Neural Computing and Applications
https://doi.org/10.1007/s00521-020-05185-y

ORIGINAL ARTICLE

Multi-view clustering via neighbor domain correlation learning

Xiaocui Li[1] · Ke Zhou[1] · Chunhua Li[1] · Xinyu Zhang[2] · Yu Liu[1] · Yangtao Wang[1]

Received: 2 February 2020 / Accepted: 11 July 2020

Abstract
With the development of data science, more and more data are presented in the form of multi-view. Compared with single-view feature learning, multi-view feature learning is more effective, and it has been successfully applied in many fields. Clustering is a core technology of computer science. Thus, many researchers start to study multi-view clustering. Recently, combining with multi-view feature learning techniques, some multi-view clustering methods have been presented. These methods mainly focus on the multiple features fusion, while most of them ignore the correlations among multiple views. Therefore, it cannot make full use of the advantages of multiple view features. In this paper, we propose a novel approach, named multi-view clustering via neighbor domain correlation learning (MCNDCL) approach. Specifically, MCNDCL learns a discriminant common space for multiple view features. Under the learned common space, the correlations of the consistent neighbor domain are maximized, and the correlations of specific neighbor domain are minimized at the same time. Extensive experimental results on four typical benchmarks, i.e., UCI Digits, Caltech7, BBCSport and CCV, validate the high effectiveness of our proposed approach.

Keywords Clustering · Multi-view learning · Neighbor domain correlation · Feature learning

1 Introduction

As one of the most important technologies of computer science, clustering technique can be applied in many real tasks, such as data mining [1–4], machine learning [5–7] and computer vision [8, 9]. Initially, clustering is conducted in the single-view feature or data, and numerous of single-view clustering methods have been proposed in the past [10–15]. However, with the development of information technology, lots of data may consist of multiple views or features. For instance, a webpage can be represented by a page view and link view. An image can be described by a variety of descriptors such as color, shape and texture.

Compared with single-view data, multi-view data contain more useful and abundant information. Different views bring not only diversity but also redundancy. Therefore, the key of multi-view features learning is how to effectively utilize the consensus and complementary information containing in the multi-view data while reducing the redundancy. Of course, we can solve this problem by converting the multi-view data to single-view data. For instance, we execute clustering process in each view and select the best view result as the final cluster results, or combine multi-view features linearly to form a new single-view features, and then feed them to the transitional clustering models. However, the crudely processing of multi-view features cannot make full use the information of data, so they usually cannot achieve satisfied clustering results [16, 17].

In order to make full use of multi-view features, more and more works focus on multi-view learning technique, and most of them obtain better performance than single-view learning technique in many application domains. Therefore, in clustering community, some researchers try to solve multi-view clustering problem by combining multi-view learning techniques, and some multi-view clustering methods have been presented recently. They can

✉ Chunhua Li
li.chunhua@hust.edu.cn

Xiaocui Li
lxc@hust.edu.cn

[1] Wuhan National Laboratory for Optoelectronics and School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China

[2] School of Computer Science, Wuhan University, Wuhan, China

Published online: 25 July 2020

be roughly categorized into three kinds: (i) Subspace learning or kernel learning-based methods [18–21]: This kind of methods tries to transform multi-view features into a common subspace feature. (ii) Graph embedding-based methods [22–24]: These methods mainly focus on maximizing the consistency of the multi-view. (iii) Matrix factorization-based methods [17]: In this community, they usually obtain a common indicator matrix by combining the multi-view feature together.

1.1 Motivation

Although existing multi-view-based clustering methods have achieved good results, there are still much room to improve. Specifically, as shown in Fig. 1, most of these methods mainly focus on the maximized clustering consistency of multi-view. Therefore, the complementary information of multi-view is not fully utilized, especially for the specific information contained in each view.

Motivated by the above analysis, we propose a novel multi-view clustering approach to fully use the complementary information of the multi-view data, which can simultaneously maximize the correlations of consistent neighbor and minimize the correlations of specific neighbor by neighbor domain correlation learning, as shown in Fig. 2. Experimental results show that our approach can achieve better clustering performance.

1.2 Contribution

The main contributions of our proposed approach lie in three aspects:

(1) We design a novel neighbor domain correlation learning to solve the multi-view clustering problem. The objective function of neighbor domain correlation learning is designed to maximize the correlations of consistent neighbor domain and minimize the correlations of specific neighbor domain.
(2) We transform the designed objective function into a convex quadratic programming problem, which can be solved by Lagrangian algorithm. Therefore, our approach does not need iterative calculation, and it can obtain the global optimal solution only by analytical solution.
(3) To verify the effectiveness of MCNDCL, extensive experiments have been carried out on four widely used clustering datasets. Experimental results show that our approach can achieve better clustering performance than existing mainstream methods. In addition, to further verify the robustness of MCNDCL, we have evaluated the impact of noise at different attributes and the influence of key parameters for the clustering performance. In

