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 as 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.

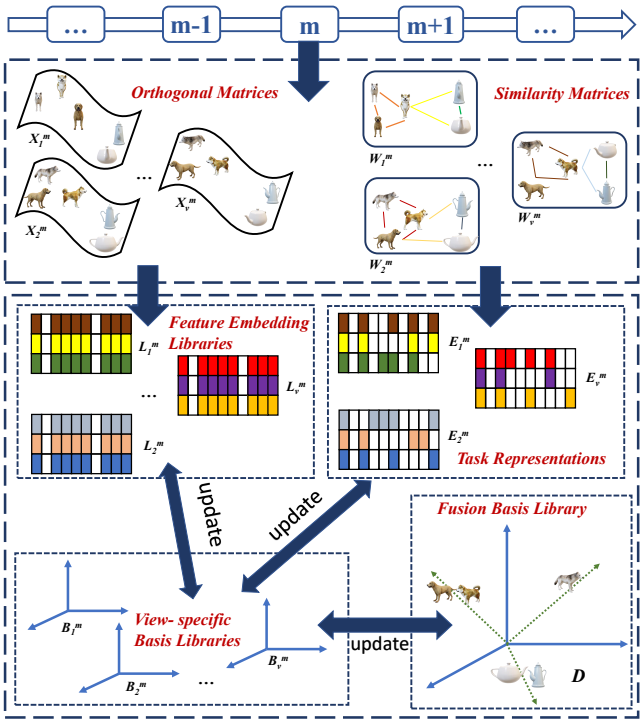


Figure 1: The demonstration of our multi-view lifelong spectral clustering model, where dogs are in the same cluster and pots are in the other. When a new clustering task X_v^m arrives, the knowledge is iteratively transferred from fusion basis library $\mathbf{D}$, view-specific basis libraries $\mathbf{B}_v^{m-1}$ and feature embedding libraries $\mathbf{L}_v^{m-1}$ 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 regualrizer 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). ϵ -nearest-neighborhood or fully connected graph. The KNN used in this paper is defined as follows:

$$w_{ij}^m = \begin{cases} \exp\left(-\frac{\left\|x_i^m - x_j^m\right\|^2}{2\sigma^2}\right), & \text{if} \quad x_i^m \in \mathcal{N}\left(x_j^m\right) \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

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

$$K^m = (D^m)^{-\frac{1}{2}} L^m (D^m)^{-\frac{1}{2}} = I - (D^m)^{-\frac{1}{2}} W^m (D^m)^{-\frac{1}{2}}, \tag{2}$$

where D^m is the diagonal matrix of W^m,$D_{ii}^m = \sum_j W_{ij}^m$. After all, we can express the final formulation of spectral clustering with normalized cut [Shi and Malik, 2000]:

$$\max_{F^m} \operatorname{tr}\left(\left(F^m\right)^\top K^m F^m\right), \text{s.t.,}\ \left(F^m\right)^\top F^m = I_k. \tag{3}$$

F^m is the optimal cluster assignment matrix, which can be calculated via the eigenvalue decomposition of matrix K^m. The final clustering labels of F^m can be achieved by post-processing, e.g., k-means or spectral rotation.

3.2 Problem Statement

Assume that there is a set of M multi-view clustering tasks $\mathcal{T}^1, \ldots, \mathcal{T}^M$. Each task T^m with V views contains n_m data samples $X_v^m \in \mathbb{R}^{d_v\times n_m}, v = 1...V$, 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 $\mathcal{T}^m, \ldots, \mathcal{T}^M$. When a new task $\mathcal{T}^m$ is coming, it arbitrarily and efficiently obtains corresponding cluster assignment matrix F_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 $\mathcal{T}^1, \ldots, \mathcal{T}^{(m-1)}$ to help the future task $\mathcal{T}^m$.

