Large-scale Unaligned Multi-View Graph Clustering

Yi Wen
yiwen23-c@my.cityu.edu.hk
City University of Hong Kong
Hong Kong, China

Xiangyu Zhao
xianzhao@cityu.edu.hk
City University of Hong Kong
Hong Kong, China

ABSTRACT

Multi-view clustering (MVC) improves clustering performance by integrating information from multiple views and has attracted significant attention. However, most methods assume fully aligned data, where sample correspondences across views are predefined, which is often unrealistic in real-world scenarios due to issues like data corruption or sensor mismatches, leading to the data-unaligned problem (DUP). Existing approaches addressing DUP have notable limitations: (1) they focus primarily on feature representation while overlooking critical structural information for clustering, and (2) their high computational complexity makes them unsuitable for large-scale data. To address these challenges, we propose Large-scale Unaligned Multi-View Graph Clustering (LUMGC), a novel method that leverages structural information to improve cross-view alignment. Instead of relying on a computationally expensive $n \times n$ full graph, LUMGC employs a more efficient $m \times n$ anchor graph, significantly reducing time complexity and enabling scalability to large datasets. Extensive experiments demonstrate the effectiveness and versatility of our framework in handling unaligned multi-view clustering scenarios.

CCS CONCEPTS

• **Computing methodologies → Cluster analysis**.

KEYWORDS

datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Yi Wen and Xiangyu Zhao. 2018. Large-scale Unaligned Multi-View Graph Clustering. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Multi-View Clustering (MVC) is a powerful unsupervised learning framework designed to handle data represented in multiple views or modalities. In many real-world scenarios, data is collected from diverse sources or measured under different conditions, resulting in multiple representations of the same underlying objects. MVC aims to leverage the complementary and consistent information across these views to achieve better clustering performance than relying on a single view alone. By integrating information from multiple perspectives, MVC enhances the ability to uncover the intrinsic structure of data.

Building upon MVC, Multi-View Graph Clustering (MVGC) adopts graph-based representations to model the relationships among data points across different views. Graph structures are particularly effective for capturing both local and global relationships within the data, making MVGC a prominent approach in multi-view learning. The process of MVGC typically includes two key stages: Individual Graph Construction: For each view, a separate graph is constructed to represent the relationships among data points within that view. These graphs capture the unique structural characteristics of each view; Graph Fusion: The individual graphs are then integrated into a unified consensus graph that reflects the shared structure across all views. This step ensures that the complementary information from different views is effectively combined. For instance, Nie et al. [4] proposed a method to derive an optimal consensus graph by linearly combining the base graphs from multiple views. By utilizing both the distinct and common structural properties of multiple views, MVGC addresses the challenge of integrating heterogeneous data sources. This not only improves clustering accuracy but also provides a more robust representation of the data's underlying relationships, making it highly effective for complex real-world applications.

While MVGC methods have demonstrated strong performance [7], they face significant challenges when dealing with unaligned multi-view data. In such cases, mismatched alignment between views can result in incorrect clustering outcomes [3]. This issue, referred to as the Data-Unaligned Problem (DUP), occurs when items with the same index in different views (e.g., x_i^v and x_i^u) correspond to entirely different objects. DUP often arises because data in different views is collected independently, without alignment information. For instance, in a tutorial video, the visual content might not sync with the narrator's text, creating misalignment between the views.

In addition to alignment issues, MVGC methods are often computationally expensive due to two main factors [6]: (i) Single-View Graph Optimization: Constructing and optimizing graphs for individual views typically requires quadratic space and cubic time to compute pairwise similarities, which is computationally intensive. (ii) The clustering process, which often involves spectral clustering or singular value decomposition (SVD), further adds to the computational burden. As the size of the data grows, these inefficiencies become more pronounced, significantly reducing the scalability and practicality of MVGC methods.

