda=2^0$ and $c = 20k$ are given, as shown in Fig. 1. The method in this paper converges between 8 and 10 iterations with stable convergence performance.

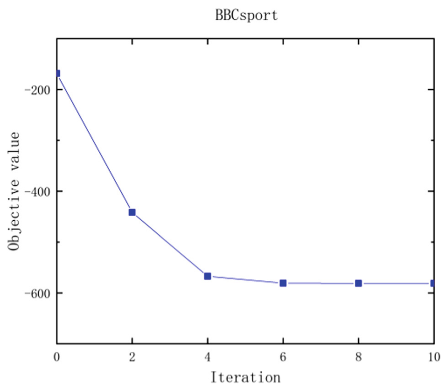

Fig. 1. Convergence validation on and BBC-Sport.

5 Conclusion

In this paper, we propose a multiview subspace clustering method for multi-kernel and low redundancy representation learning. This method uses multi-kernel learning to capture the more complex correlation information between actual data. It uses the feature decomposition method to remove the redundant information in the data to obtain the low redundant data representation. To explore the complementarity information among multiple views and the difference information among multiple views, the subspace representation matrix of each view is reorganized into tensor form, and the high-order correlation among multiple views is explored by using tensor low-rank constraint. The proposed multiview clustering model is compared with some classical multiview subspace clustering methods on three public datasets, and the convergence and parameters of the model are analyzed. Experimental results show that the effectiveness of the proposed model are superior to other methods.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China under Grant 62071157, Natural Science Foundation of Heilongjiang Province under Grant YQ2019F011 and Postdoctoral Foundation of Heilongjiang Province under Grant LBH-Q19112.

References

1. Zhao, J., Xie, X., Xu, X., et al.: Multi-view learning overview: recent progress and new challenges. Inf. Fusion **38**, 43–54 (2017)
2. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. arXiv preprint arXiv:1304.5634 (2013)
3. Chaudhuri, K., Kakade, S.M., Livescu, K., et al.: Multi-view clustering via canonical correlation analysis. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 129–136 (2009)
4. Gao, H., Nie, F., Li, X., et al.: Multi-view subspace clustering. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4238–4246 (2015)
5. Zhang, C., Hu, Q., Fu, H., et al.: Latent multi-view subspace clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4279–4287 (2017)
6. Elhamifar, E., Vidal, R.: Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **35**(11), 2765–2781 (2013)
7. Liu, G., Lin, Z., Yan, S., et al.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2012)
8. Zhang, C., Fu, H., Liu, S., et al.: Low-rank tensor constrained multiview subspace clustering. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1582–1590 (2015)
9. Abavisani, M., Patel, V.M.: Multimodal sparse and low-rank subspace clustering. Inf. Fusion **39**, 168–177 (2018)
10. Li, R., Zhang, C., Hu, Q., et al.: Flexible multi-view representation learning for subspace clustering. In: IJCAI, pp. 2916–2922 (2019)
11. Liu, J., Liu, X., Yang, Y., et al.: Multiview subspace clustering via co-training robust data representation. IEEE Trans. Neural Netw. Learn. Syst. **PP**, 1–13 (2021)
12. Kilmer, M.E., Braman, K., Hao, N., et al.: Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. SIAM J. Matrix Anal. Appl. **34**(1), 148–172 (2013)