

Auto-Weighted Multi-View Clustering for Large-Scale Data

Xinhang Wan¹, Xinwang Liu^{1*}, Jiyuan Liu¹, Siwei Wang¹, Yi Wen¹,
Weixuan Liang¹, En Zhu¹, Zhe Liu², Lu Zhou²

¹College of Computer, National University of Defense Technology, Changsha, China

²College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China
{wanxinhang, xinwangliu, liujiyuan13, wangsiwei13, wenyi21, weixuanliang, enzhu}@nudt.edu.cn, {zhe.liu, lu.zhou}@nuaa.edu.cn

Abstract

Multi-view clustering has gained broad attention owing to its capacity to exploit complementary information across multiple data views. Although existing methods demonstrate delightful clustering performance, most of them are of high time complexity and cannot handle large-scale data. Matrix factorization-based models are a representative of solving this problem. However, they assume that the views share a dimension-fixed consensus coefficient matrix and view-specific base matrices, limiting their representability. Moreover, a series of large-scale algorithms that bear one or more hyperparameters are impractical in real-world applications. To address the two issues, we propose an auto-weighted multi-view clustering (AWMVC) algorithm. Specifically, AWMVC first learns coefficient matrices from corresponding base matrices of different dimensions, then fuses them to obtain an optimal consensus matrix. By mapping original features into distinctive low-dimensional spaces, we can attain more comprehensive knowledge, thus obtaining better clustering results. Moreover, we design a six-step alternative optimization algorithm proven to be convergent theoretically. Also, AWMVC shows excellent performance on various benchmark datasets compared with existing ones. The code of AWMVC is publicly available at <https://github.com/wanxinhang/AAAI-2023-AWMVC>.

Introduction

With the rapid development of multimedia techniques, data can be described from various modalities (Xia et al. 2022c). For instance, an object could be recognized by an intelligent robot from its eyes (image), ears (transform sound into texts), and past knowledge (knowledge graph). How to uncover items' intrinsic structure and label them is crucial in many applications (Zhang et al. 2022b; Yang et al. 2022; Xia et al. 2022b), such as the recommender system and decision support system. Multi-view clustering, which explores complementary information among views and discovers the underlying structure for clustering, is an effective method to conduct a clustering process for data from multiple sources.

To the best of our knowledge, existing multi-view clustering algorithms can be divided into four categories including multi-view subspace clustering (Gao et al. 2015; Liu et al.

2021a, 2022a), multiple kernel clustering (Zhang et al. 2021; Li et al. 2022a; Zhang et al. 2022a; Wan et al. 2022), graph-based methods (Liang et al. 2020; Tang et al. 2020; Liu et al. 2022b,c; Xia et al. 2022a), and matrix factorization-based methods (Khan et al. 2019; Gao et al. 2019). Under the assumption that a linear combination of data samples can reconstruct themselves, multi-view subspace clustering obtains a reconstruction matrix upon a self-expressive framework. For instance, (Zhang et al. 2017) simultaneously explores the latent representation of data samples and the underlying complementary information to attain a more comprehensive reconstruction matrix. Multiple kernel clustering (Liu 2021) always seeks a common cluster assignment matrix by jointly maximizing the partition matrix and kernel coefficients. As pointed out by (Li et al. 2022b), graph-based methods rely on spectral clustering, which performs eigen-decomposition upon the Laplacian matrix to partition the data cloud with a linear/nonlinear relationship.

Despite the above three categories of algorithms displaying acceptable clustering performance, they encounter the square or cubic time complexity, which limits the applicability in large-scale data. To handle the issue, a series of researches (Liu et al. 2013; Gao et al. 2019) based on non-negative matrix factorization (NMF) appear due to its low time complexity. NMF-based multi-view clustering factorizes the original features into two components, i.e., a consensus coefficient matrix and view-specific base matrices. For example, (Wang et al. 2018) proposes a method to reduce the redundancy using a diversity term and solves the resultant optimization problem in linear execution time. However, the algorithm puts up with two hyperparameters, restraining the practicability in many applications. Based on the observation that the non-negative constraint leads to a less discriminatory embedding, (Liu et al. 2021b) removes the restriction and proposes an MF-based algorithm. However, it assumes that the latent embedding of multiple views is under a fixed latent dimension, which harms the expressiveness of the model and the ability to extract complementary information between views.

We propose auto-weighted multi-view clustering for large-scale data (AWMVC) to address the abovementioned drawbacks. To enhance the data embedding and gain complementary information among views, we obtain various coefficient matrices learned by base matrices under diverse di-

*Corresponding author

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

mensions, then integrate the coefficient matrices into a consensus one. It is worth noting that the diverse coefficient matrices can guide a more robust consensus item. As feedback, the consensus one promotes the quality of coefficient matrices likewise. In addition, the proposed method is parameter-free. Considering that the quality of each specific dimension exists differences, we automatically measure the importance of each one by the corresponding contributions to the final result. We develop a six-step alternate optimization algorithm with proven convergence to solve the resulting optimization problem. The linear time complexity and parameter-free property enable AWMVC to handle large-scale data. Compared with state-of-the-art multi-view clustering algorithms on various benchmark datasets, AWMVC displays effectiveness and superiority. In general, the contributions of the paper are summarized as follows:

1. We remove the non-negative constraint of NMF and obtain coefficient matrices with view-specific base matrices of different dimensions, then integrate the coefficient matrices into a consensus one and combine these two separate steps into a unified one, which filters redundant information and attains a more comprehensive knowledge.
2. The proposed algorithm can be carried out in linear time complexity with no hyperparameters, which is suitable for large-scale applications. The importance of each part is automatically tuned by its corresponding contribution to the final result. Therefore, the clustering result is superior with more discriminatory information.
3. We develop a six-step alternative optimization algorithm with proven convergence. To validate the effectiveness of AWMVC, we conduct extensive experiments on various datasets. The results demonstrate the efficiency and excellent performance of the proposed method.

