Incomplete multi-view data clustering with hidden data mining and fusion techniques

Zhenjiao Liu

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAS011

Thèse de doctorat

Incomplete Multi-view Data Clustering with Hidden Data Mining and Fusion Techniques

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom SudParis

École doctorale n°626 Ecole Doctorale de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Evry, le 18/12/2023, par

ZHENJIAO LIU

Composition du Jury :

Abdelhamid Mellouk Professeur, University of Paris-Est Creteil (UPEC) - France	Prèsident
Abdelhamid Mellouk Professeur, University of Paris-Est Creteil (UPEC) - France	Rapporteur
Lei Wang Professeur,Dalian University of Technology - China	Rapporteur
Luis Sanchez Professeur Associè, University of Cantabria - Spain	Examinateur
Gyu Myoung Lee Professeur, Liverpool John Moores University - UK	Examinateur
Noel Crespi Professeur, IP Paris, Telecom SudParis - France	Directeur de thèse
Praboda Rajapaksha Chercheuse, IP Paris, Telecom SudParis - France	Co-directeur de thèse

Doctor of Philosophy (PhD) Thesis
Institut-Mines Télécom, Télécom SudParis & Institut Polytechnique de Paris (IP Paris)

Specialization

Computer Science

presented by

Zhenjiao Liu

Incomplete Multi-view Data Clustering with Hidden Data Mining and Fusion Techniques

Committee:

Abdelhamid Mellouk	Reviewer	Professor, University of Paris-Est Creteil (UPEC) - France
Lei Wang	Reviewer	Professor, Dalian University of Technology - China
Luis Sanchez	Examiner	Associate Professor, University of Cantabria - Spain
Gyu Myoung Lee	Examiner	Professor, Liverpool John Moores University - UK
Noel Crespi	Advisor	Professor, IP Paris, Telecom SudParis - France
Praboda Rajapaksha	Co-supervisor	Researcher, IP Paris, Telecom SudParis - France

Thèse de Doctorat (PhD) de
Institut-Mines Télécom, Télécom SudParis
et l'Institut Polytechnique de Paris (IP Paris)

Spécialité

Informatique

présentée par

Zhenjiao Liu

Clustering de données multivues incomplétes à l'aide de techniques de mining de données cachées et de fusion

Jury composé de :

Abdelhamid Mellouk	Rapporteur	Professeur, University of Paris-Est Creteil (UPEC) - France
Lei Wang	Rapporteur	Professeur, Dalian University of Technology - China
Luis Sanchez	Examinateur	Professeur Associè, University of Cantabria - Spain
Gyu Myoung Lee	Examinateur	Professeur, Liverpool John Moores University - UK
Noel Crespi	Directeur de thèse	Professeur, IP Paris, Telecom SudParis - France
Praboda Rajapaksha	Co-encadrant	Chercheuse, IP Paris, Telecom SudParis - France

Dedication

To My Family

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my advisor, Professor Noel Crespi. My appreciation for him knows no bounds. I am truly thankful for the opportunity he extended to me to study in Paris. Not only did he provide me with an exceptional learning environment, but he also exposed me to the rich traditions of French culture, broadening my aesthetic perspective and contributing to my comprehensive personal growth. He possesses exceptional charisma and his influence continues to inspire me. He has taught me valuable social skills and how to connect with individuals from around the world. I am profoundly grateful for his mentorship and the impact he has had on shaping my academic and personal journey.

I would like to extend my gratitude to my Co-supervisor, Praboda Rajapaksha. She has not only provided me with invaluable academic guidance but has also shown exceptional care and concern for me in my personal life. During my time in Paris, she has been the person I trust the most. With her presence, it feels as though I have found a pillar of support that has given me both confidence and reassurance in my studies and daily life. I sincerely wish her an even brighter future ahead !

I would like to extend my gratitude to the reviewers of my paper. Your valuable feedback and constructive suggestions have been instrumental in guiding me toward making continuous improvements and refinements to my work. This process has allowed me to identify areas where I can enhance my understanding and contributions. I deeply appreciate the advice provided by the reviewers and I am committed to incorporating their insights into my future academic endeavors.

I would also like to express my gratitude to all members of my DICE research lab. I fondly remember the days when we enjoyed coffee breaks, meals, and barbecues together. I am thankful for the friends who offered me encouragement and assistance during challenging times. For instance, Amir has been incredibly supportive and caring, even though my language skills were limited. He helped me to integrate into the group. Aung, my office mate, has always been willing to assist me when I faced difficulties in my studies, offering encouragement and guidance. Lastly, I would like to thank my fellow Chinese friends, Shaowen Hao and Guanlin Li, who came to Paris to study with me. I appreciate your care, support, and guidance in both my daily life and academic pursuits.

Lastly, I want to extend special thanks to my family. I am grateful to my parents for their unwavering support in my academic pursuits, encouraging me to be true to myself and live a fulfilling life. I also want to thank my younger sister ; despite being much younger than me, her constant spirit of hard work has always been an inspiration to me. I love you all very much.

8

Zhenjiao Liu

1^{st} September 2023

Abstract

Incomplete multi-view data clustering is a research direction that attracts attention in the fields of data mining and machine learning. In practical applications, we often face situations where only part of the modal data can be obtained or there are missing values. Data fusion is an important method for incomplete multi-view information mining. Solving for incomplete multi-view information mining in a targeted manner, achieving flexible collaboration between visible views and shared hidden views, and improving the robustness have become quite challenging. This thesis focuses on three aspects: hidden data mining, collaborative fusion, and enhancing the robustness of clustering. The main contributions are as follows:

1. Hidden data mining for incomplete multi-view data: existing algorithms cannot make full use of the observation of information within and between views, resulting in the loss of a large amount of valuable information, and so we propose a new incomplete multi-view clustering model IMC-NLT (Incomplete Multi-view Clustering Based on NMF and Low-Rank Tensor Fusion) based on non-negative matrix factorization and low-rank tensor fusion. IMC-NLT first uses a low-rank tensor to retain view features with a unified dimension. Using a consistency measure, IMC-NLT captures a consistent representation across multiple views. Finally, IMC-NLT incorporates multiple learning into a unified model such that hidden information can be extracted effectively from incomplete views. We conducted comprehensive experiments on five real-world datasets to validate the performance of IMC-NLT. The overall experimental results demonstrate that the proposed IMC-NLT performs better than several baseline methods, yielding stable and promising results.

2. Collaborative fusion for incomplete multi-view data: our approach to address this issue is Incomplete Multi-view Co-Clustering by Sparse Low-Rank Representation (CCIM-SLR). The algorithm is based on sparse low-rank representation and subspace representation, in which jointly-missing data is filled using data within a modality and related data from other modalities. To improve the stability of clustering results for multi-view data with different missing degrees, CCIM-SLR uses the Γ-norm model, which is an adjustable low-rank representation method. CCIM-SLR can alternate between learning the shared hidden view, visible view, and cluster partitions within a co-learning framework. An iterative algorithm with guaranteed convergence is used to optimize the proposed objective function. Compared with other baseline models, CCIM-SLR achieved the best performance in the comprehensive experiments on the five benchmark datasets, particularly on those with varying degrees of incompleteness.

3. Enhancing the clustering robustness for incomplete multi-view data: we offer a fusion of graph convolution and information bottlenecks (Incomplete Multi-view Representation Learning Through Anchor Graph-based GCN and Information Bottleneck – IMRL-AGI). First, we introduce the information bottleneck theory to filter out the noise data with irrelevant details and retain only the most relevant feature items. Next, we integrate the graph structure information based on anchor points into the local graph information of the state

fused into the shared information representation and the information representation learning process of the local specific view, a process which can balance the robustness of the learned features and improve the robustness. Finally, the model integrates multiple representations with the help of information bottlenecks, reducing the impact of redundant information in the data. Extensive experiments are conducted on several real-world datasets, and the results demonstrate the superiority of IMRL-AGI. Specifically, IMRL-AGI shows significant improvements in clustering and classification accuracy, even in the presence of high view missing rates (e.g. 10.23% and 24.1% respectively on the ORL dataset). Furthermore, experiment results demonstrate the robustness of IMRL-AGI across different view missing rates (e.g. the NMI values ranging from 82.52% to 86.49% on the ORL dataset).

Keywords

Hidden Data Mining, Non-negative matrix factorization, Collaborative Fusion, Low-Rank Tensor, Sparse Low-Rank Representation, Information Bottleneck, Anchor Graph GCN

Résumé

Le regroupement de données multivues incomplètes est un axe de recherche majeur dans le domaines de l'exploration de données et de l'apprentissage automatique. Dans les applications pratiques, nous sommes souvent confrontés à des situations où seule une partie des données modales peut être obtenue ou lorsqu'il y a des valeurs manquantes. La fusion de données est une méthode clef pour l'exploration d'informations multivues incomplètes. Résoudre le problème de l'extraction d'informations multivues incomplètes de manière ciblée, parvenir à une collaboration flexible entre les vues visibles et les vues cachées partagées, et améliorer la robustesse sont des défis. Cette thèse se concentre sur trois aspects : l'exploration de données cachées, la fusion collaborative et l'amélioration de la robustesse du regroupement. Les principales contributions sont les suivantes :

1) Exploration de données cachées pour les données multi-vues incomplètes : les algorithmes existants ne peuvent pas utiliser pleinement l'observation des informations dans et entre les vues, ce qui entraîne la perte d'une grande quantité d'informations. Nous proposons donc un nouveau modèle de regroupement multi-vues incomplet IMC-NLT (Incomplete Multi-view Clustering Based on NMF and Low-Rank Tensor Fusion) basé sur la factorisation de matrices non négatives et la fusion de tenseurs de faible rang. IMC-NLT utilise d'abord un tenseur de faible rang pour conserver les caractéristiques des vues avec une dimension unifiée. En utilisant une mesure de cohérence, IMC-NLT capture une représentation cohérente à travers plusieurs vues. Enfin, IMC-NLT intègre plusieurs apprentissages dans un modèle unifié afin que les informations cachées puissent être extraites efficacement à partir de vues incomplètes. Des expériences sur cinq jeux de données ont validé les performances d'IMC-NLT.

2) Fusion collaborative pour les données multivues incomplètes : notre approche pour résoudre ce problème est le regroupement multivues incomplet par représentation à faible rang. L'algorithme est basé sur une représentation éparse de faible rang et une représentation de sous-espace, dans laquelle les données manquantes sont complétées en utilisant les données d'une modalité et les données connexes d'autres modalités. Pour améliorer la stabilité des résultats de clustering pour des données multi-vues avec différents degrés de manquants, CCIM-SLR utilise le modèle Γ-norm, qui est une méthode de représentation à faible rang ajustable. CCIM-SLR peut alterner entre l'apprentissage de la vue cachée partagée, la vue visible et les partitions de clusters au sein d'un cadre d'apprentissage collaboratif. Un algorithme itératif avec convergence garantie est utilisé pour optimiser la fonction objective proposée.

3) Amélioration de la robustesse du regroupement pour les données multivues incomplètes : nous proposons une fusion de la convolution graphique et des goulots d'étranglement de l'information (apprentissage de la représentation multivues incomplète via le goulot d'étranglement de l'information). Nous introduisons la théorie du goulot d'étranglement de l'information afin de filtrer les données parasites contenant des détails non pertinents et de ne conserver que les éléments les plus pertinents. Nous intégrons les informations sur

la structure du graphe basées sur les points d'ancrage dans les informations sur le graphe local. Le modèle intègre des représentations multiples à l'aide de goulets d'étranglement de l'information, réduisant ainsi l'impact des informations redondantes dans les données. Des expériences approfondies sont menées sur plusieurs ensembles de données du monde réel, et les résultats démontrent la supériorité de IMRL-AGI. Plus précisément, IMRL-AGI montre des améliorations significatives dans la précision du clustering et de la classification, même en présence de taux élevés de données manquantes par vue (par exemple, 10,23 % et 24,1 % respectivement sur l'ensemble de données ORL). De plus, les résultats des expériences démontrent la robustesse de l'IMRL-AGI à travers différents taux de vues manquantes (par exemple, les valeurs de NMI variant de 82,52 % à 86,49 % sur le jeu de données ORL).

Mots-clés

Exploration de données cachées, Factorisation matricielle non négative, Fusion coopérative, Tenseurs de bas rang, Représentation á faible rang parcimonieuse, la théorie de goulot d'étranglement de l'information, réseaux convolutifs de graphes d'ancrage

Table of contents

Chapter 1

Introduction

Contents

1.1 Motivation

In recent times, a significant volume of data has been distributed across various domains, originating from different views, perspectives, or angles. For instance, users on social media share diverse types of content, encompassing text, images, and videos. Similarly, medical imaging equipment can capture data from various views, including X-ray images and tissue density information. In the realm of autonomous driving, sensors collect data from different views, such as cameras, radars, and lidars. These examples illustrate the acquisition of multifaceted data or multi-view data, contributing to providing insights for potential applications [1].

The acquisition of multi-view data is important due to its ability to offer more comprehensive and accurate data from many perspectives. As a result, multi-view data can provide significant advantages for research and applications across diverse domains. Nonetheless, it is essential to note that each set of view data should possess distinct features, formats, and semantics. This heterogeneity introduces challenges to the analysis and processing of data. One possible approach to address this diversity is multi-view clustering. Through the application of multi-view clustering techniques, we can uncover the interrelations among distinct views, thereby gaining insights into the underlying structure of the data. By employing multi-view clustering techniques, we can present the interrelations among distinct views, thus enhancing our understanding of the data's inherent structure. Therefore, we are able to amalgamate data from different views, exploring latent data insights, and consequently, utilizing valuable insights for further research and application scenarios.

In real-world scenarios, multi-view data generation involves various factors such as occurrences of missing values, noise interference, or even manifest incompleteness. This inherent data incompleteness creates a huge challenge for traditional multi-view clustering methods, as these conventional approaches may struggle to efficiently handle the presence of missing or incomplete data. Therefore, the rationale behind introducing incomplete multi-view clustering is to effectively address the issue of missing data in real-world scenarios. Hence, we propose a novel clustering methodology that aims to optimize the utilization of available limited information, effectively uncovering the underlying structural patterns of latent information to utilize in real-world applications.

However, to overcome the challenge of clustering incomplete multi-view data, simple missing value handling approaches are not efficient. One reason is that in real-world applications, some important perspectives are undocumented or inadequately described. These hidden pieces of information may contain vital insights essential for data analysis and comprehension. Therefore, focusing on the exploration of hidden data has emerged as an important motivation within the area of incomplete multi-view clustering research. By effectively

extracting knowledge from these missing sources, we anticipate achieving enhanced precision and comprehensiveness in clustering results. This, in turn, will empower us to gain a deeper understanding of the complex relationships built in multi-view data.

Moreover, the integration of fusion clustering techniques presents another potent solution to address the challenges posed by incomplete multi-view data. Traditional single-view clustering methods pose several limitations when dealing with multi-view data, as they often fail to fully integrate the information across different views. However, the introduction of fusion clustering techniques offers promising research that strengthens data from diverse views into a cohesive representation. Fusion clustering can uncover the interconnections and associations between different views and is capable of overcoming the incompleteness of information among views, thereby facilitating effective clustering of cross-view data. Such fusion approaches can enhance the reliability and interpretability of clustering outcomes.

In summary, incomplete multi-view clustering, as a methodology to address the incompleteness of multi-view data, holds significant research significance and practical application value [2]. By surpassing the limitations of traditional multi-view clustering methods in handling missing data, incomplete multi-view clustering exhibits potent capabilities in real-world multi-view data analysis tasks. Its distinctive approach and feature fusion strategies hold the potential to bring about novel breakthroughs in fields like data mining, knowledge discovery, and pattern recognition. This paper aims to present innovative solutions to the complexities posed by issues such as complex view omissions, high missing rates, and latent information within missing views in incomplete multi-view data. Our objective is to fully exploit incomplete multi-view data and empirically validate the performance and application prospects of these methodologies across diverse domains. Through this research, we aspire to contribute to the advancement of multi-view clustering algorithms, offering valuable insights into both societal and scientific endeavours.

1.2 Objectives of the Thesis

In this section, we outline the primary research objectives of my thesis, each encompassing specific issues to be addressed. This manuscript primarily introduces three distinct solutions aimed at tackling the challenges posed by incomplete multi-view clustering. To achieve these objectives, we conducted the following key research efforts:

- To address the issue of mining latent information from incomplete multi-view data, we propose a method based on Non-Negative Matrix Factorization (NMF) and low-rank tensor decomposition.

- To devise a more flexible collaborative learning model based on visible views and

shared latent views, we introduce a strategy founded on sparse low-rank representation. This enables the construction of an incomplete multi-view joint clustering model.

- In order to enhance the robustness of the incomplete multi-view clustering approach, we explored methods based on the information bottleneck principle and the Graph Convolutional Network (GCN) with an anchor graph

1.3 Contributions of the Thesis

Incomplete multi-view clustering plays a crucial role in big data analysis and mining. However, it faces several challenges, including complex multi-view missing situations, high missing rates, and redundancy issues within view data. Complex modal data missing refers to situations where some samples may exist in certain views but are missing in others, or data in different views may have varying features, dimensions, or quality levels. High missing rates indicate that a significant amount of crucial information for many samples is severely missing across different modalities. High redundancy, on the other hand, signifies that multiple modalities contain a substantial amount of similar or repetitive information, leading to data redundancy and increased complexity in data processing and analysis.

Hence, this research focuses on addressing these three key challenges in multi-view data: complex view data missing, high missing rates, and redundancy problems. To tackle these challenges, a series of algorithms, including incomplete multi-view clustering by NMF and low-rank tensor, incomplete multiview co-clustering by sparse low-rank representation, and incomplete multi-view representation learning via information bottleneck and anchor graph GCN, have been proposed. The specific research contributions are as follows:

C1: To address the challenge of complex missing situations from incomplete multi-view data, our primary contribution resides in the proposal of a method known as incomplete multi-view clustering using Non-negative matrix factorization and a low-rank tensor (IMC-NLT). This method synergies Non-Negative Matrix Factorization (NMF) and low-rank tensor techniques, establishing a comprehensive framework for incomplete multi-view clustering. This framework adeptly navigates through multiple constraints, including lack of valuable latent information within the data, excessive sensitivity to model parameters, and the complexity of effectively handling samples with incomplete views. To provide a more detailed exposition of our research accomplishments, we further dissect our principal contribution into four pivotal sub-contributions. Each sub-contribution encompasses distinctive methods and procedural implementations.

C1.1: We introduce a novel and efficient incomplete multi-view clustering model, named IMC-NLT, designed to handle incomplete view data. To the best of our knowledge, IMC-NLT is the first incomplete multi-view clustering method that integrates the low-dimensional representation generated by a rapid and effective dimensionality reduction technique with a low-rank tensor model.

C1.2: To address missing multi-view data in various scenarios, IMC-NLT employs low-rank constraints and a tensor model constructed from incomplete multi-view data. As a result, it effectively captures the correlations among instances within views and across different views.

C1.3: The proposed approach demonstrates robustness in handling globally extracted consistency information. Particularly, consistency representation learning effectively quantifies the inconsistencies among consistent information acquired from different perspectives. Under this mechanism, IMC-NLT efficiently filters out noisy data, thereby generating accurate multi-view consistency representations.

C1.4: We conducted comprehensive experiments on multi-view benchmark datasets collected from various application domains to evaluate the effectiveness of IMC-NLT. The results indicate that IMC-NLT outperforms baseline methods. Furthermore, IMC-NLT exhibits lower sensitivity to its parameters, showcasing excellent generalization performance in the context of incomplete multi-view clustering.

C2: To design a more flexible collaborative learning model that addresses the high missing rates , we incorporate a strategy based on sparse low-rank representation. This approach is called a co-clustering method for incomplete multiview data by sparse low-rank representation (CCIM-SLR). This method adeptly addresses the intricate task of generating clustering outcomes from multi-view data featuring missing views and varying degrees of missing data points. We further divide our principal contribution into four distinct sections, offering comprehensive elaborations on each segment.

C2.1: This paper presents CCIM-SLR, a novel approach to incomplete multi-view clustering that leverages a low-rank sparse representation matrix to recover data from missing samples. CCIM-SLR utilizes association information between missing samples and observed samples within views, as well as their association information between views.

C2.2: To improve the stability of clustering results for multi-view data with different missing degrees, CCIM-SLR uses the Γ-norm model, which is an adjustable low-rank representation method. Γ-norm shows the accuracy of achieving a low-rank representation and the stability of data recovery.

C2.3: CCIM-SLR learns both a visible view and a hidden view within a co-learning framework in an end-to-end manner, using a mutual interplay between the view data recovery and a clustering process. This approach avoids the need for post-processing steps such as k-means for final clustering assignment results.

C2.4: CCIM-SLR has been validated through both theoretical proofs and experiments. Based on the experimental results, CCIM-SLR outperformed state-of-the-art approaches on the five incomplete multi-view datasets. The robustness of CCIM-SLR has been demonstrated through experiments on incomplete multi-view datasets with different missing rates of data points.

C3: In order to solve the redundancy issues of incomplete multi-view clustering methodologies, we introduce the concepts of information bottleneck and anchor graph Generative Adversarial Networks (GAN). Subsequently, we present an approach for incomplete multi-view representation learning that is grounded in information bottleneck principles and anchor graph Graph Convolutional Network (GCN). This approach adeptly handles all forms of incomplete view data, while maintaining robustness across various levels of data completeness. To provide a lucid exposition of the mechanisms intrinsic to each sub-contribution, we have diligently structured and expounded upon each distinct sub-contribution point.

C3.1: We propose a novel framework called IMRL-AGI for multi-view representation learning in the presence of incomplete data. This framework combines information bottlenecks and anchor graph GCN. To the best of our knowledge, this is the first time that information bottleneck is integrated with anchor graph GCN for addressing incomplete multi-view representation learning.

C3.2: We introduce the maximization of mutual information constraints to enhance the correlation between the view information obtained from the common representation and the view information derived through the anchor graph GCN.

C3.3: IMRL-AGI enhances the robustness and diversity of information bottleneck representations in downstream tasks by minimizing the mutual information between different views derived from the common representation and between the representation of the original view and the common representation.

C3.4: We extensively evaluated the performance of IMRL-AGI on three real-world datasets. Particularly, IMRL-AGI shows significant improvements in clustering and classification accuracy even with high view missing rates (e.g. 10.23% and 24.1% respectively on the ORL dataset). Furthermore, experiment results demonstrate the robustness of

IMRL-AGI under different view missing rates (e.g. the NMI values from 82.52% to 86.49% on ORL dataset).

1.4 Publications List

Journal Papers

- Zhenjiao Liu, Zhikui Chen, Yue Li, Liang Zhao, Tao Yang, Reza Farahbakhsh, Noel Crespi and Xiaodi Huang. "IMC-NLT: Incomplete multi-view clustering by NMF and low-rank tensor." Expert Systems with Applications (2023).
- Xu Yuan, Shaokui Gu, Liang Zhao, and Zhenjiao Liu, "Mining Multi-View Clustering Space with Interpretable Space Search Constraint." IEEE Signal Processing Letters (2023).
- Liang Zhao, Xiao Wang, Zhenjiao Liu, Hong Yuan, Jingyuan Zhao, and Shuang Zhou. "Deep probability multi-view feature learning for data clustering." Expert Systems with Applications (2023).
- Zhikui Chen, Kai Lou, Zhenjiao Liu, Yue Li, Yiming Luo, and Liang Zhao "Joint Long and Short Span Self-Attention Network for Multi-view Classification." Expert Systems with Applications (2023).
- Zhenjiao Liu, Zhikui Chen, Kai Lou, Praboda Rajapaksha, Liang Zhao, and Noel Crespi. "CCIM-SLR: Incomplete Multiview Co-Clustering by Sparse Low-Rank Representation." Multimedia Tools and Applications (2023).

Conference Papers

- Zhenjiao Liu, Xiao Wang, Xiaodi Huang, Guanlin Li, Ke Sun and Zhikui Chen. "Incomplete Multi-view Representation Learning Through Anchor Graph-based GCN and Information Bottleneck." IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE ICASSP, 2024.
- Shaokui Gu, Xu Yuan, Liang Zhao, Zhenjiao Liu, Yan Hu, and Zhikui Chen. "MVCIR-net: Multi-view Clustering Information Reinforcement Network." The 31th ACM International Conference on Multimedia (ACM MM). ACM, 2023.

Under Review

- Liang Zhao, Wang Xiao, and Zhenjiao Liu. "Learnable Graph Guided Deep Multi-view Representation Learning via Information Bottleneck." IEEE Transactions on

Neural Networks and Learning Systems.

Working Paper

- Zhenjiao Liu, Zhikui Chen, Praboda Rajapaksha, and Noel Crespi. "Incomplete Multi-view Clustering based on Dual Weight Mechanism and Cluster Separation." IEEE Transactions on Cybernetics.

1.5 Relationship of Publications with Contributions

In this section, we provide the relationships of publications with contributions.

- The publication 'IMC-NLT: Incomplete multi-view clustering by NMF and low-rank tensor' corresponds to Contributions C1, C1.1, C1.2, C1.3, and C1.4 in Chapter 3.
- The submitted paper 'CCIM-SLR: Incomplete Multi-view Co-Clustering by Sparse Low-Rank Representation' corresponds to Contributions C2, C2.1, C2.2, C2.3, and C2.4 in Chapter 4.
- The submitted paper titled 'Incomplete Multi-view Representation Learning via Information Bottleneck and Anchor Graph GCN' corresponds to Contributions C3, C3.1, C3.2, C3.3, and C3.4 in Chapter 5.

1.6 Outline of the Thesis

The thesis is structured into five chapters.

- Chapter 1 describes the background of research topics, motivation, contributions of this thesis, a summary of each chapter, and the outline of my thesis.
- Chapter 2 presents the background and related technologies relevant to the main topics of this thesis, i.e., multi-view data analysis, data missing and processing, hidden data mining, and cluster fusion techniques.
- Chapter 3 introduces a method based on NMF and low-rank tensor, which encompasses three mechanisms: incomplete multi-view data filling and decomposition, the multi-view collaborative fusion of low-rank tensor, and consensus representation learning. Performance tests were conducted on relevant incomplete multi-view datasets.
- Chapter 4 illustrates the application of sparse low-rank representation and co-clustering for addressing incomplete multi-view clustering challenges, which encompasses two

mechanisms: shared hidden subspace learning based on SRRS and incomplete multi-view co-clustering by sparse low-rank representation. We conducted numerous relevant experiments to validate the effectiveness of the method.

- Chapter 5 outlines the utilization of information bottleneck and anchor graph GCN techniques for addressing incomplete multi-view representation learning. Additionally, it aims to assess the effectiveness of incomplete multi-view representation learning across various degrees of missing data within the datasets.

- Chapter 6 summarizes the thesis and discusses possible future directions relevant to my research works.

Chapter 2

Background and Related Technologies

Contents

2.1 Overview

In this chapter, we present the background and relevant technologies concerning our paper's focus—incomplete multi-view clustering with hidden data mining and fusion clustering techniques. We will present key subsections that contain important concepts and techniques closely tied to the study of incomplete multi-view clustering. Furthermore, we will conduct an independent and detailed discussion of relevant research to thoroughly explore their practical applications and roles within the incomplete multi-view clustering techniques.

2.2 Multi-view Data Analysis

To date, multi-view data has extensively penetrated various industries. Different domains exhibit diverse data patterns, including but not limited to images, texts, audio, and more. In the domain of medical research, multi-view data analysis finds concrete applications in disease localization through image recognition, disease prediction, and the formulation of treatment strategies. In the field of finance, multi-view data is commonly employed for detecting market trends, evaluating investment risks, and conducting fraud detection in financial contexts. In manufacturing, the integration of data from different production stages facilitates quality control, production optimization, and supply chain management. In social media, multi-view data analysis aids in sentiment analysis of users and personalized recommendations. Furthermore, multi-view analysis demonstrates its applicability in diverse areas such as pollution monitoring, traffic scheduling, energy, and more. In summary, multi-view data analysis plays a pivotal role across various industries, providing substantial support to societal development.

In the preceding section, we have already discussed the significance and application domains of multi-view data analysis. Next, we explain the fundamental concepts of multi-view data, which will lay the foundation for our comprehensive understanding and application of analytical methods for multi-view data. In this section, we will explore the core concepts of multi-view data, containing the relationships and characteristics among different views. Moreover, we will investigate a key task in multi-view data analysis, namely, multi-view clustering. As an important data analysis technique, multi-view clustering amalgamates information from various views to achieve holistic data comprehension and pattern extraction. We will specify the groundwork and related studies in the following sections.

2.2.1 Fundamental Concepts of Multi-View Data

Multi-view data refers to data originating from diverse sources or views, with variations in content, feature representations, or perspectives. These views can encompass a range

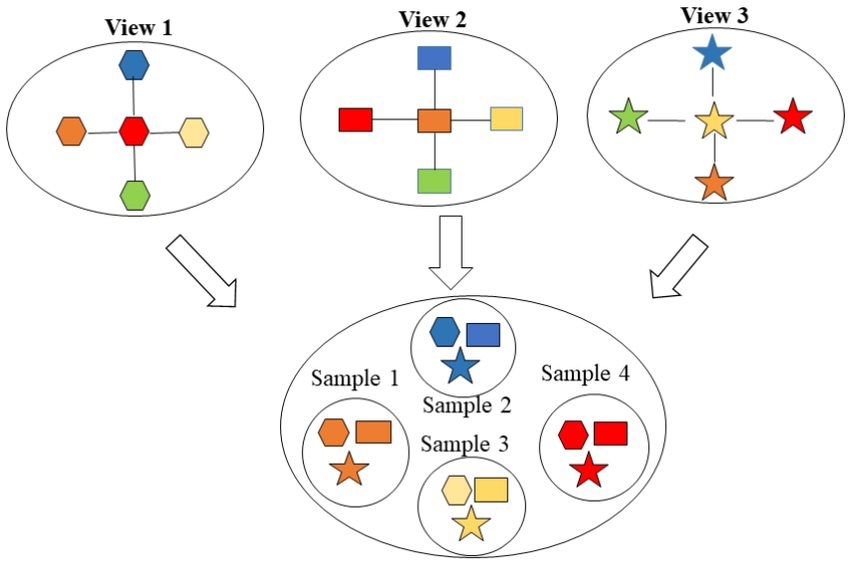


Figure 2.1: Illustration of the multi-view representation diagram. Three shapes (hexagon, rectangle, and pentagram) and four colors (blue, orange, yellow, and red) are employed to distinguish between different views and samples, with instances of the same color belonging to the same sample.

of data types, such as images, text, audio, and more. In the context of multi-view data analysis, our focus is on harnessing the interplay of these distinct views to attain a more comprehensive grasp of the data and yield more accurate analytical outcomes [3] [4] [5]. To better illustrate the characteristics of multi-view data, we have crafted Figure 2.1. As depicted in Figure 2.1, it presents a schematic representation of multi-view depiction.

2.2.2 Relevant Techniques for Multi-view Clustering

In the current era of data-driven advancements, multi-view data has found extensive applications across various domains, ranging from social network analysis to biomedical research. Multi-view data presents a wealth of information, offering opportunities for an in-depth understanding of phenomena and issues. However, with the continuous proliferation of data sources and viewpoints, multi-view data analysis faces a host of challenges.

In this chapter, we delve into technologies closely aligned with multi-view data analysis, aiming to address challenges such as data integration, feature extraction, and cross-view correlations. We will explore methods based on subspaces [6] [7] [8] [9] [10], matrix factorization [11] [12] [13] [14], graph-based integration [15] [16] [17] [18] [19] [20] [21], as well as deep learning techniques [22] [23] [24] [25]. This comprehensive approach seeks to provide a deeper and more comprehensive perspective for multi-view data analysis. Through these methodologies, we are poised to better unearth latent information within multi-view data,

thereby attaining more precise analytical outcomes and insights.

2.2.2.1 Subspace-Based Multi-view Clustering Methods

Subspace-based methods play a crucial role in the analysis of multi-view data. These techniques leverage the inherent distribution properties of data in low-dimensional subspaces to achieve clustering and correlation. By mapping data from various views onto a shared subspace, subspace-based methods capture both the commonalities and differences among multi-view data, enabling more precise analyses. In this chapter, we will provide a comprehensive overview of several prevalent subspace-based multi-view clustering approaches, including methods based on local subspaces [7], shared subspaces [8], and subspace methods involving both consistency and specificity [9] [10].

The approach based on local subspaces primarily aims to address the diversity inherent in different views. It concentrates on the local subspaces within each view. For instance, Gao et al. [7] introduced a multi-view subspace clustering method that simultaneously performs clustering on the subspace representation of each view. Additionally, the method employs a shared clustering structure to ensure consistency among different views, ultimately leading to the attainment of multi-view clustering results.

The approach based on shared subspaces has the capability to integrate information from different views into a common subspace, thus comprehensively capturing the intrinsic structure of the data. For instance, Zhang et al. [8] employed this approach to uncover underlying complementary information from multi-view data while simultaneously exploring latent representations. By harnessing the complementarity among multiple views, the learned shared latent representation depicts the data more comprehensively than individual views, thereby enhancing the accuracy and robustness of the subspace representation.

The approach based on the interplay of consistency and specificity within subspaces excels at balancing commonalities and differences during the clustering process. This equilibrium contributes to the generation of more stable and robust clustering outcomes, mitigating the instability stemming from inter-view disparities. For instance, Luo et al. [9] introduced a novel multi-view subspace clustering method (CSMSC) that leverages both consistency and specificity for subspace representation learning. The multi-view self-representation constructed by this method not only amalgamates shared consistent representations but also integrates view-specific representations. As a result, more precise clustering outcomes can be achieved.

2.2.2.2 Non-negative Matrix Factorization-Based Multi-view Clustering

The multi-view clustering methods based on non-negative matrix factorization primarily decompose the multi-view data matrices into non-negative basis matrices and weight ma-

trices, thereby mapping the multi-view data into a low-dimensional representation space. This low-dimensional representation not only aids in reducing the dimensionality of the data but also captures the shared structures and features of the multi-view data. In this section, we will provide a detailed overview of several common multi-view clustering methods based on non-negative matrix factorization, including traditional non-negative matrix factorization-based multi-view clustering [11], graph regularized non-negative matrix factorization approaches [12] [13], and deep model-based non-negative matrix factorization methods for multi-view clustering [14].

Methods based on traditional non-negative matrix factorization effectively extract features and representations from multi-view data, enabling a better capture of the data's underlying structures and patterns. This aids in revealing both shared characteristics and differences among the multi-view data. The non-negativity property typically exhibited by representations obtained from non-negative matrix factorization enhances interpretability and facilitates the understanding of data meaning and associations. For instance, Liu et al. [11] introduced a novel multi-view clustering algorithm based on non-negative matrix factorization (NMF), aiming to achieve a harmonious and consistent clustering solution across multiple perspectives. This is achieved by seeking appropriate factorizations through a constrained joint matrix factorization process, guiding the clustering solutions of each view towards a common consensus rather than directly fixing them.

The graph regularization non-negative matrix factorization approach not only effectively preserves local structural information within the data but also enables feature selection through graph structures, thus effectively filtering out noise and redundant information. Furthermore, graph regularization methods often exhibit adaptability, allowing them to flexibly learn the graph's structure, and adjust its weights, and connections based on the data's characteristics. This adaptability is beneficial in accommodating different types of multi-view data.

For instance, addressing this challenge, Li et al. [12], inspired by feature selection and graph regularization concepts, proposed a novel NMF-based unified representation learning framework. Within this framework, they incorporated two specifically designed graph regularization terms to obtain high-quality representations of multi-view data. Simultaneously, they constructed a unified NMF-based optimization problem to achieve improved consensus and complementary representations. Ultimately, they devised an alternating optimization algorithm to address this non-convex optimization problem.

The method of multi-view clustering based on non-negative matrix factorization with deep models enhances the capability of learning higher-level and abstract feature representations. This enables a more effective capture of the intricate structures and patterns within the data, facilitating the extraction of more distinctive features and ultimately im-

proving the performance of multi-view clustering. Deep models typically consist of multiple hierarchical layers, with each layer being adept at extracting features at varying levels of abstraction. Within the context of multi-view clustering, this multi-layer feature extraction contributes to a more comprehensive understanding of the data's diversity and shared characteristics. For example, Wang et al. [14] proposed a multi-view clustering algorithm based on deep learning to address challenges in multi-view clustering. In this approach, two types of activation functions were initially designed to limit the range of elements in the low-dimensional matrix, thus eliminating constraints. Subsequently, the stochastic gradient descent (SGD) algorithm was employed for element updates, with learning rates guiding parameter adjustments. After obtaining the corresponding weight and bias matrices, these were integrated with the activation functions to construct a deep network. This network was employed to update the elements within the low-dimensional matrix corresponding to each view, ultimately resulting in the derivation of the final consensus matrix.

2.2.2.3 Graph-Based Integration Multi-view Clustering

Graph fusion-based multi-view clustering methods often focus on the local structure of data, specifically the similarity relationships between samples. By preserving the local structure, these methods are better equipped to capture the underlying distribution of the data, leading to more accurate clustering outcomes. Simultaneously, the graph information from different views can complement each other. As a result, graph fusion-based approaches are capable of integrating information from multiple views to comprehensively depict the characteristics of the data, thereby enhancing the accuracy of clustering. Graph fusion-based multi-view clustering methods can be categorized into several subtypes, including weight learning-based graph fusion methods [15] [16], neighborhood graph learning-based graph fusion methods [17] [18], and one-step clustering-based graph fusion methods [19].

Graph fusion-based weight learning methods for multi-view clustering effectively leverage information from multiple views, preserve local structures, achieve information complementarity, and adaptively learn graph weights. This reduces dependence on pre-defined graph structures and yields superior performance in multi-view clustering tasks. For instance, Wang et al. [15] proposed an innovative multi-view fusion technique that automatically assigns weights to each data graph matrix to derive a unified graph matrix. A rank constraint is also imposed on the Laplacian matrix of the unified matrix without introducing tuning parameters, facilitating the natural partitioning of data points into the desired number of clusters.

Graph fusion-based neighborhood graph learning approach for multi-view clustering liberates itself from the typical constraints of predefining view-specific graph structures, a limitation of conventional multi-view clustering methods. This innovative methodology

is capable of reducing the dependence on predetermined graph structures, thus enhancing the robustness of the technique. For instance, Du et al. [17] introduced a novel multi-view clustering method known as Robust and Optimal Neighborhood Graph Learning (RONGL) for multi-view clustering. Specifically, an initial graph is constructed for each view. To better represent relationships among samples within each view, an optimal graph with k connected components is sought in the neighborhood of each initial graph, where k corresponds to the number of clusters. Subsequently, the optimal graph is reconstructed using self-representation matrices. Finally, all self-representation matrices are stacked into a tensor, and a tensor low-rank constraint is applied to maximize the enhancement of consistent features and explore high-order relationships among optimal graphs.

