Lifelong Multi-view Spectral Clustering

Hecheng Cai, Yuze Tan, Shudong Huang* and Jiancheng Lv
College of Computer Science, Sichuan University, Chengdu, China
{caihecheng, yuzetan}@stu.scu.edu.cn, {huangsd, lvjiancheng}@scu.edu.cn

Abstract

In recent years, spectral clustering has become a well-known and effective algorithm in machine learning. However, traditional spectral clustering algorithms are designed for single-view data and fixed task setting. This can become a limitation when dealing with new tasks in a sequence, as it requires accessing previously learned tasks. Hence it leads to high storage consumption, especially for multi-view datasets. In this paper, we address this limitation by introducing a lifelong multi-view clustering framework. Our approach uses view-specific knowledge libraries to capture intra-view knowledge across different tasks. Specifically, we propose two types of libraries: an Orthogonal Basis Library that stores cluster centers in consecutive tasks, and a Feature Embedding Library that embeds feature relations shared among correlated tasks. When a new clustering task is coming, the knowledge is iteratively transferred from libraries to encode the new task, and knowledge libraries are updated according to the online update formulation. Meanwhile, basis libraries of different views are further fused into a consensus library with adaptive weights. Experimental results show that our proposed method outperforms other competitive clustering methods on multi-view datasets by a large margin.

1 Introduction

The classical spectral clustering algorithm was first proposed by Ng et al. (2001), which performs dimensionality reduction by using the spectrum of the similarity matrix constructed from data before clustering. In the past decades, it has been used in many areas, such as web classification [Zhou and Burges, 2007], text mining [Janani and Vijayarani, 2019], image segmentation [Chahhou et al., 2014], speech recognition [Lin et al., 2019] and machine learning [Sun et al., 2020b]. However, most methods are only applicable to single-view data, while the number of sensors grows on the Internet-of-Things, data from different sources or styles are more common and publicly available [Huang et al., 2021]. Another key issue is that they usually rely on a fixed set of tasks. When encountering an online environment with unknown amounts of consecutive tasks, they need to repeatedly access data from previous tasks. In real-world applications, like mobile apps or web servers, this may bring high memory consumption and computing work. In this paper, we focus on applying a lifelong multi-view clustering framework to common spectral clustering, enabling it to overcome the shortcomings mentioned above.

The lifelong machine learning model is trained over a sequence of tasks, which utilizes knowledge from past tasks to help future tasks [Thrun and Mitchell, 1995], meanwhile, alleviates catastrophic forgetting on past tasks. It could be broadly categorized into three categories [De Lange et al., 2019; Masana et al., 2020]: Architecture-based, Regularization-based, and Memory-based methods. In the past decade, It has been successfully adopted into supervised learning [Ruvolo and Eaton, 2013], unsupervised learning [Liu et al., 2016], semi-supervised learning [Mitchell et al., 2018], and reinforcement learning [Ammar et al., 2015]. Inspired by [Sun et al., 2020a], a memory-based method is applicable to solve continuous clustering tasks. A school website clustering problem on text-image web data could be an example. The semantic meaning of teacher web is different from student web, so they should be divided into two clusters. Tasks from different schools can be considered as sequence tasks, the correlation information of teacher or student websites between two schools is similar, so knowledge learned from the past task could be beneficial for future tasks. Although the concept has been proposed for more than 20 years, research in multi-view clustering, a topic in data mining, has not been extensive.

Inspired by the scenario mentioned above, we consider establishing a lifelong multi-view clustering method based on spectral clustering tasks. The problem is how to use the accumulated knowledge to improve performance on future tasks and update knowledge over lifetime. There are two assumptions considered in our paper: 1) Cluster Space Correlation, multiple clustering tasks should have a consistent latent cluster space. For instance, there are two cluster centers (teacher, student) on the website of school A, while school B obviously has the same centers; 2) Feature Embedding Correlation, which sharing between different tasks should be the

*Corresponding author.
In this paper, we propose a lifelong multi-view spectral clustering model, where dogs are in the same cluster and pots are in the other. When a new clustering task $X_v^m$ arrives, the knowledge is iteratively transferred from fusion basis library $D$, view-specific basis libraries $B_v^{m-1}$ and feature embedding libraries $L_v^{m-1}$ to encode the new task.

