Cross-View Diversity Embedded Consensus Learning for Multi-View Clustering

Chong Peng$^1$, Kai Zhang$^1$, Yongyong Chen$^{2,*}$, Chenglizhao Chen$^{3,*}$, Qiang Cheng$^4$

$^1$College of Computer Science and Technology, Qingdao University, China
$^2$School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), China
$^3$College of Computer Science and Technology, China University of Petroleum (East China), China
$^4$Department of Computer Science, University of Kentucky, USA

pchong1991@163.com, zk127585@163.com, yongyongchen.cn@gmail.com, cclz123@163.com, qiang.cheng@uky.edu

Abstract

Multi-view clustering (MVC) has garnered significant attention in recent studies. In this paper, we propose a novel MVC method, named CCL-MVC. The novel method constructs a cross-order neighbor tensor of multi-view data to recover a low-rank essential tensor, which preserves noise-free, comprehensive, and complementary cross-order relationships among the samples. Furthermore, it constructs a consensus representation matrix by fusing the low-rank essential tensor with auto-adjusted cross-view diversity embedding, fully exploiting both consensus and discriminative information of the data. An effective optimization algorithm is developed, which is theoretically guaranteed to converge. Extensive experimental results confirm the effectiveness of the proposed method.

1 Introduction

As one of the fundamental tasks of machine learning, clustering has achieved notable successes in diverse domains, such as image segmentation [Ng et al., 2023], face clustering [Wu et al., 2023], community detection [Park et al., 2022], etc. Among the extensive clustering algorithms, the subspace clustering methods have drawn significant attention in recent years due to their effectiveness and elegant theory [Liu et al., 2010; Peng et al., 2017; Elhamifar and Vidal, 2009]

In real-world scenarios, the advancement of technology has led to an increase in the availability of data from multiple sources, leading to the natural emergence of multi-view data [Zhang et al., 2020]. Consequently, multi-view data are prevalent and contain a wealth of useful, discriminative, and complementary information across multiple perspectives, making them essential for enhancing clustering capability [Xia et al., 2021]. However, traditional clustering methods mainly focus on single-view data and are not suitable for analyzing multi-view data, as each view can provide a unique and valuable insight. Thus, multi-view clustering (MVC) has garnered significant attention in recent years due to its versatility and effectiveness in various domains [Xia et al., 2022; Wen et al., 2023; Brbić and Kopriva, 2018].

Due to the great success of single-view subspace clustering (SVSC) methods, they have been extensively developed for MVC [Fu et al., 2023; Zhang et al., 2020; Yang et al., 2019]. Most of these methods follow a pipeline similar to the SVSC with a two-step strategy. In particular, they first construct an affinity matrix, on which the standard spectral clustering (SPC) [Ng et al., 2001] is then performed. Since the SPC is standard, the step of affinity matrix construction has been the main focus of MVC [Chen et al., 2021; Zhang et al., 2020].

According to the way of affinity matrix construction, existing MVC methods can be mainly categorized into two types [Wu et al., 2019]. The first type is developed from the SPC-based subspace clustering methods [Zhang et al., 2015; Li et al., 2019; Zhang et al., 2020], while the second is developed from the graph-based clustering, which constructs affinity matrix using similarity matrix [Wang et al., 2020; Xia et al., 2014; Zhan et al., 2018; Kumar et al., 2011] and is the main focus of our paper. It has been revealed that the Markov random walk has a close connection with the SPC [Zhou et al., 2005; Shi and Malik, 2000], based on which a number of methods have been developed [Zhou and Burges, 2007; Xia et al., 2014; Wu et al., 2019]. For example, the transductive inference approach constructs a Markov transition probability matrix (TPM) in each view, and combines the TPMs using the Markov mixture [Zhou and Burges, 2007]. The RMSC constructs the TPMs of different views and learns a common low-rank stochastic matrix to alleviate the noise effects [Xia et al., 2014]. Different from them, the ETLMSC learns a low-rank essential tensor using the low-rank tensor recovery (LTR) approach from the TPMs [Wu et al., 2019].

The LTR approach has been quite popular and successful in recent development of MVC [Xie et al., 2018; Zhang et al., 2015; Wu et al., 2019], which has been extensively attempted in above mentioned methods [Zhou and Burges, 2007; Xia et al., 2014; Wu et al., 2019; Xie et al., 2018]. However, these methods suffer from some key issues, which may severely degrade the learning performance. First, by adopting the tensor nuclear norm (TNN), the LTR approach may suffer from inaccurate approximation issue [Yu and Yang, 2023]. Second, the high-order neighbor information of the data is rarely considered, which has been revealed essential in graph learning [Tang et al., 2015]. Third, simultaneous consensus and diversity learning is rarely considered on the low-rank essential tensor, which omits some essential properties. To this end, we develop a new method for MVC to address the above issues.

We summarize the key contributions of this paper as fol-
lows: 1) We recover a low-rank essential tensor from a cross-
order neighbor graph tensor derived from multi-view data,
which preserves comprehensive and complementary in-
formation of the data. 2) By embedding an automatically ad-
justed weighting vector with the learned noise-free cross-
order neighbor graphs, our method explicitly preserves cross-
view diversity information of the multi-view data to learn a
consensus affinity matrix. 3) We design an efficient opti-
mentation algorithm, which is theoretically guaranteed to converge
under some mild conditions. 4) Extensive experimental re-
sults confirm the effectiveness of the proposed method and
its superiority to baselines.

