Tensorized Label Learning on Anchor Graph

Jing Li[1], **Quanxue Gao**[1*], **Qianqian Wang**[1], **Wei Xia**[1]

[1]School of Telecommunications Engineering, Xidian University
jinglxd@stu.xidian.edu.cn, qxgao@xidian.edu.cn, qqwang@xidian.edu.cn, xdweixia@gmail.com

Abstract

Graph-based multimedia data clustering has attracted much attention due to the impressive clustering performance for arbitrarily shaped multimedia data. However, existing graph-based clustering methods need post-processing to get labels for multimedia data with high computational complexity. Moreover, it is sub-optimal for label learning due to the fact that they exploit the complementary information embedded in data with different types pixel by pixel. To handle these problems, we present a novel label learning model with good interpretability for clustering. To be specific, our model decomposes anchor graph into the products of two matrices with orthogonal non-negative constraint to directly get soft label without any post-processing, which remarkably reduces the computational complexity. To well exploit the complementary information embedded in multimedia data, we introduce tensor Schatten p-norm regularization on the label tensor which is composed of soft labels of multimedia data. The solution can be obtained by iteratively optimizing four decoupled sub-problems, which can be solved more efficiently with good convergence. Experimental results on various datasets demonstrate the efficiency of our model.

Introduction

Multimedia data we obtain are usually unlabeled and clustering multimedia data becomes a crucial step in its analysis. Therefore, multimedia data clustering is a hot research direction. Multimedia data typically contains information from diverse modalities, such as images, text and voice. By considering data from different modalities as distinct viewpoints, we can apply multi-view clustering (MVC) techniques to uncover the latent value in multimedia data. Multimedia data clustering has attracted considerable attention in recent years (Tao et al. 2017; Wu, Lin, and Zha 2019; Shu et al. 2023), graph-based multimedia data clustering is one of the representative methods (Zhou and Burges 2007; Saha 2013; Nie et al. 2016; Wang, Yang, and Liu 2019; Zhou, Yang, and Gao 2022), which is widely used because the impressive clustering performance for arbitrarily shaped multimedia data. But the core idea behind the graph-based approach is to construct $n \times n$ graph and eigen-decomposition of Laplacian matrix, whose computational complexity are $\mathcal{O}(Vn^2)$ and $\mathcal{O}(n^3)$, respectively, where V and n are the number of views and samples. Therefore, graph-based methods tend to fail when clustering large-scale multimedia data.

To tackle the aforementioned issues, several solutions have been proposed, of which the anchor graph-based approach is a notable example (Liu, He, and Chang 2010; Li et al. 2015; Kang et al. 2020; Yang et al. 2021). Anchor graph can represent the complex mechanisms of multimedia data by modeling the relationship between n samples and m anchors. Since the number of anchors is much smaller than that of samples, the anchor graph-based approach can reduce both the computational complexity and storage complexity compared to the graph-based approach. Hence, the anchor-based approach is effective.

On the one hand, in order to get the labels of multimedia data from anchor graph, it usually has two ideas. The first one is to perform post-processing on anchor graph to get the final labels, which is time-consuming. Another is to obtain clustering labels based on connected components, but it is highly parameter dependent and difficult to ensure K-connected components under certain parameter settings. So, this idea is not suitable in practical applications. On the other hand, in multimedia data clustering, leveraging complementary information across different data types is often necessary to improve clustering performance. However, the existing method based on anchor graph exploit the complementary information embedded in data with different types pixel by pixel, which ignores the spatial structure information between different types of data, leading to sub-optimal for label learning.

Based on the above arguments, we are concerned in this paper with a novel label learning model with good interpretability for clustering. To be specific, considering that anchor graph can reduce the computational complexity and storage complexity compared to the full graph, we first construct anchor graph. Then, inspired by the excellent clustering interpretability of the orthogonal non-negative matrix factorization, we decompose the anchor graph of each view into the products of two matrices with orthogonal non-negative constraint, which in turn can directly obtain the soft labels of clustering without any post-processing and further reduces the computational complexity. To better explore the complementary information of different views, inspired by

*Corresponding author
Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

the tensor Schatten p-norm (Gao et al. 2021; Xia et al. 2023; Yang et al. 2022), we use the soft labels of each view to form a 3rd-order tensor as illustrated in Fig. 1, and exploit tensor Schatten p-norm regularizer on this term. This helps to get the complementary information embedded in the soft labels of different views. Finally, extensive experiments demonstrate that our algorithm guarantees stable and good clustering performance.

The main contributions are summarized as follows:

- To improve clustering efficiency, we adopt the idea of anchor graph and decompose it into the products of two matrices with orthogonal no-negative constraint to get soft label directly without post-processing.
- We minimize the tensor Schatten p-norm to reduce the divergence between the clustering soft labels of the different views, and thus make full use of the structural and complementary information between views.
- We propose an effective optimization strategy based on the Augmented Lagrangian Multiplier (ALM) method. And we mathematically prove that the proposed algorithm always converges to the KKT stationary point. Extensive experiments on different datasets have shown the good performance of our proposed method.

Related Work

Graph-Based Multimedia Data Clustering

Graphs are used to represent relationships between data objects. The eigenvalues of the graph matrix reflect the connected components in the graph, which can be used to cut the constructed graph to achieve clustering. The purpose of graph-based MVC is to learn the fusion graph from the graph of each view and then use the spectral graph technique to obtain the clustering result. Zhou et al. (Zhou and Burges 2007) for the first time extended clustering from a single view to multiple views through the normalized graph cutting algorithm. In order to better represent data objects, Saha (Saha 2013) carried out feature mapping on the original data and then completed MVC using the spectral theory. Nie et al. (Nie et al. 2016, 2017) carried out research on the auto-weighted graph of different views. The proposed method can automatically weight the initial graph of each view to learn the consistent fusion graph, and can directly output the clustering results on the fusion graph. Although the graph-based MVC has achieved impressive results in many clustering tasks, it is not satisfactory when processing large-scale multimedia data due to the high computational complexity.