![Diagram of lifelong multi-view spectral clustering](image)

Figure 1: The demonstration of our multi-view lifelong spectral clustering model, where dogs are in the same cluster and pots are in the other. When a new clustering task $X_v^m$ arrives, the knowledge is iteratively transferred from fusion basis library $D$, view-specific basis libraries $B_v^{m-1}$ and feature embedding libraries $L_v^{m-1}$ to encode the new task.

same. In particular, the feature embedding of the teacher website should be the same, because the semantic meaning of the web (teacher or student) is similar between schools A and B.

In this paper, we propose a lifelong multi-view spectral clustering framework (LMSC) as shown in Figure 1. According to the mentioned two correlations, we use two view-specific knowledge libraries to transfer knowledge among consecutive tasks and alleviate catastrophic forgetting. They are the Orthogonal Basis Library and Feature Embedding Library respectively. The former contains a set of latent orthogonal cluster centers, and each sample of cluster tasks can be effectively assigned to multiple clusters with different weights. The latter can be modeled by introducing bipartite graph co-clustering, which is able to not only discover the shared manifold information among cluster tasks but also maintain the data manifold information of each individual task. When the model encounters a new multi-view clustering task, it first encodes the new task from each view via the knowledge of both two libraries. Meanwhile, libraries are updated and basis libraries of all views are fused into one. An alternating direction strategy is applied for model optimization, and finally obtains the task-specific representation and Multi-view Fusion Library with clustering center. In summary, this paper makes the following contribution:

- We focus on the lifelong clustering paradigm, which learns and transfers knowledge from previous tasks to a new task. Shared knowledge among multiple tasks is effectively mined and stored via two libraries.
- A multi-view model is proposed to learn view-specific Orthogonal Basis Library and Feature Embedding Library, which can simultaneously preserve the latent clustering centers and capture the feature embedding among different tasks, respectively. It also learns a Fusion Basis Library with adaptive weights.
- Various experiments on multi-view datasets certificate the effectiveness and superiority of our method by comparing it with state-of-the-art algorithms.

2 Related Work

The three most relevant topics are multi-task clustering, multi-view clustering, and lifelong learning.

The aim of multi-task clustering (MTC) is to leverage useful information contained in multiple related tasks to help improve the clustering performance of all the tasks [Zhang and Yang, 2021]. Multi-task spectral clustering (MTSC) [Yang et al., 2014] first attempts to apply the multi-task learning paradigm in spectral clustering. It assumes that all related tasks share a low-dimensional representation and use a $\ell_2, p$-norm regularizer to constrain the coherence among all tasks. Self-adapted multi-task clustering (SAMTC) [Zhang et al., 2016] points out that tasks are usually partially related in the real world, and automatically identify and transfer reusable instances among the tasks to avoid a negative transfer. Partially related multi-task clustering (PRMC) [Zhang et al., 2018b] extends SAMTC with a manifold regularized coding, which uses a more stable way to learn the related instances. However, in real applications, task sets are not fixed and the model may encounter new tasks at any time. Multi-task does improve clustering performance but also costs high storage and computation consumption.

With the development of data collection, more and more data are gathered from different sources or styles. Multi-view clustering (MVC) has also attracted increasing attention in recent years [Xu et al., 2013; Zhao et al., 2017; Huang et al., 2019]. It exploits complementary and consensus information across multiple views to improve clustering performance. Co-regularized multi-view spectral clustering (Co-MVC) [Kumar et al., 2011] applies classical spectral clustering framework to multi-view data. By co-regularizing the clustering hypotheses across views, Co-MVC combines multiple kernels (or similarity matrices) for the clustering problem. One-step multi-view spectral clustering (OMSC) [Zhu et al., 2018] outputs the common affinity matrix learned from low-dimensional data as the final clustering result, thus avoiding the negative influence of the two-step processing in classical spectral clustering. For the problem of spectral clustering doesn’t work well with high dimensional data in complex distribution, [Wang et al., 2018] proposes a linear space embedded method called spectral embedded adaptive neighbors clustering (SEANC). It processes the high-dimensional data with embedded representation and obtains clustering results by adaptive neighbors clustering.