2 Related Work

Given multi-view data \( \{X^{(v)}\}_{v=1}^{V} \), with superscript \((v)\) den-
noting the \(v\)-th view of multi-view data or the \(v\)-th frontal
slice of a tensor, \( X^{(v)} \in \mathbb{R}^{d_v \times n} \) being the \(v\)-th view samples,
\( d_v \) and \( n \) being the number of features and samples in the
\(v\)-th view, and \( V \) being the number of views, respectively, the
ETLMSC recovers a low-rank essential tensor \( Z \in \mathbb{R}^{K \times n \times V} \)
from a TPM-based tensor \( P \in \mathbb{R}^{K \times n \times V} \) constructed from
multi-view data by [Wu et al., 2019]:

\[
\min_{Z, E \in \mathbb{R}^{K \times n \times V}} \|Z\|_\odot + \lambda \|E\|_2, \quad s.t. \quad P = Z + E, \tag{1}
\]

where \( E \) denotes the noise, \( \| \cdot \|_\odot \) is the t-SVD-based TNN,
and \( \| \cdot \|_2 \) is the tensor \( \ell_2 \)-norm defined as the sum of
\( \ell_2 \)-norm of mode-3 fibers. Here, the tensor \( P \) is constructed
by slice by slice as \( P^{(v)} = (D_v(S^{(v)}))^{-1}S^{(v)} \), where \( D_v(\cdot) \) returns a
diagonal matrix with diagonal elements being the sum of rows
of the input matrix, and \( S \in \mathbb{R}^{n \times n \times V} \) is a similarity
tensor constructed by slice with \( S^{(v)} \) being a pair-wise simi-
larity matrix in the \(v\)-th view. Eq. (1) essentially follows the robust
tensor PCA [Lu et al., 2016], where the key difference between them is that they aim at different learning tasks with
different types of data as input.

3 The Proposed Method

Due to the great success of low-rank essential recovery ap-
proach, in this paper, we follow the framework of Eq. (1)
and develop a new method with some more desired prop-
ties for enhanced learning performance. As has been widely
revealed in literature, the convex approach to rank or spar-
sity approximation is not accurate, and nonconvex approach
may provide a more accurate approximation and enhance the
structural learning capability of the model [Peng et al., 2022a;
Peng et al., 2022b]. Therefore, in a way similar to the
t-SVD-based TNN [Wu et al., 2019] and tensor Schatten
norm [Xia et al., 2022], and inspired by [Peng et al., 2022a;
Peng et al., 2020], we expand the log-based matrix rank and
sparsity approximation to log-based tensor rank approxima-
tion (LTRA) and sparsity approximation (LTSA) as \( \|Z\|_\odot = \sum_{v=1}^{V} \sum_{i=1}^{n} \log(1 + \sigma_i(Z^{(v)})) \), and \( \|E\|_\odot = \sum_{v=1}^{V} \log(1 + \sum_{i=1}^{n} E_{i,j,v}^2)^{1/2} \), respectively, where the LTRA is def-
ined in the frequency domain, \( \hat{Z} = \text{fft}(Z, 3) \in \mathbb{R}^{n \times n \times V} \)
with \( \text{fft}(\cdot) \) being the fast Fourier transform (FFT) along
the third dimension, \( \sigma_i(\cdot) \) is the \(i\)-th largest singular value
of

the input matrix, and the LTSA is defined in a lateral slice-
wise manner that enhances the cross-view sparsity and helps
strengthen the connections among different views. Then, we
may develop Eq. (1) into the following model:

\[
\min_{Z, E \in \mathbb{R}^{n \times n \times V}} \|Z\|_\odot + \lambda \|E\|_\odot \quad s.t. \quad P = Z + E, \tag{2}
\]

in which the nonconvex approach provides more accurate ap-
proximations and helps preserve more useful information to
enhance the learning capability [Peng et al., 2022a].

In Eq. (2), \( P \) reveals the probability of a one-step random
walk from one example to another in each view, which pro-
vides soft neighbor relationships of the samples. In practice,
such relationships may be insufficient to explicitly measure
the structure of the data, because samples may have latent
higher-order neighbor relationships that are not directly pre-
hased in a first-order graph [Kang et al., 2022]. Intuitively,
a higher-order relationship can be measured by the proba-
bility of a multi-step random walk from one example to an-
other [Tang et al., 2015]. Therefore, we may incorporate the
higher-order neighbor relationships into LTR by defining a
higher-order TPM for multi-step random walks in each view as:

\[
(P^{(v)})^k = P^{(v)} \times P^{(v)} \times \ldots \times P^{(v)},
\]

with \( k \) being the step of random walks or the order of neighbor relationships. To fully exploit the cross-order neighbor relationships, we define the following fine-grained probability tensor \( P_K \in \mathbb{R}^{n \times n \times V} \) by

\[
\text{slice as}: P_K^{(v)} = \sum_{k=1}^{K} (P^{(v)})^k + ((P^{(v)})^t)^2 ,
\]

where the transpose ensures the symmetry of the neighbor relationships that is natural and essentially desired for neighbor relationships. Moreover, to exploit local structure of the data and filter the redundancy of neighbor relationships, we construct a local similarity tensor \( S \), where in each view we keep \( N \) neighbors in the neighbor graph while setting the others to zero for each sample. Consequently, we construct \( P_K \) with local structure of the data and our model becomes

\[
\min_{Z, E} \|Z\|_\odot + \lambda \|E\|_\odot \quad s.t. \quad P_K = Z + E. \tag{3}
\]