Anchor Graph-Based Multimedia Data Clustering

The anchor graph-based MVC can be efficient in clustering multimedia data on a large scale because its dimension is significantly lower than that of the traditional affinity graph. Inspired by anchor graph. Some related work have been proposed by many authors, for example, Liu (Liu, He, and Chang 2010) solved scalable semi-supervised learning by constructing anchor graphs using anchor points to cover the entire data point cloud. Inspired by this work, Li (Li et al. 2015) proposed a novel large-scale multi-view spectral clustering approach based on the bipartite graph. As a first attempt to solve the large-scale multi-view subspace clustering problem, Kang et al. applied the anchor graph-based method for subspace clustering and proposed LMVSC (Kang et al. 2020) by means of anchors and anchor graphs instead of the original data points and similarity matrix. To further improve the efficiency of multi-view clustering, the FMCNOF (Yang et al. 2021) proposed by Yang et al. replaces the original data matrix with anchor graphs and uses them for non-negative orthogonal factorization. Since the number of anchors is much less than the number of original data, it can effectively improve the clustering efficiency. However, FMCNOF assumes that the clustering labels are the same on different views, this constraint is too strong and may lead to over-fitting. Also, it uses F-norm to calculate the error between different views pixel by pixel, causing to ignore the spatial structure information between different views. Our proposed method addresses these issues.

Notations

We introduce the notations used throughout this paper. We use bold calligraphy letters for 3rd-order tensors, $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, bold upper case letters for matrices, $\mathbf{M}$, bold lower case letters for vectors, $\mathbf{m}$, and lower case letters such as m_{ijk} for the entries of $\boldsymbol{\mathcal{M}}$. Moreover, the i-th frontal slice of $\boldsymbol{\mathcal{M}}$ is $\boldsymbol{\mathcal{M}}^{(i)}$. $\overline{\boldsymbol{\mathcal{M}}}$ is the discrete Fourier transform (DFT) of $\boldsymbol{\mathcal{M}}$ along the third dimension, $\overline{\boldsymbol{\mathcal{M}}} = \mathrm{fft}(\boldsymbol{\mathcal{M}}, [], 3)$. Thus, $\boldsymbol{\mathcal{M}} = \mathrm{ifft}(\overline{\boldsymbol{\mathcal{M}}}, [], 3)$. The trace and transpose of matrix $\mathbf{M}$ are expressed as $tr(\mathbf{M})$ and $\mathbf{M}^{\mathrm{T}}$. The F-norm of $\boldsymbol{\mathcal{M}}$ is denoted by $\|\boldsymbol{\mathcal{M}}\|_F$. (To save spaces, the definitions of t-product, orthogonal tensor, f-diagonal tensor and tensor transpose are omitted (e.g. see (Kilmer and Martin 2011))).

Definition 1. *(Gao et al. 2021) Given $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $h = \min(n_1, n_2)$, the tensor Schatten p-norm of $\boldsymbol{\mathcal{M}}$ is defined as*

$$\|\mathcal{M}\|_{\mathcal{S}_p} = \left(\sum_{i=1}^{n_3} \left\| \overline{\mathcal{M}}^{(i)} \right\|_{\mathcal{S}_p}^{p} \right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n_3} \sum_{j=1}^{h} \sigma_j \left(\overline{\mathcal{M}}^{(i)} \right)^{p} \right)^{\frac{1}{p}}, \tag{1}$$

where, $0 \leq p \leq 1$, $\sigma_j(\overline{\boldsymbol{\mathcal{M}}}^{(i)})$ denotes the j-th singular value of $\overline{\boldsymbol{\mathcal{M}}}^{(i)}$.

It should be pointed out that for $0 \leq p \leq 1$, when p is appropriately chosen, the Schatten p-norm provides quite effective improvements for a tighter approximation of the rank function (Zha et al. 2020; Xie et al. 2016).

Methodology

Proposed Formula

The graph-based multimedia data clustering method first constructs the graph and then uses clustering techniques on the constructed graph to obtain the final clustering results. Considering that the construction of similar graphs using the original data directly is a huge waste of time and space, inspired by the idea of anchor graphs, we first select some of

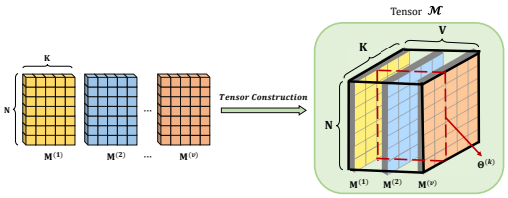

Figure 1: Tensor construction process.

the original data points from $\mathbf{X}^{(v)}$ as anchors and then construct the anchor graph $\mathbf{E}^{(v)} \in \mathbb{R}^{n\times m}$, which exploits the relationship between n original data points and m anchors, and since $m \ll n$, constructing anchor graphs can significantly reduce the computational complexity compared to the full graph.

In order to obtain clustering results from the constructed anchor graphs, inspired by FMCNOF (Yang et al. 2021), we consider implementing a non-negative matrix factorization (NMF) (Lee and Seung 1999) directly on the anchor graphs of the different views.

NMF decomposes the anchor graph of each view obtained above into the products of two non-negative matrices. As shown in Eq. (2), the product of two non-negative matrices $\mathbf{M}$, $\mathbf{Q}$ obtained after the NMF is approximately equal to the matrix $\mathbf{E}$. In order to make the matrices before and after the factorization closer, the reconstruction process can usually be formulated as an optimization problem with ℓ_2-norm or F-norm, and in this paper, the F-norm is used to construct the objective function in order to make the model easy to optimize.

$$\min_{\mathbf{M}\geqslant 0,\mathbf{Q}\geqslant 0} \left\|\mathbf{E}-\mathbf{M}\mathbf{Q}^{\mathrm{T}}\right\|_F^2 \tag{2}$$

where $\mathbf{E} \in \mathbb{R}_+^{n\times m}$, $\mathbf{M} \in \mathbb{R}_+^{n\times k}$ and $\mathbf{Q} \in \mathbb{R}_+^{m\times k}$.

With the widespread use of NMF, many variants of it have been proposed, such as $\mathbf{G}$-orthogonal NMF (Ding et al. 2006) and Semi-NMF (Ding, Li, and Jordan 2010). $\mathbf{G}$-orthogonal NMF obtains a unique factor matrix by constraining a factor matrix to be orthogonal. By constraining $\mathbf{M}$ in Eq. (2) to the orthogonal, we can obtain the objective function of the G-orthogonal NMF as:

$$\min_{\mathbf{M}\geqslant 0,\mathbf{Q}\geqslant 0,\mathbf{M}^{\mathrm{T}}\mathbf{M}=\mathbf{I}} \left\|\mathbf{E}-\mathbf{M}\mathbf{Q}^{\mathrm{T}}\right\|_F^2 \tag{3}$$

where $\mathbf{E}$ represents the anchor graph.