Graph fusion-based one-step clustering approach for multi-view clustering can directly obtain clustering results from the fused graph structure, eliminating the need for additional post-processing steps like K-means clustering. This simplifies the process and enhances efficiency. For instance, Zhan et al. [19] address the multi-view clustering problem by seamlessly integrating graph structures from different views to fully exploit the geometric properties of the underlying data structure. The proposed method is based on the assumption that the intrinsic underlying graph structure assigns corresponding connected components in each graph to the same cluster. Different graphs from multiple views are integrated using the Hadamard product, as different views often share the same underlying structure across views. Moreover, this method directly derives cluster indicators from the graph itself, without the necessity of performing further graph cutting or K-means clustering algorithms.

2.2.2.4 Deep Learning-Based Multi-view Clustering

Deep-based multi-view clustering methods, leveraging advantages such as high-level feature representation, feature fusion, shared representation learning, end-to-end learning, flexibility, adaptability, and data augmentation, have proven to be effective in enhancing the performance of multi-view clustering. These methods find applications in analyzing and utilizing complex multi-view data. Deep-based multi-view clustering approaches can be categorized into deep autoencoder-based multi-view clustering methods [20] [21] [22], deep neural network-based multi-view clustering methods [23] [24], and deep clustering-based multi-view clustering methods [25].

Deep autoencoders can be employed not only for feature learning but also for representation learning. Through the encoding layers of the autoencoder, data is mapped to a lower-dimensional representation space, which can provide more compact and discriminative feature representations for clustering. For instance, Li et al. [21] proposed a novel multi-view clustering method known as Deep Adversarial Multi-view Clustering (DAMC) network to learn the intrinsic structure embedded within multi-view data. Specifically,

this model utilizes deep autoencoders to learn latent representations shared across multiple views, while also leveraging adversarial training to further capture data distribution and disentangle the latent space.

Multi-view clustering methods based on deep neural networks can progressively extract multi-level abstract feature representations of data through multiple hidden layers. This enables the model to learn higher-level and more discriminative features, aiding in better differentiation of different data categories. Simultaneously, multi-view data can be input into different branches of neural networks, with each branch learning a view-specific representation. These representations can then be fused at higher layers of the network, thus fully utilizing complementary information from different views. For example, Zhao et al. [24] proposed a novel Deep Probability Multi-View Feature Learning (DPMFL) method to address these challenges. Specifically, this method designs a probabilistic matrix factorization (PMF) algorithm that assumes data follow Gaussian distribution during noise and dimension reduction, serving as data preprocessing. Additionally, considering the success of Deep Neural Networks (DNN) in the field of machine learning, this method integrates DNN with PMF and subspace self-representation to achieve effective consistency and specific multi-view feature learning.

The deep clustering approach based on multi-view data enables end-to-end learning, starting from raw data and progressively acquiring feature representations and clustering information through a sequence of network layers, ultimately attaining optimal clustering outcomes. This streamlines the entire process and enhances model optimization. For instance, Huang et al. [25] introduced a self-supervised graph attention network for Deep Weighted Multi-View Clustering (SGDMC), which leverages self-supervised information to enhance the effectiveness of graph-based deep Multi-View Clustering (MVC) models from two aspects. Firstly, this method devises a novel attention allocation approach that considers node attribute similarity and self-supervised information, thereby comprehensively assessing the interconnectedness of distinct nodes. Secondly, to mitigate the adverse impact of noisy samples and differences in cluster structures, the approach further devises a sample-weighting strategy based on attention graphs, along with addressing the divergence between global pseudo-labels and local cluster assignments for each individual view.

2.3 Data Missing and Processing

In the era of big data, with the continuous growth of data, the issue of missing data during the data acquisition process has become increasingly common. The presence of missing data has significant implications for practical applications, making the handling of such data a prominent research concern in the field of big data processing. This chapter aims to

systematically expound the factors leading to data incompleteness and categorize the types of data incompleteness.

2.3.1 Reasons for Data Missing

Data missing is a common issue in the era of big data applications, and its causes are diverse. Firstly, data collection processes can be prone to device malfunctions or transmission errors, resulting in data loss. Secondly, human errors during data recording, such as operational mistakes or input errors, can also lead to data gaps. Furthermore, the complexities of different domains contribute to varying missing data reasons across diverse fields. For instance, the medical field may involve privacy protection needs, while the financial sector may face complexities in data transactions. Additionally, certain data could exhibit partial information loss due to special conditions or incomplete observations. In summary, data missing can stem from various factors including technological glitches, human involvement, and domain-specific characteristics. A thorough understanding of these causes is crucial for devising effective data processing and imputation strategies [26] [27].

2.3.2 Classification of Data Missing

When dealing with data missing, a crucial approach involves categorizing the nature of the missing data, aiming to better comprehend and address different types of missingness. Data missingness can be divided into two main categories: random missing and non-random missing. Random missing refers to instances where the absence of data is unrelated to the characteristics of the observed values, possibly stemming from device malfunctions, transmission errors, and the like. Conversely, non-random missing denotes that the data's absence correlates with attributes or features of the observed values, typically occurring under specific conditions, such as certain questions being omitted by some respondents in survey questionnaires. Furthermore, data missingness can be classified as complete missing or partial missing. Complete missing implies that all attributes of a particular observed value are missing, while partial missing indicates the absence of some attributes within an observed value. By comprehending the distinct categories of data missing, tailored strategies can be employed to handle and rectify the data, ultimately enhancing the accuracy and reliability of data analysis and applications [28] [29] [30].

2.3.3 Methods for Handling Missing Values

Currently, the methods for handling missing values can be broadly categorized into three types, as illustrated in Figure 2.2: deletion, imputation, and preservation without modification [31]. Among these three approaches, the choice can be made based on the specific

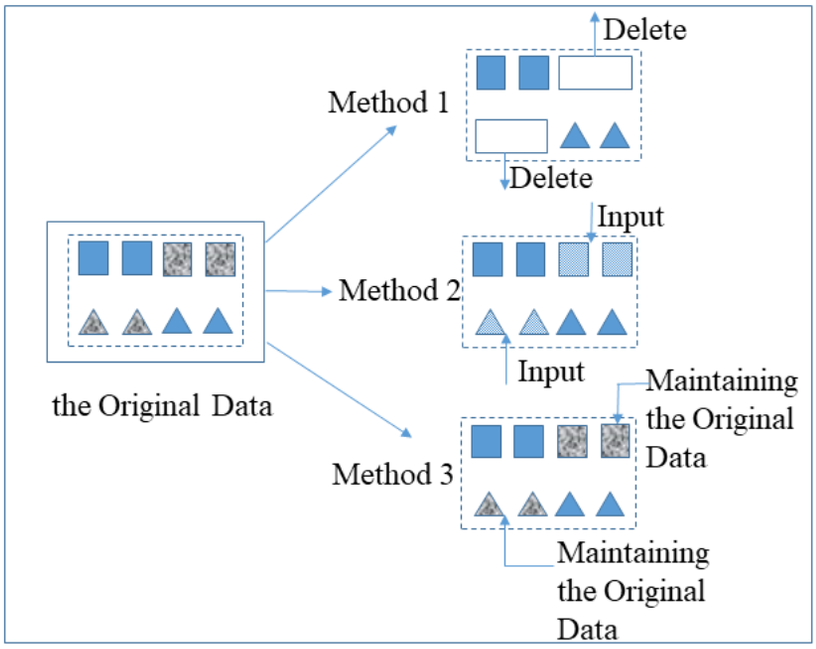


Figure 2.2: The Main Method of Missing Data Processing.

nature of the missing data.

2.3.3.1 Deletion Method

The most basic method for handling missing data mainly involves the simple deletion approach, where instances containing missing values or attribute values within instances are removed [32].

Instance Deletion involves the removal of a specific research object's data record from the dataset when it contains missing values. This approach is suitable only when the number of instances with missing values is exceedingly small, as excessive removal of instances due to missing values can lead to significant data loss and consequently compromise the dataset's integrity, thereby impacting the accuracy of subsequent experimental results.

Attribute Deletion entails the removal of a particular attribute from instances in the dataset when that attribute is missing. However, if a substantial number of instances lack attribute information, excessive attribute deletion can occur, undermining the meaningfulness of the research.

2.3.3.2 Imputation Method

Currently, various imputation methods have been proposed in the academic community to address the issue of missing data. These methods can be broadly categorized into two main classes: statistical methods and machine learning methods [33]. Statistical methods are based on assumptions made about the data itself and rely on the original dataset for imputation. However, these methods often do not consider the characteristics of the data, potentially leading to imputed values being influenced by other data objects. On the other hand, machine learning methods primarily involve clustering or classifying the missing raw data before performing imputation. In the following sections, we will provide a detailed overview of several commonly used specific methods. The methods mainly include filling manually, mean imputation, expectation maximization imputation, and clustering imputation.

The manual filling method primarily relies on the characteristics of the dataset itself for imputation. This approach is often the quickest and most accurate for filling many missing data points. However, when dealing with exceptionally large dataset sizes and a substantial amount of missing data, not only does it consume a significant amount of time, but it also increases the risk of errors. Additionally, this method's feasibility is considerably reduced for other users of the dataset.

The mean imputation method is a common approach for handling missing data, and its fundamental idea is to replace missing values with the mean of the corresponding instances or attributes from the available data. However, it's important to note that this method is suitable when the data's variables follow or approximate a normal distribution. If the distribution of the data doesn't meet this requirement, using the mode or median of instances or attributes can be considered for filling in the missing values [34]. Despite being widely used for imputing missing values, the mean imputation method has limitations and is best suited for simple research scenarios with smaller data scales and fewer missing values. In more complex research scenarios, the effectiveness of the mean imputation method becomes more limited [35] [36].

The EM algorithm (Expectation-Maximization algorithm) was originally proposed by Dempster et al. [37]. This method involves alternating between two computational steps. Firstly, in the Expectation step, existing estimates of hidden variables are used to calculate their maximum likelihood estimates. Secondly, in the Maximization step, the parameter values are computed by maximizing the maximum likelihood value obtained from the previous step. The parameter estimates found during the Maximization step are used in the next Expectation step calculation, facilitating an iterative alternating process.

Cluster-based imputation is one of the most widely employed methods in current research for handling missing data. This approach begins by grouping the dataset using clus-

tering techniques, followed by performing similarity-based imputation within each group. Taking the classic K-means clustering-based imputation algorithm as an example, the process involves initially dividing the original dataset into complete and missing data subsets. Clustering is then performed on the complete data subset, resulting in K clusters. Subsequently, the similarity between the missing data objects and the centroids of the K clusters is computed. Finally, the missing values of the target object are imputed with the mean attribute values of the most similar cluster [38].

2.3.3.3 Non-Processing Method

This non-imputation method can directly learn from the original data [39] [40]. Among these, Bayesian networks, artificial neural networks, and rough set theory are commonly used non-imputation methods.

Bayesian networks are graphical models used to represent dependencies between variables [41]. However, their utilization is constrained by the high requirement for domain knowledge.

Artificial neural networks effectively address the issue of missing values using methods such as radial basis functions. Nevertheless, the complex and less interpretable nature of the learning process in neural network models poses limitations.

On the other hand, rough set theory is a mathematical tool for handling uncertainty and incomplete information. It aids in identifying patterns, regularities, and relationships within data, thereby facilitating partial data completion [42].

2.4 Hidden Data Mining in Incomplete Multi-view Data

An incomplete view refers to the situation in which data contains missing, incomplete, or inaccurate information, resulting in the inability to present the data in its entirety [2]. As shown in Figure 2.3, an example of incomplete multiview data is presented, containing data from three samples: cat, dog, and bird. Each instance was originally described by three modalities: text, image, and video. However, upon observation, it is noted that two instances each lack data from one modality. The purpose of this chapter is to introduce the role of hidden data mining in incomplete multiview data. This approach involves analyzing and extracting latent patterns, regularities, and information from multiple modalities to fill in missing data and reveal interrelationships among the data. This method contributes to a more comprehensive understanding of the dataset and enhances the accuracy of tasks such as prediction, classification, and clustering.

Existing methods for hidden data mining in incomplete multi-view data can be categorized from two perspectives. One classification approach is based on hidden data recovery,

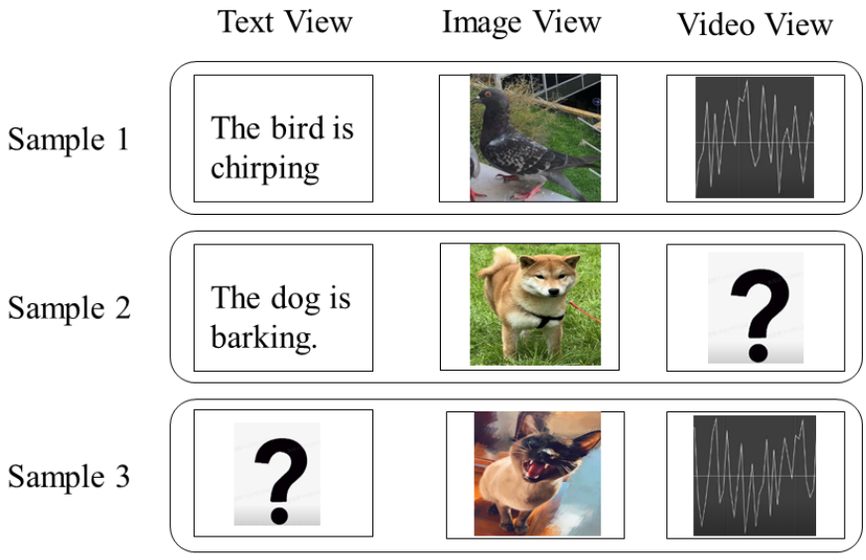


Figure 2.3: Illustration of Incomplete Multiview Data.

where the existing methods can be divided into two main groups. The first group of methods focuses on addressing the incomplete learning problem by recovering missing views or missing connections between samples. The second group of methods, on the other hand, does not specifically aim to recover missing view information; instead, they concentrate on capturing partially aligned information among the available views. Another classification approach is based on the adopted modeling perspectives, and the existing methods can be classified into four categories: matrix factorization (MF)-based incomplete multiview learning [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53], kernel learning-based incomplete multi-view learning [54] [55] [56] [57] [58] [59], graph learning-based incomplete multi-view learning [60] [61] [62] [63] [64] and deep learning-based incomplete multi-view learning [65] [66] [67].

2.4.1 Matrix Factorization (MF)-based Incomplete Multi-view Clustering

The matrix factorization-based approach for incomplete multiview clustering possesses multiple advantages. Firstly, this method effectively leverages the correlations and complementary information among multiview data by representing data in a low-dimensional latent feature space, thereby capturing the inherent structure of the data more accurately. Secondly, matrix factorization can handle scenarios with missing views, generating meaningful clustering results even when partial view data is absent. Furthermore, this approach is grounded in solid mathematical theory, providing it with high interpretability and stability

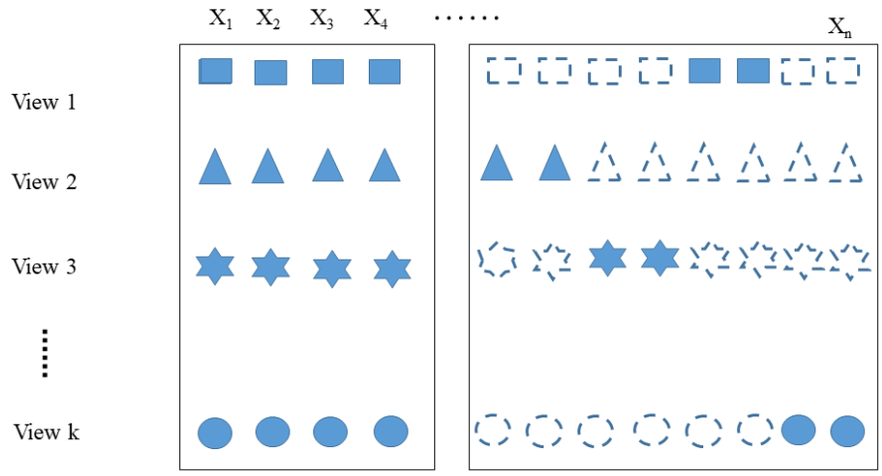


Figure 2.4: Partial Multiview Data. Geometric shapes in four different colors (i.e., rectangle, triangle, hexagram, circle) and four distinct hollow geometric shapes represent non-missing and missing samples, respectively.

in the field of multiview clustering. Lastly, the method demonstrates strong scalability, adapting well to various data types and numbers of views, offering a potent tool for analyzing complex multiview data. Consequently, the matrix factorization-based approach to incomplete multiview clustering boasts significant advantages in enhancing clustering performance, managing missing data, and maintaining stability. Next, we will present two categories of matrix factorization-based incomplete multiview clustering methods, categorized based on the different types of data being addressed.

2.4.1.1 MF-Based Approaches for Partial Multiview Data

To address this type of missing data as depicted in Figure 2.4, MF-based approaches for partial multiview data were proposed.

For instance, to address such partial data missing scenarios, a novel multi-view clustering method called Partial Multiview Subspace Clustering (PMSC) was introduced by Xu et al. [45]. This method aims to tackle the issue of partial multi-view scenarios. Unlike most of the existing partial multi-view clustering methods, the proposed approach not only learns a novel representation of the original data but also simultaneously explores the latent space and performs data reconstruction to acquire a subspace representation. The learned subspace representation uncovers the latent subspace structure embedded within the raw data, resulting in a more comprehensive data depiction. Furthermore, the method enforces non-negativity on the subspace representation, providing an intuitive weight interpretation among distinct data points.

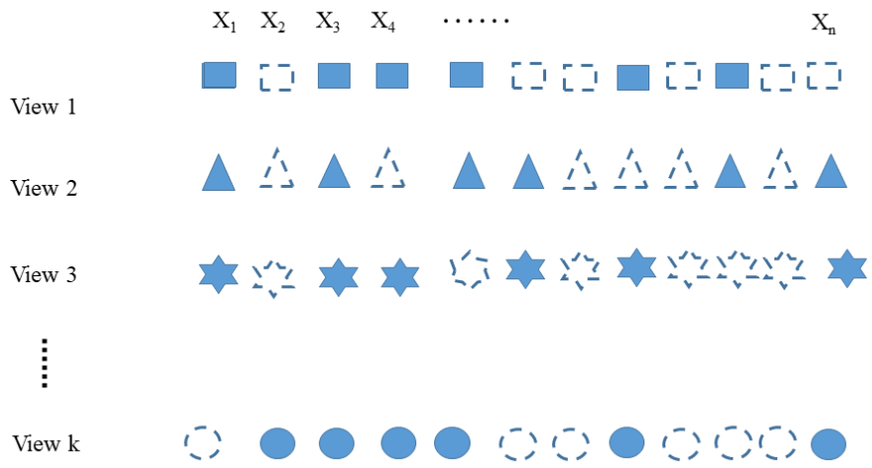


Figure 2.5: Arbitrary Missing Views. Geometric shapes in four different colors (i.e., rectangle, triangle, hexagram, circle) and four distinct hollow geometric shapes represent non-missing and missing samples, respectively.

2.4.1.2 MF-Based Approaches for Arbitrary Missing Views

To tackle this type of missing data depicted in Figure 2.5, MF-based approaches were proposed for dealing with Arbitrary Missing Views.

For instance, Hu et al. [48] proposed a doubly aligned incomplete multi-view clustering (DAIMC) based on weighted semi-nonnegative matrix factorization (semi-NMF). Specifically, DAIMC leverages the provided instance alignment information to learn a shared latent feature matrix for all views. Additionally, DAIMC employs $L_{2,1}$-norm-regularized regression to establish a consensus basis matrix, aiming to mitigate the impact of missing instances.

2.4.2 Kernel Learning-based Incomplete Multi-view Clustering

Kernel learning methods can address incomplete kernel matrices through appropriate strategies, thereby tackling the issue of missing views and effectively handling common data incompleteness in practical applications. Most kernel learning-based incomplete multiview clustering methods focus on recovering the missing rows and columns of kernel matrices to address the incomplete learning problem. Based on the primary techniques used to tackle this incompleteness, we categorize existing kernel learning-based incomplete multiview clustering methods into two groups: the first group employs Laplacian regularization and kernel canonical correlation analysis (KCCA), while the second group utilizes the multiple kernel k-means algorithm.

2.4.2.1 Laplacian Regularization and KCCA-Based Incomplete Multi-view Clustering

Incomplete multiview clustering (IMC) methods based on Laplacian regularization and Kernel Canonical Correlation Analysis (KCCA) typically start by completing the kernel matrices and then proceed to learn the latent representations of all views using KCCA. For instance, Shao et al. [55] proposed an approach called "Collective Kernel Learning" to address this issue. This method aims to infer hidden sample similarities from multiple incomplete datasets. The approach collectively completes the kernel matrices of incomplete datasets by optimizing the alignment of shared instances across the datasets.

2.4.2.2 Multiple Kernel K-means-Based Incomplete Multi-view Clustering

Compared to traditional methods, the approach based on multiple kernel K-means exhibits improved capability in handling incomplete multiview data, where certain views might exhibit missing data. By integrating multiple kernel functions, the algorithm can effectively capture existing information, thereby enhancing clustering accuracy. For instance, Liu et al. [58] identified that existing MKC algorithms struggle to effectively handle situations where certain rows and columns are missing in the base kernel matrix. Therefore, they introduced a straightforward yet effective algorithm to address this issue. Unlike previous methods that typically involve completing the incomplete kernel matrix first and then applying the standard MKC algorithm to the completed matrix, their proposed approach integrates kernel matrix completion and clustering into a unified learning process. Importantly, their algorithm doesn't mandate the presence of at least one complete base kernel matrix across all samples. Additionally, this algorithm has the capability to adaptively impute incomplete kernel matrices and fuse them to better support clustering. Furthermore, they enhanced the algorithm by encouraging mutual completion among the incomplete kernel matrices.

2.4.3 Graph Learning based Incomplete Multi-view Clustering

Graph-based learning enables the integration of available information from diverse views, contributing to the establishment of more robust and comprehensive data representations. This facilitates a better capture of the intrinsic data structure, even in the presence of incomplete views. Furthermore, graph-based approaches can accommodate varying degrees of view completeness. This category of methods can effectively impute missing information based on the available data from other views, particularly excelling when confronted with substantial data incompleteness. Regarding the diverse correlated representations of multi-view data obtained through graph-based approaches, we categorize this method primarily into two classes, denoted as graph-based incomplete multiview clustering with consensus

representation and graph-based incomplete multiview clustering with consensus graph representation.

2.4.3.1 Graph-Based Incomplete Multiview Clustering with Consensus Representation

When data is provided by multiple views with existing correlations and complementary information among these views, the approach of multiple-view consensus representation excels at amalgamating this information, thereby enhancing the cohesiveness and accuracy of data representation. This is particularly advantageous for tasks that involve multiple types of features or stem from diverse data sources. For instance, Wen et al. [63] proposed a general framework for incomplete multiview clustering. This method is the first to utilize graph learning and spectral clustering techniques to learn a consensus representation for incomplete multiview clustering. Firstly, due to the effective performance of low-rank representation in discovering the intrinsic subspace structure of data, the method employs it to adaptively construct a graph for each view. Secondly, spectral constraints are applied to achieve low-dimensional representations for each view based on spectral clustering. Thirdly, the method introduces a co-regularization term to further learn a common representation for samples across all views, followed by using k-means to partition the data into respective groups. Lastly, an efficient iterative algorithm is provided to optimize this model.

2.4.3.2 Graph-Based Incomplete Multiview Clustering with Consensus Graph Representation

When the relationships and structures of data are crucial for the success of a task, graph-based consensus representation methods excel in capturing the relationships among data points. Especially in fields such as network analysis, social networks, and bioinformatics, graph-based consensus representation can reveal intricate interaction patterns among the data. For example, Zhou et al. [62] proposed a novel graph-based method for incomplete multi-view clustering (GIMC) to address this task. GIMC is capable of effectively constructing complete graphs for each view with the assistance of other views, and automatically assigning weights to each constructed graph to learn a consensus graph, thereby obtaining the final clustering results.

2.4.4 Deep Learning-based Incomplete Multi-view Clustering

Deep learning models possess the ability to adapt to varying degrees of data incompleteness. Through training, these models can automatically learn valuable features from partial views, thereby exhibiting strong robustness when dealing with incomplete data. Furthermore, deep learning methods can tailor their network structures and loss functions to accommodate

diverse tasks and data characteristics. This flexibility enables them to address a wide range of incomplete multi-view clustering problems with excellent scalability. In this section, we will primarily introduce several existing methods that utilize deep models to tackle incomplete multi-view clustering.

2.4.4.1 Incomplete multi-view clustering via deep semantic mapping

Based on deep semantic mapping, the method is capable of learning high-level, semantically-rich feature representations by extracting valuable information from incomplete views, enabling more accurate clustering. For example, Zhao et al. [65] proposed a method known as Deep Incomplete Multi-view Clustering (DIMC), which integrates constraints from an intrinsic geometric structure with incomplete multi-view samples. To bridge the gap between each view and the common representation, they trained multi-view deep coupled networks to map high-level semantic features. Additionally, to preserve the local invariance of each view, they constructed an affinity graph-based regularizer for encoding geometric information.

2.4.4.2 Partial multi-view clustering via consistent GAN

The advantage of using Generative Adversarial Networks (GANs) is that they can generate realistic data samples to fill in missing data points from partial or incomplete views. This enhances data completeness and availability, enabling a more accurate capture of the underlying structure and patterns during the clustering process. Additionally, GAN architecture and loss functions can be flexibly adjusted based on different problems and data characteristics. This versatility allows the method to adapt to various incomplete multi-view clustering scenarios and exhibit excellent performance in diverse contexts. For instance, Wang et al. [66] introduced a novel Consistent Generative Adversarial Network (CGAN) for partial multi-view clustering. This method learns a shared low-dimensional representation that is capable of generating missing view data and capturing a more effective common structure from partial multi-view data for clustering purposes. Diverging from the majority of existing approaches, this method employs the shared representation encoded by one view to generate the missing data of the corresponding view through a generative adversarial network. Subsequently, encoder and clustering networks are employed.

2.4.4.3 Adversarial incomplete multi-view clustering

The advantage of adversarial incomplete multi-view clustering methods lies in their ability to effectively handle missing data through an adversarial learning framework. By introducing generators and discriminators, these methods can generate missing data from partial

or incomplete views, while enhancing data consistency and integrity through adversarial training. This contributes to a more accurate capture of hidden patterns and structures within the data, leading to improved clustering performance. Leveraging the power of adversarial training, these methods can adapt to various scenarios of missing data, offering robust and effective solutions for multi-view clustering. For example, Xu et al. [67] proposed a method called Adversarial Incomplete Multiview Clustering (AIMC). Unlike most existing methods that solely learn new representations based on existing views, AIMC aims to simultaneously seek the common latent space of multi-view data and perform missing data inference. Specifically, the method employs a combination of element-wise reconstruction and Generative Adversarial Networks (GANs) to infer missing data. These techniques aim to capture both the overall structure and deeper semantic understanding. Furthermore, an alignment clustering loss is designed to improve the clustering structure.

Chapter 3

Incomplete Multi-view Clustering by NMF and Low-rank Tensor

Contents

3.1 Introduction

With the development of data acquisition technology, sources and types of data are becoming more diverse. The collected data were characterized by multiple views. Complex data features can be better understood by fusing information from multi-view data. A common approach is multi-view clustering. The basic idea of this type of approach is that multi-view data are strongly related and complementary to each other in different ways [68] [69] [70] [71] [72] [73] [74] [75] . However, not all data will contain complete multi-view data. We consider land data as an example, often consisting of data with different views, such as symbols, text, and graphics. In reality, some land data may have the format of graphics but lack text or symbols. As such, incomplete multi-view data were formed [76] . That is, a dataset has an arbitrary loss of view with only some instances containing all views.

Specifically, incomplete multi-view clustering methods can be roughly grouped into four categories. Based on a filling strategy, the methods of the first category [58] [47] [77] [51] usually start by selecting a suitable padding algorithm and then applying existing multi-view learning methods to incomplete multi-view datasets populated by padding algorithms. For example, Shao et al. [77] first filled incomplete samples by averaging the eigenvalues and using regularised weighted non-negative matrix decomposition to learn a subspace. The methods in the second category [78] [79] [80] [81] [44] ignore incomplete view information in the process of learning potential representations. Quanz et al. [78] pushed clustering solutions for different views from the same example to the standard membership matrix to simultaneously generate the underlying geometric structure of the views. Learning from a single view, the third method [82] [83] attempts to learn a unified model for all views. Yuan et al. [83] built a separate classifier for each data source. This model learns different base classifiers for different data sources. Based on a classifier, the method uses a single column of prediction scores to represent each source and then combines the estimated incomplete prediction scores with the available prediction scores. Thus, a multi-source fusion model was built. The methods of the last category [84] [85] [86] [65] use deep models for incomplete multiview clustering. For example, Tran et al. [87] proposed a cascaded residual autoencoder (CRA) to complement the incomplete multi-view. By stacking the residual autoencoders, the residuals between the current prediction and the original data were obtained by iterative simulation of the algorithm.

Although many methods have been proposed to address the incomplete problem of multi-view clustering, they have some limitations. For example, as mentioned by [88], the first limitation is that most approaches do not make full use of the information of observation instances inside and between views, resulting in missing valuable information.

Second, some models are sensitive to the choice of parameters and are less robust [89], which restricts their availability in real-world scenarios. Third, existing methods cannot effectively handle samples with incomplete views, which inevitably reduces the performance of the IMC [63].

To overcome these limitations, in this paper, we propose a unified framework for incomplete multiview clustering using non-negative matrix factorization (NMF) and the low-rank tensor, called IMC-NLT. Specifically, IMC-NLT first utilizes NMF to learn a low-dimensional representation for all the views. In this way, not only is the dimensionality of the data reduced but also non-negative numerical effects with strong explanatory power can be obtained. Using the prior information of a low-rank tensor, IMC-NLT can capture the higher-order and complementary information embedded in the multi-view data. Finally, a new cost function is introduced to measure consistent information across views, using a linear kernel that measures similarities. The contributions of this study are summarised as follows.

- We propose a novel and efficient incomplete multi-view clustering model called IMC-NLT to handle incomplete view data. To the best of our knowledge, IMC-NLT is the first incomplete multi-view clustering method that combines the low-dimensional representation generated by the fast and effective dimension-reduction method with a low-rank tensor model.

- To populate missing multi-view data in various cases, IMC-NLT uses a low-rank constraint and a tensor model constructed from incomplete multi-view data. As such, it can capture the correlations among instances within and between views well.

- The proposed method is robust to globally extracted consistency information. In particular, consistent representation learning can effectively measure the disagreement between consistent information obtained from different perspectives. IMC-NLT can effectively filter out noisy data under these views, producing an accurate multi-view consistent representation.

- We conducted comprehensive experiments on multi-view benchmark datasets collected in different application fields to evaluate the effectiveness of the IMC-NLT. The results showed that IMC-NLT is superior to baseline methods. Furthermore, IMC-NLT has low sensitivity to its parameters, which demonstrates excellent generalization performance for incomplete multi-view clustering.

The remainder of this paper is organised as follows: Section 3.2 describes the related work and background. Section 3.3 describes the proposed IMC-NLT algorithm. Sections

3.4 and 3.5 present the theoretical analysis of IMC-NLT and evaluation results of the performance of IMC-NLT, respectively. In Section 3.6, we report the experiments on the parameters and convergence of IMC-NLT. Finally, Section 3.7 concludes the paper.

3.2 Related Work and Background

In this section, we present two studies closely related to the proposed method. First, we present the basics of low-rank tensor-based models for multi-view learning. Here, we review incomplete multi-view clustering based on matrix factorization.

3.2.1 Multi-view Learning Based on Low-rank Tensor

A low-rank tensor can effectively capture hidden information as a valid technique for analyzing high-dimensional data. There are many examples of low-rank tensors used for multi-view clustering [90] [91]. Zhang et al. [91] proposed a subspace clustering method called low-rank tensor-constrained multi-view subspace clustering (LT-MSC). LT-MSC can be described by the following equation:

$$\begin{aligned} &\min_{Z^{(v)},E^{(v)}} \|\mathcal{Z}\|_* + \lambda\|E\|_{2,1} \\ &s.t.\ X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, v = 1, 2, ..., V, \\ &\mathcal{Z} = \Psi(Z^{(1)}, Z^{(2)}, ..., Z^{(v)}), \\ &E = [E^{(1)}; E^{(2)}; ...; E^{(v)}], \end{aligned} \tag{3.1}$$

where $X^{(v)}$ is the v-th view of data. Note that $X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}$. The formula can learn the subspace representation matrix $Z^{(v)}$ by exploiting the self-expressive properties of the data; λ is a positive penalty parameter, and $E^{(v)}$ is the reconstruction error matrix. Imposing the $L_{2,1}$ regularizer on $E^{(v)}$ can urge the data in each column of matrix $E^{(v)}$ to be close to 0. $\|\mathcal{Z}\|_*$ is the tensor kernel norm constraint added to $\mathcal{Z}$. The tensor $\mathcal{Z}$ consists of subspace representation $Z^{(v)}$. The model obtains the low-rank tensor using a self-representation multi-view.

Similar to previous research methods, Xu et al. [90] proposed a method called low-rank tensor-constrained co-regularised multi-view spectral clustering(LTCSPC). The objective function of the LTCSPC is as follows:

$$\min_{F^{(v)}\in\mathbb{R}^{n\times c}} \sum_{v=1}^{m} \alpha^{(v)} Tr(F^{(v)^T} L^{(v)} F^{(v)}) + \|\mathcal{F}\|_{\omega,\circledast} \tag{3.2}$$

where we have

$$\alpha^{(v)} = 1 \Big/ (2\sqrt{Tr(F^{(v)^T} L^{(v)} F^{(v)})}) \tag{3.3}$$

In Equation 3.3, $\alpha^{(v)}$ is the weight of each view and ω is the singular value-weighted coefficient. LTCSPC calculates $F^{(v)}$ according to the standard spectral clustering and data $X^{(v)}$ with m views. To use the high-order structure and complement, we denote the indicator matrix as $F^{(v)} \in \mathbb{R}^{n\times c}$, where n is the number of samples, and c is the number of categories. The slice of tensor $\mathcal{F}$ consists of the indicator matrix $F^{(v)}$. $L^{(v)}$ is the Laplacian matrix. $\|\mathcal{F}\|_{\omega,\circledast}$ is the weighted nuclear norm constraint added to $\mathcal{F} \in \mathbb{R}^{n\times m\times c}$, which is defined as:

$$\|\mathcal{F}\|_{\omega,\circledast} = \sum_{i=1}^{c} \|\bar{F}^{(i)}\|_{\omega,*} = \sum_{i=1}^{c} \sum_{j=1}^{min(n,m)} \omega_j * \sigma_j(\bar{F}^{(i)}) \tag{3.4}$$

where $\bar{F}^{(i)} \in \mathbb{R}^{n\times m}$, $\sigma_j(\bar{F}^{(i)})$ is the j-th largest singular value of $\bar{F}^{(i)}$, and ω_j is the j-th element of the vector ω. Although these methods can elegantly model different views, they improve the clustering accuracy while reducing the redundancy of the learning subspace representation. In addition, they are only applicable to complete multi-view data, and cannot handle incomplete cases.

3.2.2 Incomplete Multi-view Clustering Based on Matrix Factorization

In this section, we review two classical methods: one-pass incomplete multi-view clustering(OPIMC) [50] and partial multi-view clustering(PVC) [43].

OPIMC applies regularized matrix factorization (RMF) [91] and weighted matrix factorization (WMF) [92] to produce multi-view clustering results. The objective function of the OPIMC is written as:

$$\begin{aligned} J = &\sum_{v=1}^{n_v}\{\sum_{t=1}^{[N/s]} \left\| (X_t^{(v)} - U^{(v)} V_t^T)W_t^{(v)} \right\|_F^2 + \alpha \left\| U^{(v)} \right\|_F^2\} \\ s.t.\ &V_{ik} \in \{0,1\}, \sum_{k=1}^{K} V_{ik} = 1, \forall i = 1,2,...,N \end{aligned} \tag{3.5}$$

where $X_t^{(v)}$ denotes the t-th data chunk in the v-th view. Assume that each view is composed of blocks of size s. $V \in \mathbb{R}^{N\times K}$ is a clustering indicator matrix, and K represents the number of categories of the data. If the instance belongs to one category, it is marked as 1; otherwise, it is 0. $U^{(v)}$ and V_t are the low-rank regularized factor matrix and the clustering indicator matrix for the t-th data chunk, respectively. α is a non-negative parameter. Where N denotes the number of instances. Furthermore, to achieve a high calculation cost when the number of instances and categories are large, the model applies a 1-of-K coding constraint to V, where $W_t^{(v)}$ is a diagonally weighted matrix of the t-th data chunk. OPIMC can directly obtain the clustering results at the end of the iteration.

PVC is an incomplete multi-view clustering method that uses NMF. It constructs a specific latent space for unaligned instances and a shared latent space for aligned instances. These shared and specific representations were used together for clustering. PVC was formalized as follows:

$$\begin{aligned}\min_{P_{(c)},\hat{P}^{(1)},\hat{P}^{(2)},U^{(1)},U^{(2)}} & \left\| \begin{bmatrix} {X_c}^{(1)} \\ \hat{X}^{(1)} \end{bmatrix} - \begin{bmatrix} P_c \\ \hat{P}^{(1)} \end{bmatrix} U^{(1)} \right\|_F^2 \\ & + \left\| \begin{bmatrix} {X_c}^{(2)} \\ \hat{X}^{(2)} \end{bmatrix} - \begin{bmatrix} P_c \\ \hat{P}^{(2)} \end{bmatrix} U^{(2)} \right\|_F^2 \\ & + \lambda \left\| \bar{P}^{(1)} \right\|_1 + \lambda \left\| \bar{P}^{(2)} \right\|_1 \\ & s.t. U^{(1)} \geq 0, U^{(2)} \geq 0, \bar{P}^{(1)} \geq 0 \\ & \bar{P}^{(2)} \geq 0 \end{aligned} \tag{3.6}$$

In Equation 3.6, ${X_c}^{(1)}$ and ${X_c}^{(2)}$ represent instance data that exists in both views, $\hat{X}^{(1)}$ represents instance data that exists only in the first view, and $\hat{X}^{(2)}$ represents instance data that exists only in the second view. P_c represents the low-dimensional representation of the common views after matrix decomposition. $\hat{P}^{(1)}$ and $\hat{P}^{(2)}$ represent individual potential representation parts of each view. $U^{(v)}$ is the basis matrix of the view, $\bar{P}^{(1)} = [P_c; \hat{P}^{(1)}]$ and $\bar{P}^{(2)} = [P_c \hat{P}^{(2)}]$ are the latent representation of instances in the latent space. λ is the positive tradeoff parameter. The grouping result was obtained by establishing a potential subspace.