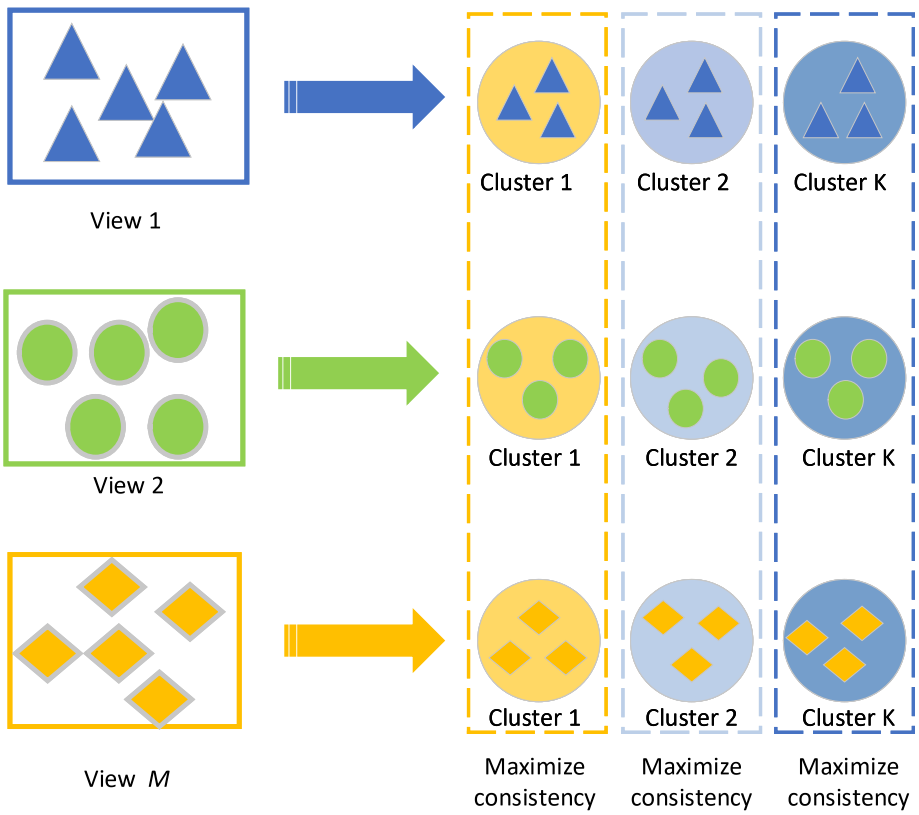

Fig. 1 The diagram of maximized clustering consistency of multi-view

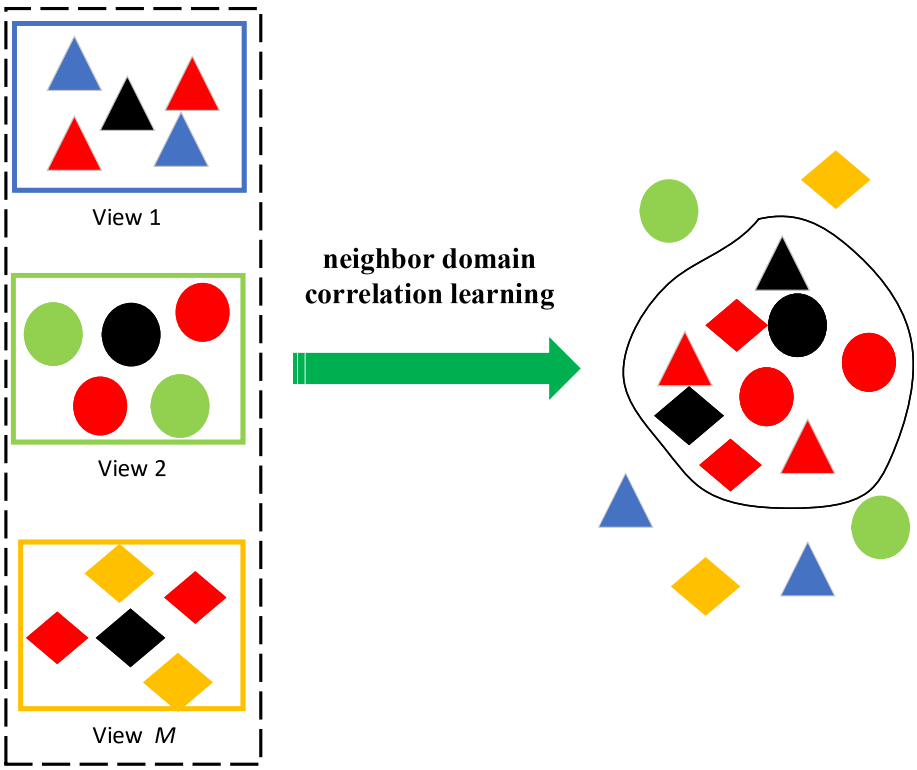

Fig. 2 The diagram of the neighbor domain correlation learning. Black block is a anchor. Red blocks are consistent neighbor domain of the anchor, and other color blocks are specific neighbor domain of the anchor

addition, we have analyzed the time performance of our approach.

2 Related work

In this section, we will briefly review existing mainstream multi-view clustering algorithm, i.e., multi-view spectral clustering, multi-view subspace clustering and multi-view nonnegative matrix factorization clustering.

2.1 Multi-view spectral clustering

As one of the most important clustering methods, spectral clustering utilizes the graph structure to solve the clustering problem [10, 25]. Given a weighted undirected graph $G = (V, E)$ as dataset where vertex set $V = \{v_1, v_2, \ldots, v_n\}$ represents the data points. Adjacency matrix of the graph is defined as W whose entry w_{ij} indicates the similarity of v_i and v_j. If vertex v_i and v_j are unconnected, $w_{ij} = 0$. Assume D is the degree matrix of this graph which is a diagonal matrix, $d_i = \sum_{j=1}^{N} w_{ij}$. Then, we can get the Laplacian matrix $L = D - W$. Normally in spectral clustering methodsthe embeddings of data points [10, 22, 26], it can be normalized as $L = D^{-1/2}WD^{-1/2}$. The object function of spectral clustering can be formulated as follows:

$$\begin{cases} \max\limits_{U \in \mathbb{R}^{N \times K}} tr(U^T L U) \\ s.t. \quad U^T U = I \end{cases} \tag{1}$$

The matrix U is the embeddings of data points which can be fed into the traditional clustering algorithms such as K-means.

In the multi-view situation, the spectral clustering process is performed on each view and then fuses their results [27]. The most classical two multi-view clustering algorithms are Co-Training [26, 28] and Co-Regularized multi-view spectral clustering [22, 29]. Co-training multi-view spectral clustering assumes that all the clustering results in each views should be consistent. The most confident clustering result in ith view is added to the jth($1 \le i, j \le m$, and $i \ne j$) view's labeled set to promote the clustering process until the clustering result in all the views is consistent.

$$\begin{cases} \max\limits_{U^1, U^2, \ldots, U^m \in \mathbb{R}^{N \times K}} \sum_{s=1}^{m} tr(U^{(s)T} L^{(s)} U^{(s)}) \\ s.t. U^{(s)T} U^{(s)} = I, \forall 1 \le s \le m. \end{cases} \tag{2}$$

While Co-Regularized multi-view spectral clustering introduce minimizing the distinction between different two views act as part of the objective function [22, 30].

$$\begin{cases} \max\limits_{U^1,U^2,\ldots,U^m\in\mathbb{R}^{N\times K}} \sum_{s=1}^{m} tr(U^{(s)T}L^{(s)}U^{(s)}) \\ \quad +\sum_{1\le s,t\le m,s\neq t} \lambda\ tr(U^{(s)}U^{(s)T}U^{(t)}U^{(t)T}) \\ s.t.\quad U^{(s)T}U^{(s)} = I, \forall 1\le s\le m. \end{cases} \tag{3}$$

where $D(U^{(s)}, U^{(t)}) = -tr(U^{(s)}U^{(s)T}U^{(t)}U^{(t)T})$ indicates the disagreement between two views, and λ is a hyperparameter.

2.2 Multi-view subspace clustering

In many practical applications, the dimension of data may be very high, while the effective dimension used to assist the clustering process is very low. In order to address this problem, subspace clustering emerged which tried to find latent effective low-dimensional subspace from high-dimensional data space and got the accurate clustering result in the identified subspaces. For the self-expressiveness property [31] of data samples, $X = XZ + E$. $Z = \{z_1, z_2, \ldots z_N\} \in \mathbb{R}^{N\times N}$ is the subspace representation matrix whose entry z_i is the embedding of data x_i mapped on subspace, and E is the error matrix. The object of subspace clustering [32] is as follows:

$$\begin{cases} \min\limits_{Z} ||X - XZ||_F^2 \\ s.t.\quad Z(i,i) = 0, Z^T 1 = 1 \end{cases} \tag{4}$$

Extending to multi-view subspace clustering, as each view's clustering result should be consistent, all the views share one common representation matrix [33–35]. Thus, the optimization problem can be formulated as

$$\begin{cases} \min\limits_{Z^{(s)},s=1,2,\ldots m} \sum_{s=1}^{m} ||X^{(s)} - X^{(s)}Z^{(s)}||_F^2 \\ \qquad +\alpha\sum_{1\le s\le t} ||Z^{(s)} - Z^{(t)}||_1 \\ s.t.\quad diag(Z^{(s)}) = 0, \forall s \in \{1,2,\ldots m\}. \end{cases} \tag{5}$$

where $||Z^{(s)} - Z^{(t)}||_1$ is the l_1-norm-based pairwise co-regularization constraint that can restrain the error problem.

After obtaining the subspace coefficient matrix Z, the similarity matrix can be computed by $\frac{|Z|+|Z|^T}{2}$ which further constructs the Laplacian matrix and executes spectral clustering [36].

2.3 Multi-view nonnegative matrix factorization clustering

As shown in Fig. 3, given a nonnegative data matrix $X \in \mathbb{R}^{N\times d}$, find two nonnegative matrix $U \in \mathbb{R}^{N\times K}$ and $V \in \mathbb{R}^{d\times K}$ to satisfy $X \approx UV^T$, where U is the indicator matrix and V is the basis matrix [37]. K represents the clusters number in clustering learning [38–40].