In particular, Eq. (3) is treated as a \( K \)-th order model. In liter-
ature, an intuitive and common way of clustering is to fuse
\( Z \) across different views to obtain a representation matrix by

\[
Z_0 = \frac{1}{\alpha} \sum_{v=1}^{V} Z^{(v)},
\]

on which the final clustering algorithm is performed for the final clustering result [Wu et al., 2019]. However, the two-step strategy omits the close connection be-
 tween the learning and fusion tasks, and the straight fusion
omits the discriminative information embedded in different
views, which might be insufficient to fully exploit structural
information of the data [Pan et al., 2023].

To build a close connection between learning and fusion,
we embed the fusion task into our model, leading to:

\[
\min_{Z, \tilde{E}, w} \|Z\|_\odot + \lambda \|E\|_\odot + \alpha \sum_{v} \|Z_0 - w \cdot Z^{(v)}\|_F^2 \quad s.t. \quad P_K = Z + \tilde{E}, \quad w_v \geq 0, \quad \sum_{v} w_v = 1, \tag{4}
\]

where \( Z_0 \) is a consensus affinity matrix that fuses cross-view
neighbor graphs, and \( w \in \mathbb{R}^V \) is a weighting vector that is
automatically adjusted to balance the discriminative informa-
tion embedded across different views. To further enhance the
cross-view diversity embedded in the weighting vector \( w \), and
thus in the consensus affinity matrix $Z_0$, it is desired to prevent the weights from being simultaneously large if the TPMs of their corresponding views are highly similar [Kang et al., 2023], which leads to the diversity embedding defined as \( \min_{w \geq 0, \sum_i w_i = 1} \sum_{i,j} w_i w_j \text{Tr}((P_K^{(i)})^T P_K^{(j)}) \). Although the embedding is straightforward and easy to solve, $P_K$ may be prone to noise effects and thus the embedding might be inaccurate. As an alternative, we may embed the diversity of $w$ with $Z$, which efficiently alleviates the noise effects while is more challenging to solve. Therefore, to explore the cross-view noise-free diversity, we finally develop Eq. (4) as:

\[
\min_{P_K = \mathcal{Z} + \mathcal{E}, w \geq 0, \sum_i w_i = 1} \|Z\|_\diamond + \lambda \|\mathcal{E}\|_\diamond + \alpha \sum_i \|Z_i - w_i Z^{(v)}\|_F^2 + \beta \sum_{i,j=1}^V w_i w_j \text{Tr}(Z^{(i)T} Z^{(j)})
\]

(5)

where $\beta > 0$ is a balancing parameter. We name the above model the Cross-view diversity embedded Consensus Learning for Multi-View Clustering (CCL-MVC). We will develop an effective optimization algorithm in the next section.

4 Optimization

In this section, we develop an effective optimization algorithm using the augmented Lagrange multiplier method (ALM). We introduce an auxiliary variable $Q \in \mathbb{R}^{n \times V \times n}$ and obtain the following augmented Lagrange function:

\[
L(Z, \mathcal{E}, Q, Z_0, w, \mathcal{Y}_1, \mathcal{Y}_2, \rho) = \|Q\|_\diamond + \lambda \|\mathcal{E}\|_\diamond + \alpha \sum_i \|Z_i - w_i Z^{(v)}\|_F^2 + \beta \sum_{i,j=1}^V w_i w_j \text{Tr}(Z^{(i)T} Z^{(j)})
\]

(6)

\[
+ \frac{\rho}{2} \|Q - Z + \mathcal{Y}_1/\rho\|_F^2 + \frac{\rho}{2} \|P_K - Z - \mathcal{E} + \mathcal{Y}_2/\rho\|_F^2,
\]

where $\mathcal{Y}_1, \mathcal{Y}_2 \in \mathbb{R}^{n \times V \times n}$ are Lagrange multipliers and $\rho$ is the penalty parameter. Then, we alternately optimize each variable with details presented as follows.

4.1 Optimization of $Z$

The subproblem associated with $Z$ is to minimize

\[
\min_{Z} \alpha \sum_i \|Z_i - w_i Z^{(v)}\|_F^2 + \beta \sum_{i,j=1}^V w_i w_j \text{Tr}(Z^{(i)T} Z^{(j)})
\]

(7)

\[
+ \frac{\rho}{2} \|Q - Z + \mathcal{Y}_1/\rho\|_F^2 + \frac{\rho}{2} \|P_K - Z - \mathcal{E} + \mathcal{Y}_2/\rho\|_F^2.
\]

