

多视图聚类研究综述

丁世飞^{1),2)} 韦修喜^{1),3),4)} 梅炎斌³⁾ 黄华娟³⁾

¹⁾(中国矿业大学计算机科学与技术学院/人工智能学院 江苏 徐州 221116)

²⁾(矿山数字化教育部工程研究中心(中国矿业大学) 江苏 徐州 221116)

³⁾(广西民族大学人工智能学院 南宁 530006)

⁴⁾(广西混杂计算与集成电路设计分析重点实验室(广西民族大学) 南宁 530006)

摘 要 随着互联网、社交网络和医学图像等多元化数据源的日益增多,传统的单视图聚类方法难以处理来自各种视图的海量数据。因此,关于多视图数据上的聚类问题变得越来越重要。本文以多视图聚类(Multi-view clustering, MVC)为对象,对多视图聚类常用的技术框架和方法进行综述。首先对多视图聚类技术框架进行分析和归类,并根据其技术特点将其分为三大类:基于启发式的 MVC、基于判别式的 MVC 和基于生成式的 MVC;其次,分别重点介绍这三种技术框架的常见方法,包括非负矩阵分解、子空间聚类、图学习、张量学习、深度嵌入聚类、对比学习、变分自动编码器和生成对抗网络;最后比较和分析现有的多视图聚类在不同数据集上的性能,总结该领域所面临的挑战,并对发展前景进行展望。

关键词 多视图聚类;非负矩阵分解;图学习;对比学习;变分自编码器

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2025.02523

A Review of Multi-View Clustering Research

DING Shi-Fei^{1,2)} WEI Xiu-Xi^{1,3,4)} MEI Yan-Bin³⁾ HUANG Hua-Juan³⁾

¹⁾(School of Computer Science & Technology/School of Artificial Intelligence, China University of Mining and Technology, Xuzhou, Jiangsu 221116)

²⁾(Mine Digitization Engineering Research Center of Ministry of Education (China University of Mining and Technology), Xuzhou, Jiangsu 221116)

³⁾(College of Artificial Intelligence, Guangxi Minzu University, Nanning 530006)

⁴⁾(Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis (Guangxi Minzu University), Nanning 530006)

Abstract Clustering is an essential technique in unsupervised learning. Its primary goal is to group data points based on their similarities and differences. With the rapid development of data mining and information acquisition, multi-view data has become increasingly common. It allows the same sample to be described from multiple perspectives. Multi-view clustering (MVC) integrates complementary and consensus information from different views. This capability enables more comprehensive and accurate clustering. As a result, MVC has attracted significant attention in recent years. Despite notable progress, existing research lacks a systematic review of MVC techniques. This limitation hinders a deeper understanding of MVC. This paper provides a detailed analysis of existing MVC methods. The study consists of five main parts. First, we introduced the fundamental concepts of MVC. We summarized four common representation methods:

收稿日期:2025-01-14;在线发布日期:2025-07-11。本课题得到国家自然科学基金面上项目(Nos. 62276265, 62206297)、国家自然科学基金地区基金(No. 62266007)、广西自然科学基金(No. 2021GXNSFAA220068)资助。丁世飞,博士,教授,博士生导师,中国计算机学会(CCF)杰出会员,主要研究领域为人工智能、模式识别、机器学习、数据挖掘。E-mail: dingsf@cumt.edu.cn。韦修喜(通信作者),博士研究生,副教授,中国计算机学会(CCF)会员,主要研究领域为机器学习、聚类分析。E-mail: weixiuxi@163.com。梅炎斌,硕士研究生,主要研究领域为机器学习、深度聚类。黄华娟(通信作者),博士,教授,硕士生导师,中国计算机学会(CCF)会员,主要研究领域为机器学习、数据挖掘。E-mail: hhj@gxmzu.edu.cn。