As discussed above, all the views should agree the final clustering result, so in multi-view *NMF* clustering, all the views share one common indicator matrix[41]. The objective of multi-view nonnegative matrix factorization clustering is as follows:

$$\begin{cases} \min\limits_{U^{(s)},V^{(s)},s=1,2,\ldots,m} \sum_{s=1}^{m} ||X^{(s)} - U^{(s)}V^{(s)T}||_F^2 \\ \qquad +\sum_{s=1}^{m} \lambda_s ||V^{(s)} - V^*||_F^2 \\ s.t.\quad \forall 1\le k\le K, ||U_{:,k}^{(s)}||_1 = 1, U^{(s)}, V^{(s)}, V^* \ge 0. \end{cases} \tag{6}$$

After getting the common indicator matrix V^*, the cluster result can be computed by $argmax_k V_{i,k}^*$ [18].

In summary, spectral clustering or graph clustering-based methods mainly focus on maximizing consistency information of multi-view data, while they do not use the specific information of each view. Most of subspace clustering or nonnegative matrix factorization-based clustering translates multi-view features into a common subspace by using linear transformation, matrix decomposition or tensor-based strategy, while ignoring the correlations of the different views.

3 Proposed method

3.1 Formulation

Suppose that the multi-view data are denoted by $X = \{(X^{(1)}, X^{(2)}, \ldots, X^{(M)}\}$, where M is the number of views, $X^{(m)} \in \mathbb{R}^{d^m\times N}$, d^m denotes the dimensionality of samples from the mth view, and N is the number of total samples. Our goal is to learn a feature projection transformation W that can project data points from different views to one discriminant common subspace, under which the correlations of consistent neighbor domain are maximized, while the correlations of specific neighbor domain are minimized at the same time.

First, we use S_i^m to represent k neighbor domain of the ith sample in the mth view. Consistent neighbor domain is the common neighbor domain of the all view, and the specific neighbor domain is the difference between sum of neighbor domain set and consistent neighbor domain set. Then, we can compute the consistent neighbor domain P_i and specific neighbor domain Q_i of each sample as follows:

$$P_i = \bigcap \sum_{m=1}^{M} S_i^m, \tag{7}$$

$$Q_i = \bigcup \sum_{m=1}^{M} S_i^m - P_i, \tag{8}$$

Therefore, the consistent neighbor domain correlations R_p and the specific neighbor domain correlations R_q can be separately defined as follows:

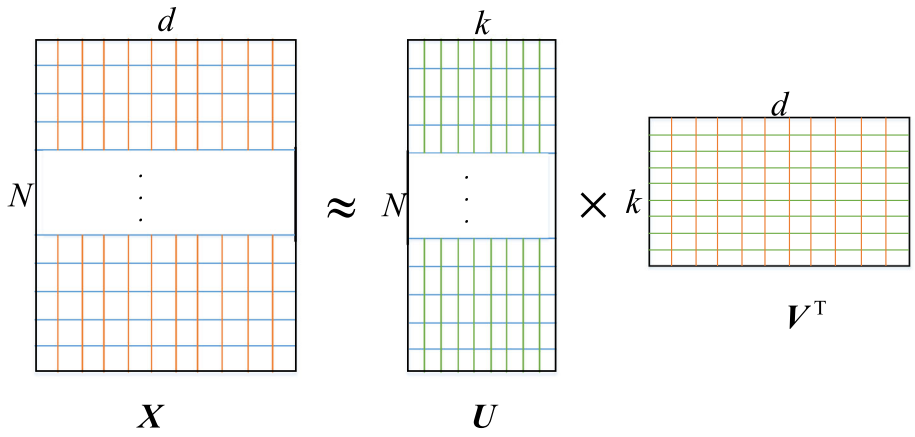

Fig. 3 The illustration of nonnegative matrix factorization

$$R_p = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|P_i|}\sum_{t=1}^{|P_i|}\frac{(p_s^m)^T WW^T(p_t^n)}{\sqrt{(p_s^m)^T WW^T(p_s^m)}\sqrt{(p_t^n)^T WW^T(p_t^n)}}, \tag{9}$$

$$R_q = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|Q_i|}\sum_{t=1}^{|Q_i|}\frac{(q_s^m)^T WW^T(q_t^n)}{\sqrt{(q_s^m)^T WW^T(q_s^m)}\sqrt{(q_t^n)^T WW^T(q_t^n)}}, \tag{10}$$

where $|P_i|$ and $|Q_i|$ indicate the number of the ith sample's consistent neighbor domain and specific neighbor domain correlations, respectively; $p_s^m(s \in [1, |P_i|])$ indicates the mth view's feature of the sth sample in the ith sample's consistent neighbor domain; $q_s^m(s \in [1, |Q_i|])$ indicates the mth view's feature of the sth sample in the ith sample's specific neighbor domain.

Combining formulas 9 and 10 , the objective function of our MCNDCL approach can be defined as:

$$\max_W f(W) = R_p - \lambda R_q \tag{11}$$

where λ is the positive tunable factor. For simplification, we set $D = WW^T$ and ($\|p_s^m\| = 1, \|q_s^m\| = 1$). Obviously, D should be a symmetric and positive semi-definite matrix, i.e., $D = D^T$ and $D \geq 0$. And then, we can relax 11 into the following equation:

$$\begin{aligned} &\max_D f(D) = R_p' - \lambda R_q' \\ &s.t. \quad D = D^T, D \geq 0, \end{aligned} \tag{12}$$

where R_p' and R_q' can be computed as follows:

$$\begin{aligned} R_p' &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|P_i|}\sum_{t=1}^{|P_i|}\frac{(p_s^m)^T WW^T(p_t^n)}{\|(p_s^m)\|\|(p_t^n)\|\|WW^T\|_F} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|P_i|}\sum_{t=1}^{|P_i|}\frac{(p_s^m)^T D(p_t^n)}{\|D\|_F}, \end{aligned} \tag{13}$$

$$\begin{aligned} R_q' &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|Q_i|}\sum_{t=1}^{|Q_i|}\frac{(q_s^m)^T WW^T(q_t^n)}{\|(q_s^m)\|\|(q_t^n)\|\|WW^T\|_F} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|Q_i|}\sum_{t=1}^{|Q_i|}\frac{(q_s^m)^T D(q_t^n)}{\|D\|_F}, \end{aligned} \tag{14}$$

To the end, Eq. 12 can be transformed to the following form:

$$\begin{aligned} &\min_D \|D\|_F \\ &s.t. \quad H \geq 1, D = D^T, D \geq 0, \end{aligned} \tag{15}$$

where

$$H = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|P_i|}\sum_{t=1}^{|P_i|}(p_s^m)^T D(p_t^n) - \frac{\lambda}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|Q_i|}\sum_{t=1}^{|Q_i|}(q_s^m)^T D(q_t^n).$$

3.2 Optimization procedure

In this subsection, we will introduce the optimization procedure of Eq. 15. To obtain the solution of Eq. 15, we design the following optimization scheme.

We will prove that D is symmetric matrix in the following contents, so we can leave the constraints $D = D^T, D \geq 0$ off. Thus, Eq. 15 can be simplified as follows:

$$\min_D \|D\|_F \quad s.t. \quad H \geq 1, \tag{16}$$

which is equivalent to the following convex quadratic optimization problem:

$$\min_D \frac{1}{2}\|D\|_F^2 \quad s.t. \quad H \geq 1, \tag{17}$$

In order to better solve the problem of Eq. 17, we introduce a slack variable ε to relax it. By doing so, Eq. 17 can be transformed as follows:

$$\begin{aligned} &\min_D \frac{1}{2}\|D\|_F^2 + \beta\varepsilon \\ &s.t. \quad H \geq 1 - \varepsilon, \varepsilon \geq 0, \end{aligned} \tag{18}$$

where λ is a regularization factor.