An intuitive way is to optimize each $Z^{(v)}$ while keeping the others fixed. However, this strategy may suffer from the following issues. First, all $Z^{(v)}$’s depend on each other and we cannot obtain the global solution for all $Z^{(v)}$’s simultaneously. Second, we may need to iteratively update all $Z^{(v)}$s within an inner loop until convergence to obtain the optimal $Z$, which lacks efficiency. To address these issues, we design an efficient optimization strategy as follows. First, we define the following augmented variables:

\[
\bar{Z} = \left[ \begin{array}{c} Z^{(1)} \\ \cdots \\ Z^{(V)} \end{array} \right], Z_0 = \left[ \begin{array}{c} Z_0^{(1)} \\ \cdots \\ Z_0^{(V)} \end{array} \right], \bar{I}_1 = \left[ \begin{array}{cccc} w_1 I & \cdots & w_V I \end{array} \right]^T, \bar{I}_2 = \left[ \begin{array}{cccc} w_1 I & \cdots & w_V I \end{array} \right]
\]

Then, with these notations, we derive an equinormal form of Eq. (7) to facilitate the optimization. For the first term of Eq. (7), we have $\alpha \sum_i \|Z_i - w_i Z^{(v)}\|_F^2 = \alpha \|Z_0 - \bar{I}_1 Z_0\|_F^2$.

For the second term, we have $\beta \sum_{i,j=1}^V w_i w_j \text{Tr}(Z^{(i)T} Z^{(j)}) = \beta \sum_{i,j} \text{Tr}(w_i Z^{(i)}(w_j Z^{(j)})) = \beta \sum_i w_i Z^{(i)}_0 = \beta \|\bar{I}_2 Z_0\|_F^2$.

For the last two terms, we may combine them and obtain $\frac{\rho}{2} \|Q - Z + \mathcal{Y}_1/\rho\|_F^2 + \frac{\rho}{2} \|P_K - Z - \mathcal{E} + \mathcal{Y}_2/\rho\|_F^2 = \rho \|\bar{Z} - \bar{I}_1 Z_0\|_F^2$, where $\bar{H} = (\bar{Q} + V^{(1)} + V^{(2)})/2 + (\bar{Y}_1 + V^{(1)} + V^{(2)})^T/2 + \cdots, (\bar{Q} + V^{(1)} + V^{(2)})^T/2 + (\bar{Y}_1 + V^{(1)} + V^{(2)})^T/2 \in \mathbb{R}^{nV \times n}$. Thus, we may convert Eq. (7) to the following quadratic problem:

\[
\min \alpha \|Z_0 - \bar{I}_1 Z_0\|_F^2 + \beta \|\bar{I}_2 Z_0\|_F^2 + \rho \|\bar{Z} - \bar{I}_1 Z_0\|_F^2.
\]

(8)

The first-order optimality condition of Eq. (10) is:

\[
2\alpha \bar{I}_1^T Z_2 + 2\beta \bar{I}_2^T Z_2 - 2\alpha \bar{I}_1^T Z_0 + 2\rho \bar{Z} - 2\rho H = 0,
\]

which leads to the following closed-form solution:

\[
\bar{Z} = (\alpha \bar{I}_1^T + \beta \bar{I}_2^T + \rho I)^{-1}(\alpha \bar{I}_1^T Z_0 + \rho \bar{H}).
\]

(10)

It is seen that the above solution involves inversion of an $nV \times nV$ matrix, which generally has $O(n^3V^2)$ complexity. Fortunately, the above solution admits a special structure, which can be further simplified by the Sherman-Morrison-Woodbury formula. First, we have

\[
(\alpha \bar{I}_1^T + \beta \bar{I}_2^T + \rho I)^{-1} = (\alpha \bar{I}_1^T + \rho I + \beta \bar{I}_2^T)^{-1}
\]

\[
= A^{-1} - \beta A^{-1} \bar{I}_2^T \left( I + \left( \sum_{i=1}^V \frac{w_i^2}{\rho + \alpha w_i^2} \right) I \right)^{-1} \bar{I}_2 A^{-1}
\]

\[
= A^{-1} - \beta A^{-1} \bar{I}_2^T \left( \frac{1}{\rho + \alpha w_1^2} \bar{I}_1 \cdots \frac{1}{\rho + \alpha w_V^2} \bar{I}_1 \right)
\]

\[
= A^{-1} - \beta \sum_{i=1}^V \frac{w_i}{\rho + \alpha w_i^2} A^{-1} \bar{I}_2 A^{-1}
\]

(11)

\[
= \left[ \begin{array}{ccc} \frac{1}{\xi_1} I & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\xi_V} I \end{array} \right] - \beta \sum_{i=1}^V \frac{w_i}{\rho + \alpha w_i^2} \xi_i \xi_i^T
\]

\[
= \left[ \begin{array}{ccc} \frac{1}{\xi_1} I & \cdots & \xi_1^T \xi_V I \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\xi_V} I \end{array} \right] - \beta \sum_{i=1}^V \frac{w_i}{\rho + \alpha w_i^2} \xi_i \xi_i^T
\]

with $\xi_i = \rho + \alpha w_i^2$ and $\xi_i^T = \frac{w_i w_j}{\rho + \alpha w_i^2}$ for $i, j = 1, \cdots, V$, and $\Lambda = \alpha \bar{I}_1^T + \rho I$. Define $B \in \mathbb{R}^{n \times n \times V}$ with $B^{(v)} = \alpha w_i Z_0 + \rho (Q^{(v)} + P^{(v)} - \mathcal{E}^{(v)}) + V^{(1)} + V^{(2)}$ for $v = 1, \cdots, V$; then, by plugging Eq. (11) and $B$ into Eq. (10), we have

\[
\bar{Z} = \left[ \begin{array}{ccc} \frac{1}{\rho + \alpha w_1^2} B^{(1)} & \cdots & \frac{1}{\rho + \alpha w_V^2} B^{(1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{\rho + \alpha w_1^2} B^{(V)} & \cdots & \frac{1}{\rho + \alpha w_V^2} B^{(V)} \end{array} \right],
\]

(12)

whose complexity surprisingly mainly comes from scatter-matrix product. Then, $\bar{Z}$ is obtained by tensoring $Z$. 