By relaxing the non-negative constraint on the input matrix $\mathbf{E}$ and one of the factor matrices $\mathbf{Q}$, we can obtain Semi-NMF, which can be applied to process a mixture of positive and negative input data. For $\mathbf{G}$-orthogonal NMF and Semi-NMF, Ding et al. (Ding, Li, and Jordan 2010) proposed:

Lemma 1. *G-orthogonal NMF and Semi-NMF are all relaxation of K-means clustering, and the main advantages are (1) Uniqueness of the solution; (2) Clustering interpretations.*

Remark 1. *G-orthogonal NMF has good clustering interpretability according to Lemma 1.* $\mathbf{M}$ *can be approximately regarded as the columns clustering label matrix of* $\mathbf{E}$ *in Eq. (2). The position of the maximum value of each row in* $\mathbf{M}$ *is the label corresponding to the data. So we can directly obtain the clustering label from* $\mathbf{M}$ *without any post-processing. Therefore, the clustering efficiency is improved.*

After applying the G-orthogonal NMF and Semi-NMF to the anchor graphs obtained in each view as described previously, we fused the matrix factorization errors of all views to obtain the following expression:

$$\begin{aligned} &\min \sum_{v=1}^{V} \frac{1}{\alpha_v}\left\|\mathbf{E}^{(v)}-\mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}}\right\|_F^2 + \beta\sum_{v=1}^{V}\mathcal{R}(\mathbf{M}^{(v)}) \\ &s.t. \quad \mathbf{M}^{(v)}\geqslant 0,\ (\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)}=\mathbf{I},\ \alpha_v\geqslant 0,\ \sum_{v=1}^{V}\alpha_v=1, \end{aligned} \tag{4}$$

where $\mathbf{E}^{(v)} \in \mathbb{R}^{n\times m}$ and $\mathbf{M}^{(v)} \in \mathbb{R}_+^{n\times k}$ are the anchor graph and the matrix of indicators, respectively. After obtaining soft labels for each view, it is necessary to make the labels from data with different types converge as much as possible by using regularization $\beta\sum_{v=1}^{V}\mathcal{R}(\mathbf{M}^{(v)})$. This enables us to leverage the complementary information from views effectively. Typically, F-norm is employed to calculate pixel by pixel, but this approach overlooks the spatial structure among the views.

In order to better exploit the complementary information and spatial structure between different views, we get inspiration from the excellent performance of the tensor Schatten p-norm (Gao et al. 2021; Yang et al. 2022; Yun et al. 2023). We introduce tensor Schatten p-norm regularization on the label tensor which is composed of soft labels of multimedia data. Our objective function is formulated as follows:

$$\begin{aligned} &\min \sum_{v=1}^{V} \frac{1}{\alpha_v}\left\|\mathbf{E}^{(v)}-\mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}}\right\|_F^2 + \beta\|\boldsymbol{\mathcal{M}}\|_{\circledS_p}^p \\ &s.t. \quad \mathbf{M}^{(v)}\geqslant 0,\ (\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)}=\mathbf{I},\ \alpha_v\geqslant 0,\ \sum_{v=1}^{V}\alpha_v=1, \end{aligned} \tag{5}$$

where $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{n\times V\times k}$, $0<p\leqslant 1$, β is the coefficient of the tensor Schatten p-norm term and is a self-defined parameter.

Remark 2. *The regularizer in the proposed objective (5) is used to explore the complementary information embedded in inter-views cluster assignment matrices* $\boldsymbol{M}^{(v)}$ *(*$v = 1, 2, \cdots, V$*). Fig. 1 shows the construction of tensor* $\boldsymbol{\mathcal{M}}$*, it can be seen that the* k*-th frontal slice* $\boldsymbol{\Theta}^{(k)}$ *describes the similarity between* N *sample points and the* k*-th cluster in different views. The idea cluster assignment matrix* $\boldsymbol{M}^{(v)}$ *should satisfy that the relationship between* N *data points and the* k*-th cluster is consistent in different views. Since different views usually show different cluster structures, we impose tensor Schatten p-norm minimization (Gao et al. 2021) constraint on* $\boldsymbol{\mathcal{M}}$*, which can make sure each* $\boldsymbol{\Theta}^{(k)}$ *has spatial low-rank structure. Thus* $\boldsymbol{\Theta}^{(k)}$ *can well characterize the complementary information embedded in inter-views.*

Optimization

Inspired by Augmented Lagrange Multiplier(ALM), we introduce two auxiliary variables $\boldsymbol{\mathcal{C}}$ and $\mathbf{W}^{(v)}$, and let $\boldsymbol{\mathcal{M}} =$

$\mathcal{C}$, $\mathbf{M}^{(v)} = \mathbf{W}^{(v)}$, respectively. Then, we rewrite the model as the following unconstrained problem:

$$\begin{aligned}
&\min \mathcal{L}(\mathbf{M}^{(v)}, \mathbf{Q}^{(v)}, \mathbf{W}^{(v)}, \alpha_v, \mathcal{C}) \\
&= \min_{\mathbf{W}^{(v)} \geqslant 0, (\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)} = \mathbf{I}} \sum_{v=1}^{V} \frac{1}{\alpha_v} \left\| \mathbf{E}^{(v)} - \mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}} \right\|_F^2 \\
&\quad + \beta \|\mathcal{C}\|_{\circledS p}^p + \langle \boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{M}} - \mathcal{C} \rangle + \frac{\mu}{2} \|\boldsymbol{\mathcal{M}} - \mathcal{C}\|_F^2 \\
&\quad + \sum_{v=1}^{V} (\langle \mathbf{Y}_2^{(v)}, \mathbf{M}^{(v)} - \mathbf{W}^{(v)} \rangle + \frac{\rho}{2} \|\mathbf{M}^{(v)} - \mathbf{W}^{(v)}\|_F^2)
\end{aligned} \tag{6}$$

where $\boldsymbol{\mathcal{Y}}_1, \mathbf{Y}_2^{(v)}$ represent Lagrange multipliers and μ, ρ are the penalty parameters. The optimisation process can therefore be separated into five steps:

•Solve $\mathbf{Q}^{(v)}$ with fixed $\mathcal{C}, \alpha_v, \mathbf{M}^{(v)}, \mathbf{W}^{(v)}$. Eq. (6) becomes:

$$\min \left\| \mathbf{E}^{(v)} - \mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}} \right\|_F^2, \tag{7}$$

Let $\mathbf{J} = \left\| \mathbf{E}^{(v)} - \mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}} \right\|_F^2$, we can obviously get the following equation:

$$\mathbf{J} = \mathrm{tr}((\mathbf{E}^{(v)})^{\mathrm{T}}\mathbf{E}^{(v)} - 2\mathrm{tr}((\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{E}^{(v)}\mathbf{Q}^{(v)} + \mathrm{tr}((\mathbf{Q}^{(v)})^{\mathrm{T}}\mathbf{Q}^{(v)}), \tag{8}$$

let $\partial \mathbf{J} / \partial \mathbf{Q}^{(v)} = 0$, we can get $-2(\mathbf{E}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)} + \mathbf{Q}^{(v)} = 0$.

So the solution of Eq. (7) is:

$$\mathbf{Q}^{(v)} = (\mathbf{E}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)}. \tag{9}$$

•Solve $\mathbf{M}^{(v)}$ with fixed $\mathbf{Q}^{(v)}, \mathbf{W}^{(v)}, \mathcal{C}, \alpha_v$. Eq. (6) becomes:

$$\begin{aligned}
&\min_{(\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)} = \mathbf{I}} \frac{1}{\alpha_v} \left\| \mathbf{E}^{(v)} - \mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}} \right\|_F^2 \\
&+ \langle \mathbf{Y}_1^{(v)}, \mathbf{M}^{(v)} - \mathbf{C}^{(v)} \rangle + \frac{\mu}{2} \left\| \mathbf{M}^{(v)} - \mathbf{C}^{(v)} \right\|_F^2 \\
&+ \langle \mathbf{Y}_2^{(v)}, \mathbf{M}^{(v)} - \mathbf{W}^{(v)} \rangle + \frac{\rho}{2} \left\| \mathbf{M}^{(v)} - \mathbf{W}^{(v)} \right\|_F^2,
\end{aligned} \tag{10}$$

which can be written as:

$$\begin{aligned}
\min_{(\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)} = \mathbf{I}} & -\frac{2}{\alpha_v} \mathrm{tr}(\mathbf{Q}^{(v)}(\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{E}^{(v)}) \\
& - \mu \mathrm{tr}((\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{D}^{(v)}) - \rho \mathrm{tr}((\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{B}^{(v)}),
\end{aligned} \tag{11}$$

where $\mathbf{D}^{(v)} = \mathbf{C}^{(v)} - \frac{\mathbf{Y}_1^{(v)}}{\mu}$ and $\mathbf{B}^{(v)} = \mathbf{W}^{(v)} - \frac{\mathbf{Y}_2^{(v)}}{\rho}$. It is clear that Eq. (11) is equivalent to

$$\begin{aligned}
\max_{(\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)} = \mathbf{I}} & \frac{2}{\alpha_v} \mathrm{tr}(\mathbf{Q}^{(v)}(\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{E}^{(v)}) \\
& + \mu \mathrm{tr}((\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{D}^{(v)}) + \rho \mathrm{tr}((\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{B}^{(v)}),
\end{aligned} \tag{12}$$

and Eq. (12) can be simplified as:

$$\max_{(\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{M}^{(v)} = \mathbf{I}} \mathrm{tr}((\mathbf{M}^{(v)})^{\mathrm{T}}\mathbf{F}^{(v)}), \tag{13}$$

where $\mathbf{F}^{(v)} = \frac{2\mathbf{E}^{(v)}\mathbf{Q}^{(v)}}{\alpha_v} + \mu \mathbf{D}^{(v)} + \rho \mathbf{B}^{(v)}$.

To solve Eq.(13), we introduce the following Theorem:

Theorem 1. *Give $\mathbf{G}$ and $\mathbf{P}$, where $\mathbf{G}(\mathbf{G})^{\mathrm{T}} = \mathbf{I}$. The optimal solution of*

$$\max_{\mathbf{G}(\mathbf{G})^{\mathrm{T}} = \mathbf{I}} tr(\mathbf{G}\mathbf{P}) \tag{14}$$

is $\mathbf{G} = \mathbf{V}[\mathbf{I}, \mathbf{0}](\mathbf{\Lambda})^{\mathrm{T}}$, where $\mathbf{V}$ and $\mathbf{\Lambda}$ can be obtained from $\mathbf{\Lambda}\mathbf{S}(\mathbf{V})^{\mathrm{T}} = svd(\mathbf{P})$

Proof. From the singular value decomposition $\mathbf{P} = \mathbf{\Lambda}\mathbf{S}(\mathbf{V})^{\mathrm{T}}$ and together with Eq. (14), it is evident that

$$\begin{aligned}
\mathrm{tr}(\mathbf{G}\mathbf{P}) &= \mathrm{tr}(\mathbf{G}\mathbf{\Lambda}^{(v)}\mathbf{S}(\mathbf{V})^{\mathrm{T}}) = \mathrm{tr}(\mathbf{S}(\mathbf{V})^{\mathrm{T}}\mathbf{G}\mathbf{\Lambda}) \\
&= \mathrm{tr}(\mathbf{S}\mathbf{H}) = \sum_i s_{ii} h_{ii},
\end{aligned} \tag{15}$$

where $\mathbf{H} = (\mathbf{V})^{\mathrm{T}}\mathbf{G}\mathbf{\Lambda}$, s_{ii} and h_{ii} are the i-th row and column elements of $\mathbf{S}$ and $\mathbf{H}$, respectively. We can simply verify $\mathbf{H}(\mathbf{H})^{\mathrm{T}} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix. Therefore $-1 \leqslant h_{ii} \leqslant 1$ and $s_{ii} \geqslant 0$, Thus we have:

$$\mathrm{tr}(\mathbf{G}\mathbf{P}) = \sum_i s_{ii} h_{ii} \leqslant \sum_i s_{ii}. \tag{16}$$

When $\mathbf{H}$ is an identity matrix, the equal sign holds. So when $\mathbf{H} = [\mathbf{I}, \mathbf{0}]$, $\mathrm{tr}(\mathbf{G}\mathbf{P})$ reaches the maximum. □