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 $E^m \in \mathbb{R}^{n_m\times k}$, as $F^m = E^m B$. Then the multi-task spectral clustering model for v-th view of M tasks can be represented as:

$$\begin{aligned} &\max_{\{E_v^m\}_{m=1}^M} \frac{1}{M}\sum_{m=1}^{M} \operatorname{tr}\left(\left(E_v^m B_v\right)^\top K_v^m E^m B_v\right), \\ &\text{s.t.,}\ B_v^\top B_v = I_k, \left(E_v^m\right)^\top E_v^m = I_k, \forall m = 1, \ldots, M, \end{aligned} \tag{4}$$

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 a 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:

$$\max_{L_v} \frac{1}{M}\sum_{m=1}^{M} \operatorname{tr}\left(L_v^\top \hat{X}_v^m E_v^m B_v\right) + \mu\|L_v\|_{2,1}, \text{ s.t., } L_v^\top L_v = I_k, \tag{5}$$

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

$$\hat{X}_v^m = (D_1^m)^{-\frac{1}{2}} X_v^m (D_2^m)^{-\frac{1}{2}}, \tag{6}$$

where $D_1^t = \operatorname{diag}(X_v^m \mathbf{1})$ and $D_2^t = \operatorname{diag}\left((X_v^m)^\top \mathbf{1}\right)$. 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]

$$\begin{aligned} &\max_{\{\mathrm{B}_v\}_{v=1}^V, \mathrm{DD}^\top = \mathrm{I}} \sum_{v=1}^{V}\left\{\alpha_v \left\|\mathrm{D} - \mathrm{B}_v\right\|_{\mathrm{F}}^2\right\}, \\ &\text{s.t.,}\ \mathrm{B}_v^\top \mathrm{B}_v = \mathrm{I}_k, \mathrm{D}^\top \mathrm{D} = \mathrm{I}_k. \end{aligned} \tag{7}$$

Algorithm 1 Lifelong multi-view spectral clustering (LSMC) model

Input: multi-view clustering task sets: $\{X_v^1, \ldots, X_v^M\}_{v=1}^V$, view-specific library:$\{B_v \leftarrow \mathbf{0}_{k\times k}, L_v \leftarrow \mathbf{0}_{d_v\times k}\}_{v=1}^V$, fusion library D,$\mu \geq 0, \lambda \geq 0, \beta \geq 0$, statistical records: $\{(M_v)_0 \leftarrow \mathbf{0}_{k\times k}, (C_v)_0 \leftarrow \mathbf{0}_{d\times k}\}_{v=1}^V$
Parameter: λ, μ and β
Output: B_v, L_v, E_v and D

```
new m-th task {X_v^m}_{v=1}^V.
compute matrices {K_v^t, X̂_v^t}_{v=1}^V.
while Not converge do
    update E_v^m by solving Eq.12;
    update B_v by solving Eq.16;
    update L_v by solving Eq.18;
    update D by solving Eq.20;
    update α_v by solving Eq.21;
    compute task representation E^m = (1/V) Σ_{v=1}^V E_v^m;
    compute assignment matrices via F^m = E^m D;
    execute K-means to obtain indicator matrices;
end while
return solution
```

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

$$\begin{aligned}\max_{\{\mathrm{B}_v,\mathrm{L}_v\}_{v=1}^V,\{\mathrm{E}_v^m\}_{m=1}^M,\mathrm{D}\mathrm{D}^\top=\mathrm{I}} \sum_{v=1}^{V}\Bigg\{\frac{1}{M}\sum_{m=1}^{M}\\ \Bigg\{\operatorname{tr}\left(\left(\mathrm{E}_v^m\mathrm{B}_v\right)^\top k_v^m\mathrm{E}_v^m\mathrm{B}_v\right)+\lambda\operatorname{tr}\left(\mathrm{L}_v^\top\widehat{\mathrm{X}}_v^m\mathrm{E}_v^m\mathrm{B}_v\right)\Bigg\}\\ +\mu\left\|\mathrm{L}_v\right\|_{2,1}-\beta\alpha_v\left\|\mathrm{D}-\mathrm{B}_v\right\|_{\mathrm{F}}^2\Bigg\},\\ \text{s.t., } \left(\mathrm{E}_v^m\right)^\top\mathrm{E}_v^m=\mathrm{I}_k,\mathrm{B}_v^\top\mathrm{B}_v=\mathrm{I}_k,\mathrm{L}_v^\top\mathrm{L}_v=\mathrm{I}_k,\mathrm{D}^\top\mathrm{D}=\mathrm{I}_k,\end{aligned} \tag{8}$$

where λ is the trade-off parameter between spectral clustering and co-clustering. If λ equals 0, the function can be reduced to common clustering centers.