4790
4.2 Optimization of $Q$

The subproblem of $Q$ is $\min_{Q} \|Q\|_0 + \frac{\rho}{2} \|Q - Z + \gamma_1 / \rho\|^2_F$. This problem is not straightforward to solve since the LTRA is defined in the frequency domain. In a way similar to the shrinkage problems in [Gao et al., 2021; Pan et al., 2023; Wu et al., 2019], the above problem can be converted to the frequency domain as

$$
\min_{Q} \sum_{i=1}^{V} \left\{ \frac{1}{2} \| \hat{K}^{(v)} - \hat{Q}^{(v)} \|^2_F + \frac{1}{\rho} \sum_{j=1}^{n} \log(1 + \sigma_j(\hat{Q}^{(v)})) \right\},
$$

where $\hat{K} = \text{fft}(Z - \gamma_1 / \rho)[,.]$. Then, the above problem can be divided into $V$ independent subproblems of $\hat{Q}^{(v)}$ with $v = 1, \ldots, V$, which are the standard log-determinant regularized shrinkage problems described in [Peng et al., 2022a]. We denote $\cup(\cdot)$ and $\vee(\cdot)$ as operators that return the left and right singular vectors of the input, and $\mathbb{D}_c(\cdot)$ as an operator that returns a diagonal matrix based on the input elements, then according to [Peng et al., 2022a; Peng et al., 2015], $\hat{Q}^{(v)}$ is obtained by

$$
\hat{Q}^{(v)} = \cup(\hat{K}^{(v)}) \mathbb{D}_c(\hat{\sigma}^{(v)}_1, \cdots, \hat{\sigma}^{(v)}_n)(\vee(\hat{K}^{(v)})^T),
$$

with $\hat{\sigma}^{(v)}_i = \arg \min_{x \geq 0} \frac{1}{2} (x - \sigma_i^{(v)})^2 + \frac{1}{\rho} \log(1 + x)$ for $i = 1, \ldots, n$. With straightforward algebra, we may obtain $\hat{\sigma}^{(v)}_i = \frac{\sigma_i^{(v)} - H(1 + \sigma_i^{(v)})^2 - \frac{1}{2}}{2}$, with $\{\cdot\}$ being an indicator function that returns 1 if the conditions in subscript hold and 0 otherwise, and $\mathbb{S}_{i,v} = \{x | x \in \mathbb{R}, f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = \frac{1}{2} (x - \sigma_i^{(v)})^2 + \frac{1}{\rho} \log(1 + x), \xi = \frac{\sigma_i^{(v)} - H(1 + \sigma_i^{(v)})^2 - \frac{1}{2}}{2} > 0$.

4.3 Optimization of $E$

The LTRA-regularized shrinkage problem of $E$ is:

$$
\min_{E} \lambda \|E\|_0 + \frac{\rho}{2} \|P_K - Z - E + \gamma_2 / \rho\|^2_F,
$$

which is a LTRA-regularized shrinkage problem. It is noted that by reforming the lateral slices of the tensors in Eq. (15) into vectors, the problem is a direct shrinkage of the $\ell_2$ norm regularized shrinkage problem in [Peng et al., 2022a]. Then, by performing a vectorization-matrixization procedure to the $\ell_2$ norm shrinkage problem in [Peng et al., 2022a], we may obtain the above solution. Thus, it is straightforward that Eq. (15) admits the following closed-form solution by lateral slice as $E_{L,j} = \frac{\delta_j - \gamma_2 / \rho + \sqrt{(1 + \delta_j^2)^2 - \frac{4\lambda}{\rho}}}{2\delta_j} \cdot \mathbb{S}_{L,j} \cdot \mathbb{S}_{E_{L,j}}$, where $(\cdot)_{L,j}$ denotes the $j$-th lateral slice of the input tensor, $\mathbb{S}_{L,j}$ denotes the $j$-th lateral slice of the input tensor, $\mathbb{K} = P_K - Z + \gamma_2 / \rho$, $\delta_j = \|P_K - Z + \gamma_2 / \rho\|_F$, and $\mathbb{S}_{L,j} = \{x | x \in \mathbb{R}, f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = \frac{1}{2} (x - \delta_j)^2 + \frac{1}{\rho} \log(1 + x), \}$.