3.3 Proposed IMC-NLT

To cluster incomplete multi-views, we present a new incomplete multi-view clustering model called IMC-NLT. It consists of three parts: incomplete multi-view data filling and decomposition, multi-view fusion of low-rank tensor, and consensus representation learning. The framework for the IMC-NLT is illustrated in Figure 3.1.

3.3.1 Incomplete Multi-view Data Filling and Decomposition

To align the dimensions of multi-view data in a way that better reflects the structure of incomplete data, we used NMF to build a multi-view data model with a unified dimension and maintain information about the data space and feature space. The model is as follows:

$$\begin{aligned} \min_{H^{(v)}} & \sum_{v=1}^{V} \sum_{i=1}^{n_o^{(v)}+n_{cp}^{(v)}} \left\| \begin{bmatrix} Y_{o_i}^{(v)} \\ Y_{cp_i}^{(v)} \end{bmatrix} - \begin{bmatrix} H_{o_i}^{(v)} \\ H_{cp_i}^{(v)} \end{bmatrix} Q_i^{(v)} \right\|_F^2 \\ & s.t.\ H^{(v)} \geq 0, Q^{(v)} \geq 0, H^{(v)} = T \begin{bmatrix} H_o^{(v)}; H_{cp}^{(v)} \end{bmatrix} \end{aligned} \tag{3.7}$$

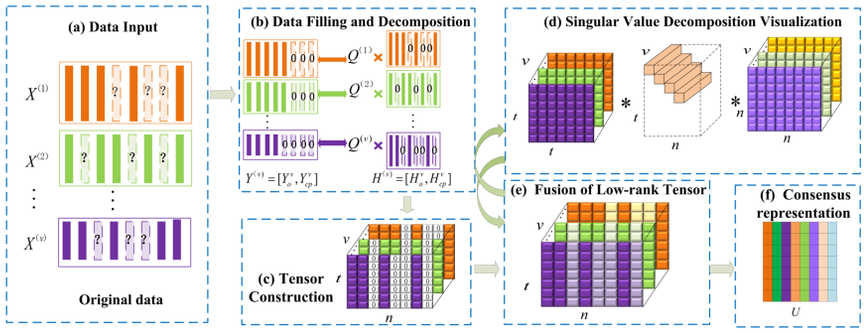


Figure 3.1: Overview of IMC-NLT. IMC-NLT consists of six major components: (a) Incomplete multi-view data input; (b) Pre-filing and dimensionality reduction of incomplete view data; (c) Construction of a tensor structure with unified view dimensions. (d) Singular value decomposition visualization; (e) The Low-rank tensor fusion implements data filling; and (f) Consensus representation.

An example is shown in Figure 3.2, where matrix $Y_o^{(v)} \in \mathbb{R}^{n_o^{(v)} \times d^{(v)}}$ is the matrix of complete instances selected from the original view $X^{(v)} \in \mathbb{R}^{n^{(v)} \times d^{(v)}}$. $n^{(v)}$ is the number of instances when the view is complete, $n_o^{(v)}$ is the number of instances in each view that are not missing, and $d^{(v)}$ represents the original feature dimension of multi-view data. $Y_{cp}^{(v)} \in \mathbb{R}^{n_{cp}^{(v)} \times d^{(v)}}$ is a matrix of incomplete instances selected from the original view. $n_{cp}^{(v)}$ is the number of instances incomplete in the multi-view data. To better reflect the structure of incomplete data and better fill in incomplete values, we filled the incomplete view data instances matrix $Y_{cp}^{(v)}$ with 0. The main purpose of dividing the available part $Y_o^{(v)}$ and missing part $Y_{cp}^{(v)}$ is to emphasize the filling of the missing part of the data using the proposed low-rank tensor model. The zero matrices represent the missing part of the data, whereas the change in the zero matrices reflects the data recovery function of the model proposed in this study. By introducing the decomposition model, we constructed the following low-dimensional representation structure: $H_o^{(v)} \in \mathbb{R}^{n_o^{(v)} \times t}$ is the low-dimensional representation matrix formed by the complete instances, where t represents the unified dimension number after the dimension reduction. The selection of the feature dimension t depends on the existing algorithms for multi-view data-sharing features based on NMF [93]. $H_{cp}^{(v)} \in \mathbb{R}^{n_{cp}^{(v)} \times t}$ is a low-dimensional representation matrix formed by the incomplete instances. $Q^{(v)} \in \mathbb{R}^{t \times d^{(v)}}$ is the coefficient matrix after non-negative factorization. T is a reconstruction operation of the modal matrix after the low-dimensional representation of each data view. The purpose of this is to arrange incomplete modal data after dimensionality reduction in the original order of the instance arrangement.

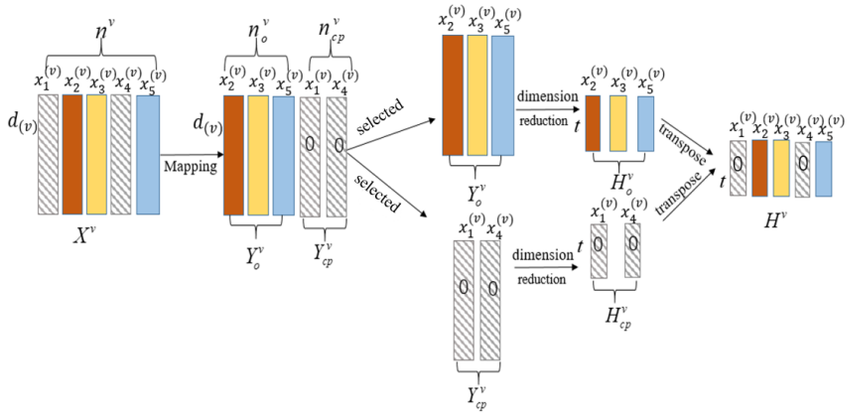


Figure 3.2: Example of incomplete multi-view data filling and decomposition.

3.3.2 Multi-view Collaborative Fusion of Low-rank Tensor

In the previous section, we completed the unified transformation of the data dimensions between views and pre-filling of missing data. In this section, we introduce a low-rank tensor-filling model that can better capture high-order correlations between viewing data. This process consists of two parts:1) the construction of the tensor, and 2) addition of low-rank constraints to the tensor.

An intuitive example is shown in Figure 3.3, where $H^{(v)}$ is the low-dimensional representation matrix after incomplete multi-view data-filling and decomposition. A third-order tensor model was constructed using the enumeration method. Its construction is as follows:

$$\mathcal{H} = \Psi(H^{(1)}, H^{(2)}, ..., H^{(v)}) \tag{3.8}$$

operation Ψ represents listing the reduced-dimensional second-order structure matrix $H^{(v)}$ in view order to construct a third-order tensor $\mathcal{H}$.

After obtaining the tensor model, we use tensor kernel parametrization to approximate the tensor low-rank representation. The optimization model is as follows:

$$\min_{\mathcal{H}} \|\mathcal{H}\|_* \tag{3.9}$$

To make the objective function separable, we introduced the auxiliary variable K_m to solve the optimization problem in Equation 3.9. The formula used is as follows:

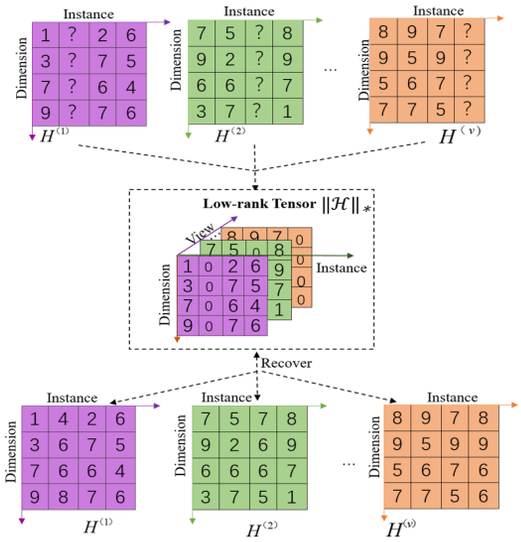


Figure 3.3: Example of restoring incomplete data based on a low-rank tensor.

$$\begin{aligned} &\min_{H^{(v)},K_m} \sum_{m=1}^{M} \sigma_m \left\|K_m\right\|_* \\ &s.t. k_m = P_m h, m = 1,2,...,M, \\ &\mathcal{H} = \Psi(H^{(1)}, H^{(2)}, ..., H^{(v)}) \end{aligned} \tag{3.10}$$

where σ_m represents the strength of the low-rank tensor constraint and k_m is the vectorization of the matrix K_m. P_m is the alignment matrix used to align the corresponding elements between $H_{(m)}$ and K_m, and $H_{(m)}$ is the matrix obtained by unfolding tensor $\mathcal{H}$ along the m-th mode, defined as unfold$_m(\mathcal{H})$= $H_{(m)} \in \mathbb{R}^{I_m \times (I_1 \times ... I_{m-1} \times I_{m+1} ... I_M)}$. The vectorization of tensor $\mathcal{H}$ is denoted as h. This process can recover incomplete data based on effective association information.

3.3.3 Consensus representation learning

Generally, an incomplete multi-view clustering algorithm based on the fusion strategy gives the clustering results a more explicit physical meaning. However, existing incomplete multi-view clustering algorithms based on fusion strategies still suffer from several shortcomings:1) during the construction of the model, the fusion results are obtained separately from each view, and the similarity features between views are ignored, resulting in incomplete extraction of essential features for multi-view data; and 2) the semantic consistency features between the fused views are not fully considered. Therefore, we introduce a consistent representation model for our algorithm, which seeks a consensus representation from different

perspectives, as follows:

$$
\begin{aligned}
&J = \min \lambda_1 \sum_{v=1}^{V} \sum_{i=1}^{n_o+n_{cp}} \left\| \begin{bmatrix} Y_{o_i}^{(v)} \\ Y_{cp_i}^{(v)} \end{bmatrix} - \begin{bmatrix} H_{o_i}^{(v)} \\ H_{cp_i}^{(v)} \end{bmatrix} Q_i^{(v)} \right\|_F^2 + \\
&\sigma_m \|K_m\|_* + \lambda_2 \Lambda(H^{(v)}, U) \\
&s.t. P_m h = k_m, m = 1, 2, ..., M, H^{(v)} = T \left[\; H_o^{(v)}; H_{cp}^{(v)} \; \right], \\
&\Re_\Omega(Y^{(v)}) = \Re_\Omega(X^{(v)}), {H^{(v)}}^T H^{(v)} = I, \\
&\mathcal{H} = \Psi(H^{(1)}, H^{(2)}, ..., H^{(v)}), U{U^{(T)}} = I, \\
&H^{(v)} \geq 0, Q^{(v)} \geq 0
\end{aligned}
\tag{3.11}
$$

where U denotes the learned consensus representation. This representation can effectively fuse each instance of information from all views. λ_1 and λ_2 are positive penalty parameters for adjusting the impact of each term in all objective functions. The function Λ produces a consistent representation matrix, and $H^{(v)}$ is the incomplete view data processed by filling the low-rank tensor model. $\Re_\Omega$ represents a mapping operation, which maps the complete instance part of the view to matrix $Y_o^{(v)}$ and the incomplete part of the view to matrix $Y_{cp}^{(v)}$.

To form a consistent representation of U, we utilised the measurement formula $\Lambda(H^{(v)}, U)$ to measure the degree of inconsistency between U and $H^{(v)}$ [94]:

$$
\Lambda(H^{(v)}, U) = \left\| \frac{S_U}{\|S_U\|_F^2} - \frac{S_{H^{(v)}}}{\|S_{H^{(v)}}\|_F^2} \right\|_F^2 \tag{3.12}
$$

where S_U is the similarity matrix of U and $S_{H^{(v)}}$ is the similarity matrix of $H^{(v)}$. This function can produce a better fusion representation by minimizing the gap between multi-view and consistent representations. Furthermore, we use the linear kernel, that is, $S_U = UU^T$, which is the standard of a similarity measurement [94]. Based on the fact that $\|S_U\|_F^2 = c$, and $\|S_{H^{(v)}}\|_F^2 = c$, the value of c in the formula is equal to the number of categories of multi-view data. Most of the algorithm settings were the same, and their effectiveness has been verified [94]. we rewrite Equation 3.12 as:

$$
\Lambda(H^{(v)}, U) = \frac{2(c - Tr({H^{(v)}}{H^{(v)}}^T U U^T))}{c^2} \tag{3.13}
$$

Because c is a constant, it can be omitted when calculating to obtain the final consistency expression in Equation 3.14

$$
\Lambda(H^{(v)}, U) = -Tr({H^{(v)}}{H^{(v)}}^T U U^T) \tag{3.14}
$$

3.3.4 Objective Function and Optimization

Putting Equations 3.7, 3.10, and 3.14 together, we can obtain the following objective function:

$$\begin{aligned} J = \min \lambda_1 \sum_{v=1}^{V} \sum_{i=1}^{n_o+n_{cp}} \left\| \begin{bmatrix} Y_{o_i}^{(v)} \\ Y_{cp_i}^{(v)} \end{bmatrix} - \begin{bmatrix} H_{o_i}^{(v)} \\ H_{cp_i}^{(v)} \end{bmatrix} Q_i^{(v)} \right\|_F^2 + \\ \sigma_m \|K_m\|_* - \lambda_2 \sum_{v=1}^{V} Tr(H^{(v)} {H^{(v)}}^T U U^T) \\ s.t. P_m h = k_m, m = 1, 2, ..., M, H^{(v)} = T \begin{bmatrix} H_o^{(v)}; H_{cp}^{(v)} \end{bmatrix} \\ \Re_\Omega(Y^{(v)}) = \Re_\Omega(X^{(v)}), {H^{(v)}}^T H^{(v)} = I, \\ \mathcal{H} = \Psi(H^{(1)}, H^{(2)}, ..., H^{(v)}), U U^{(T)} = I, \\ H^{(v)} \geq 0, Q^{(v)} \geq 0 \end{aligned} \tag{3.15}$$

This objective function is a non-convex function; therefore, it cannot be optimized directly. As such, we can minimize the objective function by iteratively solving the following subproblems: The specific process is as follows.

$H^{(v)}$-sub-problem: By fixing all the other variables, we set the derivative to $H^{(v)}$ as follows:

$$\begin{aligned} D(H^{(v)}) = -2\lambda_1 (Y^{(v)} - HQ) Q^{(T)} - 2\lambda_2 (U U^{(T)} H^{(v)}) + \\ \sum_{m=1}^{M} B_m^{(v)} - \sum_{m=1}^{M} A_m^{(v)} (MI)^{(-1)} \end{aligned} \tag{3.16}$$

$$A_m^{(v)} = \Omega^{(v)}(\alpha_m), B_m^{(v)} = \Omega^{(v)}(k_m). \tag{3.17}$$

It is difficult to use the KKT condition because of the complexity of $D(H^{(v)})$. We chose the traditional gradient descent method to update the data.

$${H^{(v)}}^{t+1} = {H^{(v)}}^t - \tau(D(H^{(v)})). \tag{3.18}$$

where τ denotes the step size. In terms of targeting, there are several M ways to expand the M-order tensor. Our model pair uses unfolding of the three modalities. α_m is the Lagrange multiplier corresponding to the constraint $P_m h = k_m$. Where I is the identity matrix. Operator $\Omega^{(v)}(.)$ only selects $N \times N$ elements corresponding to the v-th views and reshapes them to the $N \times N$ dimensional matrices $A_m^{(v)}$ and $B_m^{(v)}$ corresponding to $H^{(v)}$.

h-sub-problem: We update h directly for each element in $H^{(v)}$ by replacing it directly:

$$h_* \leftarrow H^{(v)}. \tag{3.19}$$

$Q^{(v)}$-sub-problem: When other variables are fixed, the subproblem process for updating $Q^{(v)}$ is as follows:

$$\min \sum_{v=1}^{V} \sum_{i=1}^{n_0} \left\| Y_{o_i}^{(v)} - H_{o_i}^{(v)} Q^{(v)} \right\|_F^2 \tag{3.20}$$

$$Q^{(v)} = Q^{(v)} \frac{{H_o^{(v)}}^T Y_o^{(v)}}{{H_o^{(v)}}^T H_o^{(v)} Q^{(v)}}. \tag{3.21}$$

U-sub-problem: When other variables are fixed, the subproblem process for updating U is as follows:

$$\min_{U^T U = I} -\lambda_2 \sum_{v=1}^{V} Tr(H^{(v)} {H^{(v)}}^T U U^T) \Leftrightarrow \tag{3.22}$$

$$\max_{U^T U = I} Tr(\sum_{v=1}^{V} U^T (H^{(v)} {H^{(v)}}^T) U). \tag{3.23}$$

These problems can be computed simply using eigenvalue decomposition. The best solution for variable U is the eigenvector set corresponding to the first c largest eigenvalues of the matrix $(\sum_{v=1}^{V} H^{(v)} {H^{(v)}}^T)$.

K_m-sub-problem: The formula for the sub-problem K_m is as follows:

$$\begin{aligned} K_m^* &= \underset{G_m}{\operatorname{argmin}} \sigma_m \|K_m\|_* + \mu \Phi(\alpha_m, P_m h - k_m) \\ &= prox_{\beta_m}^{tr}(\Omega_{(m)}(P_m h + \alpha_m)). \end{aligned} \tag{3.24}$$

Here, we define $\Phi(\alpha_m, P_m h - k_m) = \frac{1}{2} \|P_m h - k_m\|_F^2 + \langle \alpha_m, P_m h - k_m \rangle$, where $\langle ., . \rangle$ is the inner product of a matrix, and μ is a positive penalty parameter. The $\Omega_{(m)}(P_m h + \alpha_m)$ operator converts the vector $P_m h + \alpha_m$ into a matrix with the corresponding modal expansion. $\beta_m = (\sigma_m / \mu)$ is the threshold value for the soft-threshold operation of the spectrum. $prox_{\beta_m}^{tr}(L) = U \max(S - \beta_m, 0) V^T$ with $L = UST^T$ is the singular value decomposition (SVD) of the matrix L, and the max operation is performed element-wise. Intuitively, the solution is truncated according to the subspace representation tensor $\mathcal{H}$.

k_m-sub-problem: We update k_m by K_m:

$$k_m^* \leftarrow K_m. \tag{3.25}$$

α_m-subproblem: The variable α_m is updated by:

$$\alpha_m^* = \alpha_m + (P_m h - k_m). \tag{3.26}$$

3.4 Theoretical Analysis of IMC-NLT

The complete procedure of IMC-NLT is summarized in Algorithm 1. The values of variables $H^{(v)}$, $Q^{(v)}$, K_m, k_m, α_m, μ, and U are updated iteratively until the number of iterations reaches the maximum, or the difference between the target values in two consecutive steps is less than the set threshold ε. In the following, we analyze the computational cost and the convergence properties.

3.4.1 Computational complexity

Algorithm 1 IMC-NLT

Require: Incomplete multi-view dataset $X^{(v)}$, parameters σ_m, λ_1, λ_2, μ, maximum number of iterations t_{max}, the threshold $\varepsilon = 10^{-7}$.
Ensure: The resulting clusters.
Initialize $\rho = 1.1$,$max_\mu = 10^{10}$;
Initialize $H^{(1)}$,$H^{(2)}$...,$H^{(v)}$;
Initialize $Q^{(1)}$,$Q^{(2)}$...,$Q^{(v)}$;
Initialize K_1=0,...,K_M=0;
Initialize α_1=α_2...,α_M=0;
while not converge **do**
  for $v = 1$ to V **do**
    Update $H^{(v)}$ via Eq. (3.18);
    Update $Q^{(v)}$ by solving Eq. (3.21);
  end for
  Update K_m via Eq. (3.24);
  Update k_m by solving Eq. (3.25);
  Update α_m via Eq. (3.26);
  Update μ by $\mu = \min(\rho\mu; \max_\mu)$;
  Update U via Eq. (3.23);
  Check the convergence conditions:
  $\left|\frac{J_{t+1}-J_t}{J_t}\right| < \varepsilon$;
  Until (3.15) reaches the maximum number of iterations t_{max} or convergence.
end while
Apply K-means to U for producing the resulting clusters.

We calculated complexity in five steps. In the first step of mechanical filling, the time cost was $O(\sum_{v=1}^{V} n_o^v d_v)$. In the second step of the SVT operation on the tensor, we use Lemma 1 for the complexity calculation:

Lemma 1 [95]. Let $A = QB \in \mathbb{R}^{m\times n}$, where $Q \in \mathbb{R}^{m\times n}$ has orthonormal columns. Then, we have:

$$\mathbb{S}_\tau(A) = Q\mathbb{S}_\tau(B), \tag{3.27}$$

where $\mathbb{S}_\tau(.)$ is the SVT operator with time cost $O(n^3)$. In order to reduce the expensive cost, we perform the SVT operation on the smaller matrix $B \in \mathbb{R}^{n\times n}$ instead, if the matrix $Q \in \mathbb{R}^{m\times n}$ is available($m \geq n$). In the third step of recomposing the tensors into vectors, the time cost is $O(\sum_{v=1}^{V} n^v c)$. In the fourth step of derivative gradient descent, the vector reconstitution tensors can be approximated as $O(n^2c + cnd_v)$. The time cost is $O(n^3)$ for the fifth step of calculating eigenvalues and eigenvectors.

Therefore, the computational complexity of IMC-NLT proposed in this study is $O(\sum_{v=1}^{V} n_o^v d_v + n^v c + l(n^3 + n^2c + cnd_v + n^3))$, where l is the iteration number. Because c is usually much smaller than d_v and n, the overall complexity can be approximated as $O(ln^3)$. To ensure that the algorithm achieves faster convergence and better experimental results, we obtained $H^{(v)}$ and $Q^{(v)}$ through the non-negative matrix factorization of $Y^{(v)}$.

3.4.2 Convergence Analysis

In this section, we demonstrate the convergence of the IMC-NLT iterative algorithm.

Theorem 1. The objective function of the IMC-NLT: $J(H^{(v)}, Q^{(v)}, K_m, U, \alpha_m)$ in (3.15) is bounded. The proposed optimization algorithm monotonically reduced the value of the objective function.

Proof. Because this is the sum of norms with positive penalty parameters, problem (3.15) is bounded from below: ($H^{(v)}$, $Q^{(v)}$, U) obtained by formulas (3.18), (3.21), and (3.23) are the minimum points corresponding to subproblems (3.16), (3.20), and (3.22). For the subproblem of solving K_m, the fast method of formula (3.24) is used for the approximate matrix inversion calculation. h and k_m are updated by directly substituting the corresponding elements. Intuitively, multiplier α_m is updated according to the updating rule of multipliers.

Let $\{H^{(v)}, Q^{(v)}, K_m, U, \alpha_m\}_{t=1}$ be a bounded monotonically decreasing sequence determined by Theorem 1. According to the bounded monotone convergence theorem [96], the objective function monotonically decreases and is bounded.

Theorem 2. Algorithm 1 converges to a minimum under the updating of the value of $J(H^{(v)}, Q^{(v)}, K_m, U, \alpha_m)$ in each optimization step.

Proof. Suppose that $\{H^{(v)}, Q^{(v)}, K_m, U, \alpha_m\}_t$ and $\{H^{(v)}, Q^{(v)}, K_m, U, \alpha_m\}_{t+1}$ represent the iterative sequence of the (t) and $(t+1)$ times of problem (3.15), respectively. According to the previous sub-problem optimization steps, we can conclude that these sub-problems are not only convex optimization problems but also have closed solutions. By solving the above sub-problems one by one, we can get the following formula:

$$
\begin{aligned}
&J(H_t^{(v)}, Q_t^{(v)}, (K_m)_t, U_t, (\alpha_m)_t) \geq J(H_{t+1}^{(v)}, Q_t^{(v)}, (K_m)_t, U_t, (\alpha_m)_t) \\
\geq &J(H_{t+1}^{(v)}, Q_{t+1}^{(v)}, (K_m)_t, U_t, (\alpha_m)_t) \geq J(H_{t+1}^{(v)}, Q_{t+1}^{(v)}, (K_m)_{t+1}, U_t, (\alpha_m)_t) \\
\geq &J(H_{t+1}^{(v)}, Q_{t+1}^{(v)}, (K_m)_{t+1}, U_{t+1}, (\alpha_m)_t) \geq J(H_{t+1}^{(v)}, Q_{t+1}^{(v)}, (K_m)_{t+1}, U_{t+1}, (\alpha_m)_{t+1})
\end{aligned} \tag{3.28}
$$

Equation (3.28) proves that the decrement of the objective function is achieved through the iterative updating of variables in sequence $\left\{H^{(v)}, Q^{(v)}, K_m, U, \alpha_m\right\}_t$. So we have completed the proof of Theorem 2.

The above two theorems ensure that by using the proposed optimization method, the objective function is monotonically decreasing and bounded. Meanwhile, the sequence is continuously optimized and can converge to the minimum value.

3.5 Experimental Evaluation

In this section, we evaluate the effectiveness of the IMC-NLT in comparison with the experimental results of the baseline algorithm on five incomplete multi-view datasets.

3.5.1 Datasets

Table 3.1 lists the five datasets widely used in our experiments.

Table 3.1: Statistics of the datasets

Dataset	Clusters	Views	Instances	Features
SensIT300	3	2	300	50/50
Statlog	7	2	2310	9/10
Wisconsin	5	2	265	1703/265
WebKB	2	2	1051	1840/3000
Caltech101-7	7	6	1474	48/40/254/1984/512/928

- **SensIT300**[1]: Data collected from distributed sensors in an intelligent transportation system. A total of 300 instances were divided into three categories, which corresponded to three types of transportation in real life. Each data instance has two information views: sound information recorded by a sensor, and vibration information, in which each view contains 50-dimensional characteristic attributes.

[1] https://github.com/Liuzhenjiao123/multiview-data-sets/blob/master/sensIT300.mat

- **Statlog**[2]**:** An image segmentation dataset was randomly selected from a database of images from seven categories. The images were manually segmented to create classification for each pixel. Collected by the vision group at the University of Massachusetts, this dataset contains 2310 instances with corresponding categories under two views. The characteristic dimension of one view is nine, whereas the characteristic dimension of the other view is 10.

- **Wisconsin**[3]**:** A set of webpages collected from the University of Wisconsin website. The five types of webpages are student, project, course, staff, and faculty. Each has two views: the content view and the reference view. In the content view, each webpage consists of 1703 words. The reference view is described by the reference relationships between a page and other pages.

- **WebKB** [4]**:** A set of course and non-course documents. Each document has two representations: the text content of the webpage and the anchor text with links to other webpages pointing to the webpage. Based on the page representation, 3000 features were selected. For linked representations, 1840 features were generated.

- **Caltech101-7**[5]**:** Caltech101-7 is a subset of the real dataset Caltech101 with seven categories, which are from various categories such as football, camera, and chair. This dataset contains six views: Gabor, WM, Centrist, HOG, GIST, and LBP.

Specifically, we used the SensIT300, Starlog, Wisconsin, and WebKB datasets to build incomplete multi-view data. In our experiments, approximately 10%, 30%, and 50% of the instances were randomly deleted from each view of the four databases. For the accuracy of the results, we perform standard validation on these datasets [97], and produced the results of the average calculation. All of the experimental codes were developed using MATLAB 2015a running on an Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz with 16-GB RAM with the Win 10 system.

3.5.2 Baseline methods

Against the five datasets, IMC-NLT was compared with five IMC methods: IMSC-AGL [63], DAIMC [48], UEAF [98], IMC-GRMF [46], and HCP-IMSC [99].

- **IMSC-AGL** An algorithm first exploits low-rank representations for multi-view adaptive learning of graphs and then uses spectral constraints to obtain better low-dimensional representations.

[2]https://github.com/Liuzhenjiao123/multiview-data-sets/tree/master
[3]https://lig-membres.imag.fr/grimal/data.html
[4]https://github.com/Liuzhenjiao123/dataset4
[5]https://github.com/Liuzhenjiao123/data5

- **DAIMC** Incomplete multi-view clustering algorithm based on weighted semi-non-negative matrix factorization (semi-NMF). It exploits a weight matrix to adapt to a variety of incomplete cases and uses $L_{2,1}$ -norm regularization to obtain a cluster-friendly basis matrix shared by views.

- **UEAF** A unified and robust embedding alignment model for incomplete multi-view clustering. To maintain the consistency of the local semantics of the view and infer incomplete information, it learns the local structure shared by the views by reversing the graph regularization.

- **IMC-GRMF** Incomplete multi-view clustering method based on matrix factorization. For better integration, IMC-GRMF uses the local information of each view to facilitate fusion of the complementary information of views to obtain a shared representation. Orthogonal constraints can effectively handle out-of-sample problems.

- **HCP-IMSC** This is an incomplete multi-view clustering method that is based on hypergraph induction and tensor decomposition. It effectively uses the correlation of high-order information to recover missing data and combines the affinity matrix, tensor decomposition, and missing-view recovery into one framework.

3.5.3 Evaluation metrics

The resulting clusters by the algorithms in the experiments are evaluated by normalized mutual information (NMI) [100], clustering accuracy (ACC) [101], Adjusted Rand index (ARI) [102], and F1 Score (F1) [103]. A higher value of these metrics indicates a higher cluster quality. The NMI is defined as follows:

$$NMI = \frac{\sum_{i=1}^{C}\sum_{j=1}^{C} N_{i,j} \ln \frac{N_{i,j}}{N_i \hat{N}_j}}{\sqrt{(\sum_{i=1}^{C} N_i \ln \frac{N_i}{N})(\sum_{j=1}^{C} \hat{N}_j \ln \frac{\hat{N}_j}{N})}} \tag{3.29}$$

where N is the number of instances in a complete view, N_i is the number of instances in the i-th cluster, $\hat{N}_j$ is the number of instances of the j-th label, and $N_{i,j}$ is the number of samples that exist in both the i-th cluster and j-th clusters. The ACC measures the quality of clusters as follows:

$$ACC = \frac{\sum_{i=1}^{N} \delta(map(r_i), l_i)}{N} \tag{3.30}$$

where r_i is the cluster label of x_i, l_i is the exact class label, and N is the number of samples. When x=y, $\delta(x, y)$ is equal to 1; otherwise, it is 0. Map(r_i) is the optimal permutation mapping function obtained. The ARI is formulated as:

Table 3.2: Mean NMIs, ACCs, ARIs and F1 of different methods on SensIT300 , Statlog and Wisconsin datasets

Dataset	Method \ PER	NMI			ACC			ARI			F1		
		0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
SensIT300	IMSC-AGL	0.23	0.21	0.16	0.66	0.63	0.61	0.25	0.21	0.17	0.51	0.48	0.45
SensIT300	DAIMC	0.21	0.18	0.16	0.64	0.61	0.59	0.23	0.18	0.16	0.49	0.46	0.45
SensIT300	UEAF	0.21	0.18	0.16	0.65	0.61	0.59	0.24	0.18	0.16	0.51	0.45	0.44
SensIT300	IMC-GRMF	0.16	0.08	0.06	0.61	0.51	0.47	0.17	0.08	0.06	0.45	0.39	0.37
SensIT300	HCP-IMSC	0.32	**0.30**	0.21	**0.72**	0.65	0.57	**0.35**	**0.30**	**0.22**	**0.56**	0.50	0.48
SensIT300	IMC-NLT	**0.32**	0.28	**0.25**	0.69	**0.66**	**0.61**	0.31	0.24	0.19	0.54	**0.51**	**0.49**
Statlog	IMSC-AGL	0.44	0.42	0.38	0.55	0.54	0.48	0.79	0.31	0.29	0.43	0.41	0.41
Statlog	DAIMC	0.47	0.41	0.34	0.57	0.51	0.45	0.71	0.27	0.19	0.46	0.39	0.33
Statlog	UEAF	0.49	0.38	0.35	0.48	0.46	0.44	**0.81**	0.24	0.18	0.41	0.36	0.32
Statlog	IMC-GRMF	0.11	0.21	0.14	0.28	0.38	0.31	0.68	0.15	0.08	0.22	0.28	0.22
Statlog	HCP-IMSC	0.51	0.46	0.40	0.58	0.54	0.49	0.38	0.32	0.26	0.48	0.44	0.39
Statlog	IMC-NLT	**0.63**	**0.56**	**0.48**	**0.68**	**0.62**	**0.58**	0.54	**0.45**	**0.35**	**0.61**	**0.53**	**0.45**
Wisconsin	IMSC-AGL	0.21	0.19	0.14	0.43	0.39	0.34	0.17	0.11	0.08	0.51	0.33	0.32
Wisconsin	DAIMC	0.31	0.27	0.24	0.51	0.44	0.46	0.25	0.17	0.17	0.49	0.39	0.39
Wisconsin	UEAF	0.36	**0.41**	0.34	0.61	0.57	0.51	0.35	0.34	0.25	0.51	0.51	0.44
Wisconsin	IMC-GRMF	0.26	0.19	0.11	0.44	0.37	0.33	0.15	0.11	0.05	0.45	0.34	0.31
Wisconsin	HCP-IMSC	0.27	0.24	0.26	0.50	0.40	0.49	0.21	0.17	0.20	0.42	0.39	0.40
Wisconsin	IMC-NLT	**0.48**	0.39	**0.37**	**0.74**	**0.68**	**0.69**	**0.48**	**0.38**	**0.32**	**0.67**	**0.61**	**0.59**
WebKB	IMSC-AGL	0.66	0.31	0.5	0.95	0.83	0.91	0.79	0.41	0.68	0.93	0.77	0.88
WebKB	DAIMC	0.61	0.52	0.42	0.93	0.90	0.85	0.70	0.64	0.51	0.91	0.88	0.84
WebKB	UEAF	0.68	**0.71**	**0.65**	0.95	**0.96**	**0.95**	0.81	**0.82**	**0.79**	0.93	**0.93**	**0.93**
WebKB	IMC-GRMF	0.52	0.34	0.03	0.92	0.92	0.61	0.68	0.52	0.04	0.89	0.84	0.59
WebKB	HCP-IMSC	0.71	0.68	0.61	0.95	0.93	0.92	0.82	0.79	0.77	0.93	0.92	0.91
WebKB	IMC-NLT	**0.71**	0.65	0.53	**0.96**	0.94	0.92	**0.83**	0.77	0.66	**0.94**	**0.93**	0.90

$$ARI = \frac{RI - E[RI]}{(max(RI) - E[RI])} \tag{3.31}$$

where $E[RI]$ resents the expected value of RI and RI is a random index used to measure the similarity between two clusters. It is defined as follows:

$$RI = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.32}$$

where TP is the true positive, TN is the true negative, FP is false positive, and FN is false negative. The F1 is defined as:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{3.33}$$

where $precision = \frac{TP}{TP + FP}$, and $recall = \frac{TP}{TP + FN}$.

3.5.4 Evaluations on Clustering Performance and Discussion

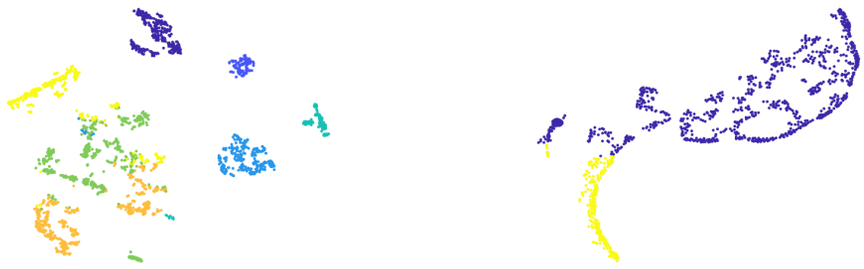
In our experiment, we selected 10%, 30%, and 50% of the total number of instances to randomly delete them from each view of the five datasets. Table 3.2 shows the average

Table 3.3: Mean NMIs, ACCs, ARIs and F1 of different clustering methods on SensIT300 , Statlog and Wisconsin datasets

Dataset	Method \ PER	NMI			ACC			ARI			F1		
		0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
SensIT300	K-means	0.32	0.28	0.25	0.69	0.66	0.61	0.31	0.24	0.19	0.54	0.51	0.49
SensIT300	Fuzzy	0.32	0.27	0.25	0.68	0.61	**0.64**	0.28	0.21	**0.22**	0.54	0.49	0.49
SensIT300	Spectral	**0.35**	**0.31**	**0.27**	**0.71**	**0.68**	0.62	**0.33**	**0.31**	0.21	**0.57**	**0.56**	**0.51**
Statlog	K-means	**0.63**	**0.56**	**0.48**	**0.68**	**0.62**	**0.58**	**0.54**	**0.45**	**0.35**	**0.61**	**0.53**	**0.45**
Statlog	Fuzzy	0.43	0.27	0.25	0.51	0.36	0.36	0.31	0.13	0.14	0.41	0.29	0.29
Statlog	Spectral	0.62	0.44	0.36	0.54	0.45	0.33	0.38	0.28	0.09	0.51	0.38	0.29
Wisconsin	K-means	**0.48**	**0.39**	**0.37**	**0.74**	**0.68**	**0.69**	**0.48**	0.38	**0.32**	**0.67**	**0.61**	**0.59**
Wisconsin	Fuzzy	0.42	0.35	0.31	0.59	0.45	0.46	0.39	0.28	0.21	0.57	0.42	0.42
Wisconsin	Spectral	0.40	0.39	0.33	0.58	0.66	0.59	0.36	**0.43**	0.32	0.54	0.61	0.51
WebKB	K-means	**0.71**	**0.65**	0.53	**0.96**	**0.94**	0.92	**0.83**	**0.77**	0.66	**0.94**	**0.93**	0.90
WebKB	Fuzzy	0.69	0.55	0.54	0.96	0.93	**0.93**	0.83	0.72	**0.71**	0.94	0.91	**0.91**
WebKB	Spectral	0.68	0.54	**0.89**	0.96	0.92	0.92	0.82	0.66	0.66	0.94	0.90	0.89

Table 3.4: Two incomplete multi-view clustering methods based on tensor models show different performances in terms of ACC, F1, Running time (seconds), and computational complexity with 90% incomplete instances of each view on the Caltech101-7 dataset

Dataset	Method	ACC	F1	Running time (seconds)	computational complexity
Caltech101-7	HCP-IMSC	0.37	0.40	106.04	$O(Vn^3 + V(n-n_o)^3 + cnVlog(V) + cn^2V)$
Caltech101-7	IMC-NLT	**0.43**	**0.45**	**92.75**	$O(n^3)$

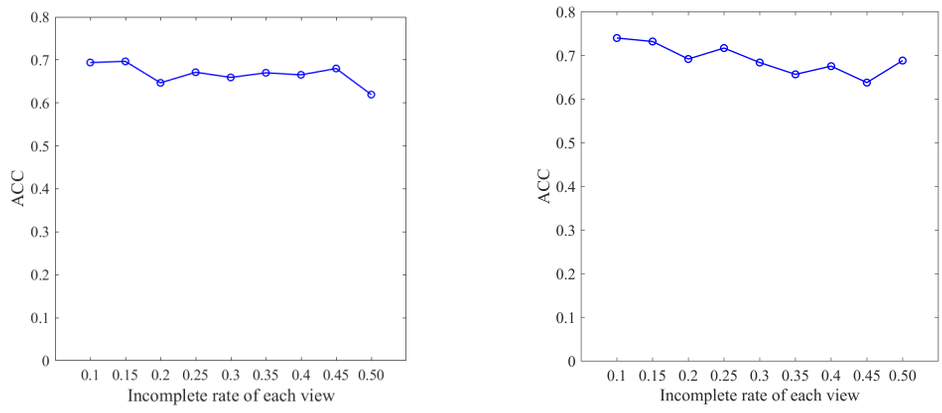


(a) (b)

Figure 3.4: Cluster structure illustration on two incomplete multi-view datasets. (a) Statlog with 10% incomplete instances of each view, (b) visualization on WebKB with 10% incomplete

performance of NMI, ACC, ARI, and F1 for the five different methods on the five incomplete multi-view datasets. From this table, we can observe the following results.