Related Work

In this section, we briefly overview MF-based single-view clustering and MF-based multi-view clustering separately.

MF-Based Single-View Clustering

Given a single-view dataset $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, where d and n denote the number of features and samples respectively, NMF seeks to factorize data matrix into two non-negative components, i.e., base matrix $\mathbf{H} \in \mathbb{R}^{d \times k}$ and coefficient matrix $\mathbf{Z} \in \mathbb{R}^{k \times n}$ with k indicating the cluster numbers. The goal of NMF is to approximate the product of \mathbf{H} and \mathbf{Z} to \mathbf{X} as follows

$$\min_{\mathbf{H} \geq 0, \mathbf{Z} \geq 0} f(\mathbf{X}, \mathbf{HZ}), \quad (1)$$

where $f(\cdot)$ is a loss function. Most papers measure the loss between the two items by adopting the Frobenius norm (Liu et al. 2013, 2020). Based on the observation that the orthogonality constraint on NMF leads to rigorous clustering interpretation (Ding et al. 2006), a large number of researchers have investigated the NMF of orthogonality constraint. By imposing orthogonality constraints on base matrices, Eq. (1) can be formulated into

$$\min_{\mathbf{H} \geq 0, \mathbf{Z} \geq 0} \|\mathbf{X} - \mathbf{HZ}\|_F^2 \quad \text{s.t. } \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k. \quad (2)$$

Noticing that the non-negative constraint produces less discriminatory information, (Liu et al. 2021b) removes the restriction and proposes an MF-based method, as shown

$$\min_{\mathbf{H}, \mathbf{Z}} \|\mathbf{X} - \mathbf{HZ}\|_F^2 \quad \text{s.t. } \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k. \quad (3)$$

After getting the coefficient matrix \mathbf{Z} , a traditional clustering method like k -means is conducted to obtain the final cluster assignments.

MF-Based Multi-View Clustering

Given the multi-view data $\{\mathbf{X}^{(v)}\}_{v=1}^V$, in which $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$ and d_v denotes its feature dimension, MF-based multi-view clustering aims to seek a consensus matrix \mathbf{Z} to extract information among views. One of the most representative formulas is as follows (Wang et al. 2018)

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{Z}} \sum_{v=1}^V \frac{1}{2} \|\mathbf{X}^{(v)} - \mathbf{H}^{(v)} \mathbf{Z}^{(v)}\|_F^2 + \\ \lambda \sum_{v \neq w} \text{DIVE}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(w)}) \end{aligned} \quad (4)$$

s.t. $1 \leq v, w \leq V, \mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{Z}^{(v)} \geq \mathbf{0}, \mathbf{Z}^{(w)} \geq \mathbf{0}$,

where $\text{DIVE}(\cdot, \cdot)$ is to ensure the diversity of the representation matrices and λ is the balanced hyperparameter. The consensus matrix \mathbf{Z} is acquired via the average value of $\mathbf{Z}^{(v)}$. On the contrary, some methods have been proposed to directly attain a common underlying matrix across views, shown as follows

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{Z}} \sum_{v=1}^V \frac{1}{2} \|\mathbf{X}^{(v)} - \mathbf{H}^{(v)} \mathbf{Z}\|_F^2 + \lambda \Phi(\mathbf{H}, \mathbf{Z}) \\ \text{s.t. } \mathbf{H}^{(v)\top} \mathbf{H}^{(v)} = \mathbf{I}_k, \end{aligned} \quad (5)$$

where $\Phi(\cdot)$ are some regularization terms on \mathbf{H} and \mathbf{Z} .

Although these above methods display acceptable performance, they suffer from two drawbacks that are hard to solve. First, most MFMVC algorithms decompose data matrices from multiple views into a fixed-dimension embedding, which ruins the representational power of the model. Meanwhile, insufficient data acquisition cannot ensure complementary knowledge among views. As a consequence, the final partition matrix fails to extract adequate information. Second, to balance the weights between the loss function and regularization terms, hyperparameters are inevitable in many methods, limiting their extensions to large-scale applications. In the next section, we propose a novel auto-weighted multi-view clustering method for large-scale data to handle these problems.

Methodology

In this section, we first provide the objective formulation of AWMVC, then offer a six-step alternate optimization algorithm to solve the resultant problem. After that, we will discuss its time complexity and convergence.

The Proposed Method

As mentioned above, current MFMVC methods are presented as Eq. (5). Although they demonstrate advantages among exiting MVC algorithms, there are still some drawbacks they fail to solve simultaneously. We observe that MFMVC encounters two main problems. First, the current MFMVC maps original features of various views into a fixed dimension. However, data matrices from multiple sources usually belong to specific latent dimensions, and the clustering result significantly depends on the dimensions. Most researchers ignore this shortcoming and merely set it equal to the number of clusters to avoid selecting the dimensions from a range of values. In contrast, others tackle this problem by introducing a hyperparameter. To deal with this issue, we project the original feature matrix of each view into several dimensions and automatically balance their weights according to their corresponding contributions. Second, state-of-the-art methods encounter one or more hyperparameters, preventing them from being applied to large-scale applications. Consequently, we eliminate the regularization term, and the optimization goal is as follows

$$\min_{\alpha, \mathbf{H}, \mathbf{Z}} \sum_{p=1}^m \sum_{v=1}^V \frac{1}{2} \alpha_p^2 \left\| \mathbf{X}^{(v)} - \mathbf{H}_p^{(v)} \mathbf{Z}_p \right\|_F^2 \quad (6)$$

s.t. $\alpha^\top \mathbf{1} = 1, \alpha \geq \mathbf{0}, \mathbf{H}_p^{(v)\top} \mathbf{H}_p^{(v)} = \mathbf{I}_k,$

where m denotes the total number of latent embedding of each view, and α_p indicates the weight of \mathbf{Z}_p .