Despite attempts to address these challenges, existing solutions have several limitations: First, many methods prioritize learning feature representations while neglecting the critical role of structural information, which is essential for effective clustering. Second,

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

the heavy reliance on resource-intensive algorithms makes these methods poorly suited for large-scale scenarios, limiting their efficiency and real-world applicability. These shortcomings highlight the need for more efficient and robust MVGC approaches that can handle unaligned data and scale effectively in large datasets.

To address the challenges of unaligned multi-view data and the inefficiencies of existing methods, we propose a novel framework called Large-scale Unaligned Multi-View Graph Clustering (LUMGC). This framework is specifically designed to handle the Data-Unaligned Problem (DUP) and to enhance scalability in large-scale scenarios. LUMGC achieves this by constructing distinct anchor graphs for each view, which serve as a shared representation space to bridge the metric differences across views. By leveraging structural information from multiple perspectives, LUMGC refines cross-view correspondences through the use of a view alignment matrix. These refined correspondences are then used to integrate the individual graphs into a unified consensus graph, capturing relationships across all views.

A key strength of LUMGC is its adaptability—it provides a unified solution for both fully unaligned and partially unaligned multi-view graph clustering tasks. It is compatible with various graph clustering methods and effectively handles scenarios where traditional approaches struggle. In summary, our contributions are as follows:

- LUMGC introduces a fresh perspective on the Data-Unaligned Problem (DUP), eliminating the need for alignment during data collection. This makes it particularly well-suited for real-world data, where cross-view samples may not naturally align.
- Instead of relying on a full $n \times n$ graph (where n is the number of samples), LUMGC employs an $m \times n$ anchor graph (where m is the number of anchors, and $m \ll n$). This dramatically reduces computational complexity, making the framework scalable to large datasets.
- By incorporating structural insights and aligning views through the alignment matrix, LUMGC effectively addresses the mismatched relationships caused by unaligned data.

2 METHODS

2.1 Problem Formulation

To address the challenges of varying feature dimensions and limited sample availability in DUP, we face a critical issue: cross-view samples exist in entirely different metric spaces, making it impossible to directly calculate distances between them. This naturally raises the question: how can we effectively refine correspondences between cross-view samples?

A simple yet indirect approach is to ensure consistency between cross-view samples and their respective graphs. However, this strategy fails to account for variations between views, potentially limiting the model's ability to represent the data accurately. Inspired by prior works [2, 5], we adopt the idea that sample correspondences can be inferred based on the similarity of their associated anchor structures. In doing so, the original problem of sample correspondence is transformed into one of structural alignment.

To solve this, we introduce the alignment matrix $\mathbf{P}_v$, which satisfies the constraint $\mathbf{P}_v^\top \mathbf{P}_v = \mathbf{I}_m$. This matrix enables an efficient solution to the structural alignment problem. By letting $\mathbf{F}$ denote the fused representation, the task of aligning anchor graphs can be formulated as follows.

Furthermore, to reduce computational overhead, we leverage anchors and anchor graphs. Anchors serve as representative points, summarizing the structure of the data in each view. These anchors, together with their corresponding graphs, provide a compact yet meaningful way to align cross-view structures without needing to process the entire dataset. This approach significantly improves efficiency while maintaining the accuracy of structural correspondences.

$$\min_{\mathbf{P}_i,\mathbf{Z}_i} \sum_{i=1}^{v} \|\mathbf{P}_i\mathbf{Z}_i - \mathbf{H}\|_F^2, \mathbf{P}_i^\top \mathbf{P}_i = \mathbf{1}_m. \tag{1}$$