Lifelong learning aims to learn new tasks while retaining its performance on the previous tasks. Elastic weight
consolidation (EWC) [Kirkpatrick et al., 2017] calculates the importance of the parameters by the Fisher information matrix and minimizes the change of important parameters when encountering a new task. Variational continual learning (VCL) [Nguyen et al., 2017] reaches the same goal with Kullback-Leibler (KL) divergence. An efficient lifelong learning algorithm (ELLA) [Ruvolo and Eaton, 2013] considers that all related tasks of the consecutive online tasks should share a common basis, each new task can be obtained by transferring knowledge from the basis. Furthermore, [Sun et al., 2018] considers that the order of tasks can affect performance, and the tasks with more unknown/novel information should be selected.

3 Method
This section presents our proposed lifelong multi-view spectral clustering model. We first review the classic single-view spectral clustering algorithm for a fixed task set and then detail the proposed LMSC.

3.1 Revisit of Spectral Clustering
Given a clustering task \( m \) with \( n_m \) samples \( X^m \in \mathbb{R}^{d \times n_m} \), where \( d \) is the dimension of samples. Spectral clustering first calculates the corresponding symmetric similarity matrix \( W^m \in \mathbb{R}^{n_m \times n_m} \) of \( X^m \), where \( w_{ij} \) represents the similarity between each pair of samples (as an element of the similarity matrix \( W \)). Three common ways are used to construct similarity matrix \( W^m \in \mathbb{R}^{n_m \times n_m} \), e.g., \( k \)-nearest-neighborhood (KNN), \( \epsilon \)-nearest-neighborhood or fully connected graph. The KNN used in this paper is defined as follows:

\[
{w}_{ij}^{m} = \begin{cases} 
\exp\left(-\frac{\|x_i^m - x_j^m\|^2}{2\sigma^2}\right), & \text{if } x_i^m \in \mathcal{N}\left(x_j^m\right) \\
0, & \text{otherwise,}
\end{cases}
\]

where \( \mathcal{N}(\cdot) \) is the function to search \( k \)-nearest neighbors, and \( \sigma \) controls the spread of the neighbors. Then apply the normalized Laplacian on \( W \):

\[
K^m = \left(D^m\right)^{-\frac{1}{2}}L^m\left(D^m\right)^{-\frac{1}{2}} = \left(I - \left(D^m\right)^{-\frac{1}{2}}W^m\left(D^m\right)^{-\frac{1}{2}}\right),
\]

where \( D^m \) is the diagonal matrix of \( W^m \).

After all, we can express the final formulation of spectral clustering with normalized cut [Shi and Malik, 2000]:

\[
\max_{F^m} \text{tr} \left( \left(F^m\right)^\top K^m F^m \right), \quad \text{s.t.,} \left(F^m\right)^\top F^m = I_k.
\]

3.2 Problem Statement
Assume that there is a set of \( M \) multi-view clustering tasks \( T^1, \ldots, T^M \). Each task \( T^m \) with \( V \) views contains \( n_m \) data samples \( X_v^m \in \mathbb{R}^{d_v \times n_m}, v = 1, \ldots, V \), with the dimension of features of \( v \)-th view as \( d_v \). Different from multi-task spectral clustering learn the correlations among all tasks, a lifelong system considers learning new tasks without access to the previously learned data. The model faces a series of consecutive clustering tasks \( T^1, \ldots, T^M \). When a new task \( T^m \) is coming, it arbitrarily and efficiently obtains corresponding cluster assignment matrix \( E_v^m \) of different views and adaptively integrates them into one. In the setting of lifelong clustering, the key issue is how to use the knowledge from each learned task \( T^1, \ldots, T^{(m-1)} \) to help the future task \( T^m \).