4 Optimization

In this section, we introduce the optimization for our method. To reduce computing consumption and memory space, all variables in Eq. (8) should be updated without accessing the previously learned tasks. In the following, the final objective function is non-convex, so an alternating iterative algorithm is given.

4.1 Update E_v^m

When B_v, L_v, D and α_v are fixed, the problem of $\{\mathrm{E}_v^m\}_{v=1}^V$ of m-th task can be express independently for different views as:

$$\begin{aligned}\max_{\mathrm{E}_v^m}\left\{\operatorname{tr}\left(\left(\mathrm{E}_v^m\mathrm{B}_v\right)^\top k_v^m\mathrm{E}_v^m\mathrm{B}_v\right)+\lambda\operatorname{tr}\left(\mathrm{L}_v^m\widehat{\mathrm{X}}_v^m\mathrm{E}_v^m\mathrm{B}_v\right)\right\},\\ \text{s.t., } \left(\mathrm{E}_v^m\right)^\top\mathrm{E}_v^m=\mathrm{I}_k,\end{aligned} \tag{9}$$

where $\left(\mathrm{E}_v^m\right)^\top\mathrm{E}_v^m=\mathrm{I}_k$ is the orthonormality constraint. And E^t can be effectively updated by Stiefel Manifold Theorem [Manton, 2002]:

Theorem 1. *Given a rank* p *matrix* $X \in \mathbb{R}^{n\times k}$ *, and the singular value decomposition of* P *is* $U\Sigma V^\top$*. On the Stiefel manifold, the projection of* P *can be calculated by:*

$$\pi(P)=\arg\min_{Q^\top}\|P-Q\|_F^2. \tag{10}$$

The projection could be expressed as $\pi(P) = UI_{n,k}V^\top$. To maximize the objective function, E^t could be updated by moving it in the direction of increasing the value of Eq. (9):

$$\mathrm{E}_v^m=\pi\left(\mathrm{E}_v^m+\eta_T\nabla g\left(\mathrm{E}_v^m\right)\right), \tag{11}$$

where η_T is step size and $\nabla g\left(\mathrm{E}_v^m\right)$ is the partial derivatives of objective function of E_v^m:

$$\nabla g\left(\mathrm{E}_v^m\right)=\frac{1}{V}\sum_{v=1}^{V}2\left(k_v^m\right)^\top\mathrm{E}_v^m\mathrm{B}_v\mathrm{B}_v^\top+\lambda\left(\hat{\mathrm{X}}_v^m\right)^\top\mathrm{L}_v\mathrm{B}_v^\top. \tag{12}$$

4.2 Update B_v

With other fixed variables, the optimization of B_v can be simplified as:

$$\begin{aligned}\max_{\mathrm{B}_v^\top\mathrm{B}_v=\mathrm{I}_k}\frac{1}{M}\sum_{m=1}^{M}\Bigg\{\operatorname{tr}\left(\left(\mathrm{E}_v^m\mathrm{B}_v\right)^\top k_v^m\mathrm{E}_v^m\mathrm{B}_v\right)\\ +\lambda\operatorname{tr}\left(\mathrm{L}_v^\top\widehat{\mathrm{X}}_v^m\mathrm{E}_v^m\mathrm{B}_v\right)\Bigg\}-\beta\alpha_v\left\|\mathrm{D}-\mathrm{B}_v\right\|_{\mathrm{F}}^2.\end{aligned} \tag{13}$$

Eq. 13 can be converted into:

$$\begin{aligned}\max_{\mathrm{B}_v^\top\mathrm{B}_v=\mathrm{I}_k}\operatorname{tr}\Bigg(\mathrm{B}_v{}^\top\Big(\frac{1}{M}\sum_{m=1}^{M}\left(\mathrm{E}_v^m\right)^\top k_v^m\mathrm{E}_v^m\\ +\frac{1}{M}\sum_{t=1}^{M}\lambda\mathrm{B}_v\mathrm{L}_v^\top\hat{\mathrm{X}}_v^m\mathrm{E}_v^m\Big)\mathrm{B}_v\Bigg)\\ -\beta\alpha_v\operatorname{tr}\left(\mathrm{B}_v^\top(2\mathrm{I}-\mathrm{B}_v\mathrm{D}^\top-\mathrm{D}\mathrm{B}_v^\top)\mathrm{B}_v\right).\end{aligned} \tag{14}$$