Obviously, Eq. 18 is a constrained Lagrangian problem. According to the Lagrange multiplier method, we construct the following Lagrangian function:

$$L(D, \varepsilon, \mu_1, \mu_2) = \frac{1}{2}\|D\|_F^2 + \beta\varepsilon - \mu_1(H - 1 + \varepsilon) - \mu_2\varepsilon, \tag{19}$$

where μ_1 and μ_2 are Lagrangian multipliers. And then, we compute the derivatives with respect to D and ε and make the derivatives equal to zeros. Thus, we can obtain:

$$\begin{cases} \dfrac{\partial L}{\partial D} = D - \mu_1 S = 0 \\ \dfrac{\partial L}{\partial \varepsilon} = \beta - \mu_1 - \mu_2 = 0 \end{cases}, \tag{20}$$

where

$$\begin{aligned} S = &\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|P_i|}\sum_{t=1}^{|P_i|}(p_s^m)(p_t^n)^T \\ &- \frac{\lambda}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|Q_i|}\sum_{t=1}^{|Q_i|}(q_s^m)(q_t^n)^T, \end{aligned} \tag{21}$$

The same as the literature [42], we can obtain the corresponding Karush–Kuhn–Tucher (KKT) conditions of this optimization problem with inequality constraints as follows:

$$\begin{cases} \mu_1(H - 1 + \varepsilon) = 0 \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \end{cases}, \tag{22}$$

At the same time, according to Eq. 20, there are

$$D = \mu_1 S, \tag{23}$$

It can be easily proved that S is a symmetric matrix, as shown in Appendix. Thus, $D = \mu_1 S$ is also a symmetric matrix. Then, we can obtain:

$$H = tr(DS^T), \tag{24}$$

where $tr(*)$ represents the trace of a square matrix.

By substituting Eqs. 20, 23 and 24 into Eq. 19, this problem can be reformulated as follows:

$$L(D, \varepsilon, \mu_1, \mu_2) = -\frac{\mu_1^2}{2}tr(SS^T) + \mu_1, \tag{25}$$

Therefore, the solution of Eq. 18 is equivalent to solving the following optimization problem:

$$\max_{\mu_1} \mu_1 - \frac{tr(SS^T)}{2}\mu_1^2, \tag{26}$$

where $tr(SS^T)$ is positive, so Eq. 26 can be regarded as a convex quadratic programming problem. If $\beta \geq 1/S$, Eq. 26 can achieve the maximum value when $\mu_1 = 1/S$. Otherwise, the solution is $\mu_1 = \beta$

At last, we substitute μ_1 into Eq. 23, and we can obtain D. And then, eigendecompose D as follows:

$$D = O\Lambda O^T, \tag{27}$$

where Λ is the diagonal eigenvalue matrix of D, and O is an orthogonal matrix corresponding to the eigenvectors of matrix D.

According to Eq. 27, we can obtain the solution of projection matrix W by:

$$W = O\sqrt{\Lambda}, \tag{28}$$

With the learned projection matrix W, we can obtain the projected features of all samples. Suppose that $Y^{(m)}$ is the projected features of view m, where $m = 1, 2, \ldots, M$, and it can be computed by:

$$Y^{(m)} = W^T X^{(m)}, \tag{29}$$

Then, we fuse all of the projected features by the following strategy:

$$Y = \left[Y^{(1)T}, Y^{(2)T}, \ldots, Y^{(M)T}\right]^T, \tag{30}$$

Finally, we use the k-means algorithm to cluster Y. The detailed optimization procedure of our proposed MCNDCL approach is summarized in Algorithm 1.

Algorithm 1 The proposed MCNDCL approach

Input: Multi-view sample sets $X = \{X^{(1)}, X^{(2)}, \ldots, X^{(M)}\}$; parameters λ, β and k
Output: *Clustering* Results C_i
1: Compute consistent neighbor domain P_i and specific neighbor domain Q_i of each sample according to the Equations 7 and 8 ;
2: Compute consistent neighbor domain correlations R_p and specific neighbor domain correlations coR_q by Equations 9 and 10 ;
3: Comupte μ_1 according to the Equation 26 ;
4: Compute D according to the Equation 27 ;
5: Compute W according to the Equation 28 ;
6: Obtain the projected sample set Y according to the learned W;
7: Perform k-means algorithm by using projected sample set Y;

4 Experiments

4.1 Datasets

In this paper, we evaluate the performance of our proposed approach on four widely used multi-view clustering datasets, including UCI Digits, Caltech7, BBCSport and CCV. Table 1 shows the details of these datasets.

UCI Digits: This dataset consists of 10 classes handwritten digits with each class having 200 different digits, and there are 2000 data points totally [43]. In our evaluation experiments, we select six published feature sets: specifically, Pix features (the 240-D pixel averages with the 2×3 sliding windows), PRO features (216-D profile correlations), FOU features (76-D) Fourier coefficients of digit shapes, KAR features (64-D Karhunen–Loeve coefficients), ZER features (47-D Zernike moment features) and MOR features (6-D morphological features).

Caltech7 This dataset is a subset of Caltech-101 [44], and it comprises of 1474 images from 7 classes, i.e., Dolla-Bill, Garfield, Face, Motorbikes, Snoopy, Stop-Sign and Windsor-Chair. In our experiments, six pattern features are extracted from all the images: specifically, 1984-D HoG features, 928-D LBP features, 512 GIST features, 254-D CENTRIST features, 48-D Gabor features and 40-D Wavelet moments features.

BBCSport This dataset[1] contains 737 documents from the BBC Sport website corresponding to sports news articles in five topical areas, including athletics, cricket, football, rugby and tennis, from 2004-2005, and among 544 samples, it appears in two views. Thus, we select these 544 samples to evaluate our approach. In particular, the feature dimension of the first view is 3203, and the second is 3183.

CCV This video dataset collects 9317 YouTube videos from 20 different semantic classes. The same with the literature [45], we choose 6773 samples of this dataset, along with three view features. To be specific, STIP features, SIFT features and MFCC features and the feature dimensions of these three views are 5000, 5000 and 4000.

4.2 Compared methods

In this paper, we compare our proposed approach with some typical clustering methods.

(1) BestView: In fact, this is a single-view-based method. Specifically, we use spectral clustering proposed in the literature [10] to perform clustering on every single-view data and report the best single-view results.
(2) Multi-view spectral clustering or graph clustering-based methods: CRSC [22], RMSC [23], MVSC [24], S-MVSC [46] and GFSC [47].
(3) Multi-view subspace clustering or kernel clustering-based methods: MultiNMF [18], SNF [19], MKKM [20] and LT-MSC [21].

4.3 Evaluation criteria

To comprehensively measure the performance of all competing methods, we adopt six used metrics which are widely used in many works [48–50], i.e., clustering accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI), Precision, Recall and F1-Score.

Let $p = [p_1, \ldots, p_N]$ be the clustering label by using proposed approach and $r = [r_1, \ldots, r_N]$ be the true clustering label, where N is the number of samples. Then, ACC can be computed by:

$$ACC = \frac{\sum_{i=1}^{N} \delta(r_i, map(p_i))}{N}, \tag{31}$$

where δ is the Dirac delta function, and *map* denotes the optimal mapping function which permutes the predicted labels to match the true labels. ACC is the most popular

[1] http://mlg.ucd.ie/datasets.

Table 1 Data descriptions used in the experiment

Datasets	# of samples	# of views	# of classes
UCI Digits	2000	6	10
Caltech7	1474	6	7
BBCSport	544	2	5
CCV	9317	3	20

clustering evaluation criteria, which represents the overall clustering accuracy.