3.3 Proposed Model
In this section, we introduce the proposed LMSC model with three parts, i.e., orthogonal basis library, Feature Embedding Library and multi-view fusion basis library.

Orthogonal Basis Library. [Lin et al., 2021] proposes a simple but effective method to store the previously accumulated experiences. An orthogonal basis clustering is applied to uncover the latent cluster centers. Specifically, the assignment matrix \( F^m \) is decomposed into two submatrices: a basis matrix \( B \in \mathbb{R}^{k \times k} \) and a task-specific representation \( E_v^m \in \mathbb{R}^{n_m \times k} \), as \( F^m = E_v^m B \). Then the multi-task spectral clustering model for \( v \)-th view of \( M \) tasks can be represented as:

\[
\max_{\{E_v^m\}_{v=1}^{V}} \frac{1}{M} \sum_{m=1}^{M} \text{tr} \left( \left(E_v^m B_v\right)^\top K_v^m E_v^m B_v \right),
\]

\[
\text{s.t., } B_v^\top B_v = I_k, \left(E_v^m\right)^\top E_v^m = I_k, \forall m = 1, \ldots, M,
\]

and \( B_v \) and \( E_v^m \) are view-specific Basis Library and \( m \) task representation, respectively.

Feature Embedding Library. Besides latent cluster center transfer across consecutive tasks, there is also common feature embedding shared among multiple tasks. [Jiang and Chung, 2012] achieved knowledge transfer between two tasks based on graph-based co-clustering. Inspired by that, with an invariant feature embedding library \( L \in \mathbb{R}^{d \times k} \) with group sparse constraint, we have graph co-clustering term for \( v \)-th view:

\[
\max_{L_v} \frac{1}{M} \sum_{m=1}^{M} \text{tr} \left( L_v^\top \hat{X}_v^m E_v^m B_v \right) + \mu \|L_v\|_{2,1}, \text{s.t., } L_v^\top L_v = I_k,
\]

where \( \hat{X}_v^m \) is defined as

\[
\hat{X}_v^m = \left(D^m\right)^{-\frac{1}{2}} X_v^m \left(D_v^m\right)^{-\frac{1}{2}},
\]

where \( D^m = \text{diag}(X^m)1 \) and \( D_v^m = \text{diag}(\left(X_v^m\right)^\top 1) \). With the sharing Embedding Library, learned tasks can facilitate the discovery of the embedding in new tasks, and the feature embedding can be transferred with each task [Argyriou et al., 2008].

Multi-view Fusion Basis Library. View-specific Orthogonal Basis Library can be learned by Eq. (4). For task with \( V \) views, we learned \( V \) basis libraries. Because the clustering centers on different views should be consistent, each library should be helpful for clustering results. We fuse these libraries to learn a consensus one with adaptive weights [Nie et al., 2018]

\[
\max_{\{B_v\}_{v=1}^{V}} \sum_{v=1}^{V} \left\{ \alpha_v \|D - B_v\|_F^2 \right\},
\]

\[
\text{s.t., } B_v^\top B_v = I_k, \|D\| = I_k.
\]
Algorithm 1 Lifelong multi-view spectral clustering (LSMC) model

Input: multi-view clustering task sets: \( \{ X^1_v, \ldots, X^M_v \}_{v=1}^V \),
view-specific library: \( \{ B_v \leftarrow 0_{d \times k}, L_v \leftarrow 0_{d \times k} \}_{v=1}^V \),
fusion library \( D, \mu \geq 0, \lambda \geq 0, \beta \geq 0 \),
statistical records: \( \{ (M_v)_0 \leftarrow 0_{d \times k}, (C_v)_0 \leftarrow 0_{d \times k} \} \) \( v=1 \)

Parameter: \( \lambda, \mu \) and \( \beta \)

Output: \( B_v, L_v, E_v \) and \( D \)