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

$$\begin{aligned}(M_v)_m=(M_v)_{m-1}+\left(E_v^m\right)^\top K_v^m E_v^m,\\ (C_v)_m=(C_v)_{m-1}+\lambda\hat{X}_v^m E_v^m.\end{aligned} \tag{15}$$

We also have $(M_v)_{M-1} = \sum_{m=1}^{M-1}\left(E_v^m\right)^\top K_v^m E_v^m$, and $(C_v)_{M-1} = \sum_{m=1}^{M-1}\lambda\hat{X}_v^m E_v^m$. Therefore, B_v in Eq. 14 can be updated by:

$$\begin{aligned}\mathrm{B}_v=\arg\max_{\mathrm{B}_v^\top\mathrm{B}_v=\mathrm{I}_k}\operatorname{tr}\left(\mathrm{B}_v^\top\left((\mathrm{M}_v)_m/m+\mathrm{B}_v\mathrm{L}^\top(\mathrm{C}_v)_m/m\right)\mathrm{B}_v\right)\\ -\beta\alpha_v\operatorname{tr}\left(\mathrm{B}_v^\top(2\mathrm{I}-\mathrm{B}_v\mathrm{D}^\top-\mathrm{D}\mathrm{B}_v^\top)\mathrm{B}_v\right)\Leftrightarrow\\ \arg\max_{\mathrm{B}_v^\top\mathrm{B}_v=\mathrm{I}_k}\operatorname{tr}\Big(\mathrm{B}_v^\top\left((\mathrm{M}_v)_m/m+\mathrm{B}_v\mathrm{L}^\top\mathrm{C}_v)_m/m\right.\\ \left.-\beta\alpha_v(2\mathrm{I}-\mathrm{B}_v\mathrm{D}^\top-\mathrm{D}\mathrm{B}_v^\top)\right)\mathrm{B}_v\Big).\end{aligned} \tag{16}$$

Finally, the B_v could be updated by the the eigen-decomposition of $\mathrm{B}_v^\top\big(m_m/m + \mathrm{B}_v\mathrm{L}^\top\mathrm{C}_m/m - \beta\alpha_v(2\mathrm{I} - \mathrm{B}_v\mathrm{D}^\top - \mathrm{D}\mathrm{B}_v^\top)\big)\mathrm{B}_v$.

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:

$$\max_{\mathrm{L}_v^\top\mathrm{L}_v=\mathrm{I}_k} \frac{1}{M}\sum_{m=1}^{M}\lambda\,\mathrm{tr}\left(\mathrm{L}_v^\top\hat{X}_v^m\mathrm{E}_v^m\mathrm{B}_v\right) + \mu\|\mathrm{L}_v\|_{2,1}. \quad (17)$$

It is equivalent to the following equations:

$$\begin{aligned}&\min_{\mathrm{L}_v^\top\mathrm{L}_v=\mathrm{I}_k} -\mathrm{tr}\left(\mathrm{L}_v^\top\big(\frac{1}{m}\sum_{m=1}^{M}\lambda_t\hat{X}_v^m\mathrm{E}_v^m\big)\mathrm{B}_v + \mu\Theta\mathrm{L}_v\right) \Leftrightarrow\\ &\min_{\mathrm{L}_v^\top\mathrm{L}_v=\mathrm{I}_k}\left\|\mathrm{L}_v - \big(\big(\frac{1}{m}\sum_{m=1}^{M}\lambda\hat{X}_v^m\mathrm{E}_v^m\big)\mathrm{B}_v + \mu\Theta^{-1}\mathrm{L}_v\big)\right\|_F^2, \Leftrightarrow\\ &\min_{\mathrm{L}_v^\top\mathrm{L}_v=\mathrm{I}_k}\left\|\mathrm{L}_v - \left(\mathrm{C}_m\mathrm{B}_v + \mu\Theta^{-1}\mathrm{L}_v\right)\right\|_F^2,\end{aligned} \quad (18)$$

where $\Theta_{ii} = \frac{1}{2\|l_i\|_2}$ and Θ is a diagonal matrix. Eq. (18) can be seen as the projection of $\mathrm{C}_m\mathrm{B}_v + \mu\Theta^{-1}\mathrm{L}_v$ on Stiefel manifold.