4.4 Optimization of $Z_0$

The subproblem of $Z_0$ is $\alpha \sum_i \|Z_i - w_i Z^{(v)}\|^2_F$, which is convex and quadratic, and admits a closed-form solution by the first-order optimality condition: $Z_0 = \frac{1}{\alpha} \sum_{v=1}^{W} w_v Z^{(v)}$. $\xi = \frac{\delta_j - \gamma_2 / \rho + \sqrt{(1 + \delta_j^2)^2 - \frac{4\lambda}{\rho}}}{2\delta_j} > 0$, $\delta_j = \|P_K - Z + \gamma_2 / \rho\|^2_F$, for ease of notation, we summarize the above procedure as

$$
E = \text{LTRA}_{\lambda / \rho}(P_K - Z + \gamma_2 / \rho).
$$

4.5 Optimization of $w$

The subproblem of $w$ is $\min_{w \geq 0, \sum w_v = 1} \alpha \sum_i \|Z_i - w_i Z^{(v)}\|^2_F + \beta \sum_{ij} w_i w_j \|T_r((Z^{(v)})^T Z^{(j)})\|$, Let $z_j$ be vectorized $Z^{(i)}$ for $i = 1, \ldots, V$, $V$, $M_1 \in \mathbb{R}^{W \times W}$ be a diagonal matrix with $M_{ii} = Z^{(i)}_i$, $M_2 \in \mathbb{R}^{W \times W}$ with $M_{ij} = Z^{(i)}_j$, and $h \in \mathbb{R}^W$ with $h_i = -\alpha \text{Tr}(Z_0^T Z^{(i)})$, then the subproblem of $w$ is equivalent to

$$
\min_{w} w^T (\alpha M_1 + M_2) w + 2 \|h\|^2_F, \text{ s.t. } w \geq 0, \sum w_i = 1.
$$

It is easy to verify that $\forall z \in \mathbb{R}^W$ and $z \neq 0$, we have $\tilde{Z}(\alpha M_1 + M_2 Z) = \alpha M_1 z + \beta \sum_{ij} z_i z_j z_i z_j = \alpha \sum_i \sum_j \frac{z_i z_j}{M_{ii} + \beta \sum_j \sum_i M_{ij} z_i z_j} = \alpha \sum_i \|z_i\|^2_2 + \beta \sum_j \|z_j\|^2_2$, and $z_i$ and $z_j$ being the vectorized $Z^{(i)}$ and $Z^{(j)}$, and $z_i$ and $z_j$ denoting the $i$-th and $j$-th elements of $z$, respectively. Thus, $\alpha M_1 + M_2 Z \geq 0, \text{ and Eq. (17) is convex. Then, w can be efficiently solved by standard convex programming technique, which is denoted as}$

$$
w = \text{quadprog}(\alpha M_1 + M_2, h).
$$

4.6 Updating of $\gamma_1$, $\gamma_2$, and $\rho$

The Lagrange multipliers and the penalty parameter are updated in a standard way as follows:

$$
\gamma_1 = \gamma_1 + \rho (Q - Z), \gamma_2 = \gamma_2 + \rho (P_K - Z - E), \rho = \rho \kappa,
$$

with $\kappa > 1$ being a parameter that keeps $\rho$ increasing. In our paper, we update the variables in the order of $Z, \mathcal{E}, Q, Z_0, w, Y_1, Y_2$, and $\rho$, which is essentially important for the convergence analysis in Section 5. After we obtain the solution, the standard SPC is applied to $Z_0$ for the final clustering result.

5 Convergence Analysis

In this section, we analyze the convergent property of the CCL-MVC. The main results are presented in the following.

**Theorem 1.** Let $t$ in the superscript denote the iteration number. Under assumptions that $\sum \frac{1}{\rho^t} < \infty$ and $\sum \frac{t+1}{(\rho^t)^2} < \infty$, and given a bounded initialization of the variables, the variable sequences $\{Z\}, \{E\}, \{Q\}, \{Z_0\}, \{w\}, \{Y_1\}, \{Y_2\}$ generated by our optimization algorithm are bounded.

**Proof:** According to the constraints of $w$, it is easy to verify that $\{w^t\}$ is bounded. At the $(t + 1)$-th iteration, according to the first-order optimality condition of $Q$, we have

$$
\partial Q L(Z^{t+1}, E^{t+1}, Q, Z_0^t, w^t, Y_1^t, Y_2^t, \rho^t) |_{Q^{t+1}}
$$
By definition of the LTSA, it is clear that \(\rho^t(Q^{t+1} - Z^{t+1}) + Y^{t+1}_1\) is bounded. Because \(Q^{t+1}, Y^{t+1}_1\) and \(\rho^{t+1}\) are all bounded, it is clear that \(Z^{t+1}\) is also bounded. Then, \(Z_0^{t+1} = \frac{1}{t} \sum_{v=1}^{V} w^v_1(Z^{t+1})^{(v)}\) is also bounded. Now, we may conclude that the sequences of \(\{Z^{t}\}\), \(\{E^t\}\), \(\{Z_0^t\}\), \(\{w^t\}\), \(\{Y^{t}_1\}\) and \(\{Y^{t}_2\}\) are all bounded.

\[\mathrm{Theorem 2.}\] Let \(\{Z^{t}, E^t, Q^t, Z_0^t, w^t, Y^{t}_1, Y^{t}_2\}\) be a sequence generated by our algorithm. Under assumptions that \(\sum \frac{\rho_{t+1}}{\rho_t} < \infty\) and \(\rho^t(Q^{t+1} - Q^t) \to 0\), the sequence \(\{Z^{t}, E^t, Q^t, Z_0^t, w^t, Y^{t}_1, Y^{t}_2\}\) that converges to \(\{Z^*, E^*, Q^*, Z_0^*, w^*, Y_1^*, Y_2^*\}\). Without loss of generality, we assume that \(\{Z^*, E^*, Q^*, Z_0^*, w^*, Y_1^*, Y_2^*\}\) is a stationary point of the optimality problem in Eq. (5).