1: new \( m \)-th task \( \{ X^m_v \}_{v=1}^V \)
2: compute matrices \( \{ K^i_v, \tilde{X}^i_v \} \), \( v=1 \)
3: while Not converge do
4: update \( E^m_v \) by solving Eq.12;
5: update \( B_v \) by solving Eq.16;
6: update \( L_v \) by solving Eq.18;
7: update \( D \) by solving Eq.20;
8: update \( \alpha_v \) by solving Eq.21;
9: compute task representation \( E^m_v = \frac{1}{V} \sum_{v=1}^V E^m_v \);
10: compute assignment matrices via \( E^m_v = E^m_v D \);
11: execute K-means to obtain indicator matrices;
12: end while
13: return solution

By combining Eq. (4), Eq. (5) and Eq. (7), the objective function is formally formulated as follows:

\[
\max_{\{ B_v, L_v \}_{v=1}^V, \{ E^m_v \}_{m=1}^M} \sum_{v=1}^V \left\{ \frac{1}{M} \sum_{m=1}^M \right. \\
\left. \tr \left( (E^m_v B_v)^\top E^m_v B_v \right) + \lambda \tr \left( L_v^\top \tilde{X}^m_v E^m_v B_v \right) \right\}
+ \mu \| L_v \|_{2,1} - \beta \| D - B_v \|_F^2.
\]

Eq. (8) can be converted into:

\[
\max_{B_v, B_v = I_k, B_v^\top B_v = I_k, L_v^\top L_v = I_k, L_v^\top D = I_k_} \sum_{v=1}^V \left\{ \tr \left( (E^m_v B_v)^\top k_v^m E^m_v B_v \right) \right\}
+ \frac{1}{M} \sum_{m=1}^M \tr \left( \frac{1}{M} \sum_{v=1}^V (E^m_v)^\top k_v^m E^m_v \right)
+ \frac{1}{M} \sum_{v=1}^V \tr \left( \frac{1}{M} \sum_{v=1}^V B_v^\top L_v^\top \tilde{X}^m_v E^m_v B_v \right)
- \beta \tr \left( B_v^\top (2I - B_v D^\top - D B_v^\top) B_v \right).
\]

Two statistical variables are constructed to represent the knowledge learned from previous tasks:

\[
(M_v)_m = (M_v)_{m-1} + (E^m_v)^\top K^m_v E^m_v, \\
(C_v)_m = (C_v)_{m-1} + \lambda \tilde{X}^m_v E^m_v.
\]

We also have \( (M_v)_{M-1} = \sum_{m=1}^{M-1} E^m_v K^m E^m_v \), and \( (C_v)_M = \sum_{m=1}^M \lambda \tilde{X}^m_v E^m_v \). Therefore, \( B_v \) in Eq. 14 can be updated by:

\[
B_v = \arg \max_{B_v, B_v = I_k} \tr \left( B_v^\top \left( (M_v)_m/m + B_v L_v^\top (C_v)_m/m \right) B_v \right)
- \beta \tr \left( B_v^\top (2I - B_v D^\top - D B_v^\top) B_v \right) \leftrightarrow \arg \max_{B_v, B_v = I_k} \tr \left( B_v^\top \left( (M_v)_m/m + B_v L_v^\top C_v)_m/m \right) B_v \right)
- \beta \tr \left( B_v^\top (2I - B_v D^\top - D B_v^\top) B_v \right).
\]
Finally, the $B_v$ could be updated by the the eigen-decomposition of $B_v^T (m_m/m + B_vL^T C_m/m - \beta \alpha_v (2I - B_vD^T - DB_v^T)) B_v$.

4.3 Update $L_v$

With fixed $B_v$ and $E_v^m$, the optimization problem for variable $L_v$ on $v$-th view can be denoted as:

$$\max_{L_v} \frac{1}{M} \sum_{m=1}^{M} \lambda \text{tr} \left( L_v^T X_m^v E_m^v B_v \right) + \mu \|L_v\|_2.1 \quad (17)$$