4.4 Update D

The part related to D in the objective function is

$$\min_{\mathrm{D}\mathrm{D}^\top=\mathrm{I}}\sum_{v=1}^{V}\left\{\alpha_v\left\|\mathrm{D} - \mathrm{B}_v\right\|_\mathrm{F}^2\right\}. \quad (19)$$

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

$$\min_{\mathrm{D}\mathrm{D}^\top=\mathrm{I}}\mathrm{tr}(\mathrm{D}^\top(\sum_{\mathrm{v}=1}^{V}\alpha_v(2\mathrm{I} - \mathrm{B}_v\mathrm{D}^\top - \mathrm{D}\mathrm{B}_v^\top))\mathrm{D}). \quad (20)$$

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

4.5 Update α_v

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

$$\alpha_v = \frac{1}{2\left\|\mathrm{D} - \mathrm{B}_v\right\|_F}. \quad (21)$$

5 Experiment

In this section, we evaluate the clustering performance of our LMSC model via a throughout 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 clsutering 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

method	task1			task2			task3			task4			avg		
	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI
SNMF	16.83	49.05	64.61	14.58	45.00	64.81	21.06	42.38	70.36	64.12	73.81	83.93	29.15	52.56	70.93
Coreg	65.46	87.86	82.38	62.82	82.86	81.84	67.29	77.86	85.77	51.43	69.76	81.31	61.75	79.58	82.83
LLMC	22.63	50.00	62.37	17.21	47.62	62.37	14.70	45.24	65.16	51.68	59.52	73.52	26.55	50.60	65.85
L2SC	21.23	52.38	66.22	18.49	45.24	66.86	21.72	40.24	70.10	59.01	70.00	80.70	30.11	51.96	70.97
DiMSC	73.93	88.10	87.99	74.15	86.19	88.80	65.51	76.19	83.97	63.21	72.86	84.88	69.20	80.83	86.41
LRMSC	26.40	57.92	66.21	16.55	47.50	63.10	36.04	65.00	71.03	30.18	50.83	65.43	27.29	55.31	66.44
AWP	47.45	73.81	68.41	62.71	80.95	79.56	46.98	64.29	75.84	56.58	64.29	76.77	53.43	70.83	75.15
WMSC	46.11	69.05	71.38	49.81	72.38	73.19	20.38	42.86	44.13	26.19	46.67	52.36	35.62	57.74	60.27
LMSC	67.20	71.42	82.85	68.40	78.57	82.17	70.93	83.33	83.44	72.45	84.33	84.19	79.75	79.41	83.16

Table 1: clustering results on 3Sources.

method	task1			task2			task3			avg		
	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI
SNMF	70.46	85.85	89.30	69.52	84.21	88.40	71.22	86.08	89.13	70.40	85.38	88.94
Coreg	64.01	80.23	85.44	84.20	93.57	95.29	76.29	91.11	93.19	74.83	88.30	91.28
LLMC	21.03	50.29	58.60	21.74	50.29	58.69	21.74	50.29	58.69	21.50	50.29	58.63
L2SC	69.28	83.63	90.38	71.43	83.98	90.80	71.75	84.09	90.88	70.82	83.90	90.66
DiMSC	74.15	86.19	88.80	65.51	76.19	83.97	63.21	72.86	84.88	67.62	78.41	85.86
LRMSC	43.18	60.12	75.22	48.33	70.41	81.41	44.53	69.01	79.60	45.35	66.51	78.74
AWP	30.72	54.39	58.64	25.02	52.63	56.44	35.36	59.06	67.64	30.37	55.36	60.89
WMSC	16.53	44.91	47.98	35.18	56.14	70.15	23.13	52.98	66.53	24.95	51.34	61.54
LMSC	73.61	88.19	90.43	77.38	90.64	92.54	73.24	88.25	90.76	74.74	89.03	91.24

Table 2: clustering results on BBC.