By solving Eq. (6), we obtain \mathbf{Z}_p with different dimensions. How to fuse them efficiently is a new issue. We address this problem via mapping them into a k -dimension space by a rotation matrix. The separate process between coefficient matrices learning and the consensus one learning may result in a sub-optimal result, so we unify them into a unified step as

$$\min \sum_{p=1}^m \sum_{v=1}^V \frac{1}{2} \alpha_p^2 \left\| \mathbf{X}^{(v)} - \mathbf{H}_p^{(v)} \mathbf{Z}_p \right\|_F^2 - \sum_{p=1}^m \beta_p \text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right) \quad (7)$$

$$\text{s.t. } \alpha^\top \mathbf{1} = 1, \sum_{p=1}^m \beta_p^2 = 1, \alpha \geq \mathbf{0}, \beta \geq \mathbf{0},$$

$$\mathbf{Z}_p \mathbf{Z}_p^\top = \mathbf{I}_{d_p}, \mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_k, \mathbf{M} \mathbf{M}^\top = \mathbf{I}_k,$$

where d_p is the dimension of p -th latent embedding, \mathbf{M} denotes the consensus coefficient matrix. To simplify the optimization process, we impose orthogonality constraint on \mathbf{Z}_p rather than $\mathbf{H}_p^{(v)}$.

After attaining \mathbf{M} , we conduct k -means on it to get the final result.

Optimization

The optimization problem in Eq. (7) is not jointly convex when considering all variables simultaneously. Hence, we propose an alternating algorithm to optimize each variable while the others maintain fixed.

$\mathbf{H}_p^{(v)}$ Subproblem Considering that the base matrices are independent of each other, we list the optimization process of p -th base matrix of v -th view $\mathbf{H}_p^{(v)}$ as an example. When other variables except for $\mathbf{H}_p^{(v)}$ are fixed, the optimization problem in Eq. (7) is reformulated as

$$\min \left\| \mathbf{X}^{(v)} - \mathbf{H}_p^{(v)} \mathbf{Z}_p \right\|_F^2. \quad (8)$$

Differentiating the objective function respecting $\mathbf{H}_p^{(v)}$ and setting the derivative to zero, it is obtained that $\mathbf{H}_p^{(v)}$ is updated by

$$\mathbf{H}_p^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}_p^\top. \quad (9)$$

\mathbf{M} Subproblem Fixing $\mathbf{H}, \mathbf{Z}, \mathbf{W}, \alpha, \beta$, Eq. (7) is reduced to

$$\max \text{Tr}(\mathbf{M} \mathbf{A}) \quad \text{s.t. } \mathbf{M} \mathbf{M}^\top = \mathbf{I}_k, \quad (10)$$

where

$$\mathbf{A} = \sum_{p=1}^m \beta_p \mathbf{Z}_p \mathbf{Z}_p^\top \mathbf{W}_p. \quad (11)$$

Eq. (10) can be effectively solved via SVD with computational complexity $\mathcal{O}(nk^2)$.

\mathbf{W}_p Subproblem With other variables fixed in Eq. (7), \mathbf{W}_p can be updated by the following formula

$$\max \text{Tr} \left(\mathbf{W}_p^\top \mathbf{Z}_p \mathbf{M}^\top \right) \quad \text{s.t. } \mathbf{W}_p^\top \mathbf{W}_p = \mathbf{I}_k. \quad (12)$$

Similar to Eq. (10), it can be effectually solved by SVD with computational complexity $\mathcal{O}(d_p k^2)$.

\mathbf{Z}_p Subproblem By dropping the irrelevant variables involved in Eq. (7), the objective formulation concerning \mathbf{Z}_p can be rewritten as

$$\min \frac{1}{2} \alpha_p^2 \sum_{v=1}^V \left\| \mathbf{X}^{(v)} - \mathbf{H}_p^{(v)} \mathbf{Z}_p \right\|_F^2 - \beta_p \text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right) \quad (13)$$

s.t. $\mathbf{Z}_p \mathbf{Z}_p^\top = \mathbf{I}_{d_p},$

which can be further transformed into

$$\max \text{Tr}(\mathbf{Z}_p \mathbf{B}) \quad \text{s.t. } \mathbf{Z}_p \mathbf{Z}_p^\top = \mathbf{I}_{d_p}, \quad (14)$$

where

$$\mathbf{B} = \alpha_p^2 \sum_{v=1}^V \mathbf{X}^{(v)\top} \mathbf{H}_p^{(v)} + \beta_p \mathbf{M}^\top \mathbf{W}_p^\top. \quad (15)$$

Same as Eq. (10), it can be efficiently solved via SVD with computational complexity $\mathcal{O}(nd_p^2)$.