Assume that the clustering results $C = \{C_i\}_{i=1}^{K}$ obtained by proposed approach, and the truth clustering results $C' = \{C'_i\}_{i=1}^{K}$, where K is the number of clusters. Then, NMI is defined as follows:

$$NMI(C, C') = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K}|C_i \cap C'_j| \log\frac{N|C_i \cap C'_j|}{|C_i||C'_j|}|}{\sqrt{\left(\sum_{i=1}^{K}|C_i| \log\frac{|C_i|}{N}\right)\left(\sum_{j=1}^{K}|C'_i| \log\frac{|C'_i|}{N}\right)}}, \tag{32}$$

NMI is also a general clustering evaluation criteria, which reflects the independence between clusters.

Let C_i^r denote the ith cluster obtained by the proposed approach, and it is re-permuted by the optimal mapping function. n_i^r and n_i^p represent the item number of C_i^r and C'_i. Thus, ARI can be measured by:

$$ARI = \frac{\sum_{i=1,j=1}^{K} M_{n_{ij}}^2 - \frac{\sum_{i=1}^{K} M_{n_i^p}^2 \sum_{i=1}^{K} M_{n_i^r}^2}{M_n^2}}{\frac{1}{2}\left(\sum_{i=1}^{K} M_{n_i^p}^2 + \sum_{i=1}^{K} M_{n_i^r}^2\right) - \frac{\sum_{i=1}^{K} M_{n_i^p}^2 \sum_{i=1}^{K} M_{n_i^r}^2}{M_n^2}}, \tag{33}$$

where M is the combination operation, and $n_{ij} = \left|C_i^r \cap C'_j\right|$. ARI is used to evaluate the consistency between clustering results and real results.

In many cases, the number of samples in each cluster is different. To evaluate the performance of clustering methods in these datasets, Precision, Recall and F1-Score are used. And they are given by:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1\text{-}Score &= \frac{2 \times Precision \cdot Recall}{Precision + Recall}, \end{aligned} \tag{34}$$

4.4 Experiments results and analysis

In this subsection, we will fully evaluate the performance of our approach. Experiments were done on a workstation with 2 NVIDIA GTX1080 GPUs and 8 i7-5960X @ 3.0GHz CPUs, and development environment is MATLAB R2016-b. In our approach, there are three parameters, i.e., k, λ and β. k and β are separately set to 5 and 1 in the all datasets. λ is set to 0.2, 0.4, 0.5, 0.4 in the UCI, Caltech7, BBCSport and CCV, respectively.

4.4.1 Evaluation on the UCI digits dataset

Table 2 shows the results of all competing methods on the UCI Digits Dataset. From the results, we can see that our MCNDCL can achieve the best performance on most of the metrics. Specifically, compared with the second best result, our approach, respectively, improves the performance by 2.12% (93.58–91.46%), 2.08% (89.32–86.34%), 1.59% (86.83–85.24%), 1.60% (88.34–86.74%) and 1.79% (87.58–85.79%) in five metrics, i.e., ACC, ARI, Precision, Recall and F1-Score, and 1.25% (88.90–87.65%) lower than the SNF in NMI metric.

Different from the BestView method which is a single-view-based clustering method actually, our MCNDCL is a multi-view-based clustering method, so our approach can fully use the information of multi-view data. Compared with spectral clustering or graph clustering-based methods, our approach can also obtain robust performance. The main reason is: Spectral clustering or graph clustering-based methods mainly focus on maximizing consistency information of multi-view data, while our approach simultaneously considers the correlations of consistent neighbor and the correlations of specific neighbor. Compared with subspace clustering or kernel clustering-based methods, our approach achieves better results than them. Of course, our approach can be also categorized to the subspace clustering-based methods. However, the major difference of our approach and these methods is: Most of these methods translate multi-view features into a common subspace by using linear transformation, matrix decomposition or tensor-based strategy, while our approach mainly focuses on the correlations of multi-view features.

4.4.2 Evaluation on the Caltech7 dataset

Table 3 shows the evaluation results of all competing methods on the Caltech7 dataset. Specifically, compared with the second best result, our approach, respectively, improves the performance by 3.21% (59.86–56.65%), 1.06% (88.72–87.66%), 2.97% (44.57–41.60%) and 3.14% (59.33–56.19%) in four metrics, such as ACC, Precision, Recall and F1-Score, and only a little lower 0.71% (59.14–58.43%) than LT-MSC in NMI metric, and 2.03% (50.56–48.53%) than GFSC in ARI metric. These indicate that our approach can obtain more competitive results than the comparison methods.

Table 2 The clustering performance of MCNDCL and compared methods on UCI Digits dataset

Methods	ACC	NMI	ARI	Precision	Recall	F1-Score
BestView	69.56 ± 4.50	64.24 ± 1.81	54.51 ± 3.00	58.13 ± 2.68	60.14 ± 2.74	59.11 ± 2.70
CRSC	91.46 ± 0.04	83.99 ± 0.05	82.01 ± 0.07	83.40 ± 0.06	84.23 ± 0.07	83.81 ± 0.07
RMSC	86.32 ± 0.03	78.03 ± 0.03	73.24 ± 0.06	75.37 ± 0.05	76.48 ± 0.05	75.92 ± 0.05
MVSC	81.80 ± 3.77	85.90 ± 1.62	76.11 ± 2.87	76.71 ± 3.94	80.57 ± 1.68	78.55 ± 2.54
S-MVSC	84.31 ± 0.36	82.53 ± 0.47	84.18 ± 0.56	83.96 ± 0.65	83.74 ± 0.29	83.87 ± 0.32
GFSC	89.45 ± 5.10	85.37 ± 1.96	86.34 ± 2.35	85.24 ± 2.66	86.69 ± 2.43	85.54 ± 1.57
SNF	88.35 ± 0.00	88.90 ± 0.00	84.21 ± 0.00	84.87 ± 0.00	86.74 ± 0.00	85.79 ± 0.00
MKKM	89.45 ± 0.00	81.74 ± 0.00	78.80 ± 0.00	80.64 ± 0.00	81.19 ± 0.00	80.92 ± 0.00
MultiNMF	77.60 ± 0.00	70.41 ± 0.00	60.31 ± 0.00	63.61 ± 0.00	65.03 ± 0.00	64.31 ± 0.00
LT-MSC	84.22 ± 0.00	82.17 ± 0.09	75.84 ± 0.11	77.07 ± 0.10	79.53 ± 0.11	78.28 ± 0.10
MCNDCL	93.58 ± 0.00	87.65 ± 0.00	89.32 ± 0.00	86.83 ± 0.00	88.34 ± 0.00	87.58 ± 0.00

Table 3 The clustering performance of MCNDCL and compared methods on Caltech7 dataset

Methods	ACC	NMI	ARI	Precision	Recall	F1-Score
BestView	41.00 ± 0.04	41.19 ± 3.87	25.82 ± 3.83	73.53 ± 4.06	29.58 ± 2.69	42.18 ± 3.41
CRSC	44.69 ± 0.30	37.22 ± 0.24	29.48 ± 0.29	78.08 ± 0.32	31.83 ± 0.20	45.22 ± 0.25
RMSC	46.06 ± 0.02	39.46 ± 0.02	33.27 ± 0.01	82.60 ± 0.03	34.14 ± 0.01	48.32 ± 0.01
MVSC	53.27 ± 6.63	52.91 ± 2.90	36.98 ± 4.87	78.49 ± 3.38	40.61 ± 4.27	53.46 ± 4.35
S-MVSC	56.28 ± 0.61	54.53 ± 1.02	50.56 ± 0.57	84.51 ± 0.87	39.53 ± 0.39	53.24 ± 0.72
GFSC	54.13 ± 2.13	55.61 ± 3.11	47.82 ± 2.64	84.25 ± 3.25	40.76 ± 1.95	54.29 ± 2.73
SNF	55.62 ± 0.81	49.08 ± 0.36	40.24 ± 1.37	83.86 ± 4.41	41.60 ± 0.89	55.55 ± 0.24
MKKM	42.06 ± 0.00	39.47 ± 0.00	31.58 ± 0.00	81.55 ± 0.00	32.69 ± 0.00	46.67 ± 0.00
MultiNMF	36.02 ± 0.00	31.56 ± 0.00	19.65 ± 0.00	64.86 ± 0.00	26.47 ± 0.00	37.60 ± 0.00
LT-MSC	56.65 ± 0.01	59.14 ± 0.73	41.82 ± 0.42	87.66 ± 0.32	41.35 ± 0.34	56.19 ± 0.37
MCNDCL	59.86 ± 0.00	58.43 ± 0.00	48.53 ± 0.00	88.72 ± 0.00	44.57 ± 0.00	59.33 ± 0.00