method	task1			task2			task3			avg		
	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI
SNMF	76.49	89.41	91.35	81.66	91.69	93.10	88.07	94.78	95.81	82.07	91.96	93.42
Coreg	84.26	93.38	92.86	73.22	87.13	91.01	76.58	89.71	92.02	78.02	90.07	91.96
LLMC	20.57	51.47	50.04	22.13	52.94	50.47	23.92	52.94	50.63	22.21	52.45	50.38
L2SC	70.62	81.47	88.08	75.25	82.72	88.92	77.83	83.38	90.02	74.57	82.52	89.01
DiMSC	49.87	71.47	76.85	34.45	55.59	70.49	27.48	53.09	64.90	37.27	60.05	70.75
LRMSC	34.54	57.54	72.51	48.33	70.41	81.41	44.53	69.01	79.60	42.47	65.65	77.84
AWP	34.96	63.24	68.77	27.94	57.35	63.37	46.46	69.12	70.92	36.45	63.24	67.69
WMSC	25.35	53.53	64.51	25.27	51.47	68.00	17.52	48.53	56.89	22.71	51.18	63.13
LMSC	87.90	95.32	96.35	83.35	93.57	94.96	81.59	92.40	93.50	84.28	93.76	94.94

Table 3: clustering results on BBCsport.

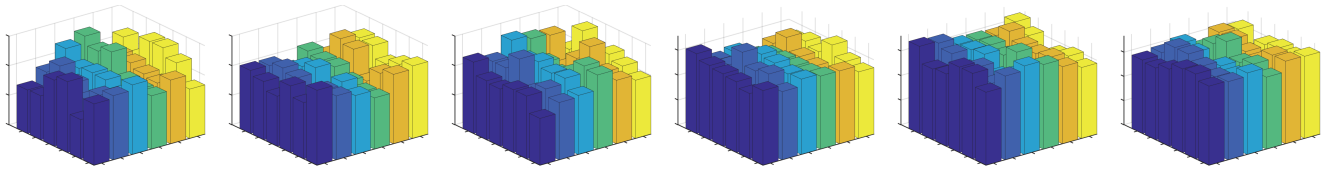
the second result is in blue. Note that our method achieves the best performance on BBCSport and Cornell, and obtains the best or second best performance on 3Sources and BBC in most cases. Specifically, as for Cornell, LMSC outperforms other competitors on all metrics with respect to different tasks. 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 better, because both the cluster centers and feature embedding are learned. Hence, LMSC has remarkable advantages over traditional multi-view clustering approaches in lifelong learning scenarios.

method	task1			task2			avg		
	NMI	Purity	RI	NMI	Purity	RI	NMI	Purity	RI
SNMF	16.15	52.29	64.02	18.82	45.62	66.83	17.48	48.95	65.42
Coreg	28.82	60.42	67.74	22.39	50.42	66.91	25.61	55.42	67.32
LLMC	13.80	54.17	55.59	15.52	47.92	56.65	14.66	51.05	56.12
L2SC	21.43	57.08	66.35	18.31	50.63	68.59	19.87	53.86	67.47
DiMSC	26.40	57.92	66.21	16.55	47.50	63.10	21.48	52.71	64.66
LRMSC	20.97	55.00	63.05	13.57	48.33	42.11	17.27	51.66	52.58
AWP	24.60	56.25	61.70	14.60	43.75	56.38	19.60	50.00	59.04
WMSC	21.96	58.33	61.26	19.76	47.92	64.98	20.86	53.12	63.12
LMSC	32.08	61.67	71.53	27.22	51.67	69.72	29.65	56.67	70.62

Table 4: clustering results on Cornell.

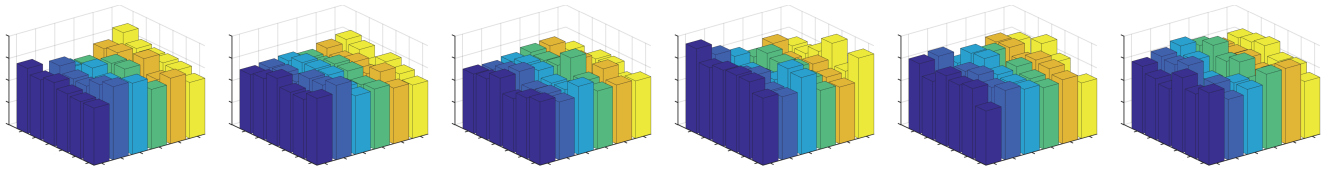
5.3 Parameter Discussion

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



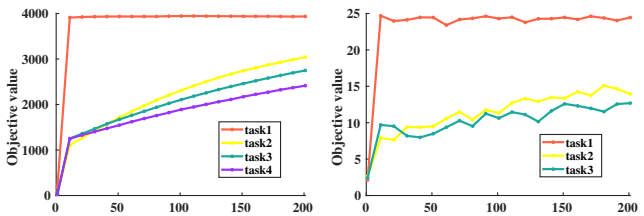
(a) $\lambda = 1e^{-3}$ (b) $\lambda = 1e^{-2}$ (c) $\lambda = 1e^{-1}$ (d) $\lambda = 1$ (e) $\lambda = 1e^{1}$ (f) $\lambda = 1e^{2}$