α_p Subproblem Given other variables, the formulation about α can be solved via optimizing the following formula

$$\min \sum_{p=1}^m \alpha_p^2 r_p^2 \quad \text{s.t. } \alpha^\top \mathbf{1} = 1, \alpha \geq \mathbf{0}, \quad (16)$$

where

$$r_p^2 = \sum_{v=1}^V \left\| \mathbf{X}^{(v)} - \mathbf{H}_p^{(v)} \mathbf{Z}_p \right\|_F^2. \quad (17)$$

Based on Cauchy-Schwarz inequality, α_p is updated by

$$\alpha_p = \frac{1}{r_p^2} \frac{1}{\sum_{p=1}^m \frac{1}{r_p^2}}. \quad (18)$$

β_p **Subproblem** Fixing \mathbf{H} , \mathbf{Z} , \mathbf{W} , \mathbf{M} , α , Eq. (7) is reduced to

$$\max \sum_{p=1}^m \beta_p \theta_p \text{ s.t. } \sum_{p=1}^m \beta_p^2 = 1, \beta \geq \mathbf{0}, \quad (19)$$

where

$$\theta_p = \text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right). \quad (20)$$

The optimal solution for Eq. (19) is

$$\beta_p = \frac{\theta_p}{\sqrt{\sum_{p=1}^m \theta_p^2}}. \quad (21)$$

Algorithm 1: Auto-Weighted Multi-View Clustering for Large-Scale Data

Input: Dataset $\{\mathbf{X}^{(v)}\}_{v=1}^V$, cluster number k .

Output: A consensus coefficient matrix \mathbf{M} .

- 1: Initialize \mathbf{Z} , \mathbf{W} , $\alpha = \mathbf{1}/m$, $\beta = \mathbf{1}/\sqrt{m}$.
 - 2: **while** not converged **do**
 - 3: Update $\mathbf{H}_p^{(v)}$ via Eq. (9).
 - 4: Update \mathbf{M} via Eq. (10).
 - 5: Update \mathbf{W}_p via Eq. (12).
 - 6: Update \mathbf{Z}_p via Eq. (14).
 - 7: Update α via Eq. (18).
 - 8: Update β via Eq. (21).
 - 9: **end while**
-

The alternate optimization process of AWMVC is summarized in Algorithm 1. After attaining the consensus coefficient matrix \mathbf{M} , we conduct k -means on it to obtain the final cluster assignments.

Discussion

Firstly, we analyze the time complexity of our proposed method. Then the convergence of AWMVC is proven theoretically. Afterward, we will discuss the extension of AWMVC.

Time complexity In the optimization process, the time complexity of updating \mathbf{M} , \mathbf{W}_p and \mathbf{Z}_p is provided as $\mathcal{O}(nk^2)$, $\mathcal{O}(d_p k^2)$ and $\mathcal{O}(nd_p^2)$, respectively. When updating $\mathbf{H}_p^{(v)}$, it costs $\mathcal{O}(d_v d_p n)$ to execute matrix multiplication to obtain the optimal $\mathbf{H}_p^{(v)}$. Updating α and β takes $\mathcal{O}(d_v d_p n)$ and $\mathcal{O}(k d_p n)$ to execute matrix multiplication, then costs $\mathcal{O}(m)$ to calculate them, separately. Thus, at each iteration, the time complexity of AWMVC is $\mathcal{O}\left(nk^2 + \sum_{p=1}^m (d_p k^2 + nd_p^2) + \sum_{v=1}^V \sum_{p=1}^m d_v d_p n\right)$, which is linear to the data number n . In addition, our proposed method is free of hyperparameters. Therefore, it is suitable for large-scale data.

Convergence The objective value in Eq. (7) decreases monotonically when one variable updates with the others fixed. Therefore, to prove the convergence of AWMVC, we merely need to show that the formulation has a lower bound.

The proven process is provided as follows: By Cauchy-Schwartz inequality, we have

$$\text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right) \leq \|\mathbf{Z}_p^\top\|_F \|\mathbf{W}_p\|_F \|\mathbf{M}\|_F. \quad (22)$$

Considering that β satisfies $\sum_{p=1}^m \beta_p^2 = 1$ and $\beta \geq \mathbf{0}$, it is easy to conclude that $\beta \leq \mathbf{1}$. Based on the above analysis, we can obtain

$$\begin{aligned} \sum_{p=1}^m \beta_p \text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right) &\leq \sum_{p=1}^m \text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right) \\ &\leq \sum_{p=1}^m \|\mathbf{Z}_p^\top\|_F \|\mathbf{W}_p\|_F \|\mathbf{M}\|_F = \sum_{p=1}^m k \sqrt{d_p}. \end{aligned} \quad (23)$$

Therefore, there is a lower bound for the formula of Eq. (7), with a specific value as follows

$$\begin{aligned} &\sum_{p=1}^m \sum_{v=1}^V \frac{1}{2} \alpha_p^2 \left\| \mathbf{X}^{(v)} - \mathbf{H}_p^{(v)} \mathbf{Z}_p \right\|_F^2 - \\ &\sum_{p=1}^m \beta_p \text{Tr} \left(\mathbf{Z}_p^\top \mathbf{W}_p \mathbf{M} \right) \geq - \sum_{p=1}^m k \sqrt{d_p}. \end{aligned} \quad (24)$$

In our alternating optimization process, the objective value monotonically decreases with each iteration. As a consequence, the algorithm is convergent in theoretic. Furthermore, we will verify the convergence of AWMVC in our experiment.