subspace representation, latent representation, graph representation, and tensor representation. We also categorized three fusion strategies: direct fusion, early fusion, and late fusion. Second, we proposed a new MVC classification framework. We further examined various MVC approaches. We classified MVC methods into heuristic MVC and deep learning-based MVC based on whether they use neural networks. For deep learning-based MVC, we further divided the methods into discriminative MVC and generative MVC, depending on whether they incorporate generative models. For heuristic MVC, we grouped existing methods by data representation strategies. These include approaches based on non-negative matrix factorization (NMF), subspace learning, graph learning, and tensor learning. For discriminative MVC, we categorized methods based on different discrimination strategies. These include deep embedding-based methods and contrastive learning-based methods. Deep embedding-based methods include approaches based on deep NMF (Deep-NMF) and autoencoders (AE). We also categorized generative MVC into two types: variational autoencoder (VAE)-based MVC and generative adversarial network (GAN)-based MVC. We conducted a detailed analysis of heuristic MVC and discriminative MVC. We also introduced the two types of generative MVC and compared classical MVC methods. We examined their advantages and disadvantages. To provide insights into performance trends, we summarized experimental results from 22 MVC methods and 7 incomplete multi-view clustering (IMVC) methods. These experiments were conducted on 9 public datasets and 3 incomplete datasets. We also analyzed the time complexity of these algorithms to help readers understand their computational efficiency. To reflect the current state of MVC research, we investigated its major applications. These include 3D reconstruction, bioinformatics, and text classification. We focused on graph-based MVC methods and deep embedding-based MVC methods in these applications. Finally, we summarized our findings and discussed potential research directions. MVC is evolving rapidly as multi-source data grows. Many previously unsolved problems are now being addressed. However, several challenges remain. These include handling incremental and uncertain data, improving computational efficiency, and enhancing model interpretability. MVC leverages complementarity and consistency to improve clustering performance. It has the potential to bring significant advancements. This paper systematically reviewed MVC techniques, performance trends, applications, and future research directions. We hope it helps new researchers enter the field and assists experts in identifying new research directions. Ultimately, we aim to promote further progress in MVC research.

Keywords multi-view clustering; non negative matrix factorization; graph learning; contrastive learning; variational autoencoder

1 引 言

聚类是无监督学习的重要技术,其核心目标是基于数据点之间的相似性和不相似性对数据进行分组^[1],以确保同一聚类内的数据点之间的相似性高于不同聚类间的数据点之间的相似性。聚类技术适用于多种类型的数据,包括社交网络信息、文本数据和生物医学数据等。常见的聚类技术,如 K-均值聚类、谱聚类和层次聚类,都是单视图聚类^[2]的基础。

然而,随着互联网、社交网络和医学图像等数据源的迅速发展,传统的单视图聚类方法已难以应对大规模数据的复杂性。因此,多视图数据的研究逐渐成为热点。多视图数据指的是从多个角度或模式收集的信息。例如,在图像和视频分析中,颜色和纹理特征可以视为不同的视图^[3];同样,网页内容、图片和超链接等信息也可作为多视角数据源进行分类分析。

近年来,多视图聚类(Multi-View Clustering, MVC)因其在实际场景中的优越表现而备受关注。

MVC 已广泛应用于社交网络中的社区检测、计算机视觉中的图像标注和识别,以及推荐系统中的跨域用户建模等^[4]。图 1 展示了多视图聚类的基本框

架:首先对多视图数据进行特征提取,其次进行特征学习与融合,最后对融合后的特征进行聚类,从而得出最终结果。

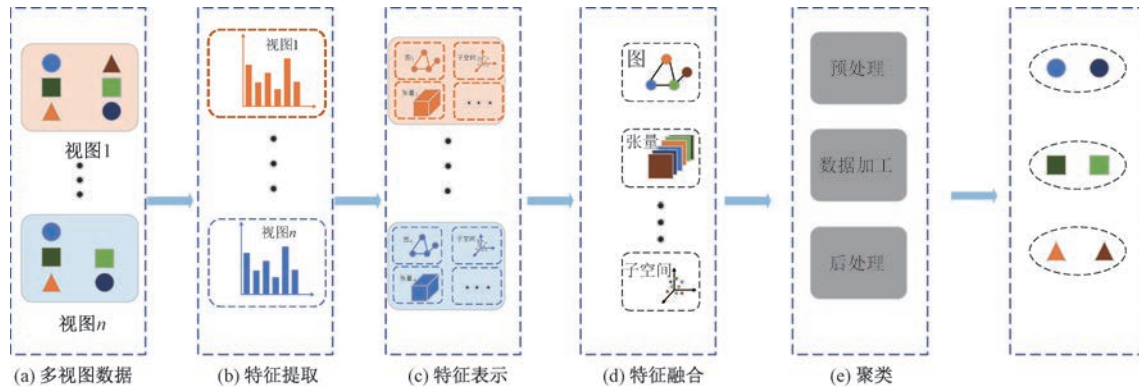


图 1 多视图聚类(MVC)的一般框架

近年来,随着 MVC 技术的深入研究,学者们提出了许多高效的算法。早期的研究通常通过将不同视图合并为单一表示,再应用传统的单视图聚类方法,如 K-均值聚类、光谱聚类和层次聚类等。但是这些方法普遍面临两个关键理论缺位:其一未能充分遵循多模态数据的互补性原则,其二缺乏对视图间一致性约束的系统建模。具体而言,互补性原则强调异构视图间存在信息互惠关系,即各视图承载着其他模态未捕获的判别性特征(如猫的侧视与正视图分别提供解剖结构的不同观测维度)。有效挖掘这种互补特性能够突破单视图的信息瓶颈,构建更具完备性的特征表达体系。一致性原则则关注跨模态的语义共识,不同视图在潜在语义层面应保持内在一致性(如不同视角的犬类图像均包含可辨识的轮廓特征与生物体征)。该原则通过建立跨视图语义对齐机制,可强化模型对样本本质属性的认知能力。