1) The performance of HCP-IMSC was better than that of other incomplete multi-view clustering methods on most datasets. This indicates that the tensor-based model can achieve incomplete multi-view clustering, which preserves the high-order correlation advantage. In addition, compared to HCP-IMSC, our IMC-NLT can produce good results from most



(a) (b)

Figure 3.5: Robust performance on two incomplete multi-view datasets: (a) SensIT300 (b) Wisconsin

datasets with high missing rates. This verifies that building a tensor model directly from prepopulated modal data can better recover missing data and reduce the impact of noise.

2) Compared with the other five algorithms, the IMC-NLT has obvious advantages. For example, on the statlog and Wisconsin datasets, according to various clustering indicators, IMC-NLT performs the best. When the missing rate of the Wisconsin dataset was set to 10%, our algorithm achieved an ACC score that was approximately 13% higher than that of the second-best method. Although the advantages of our algorithm on SensIT300 and Webkb are not as prominent as those of the above datasets, the difference between our method and other superior algorithms is not evident in most cases. The IMC-NLT is relatively stable for datasets with different missing rates.

3) From Table 3.2, we can see that our method is superior to other views based on recovery methods such as UEAF. This shows that IMC-NLT not only effectively utilises the specific information of each view but also builds a unified structure to effectively maintain the semantic relationships among the different views. Therefore, IMC-NLT can capture useful information with complex interactions between views to recover the missing data.

4) The proposed IMC-NLT method was more robust than the DAIMC, IMSC-AGL, and IMC-GRMF methods. Our method neither indirectly obtains consensus representations from individual expressions of all viewpoints, such as DAIMC, nor is it constrained by the existing incomplete data. At the same time, it prevents the IMSC-AGL from extracting hidden information from incomplete and complex data. Our method is suitable for various complex missing situations, and ensures the filling effect of complex missing situations through an effective filling mechanism.

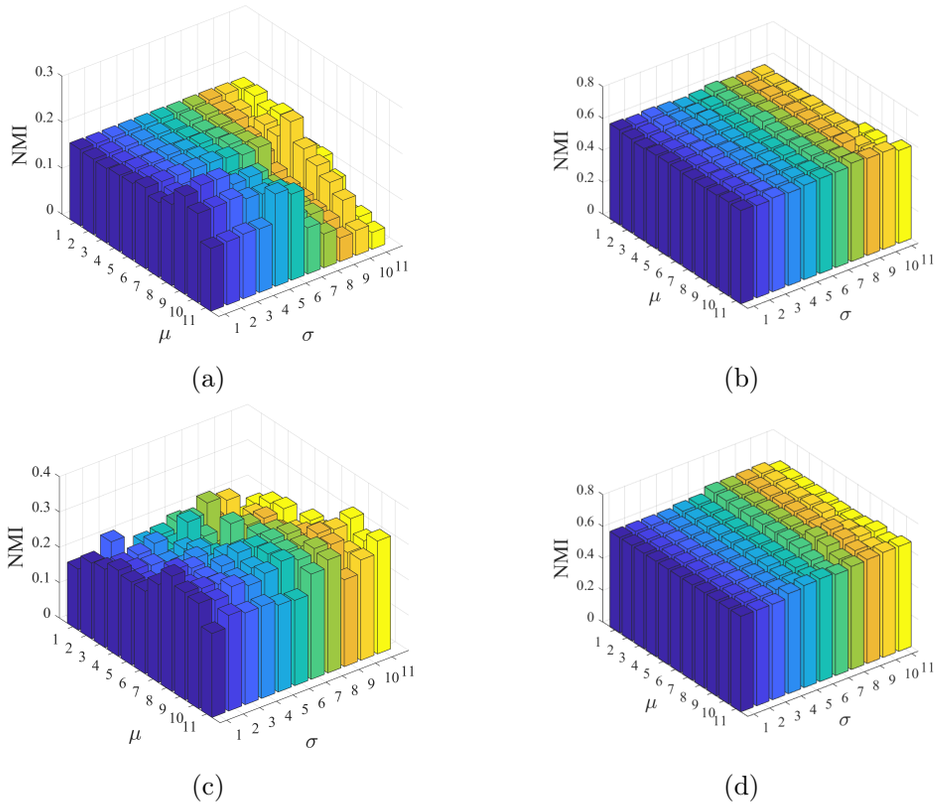


Figure 3.6: NMI versus parameters σ and μ of IMC-NLT on different datasets with various percentages of incomplete instances of each view (a) SensIT300 with 50%, (b) Statlog with 10%, (c) Wisconsin with 50%, and (d) WebKB with 10% .

3.5.5 Clustering Performance on Three Clustering Methods

Table 3.3 reports the performance of the fusion effect on three clustering algorithms. Compared with other incomplete multi-view clustering algorithms, the overall experimental result of applying K-means is the best, followed by Fuzzy clustering and spectral clustering. Again, our algorithm has demonstrated excellent performance in incomplete multi-view data filling and later fusion.

3.5.6 Visualization of Clustering Results

To further demonstrate the advantages of the IMC-NLT algorithm, we visualised the distributions of experimental clustering results. For example, on the Statlog and WebKB datasets, the IMC-NLT algorithm can produce different colour clusters with obvious grouping structures, as illustrated in Figure 3.4 .

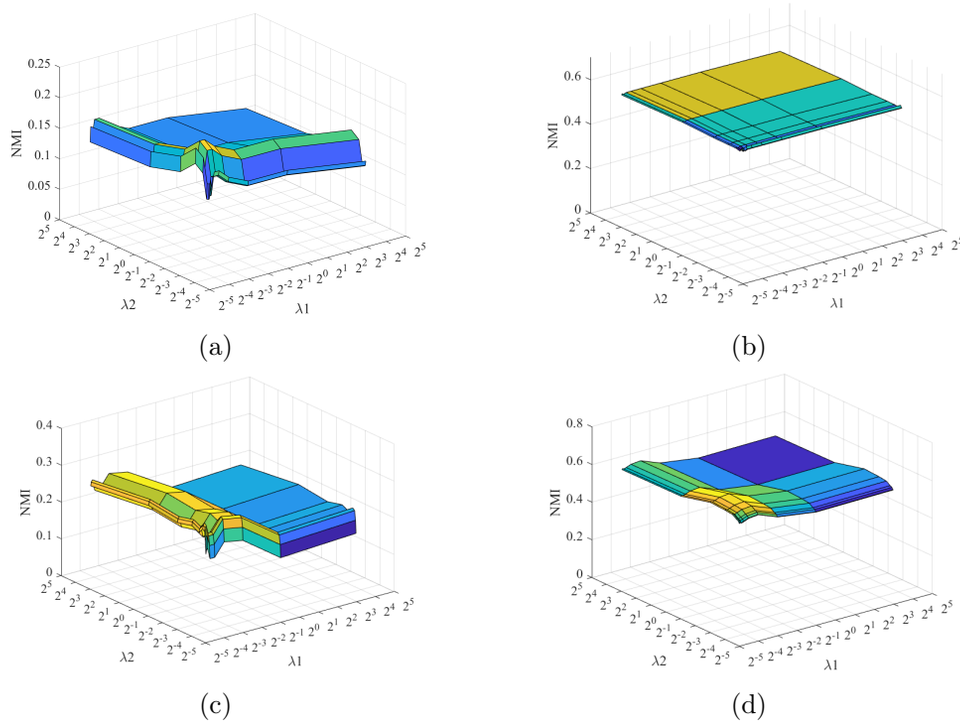


Figure 3.7: NMI versus parameters λ_1 and λ_2 of IMC-NLT on the four datasets with various percentages of incomplete instances with each view, respectively. (a) SensIT300 with 50%, (b) Statlog with 10%, (c) Wisconsin with 50%. and (d) WebKB with 10%.

3.5.7 The effectiveness of IMC-NLT on large-scale datasets

The performances of the two methods for incomplete multi-view clustering based on the tensor model are presented in Table 3.4. From this table, we observe that our method demonstrates certain advantages in terms of clustering indicators, running time, and complexity. In addition, the proposed method performed well on larger datasets.

3.5.8 Robustness Analysis

To investigate the robustness of the proposed algorithm, we examined the clustering results of the IMC-NLT algorithm on SensIT300 and Wisconsin datasets with different missing rate intervals of 5%, as shown in Figure 3.5. In terms of accuracy, our method remains relatively robust as the missing rate of data points increases. This shows the capacity of the IMC-NLT to handle missing data filling.

3.6 Parameters and Convergence of IMC-NLT

3.6.1 Parameter sensitivity analysis

This section presents several comparative experiments to demonstrate the effects of different IMC-NLT parameter values. We focus mainly on the following parameters of IMC-NLT: the non-negative matrix decomposition control parameter λ_1, feature orthogonal constraints to limit the control parameter λ_2, and low-rank force coding. We set the M parameters inside to be equal, that is, $\sigma_1 = .. = \sigma_M = \sigma$, and accordingly, tune the parameter σ and Lagrange operator control parameter μ.

1) Parameters σ and μ: We show the NMI versus the two parameters σ and μ on the datasets of SensIT300, Statlog, Wisconsin, and WebKB with different incomplete-view rates in Figure 3.6. For example, on the Statlog dataset, the experimental results showed that the best performance was achieved when σ ranges between {0,9} and μ between {1,11}. In this study, we find that our algorithm achieves the best clustering when σ takes values in the range of {1,6} and μ in the range of {8,10} by using a latticework search. Meanwhile, on the Statlog and WebKB datasets, IMC-NLT exhibited low sensitivity to its parameters.

2) Parameters λ_1 and λ_2: From Figure 3.7, we can observe that parameters λ_1 and λ_2 are insensitive to the Statlog and WebKB datasets. For SensIT300, the algorithm performs well when the parameters λ_1 and λ_2 are between $\{2^0,2^1\}$ and $\{2^{-1}, 2^1\}$, respectively. Similarly, there is a certain parameter sensitivity in the Wisconsin dataset. When the parameters λ_1 and λ_2 are between $\{2^{-2},2^0\}$ and $\{2^1, 2^2\}$, respectively, IMC-NLT achieves good performance.

The adaptive selection of various parameters for different datasets to reach their optimal values is problematic. Determine the most suitable parameters for the proposed IMC-NLT model. We solve this problem by choosing a combination strategy to find the optimal parameters. Specifically, we first fixed the insensitive parameters to set λ_1 and λ_2 with a fixed value range, and then ran IMC-NLT with different values of σ and μ. As such, the optimal parameter value set of the algorithm is obtained, experiments are conducted, and the results are compared and reported.

3.6.2 Convergence analysis

To better deal with the complex objective function, we split the objective function into several subproblems and use an iterative optimization algorithm to monotonically decrease it until convergence. The objective function values are plotted in Figure 3.8 according to the corresponding NMI with the number of iterations (within 30 iterations). As shown, the loss of the objective function decreases monotonically and converges to a stationary point. This ensures the convergence of the proposed optimization method.

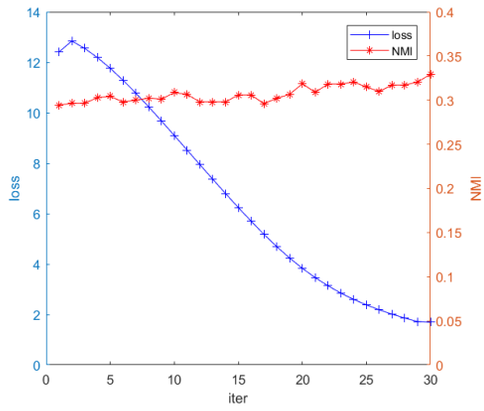


Figure 3.8: The objective function loss and NMI v.s. iterations on (a) SensIT300 with 10% incomplete instances of each view.

3.7 Conclusion

This section presents a novel algorithm for incomplete multi-view clustering, named IMC-NLT, which is based on Non-negative Matrix Factorization (NMF) and low-rank tensor fusion. IMC-NLT relies on both the modal unified dimensional structure and low-rank tensor. IMC-NLT can effectively integrate the information hidden in datasets with the same view and between views. Moreover, IMC-NLT not only deals effectively with various incomplete data but also has low sensitivity to its parameters. We have carried out comprehensive experiments on the five representative data sets by comparing IMC-NLT with state-of-the-art algorithms. The experimental results have shown that our method can achieve good clustering results with stability.

As a future extension, this research aims to integrate the low-rank tensor filling model with deep learning models to obtain more efficient view representations for incomplete multi-view clustering.

Chapter 4

Incomplete Multiview Co-Clustering by Sparse Low-Rank Representation

Contents

4.1 Introduction

In recent years, the availability of multiview data has increased significantly. However, the success of the current multiview learning approaches heavily relies on complete and consistent data from various sources [104] [105] [106] [3] [107] [108]. Unfortunately, in real-world scenarios, multi-view data often suffer from incompleteness due to various factors, including missing views and data points. Moreover, the degree of such data incompleteness varies depending on the context. For instance, during land multi-view data acquisition, sensor acquisition failure can easily lead to incomplete optical and SAR image view data. In the medical field, collecting multi-view medical data with missing records of patients with different degrees is often necessary due to privacy concerns. To tackle these issues, incomplete multiview clustering (IMVC) fuses information between multiple views of complex missing data. However, IMVC is challenging in real-world scenarios. Many approaches have been proposed to address IMVC to date, which can be grouped into shallow and deep multiview clustering models.

Shallow IMVC models are limited to low-level features, which can result in samples representing the same object being restricted to the same potential representation in the potential subspace [43] [44] [109]. In contrast, deep IMVC models employ a view-specific depth encoder and a graphics embedding strategy to effectively capture high-level features and local structure of each view simultaneously [110]. While both shallow and deep approaches to IMVC have been successful in real-world applications, there are still several limitations that need to be addressed. First, many existing approaches fail to consider both global and local incomplete multiview information. Second, although data recovery and cluster partitions are essential for IMVC, many existing methods do not combine them effectively. Third, existing deep learning-based IMVC methods have issues with training stability and clustering performance. Finally, some existing approaches to IMVC do not demonstrate their robustness on datasets with missing views and varying degrees of missing data points.

To address the limitations discussed above, we propose a novel framework called Incomplete Multiview Co-Clustering by Sparse Low-Rank Representation (CCIM-SLR). Our approach learns a low-rank sparse representation matrix for each view, which is then used to fill in missing samples within each view. This process is achieved through the use of association information between missing and observed samples within a particular view, as well as by considering the association relationships between samples in different views through learning common subspace representations. Furthermore, by introducing the filled view data and the learned implicit view data in a clustering process, the data recovery process and the clustering process can complement each other and lead to improved clustering

results. To ensure even more accurate data recovery, we use the adjustable low-rank approximation representation model Γ-norm, which replaces the traditional kernel norm that can only produce low-rank feature representations under certain conditions.

In summary, the contributions of this paper are as follows.

- This paper presents CCIM-SLR, a novel approach to incomplete multi-view clustering that leverages a low-rank sparse representation matrix to recover data from missing samples. CCIM-SLR utilizes association information between missing samples and observed samples within views, as well as their association information between views.
- To improve the stability of clustering results for multi-view data with different missing degrees, CCIM-SLR uses the Γ-norm model, which is an adjustable low-rank representation method. Γ-norm shows the accuracy of achieving a low-rank representation and the stability of data recovery.
- CCIM-SLR learns both a visible view and a hidden view within a co-learning framework in an end-to-end manner, using a mutual interplay between the view data recovery and a clustering process. This approach avoids the need for post-processing steps such as k-means for final clustering assignment results.
- CCIM-SLR has been validated through both theoretical proofs and experiments. Based on the experimental results, CCIM-SLR outperformed state-of-the-art approaches on the five incomplete multiview datasets. The robustness of CCIM-SLR has been demonstrated through experiments on incomplete multi-view datasets with different missing rates of data points.

The remainder of this paper is organized as follows: Section 4.2 provides a review of related work, while Section 4.3 outlines the proposed method and methodologies used in this work. Section 4.4 reports the experimental results of our CCIM-SLR, together with comparisons of other methods. This paper concludes in Section 4.5.

4.2 Related Work and Background

In this section, we review related work and describe some background on IMVC.

4.2.1 Related work

As mentioned before, the current approaches to IMVC can be grouped into shallow and deep ones. As a shallow approach, Guo J et al. [111] proposed a simple and easy-to-implement method that can reconstruct samples and intersample relations through anchors

and fully integrates intraview and interview similarities. In the literature, a weighted semi-nonnegative matrix factorization-based method was proposed to reduce the influence of view incompleteness in clustering [48]. Wen et al. [112] developed a new graph regularization matrix decomposition model to consider the local geometric information and the unbalanced resolution of incomplete multiview observations. Wen et al. [14] proposed introducing a local retention reconstruction term to infer missing views so that all views can be naturally aligned and adding an adaptive weighting strategy for capturing the importance of different views. In [113], feature space-based missing-view inference and manifold space-based similarity graph learning were proposed to better explore the potential information of missing views. Liang et al. [114] developed a reproduced representation; on this basis, a set of incomplete graphs was used to make full use of the geometric structure of the data. Yin et al. [115] introduced a cosine similarity metric to further enhance the preservation of the flow structure of the original multiview, called incomplete multiview clustering with cosine similarity (IMCCS).

As for deep approaches to IMVC, Xu et al. [67] designed an adversarial incomplete multiview clustering (AIMC) method that captures the overall structure and obtains a deeper semantic understanding by seeking the common potential space of multiview data and inferring incomplete data at the same time. Xu et al. [116] proposed to establish a new multi-view clustering complementarity mechanism that can obtain supplementary information and be regarded as supervisory information with high confidence. Therefore, this method achieves the consistency information of multi-view clustering. To solve the incomplete multiview problem by explicitly generating the data of missing views, Wang Q et al. [117] applied adaptive fusion and a cycle consistency generation model for incomplete multiview clustering. The deep IMVC model can take into account high representation ability and save time and space. For example, in [118], an adaptive partial graph learning and fusion (APGLF) method was proposed to capture the local data structure of both within-view and cross-view. A generative adversarial network-based model was proposed by Wang et al. [76], which can effectively generate incomplete view data and capture better common structures in IMVC.

While both shallow and deep approaches to IMVC have been successful in real-world applications, there are still several limitations that need to be addressed. First, many existing approaches fail to consider both global and local incomplete multi-view information. Second, although data recovery and cluster partitions are essential for IMVC, many existing methods do not combine them effectively. Third, existing deep learning-based IMVC methods have issues with training stability and clustering performance. Finally, some existing approaches to IMVC do not demonstrate their robustness on datasets with missing views and varying degrees of missing data points.

Except for the shallow and deep approaches to IMVC, the incomplete multiview clustering algorithm with low-rank sparsity [119] [120] and the multiview algorithm with one-step clustering [121] [122] [123] have shown good advantages in the field of the multiview study. However, these approaches can further be improved by overcoming some limitations as mentioned before. Before presenting our proposed approach of CCIM-SLR, we need to provide two background works: 1) sparse low-rank representation through multiview subspace (SRRS) learning; and 2) multiview clustering with the cooperation of visible and hidden views.

4.2.2 Sparse low-rank representation through multiview subspace (SRRS) learning

SRRS has a significant effect on incomplete multiview data recovery [119]. For a dataset of incomplete multiview $X = \{X^{(1)}, X^{(2)}, ...X^{(v)}\}$, SRRS imputes missing values of all views by the following expression:

$$\begin{aligned} &\min_{\{R^{(v)},E^{(v)}\},H} \sum_{v=1}^{s} \mathcal{B}(R^{(v)}, E^{(v)}) \\ &s.t. \forall v, \mathcal{P}_v(\widehat{R}^{(v)} {X_o}^{(v)}) = H + E^{(v)}, H^\top H = I \end{aligned} \tag{4.1}$$

where $R^{(v)} \in \mathbb{R}^{\bar{m}^{(v)} \times m^{(v)}}$ is the sparse low-rank representation matrix, $\bar{m}^{(v)}$ is the missing samples, and $m^{(v)}$ is the observed samples. $E^{(v)} \in \mathbb{R}^{m \times t}$ is the noise matrix, m is the number of all samples (m=$\bar{m}^{(v)}$+$m^{(v)}$), t is the unified dimension of the subspace, $X_o^{(v)}$ is the matrix constructed from the observed samples, and $\widehat{R}^{(v)} \in \mathbb{R}^{m \times m^{(v)}}$ is composed of the matrix $R^{(v)}$ and the identity matrix constructed from the indices of the observed samples. $\mathcal{P}_v(\widehat{R}^{(v)} X_o^{(v)})$: $\mathbb{R}^{m \times d^{(v)}} \rightarrow \mathbb{R}^{m \times t}$ represents an operator that projects the samples of all views into the corresponding subspace. SRRS learns that the common representation of all views is $H \in \mathbb{R}^{m \times t}$, and adding constraints to H can effectively avoid trivial solutions. $\mathcal{B}(R^{(v)}, E^{(v)})$ is defined as follows:

$$\mathcal{B}(R^{(v)}, E^{(v)}) = \lambda_1 \|R^{(v)}\|_1 + \lambda_2 \|R^{(v)}\|_* + \lambda_3 \|E^{(v)}\|_1 \tag{4.2}$$

where λ_1, λ_2 and λ_3 are the parameters and $\|R^{(v)}\|_1$ and $\|R^{(v)}\|_*$ represent the sparse and low-rank constraints on $R^{(v)}$, respectively. To reduce the influence of the noise matrix and make the subspace representations robust, we add a sparse constraint to E. SRRS is a technique that can impute missing values by taking into account both intraview and inter-view relations.

4.2.3 Multiview clustering with the cooperation of visible and hidden views

Another important work is the multiview clustering model, called multiview clustering with the cooperation of visible and hidden views (MV-Co-VH) [121]. MV-Co-VH is a clustering method that integrates and optimizes both visible and hidden views:

$$\begin{aligned} &\min J(U, Z, \widetilde{Z}, w) = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \|h_j - \widetilde{z}_j\|^2 \\ &+ (1-\lambda) \sum_{v=1}^{s} w_{(v)} \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \|{x_j}^{(v)} - {z_j}^{(v)}\|^2 + \eta \sum_{v=1}^{s} w_{(v)} \ln w_{(v)} \\ &s.t. \sum_{i=1}^{C} u_{ij} = 1, u_{ij} \in (0,1), 1 \leqslant j \leqslant m \\ &\sum_{v=1}^{S} w_{(v)} = 1, 0 \leqslant w_{(v)} \leqslant 1, H \geq 0 \end{aligned} \tag{4.3}$$

where λ is a parameter, matrix $U \in \mathbb{R}^{c \times m}$ is the cluster indicator matrix, c denotes the number of categories, and m represents the number of samples. If u_{ij}=1, sample j belongs to cluster i; otherwise, u_{ij} =0. Z=$\{Z^{(1)}, Z^{(2)}, Z^{(3)}.....Z^{(v)}\}$ is the cluster center matrix for each view. w=$[\omega_1, \omega_2, ...\omega_s]$ contains the weight of each view. $H \in \mathbb{R}^{m \times t}$ denotes the shared hidden view of all views. t is the sample dimension of the hidden view. $\widetilde{Z}$ represents the corresponding clustering center matrix from the hidden view. From (4.3), MV-Co-VH is a method for extracting hidden views from multiview data through nonnegative matrix factorization. It is also a multiview clustering framework that combines explicit and implicit views to obtain clustering results in one step.

4.3 The Proposed CCIM-SLR

In this section, we describe our CCIM-SLR method, which learns recovery data from the global and local structures of their original data with incomplete multiview. Apart from that, a one-step clustering strategy is also adopted to produce clustering results that combine shared hidden space and visible view effectively. CCIM-SLR comprises two parts: 1) shared hidden subspace learning based on SRRS; and 2) incomplete multiview co-clustering by SRRS.

The components and the pipeline of CCIM-SLR are illustrated in Figure 4.1. As shown, the proposed framework mainly includes three key modules, i.e., the data input module, missing data recovery module, and collaborative learning module. These modules are described in detail in the following sections.

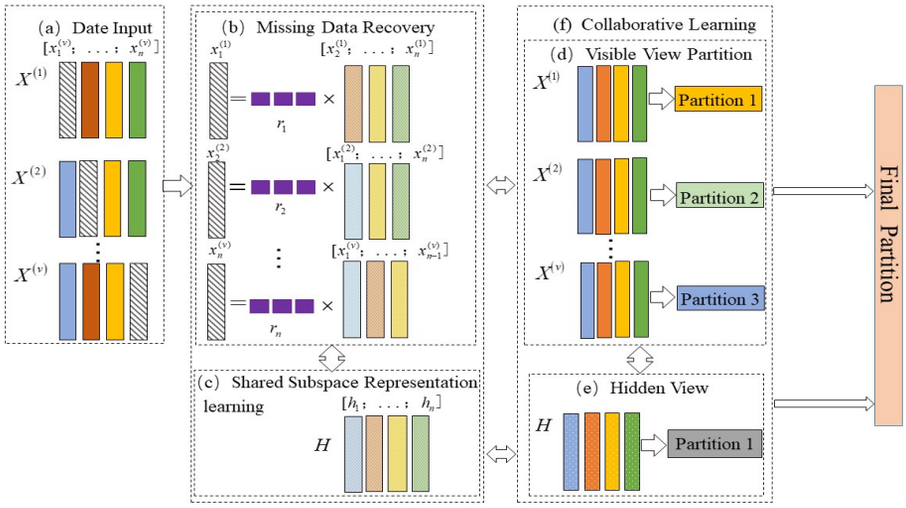


Figure 4.1: Overview of CCIM-SLR. CCIM-SLR consists of six major components: (a) Date Input; (b) Missing Data Recovery; (c) Shared Subspace Representation Learning; (d) Visible View Partition; (e) Hidden View; and (f) Collaborative Learning.

4.3.1 Shared hidden subspace learning based on SRRS

The low-rank sparse representation-based methods have widely been used to recover missing data [124] [125] [126]. As such, we introduce the γ-norm [127], which is a nonconvex low-rank representation method. Specifically, the γ-norm for matrix R is defined as follows:

$$\|R\|_{\gamma} = \sum_{i=1} \frac{(1+\gamma)\epsilon_{R_i}}{\gamma + \epsilon_{R_i}} \tag{4.4}$$

In Equation (4.4), γ is a penalty parameter, and the i-th singular value of the SVD decomposition of the matrix R is denoted as ϵ_i. However, the γ-norm applies only to the processing of single-view data. To extend to multiview data, we introduce the Γ-norm [120] to implement the low-rank constraint on matrix $R^{(v)}$, with the following formula:

$$\|R^{(v)}\|_{G} = \sum_{v=1}^{s} \frac{\epsilon_i^{(v)}}{\epsilon_i^{(v)} + \gamma} \tag{4.5}$$

where $\epsilon_i^{(v)}$ is the i-th singular value of matrix $R^{(v)}$, and if $\gamma \to 0$, we have $\|R^{(v)}\|_{G} \to$ rank $(R^{(v)})$. An appropriate value of γ (e.g., $\gamma = 0.001$) is chosen in such as way that a value that is closer to the genuine rank is obtained. Figure 4.2 shows the approximate ranks obtained with different norms. The experimental results show that when γ=0.001, the obtained results are closest to the true rank. To obtain the incomplete multiview shared hidden subspace of data, we combine the above terms into one model as follows:

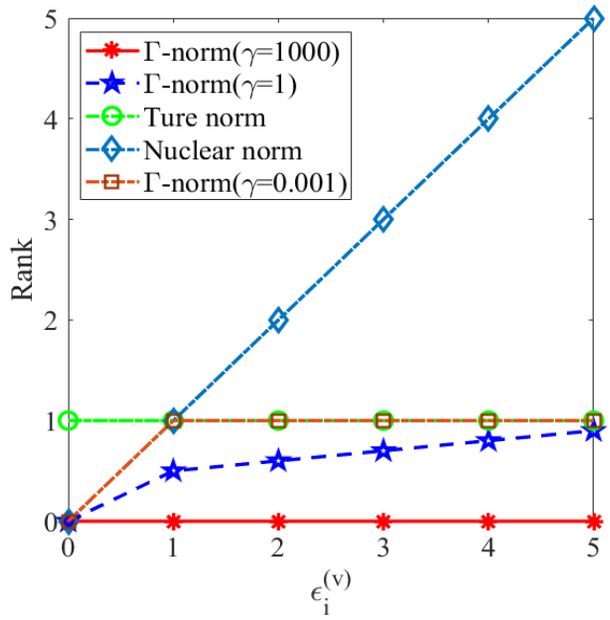


Figure 4.2: The performance of different functions on rank estimation changes with the change in positive singular value $\epsilon_i^{(v)}$ (true rank is 1).

$$\begin{aligned} &\min_{\{R^{(v)},E^{(v)},A^{(v)},H\}} \sum_{v=1}^{s} \lambda_1\|R^{(v)}\|_1 + \lambda_2\|R^{(v)}\|_G + \lambda_3\|E^{(v)}\|_{2,1} + \lambda_4\|A^{(v)}\|_{2,1} \\ &s.t. \forall v, \mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)}) = H + E^{(v)}, H^\top H = I, X^{(v)} = \widehat{R}^{(v)}X_o^{(v)} \\ &\quad H \geq 0 \end{aligned} \tag{4.6}$$

where λ_1, λ_2, λ_3 and λ_4 are the weight parameters for the data $X = \{X^{(1)}, ...X^{(v)}\}$, $X^{(v)} = [X_{cp}^{(v)}; X_o^{(v)}]$ denotes multiview data, $X_{cp}^{(v)} \in \mathbb{R}^{\bar{m}^{(v)} \times d^{(v)}}$ represents the matrix consisting of missing samples, $X_o^{(v)} \in \mathbb{R}^{m^{(v)} \times d^{(v)}}$ represents the matrix consisting of observed samples, $d^{(v)}$ is the sample dimension of the v-th view, $\bar{m}^{(v)}$ is missing samples, and $m^{(v)}$ is the observed samples. The total number of samples is m (m=$\bar{m}^{(v)}$+$m^{(v)}$). $H \in \mathbb{R}^{m \times t}$ stands for the obtained subspace representation of a hidden view, and t is the unified dimension of the subspace. The orthogonal constraint $H^\top H = I$ makes the bases independent of each other.

To fully utilize the data observed in other views to recover missing data, we formulate the missing sample linear reconstruction as Equation (4.7):

$$\forall v, X^{(v)} = \widehat{R}^{(v)}X_o^{(v)} \tag{4.7}$$

where $\widehat{R}^{(v)} \in \mathbb{R}^{m \times m^{(v)}}$ is composed of the matrix $R^{(v)}$, and the identity matrix constructed

from the indices of the observed samples. To acquire a better representation of subspace H, we utilize the following operation:

$$\begin{aligned}\mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)}) &= \widehat{R}^v X_o^{(v)} A^{(v)} \\ &= H + E^{(v)}\end{aligned} \tag{4.8}$$

where $E^{(v)} \in \mathbb{R}^{m\times t}$ is the noise matrix from the original space to subspace $H^{(v)}$, and $A^{(v)} \in \mathbb{R}^{d^{(v)}\times t}$ is the linear transformation matrix that converts the original spatial data to subspace $H^{(v)}$. To guarantee the sparsity of the data, we add L_1 constraints [128] to the matrix R. In addition, the $L_{2,1}$ norm third and fourth terms of Equation (4.6)) are added to discard irrelevant features.

Similar to the method proposed by [119], our proposed model (Equation (4.6)) recovers data using constraints on linear representations. Most importantly, Equation (4.6) introduces the Γ-norm to impose low-rank constraints on the linear representation matrix. Such a cutting-edge nonconvex low-rank representation method can better capture the correlation between samples than the kernel norm.

4.3.2 Incomplete multiview co-clustering by sparse low-rank representation

Most of the existing incomplete multiview learning methods based on subspaces are two-step multiview clustering methods (Step 1: subspace acquisition, and Step 2: clustering). These methods do not effectively combine the clustering process with the filling process. Therefore, we integrate clustering results and shared hidden subspace learning into the same objective function as expressed in Equation (4.9)).

$$\begin{aligned}&\min_{\{R^{(v)},E^{(v)},A^{(v)},H,Z^{(v)},\widetilde{Z}\}} \lambda\sum_{i=1}^{c}\sum_{j=1}^{m} u_{ij}\|h_j - \widetilde{z}_i\|^2 \\ &+ (1-\lambda)\sum_{v=1}^{s}\sum_{i=1}^{c}\sum_{j=1}^{m} u_{ij}\|{x_j}^{(v)} - {z_i}^{(v)}\|^2 \\ &+ \sum_{v=1}^{s}(\lambda_1\|R^{(v)}\|_1 + \lambda_2\|R^{(v)}\|_G + \lambda_3\|E^{(v)}\|_{2,1} + \lambda_4\|A^{(v)}\|_{2,1}) \\ s.t.&\forall v, \mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)}) = H + E^{(v)}, H^\top H = I, X^{(v)} = \widehat{R}^{(v)}X_o^{(v)} \\ &H \geq 0, \sum_{i=1}^{c} u_{ij} = 1, u_{ij} \in (0,1), 1 \leqslant j \leqslant m\end{aligned} \tag{4.9}$$

where λ is the clustering model weight parameter, $U \in \{0,1\}^{c*m}$ is the partition matrix, c is the number of clusters, and m is the number of complete samples. When sample j belongs

to class i, u_{ij}=1, and 0 otherwise. $\widetilde{Z}$= $[\widetilde{z}_1, \widetilde{z}_2, ...\widetilde{z}_c]$ and $Z^{(v)}$= $\left[z_1^{(v)}, z_2^{(v)}, ...z_c^{(v)}\right]$ are the cluster center of the hidden view and cluster centers of each view, respectively.

In Equation (4.9), the first two terms are about the partition matrix with a clustering algorithm, the goal of which is to obtain the global and local partitions of all incomplete views. With the joint optimization model, we can capture the correlations between and within the intraview and interview samples through their sparse low-rank representation and the hidden view, respectively. By alternating the process of data recovery and clustering, we can obtain a high-performance padding matrix.

4.3.3 Optimization procedure

This section presents the Alternating Direction Method of Multipliers (ADMMs) as a solution for the problem stated in Equation (4.9). For that, we introduce several auxiliary variables to transform Equation (4.9) into the following expression:

$$
\begin{aligned}
&\min_{\{R^{(v)},E^{(v)},A^{(v)},H,Z^{(v)},\widetilde{Z}\}} \lambda \sum_{i=1}^{c}\sum_{j=1}^{m} u_{ij}\|h_j - \widetilde{z}_i\|^2 \\
&+ (1-\lambda)\sum_{v=1}^{s}\sum_{i=1}^{c}\sum_{j=1}^{m} u_{ij}\|{x_j}^{(v)} - {z_i}^{(v)}\|^2 \\
&+ \sum_{v=1}^{s}(\lambda_1\|Q^{(v)}\|_1 + \lambda_2\|M^{(v)}\|_G + \lambda_3\|E^{(v)}\|_{2,1} + \lambda_4\|A^{(v)}\|_{2,1}) \qquad (4.10)\\
s.t. &\forall v, \mathcal{P}_v(\widehat{R}^{(v)}X_o^{(v)}) = H + E^{(v)}, H^\top H = I, X^{(v)} = \widehat{R}^{(v)}X_o^{(v)} \\
&H \geq 0, \sum_{i=1}^{C} u_{ij} = 1, u_{ij} \in (0,1), 1 \leqslant j \leqslant m \\
&R^{(v)} = Q^{(v)}, R^{(v)} = M^{(v)}
\end{aligned}
$$

Therefore, the augmented Lagrangian function of Equation (4.10) is equivalent to the following function:

$$
\begin{aligned}
L &= \lambda \sum_{i=1}^{c}\sum_{j=1}^{m} u_{ij}\|h_j - \widetilde{z}_i\|^2 + (1-\lambda)\sum_{v=1}^{s}\sum_{i=1}^{c}\sum_{j=1}^{m} u_{ij}\|{x_j}^{(v)} - {z_i}^{(v)}\|^2 \\
&+ \sum_{v=1}^{s}(\lambda_1\|Q^{(v)}\|_1 + \lambda_2\|M^{(v)}\|_G + \lambda_3\|E^{(v)}\|_{2,1} + \lambda_4\|A^{(v)}\|_{2,1}) \\
&+ \frac{\mu}{2}\sum_{v=1}^{s}(\left\|R^{(v)} - Q^{(v)} + \frac{C_2^{(v)}}{\mu}\right\|_F^2 + \left\|R^{(v)} - M^{(v)} + \frac{C_3^{(v)}}{\mu}\right\|_F^2 \\
&+ \left\|\widehat{R}^{(v)}X_o^{(v)}A^{(v)} - H - E^{(v)} + \frac{C_1^{(v)}}{\mu}\right\|_F^2)
\end{aligned}
\tag{4.11}
$$

where $C_1^{(v)}$, $C_2^{(v)}$, $C_3^{(v)}$ are Lagrange multipliers and μ is a penalty parameter. Then, we can solve all unknown variables in the objective function ((4.11)) by alternative optimization as follows:

Update variable $R^{(v)}$: By removing irrelevant terms and fixing the other variables, the function becomes:

$$
\begin{aligned}
L(R^{(v)}) &= \left\|R^{(v)} - Q^{(v)} + \frac{C_2^{(v)}}{\mu}\right\|_F^2 + \left\|R^{(v)} - M^{(v)} + \frac{C_3^{(v)}}{\mu}\right\|_F^2 \\
&+ \left\|\widehat{R}^{(v)}X_o^{(v)}A^{(v)} - H - E^{(v)} + \frac{C_1^{(v)}}{\mu}\right\|_F^2
\end{aligned}
\tag{4.12}
$$

The partial derivative of $L(R^{(v)})$ with respect to $R^{(v)}$ is given by:

$$
\begin{aligned}
\frac{\partial L(R^{(v)})}{\partial(R^{(v)})} &= \frac{\partial(\left\|R^{(v)} - S_1^{(v)}\right\|_F^2)}{\partial(R^{(v)})} + \frac{\partial\left\|R^{(v)} - S_2^{(v)}\right\|_F^2}{\partial(R^{(v)})} \\
&\quad + \frac{\partial\left\|R^{(v)}X_o^{(v)}A^{(v)} + \iota(S_3^{(v)})\right\|_F^2}{\partial(R^{(v)})} \\
&= 2(R^{(v)} - S_1^{(v)}) + 2(R^{(v)} - S_2^{(v)}) + 2R^{(v)}X_o^{(v)}A^{(v)}{A^{(v)}}^T{X_o}^{(v)^T} \\
&\quad + 2\iota(S_3^{(v)}){A^{(v)}}^T{X_o}^{(v)^T}
\end{aligned}
\tag{4.13}
$$

where $S_1^{(v)} = Q^{(v)} - \frac{C_2^{(v)}}{\mu}$, $S_2^{(v)} = M^{(v)} - \frac{C_3^{(v)}}{\mu}$, $S_3^{(v)} = -H - E^{(v)} + \frac{C_1^{(v)}}{\mu}$, and the ι operation refers to selecting the row to represent the missing sample from the matrix according to the index of the missing sample.