Figure 2: The influence of parameters λ, μ and β on 3Sources.



(a) $\lambda = 1e^{-3}$ (b) $\lambda = 1e^{-2}$ (c) $\lambda = 1e^{-1}$ (d) $\lambda = 1$ (e) $\lambda = 1e^{1}$ (f) $\lambda = 1e^{2}$

Figure 3: The influence of parameters λ, μ and β on Cornell.



(a) 3Sources (b) BBCSport

Figure 4: Convergence analysis of our proposed LMSC model on (a) 3Sources and (b) BBCSport datasets, where lines with different colors denote different tasks in each dataset.

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 μ and β in $[1e^{-3}, 1e^{-1}, \ldots, 1e^{3}]$, respectively. We find that a high value of λ is beneficial to clustering results. From a global perspective, our proposed method is not greatly affected by μ and β in most cases. In detail, the performance demonstrates a bell-shaped curve since it first increases and then decreases as μ and β vary. From Fig. 3, it is generally the same in the case of Cornell. We observe that our method achieves consistently better performance when λ 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 λ, μ, and β. 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

[Ammar *et al.*, 2015] Haitham Bou Ammar, Rasul Tutunov, and Eric Eaton. Safe policy search for lifelong reinforcement learning with sublinear regret. In *International Conference on Machine Learning*, pages 2361–2369. PMLR, 2015.

[Argyriou *et al.*, 2008] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. Convex multi-task feature learning. *Machine Learning*, 73(3):243–272, 2008.

[Cao *et al.*, 2015] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*, pages 586–594, 2015.

[Chahhou *et al.*, 2014] Mohamed Chahhou, Lahcen Moumoun, Mohamed El Far, and Taoufiq Gadi. Segmentation of 3d meshes usingp-spectral clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1687–1693, 2014.

[De Lange *et al.*, 2019] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. Continual learning: A comparative study on how to defy forgetting in classification tasks. *arXiv preprint arXiv:1909.08383*, 2(6):2, 2019.

[Huang *et al.*, 2019] Shudong Huang, Zhao Kang, Ivor W Tsang, and Zenglin Xu. Auto-weighted multi-view clustering via kernelized graph learning. *Pattern Recognition*, 88:174–184, 2019.

[Huang *et al.*, 2021] Shudong Huang, Ivor W Tsang, Zenglin Xu, and Jiancheng Lv. Measuring diversity in graph learning: a unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5869–5883, 2021.

[Janani and Vijayarani, 2019] R Janani and S Vijayarani. Text document clustering using spectral clustering algorithm with particle swarm optimization. *Expert Systems with Applications*, 134:192–200, 2019.

[Jiang and Chung, 2012] Wenhao Jiang and Fu-lai Chung. Transfer spectral clustering. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 789–803. Springer, 2012.

[Kirkpatrick *et al.*, 2017] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

[Kuang *et al.*, 2012] Da Kuang, Chris Ding, and Haesun Park. Symmetric nonnegative matrix factorization for graph clustering. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 106–117. SIAM, 2012.

[Kumar *et al.*, 2011] Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. *Advances in Neural Information Processing Systems*, 24, 2011.

[Lin *et al.*, 2019] Qingjian Lin, Ruiqing Yin, Ming Li, Hervé Bredin, and Claude Barras. Lstm based similarity measurement with spectral clustering for speaker diarization. *arXiv preprint arXiv:1907.10393*, 2019.