4.4.3 Evaluation on the BBCSport dataset

Table 4 reports the experimental results of our MCNDCL and other compared methods on the BBCSport dataset. From the results, it is obvious that our MCNDCL approach can achieve better results than all compared method in all six metrics. Specifically, compared with the second best result, our approach separately improves the performance by 6.18% (96.51–90.33%), 3.46% (89.18–85.72%), 7.05% (96.39–89.34 %), 5.31% (91.89–86.58%), 2.84% (93.07–90.23%) and 4.85% (92.48–87.63%) in ACC, NMI, ARI, Precision, Recall and F1-Score. These results further verify the effectiveness of the proposed approach.

4.4.4 Evaluation on the CCV dataset

To fully verify superiority of our approach, we have tested it on this larger dataset. The experimental results of our MCNDCL and compared methods on the CCV dataset can be seen in Table 5. From the results, we can see that all methods including our approach are failure. The main reasons should be twofold: (i) The quality of extracted features is not high on all views, which can be proven by the results of BestView method. (ii) This dataset is larger with more number of views (i.e., 6), samples (i.e., 9317) and clusters (i.e., 20). These challenges increase the difficulty of the clustering.

4.5 Further discussion

4.5.1 Robustness evaluation

To evaluate the robustness of our proposed approach, we have investigated the impact of noise at different attributes. Referring to the literature [51], we have noised the BBCSport dataset by adding 2 features, 4 features and 8 features composed of uniformly random values, respectively, and record as BBCSport+2NF, BBCSport+4NF and BBCSport+8NF. The results of the comparative experiment can be seen in Table 6. From the results, we can see that the ACC only reduces 0.76% (93.58–92.82%)

Table 4 The clustering performance of MCNDCL and compared methods on BBCSport dataset

Methods	ACC	NMI	ARI	Precision	Recall	F1-Score
BestView	43.00 ± 0.04	17.97 ± 1.26	9.73 ± 1.88	28.58 ± 1.08	65.49 ± 5.79	39.68 ± 0.17
CRSC	51.40 ± 3.35	32.83 ± 6.17	20.63 ± 4.89	35.78 ± 3.07	62.76 ± 2.22	44.10 ± 2.43
RMSC	82.15 ± 6.34	80.13 ± 2.48	77.41 ± 5.87	85.19 ± 1.74	80.32 ± 7.39	82.59 ± 4.68
MVSC	83.27 ± 1.66	79.28 ± 2.53	80.13 ± 1.72	83.24 ± 2.34	83.55 ± 2.68	83.39 ± 2.57
S-MVSC	85.29 ± 0.58	83.36 ± 0.49	85.24 ± 0.71	84.46 ± 0.70	86.28 ± 0.64	85.47 ± 0.82
GFSC	84.31 ± 2.16	81.25 ± 1.78	83.59 ± 1.63	81.34 ± 3.14	82.55 ± 1.83	81.64 ± 2.37
SNF	85.89 ± 0.85	82.46 ± 1.03	86.82 ± 0.56	86.58 ± 0.82	88.71 ± 0.64	87.63 ± 0.63
MKKM	90.33 ± 0.00	85.72 ± 0.00	89.34 ± 0.00	84.32 ± 0.00	90.23 ± 0.00	87.17 ± 0.00
MultiNMF	44.67 ± 0.00	30.17 ± 0.00	14.71 ± 0.00	32.46 ± 0.00	50.16 ± 0.00	39.41 ± 0.00
LT-MSC	71.69 ± 0.00	55.65 ± 0.00	49.58 ± 0.00	55.24 ± 0.00	74.33 ± 0.00	63.38 ± 0.00
MCNDCL	96.51 ± 0.00	89.18 ± 0.00	96.39 ± 0.00	91.89 ± 0.00	93.07 ± 0.00	92.48 ± 0.00

Table 5 The clustering performance of MCNDCL and compared methods on CCV dataset

Methods	ACC	NMI	ARI	Precision	Recall	F1-Score
BestView	19.43 ± 0.15	17.32 ± 0.20	20.18 ± 0.14	21.34 ± 0.08	22.41 ± 0.32	21.64 ± 0.19
CRSC	18.86 ± 1.37	18.14 ± 0.82	20.65 ± 1.12	20.87 ± 0.65	21.97 ± 2.01	21.38 ± 1.65
RMSC	21.40 ± 1.26	18.85 ± 1.31	21.36 ± 0.67	22.25 ± 1.20	24.63 ± 0.76	23.32 ± 1.45
MVSC	19.35 ± 1.66	19.20 ± 1.12	20.58 ± 1.23	21.83 ± 0.97	23.45 ± 0.58	22.80 ± 1.06
S-MVSC	22.36 ± 0.21	21.03 ± 0.09	24.81 ± 0.11	24.55 ± 0.08	26.79 ± 0.12	25.83 ± 0.14
GFSC	23.17 ± 2.58	19.57 ± 1.62	23.06 ± 2.26	25.31 ± 1.83	27.88 ± 1.23	26.09 ± 2.54
SNF	22.86 ± 0.85	18.76 ± 0.63	23.70 ± 0.34	26.34 ± 0.26	26.77 ± 0.51	26.55 ± 0.47
MKKM	21.74 ± 0.00	17.88 ± 0.00	24.12 ± 0.00	23.34 ± 0.00	25.64 ± 0.00	24.44 ± 0.00
MultiNMF	17.82 ± 0.00	17.45 ± 0.00	18.56 ± 0.00	21.57 ± 0.00	22.52 ± 0.00	22.03 ± 0.00
LT-MSC	19.98 ± 0.00	18.24 ± 0.00	21.47 ± 0.00	22.78 ± 0.00	23.46 ± 0.00	23.11 ± 0.00
MCNDCL	24.22 ± 0.00	19.33 ± 0.00	26.16 ± 0.00	25.83 ± 0.00	27.56 ± 0.00	26.67 ± 0.00

when adding 2 noise features, so our approach has good ability to resist small-scale noise.

Of course, with the increase in noise scale, the performance of our approach decreases significantly. Specifically, with adding 8 noise features, our approach separately decreases the performance by 8.35% (93.58–85.23%), 11.16% (87.65–76.49%), 11.4% (89.32–77.92%), 9.45% (86.83–77.38%), 10.1% (88.34–78.24%) and 10.25% (87.58–77.33%) in ACC, NMI, ARI, Precision, Recall and F1-Score.