By setting $\partial L(R^{(v)})/\partial(R^{(v)}) = 0$, the optimal $R^{(v)}$ can be obtained as follows:

$$R^{(v)} = \frac{-\iota(S_3^{(v)})A^{(v)^T}X_o^{(v)^T} + S_1^{(v)} + S_2^{(v)}}{2I + X_o{}^{(v)}A^{(v)}A^{(v)^T}X_o{}^{(v)^T}} \tag{4.14}$$

Update variable $Q^{(v)}$: By eliminating irrelevant terms and holding the other variables constant, $Q^{(v)}$ can be computed as follows:

$$\min_{\{Q^{(v)}\}} \lambda_1 \left\|Q^{(v)}\right\|_1 + \frac{\mu}{2}\left\|R^{(v)} - Q^{(v)} + \frac{C_2^{(v)}}{\mu}\right\|_F^2 \tag{4.15}$$

Equation (4.15) can be computed as [129] :

$$Q^{(v)} = \vartheta_{\frac{\lambda_1}{\mu}}(R^{(v)} + \frac{C_2^{(v)}}{\mu}) \tag{4.16}$$

Update variable $M^{(v)}$: By removing irrelevant terms and fixing the other variables in Equation (4.11), we can calculate $M^{(v)}$ as follows:

$$M^{(v)} = \underset{\{M^{(v)}\}}{\operatorname{argmin}} \lambda_2\|M^{(v)}\|_G + \frac{\mu}{2}\left\|M^{(v)} - P^{(v)}\right\|_F^2 \tag{4.17}$$

where $P^{(v)} = R^{(v)} + \frac{C_3^{(v)}}{\mu}$. We set the nonconvex surrogate of $\operatorname{rank}(M^{(v)})$ as $\|M^{(v)}\|_G$. It is difficult to obtain the solution from Equation (4.17) because it is a nonconvex function. It can be solved for a nonconvex function through regularization techniques and the difference of convex (DC) programming proposed by Moreau-Yosida [130]. Hence, the subproblem becomes:

$$M^{(v)^{t+1}} = \underset{\{M^{(v)^t}\}}{\operatorname{argmin}} \lambda_2\|M^{(v)^t}\|_G + \frac{\mu^t}{2}\left\|M^{(v)^t} - P^{(v)^t}\right\|_F^2 \tag{4.18}$$

To solve Equation (4.18), we develop Theorem 1 and provide the proof as below.

Theorem 1. Let $P = U\Sigma_P V^T$ be the singular value decomposition(SVD) of P, where $\Sigma_P = diag(\sigma_P)$. Set $F(M^{(v)}) = \|M^{(v)^t}\|_G = f \circ \sigma_M$.

$$\min_{\{M^{(v)}\}} F(M^{(v)}) + \frac{\mu}{2}\left\|M^{(v)} - P\right\|_F^2 \tag{4.19}$$

Therefore, the problem of the next optimal solution is transformed into $M^* = U\Sigma_M^* V^T$, where $\Sigma_M^* = diag(\sigma^*)$ and $\sigma^* = prox_{f,\mu}(\sigma_P)$, and $prox_{f,\mu}(\sigma_P)$ is the MoreauYosida operator, as follows:

$$prox_{f,\mu}(\sigma_P) = \underset{\{\sigma\}}{\operatorname{argmin}} f(\sigma) + \frac{\mu}{2}\|\sigma - \sigma_P\|_2^2 \tag{4.20}$$

Proof. Given $P = U\Sigma_P V^T$, $\Sigma_P = U^T PV$, and recording $D^{(v)} = U^{(v)^T} M^{(v)} V^{(v)}$. Since it has the same singular value as $M^{(v)}$, the formula is converted as follows:

$$F(M^{(v)}) + \frac{\mu}{2}\left\|M^{(v)} - P\right\|_F^2, \tag{4.21}$$

$$=F(D^{(v)}) + \frac{\mu}{2}\left\|D^{(v)} - \Sigma_P\right\|_F^2, \tag{4.22}$$

$$\geq F(\Sigma_D^{(v)}) + \frac{\mu}{2}\left\|\Sigma_D^{(v)} - \Sigma_P\right\|_F^2, \tag{4.23}$$

$$=F(\Sigma_M^{(v)}) + \frac{\mu}{2}\left\|\Sigma_M^{(v)} - \Sigma_P\right\|_F^2, \tag{4.24}$$

$$=f(\sigma) + \frac{\mu}{2}\left\|\sigma - \sigma_P\right\|_2^2 \quad , \tag{4.25}$$

$$\geq f(\sigma^*) + \frac{\mu}{2}\left\|\sigma^* - \sigma_P\right\|_2^2 \quad , \tag{4.26}$$

It should be noted that Equation (4.22) is valid because the Frobenius norm is unitarily invariant. Equation (4.23) is based on the Hoffman-Wielandt inequality and Equation (4.24) holds as we have $\Sigma_M^{(v)} = \Sigma_D$. Thus, Equation (4.24) is the lower bound of Equation (4.21) as $\Sigma_D^{(v)} = \Sigma_M^{(v)} = M^{(v)} = U^{(v)^T} D^{(v)} V^{(v)}$ holds, and the SVD of $D^{(v)}$ is $D^{(v)} = U^{(v)^T} \Sigma_D^{(v)} V^{(v)}$. When we perform a minimization operation on Equation (4.25), we obtain σ^*. Therefore, we have $D^* = Udiag(\sigma^*)V^T$, the optimal solution to Equation (4.19). We have completed the proof of Theorem 1.

Through the inspiration from the Moreau-Yosida regularization technique and the difference of convex (DC) programming, we transform Equation (4.18) to address the difference between two convex functions. The concave term is iterated for optimization in each iteration. Then, the optimization formula is as follows:

$$\sigma^{t+1} = \operatorname{argmin} f(\sigma^t) + \frac{\mu^t}{2}\left\|\sigma^t - \sigma_P^t\right\|_2^2 \tag{4.27}$$

which admits a closed-form solution [131], as shown in Equation (4.28).

$$\sigma^{t+1} = (\sigma_T^t - \frac{\varphi_t}{\mu_t})_+ \tag{4.28}$$

at point σ^t, the gradient representation of $f(\cdot)$ is denoted as $\varphi_t = \partial f(\sigma^t)$ and $U^{(v)} diag(\sigma_P^t)^{(v)} V^{(v)^T}$ is the singular value decomposition of $(R^{(v)} + \frac{C_3^{(v)}}{\mu})$. Through optimization iterations, the final convergence obtains the best advantage σ^*. The solution can be derived as follows:

$$M^{(v)^{t+1}} = U^{(v)} diag(\sigma^{(v)^*}) V^{(v)^T}. \tag{4.29}$$

Update variable $E^{(v)}$: after removing the irrelevant terms and fixing the other variables, the subproblem becomes:

$$L(E^{(v)}) = \lambda_3 \left\| E^{(v)} \right\|_{2,1} + \frac{\mu}{2} \left\| E^{(v)} - S_4^{(v)} \right\|_F^2 \tag{4.30}$$

where $S_4^{(v)} = \widehat{R}^v X_o^{(v)} A^{(v)} - H + \frac{C_1^{(v)}}{\mu}$, by setting $\partial L(E^{(v)})/\partial(E^{(v)}) = 0$, the solution can be derived as follows:

$$E^{(v)} = \frac{\mu S_4^{(v)}}{\lambda_3 Y^{(v)} + \mu I} \tag{4.31}$$

where $Y^{(v)} = diag\left\{ \frac{1}{\left\| e_1^{(v)} \right\|_2}, \frac{1}{\left\| e_2^{(v)} \right\|_2}, \cdots \frac{1}{\left\| e_1^{(v)} \right\|_n} \right\}$, $e_i^{(v)}$ represents the i-th row vector of $E^{(v)}$.

Update variable $A^{(v)}$: by removing the irrelevant terms and fixing the other variables, we can calculate $A^{(v)}$ as follows:

$$L(A^{(v)}) = \lambda_4 \left\| A^{(v)} \right\|_{2,1} + \frac{\mu}{2} \left\| \widehat{R}^{(v)} X_o^{(v)} A^{(v)} + S_3^{(v)} \right\|_F^2 \tag{4.32}$$

where $S_3^{(v)} = -H - E + \frac{C_1^{(v)}}{\mu}$. By setting $\partial L(A^{(v)})/\partial(A^{(v)}) = 0$, we can obtain the optimal $A^{(v)}$ as follows:

$$A^{(v)} = \frac{-\mu X_o^{(v)^T} \widehat{R}^{(v)^T} S_3^{(v)}}{\lambda_4 G^{(v)} + \mu X_o^{(v)^T} \widehat{R}^{(v)^T} \widehat{R}^{(v)} X_o^{(v)}} \tag{4.33}$$

where $G^{(v)} = diag\left\{ \frac{1}{\left\| a_1^{(v)} \right\|_2}, \frac{1}{\left\| a_2^{(v)} \right\|_2}, \cdots \frac{1}{\left\| a_1^{(v)} \right\|_n} \right\}$, and $a_i^{(v)}$ is the i-th row vector of $A^{(v)}$.

Update variable H: by removing the irrelevant terms and fixing the other variables, we can calculate H as follows:

$$L(H) = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| h_j - \widetilde{z}_i \right\|^2 + \frac{\mu}{2} \sum_{v=1}^{s} \left(\left\| B^{(v)} - H \right\|_F^2 \right) \tag{4.34}$$

where $B^{(v)} = \widehat{R}^v X_o^{(v)} A^{(v)} - E^{(v)} + \frac{C_1^{(v)}}{\mu}$. By setting $\partial L(H)/\partial(H) = 0$, we can obtain the optimal H as follows:

$$H = \frac{2\lambda \sum_{i=1}^{c} U_{in} Z_{it} + \mu \sum_{v=1}^{s} B_{nt}^{(v)}}{2\lambda \sum_{i=1}^{c} U_{in} + \mu s} \tag{4.35}$$

Update variable U: We also use the K-means algorithm and Euclidean distance to measure the similarity between samples. If the distance from the i-th sample to the j-th cluster center is smaller than the distance to other cluster centers, the element in the matrix u_{ij}

is 1, and 0 otherwise. According to our proposed model, the distance D_{ij} can be expressed as:

$$D_{ij} = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| h_j - \widetilde{z}_i \right\|^2 + (1-\lambda) \sum_{i=1}^{s} \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| x_j^{(v)} - z_i^{(v)} \right\|^2 \tag{4.36}$$

According to the K-means algorithm, the specific update method of matrix U is as follows:

$$u_{ij} = \begin{cases} 1, \forall k \in [1,m] \ and \ k \neq j, & D_{ij} \leq D_{ik} \\ 0, \qquad\qquad \exists k \in [1,m], & D_{ij} \geq D_{ik} \end{cases} \tag{4.37}$$

Update variable $Z^{(v)}$: To update the variable $Z^{(v)}$, we can calculate it by removing irrelevant terms and holding other variables constant, as shown below:

$$L(Z^{(v)}) = (1-\lambda) \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| x_j^{(v)} - z_i^{(v)} \right\|^2 \tag{4.38}$$

$Z^{(v)}$ can be solved as follows:

$$z_i^{(v)} = \frac{\sum_j^N u_{ij} x_j^{(v)}}{\sum_{j=1}^N u_{ij}} \tag{4.39}$$

Update variable $\widetilde{Z}$: by removing the irrelevant terms and fixing the other variables, we can calculate $\widetilde{Z}$ as follows:

$$L(\widetilde{Z}) = \lambda \sum_{i=1}^{c} \sum_{j=1}^{m} u_{ij} \left\| h_j - \widetilde{z}_i \right\|^2 \tag{4.40}$$

$\widetilde{Z}$ can be solved as follows:

$$\widetilde{z}_i = \frac{\sum_j^N u_{ij} h_j}{\sum_{j=1}^N u_{ij}} \tag{4.41}$$

Update Variables $C_1^{(v)}$, $C_2^{(v)}$, $C_3^{(v)}$: We update $C_1^{(v)}$, $C_2^{(v)}$, $C_3^{(v)}$ as follows:

$$C_1^{(v)} = C_1^{(v)} + \mu(\widehat{R}^{(v)} X_o^{(v)} A^{(v)} - H - E^{(v)}) \tag{4.42}$$

$$C_2^{(v)} = C_2^{(v)} + \mu(R^{(v)} - Q^{(v)}) \tag{4.43}$$

$$C_3^{(v)} = C_3^{(v)} + \mu(R^{(v)} - M^{(v)}) \tag{4.44}$$

Update Variables μ: we update μ by:

$$\mu = min(\rho\mu, \mu_0) \tag{4.45}$$

where ρ and μ_0 represent preset parameters.

4.3.4 Complexity analysis of CCIM-SLR

As discussed in Section 4.3.3, the computational expense of our algorithm is mainly due to operations such as matrix inversion and singular value decomposition of self-matrices. The algorithm of CCIM-SLR is summarized in Algorithm 1. The computational complexities of Steps 8, 10, and 11 in Algorithm 2 are approximately $O({m^{(v)}}^3)$, $O(\bar{m}^{(v)}{m^{(v)}}^2)$, and $O({m^{(v)}}^3)$, respectively. Therefore, the complexity of the entire optimization of the algorithm is approximately $O(\tau(2{m^{(v)}}^3 + \bar{m}^v{m^{(v)}}^2))$, where τ is the number of iterations, $\bar{m}^{(v)}$ is the number of missing samples, and $m^{(v)}$ is the number of observed samples of the view.

Algorithm 2 CCIM-SLR

Require: Incomplete multi-view dataset $X^{(v)}$, and parameters λ, λ_1, λ_2, λ_3, λ_4, ρ, and μ_0
Ensure: The resulting clusters
1: Initialize H;
2: Initialize U;
3: Initialize $E^{(v)}$;
4: Initialize $V^{(v)}$;
5: Initialize $\widetilde{V}$;
6: **while** not converge **do**
7: **for** $v = 1$ to V **do**
8: Update $R^{(v)}$ via Eq. (4.14);
9: Update $Q^{(v)}$ via Eq. (4.16);
10: Update $M^{(v)}$ via Eq. (4.29);
11: Update $E^{(v)}$ via Eq. (4.31);
12: Update $A^{(v)}$ via Eq. (4.33);
13: Update $Z^{(v)}$ via Eq. (4.39);
14: Update $C_1^{(v)}$ via Eq. (4.42);
15: Update $C_2^{(v)}$ via Eq. (4.43);
16: Update $C_3^{(v)}$ via Eq. (4.44);
17: **end for**
18: Update H via Eq. (4.35);
19: Update U via Eq. (4.37);
20: Update $\widetilde{Z}$ by solving Eq. (4.41);
21: Update μ by $\mu = \min(\rho\mu; \max_\mu)$;
22: **end while**

4.3.5 Convergence analysis of CCIM-SLR

In addition to the update step of $M^{(v)}$, the other steps are easily found to be bounded. Therefore, we analyze the effect of the update step of $M^{(v)}$ on our objective function.

We write $\|M^{(v)}\|_G$ as $K(M^{(v)})$ in Eq. (4.9):

$$\begin{aligned}\min J(M^{(v)}, R^{(v)}, C_3^{(v)}, \mu) = \sum_{v=1}^{s}(K(M^{(v)}) + \frac{\mu}{2}\left\|M^{(v)} - R^{(v)}\right\|_F^2 \\ + < \frac{C_3^{(v)}}{\mu}, M^{(v)} - R^{(v)} >)\end{aligned} \tag{4.46}$$

where $< ., . >$ represents the sum of the products of corresponding components between two matrices.

Lemma 1. $M^{(v)^t}$ and $R^{(v)^t}$ are bounded if $\sum_{t=1}^{\infty} \frac{(\mu^t - \mu^{t-1})}{2(\mu^{t-1})^2} < \infty$.

Proof. With some algebra, we can obtain:

$$\begin{aligned}&J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}, \mu^t) \\ =& J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^{t-1}}, \mu^{t-1}) \\ &+ \frac{(\mu^t - \mu^{t-1})}{2}\left\|M^{(v)} - R^{(v)}\right\|_F^2 \\ &+ Tr[(C_3^{(v)^t} - C_3^{(v)^{t-1}})(M^{(v)} - R^{(v)})] \\ =& J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^{t-1}}, \mu^{t-1}) \\ &+ \frac{(\mu^t - \mu^{t-1})}{2(\mu^{t-1})^2}\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2\end{aligned} \tag{4.47}$$

Then,

$$\begin{aligned}&J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t) \\ \leq& J(M^{(v)^{t+1}}, R^{(v)^t}, C_3^{(v)^t}, \mu^t) \\ \leq& J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}, \mu^t) \\ \leq& J(M^{(v)^t}, R^{(v)^t}, C_3^{(v)^{t-1}}, \mu^{t-1}) \\ &+ \frac{(\mu^t - \mu^{t-1})}{2(\mu^{t-1})^2}\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2\end{aligned} \tag{4.48}$$

By iterating the above inequality (4.48) t times, we obtain:

$$\begin{aligned}&J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t) \\ \leq& J(M^{(v)^1}, R^{(v)^1}, C_3^{(v)^0}, \mu^0) \\ &+ \sum_{i=1}^{t} \frac{(\mu^i - \mu^{i-1})}{2(\mu^{i-1})^2}\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2\end{aligned} \tag{4.49}$$

As $\left\|(C_3^{(v)^t} - C_3^{(v)^{t-1}})\right\|_F^2$ is bounded, the other terms included in the right-hand side of the inequality are also bounded. Therefore, $J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t)$ is upper bounded.

In addition, we have

$$\begin{aligned} & J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t}, \mu^t) + \frac{1}{2\mu^t} \left\| (C_3^{(v)^t}) \right\|_F^2 \\ =& K(M^{(v)^{t+1}}) + \frac{\mu^t}{2} \left\| M^{(v)^{t+1}} - R^{(v)^{t+1}} + \frac{C_3^{(v)^t}}{\mu^t} \right\|_F^2 \end{aligned} \tag{4.50}$$

By observing several terms on the right side of Equation (4.50), we find that each of them is finite, so $M^{(v)^{t+1}}$ and $R^{(v)^{t+1}}$ are also finite respectively. Therefore, $\left\{M^{(v)^t}\right\}$ and $\left\{R^{(v)^t}\right\}$ are also bounded.

Lemma 2. Let $\left\{M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}\right\}$ be the sequence and $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$ be an accumulation point. Then $\left\{M^{(v)^*}, R^{(v)^*}\right\}$ is a stationary point if we have $\lim_{t\to\infty} \mu^t \left(\boldsymbol{R}^{(v)^{t+1}} - \boldsymbol{R}^{(v)^t}\right) \to 0$.

Proof. The sequence $\left\{M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}\right\}$ is bounded as shown in Lemma 2. By the Bolzano-Weierstrass theorem, at least one accumulation point must exist in this sequence, e.g., $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$. Therefore, we presume that $\left\{M^{(v)^t}, R^{(v)^t}, C_3^{(v)^t}\right\}$ itself converges to $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$.

Since $R^{(v)^t} - M^{(v)^t} = (C_3^{(v)^t} - C_3^{(v)^{t-1}})/\mu^{t-1}$ holds, we have $\lim_{t\to\infty} R^{(v)^t} - M^{(v)^t} = 0$. Therefore, the primal feasibility condition is fulfilled.

For $M^{(v)^{t+1}}$, it holds that

$$\begin{aligned} & \partial_M \left(M^{(v)^{t+1}}, R^{(v)^t}, C_3^{(v)^t}, \mu^t\right) |_{M(v)^{t+1}} \\ =& \partial_M K \left(M^{(v)^{t+1}}\right) + {C_3}^{(v)^t} + \mu^t \left(R^{(v)^t} - M^{(v)^t}\right) \\ =& \partial_M K \left(M^{(v)^{t+1}}\right) + {C_3}^{(v)^{t+1}} + \mu^t \left(R^{(v)^{t+1}} - R^{(v)^t}\right) = 0 \end{aligned} \tag{4.51}$$

If the singular value decomposition of $M^{(v)}$ is $\boldsymbol{U}^{(v)}$ diag $\left(\sigma_i^{(v)}\right) V^{(v)^T}$ according to Theorem 1,

$$\partial_M K \left(M^{(v)^{t+1}}\right) |_{M(v)^{t+1}} = U \text{ diag} \left(\tau^{(v)}\right) V^{(v)^T}, \tag{4.52}$$

where $\tau_i = \gamma/(\gamma + \sigma_i)^2$ when $\sigma_i \neq 0$; else, it acts as $1/\gamma$. Since $\sigma_i \in (0, 1/\gamma]$ is finite, $\partial_M K \left(M^{(v)^{t+1}}\right) |_{M(v)^{t+1}}$ is bounded. ${C_3}^{(v)^t}$ is bounded as a Lagrange multiplier. $\mu^t \left(R^{(v)^{t+1}} - R^{(v)^t}\right)$ is bounded. Under the assumption that $\lim_{t\to\infty} \mu^t \left(R^{(v)^{t+1}} - R^{(v)^t}\right) \to 0$,

$$\partial_M K \left(M^{(v)^*}\right) + {C_3}^{(v)^*} = 0 \tag{4.53}$$

Hence, $\left\{M^{(v)^*}, R^{(v)^*}, C_3^{(v)^*}\right\}$ satisfies the Karush–Kuhn–Tucker conditions of $J(M^{(v)^{t+1}}, R^{(v)^{t+1}}, C_3^{(v)^t})$. Therefore, $\left\{M^{(v)^*}, R^{(v)^*}\right\}$ is the point satisfying the condition.

4.4 Experiments

This section reports in detail the performance evaluation of CCIM-SLR by comparing it with the state-of-the-art methods against five real-world datasets. Furthermore, we present experimental results on the proposed optimization approach, and its convergence property to demonstrate the efficiency of CCIM-SLR and the robustness of CCIM-SLR.

4.4.1 Datasets

To validate the clustering performance of the proposed method under different data dimensions, we used the five representative datasets in our experiments. The statistics of the datasets are listed in Table 4.1.

Table 4.1: Statistics of the datasets

Dataset	Clusters	Views	samples	Features
SensIT300	3	2	300	50/50
Statlog	7	2	2310	9/10
Wisconsin	5	2	265	1703/265
WebKB	2	2	1051	1840/3000
Yale	15	3	165	4096/3304/ 6750

- **SensIT300**[1] [132]: SensIT300 contains sensory data collected from an intelligent transportation system targeting three vehicle types. This is one of the main datasets used in many research papers to evaluate the performance of clustering algorithms. This dataset consists of 300 samples under three different classes with two views, and the sample data in each view consists of features of 50 dimensions respectively. The three classes are three types of transportation, while the views are split into vibration information and sound view obtained through sensor transmission.
- **Statlog**[2] [133]: The Statlog dataset was collected by the Vision Group, University of Massachusetts. The total number of samples in this dataset is 2310. The dataset contains seven kinds of outdoor images that were hand-segmented to create a classification for every pixel. Each sample in the dataset has feature dimensions of either 9 or 10.
- **Wisconsin** [3] [134]: The Wisconsin dataset is a real-world multiview dataset that contains 256 samples with different descriptions from 5 different categories (Student

[1] https://github.com/Liuzhenjiao123/multiview-data-sets/blob/master/sensIT300.mat
[2] https://github.com/Liuzhenjiao123/multiview-data-sets/tree/master
[3] https://lig-membres.imag.fr/grimal/data.html

Pages, Program Pages, Course Pages, Staff Pages, and Faculty and Staff Pages). The content view and reference view are the two types of views identified from each sample, with 1703 and 265 feature dimensions, respectively.

- **WebKB** [4] [135]: The WebKB dataset consists of 1051 samples under 2 classes. Each sample in this dataset corresponds to two types of features: i) those derived from the textual content of the web page, and ii) those derived from the anchor text containing links to other web pages. In this dataset, the dimension of the linked representation is 1840, while the other dimension is 3000.

- **Yale** [5] [136]: The Yale dataset is a collection of 165 pictures from 15 people. The pictures are distinguished by different expressions, gestures, and lights. The Yale dataset contains three views. The feature dimensions of each view are 4096, 3304, and 6750, respectively.

4.4.2 Baseline approaches

To validate the performance of the proposed CCIM-SLR, we compared it with five IMC methods: IMC-GRMF [46], IMSC-AGL [63], UEAF [98], DAIMC [48], and HCP-IMSC [99].

- **IMC-GRMF**: The IMC-GRMF method uses the orthogonal matrix factorization technique to learn the latent subspace. The local information of each view is incorporated to help fuse the complementary information of views, which results in a better-shared representation.

- **IMSC-AGL**: The IMSC-AGL method utilizes low-rank representations of adaptive learning of graphs in a multiview scenario. To obtain more refined low-dimensional representations, this model employs a number of spectral constraints.

- **UEAF**: The UEAF model is designed for incomplete multiview clustering and serves as a unified and robust embedding alignment approach. Differing from other methods, UEAF infers incomplete information by maintaining the consistency of the local structure of the views and learning the local structure shared among multiple views through reversing graph regularization.

- **DAIMC**: The DAIMC method is characterized by weighted semi-NMF: semi-nonnegative matrix factorization, which learns a weight matrix that can be adapted to multiple incomplete cases. This model performs an $L_{2,1}$ regularization to obtain a supplemental cluster-friendly matrix representation that can be shared by views.

[4] https://github.com/Liuzhenjiao123/dataset4
[5] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

- **HCP-IMSC**: The HCP-IMSC method uses higher-order information to improve the clustering performance of incomplete multiview. Tensor decomposition is adopted in the process of capturing higher-order association relations. Then, under hypergraph-induced superLaplace regularization, the missing view samples are restricted to be reconstructed by neighboring samples.

4.4.3 Experimental setups

We removed 10%, 30%, 50%, 70%, and 90% of the sample data in each view from the five datasets with the incomplete view. Specifically, all IMC-GRMF, IMSC-AGL, UEAF, DAIMC, and HCP-IMSC perform post-clustering operations (e.g., K-means) based on obtained latent representations to produce their final clustering results. Considering that the clustering results of K-means are affected by the initialization of seed points, we performed K-means 10 times in the experiment to obtain the average value. For setting the parameters of the compared methods, we choose the values within the parameter ranges specified in the original papers.

4.4.4 Evaluation metrics

In our experiments, we used four performance metrics to evaluate the clustering performances: NMI - Normalized mutual information [99], ACC - Accuracy [101], ARI - Adjusted Rand index [102], and F-scores [103].

- **NMI - Normalized mutual information**: The NMI indicator measures the quality of clusters defined as:

$$NMI = \frac{\sum_{i=1}^{C}\sum_{j=1}^{C} N_{i,j} \ln \frac{N_{i,j}}{N_i \hat{N}_j}}{\sqrt{(\sum_{i=1}^{C} N_i \ln \frac{N_i}{N})(\sum_{j=1}^{C} \hat{N}_j \ln \frac{\hat{N}_j}{N})}} \tag{4.54}$$

  where N is the number of samples in a complete view, N_i and $\hat{N}_j$ are the numbers of samples in the i-th cluster and the number of samples of the j-th label, respectively. The number of samples in the intersection between the i-th cluster and j-th label is represented by $N_{i,j}$.

- **ACC - Accuracy**: ACC measures the cluster quality. ACC is estimated by:

$$ACC = \frac{\sum_{i=1}^{N} \delta(map(r_i), l_i)}{N} \tag{4.55}$$

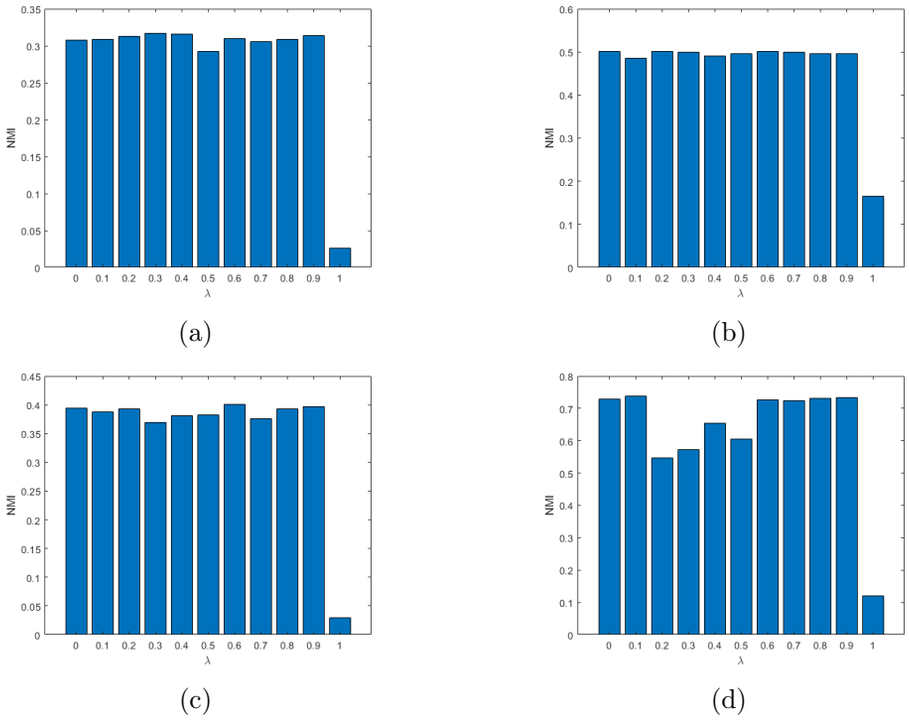


Figure 4.3: NMI (%) versus parameter λ of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view, (b) Statlog dataset with 10% incomplete samples of each view, (c) Wisconsin dataset with 10% incomplete samples of each view, and (d) WebKB dataset with 10% incomplete samples of each view.

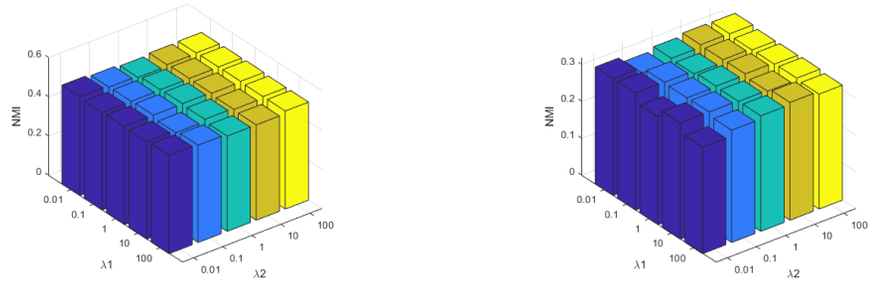
where N is the number of samples, r_i and l_i are a predicted cluster label of x_i and the corresponding ground-true label, respectively. If x=y, then $\delta(x, y) = 1$, and 0 otherwise. map(r_i) represents the function of the optimal permutation mapping.

- **ARI - Adjusted Rand index (ARI)**: ARI is a performance evaluation indicator of the clustering model. A larger value indicates a better clustering result. ARI is calculated using the following four indicators: 1) A true positive (TP) represents true positive, 2) A true negative (TN) means true negative, 3) A false-positive (FP) is false-positive, and 4) A false-negative (FN) is false-negative. ARI is formulated as:

$$ARI = \frac{2 \times (TP \cdot TN - FN \cdot FP)}{(TP + FN)(TN + FN) + (TP + FP)(FP + TN)} \tag{4.56}$$

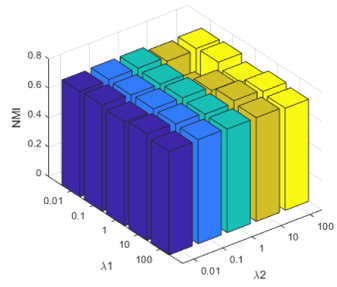
- **F-scores**: The F-score integrates the recall and precision of a classifier into a single metric that compares the performances of two classifiers:

$$F - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4.57}$$

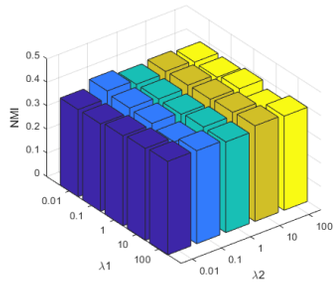


(a)

(b)



(c)



(d)

Figure 4.4: NMI (%) versus parameters λ_1 and λ_2 of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view, (b) Statlog dataset with 10% incomplete samples of each view, (c) Wisconsin dataset with 10% incomplete samples of each view, and (d) WebKB dataset with 10% incomplete samples of each view.

$$\text{where } precision = \frac{TP}{TP + FP}, \text{ and } recall = \frac{TP}{TP + FN}.$$

4.4.5 Comparisons of the performance of clustering and discussion

Table 4.2 lists the evaluation scores of NMI, ACC, ARI, and F-score and the results of different IMC baseline methods and our proposed method on the five datasets with different missing ratios. From this table, we can make the following important observations.

1) Based on the experiments, our proposed CCIM-SLR achieved the best performance compared with all other state-of-the-art methods. In particular, our method achieved 96.63% accuracy on the WebKB dataset with missing 10% samples. Compared to the proposed CCIM-SLR, IMSC-AGL obtained comparable results on the Yale dataset. However, CCIM-SLR still shows its superiority on other datasets with an increased missing rate.

2) Although UEAF exhibits good performance in terms of metric scores, it lacks robustness. The main reason for this is that performing well, UEAF must satisfy the condition

Table 4.2: Mean NMIs(%), ACCs(%), ARIs(%) and F-score(%) of different methods on SensIT300 , Statlog , Wisconsin , WebKB and Yale datasets

Dataset	Method \ PER	NMI					ACC					ARI					F-score				
		0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
SensIT300	IMSC-AGL	22.50	19.86	16.20	14.40	15.00	65.74	62.73	60.00	58.00	56.33	24.78	21.31	17.24	14.56	15.18	49.85	47.69	44.93	43.37	43.60
SensIT300	DAIMC	20.44	17.29	15.67	10.58	8.89	64.08	60.63	59.18	51.90	50.12	22.54	17.87	16.27	10.43	8.28	48.68	45.73	44.80	41.28	40.59
SensIT300	UEAF	21.23	17.83	15.85	14.09	15.00	65.00	60.67	59.33	58.00	56.00	23.74	17.93	16.05	15.11	12.92	49.52	45.35	44.10	43.32	41.92
SensIT300	IMC-GRMF	15.50	8.36	5.65	4.74	2.19	60.20	51.00	47.33	39.13	35.67	16.51	8.45	5.72	4.18	0.07	44.85	38.98	36.99	36.57	**47.59**
SensIT300	HCP-IMSC	32.17	**30.08**	21.22	18.67	15.32	**72.12**	65.17	57.86	57.33	56.67	**32.17**	**30.16**	21.35	18.67	15.12	**56.12**	50.15	48.05	46.59	43.88
SensIT300	Ours	**32.70**	26.88	**24.65**	**19.93**	**17.37**	68.13	**67.00**	**64.07**	**61.27**	**57.27**	29.27	26.84	**24.51**	**19.86**	**15.21**	53.79	**51.57**	**50.30**	**47.12**	44.39
Statlog	IMSC-AGL	10.81	20.51	13.51	8.76	3.08	27.45	37.69	30.95	24.11	17.92	5.91	15.09	7.63	2.28	0.27	21.71	27.89	21.59	18.96	18.77
Statlog	DAIMC	47.35	39.59	34.15	30.49	26.61	56.66	51.29	44.52	41.19	35.26	36.11	27.27	18.80	18.01	13.56	45.98	38.65	32.63	30.99	27.52
Statlog	UEAF	48.83	37.61	34.98	32.61	29.83	48.09	46.10	43.51	37.40	39.18	25.62	23.60	18.28	13.48	10.42	40.48	35.90	32.40	29.54	27.34
Statlog	IMC-GRMF	43.93	41.76	38.21	31.07	30.31	54.68	53.84	47.45	36.36	40.61	33.85	30.46	28.53	18.14	16.97	43.43	40.69	39.45	30.62	29.41
Statlog	HCP-IMSC	50.43	**46.16**	40.21	36.05	32.51	57.12	53.84	49.45	45.06	39.56	38.12	32.11	26.31	20.17	17.32	48.23	44.19	39.25	34.42	30.25
Statlog	Ours	**50.53**	45.22	**41.43**	**37.39**	**33.84**	**57.52**	**53.97**	**49.76**	**45.36**	**41.40**	**39.02**	**34.09**	**29.98**	**24.93**	**20.09**	**48.52**	**44.45**	**40.66**	**36.14**	**32.15**
Wisconsin	IMSC-AGL	20.98	18.81	14.10	13.64	12.24	42.66	38.87	34.34	34.72	32.57	16.56	11.10	8.29	5.01	5.76	41.04	33.01	32.02	32.11	32.57
Wisconsin	DAIMC	29.87	26.52	24.04	22.10	16.75	51.39	43.77	45.66	48.83	44.00	24.96	16.95	16.52	15.88	12.70	46.01	39.37	39.02	40.87	36.07
Wisconsin	UEAF	35.94	**40.14**	34.25	**33.71**	**29.21**	60.75	57.35	50.56	**55.47**	**45.28**	34.56	34.00	25.34	**26.25**	**16.06**	53.18	50.78	44.07	**47.57**	41.32
Wisconsin	IMC-GRMF	25.69	18.57	10.63	7.9	6.2	43.69	37.21	33.28	39.17	44.10	14.52	9.87	5.22	4.80	0.51	40.04	33.78	30.51	34.16	**45.97**
Wisconsin	HCP-IMSC	27.21	24.13	26.63	25.53	23.63	50.16	40.11	49.32	44.15	38.11	21.12	17.41	20.32	17.01	14.13	42.14	39.18	40.21	14.78	37.42
Wisconsin	Ours	**42.33**	35.35	**36.45**	28.90	24.68	**65.51**	**60.98**	**54.26**	44.45	43.92	**42.32**	**35.57**	**27.56**	17.93	14.24	**59.10**	**54.17**	**47.38**	39.72	40.14
WebKB	IMSC-AGL	65.50	30.11	50.13	35.22	7.1	95.05	82.78	91.34	85.06	62.13	79.36	40.73	67.53	46.84	5.83	92.74	76.54	87.62	79.25	59.30
WebKB	DAIMC	60.57	52.35	41.54	44.96	38.16	93.14	90.67	84.64	90.56	88.79	69.79	63.84	50.72	59.51	53.33	90.89	88.30	83.63	87.80	85.73
WebKB	UEAF	68.41	70.05	64.51	59.63	**61.76**	95.43	95.62	94.86	94.10	**94.57**	80.97	81.77	78.60	75.31	**76.99**	93.23	93.48	92.47	91.55	**92.26**
WebKB	IMC-GRMF	51.70	34.37	2.8	9.1	2.9	92.01	87.82	61.08	78.21	71.36	67.85	52.22	4.36	21.92	10.11	88.59	83.80	59.61	74.87	68.52
WebKB	HCP-IMSC	71.11	68.23	61.12	53.92	45.72	95.12	93.21	92.08	91.72	86.86	82.12	79.21	77.12	67.45	52.75	93.12	92.60	91.13	87.23	81.23
WebKB	Ours	**73.31**	**72.25**	**69.57**	**67.52**	57.99	**96.63**	**96.44**	**95.87**	**95.61**	93.82	**85.53**	**84.77**	**82.61**	**81.33**	73.85	**94.99**	**93.88**	**94.77**	**93.60**	91.30
Yale	IMSC-AGL	**68.49**	65.67	**70.38**	65.71	**68.80**	**67.09**	62.55	**68.97**	61.33	**64.36**	**44.52**	43.71	**47.32**	44.46	**46.97**	**59.75**	56.67	62.06	56.33	**60.12**
Yale	DAIMC	57.16	53.68	54.26	44.24	41.36	53.21	50.18	48.79	37.21	34.42	32.44	26.87	25.21	14.49	11.36	44.83	42.32	41.42	31.36	28.92
Yale	UEAF	61.21	61.87	61.02	60.84	61.97	55.21	55.88	54.85	55.15	55.58	37.81	38.70	37.91	36.84	38.31	50.85	50.93	49.89	50.02	51.14
Yale	IMC-GRMF	63.67	64.12	56.58	46.55	45.27	57.64	58.24	48.30	37.82	36.06	41.75	42.22	32.40	19.04	17.38	53.65	53.71	44.67	33.52	31.67
Yale	HCP-IMSC	64.42	63.77	58.51	53.36	47.92	56.85	60.85	54.79	53.36	47.92	41.60	41.38	35.07	27.81	20.56	52.28	53.12	47.07	40.88	33.90
Yale	Ours	66.58	**72.37**	69.76	**65.74**	64.12	65.55	**63.79**	67.73	**62.41**	60.21	43.12	**54.19**	46.13	**45.02**	43.46	57.79	**63.87**	**62.14**	**56.95**	56.78

that the feature dimensions of all views are larger than the cluster number. From this perspective, our method has stronger robustness in handling complex types of incomplete multiview.