值得注意的是,尽管多源视图在原始特征空间呈现显著异构性(如成像模态差异导致的特征分布偏移),其潜在语义表征往往具有隐式的低维流形结构。这引出了多视图学习的核心科学问题:如何构建统一的嵌入空间,在保留视图特异性的同时提取跨模态共享的深度表征,并基于此实现鲁棒的聚类决策。当前该领域的研究仍面临特征解耦、语义对齐、噪声抑制等多重技术挑战^[5]。

目前,MVC 的研究逐渐聚焦于以下关键问题:

(1) 不确定数据

传统的 MVC 方法通常采用硬聚类方式,这种方法难以捕捉样本与聚类之间的不确定性^[6]。而且,现实世界的数据集往往包含噪声,这进一步增加

了数据的不确定性^[7]。因此,设计能够有效处理不确定数据的方法成为研究热点。

(2) 不完整数据

由于设备故障或数据传输问题,数据缺失现象时有发生,这类数据被称为不完整数据^[8]。如何处理缺失数据是不完整多视图聚类(IMVC)研究的核心问题。例如,Liu 等人提出使用样本数据的线性组合来弥补缺失数据^[9];另一方面,Hao 利用低秩约束方法对缺失数据进行重建,从而提高数据完整性^[10]。

(3) 增量多视图

当前的多视图学习方法通常难以适应动态变化的数据环境,例如新视图的引入。为应对这一问题,Zhou 提出了一种多视图图集聚类方法,通过从有限视图中学习初始模型,并在新视图添加后更新模型^[11]。Yin 等人和 Huang 等人分别提出了对特征矩阵进行增量更新的方法,以适应动态数据的需求^[12-13]。

(4) 无映射多视图

无映射数据是指同一样本在不同视图中的位置可能不一致的情况。传统 MVC 方法通常假设样本在不同视图中的映射是一致的,因此对无映射数据的处理成为一大挑战。图 2 (a)展示了映射数据的情况,而图 2 (b)展示了无映射数据的情形。针对这一问题,Yang 等人提出了一种类别级对齐策略,通过重新对齐跨视图的样本来解决映射不一致的问题^[14]。同时,Cai 等人采用 Calinski-Harabasz 指数确定最佳视图,并利用配准矩阵将其他视图中的样本与最佳视图进行配准^[15]。

解决这些问题是提升 MVC 方法性能的关键,

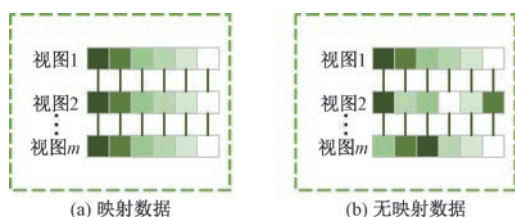


图2 两种不同的数据展示方式

同时也是推动多视图聚类技术在实际应用中更广泛普及的重要方向。

近年来,多视图聚类(MVC)的研究受到广泛关注,多篇综述文章从不同角度对其进行了总结与分类^[5,16-17]。例如,Chao等人基于机器学习方法将MVC分为生成类和判别类^[17];Yang等人从机制和原理出发,将其分为协同训练聚类、多核学习、多视图图聚类、多视图子空间聚类和多任务多视图聚类^[18];Fang等人则从技术机制角度出发,将MVC分为启发式方法和基于神经网络的聚类方法^[19]。Zhou等人根据数据层次性,将MVC分为完全(MVC)、不完全(IMVC)、不确定和动态多视图聚类^[5];Tang等人进一步细化不完全多视图聚类,划分为生成类和判别类方法^[16]。

尽管这些综述为MVC的发展提供了宝贵的总结,但仍存在一些不足之处。例如,Zhou等人提出的多视图数据分类方法缺乏对未映射数据的准确描述^[5];Tang等人将对比学习简单地归为生成方法,这是不合理的^[16];Chao等人则忽视了神经网络在多视图聚类中的重要作用^[17]。Yang等人和Fang等人尽管对多视图数据的复杂性进行了总结,但未充分考虑不确定数据和增量多视图对聚类性能的潜在影响^[18-19]。