Extension Our proposed method gives the inspiration to map data features into different latent embedding under specific dimensions and fuse the embedding efficiently. This way, the consensus matrix extracts more complementary and comprehensive information than a fixed dimension. In addition, the model is free of hyperparameters and fitting to handle large-scale datasets. Also, this idea can easily extend to various fields, including multi-view clustering.

Experimental Results

In this section, we conduct comprehensive experiments on various large-scale benchmark datasets to verify the excellent performance of AWMVC, involving clustering performance comparison, convergence and evolution, and running time comparison. Moreover, we analyze dimension weights and conduct an ablation study to evaluate the validity of our proposed algorithm.

Experiment Settings

Seven benchmark datasets are adopted to verify the promising performance of AWMVC, and the maximum number of samples used is more than 100,000. The datasets used in our experiment include Flower17¹, AwA², Caltech256³,

¹<https://www.robots.ox.ac.uk/vgg/data/flowers/>

²<https://cvml.ist.ac.at/AwA/>

³<https://www.kaggle.com/datasets/jessicali9530/caltech256/>

Datasets	Samples	Views	Clusters
Flower17	1360	7	17
AwA	30475	6	50
Caltech256	30607	4	257
MNIST	60000	3	10
VGGFace2	72283	4	200
TinyImageNet	100000	4	200
YouTubeFace50	126054	4	50

Table 1: Datasets used in our experiments.

MNIST⁴, VGGFace2⁵, TinyImageNet⁶, YouTubeFace50⁷. The detailed information of each dataset is listed in Table 1.

Our proposed algorithm is compared with eight state-of-the-art methods, and the comparative algorithms are summarized as follows

1. **Parameter-free auto-weighted multiple graph learning (AMGL)** (Nie, Li, and Li 2016). The algorithm automatically learns an optimal weight for each graph and obtains an optimal global result.
2. **Unified One-step Multi-view Spectral Clustering (UOMVSC)** (Tang et al. 2022). This work integrates spectral embedding and k -means into a unified framework.
3. **Multi-view clustering via joint nonnegative matrix factorization (MNMF)** (Liu et al. 2013). MNMF proposes a matrix-factorization framework to push clustering results of each view into a consensus one.
4. **Scalable Multi-view Subspace Clustering with Unified Anchors (SMVSC)** (Sun et al. 2021). SMVSC obtains consensus anchor points to get more discriminative information and puts anchor learning and subspace learning into a unified optimization framework.
5. **Binary Multi-View Clustering (BMVC)** (Zhang et al. 2019). This paper puts discrete representation learning and binary clustering structure learning together.
6. **Large-scale Multi-view Subspace Clustering in Linear Time (LMVSC)** (Kang et al. 2020). LMVSC integrates anchor graphs and conducts spectral clustering on a smaller graph.
7. **Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance (FPMVSCAG)** (Wang et al. 2022). This paper simultaneously conducts anchor selection and subspace graph construction and proposes a parameter-free algorithm.
8. **One-pass Multi-view Clustering for Large-scale Data (OPMC)** (Liu et al. 2021b). OPMC removes the non-negativity constraint of NMFVSC and proposes a method to obtain a discrete clustering partition matrix.

⁴<http://yann.lecun.com/exdb/mnist/>

⁵http://www.robots.ox.ac.uk/~vgg/data/vgg_face2/

⁶<http://cs231n.stanford.edu/tiny-imagenet-200.zip>

⁷<http://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset>

The implementations of the above methods are public on the corresponding papers, and we run them without any changes. Considering that all the methods need to conduct k -means to attain final cluster assignments, we run 50 times k -means to eliminate the randomness in initialization. For methods with hyperparameters, we tune them with grid-search recommended in their papers and report the best results. In our experiment, we simply set $m = 3$ and the dimensions range from k to mk . All the experiments are conducted on a desktop computer with Intel(R) Core(TM) i9-10850K CPU and 96G RAM.

Experiment Results

We compare AWMVC with eight algorithms on seven widely used benchmark datasets based on four clustering metrics, including accuracy (ACC), normalized mutual information (NMI), purity, and Fscore. The results are shown in Table 2. It is worth noting that the best outcome is marked in bold, and the second best is underlined. In addition, '-' indicates that the algorithm fails to run smoothly due to an out-of-memory error. From the table, it is observed that

1. The proposed method consistently exceeds the competitors on Flower17, Caltech256, VGGFace2, and YouTubeFace50. On other datasets, AWMVC also demonstrates comparable results. For instance, AWMVC outperforms the second best algorithm on all datasets by 21.67%, 1.73%, 13.84%, 0.01%, 55.45%, 0.39%, and 3.86% in terms of ACC, respectively. The improvements on others clustering indicators are similar. The superior clustering performance verifies the superiority of AWMVC.
2. Most large-scale algorithms fail to handle regular datasets, and AWMVC overcomes this drawback. For instance, anchor-based methods always utilize anchor points to represent the data cloud. However, cause of the information loss between anchors and the data cloud, this method is hard to work well on regular datasets. AWMVC is suitable for diverse applications. Our approach attains promising results for routine datasets like Flower17 and Caltech256, and on large-scale scenarios such as YouTubeFace50, the results are the same excellent.
3. Compared with SMVSC, BMVC, and LMVSC, which suffer from hyperparameters, AWMVC achieves better results. In practical applications, these methods ought to tune the hyperparameters many times to choose the best one, resulting in huge time and space resource consumption. In contrast, our proposed method is parameter-free and more suitable for practical situations.