4.5.2 Parameter evaluation

In this subsection, we will evaluate the impact of important parameters for experimental performance. In our approach, there are three parameters in total, i.e., k, λ and β. However, we found that the performance is not sensitive to the parameter β. Thus, we mainly discuss the effect of parameters k and λ, where k is used to compute the neighbor domain, and λ is a positive tunable factor in objective function. Figure 4 displays the evaluation results of k ranging from 3 to 8. From the results, we can observe that our approach can achieve the best performance when $k = 5$. It should be noted that F1-Score reaches the highest value when $k = 6$ not $k = 4$, but ACC and NMI both reach the highest value when $k = 5$. Figure 5 shows the evaluation results of λ in the range of [0.1,1]. The results indicate that the proposed approach can obtain stable performance, which further enhance the effectiveness of our MCNDCL. Similar phenomenon can be observed in other datasets.

Table 6 The clustering performance of MCNDCL with different noise attributes

Datasets	ACC	NMI	ARI	Precision	Recall	F1-Score
BBCSport	93.58	87.65	89.32	86.83	88.34	87.58
BBCSport+2NF	92.82	84.23	87.18	83.95	86.17	84.46
BBCSport+4NF	90.46	81.37	83.73	80.56	82.66	81.24
BBCSport+8NF	85.23	76.49	77.92	77.38	78.24	77.33

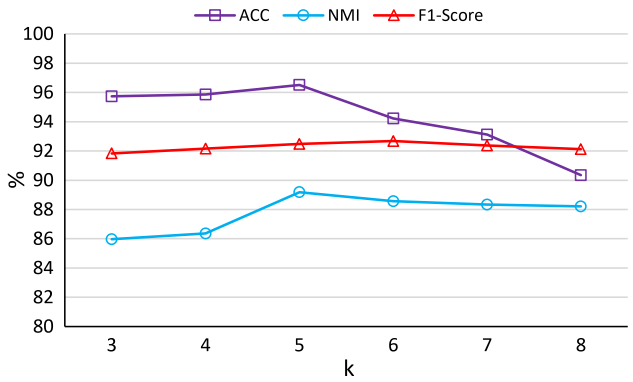

Fig. 4 Clustering performance of MCNDCL versus different values of parameter k

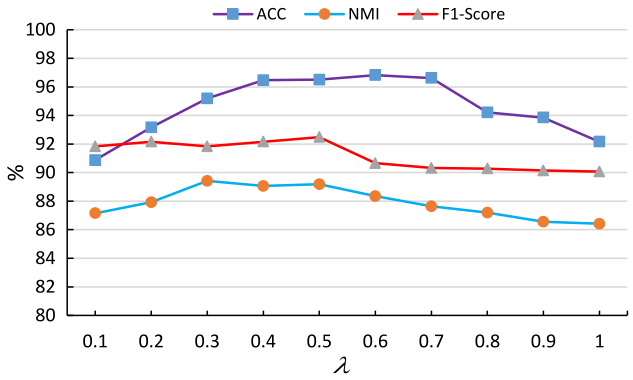

Fig. 5 Clustering performance of MCNDCL versus different values of parameter λ

Table 7 The time cost (s:seconds) of our MCNDCL and four relevant methods

Datasets	BestView	S-MVSC	MKKM	LT-MSC	MCNDCL
UCI Digits	1.4	5.7	20.8	9.2	13.6
Caltech7	0.5	3.4	12.4	7.3	7.8
BBCSport	0.3	1.2	4.7	3.5	2.8
CCV	2.3	7.6	80.9	39.1	50.6

4.5.3 Computation time

In this subsection, we have compared the time performance of our approach with some methods. Experiments were done on a workstation with 2 NVIDIA GTX1080 GPUs and 8 i7-5960X @ 3.0GHz CPUs. Table 7 reports the detailed training time of our MCNDCL approach and four relevant methods on all datasets. It should be noted that: For BestView, the reported time cost is only for one view. From the results, we can see that the time cost of our approach is comparable to other methods.

5 Conclusion and future work

In this paper, we propose a novel multi-view clustering approach named MCNDCL. MCNDCL can maximize the correlations of consistent neighbor domain and minimize the correlations of specific neighbor domain at the same time and translates the multi-view features into a common subspace. Extensive experimental results and verification analysis on four public datasets show that the proposed MCNDCL approach can achieve better performance than the existing mainstream methods.

With the development of deep learning technique, lots of works have combined with it and achieved satisfied results in many fields. Therefore, we will incorporate deep learning technique into multi-view clustering to further improve the performance of our approach in the future.

Acknowledgements The authors would like to thank the editor, the associate editor and anonymous reviewers for their constructive comments in helping improve our work. This work is supported by the National Natural Science Foundation of China No. 61902135, the Innovation Group Project of the National Natural Science Foundation of China No. 61821003, and the Scientific Research Projects of Hunan Education Department under Grant No. 14C0304.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

Proof

$$
\begin{aligned}
S^T = &\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|P_i|}\sum_{t=1}^{|P_i|}(p_s^m)(p_t^n)^T\right)^T \\
&-\left(\frac{\lambda}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{s=1}^{|Q_i|}\sum_{t=1}^{|Q_i|}(q_s^m)(q_t^n)^T\right)^T \\
=&\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i|}\sum_{n=1}^{M}\sum_{m=1}^{M}\sum_{t=1}^{|P_i|}\sum_{s=1}^{|P_i|}(p_t^n)(p_s^m)^T \\
&-\frac{\lambda}{N}\sum_{i=1}^{N}\frac{1}{|Q_i|}\sum_{n=1}^{M}\sum_{m=1}^{M}\sum_{t=1}^{|Q_i|}\sum_{s=1}^{|Q_i|}(q_t^n)(q_s^m)^T \\
=&\ S
\end{aligned}
,
$$