So the solution of Eq. (13) is:

$$\mathbf{M}^{(v)} = \mathbf{\Lambda}^{(v)}[\mathbf{I}, \mathbf{0}](\mathbf{V}^{(v)})^{\mathrm{T}}. \tag{17}$$

•Solve $\mathbf{W}^{(v)}$ with fixed $\mathbf{Q}^{(v)}, \mathbf{M}^{(v)}, \mathcal{C}, \alpha_v$. Eq. (6) becomes:

$$\min_{\mathbf{W}^{(v)} \geqslant 0} \langle \mathbf{Y}_2^{(v)}, \mathbf{M}^{(v)} - \mathbf{W}^{(v)} \rangle + \frac{\rho}{2} \left\| \mathbf{M}^{(v)} - \mathbf{W}^{(v)} \right\|_F^2, \tag{18}$$

which can be rewritten as:

$$\min_{\mathbf{W}^{(v)} \geqslant 0} \frac{\rho}{2} \left\| \mathbf{M}^{(v)} - \mathbf{W}^{(v)} + \frac{Y_2^{(v)}}{\rho} \right\|_F^2. \tag{19}$$

The solution of Eq. (19) is:

$$\mathbf{W}^{(v)} = (\mathbf{M}^{(v)} + \frac{Y_2^{(v)}}{\rho})_+. \tag{20}$$

•Solve $\mathcal{C}$ with fixed $\mathbf{Q}^{(v)}, \mathbf{M}^{(v)}, \mathbf{W}^{(v)}, \alpha_v$. After completing the square regarding $\mathcal{C}$, we can deduce

$$\begin{aligned}
\mathcal{C}^* &= \arg\min \beta \|\mathcal{C}\|_{\circledS p}^p + \langle \boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{M}} - \mathcal{C} \rangle + \frac{\mu}{2} \|\boldsymbol{\mathcal{M}} - \mathcal{C}\|_F^2 \\
&= \arg\min \frac{\beta}{\mu} \|\mathcal{C}\|_{\circledS p}^p + \frac{1}{2} \left\| \boldsymbol{\mathcal{M}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\mu} - \mathcal{C} \right\|_F^2,
\end{aligned} \tag{21}$$

which has a closed-form solution as Lemma 2 (Gao et al. 2021):

Lemma 2. *Suppose $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $h = \min(n_1, n_2)$, let $\mathcal{Z} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$. For the following equation*

$$\min_{\mathcal{X}} \tfrac{1}{2} \|\mathcal{X} - \mathcal{Z}\|_F^2 + \tau \|\mathcal{X}\|_{\circledS p}^p. \tag{22}$$

the optimal solution $\mathcal{X}^*$ is $\mathcal{X}^* = \Gamma_\tau(\mathcal{Z}) = \mathcal{U} * ifft(P_\tau(\overline{\mathcal{Z}})) * \mathcal{V}^{\mathrm{T}}$, where $P_\tau(\overline{\mathcal{Z}})$ is a f-diagonal tensor, whose diagonal elements can be found by using the GST algorithm introduced in (Gao et al. 2021).

Algorithm 1: Tensorized Label Learning on Anchor Graph (TLL-AG)

Input: Data matrices $\{\mathbf{X}^{(v)}\}_{v=1}^{V} \in \mathbb{R}^{N\times d_v}$; anchors numbers m; cluster number K.
Output: Cluster labels $\mathbf{Y}$ of each data points.

Initialize: $\alpha_v = 1/V$, $\mu = 10^{-5}$, $\rho = 10^{-5}$, $\eta = 1.1$, $\boldsymbol{\mathcal{Y}}_1 = 0$, $\mathbf{Y}_2^{(v)} = 0$, and $\mathbf{M}^{(v)}$ is an identity matrix;
Compute graph matrix $\mathbf{E}^{(v)}$ of each views;
while not condition **do**
  Update $\mathbf{Q}^{(v)}$ by solving Eq. (9);
  Update $\mathbf{M}^{(v)}$ by solving Eq. (17);
  Update $\mathbf{W}^{(v)}$ by solving Eq. (20);
  Update $\boldsymbol{\mathcal{C}}$ by using Eq. (21);
  Update α_v by using Eq. (26);
  Update $\boldsymbol{\mathcal{Y}}_1$, $\mathbf{Y}_2^{(v)}$, μ and ρ: $\boldsymbol{\mathcal{Y}}_1 = \boldsymbol{\mathcal{Y}}_1 + \mu(\boldsymbol{\mathcal{M}} - \boldsymbol{\mathcal{C}})$, $\mathbf{Y}_2^{(v)} = \mathbf{Y}_2^{(v)} + \rho(\mathbf{M}^{(v)} - \mathbf{W}^{(v)})$, $\mu = \min(\eta\mu, 10^{13})$, $\rho = \min(\eta\rho, 10^{13})$;
end while
Calculate the K clusters by using $\mathbf{M} = \sum_{v=1}^{V} \frac{\mathbf{M}^{(v)}}{\alpha_v} / \sum_{v=1}^{V} \frac{1}{\alpha_v}$;
return Clustering result.

Now the solution of Eq. (21) is:

$$\boldsymbol{\mathcal{C}}^* = \Gamma_{\frac{\beta}{\mu}}(\boldsymbol{\mathcal{M}} + \frac{\boldsymbol{\mathcal{Y}}_1}{\mu}). \tag{23}$$

•**Solve α_v with fixed $\mathbf{Q}^{(v)}, \mathbf{M}^{(v)}, \mathbf{W}^{(v)}, \boldsymbol{\mathcal{C}}$.** In this case, Eq. (6) becomes:

$$\min \sum_{v=1}^{V} \frac{(z^{(v)})^2}{\alpha_v} \quad \text{s.t.} \quad \alpha_v \geqslant 0, \sum_{v=1}^{V} \alpha_v = 1 \tag{24}$$

where $z^{(v)} = \left\|\mathbf{E}^{(v)} - \mathbf{M}^{(v)}(\mathbf{Q}^{(v)})^{\mathrm{T}}\right\|_F$.