根据现有多视图聚类方法是否依赖传统神经网络模型以及是否基于数据生成假设,多视图聚类方法可以划分为基于启发式的多视图聚类方法和基于深度学习的多视图聚类方法。其中,基于启发式的方法通常通过矩阵分解、图学习等非深度模型设计启发性目标函数来实现多视图数据的一致性与互补性,不依赖于深度神经网络或数据生成过程。进一步地,针对基于深度学习的多视图聚类方法,根据其是否引入生成模型的假设,又可细分为基于判别式的多视图聚类方法和基于生成式的多视图聚类方法。需要特别指出的是,尽管判别式与生成式方法均通过数据的潜在空间表示进行聚类,但二者在建模思想上存在本质区别:判别式方法通常通过自编码器等深度网络直接从观测数据中学习潜在表示,

用于提升聚类性能,而生成式方法则假设存在一个共享的先验潜变量,所有视图的数据均由该潜在变量生成,从而通过建模数据生成过程来发现潜在的聚类结构。为了更全面地概述MVC的研究现状,本文在分析多视图聚类的技术机制的基础上,将现有MVC方法归纳为三大类:启发式MVC、判别式MVC和生成式MVC。

(1) 启发式 MVC 方法

启发式MVC方法是指基于问题特性和工程经验,通过设计一致性或互补性目标函数,利用优化策略获得聚类结果的方法。这类方法通常不假设数据的生成过程,也不建模数据分布,而是依托于如矩阵分解、图学习、低秩约束等机制,解决多视图数据之间的信息融合问题。本文重点分析了四种方法,包括非负矩阵因式分解(NMF)、子空间学习、基于图的学习以及张量分解方法。

(2) 判别式 MVC 方法

判别式多视图聚类方法侧重从数据出发,直接学习有利于区分不同类别的潜在表示,通常结合一致性约束、对比学习等提升嵌入质量,从而实现高效聚类。本文对判别式方法进行了分类,主要涵盖多视图深度嵌入聚类 and 对比学习两种技术。

(3) 生成式 MVC 方法

生成式多视图聚类方法假设多视图数据由共享的潜在变量生成,旨在刻画多视图数据的联合分布,通过学习生成机制获得潜在表示,并在该空间完成聚类。本文讨论了基于生成式对抗网络(GAN)和变分自编码器(VAE)框架的多视图聚类方法。

图3展示了本文对现有MVC方法的分类与总结。这种新的分类方法试图弥补现有综述的不足,强调MVC方法在数据复杂性、多视图动态性以及技术机制上的关键进展。



图3 现有 MVC 方法概述

最后,为了比较和分析现有的多视图聚类在不同数据集上的性能,对一些主要方法进行了实验对

比分析,总结各类算法的优势与不足,提出该领域所面临的挑战,并对发展前景进行展望。

2 基本理论

2.1 多视图聚类

假设一个由 m 个视图组成的数据集,记为 $X = \{X^1, X^2, \dots, X^m\}$, $X^i = \{x_1^i, x_2^i, \dots, x_n^i\} \in \mathbb{R}^{d \times n}$, X^i 表示第 i 个视图数据, n 表示样本数, x_i^v 表示第 v 个视图上第 i 个样本数据的特征向量,其特征维度为 d 。设 $C = \{C_1, C_2, \dots, C_k\}$ 为一个包含 k 个聚类的聚类集,多视图聚类(Multi-View Clustering, MVC)的目标是通过无监督学习整合多个视图的特征,将 n 个样本归入 k 个聚类中,同时最大化簇内的相似性,并最小化簇间的相似性。

2.2 多视图聚类表示

MVC 方法的实现通常因多视图数据的表示类型不同而有所不同。例如,Zhang 提出了一种基于潜在子空间表示的多视图数据聚类方法^[20],而 Kang 提出了基于图表示的聚类方法^[21]。总体而言,多视图聚类算法主要集中在以下几种数据表示方式:潜在表示、子空间表示、图表示和张量表示。

在后续章节中,我们将结合 MVC 的具体技术机制,系统介绍这些数据表示方式的基本概念及特点。多视图聚类中常用符号及其说明如表 1 所示。

表 1 常用的符号及描述

| 符号 | 描述 |
|--------------|-------------------------------------|
| n | 数据集中数据全部数量 |
| m | 视图的数量 |
| d | 表示数据样本的维度大小 |
| X | 全部视图的所有数据的特征向量集合 |
| X^i | 表示第 i 个视图中所有数据的特征向量集合 |
| x_i^v | 表示第 v 个视图第 i 个数据的特征向量 |
| Z_i | 第 i 个样本的低维表示 |
| $\mathbf{1}$ | 元素全为 1 的向量 |
| k | 聚类簇的个数 |
| c_i | 表示第 i 个簇的集合 |
| C | 包含 k 个聚类簇的聚类集合 |
| H | 表示共享的潜在表示 |
| Z | 表示共享潜在表示的自表达矩阵 |
| L^i | 表示第 i 个视图的拉普拉斯矩阵 |
| s_{ij}^m | 表示第 m 个视图中第 i 个样本与第 j 个样本的相似度 |
| S^i | 表示第 i 个视图的相似性矩阵 |
| S | 表示每个视图相似矩阵生成的共识矩阵 |
| $Reg(\cdot)$ | 表示正则项 |
| I | 表示单位矩阵 |
| χ | 表示多个矩阵堆叠的张量 |
| g | 表示一种核心张量 |