Convergence and Evolution

As proved above, our proposed algorithm is convergent in theoretic. To verify it in practice and discuss the convergence rate, we plot the objective value of AWMVC changes with iterations on AwA and VGGFace2, shown in Figure 1. From the figure, it is seen that the objective value decreases monotonously and converges in less than ten iterations. In

Datasets	AMGL	UOMVSC	MNMF	SMVSC	BMVC	LMVSC	FPMVS-CAG	OPMC	AWMVC
Hyperparameter number	0	1	1	1	4	1	0	0	0
ACC									
Flower17	9.70	36.32	35.81	27.94	26.99	40.37	26.03	32.13	49.12
AwA	-	-	7.72	9.19	8.67	7.97	9.11	9.26	9.42
Caltech256	-	-	2.71	9.67	8.63	9.87	9.49	11.20	12.75
MNIST	-	-	98.06	98.75	45.95	98.57	98.84	98.58	98.85
VGGFace2	-	-	0.90	3.08	3.99	4.40	3.15	3.88	6.84
TinyImageNet	-	-	0.50	3.03	4.09	4.39	2.93	5.15	5.17
YouTubeFace50	-	-	4.82	67.79	66.00	72.73	66.31	69.36	75.54
NMI									
Flower17	10.25	34.95	35.25	25.12	25.62	38.29	26.04	29.69	49.78
AwA	-	-	9.65	10.68	11.95	9.32	10.84	12.22	11.31
Caltech256	-	-	1.96	24.42	31.83	32.14	22.03	32.72	34.61
MNIST	-	-	94.74	96.27	39.59	95.93	96.51	95.85	96.47
VGGFace2	-	-	0.61	10.35	15.04	14.04	9.57	13.29	18.59
TinyImageNet	-	-	0.41	10.64	13.75	13.32	10.16	16.13	14.54
YouTubeFace50	-	-	0.08	82.76	81.90	83.98	83.51	82.36	85.97
Purity									
Flower17	10.76	37.94	37.06	29.26	29.41	41.69	27.43	33.60	51.18
AwA	-	-	9.99	10.02	10.94	10.25	9.69	11.19	11.51
Caltech256	-	-	3.54	11.42	14.94	16.52	11.07	16.93	18.83
MNIST	-	-	98.06	98.75	47.66	98.57	98.84	98.58	98.85
VGGFace2	-	-	1.17	3.15	4.66	5.05	3.22	4.43	7.57
TinyImageNet	-	-	0.70	3.18	4.69	5.05	2.96	5.88	5.77
YouTubeFace50	-	-	4.85	71.09	73.64	78.24	69.32	72.53	79.31
Fscore									
Flower17	11.49	19.69	23.00	16.00	16.61	26.12	16.65	12.85	34.60
AwA	-	-	4.15	6.15	4.32	3.99	6.40	2.47	4.56
Caltech256	-	-	1.18	5.09	6.25	6.54	5.65	9.26	11.03
MNIST	-	-	96.19	97.51	33.57	97.19	97.68	96.85	97.71
VGGFace2	-	-	1.06	1.46	1.46	1.43	1.46	0.72	2.43
TinyImageNet	-	-	0.99	1.53	1.55	1.38	1.75	1.30	1.70
YouTubeFace50	-	-	4.39	60.74	57.09	67.14	61.99	62.06	70.49

Table 2: Empirical evaluation and comparison of AWMVC with eight baseline methods on 7 benchmark datasets in terms of clustering accuracy (ACC), normalized mutual information (NMI), Purity, and Fscore.

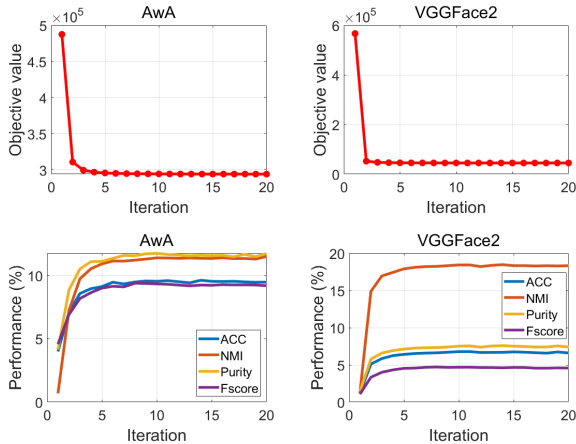


Figure 1: The objective values of AWMVC vary with iterations (above) and the evolution of the consensus coefficient matrix M (below). The results on other datasets are similar and omitted due to space limitations.

addition, to investigate the evolution of the consensus coefficient matrix of AWMVC, in each iteration, we take M as the input to conduct k-means on the same datasets. The results are also shown in Figure 1. From the figure, it is obtained that the clustering performance gradually gets better and then stays stable, which shows the effectiveness of our proposed method.

Dimension Weights

We plot the optimized weights for different dimensions on seven datasets in Figure 2. Considering that the embedding with higher dimensions contains richer information, their final weights ought to be larger, and the results on all datasets are as expected in our experiments. Consequently, our model can effectually integrate diverse and complementary information among views in a discriminatory way.