According to Cauchy-Schwartz inequality, we have

$$\sum_{v=1}^{V} \frac{(z^{(v)})^2}{\alpha_v} \overset{\text{(i)}}{=} \left(\sum_{v=1}^{V} \frac{(z^{(v)})^2}{\alpha_v}\right)\left(\sum_{v=1}^{V} \alpha_v\right) \overset{\text{(ii)}}{\geqslant} \left(\sum_{v=1}^{V} z^{(v)}\right)^2 \tag{25}$$

where (i) holds because $\sum_{v=1}^{V} \alpha_v = 1$, (ii) holds because $\sqrt{\alpha_v} \propto \frac{z^{(v)}}{\sqrt{\alpha_v}}$. The optional α_v is

$$\alpha_v^+ = z^{(v)} / \sum_{v=1}^{V} z^{(v)}. \tag{26}$$

Finally, the optimization procedure for Tensorized Label Learning on Anchor Graph (TLL-AG) is outlined in Algorithm 1.

Convergence Analysis

Theorem 2. *[Convergence Analysis of Algorithm 1] Let* $\mathcal{P}_k = \{\boldsymbol{\mathcal{Q}}_k, \boldsymbol{\mathcal{W}}_k, \boldsymbol{\mathcal{M}}_k, \boldsymbol{\mathcal{C}}_k, \boldsymbol{\mathcal{Y}}_{1,k}, \boldsymbol{\mathcal{Y}}_{2,k}\}$, $1 \leq k < \infty$ *in* (6) *be a sequence generated by* ***Algorithm 1****, then*

1. *$\mathcal{P}_k$ is bounded with the assumption* $\lim_{k\to 0} \rho_k(\mathbf{W}_{k+1}^{(v)} - \mathbf{W}_k^{(v)}) = 0$;
2. *Any accumulation point of $\mathcal{P}_k$ is a stationary KKT point of* (6).

The proof will be provided in the appendix (in the Supplementary Material) and we need to mention that the KKT conditions can be used to determine the stop conditions for Algorithm 1, which are $\|\mathbf{M}_k^{(v)} - \mathbf{W}_k^{(v)}\|_\infty \leq \varepsilon$, $\|\boldsymbol{\mathcal{M}}_k - \boldsymbol{\mathcal{C}}_k\|_\infty \leq \varepsilon$.

Complexity Analysis

TLL-AG consists of two stages: 1) Construction of $\{\mathbf{E}^{(v)}\}_{v=1}^{V}$; 2) Iterative updating Eq. (6). The first stage takes $\mathcal{O}(Vnmd + Vnm\log(m))$. where V, m, n and d are the number of views, anchors, samples and the sum of features on each view, respectively. The second stage mainly focuses on solving $\mathbf{M}^{(v)}$, $\mathbf{Q}^{(v)}$ and $\boldsymbol{\mathcal{C}}$. The complexity in updating these variables iteratively are $\mathcal{O}(Vm^2k + Vmk^2)$, $\mathcal{O}(Vnkm + Vnk\log(k))$ and $\mathcal{O}(2Vnk\log(Vk) + V^2kn)$. Due to $m \ll n$, and k, V are small constants, the main computational complexity of solving Eq. (6) is $\mathcal{O}(Vnkm + Vm^2k)$. Finally, the total computational complexity of the proposed method is $\mathcal{O}(Vnmd + Vm^2k)$. As a result, our method is certain to scale well with data size.

#Dataset	#Samples	#View	#Class
MSRC	210	5	7
HandWritten4	2000	4	10
Mnist4	4000	3	4
Reuters	18758	5	6
NoisyMNIST	50000	2	10

Table 1: Multi-view datasets used in our experiments

Anchor Selection and Graph Construction

Detailed methods for anchor selection and anchor graph construction are provided in the appendix (appendix is in the Supplementary Material).

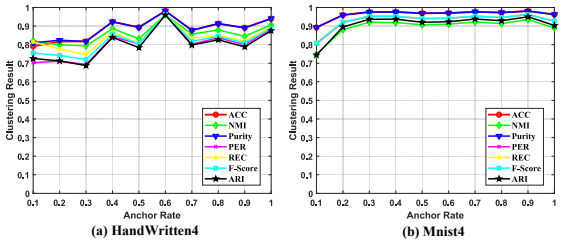

Figure 2: Clustering results of our algorithm with different anchor rate on HandWritten4 and Mnist4.

Experiments

In this section, we estimate the performance of our algorithm on 5 extensively used multi-view datasets and compare our algorithm with 7 state-of-art multi-view clustering

Dataset	MSRC			HandWritten4			Mnist4		
Metrics	ACC	NMI	PURITY	ACC	NMI	PURITY	ACC	NMI	PURITY
FMCNOF (Yang et al. 2021)	0.440	0.345	0.449	0.388	0.368	0.389	0.697	0.490	0.711
LMVSC (Kang et al. 2020)	0.814	0.717	0.814	0.904	0.831	0.904	0.892	0.726	0.892
GMC (Wang, Yang, and Liu 2019)	0.895	0.809	0.895	0.861	0.859	0.861	0.920	0.807	0.920
SMSC (Hu et al. 2020)	0.766	0.717	0.804	0.742	0.781	0.759	0.909	0.774	0.909
FPMVS-CAG (Wang et al. 2021)	0.786	0.686	0.786	0.744	0.753	0.744	0.885	0.715	0.885
ETLMSC (Wu, Lin, and Zha 2019)	0.962	0.937	0.962	0.938	0.893	0.938	0.934	0.847	0.934
MSC-BG (Yang et al. 2022)	0.981	0.960	0.981	0.889	0.922	0.889	0.938	0.861	0.938
ours	**0.986**	**0.968**	**0.986**	**0.982**	**0.962**	**0.982**	**0.981**	**0.935**	**0.981**

Table 2: Clustering performance on MSRC, HandWritten4 and Mnist4. (The best result is in bold, and the second-best result is underlined.)

Dataset	Reuters				NoisyMnist			
Metrics	ACC	NMI	Purity	Running Time	ACC	NMI	Purity	Running Time
FMCNOF (Yang et al. 2021)	0.365	0.106	0.405	186.45	0.333	0.237	0.340	192.60
LMVSC (Kang et al. 2020)	0.589	0.335	0.615	**150.51**	0.388	0.344	0.434	**151.14**
GMC (Wang, Yang, and Liu 2019)	-	-	-	-	-	-	-	-
SMSC (Hu et al. 2020)	OM	OM	OM	OM	OM	OM	OM	OM
FPMVS-CAG (Wang et al. 2021)	0.576	0.359	0.637	2252.20	0.554	0.513	0.567	2258.06
ETLMSC (Wu, Lin, and Zha 2019)	OM	OM	OM	OM	OM	OM	OM	OM
MSC-BG (Yang et al. 2022)	0.640	0.484	0.686	462.33	OM	OM	OM	OM
ours	**0.686**	**0.554**	**0.740**	406.53	**0.605**	**0.593**	**0.627**	415.93

Table 3: Clustering results and running time (in seconds) on Reuters and NoisyMNIST. (The best result is in bold, and the second-best result is underlined. "OM" means out of memory and "-" means the algorithm takes more than three hours to calculate.)