3) In general, DAIMC ignores padding for missing views. As a result, the achieved NMI score of our proposed CCIM-SLR is 28.03% higher than that of DAIMC in a case in which 50% of samples in the Webkb dataset are missing. CCIM-SLR uses an advanced filling mechanism. The experimental results show that this mechanism can result in better clustering performance.

4) Compared with IMC-GRMF, CCIM-SLR can maintain meaningful semantic relation-

Table 4.3: Ablation study of the CCIM-SLR performance (%) on SensIT300 , Wisconsin, and WebKB datasets

Dataset	Method \ PER	NMI					ACC					F-score				
		0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
SensIT300	Ablation-1	19.84	15.22	12.78	10.11	6.06	61.82	57.13	55.06	51.06	46.13	47.60	44.18	41.81	39.90	37.46
SensIT300	Ablation-2	20.19	15.41	13.14	7.1	4.2	56.82	53.61	47.13	42.41	56.67	47.85	44.24	42.18	37.55	35.78
SensIT300	Ours	**32.70**	**26.88**	**24.65**	**19.93**	**17.37**	**68.13**	**67.00**	**64.07**	**61.27**	**57.27**	**53.79**	**51.57**	**50.30**	**47.12**	**44.39**
Wisconsin	Ablation-1	39.75	36.18	33.71	27.13	24.19	62.64	53.50	51.84	42.86	42.19	53.86	47.25	47.13	39.05	37.98
Wisconsin	Ablation-2	37.20	**38.01**	31.43	27.67	23.25	60.75	55.39	51.92	44.08	41.88	52.39	49.59	45.25	39.70	38.21
Wisconsin	Ours	**42.33**	35.35	**32.45**	**28.90**	**24.68**	**65.51**	**60.98**	**54.26**	**44.45**	**43.92**	**59.10**	**54.17**	**47.38**	**39.72**	**40.14**
WebKB	Ablation-1	68.33	57.15	49.34	41.11	4.1	94.53	88.69	85.11	56.17	50.12	92.09	84.78	80.71	57.82	56.77
WebKB	Ablation-2	63.82	41.49	26.64	14.12	5.23	93.92	83.93	73.26	57.51	57.01	91.32	78.06	66.19	57.84	61.71
WebKB	Ours	**73.31**	**72.25**	**69.57**	**67.52**	**57.99**	**96.63**	**96.44**	**95.87**	**95.61**	**93.82**	**94.99**	**93.88**	**94.77**	**93.60**	**91.30**

Table 4.4: Two incomplete multi-view clustering methods based on association models exhibit differences in ACCs(%), F-scores(%), running time (seconds), and computational complexity when applied to datasets SensIT300 and Statlog, with 50% incomplete samples in each view. m_o denotes the number of observed samples.

Dataset	Method	ACC	F-score	Running time (seconds)	computational complexity
SensIT300	HCP-IMSC	57.86	48.05	1.4312	$O(sm^3 + s(m-m_o)^3 + cmslog(s) + cm^2s)$
SensIT300	CCIM-SLR	**64.07**	**50.30**	**1.4284**	$O(\tau(2{m^{(v)}}^3 + \bar{m}^v{m^{(v)}}^2))$
Statlog	HCP-IMSC	49.45	39.25	219.9447	$O(sm^3 + s(m-m_o)^3 + cmslog(s) + cm^2s)$
Statlog	CCIM-SLR	**49.76**	**40.66**	**62.7427**	$O(\tau(2{m^{(v)}}^3 + \bar{m}^v{m^{(v)}}^2))$

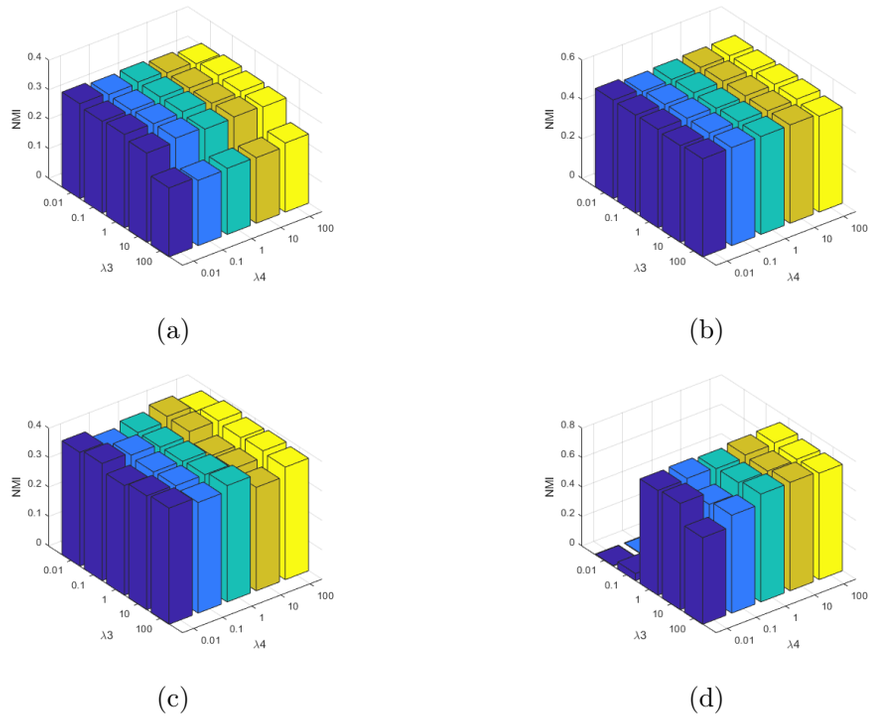


Figure 4.5: NMI (%) versus parameters λ_3 and λ_4 of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view; (b) Statlog dataset with 10% incomplete samples of each view; (c) Wisconsin dataset with 10% incomplete samples of each view; and (d) WebKB dataset with 10% incomplete samples of each view.

ships between the original view by building a consistent structure. In particular, clustering performance on the datasets with large differences in sample dimensions between Wisconsin and WebKB. CCIM-SLR can handle all kinds of incomplete data, which is much better than IMC-GRMF.

5) From Table 4.2, the performance of HCP-IMSC in terms of several metric scores is superior to other compared methods, which shows the advantage of capturing high-order correlation. In a case of an increasing missing rate, CCIM-SLR, however, can produce

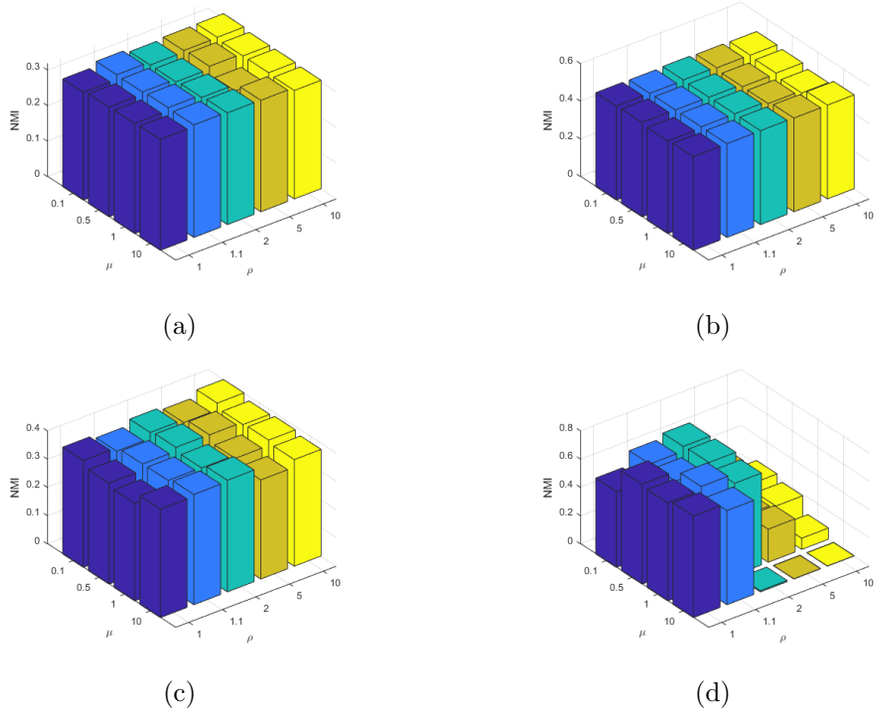


Figure 4.6: NMI (%) versus parameters μ and ρ of the proposed CCIM-SLR on the (a) SensIT300 dataset with 10% incomplete samples of each view, (b) Statlog dataset with 10% incomplete samples of each view, (c) Wisconsin dataset with 10% incomplete samples of each view, and (d) WebKB dataset with 10% incomplete samples of each view.

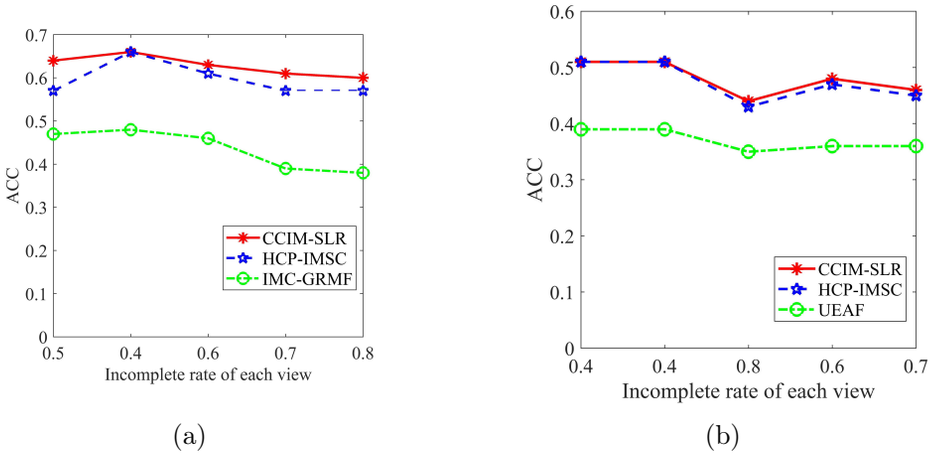


Figure 4.7: Comparisons of robustness experiments on SensIT300 and Statlog Datasets.

excellent incomplete multiview clustering results.

4.4.6 Parameter sensitivity of CCIM-SLR

Comparative experiments are performed in this section to investigate the sensitivity of each parameter of CCIM-SLR. Because the NMI index can objectively evaluate the accuracy of the comparison between a community division and the standard division, we use NMI to determine the range of parameters under satisfactory clustering results. Our main focus is on the following parameters in Equation (4.9): clustering model weight parameter λ, sparsity term parameter λ_1, low-rank term parameter λ_2, noise term parameter λ_3, flexible term parameter λ_4, and Lagrange operator control parameters μ and ρ.

1) Parameter λ: Figure 4.3 shows the NMI (%) scores for different scales of λ parameters. Our proposed CCIM-SLR achieved satisfactory performances on the SensIT300, Statlog, Wisconsin, and WebKB datasets when λ was in the range of [0, 0.9], [0, 0.9], [0, 0.9], and [0.6, 0.9], respectively. Based on the results, the best values for the λ parameter should range between [0.6, 0.9]. These values are used in further experiments.

2) Parameters λ_1 and λ_2: We explored the details of parameters λ_1 and λ_2 in Equation (4.9) by applying CCIM-SLR to the SensIT300, Starlog, Wisconsin, and WebKB datasets with a 10% incomplete-view rate, as shown in Figure 4.4. In our analyses, the performance of our algorithm is shown to be insensitive to λ_1 and λ_2 parameters. As shown in Figure 4.4, the values of the indicators of clustering do not change significantly as the values of the parameters change.

3) Parameters λ_3 and λ_4: We evaluated the NMI score for different values of λ_3 and λ_4 parameters in Equation (4.9) on the SensIT300, Statlog, Wisconsin, and WebKB datasets with an incomplete-view rate of 10%. Figure 4.5 depicts the best clustering results. Based on the experiments, the most suitable values for the candidate parameters λ_3 and λ_4 range from [1,10] and [0.1,100], respectively.

4) Parameters μ and ρ: Figure 4.6 shows the NMI versus μ and ρ parameters in Equation (4.45) on the SensIT300, Statlog, Wisconsin, and WebKB datasets with an incomplete-view rate of 10%. The experimental results show that if μ ranges from [0.5,10] and ρ from [1,1.1], our CCIM-SLR performed best.

4.4.7 Ablation study

To investigate the impact of each component of CCIM-SLR on its overall performance, we performed two ablation experiments on the three datasets. Specifically, we removed the hidden view from the clustering part in Ablation 1. Ablation 2 replaced the adjustable Γ-norm with the traditional kernel norm to produce a low-rank representation. The experiment results are reported in Table 4.3. From this table, we can find that although the NMI score of Ablation 2 is 2.66% higher than that of CCIM-SLR at a missing rate of

30%, CCIM-SLR still performed well in all other cases. This indicates that the adjustable low-rank representation Γ-norm can obtain better filling performance than the traditional kernel norm. In addition to this, we can find that clustering with the removed hidden views produced the worst results, which indicates that the fusion results of multiview directly affect the clustering results.

4.4.8 Robustness experiments

To further demonstrate the robustness of the proposed CCIM-SLR algorithm, we conducted experiments on two datasets, SensIT300 and Statlog, with varying degrees of missing data. Specifically, we set missing rates to (50%, 40%, 60%, 70%, and 80%) and (40%, 40%, 80%, 60%, and 70%) for SensIT300 and Statlog, respectively.

In Figure 4.7a, we compare the clustering performance of CCIM-SLR, HCP-IMSC, and IMC-GRMF on the SensIT300 dataset across the different missing rates. Our proposed algorithm is the most stable, except for the 40% missing rate. Similarly, in Figure 4.7b, we show the clustering performance of CCIM-SLR, HCP-IMSC, and UEAF on the Statlog dataset with the above-mentioned missing rates. Our proposed algorithm still performs the best, except for the 40% missing rate.

Overall, our experimental results demonstrate that CCIM-SLR is robust and performs well on datasets with varying degrees of missing data.

4.4.9 Experiments on the convergence of CCIM-SLR

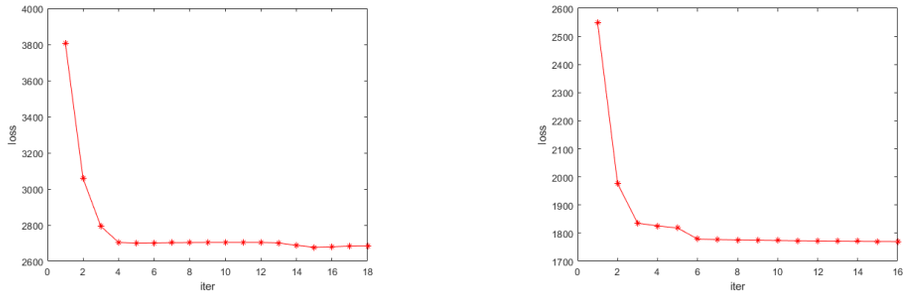
As presented in Section 4.3.3, the objective function of CCIM-SLR is divided into several subproblems, with each subproblem being analytically solved. The CCIM-SLR algorithm adopts an alternating iterative optimization procedure. The objective function monotonically decreases until it converges. As shown in Figure 4.8, the experimental results have demonstrated the correctness of the theoretical proofs in Section 4.3.5.

4.4.10 The effectiveness of CCIM-SLR on the datasets

As shown in Table 4.4, the performance of two methods that utilize association information to constrain the reconstruction of missing view samples is presented. This is evaluated in terms of clustering metrics, execution time, and complexity. The experimental results indicate that CCIM-SLR has the best overall efficiency.

4.5 Conclusion

In reality, datasets that are collected often contain data samples with incomplete multiview, and the number of such samples varies significantly. This presents a challenge for cluster-



(a) (b)

Figure 4.8: Objective function values versus the iteration steps of CCIM-SLR on the (a) Statlog and (b) WebKB face databases, in which 10% samples are randomly selected as the paired samples.

ing methods. To address this issue, a novel incomplete multiview clustering method called CCIM-SLR has been presented in this paper based on a sparse low-rank representation. In particular, CCIM-SLR measures the correlations between samples with the same views using sparse low-rank learning, while also capturing the correlations between different views through shared hidden view learning. Moreover, the proposed method learns the shared hidden space, visible view, and cluster partition alternatively, thus avoiding the sensitivity of postprocessing methods like K-means to initial parameter values. This improves performance, as demonstrated through both theoretical proof and experimental comparison with advanced methods for IMVC on five representative datasets. The experimental results showed that our CCIM-SLR achieved good performance, especially on datasets with an increasing number of incomplete samples.

Chapter 5

Incomplete Multi-view Representation Learning Through Anchor Graph-based GCN and Information Bottleneck

Contents

5.1 Introduction

In the era of big data, the presence of multiple modalities in data is prevalent [137] [138] [139] [140] [141]. However, it is common for data to be missing during the data collection process [142]. As depicted in Figure 5.1(a) and Figure 5.1(b), two common types of missing data patterns can be observed [2] [63] [143]. One type is characterized by many instances containing all views, while some instances have only one view. The other type involves randomly missing views. Therefore, a number of algorithms for incomplete multiview representation learning (IMRL) have been proposed to learn consistent representations in the presence of missing views [144] [119].

The existing algorithms for IMRL can be broadly categorized into two categories based on their key strategies for handling missing data. The first category entails the removal of samples that lack particular views, followed by the clustering of the remaining samples that have complete observations of all views [2]. However, this approach has limitations as it completely excludes samples with missing views from data analysis or clustering algorithms. As a result, subsequent clustering algorithms may fail to capture the underlying correlations among certain instances. For example, in the context of medical diagnosis, we consider various examinations and clinical data as different views. However, due to uncontrollable factors, patients may have missing data in certain views. If we employ the first category of methods to address this issue, it has the following drawbacks: inaccurate diagnosis of patients with missing view data and further impact on the diagnostic results of other patients due to the loss of relevant data.

The second category involves filling the missing views with either 0 or the measured value of instances, followed by processing the data using conventional multiview clustering methods [44] [98] [114]. For example, Wen et al. [98] introduce reconstruction items to fill in missing data and ensure natural alignment among views. This approach leverages available information from non-missing views and hidden information from missing views for shared representation learning. However, the limitation of this category is that setting missing views to fixed values may introduce bias and result in the loss of true information contained in the missing views, consequently impacting the quality of clustering results.

In summary, both the deletion and imputation methods commonly used for handling incomplete multi-view data suffer from the same problem: they disregard the underlying real information within the missing data. As depicted in Figure 5.1(c), these methods solely rely on simple processing techniques for fusing multi-view data [2] [98] [55].

In the real world, datasets often not only contain incomplete data but also exhibit varying degrees of incompleteness. This poses a significant challenge in developing algorithms that can effectively utilize the available information within the missing data instead of dis-

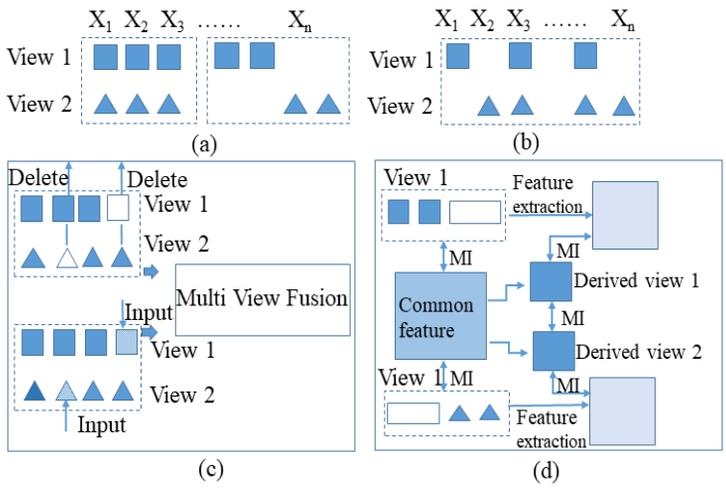


Figure 5.1: (a) Many instances contain all views, while some instances have only one view. (b) Incomplete multiview data with arbitrary missing instances. (c) Most previous IMRL methods. (d) Our Method.

carding it. Moreover, these algorithms need to be robust enough to handle different levels of data completeness.

To address this challenge, we propose a novel approach to IMRL from a data compression perspective (Figure 5.1(d)). Instead of deleting instances or filling in missing instances, our method learns a shared low-dimensional representation by minimizing or maximizing the mutual information between different views. This approach offers several advantages: Firstly, compressing the data allows for the preservation of a greater amount of useful information while mitigating redundancy and noise interference. Secondly, our approach mitigates the impact on clustering outcomes by avoiding the introduction of inaccurate information through instance imputation methods. Finally, compressed data fusion methods typically employ dimensionality reduction techniques for representation and fusion, thereby reducing data complexity and storage requirements.

In summary, the main contributions of this paper are as follows:

- We propose a novel framework called IMRL-AGI for multi-view representation learning in the presence of incomplete data. This framework combines information bottlenecks and anchor graph GCN. To the best of our knowledge, this is the first time that information bottleneck is integrated with anchor graph GCN for addressing incomplete multi-view representation learning.

- We introduce the maximization of mutual information constraints to enhance the correlation between the view information obtained from the common representation

and the view information derived through anchor graph GCN.

- IMRL-AGI enhances the robustness and diversity of information bottleneck representations in downstream tasks by minimizing the mutual information between different views derived from the common representation and between the representation of the original view and the common representation
- We extensively evaluated the performance of IMRL-AGI on three real-world datasets. Particularly, IMRL-AGI shows significant improvements in clustering and classification accuracy even with high view missing rates (e.g. 10.23% and 24.1% respectively on the ORL dataset). Furthermore, experiment results demonstrate the robustness of IMRL-AGI under different view missing rates (e.g. the NMI values from 82.52% to 86.49% on ORL dataset).

The rest of this paper is organized as follows. Section 5.2 presents an overview of the related work. In Section 5.3 our proposed IMRL-AGI is described in detail. The experimental results and analysis are presented in Section 5.4. Finally, a summary is provided in Section 5.5.

5.2 Related Work

The methods for incomplete multiview representation learning can be roughly divided into shallow methods and deep methods.

The shallow methods can be further classified into three categories. The first category is graph-based learning methods [145] [146]. These methods construct affinity matrices or graphs using the relationships between instances and assign clusterings through the relationships in the graphs. The second category is based on non-negative matrix factorization (NMF) methods [3] [43]. The third category is kernel-based methods [147] [55], which use the kernel matrix of complete views to complete the kernel matrix of incomplete views, thereby completing clustering allocation.

The deep methods typically use interpolation strategies to fill in missing data values, which can be roughly divided into two categories. The first category of deep methods mainly generates missing view data by generative adversarial networks [67] [148]. The second category of deep methods is to recover the missing multi-view data through contrastive learning [149] [150].

In addition to shallow and deep methods, the anchor graph-based shallow methods have demonstrated their effectiveness in approximating the complete instance graph. They have been widely utilized in several incomplete multi-view clustering methods [111] [151] [152]. In particular, Guo et al. [111] propose an Anchor-based Partial Multi-view Clustering (APMC)

by reconstructing inter-instance relationships. This approach provides a straightforward yet powerful strategy for leveraging anchors for clustering tasks.

Methods based on information bottleneck are widely recognized as an effective deep method for obtaining a minimal and sufficient representation of multi-view data [153] [152] [154]. CMIB-Nets [153] is one of the representative models that address multi-view representation learning problems using information bottlenecks. On the other hand, GCN is a deep learning model that extracts features from graph data. Additionally, GCN demonstrates its capability in effectively handling data with missing values [155].

However, the anchor-based methods still have certain limitations as they fail to consider the redundancy present in the views. Similarly, the information bottleneck-based methods have not been effectively applied to handling missing views. To address these gaps, we propose a novel framework that combines information bottlenecks and anchor Graph GCN. Our framework aims to effectively reduce redundant information in view representations while also tackling the challenges posed by various missing views in the data.

5.3 Proposed IMRL-AGI

In this section, we describe the proposed IMRL-AGI in detail. As shown in Figure 5.2, IMRL-AGI mainly consists of three components: 1) view-specific representations; 2) shared representations; and 3) incomplete multi-view representation with information bottlenecks. We start by explaining the relevant symbols.

Let $X = \{X^{(1)}, X^{(2)}, ...X^{(v)}\}$ represent the original incomplete multi-view dataset. Here, $X^{(v)} \in \mathbb{R}^{n_{(v)} \times d_{(v)}}$ represents the data of the v-th view, where $n_{(v)}$ and $d_{(v)}$ represent the number of instances and dimensions of each view, respectively. Given two random variables B and C, their mutual correlation is quantified using mutual information $I(B, C)$ [156].

5.3.1 View-specific representation construction

In order to better learn a specific representation of views, we construct an affinity matrix for each view, which is based on the anchor information between the views. In addition, we introduce GCN to facilitate feature learning of each view. Moreover, the construction method can be easily extended to encompass more than two views.

5.3.1.1 Anchor-based similarity reconstruction

Although the sources of multiview data are different, there can be consistent similarities between certain data points within a view. As a result, we identify the common instances that appear in both views as anchors and compute similarity matrices for each view based on these anchors. This approach differs from previous methods that calculate a unified

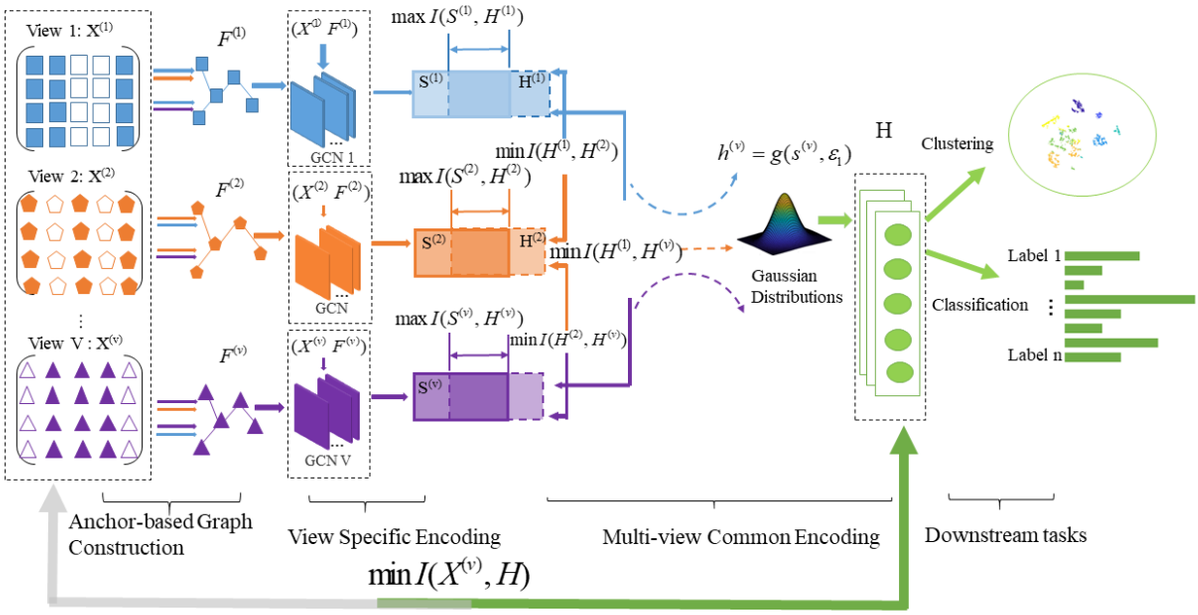


Figure 5.2: Overview of IMRL-AGI. It consists of four major components: (a) graph construction ; (b) view-specific encoding; (c) multi-view encoding; and (d) downstream task finetuning.

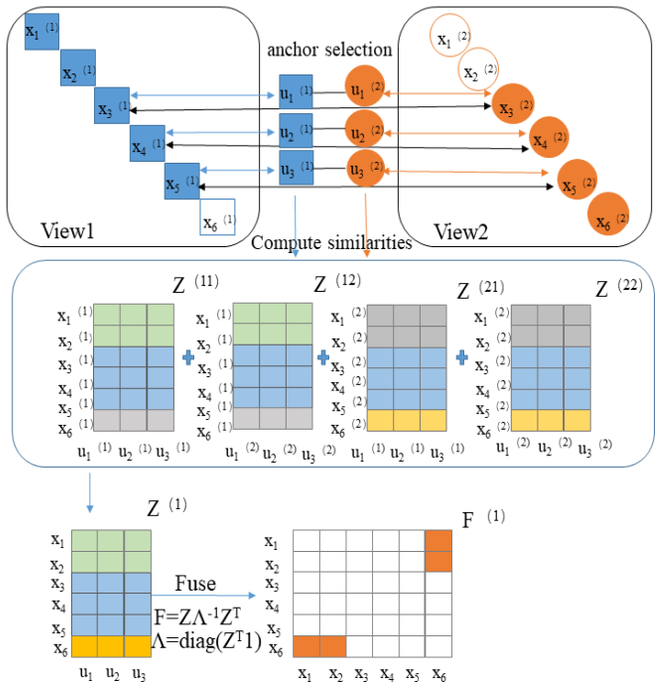


Figure 5.3: Construction of the Anchor-based similarity matrix.

similarity matrix using anchors [111]. Given the uniqueness of each view, our objective is to preserve the intrinsic feature information while ensuring consistency among multiple views as much as possible. Thus, we construct separate similarity graph structures for each view. Taking the similarity matrix of view 1 as an example, the construction process is illustrated in Figure 5.3.

We choose the instances x_3, x_4 and x_5 present in both view 1 and view 2 as anchor points, which we place into the corresponding anchor point set $U^{(v)}$ of the respective view. This set is denoted as $\left\{U_i^{(v)}\right\}_{i=1}^{l}$, where l represents the number of common instances occurring simultaneously in both views.

In order to obtain the final similarity matrix of view 1, we follow a four-step process. The first step is to calculate the similarity matrix $Z^{(11)}$ between all instances of view 1 and their own anchor points, which are selected from their own instances. In the second step, we calculate the similarity matrix $Z^{(12)}$ between all instances of view 1 and the anchor points from view 2. Finally, the third step is to calculate the similarity matrix $Z^{(21)}$ between all instances of view 2 and the anchor points from view 1. The fourth step is to calculate the similarity matrix $Z^{(22)}$ between all instances in view 2 and the anchor points from view 2. The similarity calculation process for instances from the same view and their corresponding anchors is as follows:

$$Z_{ij}^{(vv)} = \begin{cases} \frac{\exp\left(-\mathcal{D}^2\left(x_i^{(v)}, u_j^{(v)}\right)/\sigma^2\right)}{\sum_{j\in\langle i\rangle^v}\exp\left(-\mathcal{D}^2\left(x_i^{(v)} u_j^{(v)}\right)/\sigma^2\right)} & \forall j \in \langle i\rangle^v \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Here, $\langle i\rangle^{(v)}$ is the index set of the view instance corresponding to the selected anchor. The function $\mathcal{D}\left(x_i^{(v)}, u_j^{(v)}\right)$ is the distance function that measures the instance and the anchor. The kernel method is capable of capturing the nonlinear relationships within data, thereby enhancing the accuracy and efficiency of clustering algorithms. For the truncated similarity, we employ a Gaussian kernel function $\mathcal{K}_\sigma(\cdot)$, which is defined as follows:

$$\mathcal{K}_\sigma(x_i^{(v)}, u_j^{(v)}) = \exp\left(-\mathcal{D}^2\left(x_i^{(v)}, u_j^{(v)}\right)/\sigma^2\right) \tag{5.2}$$

For the sake of universality, we set σ to 1. In a similar manner to calculating $Z^{(11)}$, we use the same method to obtain $Z^{(12)}$, $Z^{(21)}$, and $Z^{(22)}$. As shown in Figure 5.3, we can obtain the final instance-anchor similarity matrix $Z^{(1)}$ by fusing the matrices as follows: $Z^{(1)} = [\tilde{Z}^{(12)}; \tilde{Z}^{(1)}; \tilde{Z}^{(21)}]$. Here, $\tilde{Z}^{(1)}$ represents the average of the similarity measures in $Z^{(11)}$, $Z^{(12)}$, $Z^{(21)}$, and $Z^{(22)}$, which contain common instances and different sets of anchor points chosen for different views. $\tilde{Z}^{(12)}$ and $\tilde{Z}^{(21)}$ represent the average similarity values between the anchor and the exclusively shared instances of $Z^{(11)}$ and $Z^{(12)}$ and between the anchors and the exclusively shared instances of $Z^{(21)}$, and $Z^{(22)}$, respectively.

In order to obtain the similarity matrix between all instances based on anchor graphs, we employ the low-rank approximation method mentioned in [111]. The following function is used for this calculation:

$$F^{(v)} = Z^{(v)} \Lambda^{(v)^{-1}} Z^{(v)^T} \tag{5.3}$$

where $\Lambda^{(v)} = diag\left(Z^{(v)^T} 1\right)$ is a diagonal matrix.

5.3.1.2 Feature learning of graph convolutional neural network

To encode each incomplete multiview data, we use a graph convolutional neural network $f_{gcn}^{(v)}$ to capture view-specific information. This network operates on the input data $X^{(v)}$ from the incomplete view itself, along with an anchor-based similarity graph structure $F^{(v)}$. By employing this approach, we can obtain the specific representation $S^{(v)}$ for the v-th view through the following function:

$$S^{(v)} = f_{gcn}^{(v)}\left(X^{(v)}, F^{(v)}\right) \tag{5.4}$$

Equation (5.4) mainly focuses on extracting graph structure information using anchor similarity, thereby ensuring consistent information across views. Consequently, to ensure that the learned view-specific representation contains more structural information about the views, we incorporate a loss function based on the decoder network structure $f_{de}^{(v)}$. This loss function aims to minimize the distance between the view representation and the structural representation, both derived from the anchor similarity graph in the metric space. The specific form of the loss function for learning $S^{(v)}$ is as follows:

$$\mathcal{L}_s = \sum_{v=1}^{V} \left\| X^{(v)} - f_{de}^{(v)}\left(S^{(v)}\right) \right\|_F^2 \tag{5.5}$$

In our model, we aim to learn specific view information $S^{(v)}$ that preserves inter-view consistency and intra-view specificity. This is achieved by utilizing Equation (5.4) and Equation (5.5).

5.3.2 Construct shared representation

In order to obtain a more representative, robust, and generalizable data representation that takes into account information bottlenecks, we introduced the incomplete multiview common representation denoted as H. At the same time, to better capture the data distribution of shared view H within a specific view, we use a reconstruction network $f^{(v)}$. This network facilitates the conversion between shared and multiview representations. The conversion function can be expressed as follows:

$$H^{(v)} = f^{(v)}(H) \tag{5.6}$$

Equation (5.6) focuses on obtaining the corresponding multiview representation information using the common representation of multiview. In order to ensure that the shared representation contains as much meaningful information as possible from the view-specific information $S^{(v)}$, we use the following reconstruction loss L_h to enhance the consistency between specific view information and specific view information obtained based on the common representations. The formulation of the reconstruction loss is as follows:

$$L_h = \min_{H} \sum_{v=1}^{V} \left\| S^{(v)} - f^{(v)}(H) \right\|_F^2 \tag{5.7}$$

5.3.3 Incomplete multiview information bottleneck representation

This section mainly consists of three modules: extracting view-specific via an information bottleneck, learning the diversity of multiview information via an information bottleneck, and reducing redundant features via an information bottleneck.