Running Time Comparison

To compare the computational efficiency of the proposed algorithm, we report the execution time of AWMVC and the compared methods on seven benchmark datasets in Table 3. From the table, it can be seen that:

Datasets	AMGL	UOMVSC	MNMF	SMVSC	BMVC	LMVSC	FPMVS-CAG	OPMC	AWMVC
Flower17	378.82	192.57	176.95	77.38	1.72	2.55	100.89	45.48	42.76
AwA	-	-	3532.20	1061.88	67.87	206.83	2834.32	693.25	434.70
Caltech256	-	-	1435.04	1716.02	53.09	797.29	2588.30	212.99	796.32
MNIST	-	-	367.46	604.59	13.85	342.26	607.07	29.07	562.31
VGGFace2	-	-	2526.49	2378.45	74.53	1631.52	4058.34	499.50	1733.50
TinyImageNet	-	-	6464.19	7992.69	138.87	2436.29	5870.46	972.72	3165.95
YouTubeFace50	-	-	2517.03	1818.27	98.71	1801.96	2690.21	101.92	4160.96

Table 3: Running time comparison with state-of-the-art algorithms on seven benchmark datasets.

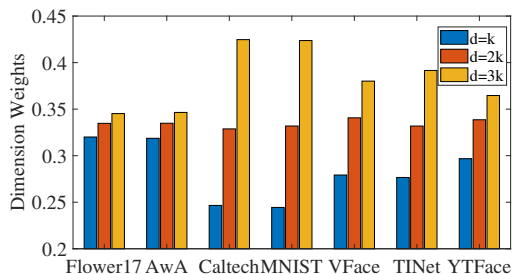


Figure 2: Dimension weights of AWMVC on seven benchmark datasets. From left to right, the dimensions are k , $2k$ and $3k$, respectively.

1. Compared with graph-based methods such as AMGL and UOMVSC, AWMVC is time-consuming and can handle large-scale situations. Also, AWMVC shows superior clustering performance, demonstrating its efficiency.
2. In comparison to MF-based methods like MNMF and OPMC, AWMVC is parameter-free. MNMF undergoes a hyperparameter and needs to run the implementation multiple times to choose a 'suitable' value. Although OPMC is parameter-free, the initialization significantly impacts the final result. Therefore, the algorithm has to be repeated many times to attain a minor loss. In contrast, we merely perform AWMVC once to get the clustering result. Thus, our proposed method is more time-economic in total.
3. The difference in running time between multi-view subspace clustering and AWMVC is slight. For instance, compared to SMVSC and FPMVS-CAG, AWMVC consumes less time on almost all datasets except YouTubeFace50. Furthermore, AWMVC attains better cluster results. Therefore, the computation cost of AWMVC is more worthwhile.

Ablation Study

Our proposed algorithm factorizes data matrices of multiple views into coefficient matrices under diverse dimensions, then auto-weights them to learn a consensus matrix. To evaluate the validity of our model, we formulate two algorithms for comparison, including AWMVC-P and AWMVC- α . To investigate the availability of mapping data matrices into diverse dimensions, we develop the AWMVC-P method, whose embedding number equals 1 and the corresponding

Datasets	AWMVC-P	AWMVC- α	AWMVC
Flower17	46.53	48.03	49.12
AwA	9.44	9.32	9.42
Caltech256	11.68	12.38	12.75
MNIST	96.84	98.15	98.85
VGGFace2	6.53	6.39	6.84
TinyImageNet	5.02	5.05	5.17
YouTubeFace50	73.71	72.63	75.54

Table 4: The ablation study of our proposed method on seven benchmark datasets in terms of ACC. The best results are marked in bold.

dimension $d = k$. For AWMVC- α method, we allocate each embedding with the same weights to research the influence of the discriminatory factor α on AWMVC. The results of our ablation study are reported in Table 4. The table shows that the clustering performance descends when one of the critical components of AWMVC is dropped out. Therefore, the success of our work is related to the mapping of data to multiple dimensions and the obtained discriminatory information.

Conclusion

This paper proposes a novel matrix factorization-based method termed auto-weighted multi-view clustering for large-scale data (AWMVC). Distinct from existing large-scale multi-view clustering algorithms, AWMVC maps data matrices into diverse latent spaces and combines the coefficient matrices from different spaces into a consensus one utilizing discriminatory information. To solve the resultant problem, we develop a six-step alternating algorithm with proven convergence, both theoretical and experimental. The linear complexity and parameter-free property enable it to handle large-scale datasets. Comprehensive experiments demonstrate its superior performance, and the ablation study verifies the effectiveness of the critical components of AWMVC. Besides, mapping data matrices into diverse dimensions can be easily used with existing multi-view clustering methods. In future work, we intend to extend it to handle incomplete views.

Acknowledgments

This work was supported by the National Key R&D Program of China 2020AAA0107100 and the National Natural Science Foundation of China (project no. 61872371, 61922088 and 61976196).