2.2 Proposed Method

In summary, the proposed Large-scale Unaligned Multi-View Graph Clustering(LUMGC) can be represented as follows:

$$\begin{aligned} &\min_{\mathbf{A}_i,\mathbf{Z}_i,\mathbf{H},\mathbf{P}_i} \sum_{i=1}^{v} \underbrace{\|\mathbf{X}_i - \mathbf{A}_i\mathbf{Z}_i\|_F^2}_{\textit{Anchor Learning}} + \underbrace{\lambda\|\mathbf{P}_i\mathbf{Z}_i - \mathbf{H}\|_F^2}_{\textit{Structure Alignment}}, \\ &s.t.\ \mathbf{H}\mathbf{H}^\top = \mathbf{1}_k,\ \mathbf{A}_i^\top\mathbf{A}_i = \mathbf{1}_m, \mathbf{Z}_i \geq 0, \\ &\quad \mathbf{Z}_i^\top \mathbf{1} = \mathbf{1},\ \mathbf{P}_i^\top\mathbf{P}_i = \mathbf{1}_m, \forall i \in \{1,2,\cdots v\}, \end{aligned} \tag{2}$$

where $\mathbf{A}_i \in \mathbb{R}^{d_i \times m}$ denotes learned anchors in the i-the view, $\mathbf{Z}_i \in \mathbb{R}^{m \times n}$ is the anchor graph, $\boldsymbol{\beta} \in \mathbb{R}^{v \times 1}$ contributes to view weights and $\mathbf{H}$ is the fused spectral embedding. After the anchor learning and structure fusion stage, the learned $\mathbf{H}$ is put into k-means to get the clustering labels.

3 DISCUSSIONS

In this section, we analyze the convergence and computational complexity of LUMGC.

3.1 Computation Complexity

The computational complexity is composed of five processes altogether. Suppose that we select m anchors in the i-th view, the first step to update anchor matrices needs $O(\sum_{i=1}^{v} nd_im + d_im^2)$. Then updating anchor graphs $O(\sum_{i=1}^{v} n(d_i + k + m)m)$ and the view alignment matrices $\{\mathbf{P}_i\}_{i=1}^{v}$ needs $O((\sum_{i=1}^{v} d_i)(nk + k^2) = ndk + dk^2)$. The last two steps to obtain consensus embedding $\mathbf{H}$ and the view weights should include $O(ndk + dk^2 + nk(\sum_{i=1}^{v} m))$ and $O(n\sum_{i=1}^{v} d_im)$

3.2 Space Complexity

In the optimization process, the space cost for storing those matrices $\{\mathbf{A}_i\}_{i=1}^{v}, \{\mathbf{Z}_i\}_{i=1}^{v}, \{\mathbf{P}_i\}_{i=1}^{v}, \mathbf{H}$ is $O(\sum_{i=1}^{v}((d_i + n)m + km) + nk)$.

Our method inherits the linear complexity advantage of conventional anchor graph competitors, rendering it time-efficient and resource-conserving for handling large-scale datasets.

3.3 Convergence

With each iteration of our algorithm, the value of the objective function in Eq. (2) progressively decreases and remains bounded below by 0. Thus, according to [1], convergence to a local minimum is theoretically guaranteed.

REFERENCES

[1] James C. Bezdek and Richard J. Hathaway. 2003. Convergence of Alternating Optimization. *Neural, Parallel Sci. Comput.* 11, 4 (2003), 351–368.

[2] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. 2020. Partially view-aligned clustering. *Advances in Neural Information Processing Systems* 33 (2020), 2892–2902.

[3] Jia-Qi Lin, Man-Sheng Chen, Chang-Dong Wang, and Haizhang Zhang. 2022. A Tensor Approach for Uncoupled Multiview Clustering. *IEEE Transactions on Cybernetics* (2022).

[4] Feiping Nie, Jing Li, Xuelong Li, et al. 2016. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification.. In *IJCAI*. 1881–1887.

[5] Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. 2022. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. *Advances in Neural Information Processing Systems* 35 (2022), 5882–5895.

[6] Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao, and En Zhu. 2022. Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance. *IEEE Trans. Image Process.* 31 (2022), 556–568.

[7] Guang-Yu Zhang, Yu-Ren Zhou, Chang-Dong Wang, Dong Huang, and Xiao-Yu He. 2021. Joint representation learning for multi-view subspace clustering. *Expert Systems with Applications* 166 (2021), 113913.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009