Proof: By Theorem 1, we know \(\{Z^{t}, E^t, Q^t, Z_0^t, w^t, Y^{t}_1, Y^{t}_2\}\) is bounded. Then, according to the Bolzano-Weierstrass theorem, the sequence has at least one accumulation point, denoted as \(\{Z^*, E^*, Q^*, Z_0^*, w^*, Y_1^*, Y_2^*\}\). Next, we will show that \(\{Z^*, E^*, Q^*, Z_0^*, w^*, Y_1^*, Y_2^*\}\) is a stationary point of Eq. (5).

By the assumption that \(\sum \frac{\rho_{t+1}}{\rho_t} < \infty\), it is clear that \(\rho^t \to 0\). Then, by the boundedness of \(\{Y^{t}_1\}\) and \(\{Y^{t}_2\}\), we have

\[Q^* - Z^* = \lim_{t \to \infty} Q^{t+1} - Z^{t+1} = \lim_{t \to \infty} (Q^{t+1}_1 - Y^{t}_1)/\rho^t = 0\]

\[P_K - Z^{*} - E^* = \lim_{t \to \infty} P_K - Z^{t+1} - E^{t+1} = \lim_{t \to \infty} (Y^{t}_2 - \rho^t)/\rho^t = 0\]

At the \((t+1)\)-th iteration, according to the optimality conditions of \(Q^t\) and \(E^t\), we have the following equalities:

\[\lim_{t \to \infty} \frac{\partial Q}{\partial Q}(Z^{t+1}, E^{t+1}, Q, Z_0^t, w^t, Y_1^t, Y_2^t) = 0\]

\[\lim_{t \to \infty} \frac{\partial P_{K}}{\partial Q}(Q^{t+1} + \rho^t(Q^{t+1} - Z^{t+1}) + Y^{t}_2) = 0\]

\[\lim_{t \to \infty} \frac{\partial E}{\partial E}(Z^{t+1}, E, Q^t, Z_0^t, w^t, Y_1^t, Y_2^t) = 0\]

\[\lim_{t \to \infty} \frac{\partial P_{K}}{\partial P_{K}}(Q^{t+1} + \rho^t(Q^{t+1} - Z^{t+1}) + Y^{t}_2) = 0\]

For ease of notation, we denote \(G^{t+1} = \partial(\alpha \sum_{v=1}^{V} ||Z_0^t - w^v_1/Z^{(v)}||_F^2 + \beta \sum_{j=1}^{V} w^v_j \text{Tr}((Z^{(v)}^T Z^{(v)})/Z_0^t)\) which may be obtained by tensorizing the first three terms of Eq. (9). Then, at the \((t+1)\)-th iteration, according to the optimality condition of \(Z^{t+1}\), and under the assumptions that \(\rho^t(Q^{t+1} - Q^t) \to 0\) and \(\rho^t(E^{t+1} - E^t) \to 0\), it is clear that

\[\lim_{t \to \infty} (G^{t+1} + \rho^t(Q^{t+1} - Q^t) + \rho^t(E^{t+1} - E^t) - P_K) = 0\]
are optimized using the standard convex programming techniques, the corresponding Karush-Kuhn-Tucker (KKT) conditions are satisfied by \( w^* \) and \( Z^* \).

Therefore, \( \{Z^*, \tilde{Z}^*, \tilde{Z}_w^*, \tilde{Z}_v^*, \tilde{Z}_v^* \} \) satisfies KKT conditions of Eq. (6) and thus \( \{Z^*, \tilde{Z}^*, \tilde{Z}_w^*, \tilde{Z}_v^*, \tilde{Z}_v^* \} \) is a stationary point of the original problem Eq. (5).

6 Experiments

In this section, we conduct extensive experiments to evaluate the proposed method. In particular, we use six benchmark data sets, including the BBC-4view, BBC-Sport, Flowers, UCI-3view, StillDB, and MITIndoor, and four evaluation metrics, including the clustering accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI), and F-Score, of which the detailed descriptions can be found in [Wu et al., 2019; Larson, 2019], respectively. Seventeen state-of-the-art methods are adopted as baselines for comparison, including the AWP [Nie et al., 2018b], MvDSCN [Zhu et al., 2019], MLAN [Nie et al., 2018a], UOMvSC [Tang et al., 2023], EOMSC-CA [Liu et al., 2022], LMSC [Zhang et al., 2020], GMC [Wang et al., 2020], MLCES [Chen et al., 2020], FPMVS-CAG [Wang et al., 2022], CLR-MVP [Kang et al., 2019], t-SVDS-MSC [Zhang et al., 2020], AW2 [Yang et al., 2019], ETLMSC [Wu et al., 2019], RML [Li et al., 2019], and OMFVC-LICAG [Zhang et al., 2024], among which six are developed within the last two years.