□

References

1. Sato Y, Izui K, Yamada T, Nishiwaki S (2019) Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. Expert Syst Appl 119:247–261
2. Li X, Wang Y, Song J, Liu Y, Zhang X, Zhou K, Li C (2020) A low cost and un-cancelled laplace noise based differential privacy algorithm for spatial decompositions. World Wide Web 23(1):549–572
3. Zhang J, Liu Y, Zhou K, Li G, Xiao Z, Cheng B, Xing J, Wang Y, Cheng T, Liu L, Ran M, Li Z (2019) An end-to-end automatic cloud database tuning system using deep reinforcement learning. In: Boncz PA, Manegold S, Ailamaki A, Deshpande A, Kraska T (eds) Proceedings of the international conference on management of data, SIGMOD, Amsterdam, The Netherlands, 2019, ACM, 2019, pp 415–432
4. Liu Y, Wang Y, Zhou K, Yang Y, Liu Y (2020) Semantic-aware data quality assessment for image big data. Future Gener Comput Syst 102:53–65
5. Tan TY, Zhang L, Lim CP (2020) Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks. Knowl-Based Syst 187:1–26
6. Hu S, Yan X, Ye Y (2020) Dynamic auto-weighted multi-view co-clustering. Pattern Recogn 99:1–12
7. Zhou K, Sun S, Wang H, Huang P, He X, Lan R, Li W, Liu W, Yang T (2019) Improving cache performance for large-scale photo stores via heuristic prefetching scheme. IEEE Trans Parallel Distrib Syst 30(9):2033–2045
8. Netto SMB, Diniz JOB, Silva AC, de Paiva AC, Nunes RA, Gattass M (2019) Modified quality threshold clustering for temporal analysis and classification of lung lesions. IEEE Trans Image Process 28(4):1813–1823
9. Wang Q, Yin H, Wang W, Huang Z, Guo G, Nguyen QVH (2019) Multi-hop path queries over knowledge graphs with neural memory networks. In: DASFAA, pp 777–794
10. Ng AY, Jordan MI, Weiss Y (2001) On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS, December 3-8, 2001, Vancouver, British Columbia, Canada], 2001, pp 849–856
11. Li X, Yin H, Zhou K, Zhou X (2019) Semi-supervised clustering with deep metric learning and graph embedding. World Wide Web, pp 1–18
12. Ahmad A, Khan SS (2019) Survey of state-of-the-art mixed data clustering algorithms. IEEE Access 7:31883–31902
13. Liu Q, Zhang R, Hu R, Wang G, Wang Z, Zhao Z (2019) An improved path-based clustering algorithm. Knowl-Based Syst 163:69–81
14. He Z, Ho C (2019) An improved clustering algorithm based on finite gaussian mixture model. Multimedia Tools Appl 78(17):24285–24299
15. Li X, Yin H, Zhou K, Zhou X (2020) Semi-supervised clustering with deep metric learning and graph embedding. World Wide Web 23(2):781–798
16. Wang Y, Wu L, Lin X, Gao J (2018) Multiview spectral clustering via structured low-rank matrix factorization. IEEE Trans Neural Netw Learn Syst 29(10):4833–4843
17. Yin M, Gao J, Xie S, Guo Y (2019) Multiview subspace clustering via tensorial t-product representation. IEEE Trans Neural Netw Learn Syst 30(3):851–864
18. Gao J, Han J, Liu J, Wang C (2013) Multi-view clustering via joint nonnegative matrix factorization. In: Proceedings of the 13th siam international conference on data mining, 2013. Austin, Texas, USA, pp 252–260
19. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A (2014) Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 11(3):333
20. Liu X, Dou Y, Yin J, Wang L, Zhu E (2016) Multiple kernel *k*-means clustering with matrix-induced regularization. In: Proceedings of the thirtieth AAAI conference on artificial intelligence, 2016, Phoenix, Arizona, USA, pp 1888–1894
21. Zhang C, Fu H, Liu S, Liu G, Cao X (2015) Low-rank tensor constrained multiview subspace clustering. In: 2015 IEEE international conference on computer vision, ICCV, Santiago, Chile, 2015, pp 1582–1590
22. Kumar A, Rai P, III HD (2011) Co-regularized multi-view spectral clustering. In: Advances in neural information processing systems, 25th annual conference on neural information processing systems. Proceedings of a meeting held 12-14 December 2011, Granada, Spain, vol 24, pp 1413–1421
23. Xia R, Pan Y, Du L, Yin J (2014) Robust multi-view spectral clustering via low-rank and sparse decomposition. In: Proceedings of the twenty-eighth AAAI conference on artificial intelligence, 2014, Québec City, Québec, Canada, pp 2149–2155
24. Li Y, Nie F, Huang H, Huang J (2015) Large-scale multi-view spectral clustering via bipartite graph, In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence, 2015, Austin, Texas, USA, pp 2750–2756
25. von Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416
26. Kumar A, III HD (2011) A co-training approach for multi-view spectral clustering. In: Proceedings of the 28th international conference on machine learning, ICML, Bellevue, Washington, USA, 2011, pp 393–400
27. Zhou D, Burges CJC (2007) Spectral clustering and transductive learning with multiple views. In: Machine learning, proceedings of the twenty-fourth international conference (ICML), Corvallis, Oregon, USA, 2007, pp 1159–1166
28. Long B, Yu PS, Zhang ZM (2008) A general model for multiple view unsupervised learning. In: Proceedings of the SIAM international conference on data mining, SDM, 2008, Atlanta, Georgia, USA, pp 822–833
29. Tsivtsivadze E, Borgdorff H, van de Wijgert J, Schuren FHJ, Verhelst R, Heskes T (2013) Neighborhood co-regularized multi-view spectral clustering of microbiome data. In: Partially supervised learning–second IAPR international workshop, PSL, Nanjing, China, 2013 Revised Selected Papers, pp 80–90
30. Cai X, Nie F, Huang H, Kamangar F (2011) Heterogeneous image feature integration via multi-modal spectral clustering. In: The 24th IEEE conference on computer vision and pattern recognition, CVPR, Colorado Springs, CO, USA, 2011, pp 1977–1984
31. Elhamifar E, Vidal R (2013) Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans Pattern Anal Mach Intell 35(11):2765–2781
32. Agrawal R, Gehrke J, Gunopulos D, Raghavan P (1998) Automatic subspace clustering of high dimensional data for data mining applications. In: SIGMOD, Proceedings ACM SIGMOD international conference on management of data, 1998, Seattle, Washington, USA, pp 94–105
33. Li S, Jiang Y, Zhou Z (2014) Partial multi-view clustering. In: Proceedings of the twenty-eighth AAAI conference on artificial intelligence, 2014, Québec City, Québec, Canada, pp 1968–1974
34. Zhao H, Liu H, Fu Y (2016) Incomplete multi-modal visual data grouping. In: Proceedings of the twenty-fifth international joint conference on artificial intelligence, IJCAI, New York, NY, USA (2016), pp 2392–2398
35. Yin Q, Wu S, Wang L (2015) Incomplete multi-view clustering via subspace learning. In: Proceedings of the 24th ACM

international conference on information and knowledge management, CIKM, Melbourne, VIC, Australia, 2015, pp 383–392
36. Chao G, Sun S, Bi J A survey on multi-view clustering, CoRR abs/1712.06246
37. Lee DD, Seung HS (2000) Algorithms for non-negative matrix factorization. In: Advances in neural information processing systems 13, Papers from neural information processing systems (NIPS), Denver, CO, USA, pp 556–562
38. Lazar C, Doncescu A (2009) Non negative matrix factorization clustering capabilities; application on multivariate image segmentation. In: 2009 international conference on complex, intelligent and software intensive systems, CISIS, Fukuoka, Japan, 2009, pp 924–929
39. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788–791
40. Brunet J-P, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. Proc Nat Acad Sci 101(12):4164–4169
41. Akata Z, Thurau C, Bauckhage C (2011) Non-negative matrix factorization in multimodality data for segmentation and label prediction. In: Wendel A, Sternig S, Godec M (eds) 16th computer vision winter workshop. Mitterberg, Austria
42. Chen X, Chen S, Xue H (2011) Large correlation analysis. Appl Math Comput 217(22):9041–9052
43. Rai N, Negi S, Chaudhury S, Deshmukh O (2016) Partial multi-view clustering using graph regularized NMF. In: 23rd international conference on pattern recognition, ICPR, Cancún, Mexico, 2016, pp 2192–2197
44. Li F, Fergus R, Perona P (2007) Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Comput Vis Image Underst 106(1):59–70
45. Jiang Y, Ye G, Chang S, Ellis DPW, Loui AC (2011) Consumer video understanding: a benchmark database and an evaluation of human and machine performance. In: Proceedings of the 1st international conference on multimedia retrieval, ICMR, Trento, Italy, 2011, ACM, pp 1–8
46. Hu Z, Nie F, Chang W, Hao S, Wang R, Li X (2020) Multi-view spectral clustering via sparse graph learning. Neurocomputing 384:1–10
47. Kang Z, Shi G, Huang S, Chen W, Pu X, Zhou JT, Xu Z (2020) Multi-graph fusion for multi-view spectral clustering. Knowl Based Syst 189:1–9
48. Yuan T, Deng W, Hu J, An Z, Tang Y (2019) Unsupervised adaptive hashing based on feature clustering. Neurocomputing 323:373–382
49. Sui XL, Xu L, Qian X, Liu T (2018) Convex clustering with metric learning. Pattern Recogn 81:575–584
50. Yang B, Fu X, Sidiropoulos ND, Hong M (2017) Towards k-means-friendly spaces: simultaneous deep learning and clustering. In: Proceedings of the 34th international conference on machine learning, ICML, Sydney, NSW, Australia, 2017, pp 3861–3870
51. de Amorim RC, Hennig C (2015) Recovering the number of clusters in data sets with noise features using feature rescaling factors. Inf Sci 324:126–145