(续表)

| 符号 | 描述 |
|------------------|------------------|
| $L_r(\cdot)$ | 与数据重构相关的损失函数 |
| $L_h(\cdot)$ | 与潜在表示相关的损失函数 |
| L_c | 表示一种对比损失 |
| $Tr(\cdot)$ | 表示矩阵的迹 |
| G^i | 第 i 个视图对应的锚点图 |
| G | 表示图融合之后的共识锚点图 |
| M | 表示马尔可夫概率转移矩阵 |
| $\ \cdot\ _1$ | 表示 l_1 范数 |
| $\ \cdot\ _2$ | 表示 l_2 范数 |
| $\ \cdot\ _F$ | 表示 Frobenious 范数 |
| $\ \cdot\ _*$ | 表示张量核范数 |
| τ_1, τ_2 | 是一对正的平衡参数 |

下面,介绍几种常用的多视图数据表示方法。

多视图子空间表示:子空间表示指通过确定一个低维子空间,在该子空间中对原始数据进行有效表示。作为 MVC 中广泛使用的一种技术^[21],多视图子空间聚类的目标是为每个视图学习一个或多个低维子空间,从而更好地捕捉数据的内在结构。其数学表示形式如下:

$$\min_{Z^i} \sum_{i=1}^m \|X^i - X^i Z^i\|_F^2 + \partial Reg(Z^i) \quad (1)$$

$$\text{s. t. } Z^i \geq 0, Z^i \mathbf{1} = \mathbf{1}$$

其中, Z^i 表示第 i 个视图的自表达矩阵,矩阵大小为 $n \times n$, ∂ 是平衡参数, $\mathbf{1}$ 是所有元素都为 1 的向量, $Reg(\cdot)$ 是相似性矩阵的正则化函数。

多视图潜在表示:该方法通过提取各视图的共同特征和独特特征,将来自多个视图的数据嵌入到一个统一的潜在空间中,从而增强聚类性能^[19]。多视图潜在表示模型的数学形式可描述如下:

$$\min_{P, H, Z} L_h(X, PH) + \tau_1 L_r(H, HZ) + \tau_2 Reg(Z)$$

$$\text{s. t. } X = PH, H = HZ, P^T P = I \quad (2)$$

其中, $P = \{P^1, P^2, \dots, P^m\}$ 是特定视图的重构模型集合,它将多个视图上的数据表示与共享的潜表征一起重构, $H = \{h_i\}_{i=1}^n$ 表示共享的潜在表示, Z 是重构的潜表征的自表达矩阵, τ_1, τ_2 是 2 个平衡参数, $Reg(Z)$ 是自表达矩阵 Z 的正则化项。

图表示法:在图表示法中,每个视图被建模为一个亲和图,其中节点表示对象,边表示对象之间的关系或相似性。亲和图可表示为 $G(V, E)$, 其中 V 表示数据点(节点), E 表示节点之间的边(关系)。数据点的特征可以用 x_i 表示。图 4 展示了亲和图的构造过程。

通常情况下,如果顶点 v_i 位于顶点 v_j 的一阶邻

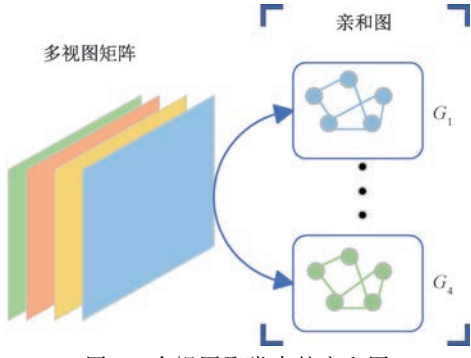


图4 多视图聚类中的亲和图

域内,则认为顶点 v_i 与 v_j 之间存在连接(记为 e_{ij})。然而,这种全连接方式的计算复杂度较高^[19],因此当前的研究更倾向于使用 k-近邻(k-nearest neighbors, k-NN)方法来构建亲和图,从而降低计算成本。