[Lin *et al.*, 2021] Xiaochang Lin, Jiewen Guan, Bilian Chen, and Yifeng Zeng. Unsupervised feature selection via orthogonal basis clustering and local structure preserving. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[Liu *et al.*, 2016] Qian Liu, Bing Liu, Yuanlin Zhang, Doo Soon Kim, and Zhiqiang Gao. Improving opinion aspect extraction using semantic similarity and aspect associations. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[Manton, 2002] Jonathan H Manton. Optimization algorithms exploiting unitary constraints. *IEEE transactions on signal processing*, 50(3):635–650, 2002.

[Masana *et al.*, 2020] Marc Masana, Xialei Liu, Bartlomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost van de Weijer. Class-incremental learning: survey and performance evaluation on image classification. *arXiv preprint arXiv:2010.15277*, 2020.

[Mitchell *et al.*, 2018] Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Bishan Yang, Justin Betteridge, Andrew Carlson, Bhavana Dalvi, Matt Gardner, Bryan Kisiel, et al. Never-ending learning. *Communications of the ACM*, 61(5):103–115, 2018.

[Ng *et al.*, 2001] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 2001.

[Nguyen *et al.*, 2017] Cuong V Nguyen, Yingzhen Li, Thang D Bui, and Richard E Turner. Variational continual learning. *arXiv preprint arXiv:1710.10628*, 2017.

[Nie *et al.*, 2017] Feiping Nie, Guohao Cai, and Xuelong Li. Multi-view clustering and semi-supervised classification with adaptive neighbours. In *Thirty-first AAAI Conference on Artificial Intelligence*, 2017.

[Nie *et al.*, 2018] Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & data mining*, pages 2022–2030, 2018.

[Ruvolo and Eaton, 2013] Paul Ruvolo and Eric Eaton. Ella: An efficient lifelong learning algorithm. In *International Conference on Machine Learning*, pages 507–515. PMLR, 2013.

[Shi and Malik, 2000] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[Sun *et al.*, 2018] Gan Sun, Yang Cong, and Xiaowei Xu. Active lifelong learning with" watchdog". In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Sun *et al.*, 2020a] Gan Sun, Yang Cong, Qianqian Wang, Jun Li, and Yun Fu. Lifelong spectral clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5867–5874, 2020.

[Sun *et al.*, 2020b] Gan Sun, Yang Cong, Qianqian Wang, Bineng Zhong, and Yun Fu. Representative task self-selection for flexible clustered lifelong learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[Thrun and Mitchell, 1995] Sebastian Thrun and Tom M Mitchell. Lifelong robot learning. *Robotics and Autonomous Systems*, 15(1-2):25–46, 1995.

[Wang *et al.*, 2018] Qi Wang, Zequn Qin, Feiping Nie, and Xuelong Li. Spectral embedded adaptive neighbors clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 30(4):1265–1271, 2018.

[Xu *et al.*, 2013] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

[Yang *et al.*, 2014] Yang Yang, Zhigang Ma, Yi Yang, Feiping Nie, and Heng Tao Shen. Multitask spectral clustering by exploring intertask correlation. *IEEE Transactions on Cybernetics*, 45(5):1083–1094, 2014.

[Zhang and Yang, 2021] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[Zhang *et al.*, 2016] Xianchao Zhang, Xiaotong Zhang, and Han Liu. Self-adapted multi-task clustering. In *IJCAI*, pages 2357–2363, 2016.

[Zhang *et al.*, 2018a] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. Generalized latent multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):86–99, 2018.

[Zhang *et al.*, 2018b] Xiaotong Zhang, Xianchao Zhang, Han Liu, and Xinyue Liu. Partially related multi-task clustering. *IEEE Transactions on Knowledge and Data Engineering*, 30(12):2367–2380, 2018.

[Zhao *et al.*, 2017] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.

[Zhong and Pun, 2022] Guo Zhong and Chi-Man Pun. Local learning-based multi-task clustering. *Knowledge-Based Systems*, 255:109798, 2022.

[Zhou and Burges, 2007] Dengyong Zhou and Christopher JC Burges. Spectral clustering and transductive learning with multiple views. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1159–1166, 2007.

[Zhu *et al.*, 2018] Xiaofeng Zhu, Shichao Zhang, Wei He, Rongyao Hu, Cong Lei, and Pengfei Zhu. One-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*, 31(10):2022–2034, 2018.

[Zong *et al.*, 2018] Linlin Zong, Xianchao Zhang, Xinyue Liu, and Hong Yu. Weighted multi-view spectral clustering based on spectral perturbation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.