References

- Ding, C.; Tao, L.; Wei, P.; and Park, H. 2006. Orthogonal nonnegative matrix tri-factorizations for clustering. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, August 20-23, 2006*.
- Gao, H.; Nie, F.; Li, X.; and Huang, H. 2015. Multi-view Subspace Clustering. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 4238–4246.
- Gao, S.; Yu, Z.; Jin, T.; and Yin, M. 2019. Multi-view low-rank matrix factorization using multiple manifold regularization. *Neurocomputing*, 335: 143–152.
- Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale Multi-view Subspace Clustering in Linear Time. In *AAAI*.
- Khan, G. A.; Hu, J.; Li, T.; Diallo, B.; and Huang, Q. 2019. Weighted Multi-View Data Clustering via Joint Non-Negative Matrix Factorization. In *2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 1159–1165.
- Li, L.; Wang, S.; Liu, X.; Zhu, E.; Shen, L.; Li, K.; and Li, K. 2022a. Local Sample-Weighted Multiple Kernel Clustering With Consensus Discriminative Graph. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.
- Li, X.; Zhang, H.; Wang, R.; and Nie, F. 2022b. Multiview Clustering: A Scalable and Parameter-Free Bipartite Graph Fusion Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 330–344.
- Liang, W.; Zhou, S.; Xiong, J.; Liu, X.; Wang, S.; Zhu, E.; Cai, Z.; and Xu, X. 2020. Multi-View Spectral Clustering with High-Order Optimal Neighborhood Laplacian Matrix. In *IEEE Transactions on Knowledge and Data Engineering (TKDE)*.
- Liu, J.; Liu, X.; Yang, Y.; Guo, X.; Kloft, M.; and He, L. 2021a. Multiview Subspace Clustering via Co-Training Robust Data Representation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.
- Liu, J.; Liu, X.; Yang, Y.; Liu, L.; Wang, S.; Liang, W.; and Shi, J. 2021b. One-pass Multi-view Clustering for Large-scale Data. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12324–12333.
- Liu, J.; Wang, C.; Gao, J.; and Han, J. 2013. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM international conference on data mining*, 252–260. SIAM.
- Liu, J.-W.; Wang, Y.-F.; Lu, R.-K.; and Luo, X.-L. 2020. Multi-View Non-negative Matrix Factorization Discriminant Learning via Cross Entropy Loss. In *2020 Chinese Control And Decision Conference (CCDC)*, 3964–3971.
- Liu, S.; Wang, S.; Zhang, P.; Xu, K.; Liu, X.; Zhang, C.; and Gao, F. 2022a. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7576–7584.
- Liu, X. 2021. Incomplete Multiple Kernel Alignment Maximization for Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1.
- Liu, Y.; Tu, W.; Zhou, S.; Liu, X.; Song, L.; Yang, X.; and Zhu, E. 2022b. Deep Graph Clustering via Dual Correlation Reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7603–7611.
- Liu, Y.; Yang, X.; Zhou, S.; and Liu, X. 2022c. Simple Contrastive Graph Clustering. *arXiv preprint arXiv:2205.07865*.
- Nie, F.; Li, J.; and Li, X. 2016. Parameter-Free Auto-Weighted Multiple Graph Learning: A Framework for Multiview Clustering and Semi-Supervised Classification. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI'16*, 1881–1887. AAAI Press. ISBN 9781577357704.
- Sun, M.; Zhang, P.; Wang, S.; Zhou, S.; Tu, W.; Liu, X.; Zhu, E.; and Wang, C. 2021. Scalable Multi-view Subspace Clustering with Unified Anchors. *Proceedings of the 29th ACM International Conference on Multimedia*.
- Tang, C.; Li, Z.; Wang, J.; Liu, X.; Zhang, W.; and Zhu, E. 2022. Unified One-step Multi-view Spectral Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 1–1.
- Tang, C.; Liu, X.; Zhu, X.; Zhu, E.; Luo, Z.; Wang, L.; and Gao, W. 2020. CGD: Multi-View Clustering via Cross-View Graph Diffusion. In *AAAI Conference on Artificial Intelligence (AAAI)*, 5924–5931.
- Wan, X.; Liu, J.; Liang, W.; Liu, X.; Wen, Y.; and Zhu, E. 2022. Continual Multi-View Clustering. In *Proceedings of the 30th ACM International Conference on Multimedia, MM '22*, 3676–3684. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392037.
- Wang, J.; Tian, F.; Yu, H.; Liu, C. H.; Zhan, K.; and Wang, X. 2018. Diverse Non-Negative Matrix Factorization for Multiview Data Representation. *IEEE Transactions on Cybernetics*, 48(9): 2620–2632.
- Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2022. Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance. *IEEE Transactions on Image Processing*, 31: 556–568.
- Xia, W.; Gao, Q.; Wang, Q.; Gao, X.; Ding, C.; and Tao, D. 2022a. Tensorized Bipartite Graph Learning for Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Xia, W.; Wang, Q.; Gao, Q.; Yang, M.; and Gao, X. 2022b. Self-consistent Contrastive Attributed Graph Clustering with Pseudo-label Prompt. *IEEE Transactions on Multimedia*.
- Xia, W.; Wang, Q.; Gao, Q.; Zhang, X.; and Gao, X. 2022c. Self-Supervised Graph Convolutional Network for Multi-View Clustering. *IEEE Trans. Multim.*, 24: 3182–3192.
- Yang, X.; Liu, Y.; Zhou, S.; Liu, X.; and Zhu, E. 2022. Interpolation-based Correlation Reduction Network for Semi-Supervised Graph Learning. *arXiv preprint arXiv:2206.02796*.
- Zhang, C.; Hu, Q.; Fu, H.; Zhu, P.; and Cao, X. 2017. Latent Multi-view Subspace Clustering. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4333–4341.

Zhang, J.; Li, L.; Wang, S.; Liu, J.; Liu, Y.; Liu, X.; and Zhu, E. 2022a. Multiple Kernel Clustering with Dual Noise Minimization. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, 3440–3450. New York, NY, USA: Association for Computing Machinery. ISBN 9781450392037.

Zhang, T.; Liu, X.; Gong, L.; Wang, S.; Niu, X.; and Shen, L. 2021. Late Fusion Multiple Kernel Clustering with Local Kernel Alignment Maximization. *IEEE Transactions on Multimedia*, 1–1.

Zhang, T.; Liu, X.; Zhu, E.; Zhou, S.; and Dong, Z. 2022b. Efficient Anchor Learning-based Multi-view Clustering—A Late Fusion Method. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3685–3693.

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2019. Binary Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7): 1774–1782.