For the baseline methods, we follow the parameters in the original papers. For the proposed method, we set the parameters in the following way. For all balancing parameters, we tune them within the set \( \{0.001,0.01,0.1,1,10,100,1000\} \). For \( p, \kappa, \) and \( N \), we fix them to 0.001, 1.5, and 5 throughout the paper. If not otherwise clarified, we use a third-order CCL-MVC in the experiment. For all methods, the final clustering step is repeated 10 times and we report the averaged results with parameters tuned to the best.

6.1 Clustering Performance

We compare the CCL-MVC with the baseline methods and report the clustering results in Section 5. In general, the CCL-MVC has the best performance among all methods, where it obtains the best results in all 24 cases. Among the baseline methods, the RML, SM²SC, ETLMSC and t-SVDS-MSC are among the most competitive ones, which obtain the top three results in 5, 11, 19, and 8 out of a total number of 24 cases, respectively. Compared with the baseline methods, the CCL-MVC has significantly improved performance, where it improves the performance by about 0.04-0.12, 0.01-0.04, 0.02-0.07, 0.01-0.03, 0.14-0.17, and 0.05-0.11 in different metrics on the BBC-4view, BBC-Sport, Flowers, UCI-3view, StillDB, and MITIndoor data sets, respectively. Among these data sets, the StillDB data set is considered as a “difficult” one, on which the baseline methods rarely obtain results higher than 0.4. On this data set, the CCL-MVC improves the performance by about 0.14-0.17 and 0.29-0.35 in different
metrics, compared with the top second and third methods, i.e., the ETLMSC and SM$^2$SC, respectively. Such improvements are indeed significant. Moreover, the CCL-MVC has superior performance in terms of its stability. In particular, the CCL-MVC has the best performance on all data sets, while none of the baseline methods have the top three performance on all data sets. For example, as the most competitive method among the baselines, the ETLMSC is not among the top three on the BBC-4view data set; although the RMSL has the top second performance on the BBC-4view data set, it fails on several other data sets. These observations confirm the effectiveness and stability of the CCL-MVC on these data sets.

6.2 Ablation Study

We conduct ablation study from two perspectives, using the BBC-Sport and StillDB data sets without loss of generality. First, we validate the significance of adopting cross-view diversity and present the results in Fig. 1. In particular, we test how the CCL-MVC performs when the cross-view diversity embedding is eliminated from the model, for which we set $\beta = 0$ and tune the other parameters in the same way as described in Section 6.1. From Fig. 1, we may observe that the performance of the CCL-MVC is significantly degraded when the diversity is not considered, which confirms the significance of cross-view diversity embedding.

Then, we verify the significance of adopting cross-order neighbor information and show the results in Fig. 2. Compared with the first-order CCL-MVC, the second and third-order models have significantly improved performance, which confirms the significance of adopting cross-order neighbor information in our model. Moreover, when the order is higher than 3, the CCL-MVC cannot be further improved, or may even degrade. This may be explained by the fact that a very high order may introduce redundancy or noise to the relationships. Given the difficulty in finding a suitable $K$, and considering the significance of incorporating the third-order information, it is reasonably convincing to recommend utilizing a third-order CCL-MVC in practical applications.

6.3 Convergence Study

Besides the theoretical results about convergence provided in Section 5, we further show some empirical results to illustrate the convergent behavior of the CCL-MVC. Without loss of generality, we show the results on the BBC-4view and StillDB data sets in Fig. 3. Due to space limit, rather than plotting the curves for each variable, we plot the curve of error sequence, where at the $t$-th iteration the error is defined as $\max\{\|Z^t - Z^{t-1}\|_F, \|E^t - E^{t-1}\|_F, \|Q^t - Q^{t-1}\|_F, \|Z_0 - Z_0^{t-1}\|_F, \|w^t - w^{t-1}\|_2, \|P_K - Z^t - E^t\|_F, \|Z^t - Q^t\|_F\}$. From the results, it is seen that the curves converge to zero within about 50 iterations, which is quite efficient. Similar observations can be found on other data sets as well, which is convincing to claim the convergent behavior of the CCL-MVC.

7 Conclusion

In this paper, we propose a novel CCL-MVC method for multi-view clustering. The CCL-MVC incorporates cross-view diversity to learn a consensus affinity matrix by fusing a low-rank essential tensor recovered from a fine-grained neighbor tensor that encompasses comprehensive and complementary cross-order information of multi-view data. We develop an effective optimization algorithm for the CCL-MVC, which is proved to converge with theoretical guarantee. Extensive experimental results show the superiority of the CCL-MVC.

Acknowledgements

Y.C. and C.C. are corresponding authors. This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62276147, 62172246, and 62106063; in part by the Shandong Province Colleges and Universities Youth Innovation Technology Plan Innovation Team Project under Grants 2022KJ149, 2021KJ062, and 2020KJN011; and in part by the Guangdong Major Project of Basic and Applied Basic Research under Grant 2023B030300010.
References


[Peng et al., 2022a] Chong Peng, Yang Liu, Kehan Kang, Yongyong Chen, Xinxing Wu, Andrew Cheng, Zhao