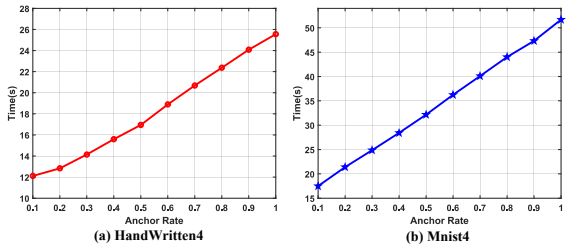

Figure 3: Computation time of our algorithm with different anchor rate on HandWritten4 and Mnist4 (in seconds).

methods. We estimate the clustering performance using metrics accuracy (ACC), normalized mutual information (NMI) and cluster purity (Purity). For each of these metrics, higher values mean better clustering performance. **Detailed experimental configurations and results are in the appendix**. It's available at https://github.com/xdjingli/TLL-AG.

Datasets and Baselines

The proposed algorithm was evaluated on 4 widely used multi-view datasets. **MSRC** (Winn and Jojic 2005), **HandWritten4** (Dua and Graff 2017), **Mnist4** (LeCun et al. 1998), **Reuters**(Apté, Damerau, and Weiss 1994), **Noisy MNIST** (Wang et al. 2015);. The details of the above datasets are shown in Table 1.

Our proposed algorithm is compared with the following multi-view clustering methods: **FMCNOF**, **LMVSC**, **GMC**, **SMSC**, **FPMVS-CAG**, **ETLMSC**, **MSC-BG**.

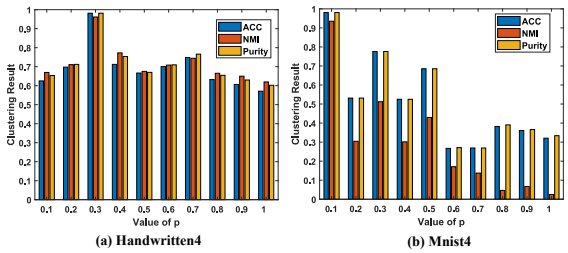

Figure 4: The influence of p on clustering results on HandWritten4 and Mnist4.

Experiments Result

Table 2 shows the performance of our proposed methods with other comparison algorithms on medium-scale datasets. Table 3. shows the clustering performance and running time of our proposed methods with other comparison algorithms on large-scale datasets.

It is clear that our proposed algorithm outperforms other comparison algorithms in terms of clustering metrics, and

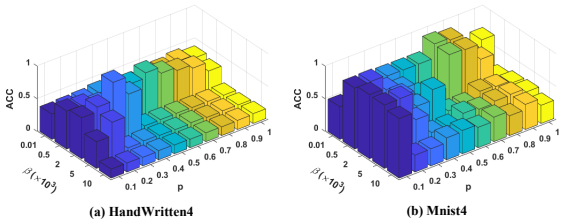

Figure 5: The influence of β on clustering results on HandWritten4 and Mnist4.

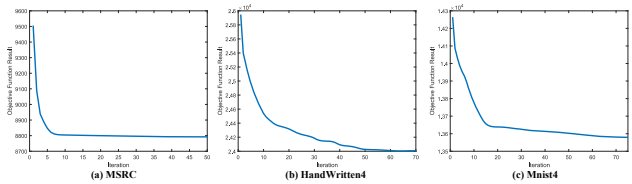

Figure 6: Convergence experiments on MSRC, HandWritten4 and Mnist4.

has achieved an acceptable running time. This may be due to the fact that the tensor Schatten p-norm regularization can be good for mining complementary information and spatial structure information between different views. And also due to the direct orthogonal non-negative matrix factorization of the anchor graph, the clustering labels can be obtained directly without post-processing while having good interpretability. Notably, unlike FMCNOF, we used VDA to select anchors to avoid unstable clustering results caused by random initial of K-means.

Impact of Parameters

In this subsection, we will analyze the effect of variable parameters on HandWritten4 and Mnist4.

Effect of Anchor Rate Our model uses anchor graph to perform semi-nonnegative matrix factorization. Anchor rate can affect the clustering results significantly, so it is a very important parameter. Its influence on the clustering results is mainly reflected in its influence on the clustering metrics and the algorithm running time. In the following, we illustrate each of them separately.

We set the anchor rate as the proportion of the anchors take in the total number of data points and turn the anchor rate from 0.1 to 1.0 with the interval of 0.1. The variation of seven metrics with anchor rate is shown in Fig 2. It can be found that when the anchor rate reaches 0.6 and 0.9, HandWritten4 and Mnist4 can achieve the best clustering performance, respectively. In order to obtain acceptable clustering metrics, we do not need to select numerous anchors.

To examine the effect of the number of anchors on the running time of the algorithm, we made the anchor rate vary from 0.1 to 1 at intervals of 0.1 and recorded the running time of our proposed method on HandWritten4 and Mnist4. The final results are shown in Fig 3. From the curves in the figure, it can be seen that the running time of the algorithm increases approximately linearly with the increase of the anchor rate. It indicates that in order to make the algorithm run faster, the number of anchors (anchor rate) we choose should be as small as possible.

However, given that clustering at lower anchor rates does not perform particularly well (as shown in Fig 2), the choice of the number of anchors needs to be traded off. In general, we generally choose the anchor rate at which the clustering performs well and the running time is within an acceptable range.

Effect of Parameter p We set the value of p to vary from 0.1 to 1 in intervals of 0.1. We obtained three clustering metrics (ACC, NMI, Purity) on HandWritten4 and Mnist4 as shown in Fig 4. It can be found that the value of p has a significant influence on the clustering performance. This is probably because that the Schatten p-norm approximates the target rank well.

Effect of Parameter β We set the value of β from $\{10, 500, 2000, 5000, 10000\}$ for HandWritten4 and Mnist4, and then we run experiment with varying values of β and p to determine the effect of β and p on the clustering metric ACC.