It is equivalent to the following equations:

$$\min_{L_v} \left\| L_v \right\| - \left( \frac{1}{M} \sum_{m=1}^{M} \lambda \hat{X}_m^v E_m^v B_v + \mu \Theta L_v \right) \Leftrightarrow$$

$$\min_{L_v} \left\| L_v \right\| - \left( \frac{1}{M} \sum_{m=1}^{M} \lambda \hat{X}_m^v E_m^v B_v + \mu \Theta^{-1} L_v \right) \Leftrightarrow$$

$$\min_{L_v} \left\| L_v \right\| - \left( C_m B_v + \mu \Theta^{-1} L_v \right) \Leftrightarrow$$

$$\min_{L_v} \left\| L_v \right\| - \left( C_m B_v + \mu \Theta^{-1} L_v \right) \Leftrightarrow$$

where $\Theta_{ii} = \frac{1}{2\|v_i\|^2}$ and $\Theta$ is a diagonal matrix. Eq. (18) can be seen as the projection of $C_m B_v + \mu \Theta^{-1} L_v$ on Stiefel manifold.

4.4 Update D

The part related to $D$ in the objective function is

$$\min_{D} \left\| \sum_{v=1}^{V} \alpha_v \|D - B_v\|^2_F \right\|.$$  \quad (19)

According to the formula of F-norm, Eq. (20) can be converted into:

$$\min_{D} \text{tr}(D^T (\sum_{v=1}^{V} \alpha_v (2I - B_vD^T - DB_v^T)) D) \quad (20)$$

The final solution of $D$ could be obtained by the eigen-decomposition of $\sum_{v=1}^{V} \alpha_v (2I - B_vD^T - DB_v^T)$.

4.5 Update $\alpha_v$

Inspired by [Nie et al., 2017], the adaptive weight of each view $\alpha_v$ could be calculated via the following formulation:

$$\alpha_v = \frac{1}{2 \|D - B_v\|^2_F} \quad (21)$$

5 Experiment

In this section, we evaluate the clustering performance of our LMSC model via a thorough empirical comparisons. Starting with a brief introduction to the benchmark datasets and several SOTA methods we adopted, we demonstrate the clustering results and followed by convergence analysis and parameter analysis of our model.

The experimental environment of the paper is AMD Ryzen 5 2600X, Windows 10 Operating System, 16 GB Main Memory, and the experimental platform is MATLAB R2022b.

5.1 Experiment Setup

Aiming at thoroughly examining the clustering performance of our method, several real-world datasets are utilized in our experiment. All datasets are divided into two, three, or four task groups such that each task contains all clusters.

- **3Sources** comprises of 3 common online news sources i.e., The Guardian, Reuters, and BBC. 169 different news stories are gathered from the agencies.

- **BBC dataset** is composed of news stories in five different labels: politics, entertainment, business, tech and sport. We use 685 samples from 4 sources.

- **BBCSport** contains 544 archives collected from the BBCSport website, where each document is divided into 2 kinds of features.

- **Cornell** dataset is a popular benchmark for multi-view clustering. It has web pages collected from computer science departments of Cornell University and consists of 195 web pages with two different views.

We also choose some single task multi-view clustering models, multi-task clustering models, and lifelong clustering models as powerful competitors, which are

- **SNMF** [Kuang et al., 2012]: a graph clustering framework based on NMF.

- Co-regularized multi-view clustering (Coreg) [Kumar et al., 2011] find clustering results that are consistent across the different views.

- Local Learning-based Multi-task Clustering (LLMC) [Zhong and Pun, 2022]: multi-task clustering with shared low-dimensional subspace information.

- Lifelong Spectral Clustering (L2SC) [Sun et al., 2020a]: single-view lifelong spectral clustering.

- Diversity-induced multi-view subspace clustering [Cao et al., 2015] (DiMSC): explores the enhanced complementarity of multi-view representations:

- Generalized latent multi-view subspace clustering (LRMSC) [Zhang et al., 2018a]: multi-view clustering with latent representation of each view.

- Multiview clustering via adaptively weighted procrustes (AWP) [Nie et al., 2018]: weights each view with its clustering capacities.

- Weighted multi-view spectral clustering (WMSC) [Zong et al., 2018]: employs the spectral perturbation to learn the weight of each view.