基于图表示的多视图聚类可以通过以下公式表示^[22]:

$$\begin{aligned} \min \sum_{\{s^v\}} \sum_{v=1}^m \sum_{i,j=1}^n \|x_i^v - x_j^v\|_2^2 s_{ij}^v + \partial \sum_{v=1}^m \sum_{i=1}^n \|s_i^v\|_1 \\ \text{s. t. } \forall v, s_{ii}^v = 0, s_{ij}^v \geq 0 \end{aligned} \quad (3)$$

其中, $\|x_i^v - x_j^v\|_2^2$ 表示边 e_{ij} 的权重, s_{ij}^v 表示第 v 个视图中数据点 i 和数据点 j 之间的相似度, \mathbf{S}^v 表示第 v 个视图的相似度矩阵。

张量表示法:张量是一种多维数组,通过堆叠多个矩阵可以构成三阶张量^[23]。图5展示了通过堆叠多个多视图矩阵构建三维张量的过程。在多视图聚类(MVC)中,数据常以堆叠不同视图的矩阵形成的三阶张量表示。例如,一个多视图数据集 X 可以划分为三个不同的视图矩阵 $\mathbf{O} = \{O_1, O_2, \dots, O_r\}$, $\mathbf{H} = \{H_1, H_2, \dots, H_r\}$, $\mathbf{W} = \{W_1, W_2, \dots, W_r\}$, 它们可以用一个张量 $\chi(\mathbf{O}, \mathbf{H}, \mathbf{W})$ 来统一表示^[19]。

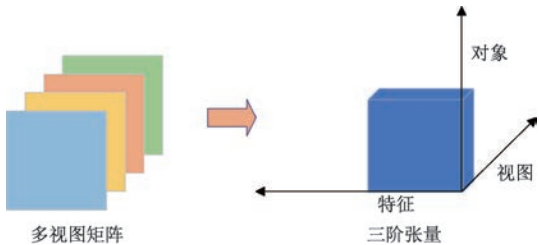


图5 多视图聚类中的三阶张量表示法

假设处理的多视图数据集由来自不同传感器的多维时间序列数据组成,例如音频信号、图像帧和文本描述,张量分解方法能够有效整合这些异构数据类型,从而提升聚类性能。其中,CP分解^[24]是一种

广泛使用的张量分解技术,其数学表达式如下:

$$\chi \approx \sum_{r=1}^m o_r \circ h_r \circ w_r \quad (4)$$

其中, $o_r \in \mathbb{R}^n$ 表示对象, $h_r \in \mathbb{R}^d$ 表示视图的特征, w_r 表示某一个特定视图。还有一种名为塔克分解的高阶奇异值分解(HSVD),它允许将张量 $\chi \in \mathbb{R}^{n \times d \times m}$ 分解为被一组更低维的张量包围的核心张量 $Q \in \mathbb{R}^{P \times Q \times R}$, 它可表示为

$$\chi = Q \times_1 \mathbf{O} \times_2 \mathbf{H} \times_3 \mathbf{W} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} o_p \circ h_q \circ w_r \quad (5)$$

其中, $g_{pqr} \in Q$ 是 χ 的核心张量, \times_n 表示 n 模式的乘积。由于高阶张量包含较大的数据量,直接计算难度较大,通过张量分解的方法有效降低计算的复杂度,同时还提高了模型的解释能力。

2.3 多视图融合策略

根据不同的多视图数据表示,整合不同视图间的信息策略主要有三种:直接融合,早期融合和后期融合,也可以分别称为数据级融合,特征级融合以及决策级融合^[5]。为了更直观地区分这三种不同的融合策略,我们用图示的方法对它们进行展示。直接融合的方法通常涉及优化特定的损失函数,将多视图数据直接融合到聚类过程中,如图6所示。早期融合将多视图数据的多个特征表示融合成跨多个视图的单一表示或共识亲和图,然后采用单视图聚类分析算法,比如,K-均值用来区分数据样本,但是大部分方法都是通过学习不同视图下的表示,然后构建共识亲和图,如图7所示。后期融合的方法首先对每个视图单独进行聚类,然后融合所有视图的聚类结果,获得具有共识性的聚类结果,如图8所示。后期融合的优势在于它减少了其他信息通道对每个单独分区的干扰,从而减少了随机噪声的干扰。