5.3.3.1 Extracting view-specific via an information bottleneck

Based on previous studies [153], maximizing the information bottleneck of two variables involves seeking a shared representation that maximizes the relevant information of these two variables. In Figure 5.2, we ensure the accuracy of extracting specific view information by maximizing $H^{(v)}$ and $S^{(v)}$ as follows:

$$\max_{H} \sum_{v=1}^{V} I\left(H^{(v)}, S^{(v)}\right) \tag{5.8}$$

With respect to the initial term $I\left(H^{(v)}, S^{(v)}\right)$, we can apply the definition of the mutual information as follow:

$$\begin{aligned} I\left(H^{(v)}, S^{(v)}\right) &= \int dsd\ \mathrm{h}p\left(\ \mathrm{s}^{(v)}, \mathrm{h}^{(v)}\right) \log \frac{p\left(\ \mathrm{s}^{(v)}, \mathrm{h}^{(v)}\right)}{p\left(\ \mathrm{s}^{(v)}\right) p\left(\ \mathrm{h}^{(v)}\right)} \\ &= \int dsd\ \mathrm{h}p\left(\ \mathrm{s}^{(v)}, \mathrm{h}^{(v)}\right) \log \frac{p\left(\ \mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right)}{p\left(\ \mathrm{s}^{(v)}\right)} \end{aligned} \tag{5.9}$$

Let $q\left(\ \mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right)$ be a variational approximation of $p\left(\ \mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right)$. Since the Kullback-Leibler divergence is always positive, we have:

$$
\begin{aligned}
&\mathrm{KL}\left[p\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right), q\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right)\right] \geq 0 \Rightarrow \\
&\int d\,\mathrm{s}^{(v)} p\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right) \log \frac{p\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right)}{q\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right)} \geq 0 \Rightarrow \\
&\int d\,\mathrm{s}^{(v)} p\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right) \log p\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right) \geq \\
&\int d\,\mathrm{s}^{(v)} p\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right) \log q\left(\mathrm{s}^{(v)} \mid \mathrm{h}^{(v)}\right).
\end{aligned}
\tag{5.10}
$$

So, we have a lower bound:

$$
\begin{aligned}
&I\left(H^{(v)}, S^{(v)}\right) \\
&\geq \int d\mathbf{s}^{(v)} d\mathbf{h}^{(v)} p\left(\mathrm{s}^{(v)}, \mathbf{h}^{(v)}\right) \log \frac{q\left(\mathrm{s}^{(v)} | \mathbf{h}^{(v)}\right)}{p\left(\mathrm{s}^{(v)}\right)} \\
&= \int d\mathbf{s}^{(v)} d\mathbf{h}^{(v)} p\left(\mathrm{s}^{(v)}, \mathbf{h}^{(v)}\right) \log q\left(\mathrm{s}^{(v)} \mid \mathbf{h}^{(v)}\right) + S\left(\mathrm{s}^{(v)}\right) \\
&\geq \int d\,\mathrm{s}^{(v)} d\mathbf{h}^{(v)} p\left(\mathrm{s}^{(v)}, \mathbf{h}^{(v)}\right) \log q\left(\mathrm{s}^{(v)} \mid \mathbf{h}^{(v)}\right) \\
&= \int d\,\mathrm{s}^{(v)} p\left(\mathrm{s}^{(v)}\right) \int d\mathbf{h}^{(v)} p\left(\mathrm{h}^{(v)} \mid \mathbf{s}^{(v)}\right) \log q\left(\mathrm{s}^{(v)} \mid \mathbf{h}^{(v)}\right)
\end{aligned}
\tag{5.11}
$$

In the actual calculation, maximizing mutual information can be regarded as a lower-bound problem [156].

5.3.3.2 Learning diversity of multiview information via information bottleneck

In order to ensure the diversity of multiview information, we expect the representation information derived from different views within a shared representation to be statistically independent of each other. As illustrated in Figure 5.2, given v views ($v \geq 2$), we establish pairwise relationships denoted as $\binom{v}{2}$ for all views. We denote the constructed view pairs as (i, m), where $i \neq m$. To achieve this independence, we minimize the mutual information between the specific view representation $H^{(m)}$ and the specific view representation $H^{(i)}$ derived from the shared representation. The corresponding function is as follows:

$$
\min I\left(H^{(i)}, H^{(m)}\right) \tag{5.12}
$$

Based on the definition of mutual information calculation, we can obtain the following formula:

$$
\begin{aligned}
&I\left(H^{(i)}, H^{(m)}\right) \\
&= \int d\,\mathrm{h}^{(m)} d\,\mathrm{h}^{(i)} p\left(H^{(i)}, H^{(m)}\right) \log \frac{p\left(H^{(i)} | H^{(m)}\right)}{p\left(H^{(i)}\right)}
\end{aligned}
\tag{5.13}
$$

Let $r\left(H^{(i)}\right)$ represent the variational approximation of this margin. Thus, we have:

$$
\begin{aligned}
&KL\left[p\left(H^{(i)}\right), r\left(H^{(i)}\right)\right] \geq 0 \\
&\Rightarrow \int d\,\mathrm{h}^{(m)} p\left(H^{(i)}\right) \log p\left(H^{(i)}\right) \\
&\geq \int d\,\mathrm{h}^{(m)} p\left(\mathrm{h}^{(m)}\right) \log r\left(\mathrm{h}^{(m)}\right)
\end{aligned}
\tag{5.14}
$$

Hence, we can compute the upper bound using the following expression:

$$\begin{aligned} & I\left(H^{(i)}, H^{(m)}\right) \\ & \leq \int d\,\mathrm{h}^{(m)} d\,\mathrm{h}^{(i)} p\left(H^{(m)}\right) p\left(H^{(i)} \mid H^{(m)}\right) \log \frac{p\left(H^{(i)} \mid H^{(m)}\right)}{r\left(H^{(i)}\right)} \\ & = \int d\ \mathrm{h}^{(m)} p\left(H^{(m)}\right) \int d\,\mathrm{h}^{(i)} p\left(H^{(i)} \mid H^{(m)}\right) \log \frac{p\left(H^{(i)} \mid H^{(m)}\right)}{r\left(H^{(i)}\right)} \end{aligned} \tag{5.15}$$

Since mutual information is always non-negative, the problem of minimizing mutual information can be viewed as seeking a tighter upper bound [156]. We cannot decrease the value of mutual information to a negative number or an arbitrarily small value. Therefore, the problem of minimizing mutual information can be interpreted as the pursuit of an upper bound.

5.3.3.3 Reducing unnecessary redundant features via information bottleneck

Algorithm 3 IMRL-AGI

Require: Incomplete multi-view dataset X=$\{X^{(v)}, ..., X^{(v)}\}$.
Ensure: The incomplete multi-view common representation H.

```
Generate the anchor point set U^(v);
Construct the similarity matrix Z^(v) between each instance in the view and the anchor point set;
Construct the similarity matrix F^(v) between each view instance by Eq. (5.3);
Initialize the incomplete multi-view common representation H;
Initialize the neural networks f_gcn^(v), f_de^(v), and f^(v) with appropriate parameters;
while Eq. (5.20) not convergent do do
    Calculate S^(v) by Eq. (5.4);
    Calculate H^(v) by Eq. (5.6);
    Calculate the loss of incomplete multi-view representation learning by Eq. (5.20);
    Update f_gcn^(v), f_de^(v), and f^(v) using backpropagation;
    Update H using backpropagation;
end while
Return the incomplete multi-view common representation H.
```

To enhance feature extraction and improve downstream tasks, we minimize the mutual information between the original view and the common representation. This process helps reduce noise and irrelevant information, allowing for the extraction of crucial and distinctive features from raw data. As a result, it improves the performance and generalization ability of downstream clustering and classification. Referring to Figure 5.2, we minimize the mutual information between the original view data $X^{(v)}$ and H in order to discard redundant information. Since our goal is to minimize mutual information, we utilize the calculation method described in the previous section. As such, we arrive at the following equation:

$$I\left(H, X^{(v)}\right) \leq \int dx p\left(\mathrm{x}^{(v)}\right) \int d\,\mathrm{h} p\left(\mathrm{~h} \mid \mathrm{x}^{(v)}\right) \log \frac{p\left(H \mid X^{(v)}\right)}{t(H)} \tag{5.16}$$

5.3.3.4 The Overall Objective Function of IMRL-AGI

To ensure the coherence of the three modules aforementioned, we integrate them into a unified learning model as follows:

$$\begin{aligned}
&\max_{H} \sum_{v=1}^{V} I\left(H^{(v)}, S^{(v)}\right) \\
&-\left(\sum_{i=1}^{V} \sum_{m=1}^{V} I\left(H^{(i)}, H^{(m)}\right)+\sum_{v=1}^{V}\left(H, X^{(v)}\right)\right) \\
\geq & \sum_{v=1}^{V} \int d \mathbf{s}^{(v)} p\left(\mathrm{~s}^{(v)}\right) \int d\ \mathrm{h}^{(v)} p\left(\mathrm{~h}^{(v)} \mid \mathrm{s}^{(v)}\right) \log q\left(\mathrm{~s}^{(v)} \mid \mathrm{h}^{(v)}\right) \\
&-\sum_{i=1}^{v} \sum_{m=1}^{v} \int d\ \mathrm{h}^{(m)} p\left(H^{(m)}\right) \int d\ \mathrm{h}^{(i)} p\left(H^{(i)} \mid H^{(m)}\right) \log \frac{p\left(H^{(i)} \mid H^{(m)}\right)}{r\left(H^{(i)}\right)} \\
&-\sum_{v=1}^{V} \int dx p\left(\mathrm{x}^{(v)}\right) \int d\ \mathrm{h} p\left(\mathrm{~h} \mid \mathrm{x}^{(v)}\right) \log \frac{p\left(H \mid X^{(v)}\right)}{t(H)}
\end{aligned} \tag{5.17}$$

The integration of $S^{(v)}$, $H^{(m)}$, and $X^{(v)}$ can be approximated using Monte Carlo sampling [157]. Here is the equation for the approximation:

$$\begin{aligned}
&\max_{H} \sum_{v=1}^{V} I\left(H^{(v)}, S^{(v)}\right) \\
&-\left(\sum_{i=1}^{v} \sum_{m=1}^{v} I\left(H^{(i)}, H^{(m)}\right)+\sum_{v=1}^{V}\left(H, X^{(v)}\right)\right) \\
\approx & \frac{1}{N} \sum_{n=1}^{N}\left\{\sum_{v=1}^{V} \int d\ \mathrm{h}^{(v)} p\left(\mathrm{~h}^{(v)} \mid \mathrm{S}_{n}{}^{(v)}\right) \log q\left(\mathrm{~S}_{n}^{(v)} \mid \mathrm{h}^{(v)}\right)\right. \\
&-\sum_{i=1}^{v} \sum_{m=1}^{v} \int d\ \mathrm{h}^{(i)} p\left(\mathrm{~h}^{(i)} \mid \mathrm{h}_{n}{}^{(m)}\right) \log \frac{p\left(\mathrm{~h}^{(i)} \mid \mathrm{h}_{n}{}^{(m)}\right)}{r\left(\mathrm{~h}^{(i)}\right)} \\
&\left.-\sum_{v=1}^{V} \int d\ \mathrm{h} p\left(\mathrm{~h} \mid \mathrm{x}_{n}{}^{(v)}\right) \log \frac{p\left(\mathrm{~h} \mid \mathrm{x}^{(v)}\right)}{t(\mathrm{~h})}\right\}
\end{aligned} \tag{5.18}$$

Next, we apply the reparameterization technique [158] to rewrite the function:

$$p\left(\mathrm{~h}^{(v)} \mid \mathrm{S}_{n}{}^{(v)}\right) d\ \mathrm{h}^{(v)}=p\left(\varepsilon_{1}^{(v)}\right) d \varepsilon_{1}^{(v)} \tag{5.19}$$

where $h^{(v)}{=}g_1\left(\mathrm{~s}^{(v)}, \varepsilon_1\right)$. Therefore, we learn incomplete multiview representations using

information bottlenecks, with the objective function $\mathcal{L}_r$ formulated as follows:

$$
\begin{aligned}
=&\frac{1}{N}\sum_{n=1}^{N}\Big\{-\mathrm{E}_{\varepsilon_1}\log \mathrm{q}\left(\mathrm{s}^{(v)} \mid g_1\left(\mathrm{s}^{(v)}, \varepsilon_1\right)\right) \\
&+\sum_{i=1}^{v}\sum_{m=1}^{v} D_{KL}\left[p\left(\mathrm{h}^{(i)} \mid \mathrm{h}_n^{(m)}\right), r\left(\mathrm{h}^{(i)}\right)\right] \\
&+\sum_{v=1}^{V} D_{KL}\left[p\left(\mathrm{h} \mid \mathrm{x}^{(v)}\right), t(\mathrm{h})\right] \\
&\}
\end{aligned}
\tag{5.20}
$$

In Algorithm 5.3.3.3, we outline the complete updating procedures of the proposed IMRL-AGI.

5.4 Experiments

In this section, we conducted experiments to compare the performance of the proposed algorithm, IMRL-AGI, against other comparative baselines on commonly used datasets. We also provided a comprehensive analysis of the experimental results. The IMRL-AGI algorithm was implemented using the Basesklearn and PyTorch libraries. In addition, all of the implemented algorithms were executed on a Windows computing platform equipped with an NVIDIA 3090 GPU.

5.4.1 Descriptions of datasets

Table 5.1: Mean ACCs (%) of different methods on the Notting-Hill, ORL, and Yale datasets

Dataset	Rate	HCP_IMSC	DAIMC	UEAF	IMC-GRMF	Ours
Notting-Hill	0.1	73.45±0.00	73.76±6.15	81.78±0.11	77.58±0.08	**81.80±0.32**
	0.3	79.64±0.00	74.69±7.22	**82.18±0.00**	69.27±0.00	78.89±0.72
	0.5	80.45±0.19	73.64±4.29	82.73±0.00	70.36±0.00	**83.69±0.18**
	0.7	80.18±0.00	71.49±4.06	82.13±0.00	71.25±0.05	**82.15±1.54**
	0.9	82.18±0.00	75.82±6.27	82.36±0.00	75.82±0.00	**85.64±3.83**
ORL	0.1	58.33±1.31	65.42±1.68	53.05±2.05	33.62±0.95	**72.17±2.02**
	0.3	60.88±1.71	66.50±2.41	55.58±1.98	40.00±2.90	**72.38±2.15**
	0.5	65.03±1.48	67.30±4.92	55.98±2.18	52.33±1.63	**74.10±1.99**
	0.7	64.08±2.27	67.67±2.23	55.15±1.94	59.38±2.21	**77.90±1.81**
	0.9	68.40±2.00	68.53±2.51	55.67±2.31	69.30±3.02	**75.05±1.75**
Yale	0.1	42.48±2.48	34.42±3.40	55.58±4.46	36.06±1.06	**59.52±1.55**
	0.3	48.48±3.22	37.21±2.30	55.15±3.08	37.82±2.17	**59.09±2.32**
	0.5	54.79±3.09	48.79±4.17	54.85±2.90	48.30±3.36	**60.85±1.28**
	0.7	60.25±3.65	50.18±3.06	55.88±3.10	58.24±1.42	**60.30±0.82**
	0.9	56.85±1.93	53.21±4.40	55.21±4.16	57.64±1.03	**58.55±0.91**

Table 5.2: Mean NMIs (%) of different methods on the Notting-Hill, ORL and Yale datasets

Dataset	Rate	HCP_IMSC	DAIMC	UEAF	IMC-GRMF	Ours
Notting-Hill	0.1	71.25±0.00	73.00±3.83	74.26±0.31	72.11±0.18	**74.28±0.63**
	0.3	68.74±0.13	72.47±3.13	73.75±0.00	72.51±0.00	**75.12±1.05**
	0.5	77.17±0.37	74.70±3.02	76.08±0.00	75.81±0.00	**78.63±0.80**
	0.7	76.90±0.00	70.61±2.56	76.36±0.00	76.94±0.19	**78.99±1.43**
	0.9	78.89±0.00	74.03±3.98	76.29±0.00	79.65±0.00	**82.17±3.21**
ORL	0.1	75.49±0.70	80.05±0.86	72.68±1.34	55.82±0.89	**82.52±0.86**
	0.3	77.38±0.82	80.90±0.66	75.25±1.26	61.04±1.77	**83.54±0.66**
	0.5	80.81±0.88	83.43±2.17	75.67±1.33	70.51±1.01	**84.49±0.85**
	0.7	81.11±1.01	84.18±1.21	75.35±0.73	76.38±1.21	**87.50±0.67**
	0.9	83.91±0.92	84.79±0.96	75.81±0.95	82.79±1.06	**86.49±0.41**
Yale	0.1	47.92±1.59	41.36±3.43	61.97±2.93	45.27±0.60	**63.62±1.14**
	0.3	53.36±2.49	44.24±1.75	60.84±1.64	46.55±1.73	**63.91±1.34**
	0.5	54.79±3.09	54.26±2.75	61.02±1.10	56.58±2.17	**64.99±1.36**
	0.7	63.77±2.43	53.68±1.83	61.87±2.24	64.12±1.73	**64.32±1.03**
	0.9	**64.42±1.11**	57.16±3.99	61.21±2.32	63.67±0.99	62.65±0.74

- **Notting-Hill:**[1] The Notting Hill dataset is a facial datasetextracted from the movie "Notting Hill" dataset. It mainly consists of facial images of 5 actors, with each actor having 110 images. The datapoint in the dataset contains three views, each with dimensions of 2000, 3304, and 6750.

- **ORL:**[2] The ORL facial dataset is a collection of 400 images, divided into 40 categories. The dataset is organized based on variations in lighting, facial expressions, and facial details. It also has three views, each with dimensions of 4096, 3304, and 6750.

- **Yale:**[3] The Yale dataset contains 165 images of 15 individuals. Each person has 11 facial images with different expressions, postures, and lighting conditions. Similar to the ORL dataset, the Yale dataset also has three views with the same feature dimension as the ORL dataset views.

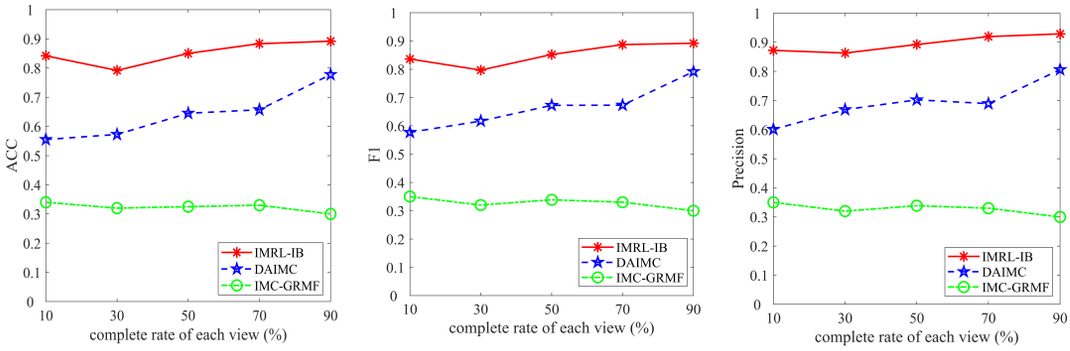
5.4.2 Incomplete percentage settings

We constructed different incomplete percentage settings for the above multi-view datasets. We denoted the percentage proportion of randomly selected instances as p. The datasets were set with p vales of {10%, 30%, 50%, 70%, 90%}. We saved the instances associated with the indexes in all views using the randomly selected instance indexes. Then, we evenly distributed the remaining instances to each view based on the number of views.

[1]https://github.com/Multiviewdate?tab=repositories
[2]https://www.kaggle.com/datasets/att-database-of-faces
[3]http://www.uk.research.att.com/facedatabase.html

5.4.3 Compared methods

To evaluate the clustering performance of representations obtained by IMRL-AGI, we compared it with four state-of-the-art IMC methods: HCP_IMSC [99], DAIMC [48], IMC_GRMF [46], and UEAF [98].



(a) (b) (c)

Figure 5.4: Performance comparison of different methods on the ORL dataset in terms of classification

Table 5.3: Ablation study of different representations in relation to the clustering task

Metrics	RATE	$F^{(*)}$	$S^{(1)}$	$S^{(2)}$	$S^{(3)}$	H
ACCs	0.1	60.55±0.00	77.43±0.25	66.42±1.24	66.60±2.06	**81.80±0.32**
	0.3	57.27±0.00	76.78±0.08	55.11±1.18	70.73±0.00	**78.89±0.72**
	0.5	40.18±0.00	82.00±0.00	65.24±5.61	**84.13±1.65**	83.62±1.09
	0.7	57.58±0.50	82.00±0.00	57.04±0.49	74.95±0.11	**83.20±0.59**
	0.9	57.45±2.36	83.27±0.00	74.51±8.00	83.20±2.92	**85.64±3.83**
NMIs	0.1	57.45±0.00	66.29±0.45	54.49±1.55	62.66±1.07	**74.28±0.63**
	0.3	49.00±0.00	70.46±0.16	37.31±1.13	62.45±0.00	**75.12±1.05**
	0.5	24.00±0.00	73.24±0.00	65.14±2.34	72.40±1.97	**78.17±0.30**
	0.7	48.27±2.22	74.10±0.24	42.05±0.41	66.31±0.01	**76.33±0.34**
	0.9	52.21±6.43	78.33±0.00	71.29±1.96	76.00±1.68	**82.17±3.21**
F1s	0.1	63.88±0.00	72.14±0.36	59.09±1.43	65.70±1.05	**77.66±0.44**
	0.3	62.18±0.00	76.62±0.13	47.01±1.43	66.61±0.00	**79.60±1.16**
	0.5	47.83±0.00	77.01±0.00	67.52±2.15	75.71±2.42	**81.03±0.49**
	0.7	59.05±1.25	76.69±0.08	48.42±0.41	69.93±0.00	**78.75±0.30**
	0.9	62.23±3.26	80.10±0.00	72.51±1.89	78.39±2.04	**83.62±3.55**
ARIs	0.1	46.45±0.00	67.75±0.99	45.28±2.13	54.34±2.04	**74.18±0.56**
	0.3	29.24±0.00	67.44±0.24	27.48±0.83	57.51±0.00	**71.77±0.97**
	0.5	6.40±0.00	73.24±0.00	51.93±3.82	72.55±4.02	**77.15±0.87**
	0.7	33.35±3.12	73.70±0.12	29.48±0.65	63.58±0.01	**75.93±0.34**
	0.9	35.56±7.93	76.74±0.00	60.97±3.61	74.63±3.57	**81.36±4.20**

- **HCP_IMSC:** HCP_IMSC focuses on learning low-rank tensor representations in tensors constructed from incomplete multi-view affinity matrices. It employs weighted

fusion to combine the affinity matrices of different views into a consistent affinity matrix. Additionally, hypergraphs are used to effectively recover missing data.

- **DAIMC:** DAIMC is a decomposition-based method that learns a shared latent feature space using instance alignment information. It incorporates a weight indicator matrix to reveal the missing information about the internal instances within each view. The introduction of the $L_{2,1}$-Norm on the regression coefficient matrix ensures consistency in the learned basis matrix.

- **IMC_GRMF:** IMC_GRMF introduces reconstruction operations to preserve the original data geometry and incorporates binary weights for data normalization. It includes orthogonalization processing to handle out-of-instance data and adds a regularisation term to enforce consistency in the representation of paired instances across different views.

- **UEAF:** UEAF first introduces reconstruction items to fill in missing data and ensure natural alignment among views. It leverages available information from non-missing views and hidden information from missing views for shared representation learning. The method incorporates reverse graph rules to capture the local structure information and employs an adaptive weighting strategy to capture the importance of different views.

5.4.4 Evaluation metrics

In our experiments, we utilized five commonly used indicators to measure the performance of clustering and classification, including Accuracy(ACC) [101], Normalized Mutual Information(NMI) [100], Adjusted Rand Index (ARI) [102], F1 score (F1) [103], and precision [104].

5.4.5 Evaluation on cluster performance

In Table 5.1 to 5.2, we report the experimental results on three datasets. The overall results demonstrated that our method consistently outperformed the other methods, especially when the Yale dataset had only 10% completeness. Compared with the latest algorithm HCP_IMSC, our method IMRL-AGI achieved a 17.04% higher accuracy (ACC) value. This indicates that the incomplete multi-view representation data obtained through information bottlenecks and anchor graph GCN is beneficial for image clustering tasks.

Furthermore, as the completeness percentage increased, our method still maintained stable and excellent performance across different completeness levels, while other methods showed relatively large fluctuations. For example, on the ORL dataset, the NMI value of IMRL-AGI remained stable in the range of 82.52%-86.49% as the completeness varied. This

stability can be attributed to the minimization of different view information in the common representation and the mutual information between the original view information and the multi-view common representation. As a result, the obtained representation by IMRL-AGI is robust and generalizable for the subsequent clustering tasks.

According to Table 5.1 to 5.2, we can observe that as the integrity of the dataset changes, the clustering performances of HCP_IMSC, UEAF, and our proposed method IMRL-AGI are better than these of DAIMC and IMC_GRMF. This indicates that introducing affinity matrices to capture instances and view feature correlations is an effective method for obtaining view feature information and beneficial for the subsequent clustering tasks.

5.4.6 Evaluation on classification performance

In this experiment, we evaluated the classification performance of our method, DAIMC and IMC-GRMF using K-Nearest Neighbor classifiers on the ORL dataset with three incomplete multi-view representations obtained. The completeness levels were set from 10% to 90% with an interval of 20%. We used ACC, F-score, and accuracy as evaluation indicators to assess the effectiveness of classification. According to Figure 5.4, our method consistently performed the best in classification across all completeness levels. Moreover, our method exhibits better overall robustness while maintaining excellent performance. For example, considering accuracy, the difference between the highest and lowest values obtained by our method is 6.57%, whereas the difference for DAIMC is 21%. This indicates that our method has significantly better robustness compared to DAIMC. Therefore, our method not only achieves superior classification results but also demonstrates more stability and consistency across different completeness levels.

5.4.7 Ablation study

We conducted ablation experiments to investigate the role of the feature values obtained in each step. Clustering operations were performed on the selected optimal graph structure data $F^{(*)}$, specific view feature representation data $S^{(v)}$, and shared feature representation H, respectively. The results of these experiments are presented in Table 5.3. From the experimental results, it is evident that the shared feature representations learned through the IMRL-AGI method exhibit the best overall clustering performance. The second-best performance is observed with the specific view feature representation data $S^{(v)}$.

On the other hand, the obtained optimal graph structure data $F^{(*)}$ yields the worst performance. This suggests that the combination of anchor graph and GCN plays a crucial role in obtaining the best-shared feature representation H based on information bottlenecks. The shared feature representation obtained through the IMRL-AGI method proves to be the most effective in improving clustering performance.

5.5 Conclusions

This paper presents a novel framework for learning incomplete multi-view common representations from the perspective of information theory. The proposed method, IMRL-AGI, follows a two-step process to reconstruct the instances relationships and extract feature information from views. In the first step, IMRL-AGI utilizes anchors to reconstruct the relationships between instances. In the second step, a graph convolutional neural network (GCN) based on anchor graphs is designed to extract feature information from views. To further improve the extraction of view information, IMRL-AGI maximizes the mutual information representation between the view information derived from the shared representation and the view modification information obtained using anchor graph GCN. In addition, IMRL-AGI imposes constraints on the minimum mutual information between different extracted view information and between the original view information and the multi-view shared representation. Experimental results show that our method has achieved good performance compared to existing methods in both clustering and classification tasks.

Chapter 6

Discussion

Contents

6.1 Implications of Findings

In my research, we have identified various challenges and possible solutions for incomplete multi-view clustering, which has significant relevance for data analysis and mining across various domains. The following outlines the key significance of our research.

1) Pioneering Innovative Solutions: Incomplete multi-view clustering, as a relatively emerging field, has not yet received comprehensive research attention. My research addresses this gap by concentrating on resolving three key issues: complex missing situations from incomplete multi-view, high missing data rate handling, and the redundancy issues of incomplete multi-view. Consequently, it furnishes a broader array of methods and techniques for dealing with incomplete multi-view data.

2) Practical Applications: In real-life scenarios, a substantial amount of data is often incomplete, particularly when dealing with multi-view data. My research outcomes can be readily applied across various domains, including healthcare, social network analysis, and financial risk management, thereby enhancing the efficiency of addressing practical problems.

3) Advancements in Data Mining: Research in incomplete multi-view clustering contributes to the progress of the data mining field. My proposed methods and algorithms in this thesis offer valuable tools to fellow researchers, assisting them in the more effective handling of incomplete multi-view data and gaining profound insights from it.

4) Inspiration for Future Research: This study provides guidance for future research initiatives. Our work highlights key issues within the field of incomplete multi-view clustering and offers solutions. This will encourage other researchers to delve deeper and enhance existing methods.

6.2 Challenges and Limitations

In the era of big data, incomplete multi-view clustering has emerged as an important field of research. Focusing on the complexities of multi-view data such as missing data, high rates of missing information, and excessive redundancy, this research confronts numerous issues and challenges. My research initiated to introduce several approaches for handling complex missing data, high rates of missing information, and excessive redundancy in incomplete multi-view data and experimented with related data fusion methods. Despite making some progress in the field of incomplete multi-view clustering, we still face many challenges and limitations that need to be further addressed in the future.

1) The tensor model established in Chapter III effectively integrates high-order representations from different views, resulting in a consistent final representation. However, in the subsequent clustering process, it is susceptible to the influence of initial values due to being

executed separately. Therefore, better integration of the tensor model and clustering model becomes one of the research directions that needs exploration in the future of incomplete multi-view clustering research.

2) In Chapter IV of this thesis, a low-rank and sparse-based incomplete multi-view clustering model is constructed. This method not only applies low-rank sparse representation but also takes into account both local and global multi-view structures. However, in many practical applications, obtaining labels for missing data may be challenging, leading to the issue of data mismatch. Therefore, considering both incompleteness and data mismatch comprehensively is expected to be a significant focus of future research.

3) Faced with the redundancy issue in incomplete multi-view data, Chapter IV of this dissertation proposes an Information Bottleneck-based approach for incomplete multi-view clustering. This method further obtains a consistent final representation by maximizing and minimizing different multi-view correlations. Multi-view data typically consists of high-dimensional features, thereby increasing the computational complexity of clustering. Therefore, future research may consider combining the Information Bottleneck method with dimensionality reduction techniques, such as Principal Component Analysis (PCA) or t-SNE, to reduce data dimensionality while preserving essential information.

4) Incomplete multi-view clustering faces several challenges and limitations in terms of data. Firstly, data from different views may exhibit disparities in feature representation, scale, or quality, potentially leading to inconsistent or inaccurate clustering results. Secondly, the collection and integration of data can be constrained, as some view-specific data may be challenging to acquire or subject to restrictions due to privacy or security concerns. These factors collectively impact the effectiveness and feasibility of incomplete multi-view clustering.

6.3 Future Research Directions

While methods for incomplete multi-view clustering (IMC) have become a focal point of research for many, there are still some issues that have not been effectively addressed.

1) In-depth Theoretical Research: Further in-depth research on the theory of IMC is needed, exploring issues related to its properties, convergence, and other aspects.

2) Enhancing Method Robustness and Efficiency: In the future, it is imperative to conduct in-depth research on enhancing the robustness of IMC methods. This will facilitate better handling of diverse forms of data missingness and noise. Furthermore, optimizing methods to enhance computational efficiency represents a pivotal avenue for future exploration.

3) The Challenge of Scalability: In practical clustering tasks, the volume of data samples

involved is often substantial. However, many IMC methods encounter significant challenges in terms of computational complexity and memory costs, especially those based on kernel and graph techniques. As a result, existing IMC approaches are ill-suited for handling large-scale datasets. Therefore, it is of paramount importance to devise efficient methodologies that simultaneously address efficiency and performance for large-scale IMC tasks.

4) Imbalance and Mismatch in Data: The issue of imbalance and mismatch in data poses a challenge. Due to the random absence of views, instances across different views often exhibit not only varying quantities but also uncertain degrees of correspondence. Furthermore, substantial disparities in feature dimensions and magnitudes exist among different views. Regrettably, prevailing IMC methods have consistently overlooked the holistic consideration of this array of factors, detrimentally impacting clustering outcomes. Therefore, there is a burgeoning research direction within the IMC domain to devise more robust clustering models by accounting for the imbalance and mismatch properties of data.

6.4 Conclusion

This study addresses the issue of incomplete multi-view clustering by introducing three innovative approaches. Firstly, we propose a novel algorithm named IMC-NLT, which relies on Non-Negative Matrix Factorization (NMF) and low-rank tensor fusion. IMC-NLT effectively integrates information within the same view and across views, leveraging modal unified dimensionality and low-rank tensors. This method not only handles various types of incomplete data proficiently but also exhibits reduced sensitivity to its parameters.

Secondly, we present a method termed CCIM-SLR, based on sparse matrix techniques, to tackle multi-view clustering challenges. By employing sparse low-rank learning, CCIM-SLR quantifies correlations among samples within the same view and captures correlations across different views, thus achieving lower-dimensional representations. CCIM-SLR learns shared latent space and visible views, while mitigating post-processing sensitivity, such as initial parameter values for K-means, through alternating partitioning clustering.

Lastly, we introduce an innovative framework, IMRL-AGI, for learning joint representations from incomplete multi-views. IMRL-AGI follows a two-step strategy: initially, it reconstructs instance relationships using anchor points; subsequently, it employs a Graph Convolutional Network (GCN) based on the anchor graph to extract feature information from the views. To enhance view information extraction, IMRL-AGI maximizes the mutual information between shared representations derived view information and view refinement information obtained through anchors. Moreover, IMRL-AGI considers minimal mutual information among different extraction views, as well as the relationship between original view information and multi-view shared representations.

In conclusion, this study's three proposed methods offer innovative solutions for incomplete multi-view clustering challenges. These methods effectively capture inter-view correlations and shared features, making valuable contributions to the field of multi-view data analysis.

References

[1] S. Shi, F. Nie, R. Wang, and X. Li, "Multi-view clustering via nonnegative and orthogonal graph reconstruction," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 1, pp. 201–214, 2023. [Online]. Available: https://doi.org/10.1109/TNNLS.2021.3093297

[2] J. Wen, Z. Zhang, L. Fei, B. Zhang, Y. Xu, Z. Zhang, and J. Li, "A survey on incomplete multiview clustering," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 53, no. 2, pp. 1136–1149, 2023. [Online]. Available: https://doi.org/10.1109/TSMC.2022.3192635

[3] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, S. Singh and S. Markovitch, Eds. AAAI Press, 2017, pp. 2921–2927. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14647

[4] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, no. 7-8, pp. 2031–2038, 2013. [Online]. Available: https://doi.org/10.1007/s00521-013-1362-6

[5] M. Brbic and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognit.*, vol. 73, pp. 247–258, 2018. [Online]. Available: https://doi.org/10.1016/j.patcog.2017.08.024

[6] Y. Qin, G. Feng, Y. Ren, and X. Zhang, "Consistency-induced multiview subspace clustering," *IEEE Trans. Cybern.*, vol. 53, no. 2, pp. 832–844, 2023. [Online]. Available: https://doi.org/10.1109/TCYB.2022.3165550

[7] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4238–4246.

[8] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, 2020. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2877660

[9] S. Luo, C. Zhang, W. Zhang, and X. Cao, "Consistent and specific multi-view subspace clustering," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, S. A. McIlraith and K. Q. Weinberger, Eds. AAAI Press, 2018, pp. 3730–3737. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16212

[10] X. Cai, D. Huang, G. Zhang, and C. Wang, "Seeking commonness and inconsistencies: A jointly smoothed approach to multi-view subspace clustering," *Inf. Fusion*, vol. 91, pp. 364–375, 2023. [Online]. Available: https://doi.org/10.1016/j.inffus.2022.10.020

[11] J. Gao, J. Han, J. Liu, and C. Wang, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA*. SIAM, 2013, pp. 252–260. [Online]. Available: https://doi.org/10.1137/1.9781611972832.28

[12] G. Li, D. Song, W. Bai, K. Han, and R. Tharmarasa, "Consensus and complementary regularized non-negative matrix factorization for multi-view image clustering," *Inf. Sci.*, vol. 623, pp. 524–538, 2023. [Online]. Available: https://doi.org/10.1016/j.ins.2022.12.063

[13] J. Li, G. Zhou, Y. Qiu, Y. Wang, Y. Zhang, and S. Xie, "Deep graph regularized non-negative matrix factorization for multi-view clustering," *Neurocomputing*, vol. 390, pp. 108–116, 2020. [Online]. Available: https://doi.org/10.1016/j.neucom.2019.12.054

[14] D. Wang, T. Li, W. Huang, Z. Luo, P. Deng, P. Zhang, and M. Ma, "A multi-view clustering algorithm based on deep semi-nmf," *Inf. Fusion*, vol. 99, p. 101884, 2023. [Online]. Available: https://doi.org/10.1016/j.inffus.2023.101884

[15] H. Wang, Y. Yang, and B. Liu, "GMC: graph-based multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1116–1129, 2020. [Online]. Available: https://doi.org/10.1109/TKDE.2019.2903810

[16] R. Wang, F. Nie, Z. Wang, H. Hu, and X. Li, "Parameter-free weighted multi-view projected clustering with structured graph learning," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2014–2025, 2020. [Online]. Available: https://doi.org/10.1109/TKDE.2019.2913377

[17] Y. Du, G. Lu, and G. Ji, "Robust and optimal neighborhood graph learning for multi-view clustering," *Inf. Sci.*, vol. 631, pp. 429–448, 2023. [Online]. Available: https://doi.org/10.1016/j.ins.2023.02.089

[18] X. Peng, Z. Huang, J. Lv, H. Zhu, and J. T. Zhou, "COMIC: multi-view clustering without parameter selection," in *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 2019, pp. 5092–5101. [Online]. Available: http://proceedings.mlr.press/v97/peng19a.html

[19] K. Zhan, C. Niu, C. Chen, F. Nie, C. Zhang, and Y. Yang, "Graph structure fusion for multiview clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1984–1993, 2019. [Online]. Available: https://doi.org/10.1109/TKDE.2018.2872061

[20] J. Ngiam, A. Khosla, M. Kim, J. Nam, and A. Y. Ng, "Multimodal deep learning," in *International Conference on Machine Learning*, 2009.

[21] Z. Li, Q. Wang, Z. Tao, Q. Gao, and Z. Yang, "Deep adversarial multi-view clustering network," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 2952–2958. [Online]. Available: https://doi.org/10.24963/ijcai.2019/409

[22] J. Xu, Y. Ren, H. Tang, X. Pu, X. Zhu, M. Zeng, and L. He, "Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering," in *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*. IEEE, 2021, pp. 9214–9223. [Online]. Available: https://doi.org/10.1109/ICCV48922.2021.00910

[23] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *International conference on machine learning*. PMLR, 2013, pp. 1247–1255.

[24] L. Zhao, X. Wang, Z. Liu, H. Yuan, J. Zhao, and S. Zhou, "Deep probability multi-view feature learning for data clustering," *Expert Syst. Appl.*, vol. 217, p. 119458, 2023. [Online]. Available: https://doi.org/10.1016/j.eswa.2022.119458

[25] Z. Huang, Y. Ren, X. Pu, S. Huang, Z. Xu, and L. He, "Self-supervised graph attention networks for deep weighted multi-view clustering," in *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, B. Williams, Y. Chen, and J. Neville, Eds. AAAI Press, 2023, pp. 7936–7943. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/25960

[26] N. Zhang, "Methodolgical progress note: Handling missing data in clinical research," *Journal of Hospital Medicine*, vol. 14, no. 2019-11-20 ONLINE FIRST, 2019.