5.2 Clustering Results

To achieve fairness, with the authors’ suggested parameter settings, each approach is conducted ten times on every dataset with several tasks. The average values of each task and all tasks are adopted. The task sequences fed into multi-view models are the same as a multi-task model and a lifelong model. Three widely used criteria are utilized: Normalized Mutual Information (NMI), Purity, and Rand Index (RI). The clustering results of our method and competitors are demonstrated in Table 1 to Table 4, the best result is in red and...
the second result is in blue. Note that our method achieves

the best performance on BBCSport and Cornell, and obtains

the best or second best performance on 3Sources and BBC

in most cases. Specifically, as for Cornell, LMSC outper-

forms other competitors on all metrics with respect to differ-

ent tasks. Overall, the average evaluation indicator of all tasks

is good, which showcases the efficiency and superiority of our

LMSC. It is worth noting that multi-view clustering models

only utilize the information in the current task, whereas the

LMSC exploits the knowledge shared among a sequence of

tasks. Compared to multi-task model, LMSC performs bet-

ter, because both the cluster centers and feature embedding

are learned. Hence, LMSC has remarkable advantages over

traditional multi-view clustering approaches in lifelong learn-

ing scenarios.

5.3 Parameter Discussion

To explore the effect on three parameters in Eq. (8), we tune

λ, μ and β within the range $[1e^{-3}, 1e^{-1}, \ldots, 1e^{3}]$. We see

the parameters of LMSC are tuned roughly. Better parameter
tuning would achieve better clustering performance than that recorded in this paper. As shown in Fig. 2, the vertical axis is NMI in [20%, 40%, 60%] and the horizontal axes are $\mu$ and $\beta$ in $[10^{-3}, 10^{-1}, \ldots, 10^3]$, respectively. We find that a high value of $\lambda$ is beneficial to clustering results. From a global perspective, our proposed method is not greatly affected by $\mu$ and $\beta$ in most cases. In detail, the performance demonstrates a bell-shaped curve since it first increases and then decreases as $\mu$ and $\beta$ vary. From Fig. 3, it is generally the same in the case of Cornell. We observe that our method achieves consistently better performance when $\lambda$ is around 1 and the performance is pretty stable under the change of other parameter settings.

5.4 Convergence Analysis

It is worth noticing that the optimization algorithm of our objective function is essentially a non-convex problem. Thus, it is rather critical to validate its convergence property. As shown in Fig. 4, we plot the value of the objective function with respect to each new task on 3Sources. Note that our objective values increase sharply with 200 iterations on both datasets and approach to a boundary. Owing to the limited space, we did not theoretically prove the convergence property of our algorithm, but we still find it converges asymptotically on real-world datasets.

6 Conclusion

In this paper, we introduce a novel lifelong multi-view clustering framework termed lifelong multi-view spectral clustering (LMSC) to deal with tasks involved with multi-view data in a sequence. Specifically, two types of libraries are proposed: 1) orthogonal basis libraries which retain cluster centers for each view, and 2) view-specific feature embedding libraries which embrace feature relationships among tasks in the same sequence. As a new multi-view spectral clustering task arrives, LMSC is able to transfer knowledge embedded in the shared knowledge libraries to encode the coming spectral clustering task and update the libraries with respect to different views. Moreover, an adaptive weighing strategy is utilized to integrate multiple orthogonal basis libraries into a fusion orthogonal basis library. Extensive experiments are conducted to evaluate the superiority of the LMSC: 1) as for clustering results, LMSC outperforms other competitors in the majority of cases. 2) As for parameter analysis, our method is relatively stable with respect to different $\lambda$, $\mu$, and $\beta$. 3) The convergence analysis proves the effectiveness of our optimization algorithm. In the future, we consider using a multi-layer library to capture the nonlinear correlation of each view.

Acknowledgements

This work is supported by the Key Program of National Science Foundation of China (Grant No. 61836006), the National Science Foundation of China under Grant 62106164, and the Sichuan Science and Technology Program under Grants 2021ZDZX0011 and 2022YFG0188.
References