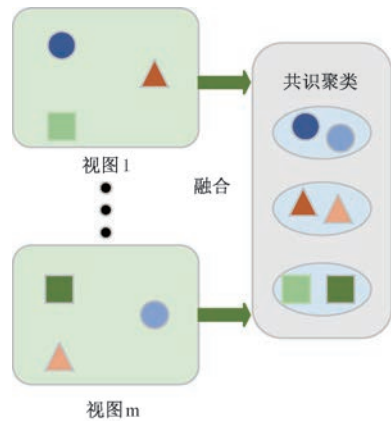


图6 多视图聚类中直接融合策略

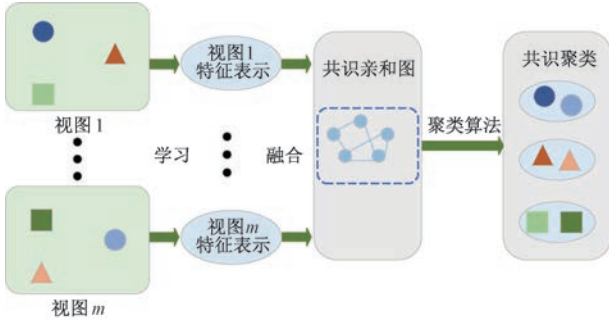


图7 多视图聚类中早期融合策略

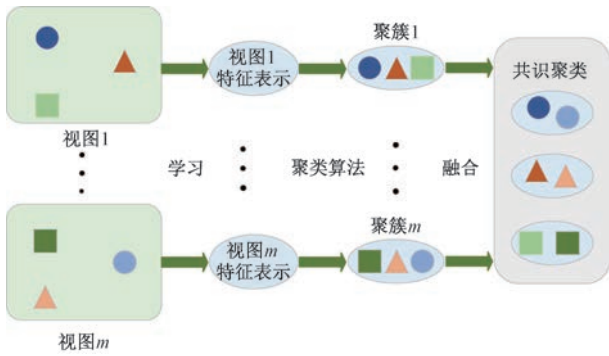


图8 多视图聚类中后期融合策略

3 MVC 最新研究进展

3.1 基于启发式的多视图聚类

在本小节中,我们将重点介绍四种典型的基于启发式的多视图聚类(MVC)方法:基于非负矩阵因式分解(NMF)的多视图聚类方法、子空间学习法、图学习法和张量学习方法。在每个小节中,我们深入不同方法的技术机制并且总结几种具有代表性的先进方法,以及讨论每种方法的优缺点。

3.1.1 非负矩阵因式分解

非负矩阵因式分解(NMF)是一种用于降低非负数据维度的技术。给定原始 d 维数据矩阵 $\mathbf{X} = [\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^n] \in \mathbb{R}^{d \times n}$, NMF 的目标是找到两个非负矩阵 $\mathbf{U} \in \mathbb{R}^{d \times k}$ 和 $\mathbf{V} \in \mathbb{R}^{n \times k}$, 使得它们的乘积能够最大程度地逼近原始数据^[25], NMF 的数学公式可以表示为: $\mathbf{X} \approx \mathbf{UV}^T$ 。NMF 通过优化以下目标函数来学习矩阵:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \\ \text{s. t. } \mathbf{U} \geq 0, \mathbf{V} \geq 0 \end{aligned} \quad (6)$$

但是,大部分基于 NMF 的聚类,基矩阵 \mathbf{U} 和指示矩阵 \mathbf{V} 是非负的,在真实世界中的数据可能是有符号的,因此,传统的 NMF 聚类算法不适用于具有混合符号的数据 \mathbf{X} , 为了扩大 NMF 适用范围,Ding 等

人提出一种 Semi-NMF(SNMF)的框架,该框架将 K-均值聚类目标视为矩阵近似函数,它可以用以下式子表述^[26]:

$$J_{K\text{-means}} = \sum_{i=1}^n \sum_{j=1}^k c_{ik} \|x_i - u_k\|^2 = \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \quad (7)$$

其中, $c_{ik}=1$ 表示样本 i 属于第 k 个聚簇,否则 $c_{ik}=0$; u_k 是第 k 个聚簇中心的表示。如果放宽优化,允许 c_{ik} 的取值为 $(0,1)$ 或者为 $(0,\infty)$, 那么这就是 SNMF 矩阵分解。它允许基矩阵 \mathbf{U} 是有符号的,但指示矩阵 \mathbf{V} 是非负的。

NMF 因其紧凑的学习能力和可解释性而广泛应用于多视图聚类,它用数学表达式可以描述为

$$\sum_{i=1}^m \|\mathbf{X}^i - (\mathbf{U}^i)(\mathbf{V}^i)^T\|_F^2 \quad (8)$$

基于 NMF 的 MVC 方法它本质上是一种全局分解方法,旨在通过矩阵分解最小化整体结构的误差,所以关于局部数据点之间的关系并没有保留。为了解决这个问题,Cai 等人将图形正则化约束条件引入到 NMF 中,提出了一种图形正则化 NMF (GNMF)方法^[27]。图形正则化约束可以用以下式子表示:

$$\begin{aligned} R &= \frac{1}{2} \sum_{j,l} \|\mathbf{Z}_j - \mathbf{Z}_l\| \mathbf{W}_{jl} \\ &= \sum_{j=1}^n \mathbf{Z}_j^T \mathbf{Z}_j \mathbf{D}_{jj} - \sum_{j,l=1}^n \mathbf{Z}_j^T \mathbf{Z}_l \mathbf{W}_{jl} \\ &= \text{Tr}(\mathbf{V}^T \mathbf{D} \mathbf{V}) - \text{Tr}(\mathbf{V}^T \mathbf{W} \mathbf{V}) \\ &= \text{Tr}(\mathbf{V}^T \mathbf{L} \mathbf{V}) \end{aligned} \quad (9)$$