[27] R. Gomila and C. S. Clark, "Missing data in experiments: Challenges and solutions." *Psychological methods*, vol. 27, no. 2, p. 143, 2022.

[28] W. Lin and C. Tsai, "Missing value imputation: a review and analysis of the literature (2006-2017)," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 1487–1509, 2020. [Online]. Available: https://doi.org/10.1007/s10462-019-09709-4

[29] S. Fletcher Mercaldo and J. D. Blume, "Missing data and prediction: the pattern submodel," *Biostatistics*, vol. 21, no. 2, pp. 236–252, 2020.

[30] M. S. Santos, R. C. Pereira, A. F. Costa, J. P. Soares, J. A. M. Santos, and P. H. Abreu, "Generating synthetic missing data: A review by missing mechanism," *IEEE Access*, vol. 7, pp. 11 651–11 667, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2891360

[31] P. Patidar and A. Tiwari, "Handling missing value in decision tree algorithm," *International Journal of Computer Applications*, vol. 70, no. 13, 2013.

[32] N. J. Horton and N. M. Laird, "Maximum likelihood analysis of generalized linear models with missing covariates," *Statistical Methods in Medical Research*, vol. 8, no. 1, pp. 37–50, 1999.

[33] D. Bertsimas, C. Pawlowski, and Y. D. Zhuo, "From predictive methods to missing data imputation: an optimization approach." *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 7133–7171, 2017.

[34] K. Maheswari, P. P. A. Priya, S. Ramkumar, and M. Arun, "Missing data handling by mean imputation method and statistical," in *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing: BDCC 2018*. Springer Nature, 2019, p. 137.

[35] S. Wang, M. Li, N. Hu, E. Zhu, J. Hu, X. Liu, and J. Yin, "K-means clustering with incomplete data," *IEEE Access*, vol. 7, pp. 69 162–69 171, 2019.

[36] G. Kabir, S. Tesfamariam, J. Hemsing, and R. Sadiq, "Handling incomplete and missing data in water network database using imputation methods," *Sustainable and Resilient Infrastructure*, vol. 5, no. 6, pp. 365–377, 2020.

[37] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society: series B (methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[38] P. Raja and K. Thangavel, "Soft clustering based missing value imputation," in *Digital Connectivity–Social Impact: 51st Annual Convention of the Computer Society of India, CSI 2016, Coimbatore, India, December 8-9, 2016, Proceedings 51*. Springer, 2016, pp. 119–133.

[39] M. Smieja, L. Struski, J. Tabor, B. Zielinski, and P. Spurek, "Processing of missing data by neural networks," in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., 2018, pp. 2724–2734. [Online]. Available: https://proceedings.neurips.cc/paper/2018/hash/411ae1bf081d1674ca6091f8c59a266f-Abstract.html

[40] G. C. Wei and M. A. Tanner, "A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms," *Journal of the American statistical Association*, vol. 85, no. 411, pp. 699–704, 1990.

[41] C. Ye, H. Wang, W. Lu, and J. Li, "Effective bayesian-network-based missing value imputation enhanced by crowdsourcing," *Knowledge-Based Systems*, vol. 190, p. 105199, 2020.

[42] J. Tang, X. Zhang, W. Yin, Y. Zou, and Y. Wang, "Missing data imputation for traffic flow based on combination of fuzzy neural network and rough set theory," *J. Intell. Transp. Syst.*, vol. 25, no. 5, pp. 439–454, 2021. [Online]. Available: https://doi.org/10.1080/15472450.2020.1713772

[43] S. Li, Y. Jiang, and Z. Zhou, "Partial multi-view clustering," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, C. E. Brodley and P. Stone, Eds. AAAI Press, 2014, pp. 1968–1974. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8241

[44] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, S. Kambhampati, Ed. IJCAI/AAAI Press, 2016, pp. 2392–2398. [Online]. Available: http://www.ijcai.org/Abstract/16/341

[45] N. Xu, Y. Guo, X. Zheng, Q. Wang, and X. Luo, "Partial multi-view subspace clustering," in *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22-26, 2018*, S. Boll, K. M. Lee, J. Luo, W. Zhu, H. Byun, C. W. Chen, R. Lienhart, and T. Mei, Eds. ACM, 2018, pp. 1794–1801. [Online]. Available: https://doi.org/10.1145/3240508.3240679

[46] J. Wen, Z. Zhang, Y. Xu, and Z. Zhong, "Incomplete multi-view clustering via graph regularized matrix factorization," in *Computer Vision - ECCV 2018 Workshops - Munich, Germany, September 8-14, 2018, Proceedings, Part IV*, ser. Lecture Notes in Computer Science, L. Leal-Taixé and S. Roth, Eds., vol. 11132. Springer, 2018, pp. 593–608. [Online]. Available: https://doi.org/10.1007/978-3-030-11018-5_47

[47] W. Shao, L. He, C. Lu, and P. S. Yu, "Online multi-view clustering with incomplete views," in *2016 IEEE International Conference on Big Data (IEEE BigData 2016), Washington DC, USA, December 5-8, 2016*, J. Joshi, G. Karypis, L. Liu, X. Hu, R. Ak, Y. Xia, W. Xu, A. Sato, S. Rachuri, L. H. Ungar, P. S. Yu, R. Govindaraju, and T. Suzumura, Eds. IEEE Computer Society, 2016, pp. 1012–1017. [Online]. Available: https://doi.org/10.1109/BigData.2016.7840701

[48] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, J. Lang, Ed. ijcai.org, 2018, pp. 2262–2268. [Online]. Available: https://doi.org/10.24963/ijcai.2018/313

[49] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, "Partial multi-view clustering using graph regularized nmf," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 2192–2197.

[50] M. Hu and S. Chen, "One-pass incomplete multi-view clustering," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 3838–3845. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.33013838

[51] C. Xu, D. Tao, and C. Xu, "Multi-view learning with incomplete views," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5812–5825, 2015. [Online]. Available: https://doi.org/10.1109/TIP.2015.2490539

[52] J. Wen, H. Sun, L. Fei, J. Li, Z. Zhang, and B. Zhang, "Consensus guided incomplete multi-view spectral clustering," *Neural Networks*, vol. 133, pp. 207–219, 2021. [Online]. Available: https://doi.org/10.1016/j.neunet.2020.10.014

[53] J. Wu, H. Liu, H. Xiong, J. Cao, and J. Chen, "K-means-based consensus clustering: A unified view," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 1, pp. 155–169, 2015. [Online]. Available: https://doi.org/10.1109/TKDE.2014.2316512

[54] A. Trivedi, P. Rai, H. Daumé III, and S. L. DuVall, "Multiview clustering with incomplete views," in *NIPS workshop*, vol. 224. Citeseer, 2010, pp. 1–8.

[55] W. Shao, X. Shi, and P. S. Yu, "Clustering on multiple incomplete datasets via collective kernel learning," in *2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA, December 7-10, 2013*, H. Xiong, G. Karypis, B. Thuraisingham, D. J. Cook, and X. Wu, Eds. IEEE Computer Society, 2013, pp. 1181–1186. [Online]. Available: https://doi.org/10.1109/ICDM.2013.117

[56] Y. Ye, X. Liu, Q. Liu, and J. Yin, "Consensus kernel k-means clustering for incomplete multiview data," *Comput. Intell. Neurosci.*, vol. 2017, pp. 3 961 718:1–3 961 718:11, 2017. [Online]. Available: https://doi.org/10.1155/2017/3961718

[57] X. Zhu, X. Liu, M. Li, E. Zhu, L. Liu, Z. Cai, J. Yin, and W. Gao, "Localized incomplete multiple kernel k-means," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, J. Lang, Ed. ijcai.org, 2018, pp. 3271–3277. [Online]. Available: https://doi.org/10.24963/ijcai.2018/454

[58] X. Liu, M. Li, L. Wang, Y. Dou, J. Yin, and E. Zhu, "Multiple kernel k-means with incomplete kernels," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, S. Singh and S. Markovitch, Eds. AAAI Press, 2017, pp. 2259–2265. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14758

[59] X. Liu, X. Zhu, M. Li, C. Tang, E. Zhu, J. Yin, and W. Gao, "Efficient and effective incomplete multi-view clustering," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 4392–4399. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.33014392

[60] H. Gao, Y. Peng, and S. Jian, "Incomplete multi-view clustering," in *Intelligent Information Processing VIII: 9th IFIP TC 12 International Conference, IIP 2016, Melbourne, VIC, Australia, November 18-21, 2016, Proceedings 9*. Springer, 2016, pp. 245–255.

[61] H. Wang, L. Zong, B. Liu, Y. Yang, and W. Zhou, "Spectral perturbation meets incomplete multi-view data," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 3677–3683. [Online]. Available: https://doi.org/10.24963/ijcai.2019/510

[62] J. Zhang, L. Fei, S. Teng, Q. Zhu, I. Rida, and J. Wen, "Consensus learning with complete graph regularization for incomplete multi-view clustering," in *IEEE Symposium Series on Computational Intelligence, SSCI 2022, Singapore, December 4-7, 2022*, H. Ishibuchi, C. Kwoh, A. Tan, D. Srinivasan, C. Miao, A. Trivedi, and K. A. Crockett, Eds. IEEE, 2022, pp. 1485–1492. [Online]. Available: https://doi.org/10.1109/SSCI51031.2022.10022269

[63] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, 2020. [Online]. Available: https://doi.org/10.1109/TCYB.2018.2884715

[64] W. Zhuge, C. Hou, X. Liu, H. Tao, and D. Yi, "Simultaneous representation learning and clustering for incomplete multi-view data," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 4482–4488. [Online]. Available: https://doi.org/10.24963/ijcai.2019/623

[65] L. Zhao, Z. Chen, Y. Yang, Z. J. Wang, and V. C. M. Leung, "Incomplete multi-view clustering via deep semantic mapping," *Neurocomputing*, vol. 275, pp. 1053–1062, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2017.07.016

[66] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, "Partial multi-view clustering via consistent GAN," in *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*. IEEE Computer Society, 2018, pp. 1290–1295. [Online]. Available: https://doi.org/10.1109/ICDM.2018.00174

[67] C. Xu, Z. Guan, W. Zhao, H. Wu, Y. Niu, and B. Ling, "Adversarial incomplete multi-view clustering," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 3933–3939. [Online]. Available: https://doi.org/10.24963/ijcai.2019/546

[68] S. El Hajjar, F. Dornaika, F. Abdallah, and N. Barrena, "Consensus graph and spectral representation for one-step multi-view kernel based clustering," *Knowledge-Based Systems*, vol. 241, p. 108250, 2022.

[69] Y. Hu, Z. Song, B. Wang, J. Gao, Y. Sun, and B. Yin, "Akm^{3}c: Adaptive k-multiple-means for multi-view clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4214–4226, 2021. [Online]. Available: https://doi.org/10.1109/TCSVT.2020.3049005

[70] J. Li, Y. Wu, J. Zhao, and K. Lu, "Low-rank discriminant embedding for multiview learning," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3516–3529, 2017. [Online]. Available: https://doi.org/10.1109/TCYB.2016.2565898

[71] S. Shi, F. Nie, R. Wang, and X. Li, "Self-weighting multi-view spectral clustering based on nuclear norm," *Pattern Recognit.*, vol. 124, p. 108429, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.108429

[72] X. Si, Q. Yin, X. Zhao, and L. Yao, "Consistent and diverse multi-view subspace clustering with structure constraint," *Pattern Recognit.*, vol. 121, p. 108196, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.108196

[73] ——, "Consistent and diverse multi-view subspace clustering with structure constraint," *Pattern Recognit.*, vol. 121, p. 108196, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.108196

[74] W. K. Wong, N. Han, X. Fang, S. Zhan, and J. Wen, "Clustering structure-induced robust multi-view graph recovery," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3584–3597, 2020. [Online]. Available: https://doi.org/10.1109/TCSVT.2019.2945202

[75] G. Zhang, Y. Zhou, C. Wang, D. Huang, and X. He, "Joint representation learning for multi-view subspace clustering," *Expert Syst. Appl.*, vol. 166, p. 113913, 2021. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.113913

[76] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, "Partial multi-view clustering via consistent GAN," in *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*. IEEE Computer Society, 2018, pp. 1290–1295. [Online]. Available: https://doi.org/10.1109/ICDM.2018.00174

[77] W. Shao, L. He, and P. S. Yu, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with l_2, 1 regularization," in *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I*, ser. Lecture Notes in Computer Science, A. Appice, P. P. Rodrigues, V. S. Costa, C. Soares, J. Gama, and A. Jorge, Eds., vol. 9284. Springer, 2015, pp. 318–334. [Online]. Available: https://doi.org/10.1007/978-3-319-23528-8_20

[78] B. Quanz and J. Huan, "Conet: feature generation for multi-view semi-supervised learning with partially observed views," in *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012*, X. Chen, G. Lebanon, H. Wang, and M. J. Zaki, Eds. ACM, 2012, pp. 1273–1282. [Online]. Available: https://doi.org/10.1145/2396761.2398429

[79] H. Tao, C. Hou, D. Yi, J. Zhu, and D. Hu, "Joint embedding learning and low-rank approximation: A framework for incomplete multiview learning," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1690–1703, 2021. [Online]. Available: https://doi.org/10.1109/TCYB.2019.2953564

[80] J. Yin and S. Sun, "Incomplete multi-view clustering with cosine similarity," *Pattern Recognit.*, vol. 123, p. 108371, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.108371

[81] Q. Yin, S. Wu, and L. Wang, "Unified subspace learning for incomplete and unlabeled multi-view data," *Pattern Recognit.*, vol. 67, pp. 313–327, 2017. [Online]. Available: https://doi.org/10.1016/j.patcog.2017.01.035

[82] E. Eaton, M. desJardins, and S. Jacob, "Multi-view clustering with constraint propagation for learning with an incomplete mapping between views," in *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, J. X. Huang, N. Koudas, G. J. F. Jones, X. Wu, K. Collins-Thompson, and A. An, Eds. ACM, 2010, pp. 389–398. [Online]. Available: https://doi.org/10.1145/1871437.1871489

[83] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, and J. Ye, "Multi-source learning for joint analysis of incomplete multi-modality neuroimaging data," in *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Beijing, China, August 12-16, 2012*, Q. Yang, D. Agarwal, and J. Pei, Eds. ACM, 2012, pp. 1149–1157. [Online]. Available: https://doi.org/10.1145/2339530.2339710

[84] L. Cai, Z. Wang, H. Gao, D. Shen, and S. Ji, "Deep adversarial learning for multi-modality missing data completion," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, Y. Guo and F. Farooq, Eds. ACM, 2018, pp. 1158–1166. [Online]. Available: https://doi.org/10.1145/3219819.3219963

[85] M. Xie, Z. Ye, G. Pan, and X. Liu, "Incomplete multi-view subspace clustering with adaptive instance-sample mapping and deep feature fusion," *Appl. Intell.*, vol. 51, no. 8, pp. 5584–5597, 2021. [Online]. Available: https://doi.org/10.1007/s10489-020-02138-9

[86] C. Li, S. Wang, D. Yang, P. S. Yu, Y. Liang, and Z. Li, "Adversarial learning for multi-view network embedding on incomplete graphs," *Knowl. Based Syst.*, vol. 180, pp. 91–103, 2019. [Online]. Available: https://doi.org/10.1016/j.knosys.2019.05.021

[87] L. Tran, X. Liu, J. Zhou, and R. Jin, "Missing modalities imputation via cascaded residual autoencoder," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 2017, pp. 4971–4980. [Online]. Available: https://doi.org/10.1109/CVPR.2017.528

[88] J. Wen, K. Yan, Z. Zhang, Y. Xu, J. Wang, L. Fei, and B. Zhang, "Adaptive graph completion based incomplete multi-view clustering," *IEEE Trans. Multim.*, vol. 23, pp. 2493–2504, 2021. [Online]. Available: https://doi.org/10.1109/TMM.2020.3013408

[89] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2013. [Online]. Available: https://doi.org/10.1109/TPAMI.2012.88

[90] H. Xu, X. Zhang, W. Xia, Q. Gao, and X. Gao, "Low-rank tensor constrained co-regularized multi-view spectral clustering," *Neural Networks*, vol. 132, pp. 245–252, 2020. [Online]. Available: https://doi.org/10.1016/j.neunet.2020.08.019

[91] S. Gunasekar, B. E. Woodworth, S. Bhojanapalli, B. Neyshabur, and N. Srebro, "Implicit regularization in matrix factorization," in *2018 Information Theory and Applications Workshop, ITA 2018, San Diego, CA, USA, February 11-16, 2018*. IEEE, 2018, pp. 1–10. [Online]. Available: https://doi.org/10.1109/ITA.2018.8503198

[92] Y. Kim and S. Choi, "Weighted nonnegative matrix factorization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009, 19-24 April 2009, Taipei, Taiwan*. IEEE, 2009, pp. 1541–1544. [Online]. Available: https://doi.org/10.1109/ICASSP.2009.4959890

[93] L. Zong, X. Zhang, L. Zhao, H. Yu, and Q. Zhao, "Multi-view clustering via multi-manifold regularized non-negative matrix factorization," *Neural Networks*, vol. 88, pp. 74–89, 2017. [Online]. Available: https://doi.org/10.1016/j.neunet.2017.02.003

[94] A. Kumar, P. Rai, and H. D. III, "Co-regularized multi-view spectral clustering," in *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, Eds., 2011, pp. 1413–1421. [Online]. Available: https://proceedings.neurips.cc/paper/2011/hash/31839b036f63806cba3f47b93af8ccb5-Abstract.html

[95] T. H. Oh, Y. Matsushita, Y. Tai, and I. Kweon, "Fast randomized singular value thresholding for nuclear norm minimization," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*. IEEE Computer Society, 2015, pp. 4484–4493. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7299078

[96] W. Rudin, *Principles of mathematical analysis*, 1953.

[97] J. D. Rodríguez, A. P. Martínez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010. [Online]. Available: https://doi.org/10.1109/TPAMI.2009.187

[98] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 5393–5400. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.33015393

[99] Z. Li, C. Tang, X. Zheng, X. Liu, W. Zhang, and E. Zhu, "High-order correlation preserved incomplete multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 31, pp. 2067–2080, 2022. [Online]. Available: https://doi.org/10.1109/TIP.2022.3147046

[100] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009. [Online]. Available: https://doi.org/10.1109/TNN.2008.2005601

[101] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, 2005. [Online]. Available: https://doi.org/10.1109/TKDE.2005.198

[102] S. Romano, X. V. Nguyen, J. Bailey, and K. Verspoor, "Adjusting for chance clustering comparison measures," *J. Mach. Learn. Res.*, vol. 17, pp. 134:1–134:32, 2016. [Online]. Available: http://jmlr.org/papers/v17/15-627.html

[103] Y. Wang, J. Li, Y. Li, R. Wang, and X. Yang, "Confidence interval for f_1 measure of algorithm performance based on blocked $3\times$ 2 cross-validation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 651–659, 2015. [Online]. Available: https://doi.org/10.1109/TKDE.2014.2359667

[104] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, C. Sierra, Ed. ijcai.org, 2017, pp. 2564–2570. [Online]. Available: https://doi.org/10.24963/ijcai.2017/357

[105] ——, "Self-weighted multiview clustering with multiple graphs," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, C. Sierra, Ed. ijcai.org, 2017, pp. 2564–2570. [Online]. Available: https://doi.org/10.24963/ijcai.2017/357

[106] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multi-view clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1774–1782, 2018.

[107] J. Chang, L. Wang, G. Meng, S. Xiang, and C. Pan, "Deep adaptive image clustering," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 2017, pp. 5880–5888. [Online]. Available: https://doi.org/10.1109/ICCV.2017.626

[108] G. Liu, H. Ge, S. Su, and S. Wang, "Multi-view clustering via dual-norm and hsic," *Multimedia Tools and Applications*, pp. 1–20, 2022.

[109] Q. Yin, S. Wu, and L. Wang, "Incomplete multi-view clustering via subspace learning," in *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, J. Bailey, A. Moffat, C. C. Aggarwal, M. de Rijke, R. Kumar, V. Murdock, T. K. Sellis, and J. X. Yu, Eds. ACM, 2015, pp. 383–392. [Online]. Available: https://doi.org/10.1145/2806416.2806526

[110] J. Wen, Z. Wu, Z. Zhang, L. Fei, B. Zhang, and Y. Xu, "Structural deep incomplete multi-view clustering network," in *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, G. Demartini, G. Zuccon, J. S. Culpepper, Z. Huang, and H. Tong, Eds. ACM, 2021, pp. 3538–3542. [Online]. Available: https://doi.org/10.1145/3459637.3482192

[111] J. Guo and J. Ye, "Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 118–125. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.3301118

[112] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE Trans. Cybern.*, vol. 51, no. 1, pp. 101–114, 2021. [Online]. Available: https://doi.org/10.1109/TCYB.2020.2987164

[113] J. Wen, Z. Zhang, Z. Zhang, L. Zhu, L. Fei, B. Zhang, and Y. Xu, "Unified tensor framework for incomplete multi-view clustering and missing-view inferring," in *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021, pp. 10 273–10 281. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17231

[114] N. Liang, Z. Yang, Z. Li, and W. Han, "Incomplete multi-view clustering with incomplete graph-regularized orthogonal non-negative matrix factorization," *Appl. Intell.*, vol. 52, no. 13, pp. 14 607–14 623, 2022. [Online]. Available: https://doi.org/10.1007/s10489-022-03551-y

[115] J. Yin and S. Sun, "Incomplete multi-view clustering with cosine similarity," *Pattern Recognit.*, vol. 123, p. 108371, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.108371

[116] J. Xu, C. Li, Y. Ren, L. Peng, Y. Mo, X. Shi, and X. Zhu, "Deep incomplete multi-view clustering via mining cluster complementarity," in *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 2022, pp. 8761–8769. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/20856

[117] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, "Generative partial multi-view clustering with adaptive fusion and cycle consistency," *IEEE Trans. Image Process.*, vol. 30, pp. 1771–1783, 2021. [Online]. Available: https://doi.org/10.1109/TIP.2020.3048626

[118] X. Zheng, X. Liu, J. Chen, and E. Zhu, "Adaptive partial graph learning and fusion for incomplete multi-view clustering," *Int. J. Intell. Syst.*, vol. 37, no. 1, pp. 991–1009, 2022. [Online]. Available: https://doi.org/10.1002/int.22655

[119] W. Yang, Y. Shi, Y. Gao, L. Wang, and M. Yang, "Incomplete-data oriented multiview dimension reduction via sparse low-rank representation," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 12, pp. 6276–6291, 2018. [Online]. Available: https://doi.org/10.1109/TNNLS.2018.2828699

[120] X. Fang, Y. Hu, P. Zhou, and D. O. Wu, "Unbalanced incomplete multi-view clustering via the scheme of view evolution: Weak views are meat; strong views do eat," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 4, pp. 913–927, 2022. [Online]. Available: https://doi.org/10.1109/TETCI.2021.3077909

[121] Z. Deng, R. Liu, P. Xu, K. Choi, W. Zhang, X. Tian, T. Zhang, L. Liang, B. Qin, and S. Wang, "Multi-view clustering with the cooperation of visible and hidden views," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 803–815, 2022. [Online]. Available: https://doi.org/10.1109/TKDE.2020.2983366

[122] S. E. Hajjar, F. Dornaika, and F. Abdallah, "One-step multi-view spectral clustering with cluster label correlation graph," *Inf. Sci.*, vol. 592, pp. 97–111, 2022. [Online]. Available: https://doi.org/10.1016/j.ins.2022.01.017

[123] G. Niu, Y. Yang, and L. Sun, "One-step multi-view subspace clustering with incomplete views," *Neurocomputing*, vol. 438, pp. 290–301, 2021. [Online]. Available: https://doi.org/10.1016/j.neucom.2021.01.080

[124] X. Niu, L. Fu, W. Zhang, and Y. Li, "Seismic data interpolation based on simultaneously sparse and low-rank matrix recovery," *IEEE Trans. Geosci. Remote. Sens.*, vol. 60, pp. 1–13, 2022. [Online]. Available: https://doi.org/10.1109/TGRS.2021.3110600

[125] L. Wang, B. Wang, Z. Zhang, Q. Ye, L. Fu, G. Liu, and M. Wang, "Robust auto-weighted projective low-rank and sparse recovery for visual representation," *Neural Networks*, vol. 117, pp. 201–215, 2019. [Online]. Available: https://doi.org/10.1016/j.neunet.2019.05.007

[126] H. Du, X. Zhang, Q. Hu, and Y. Hou, "Sparse representation-based robust face recognition by graph regularized low-rank sparse representation recovery," *Neurocomputing*, vol. 164, pp. 220–229, 2015.

[127] Z. Kang, C. Peng, and Q. Cheng, "Robust pca via nonconvex rank approximation," in *2015 IEEE International Conference on Data Mining*. IEEE, 2015, pp. 211–220.

[128] E. Richard, P.-A. Savalle, and N. Vayatis, "Estimation of simultaneously sparse and low rank matrices," *arXiv preprint arXiv:1206.6474*, 2012.

[129] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 390–403, 2018.

[130] P. D. Tao and L. H. An, "Convex analysis approach to dc programming: theory, algorithms and applications," *Acta mathematica vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[131] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, 2013.

[132] M. F. Duarte and Y. H. Hu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826–838, 2004.

[133] Y. P. Qin, P. D. Qin, Y. Wang, and S. X. Lun, "A new optimal binary tree svm multi-class classification algorithm," *Applied Mechanics and Materials*, vol. 373, pp. 1085–1088, 2013.

[134] J.-W. Liu, Y.-F. Wang, R.-K. Lu, and X.-L. Luo, "Multi-view non-negative matrix factorization discriminant learning via cross entropy loss," in *2020 Chinese Control And Decision Conference (CCDC)*. IEEE, 2020, pp. 3964–3971.

[135] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery, "Learning to extract symbolic knowledge from the world wide web," *AAAI/IAAI*, vol. 3, no. 3.6, p. 2, 1998.

[136] M. Sharif, S. Mohsin, M. J. Jamal, and M. Raza, "Illumination normalization preprocessing for face recognition," in *2010 the 2nd conference on environmental science and information application technology*, vol. 2. IEEE, 2010, pp. 44–47.

[137] M. Yang, X. Liu, C. Mao, and B. Hu, "Graph convolutional networks with dependency parser towards multiview representation learning for sentiment analysis," in *IEEE International Conference on Data Mining Workshops, ICDM 2022 - Workshops, Orlando, FL, USA, November 28 - Dec. 1, 2022*, K. S. Candan, T. N. Dinh, M. T. Thai, and T. Washio, Eds. IEEE, 2022, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ICDMW58026.2022.00070

[138] Q. Zhang, L. Zhang, B. Du, W. Zheng, W. Bian, and D. Tao, "Mmfe: Multitask multiview feature embedding," in *2015 IEEE International Conference on Data Mining*. IEEE, 2015, pp. 1105–1110.

[139] Y. Feng, P. Zhou, D. O. Wu, and Y. Hu, "Accurate content push for content-centric social networks: A big data support online learning approach," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 2, no. 6, pp. 426–438, 2018. [Online]. Available: https://doi.org/10.1109/TETCI.2018.2804335

[140] Q. Zhang, Z. Sun, W. Hu, M. Chen, L. Guo, and Y. Qu, "Multi-view knowledge graph embedding for entity alignment," *arXiv preprint arXiv:1906.02390*, 2019.

[141] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, "Multi-modal curriculum learning for semi-supervised image classification," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3249–3260, 2016.

[142] G. Ji and G.-F. Lu, "One-step incomplete multiview clustering with low-rank tensor graph learning," *Information Sciences*, vol. 615, pp. 209–225, 2022.

[143] N. Zhang and S. Sun, "Incomplete multiview nonnegative representation learning with multiple graphs," *Pattern Recognit.*, vol. 123, p. 108412, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2021.108412

[144] S. Zhao, Z. Cui, L. Wu, Y. Xu, Y. Zuo, and L. Fei, "Salient and consensus representation learning based incomplete multiview clustering," *Appl. Intell.*, vol. 53, no. 3, pp. 2723–2737, 2023. [Online]. Available: https://doi.org/10.1007/s10489-022-03530-3

[145] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, and E. Zhu, "Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9776–9785.

[146] L. Li, Z. Wan, and H. He, "Incomplete multi-view clustering with joint partition and graph learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 589–602, 2023. [Online]. Available: https://doi.org/10.1109/TKDE.2021.3082470

[147] X. Liu, "Incomplete multiple kernel alignment maximization for clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[148] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, "Generative partial multi-view clustering with adaptive fusion and cycle consistency," *IEEE Transactions on Image Processing*, vol. 30, pp. 1771–1783, 2021.

[149] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng, "COMPLETER: incomplete multi-view clustering via contrastive prediction," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*. Computer Vision Foundation / IEEE, 2021, pp. 11 174–11 183. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Lin_COMPLETER_Incomplete_Multi-View_Clustering_via_Contrastive_Prediction_CVPR_2021_paper.html

[150] Y. Lin, Y. Gou, X. Liu, J. Bai, J. Lv, and X. Peng, "Dual contrastive prediction for incomplete multi-view representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4447–4461, 2023. [Online]. Available: https://doi.org/10.1109/TPAMI.2022.3197238

[151] X. Yu, H. Liu, Y. Lin, Y. Wu, and C. Zhang, "Auto-weighted sample-level fusion with anchors for incomplete multi-view clustering," *Pattern Recognit.*, vol. 130, p. 108772, 2022. [Online]. Available: https://doi.org/10.1016/j.patcog.2022.108772

[152] W. He, Z. Zhang, Y. Chen, and J. Wen, "Structured anchor-inferred graph learning for universal incomplete multi-view clustering," *World Wide Web (WWW)*, vol. 26, no. 1, pp. 375–399, 2023. [Online]. Available: https://doi.org/10.1007/s11280-022-01012-7

[153] Z. Wan, C. Zhang, P. Zhu, and Q. Hu, "Multi-view information-bottleneck representation learning," in *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021, pp. 10 085–10 092. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/17210

[154] Q. Wang, C. Boudreau, Q. Luo, P. Tan, and J. Zhou, "Deep multi-view information bottleneck," in *Proceedings of the 2019 SIAM International Conference on Data Mining, SDM 2019, Calgary, Alberta, Canada, May 2-4, 2019*, T. Y. Berger-Wolf and N. V. Chawla, Eds. SIAM, 2019, pp. 37–45. [Online]. Available: https://doi.org/10.1137/1.9781611975673.5

[155] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. [Online]. Available: https://openreview.net/forum?id=SJU4ayYgl

[156] J. A. Thomas, "Elements of information theory," 1991.

[157] S. G. Henderson and B. L. Nelson, Eds., *Simulation*, ser. Handbooks in Operations Research and Management Science. North-Holland, 2006, vol. 13. [Online]. Available: https://doi.org/10.1016/s0927-0507(06)x1300-2

[158] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2014. [Online]. Available: http://arxiv.org/abs/1312.6114

List of figures

List of tables

DOCTORAL SCHOOL

Title : Incomplete Multi-view Data Clustering with Hidden Data Mining and Fusion Techniques

Keywords : Hidden Data Mining, Non-negative matrix factorization, Collaborative Fusion, Low-Rank Tensor, Sparse Low-Rank Representation, Information Bottleneck, Anchor Graph GCN

Abstract :
Incomplete multi-view data clustering is a research direction that attracts attention in the fields of data mining and machine learning. In practical applications, we often face situations where only part of the modal data can be obtained or there are missing values. Data fusion is an important method for incomplete multi-view information mining. Solving for incomplete multi-view information mining in a targeted manner, achieving flexible collaboration between visible views and shared hidden views, and improving the robustness have become quite challenging. This thesis focuses on three aspects : hidden data mining, collaborative fusion, and enhancing the robustness of clustering. The main contributions are as follows :
1. Hidden data mining for incomplete multi-view data : existing algorithms cannot make full use of the observation of information within and between views, resulting in the loss of a large amount of valuable information, and so we propose a new incomplete multi-view clustering model IMC-NLT (Incomplete Multi-view Clustering Based on NMF and Low-Rank Tensor Fusion) based on non-negative matrix factorization and low-rank tensor fusion. IMC-NLT first uses a low-rank tensor to retain view features with a unified dimension. Using a consistency measure, IMC-NLT captures a consistent representation across multiple views. Finally, IMC-NLT incorporates multiple learning into a unified model such that hidden information can be extracted effectively from incomplete views. We conducted comprehensive experiments on five real-world datasets to validate the performance of IMC-NLT. The overall experimental results demonstrate that the proposed IMC-NLT performs better than several baseline methods, yielding stable and promising results.
2. Collaborative fusion for incomplete multi-view data : our approach to address this issue is Incomplete Multi-view Co-Clustering by Sparse Low-Rank Representation (CCIM-SLR). The algorithm is based on sparse low-rank representation and subspace representation, in which jointly-missing data is filled using data within a modality and related data from other modalities. To improve the stability of clustering results for multi-view data with different missing degrees, CCIM-SLR uses the Γ-norm model, which is an adjustable low-rank representation method. CCIM-SLR can alternate between learning the shared hidden view, visible view, and cluster partitions within a co-learning framework. An iterative algorithm with guaranteed convergence is used to optimize the proposed objective function. Compared with other baseline models, CCIM-SLR achieved the best performance in the comprehensive experiments on the five benchmark datasets, particularly on those with varying degrees of incompleteness.
3. Enhancing the clustering robustness for incomplete multi-view data : we offer a fusion of graph convolution and information bottlenecks (Incomplete Multi-view Representation Learning Through Anchor Graph-based GCN and Information Bottleneck – IMRL-AGI). First, we introduce the information bottleneck theory to filter out the noise data with irrelevant details and retain only the most relevant feature items. Next, we integrate the graph structure information based on anchor points into the local graph information of the state fused into the shared information representation and the information representation learning process of the local specific view, a process which can balance the robustness of the learned features and improve the robustness. Finally, the model integrates multiple representations with the help of information bottlenecks, reducing the impact of redundant information in the data. Extensive experiments are conducted on several real-world datasets, and the results demonstrate the superiority of IMRL-AGI. Specifically, IMRL-AGI shows significant improvements in clustering and classification accuracy, even in the presence of high view missing rates (e.g. 10.23% and 24.1% respectively on the ORL dataset). Furthermore, experiment results demonstrate the robustness of IMRL-AGI across different view missing rates (e.g. the NMI values ranging from 82.52% to 86.49% on the ORL dataset).

Titre : Clustering de données multivues incomplétes à l'aide de techniques de mining de données cachées et de fusion

Mots clés : Exploration de données cachées, Factorisation matricielle non négative, Fusion coopérative, Tenseurs de bas rang, Représentation á faible rang parcimonieuse, la théorie de goulot d'étranglement de l'information, réseaux convolutifs de graphes d'ancrage

Résumé : Le regroupement de données multivues incomplètes est un axe de recherche majeur dans le domaines de l'exploration de données et de l'apprentissage automatique. Dans les applications pratiques, nous sommes souvent confrontés à des situations où seule une partie des données modales peut être obtenue ou lorsqu'il y a des valeurs manquantes. La fusion de données est une méthode clef pour l'exploration d'informations multivues incomplètes. Résoudre le problème de l'extraction d'informations multivues incomplètes de manière ciblée, parvenir à une collaboration flexible entre les vues visibles et les vues cachées partagées, et améliorer la robustesse sont des défis. Cette thèse se concentre sur trois aspects : l'exploration de données cachées, la fusion collaborative et l'amélioration de la robustesse du regroupement. Les principales contributions sont les suivantes :

1) Exploration de données cachées pour les données multi-vues incomplètes : les algorithmes existants ne peuvent pas utiliser pleinement l'observation des informations dans et entre les vues, ce qui entraîne la perte d'une grande quantité d'informations. Nous proposons donc un nouveau modèle de regroupement multi-vues incomplet IMC-NLT (Incomplete Multi-view Clustering Based on NMF and Low-Rank Tensor Fusion) basé sur la factorisation de matrices non négatives et la fusion de tenseurs de faible rang. IMC-NLT utilise d'abord un tenseur de faible rang pour conserver les caractéristiques des vues avec une dimension unifiée. En utilisant une mesure de cohérence, IMC-NLT capture une représentation cohérente à travers plusieurs vues. Enfin, IMC-NLT intègre plusieurs apprentissages dans un modèle unifié afin que les informations cachées puissent être extraites efficacement à partir de vues incomplètes. Des expériences sur cinq jeux de données ont validé les performances d'IMC-NLT.

2) Fusion collaborative pour les données multivues incomplètes : notre approche pour résoudre ce problème est le regroupement multivues incomplet par représentation à faible rang. L'algorithme est basé sur une représentation éparse de faible rang et une représentation de sous-espace, dans laquelle les données manquantes sont complétées en utilisant les données d'une modalité et les données connexes d'autres modalités. Pour améliorer la stabilité des résultats de clustering pour des données multi-vues avec différents degrés de manquants, CCIM-SLR utilise le modèle Γ-norm, qui est une méthode de représentation à faible rang ajustable. CCIM-SLR peut alterner entre l'apprentissage de la vue cachée partagée, la vue visible et les partitions de clusters au sein d'un cadre d'apprentissage collaboratif. Un algorithme itératif avec convergence garantie est utilisé pour optimiser la fonction objective proposée.

3) Amélioration de la robustesse du regroupement pour les données multivues incomplètes : nous proposons une fusion de la convolution graphique et des goulots d'étranglement de l'information (apprentissage de la représentation multivues incomplète via le goulot d'étranglement de l'information). Nous introduisons la théorie du goulot d'étranglement de l'information afin de filtrer les données parasites contenant des détails non pertinents et de ne conserver que les éléments les plus pertinents. Nous intégrons les informations sur la structure du graphe basées sur les points d'ancrage dans les informations sur le graphe local. Le modèle intègre des représentations multiples à l'aide de goulets d'étranglement de l'information, réduisant ainsi l'impact des informations redondantes dans les données. Des expériences approfondies sont menées sur plusieurs ensembles de données du monde réel, et les résultats démontrent la supériorité de IMRL-AGI. Plus précisément, IMRL-AGI montre des améliorations significatives dans la précision du clustering et de la classification, même en présence de taux élevés de données manquantes par vue (par exemple, 10,23 % et 24,1 % respectivement sur l'ensemble de données ORL). De plus, les résultats des expériences démontrent la robustesse de l'IMRL-AGI à travers différents taux de vues manquantes (par exemple, les valeurs de NMI variant de 82,52 % à 86,49 % sur le jeu de données ORL).

Institut Polytechnique de Paris
91120 Palaiseau, France