The experimental results for the variation of ACC with the values of β and p are shown in Fig 5. From this, it can be found that when the value of p is fixed, the impact of different values of β on the clustering performance is significant, and when the value of β is fixed, in most cases, different values of p taking on the clustering results have an effect. This means that the appropriate β value has to be chosen for the Schatten p-norm to work to some extent.

Experiments of Convergence

In this subsection we will verify the convergence of the proposed method.

We obtain the clustering results by minimising the objective function 5, so we can calculate the value of the objective function through different numbers of iterations to experimentally analyse the convergence of our model. Taking MSRC, HandWritten4 and Mnist4 as examples, we plotted the curve between the value of the objective function and the number of iterations, as shown in Fig 6. These experimental results show that the convergence of our proposed method satisfies our previous theoretical analysis.

Conclusion

We are concerned in this paper with a novel label learning model with good interpretability for clustering. To be specific, our method improves the efficiency of clustering by using anchor graphs and introducing orthogonal non-negative matrix factorization to obtain the clustering labels directly. In addition, by minimizing the tensor Schatten p-norm composed of the clustering label matrices, the complementary information and representation space structure embedded in data with different types can be well utilised to make the clustering results more reliable. Furthermore, we provide a solving algorithm for our model and it was proved to converge to the stationary KKT point. Finally, extensive experiments on different datasets demonstrate the effectiveness of our algorithm.

Acknowledgments

This work was supported by the Science and technology project of Xi'an (Grant 2022JH-JSYF-0009), by the Natural Science Basic Research Program of Shaanxi Province (Grant 2023-JC-YB-534), by the National Natural Science Foundation of China under Grants 62176203, by the Natural Science Foundation of Shandong Province under Grant ZR202102180986, by the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University, by the Guangxi Key Laboratory of Digital Infrastructure under Grant GXDIOP2023010, by the Natural Science Foundation of Guangdong Province, 2023A1515011845.

References

Apté, C.; Damerau, F.; and Weiss, S. M. 1994. Automated Learning of Decision Rules for Text Categorization. *ACM Trans. Inf. Syst.*, 12(3): 233–251.

Ding, C.; Li, T.; and Jordan, M. I. 2010. Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence*, 32(1): 45–55.

Ding, C.; Li, T.; Peng, W.; and Park, H. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 126–135.

Dua, D.; and Graff, C. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml.

Gao, Q.; Zhang, P.; Xia, W.; Xie, D.; Gao, X.; and Tao, D. 2021. Enhanced tensor RPCA and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6): 2133–2140.

Hu, Z.; Nie, F.; Wang, R.; and Li, X. 2020. Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding. *Information Fusion*, 55: 251–259.

Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI conference on Artificial Intelligence*, 4412–4419.

Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.

Lee, D. D.; and Seung, H. S. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788–791.

Li, Y.; Nie, F.; Huang, H.; and Huang, J. 2015. Large-scale multi-view spectral clustering via bipartite graph. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Liu, W.; He, J.; and Chang, S.-F. 2010. Large graph construction for scalable semi-supervised learning. In *ICML*.

Nie, F.; Li, J.; Li, X.; et al. 2016. Parameter-free auto-weighted multiple graph learning: a framework for multi-view clustering and semi-supervised classification. In *IJCAI*, 1881–1887.

Nie, F.; Li, J.; Li, X.; et al. 2017. Self-weighted Multiview Clustering with Multiple Graphs. In *IJCAI*, 2564–2570.

Saha, M. 2013. A graph based approach to multiview clustering. In *International Conference on Pattern Recognition and Machine Intelligence*, 128–133. Springer.

Shu, X.; Zhang, X.; Gao, Q.; Yang, M.; Wang, R.; and Gao, X. 2023. Self-Weighted Anchor Graph Learning for Multi-View Clustering. *IEEE Trans. Multim.*, 25: 5485–5499.

Tao, Z.; Liu, H.; Li, S.; Ding, Z.; and Fu, Y. 2017. From ensemble clustering to multi-view clustering. In *IJCAI*.

Wang, H.; Yang, Y.; and Liu, B. 2019. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6): 1116–1129.

Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2021. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31: 556–568.

Wang, W.; Arora, R.; Livescu, K.; and Bilmes, J. A. 2015. On Deep Multi-View Representation Learning. In *ICML*, volume 37, 1083–1092.

Winn, J. M.; and Jojic, N. 2005. LOCUS: Learning Object Classes with Unsupervised Segmentation. In *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, 756–763.

Wu, J.; Lin, Z.; and Zha, H. 2019. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12): 5910–5922.

Xia, W.; Gao, Q.; Wang, Q.; Gao, X.; Ding, C.; and Tao, D. 2023. Tensorized Bipartite Graph Learning for Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 5187–5202.

Xie, Y.; Gu, S.; Liu, Y.; Zuo, W.; Zhang, W.; and Zhang, L. 2016. Weighted Schatten p-norm minimization for image denoising and background subtraction. *IEEE Trans. Image Process.*, 25(10): 4842–4857.

Yang, B.; Zhang, X.; Nie, F.; Wang, F.; Yu, W.; and Wang, R. 2021. Fast Multi-View Clustering via Nonnegative and Orthogonal Factorization. *IEEE transactions on image processing*, 30: 2575–2586.

Yang, H.; Gao, Q.; Xia, W.; Yang, M.; and Gao, X. 2022. Multiview spectral clustering with bipartite graph. *IEEE Transactions on Image Processing*, 31: 3591–3605.

Yun, Y.; Li, J.; Gao, Q.; Yang, M.; and Gao, X. 2023. Low-rank discrete multi-view spectral clustering. *Neural Networks*, 166: 137–147.

Zha, Z.; Yuan, X.; Wen, B.; Zhou, J.; Zhang, J.; and Zhu, C. 2020. A benchmark for sparse coding: When group sparsity meets rank minimization. *IEEE Transactions on Image Processing*, 29: 5094–5109.

Zhou, D.; and Burges, C. J. 2007. Spectral clustering and transductive learning with multiple views. In *Proceedings of the 24th international conference on Machine learning*, 1159–1166.

Zhou, Q.; Yang, H.; and Gao, Q. 2022. Low-rank constraint bipartite graph learning. *Neurocomputing*, 511: 426–436.