其中, $\mathbf{Z}_j = [z_{j1}, z_{j2}, \dots, z_{jk}]^T$ 是第 j 个样本 x_j 的低维表示, \mathbf{W}_{jl} 是连接数据点 j 和数据点 l 的边的权重, \mathbf{D} 是对角矩阵,其条目是 \mathbf{W} 和 $\mathbf{D}_{jj} = \sum_j \mathbf{W}_{jl}$ 的列(或行,因为它是对称的)的总和。 $\mathbf{L} = \mathbf{D} - \mathbf{W}$, 它是图的拉普拉斯矩阵。这种方法使得属于同一聚簇下的不同数据点分布更紧密。

为了获取不同视图之间互补且一致的信息,Liu 等人利用 NMF 和 PLSA (Probabilistic Latent Semantic Analysis, 概率潜在语义分析)之间的关系,通过软正则的方法将不同视图下的指示矩阵 \mathbf{V} 推向共识^[28],其中软正则化的表达式为

$$\sum_{i=1}^m \lambda_i \|\mathbf{V}^i - \mathbf{V}^*\|_F^2 \quad (10)$$

其中, λ_i 可以调节不同视图之间的权重, \mathbf{V}^* 表示不同视图下的共识指示矩阵,将公式(10)代入到多视图 NMF 的目标函数中可以得到有约束条件下的目

标优化表达式:

$$\begin{aligned} & \sum_{i=1}^m \| \mathbf{X}^i - \mathbf{U}^i (\mathbf{V}^i)^T \|_F^2 + \sum_{i=1}^m \lambda_i \| \mathbf{V}^i - \mathbf{V}^* \|_F^2 \\ & \text{s. t. } \mathbf{U}^i, \mathbf{V}^i, \mathbf{V}^* \geq 0 \end{aligned} \quad (11)$$

大部分基于 NMF 的 MVC 方法普遍隐含视图间样本严格对齐的前提假设,然而实际场景中跨视图样本可能呈现部分对齐或完全未对齐状态。这种对齐缺失会导致各视图学得的簇结构空间分布不一致,进而导致不同视图下的样本集群结构不一致。Zong 等人通过定义视图间的两种约束 $ML^{(a,b)}$ (视图 a, b 的样本对属于同一聚簇)以及 $CL^{(a,b)}$ (视图 a, b 的样本对不属于同一聚簇)来最小化相关样本的差异达到获取样本一致性的集群排列结构的目的^[29]。

为了更好地利用多视图数据的多样性, Wang 等人首先通过 NMF 获得成对视图的指示向量 \mathbf{V}^a 和 \mathbf{V}^b , 然后令这两个指示向量的点积为 0, 记为 $\| \mathbf{V}^a \circ \mathbf{V}^b \|_0$, 以此保证两个向量之间的多样性。其中 $\| \cdot \|_0$ 表示 l_0 范数, 它代表非零元素的个数, 由于 l_0 范数具有非凸性和不连续性, 因此通过 l_1 范数来放宽优化, 可以得到一个多样性约束 (DIVE), 它可以表示为

$$DIVE(\mathbf{V}^a, \mathbf{V}^b) = \sum_{i=1}^n \sum_{j=1}^k \mathbf{v}_{ji}^a \cdot \mathbf{v}_{ji}^b = Tr(\mathbf{V}^a \cdot (\mathbf{V}^b)^T) \quad (12)$$

其中, $Tr(\cdot)$ 表示矩阵的迹。最后通过最小化 $Tr(\mathbf{V}^a, (\mathbf{V}^b)^T)$ 来激励指示矩阵 $\mathbf{V}^a, \mathbf{V}^b$ 相互正交, 以此来最大化视图间多样性。将 DIVE 融入多视图模型中得到具有多样性的非负矩阵分解 (DiNMF) 模型的目标函数^[30]:

$$\begin{aligned} & \sum_{i=1}^m \| \mathbf{X}^i - \mathbf{U}^i (\mathbf{V}^i)^T \|_F^2 + \vartheta \sum_{a \neq b} DIVE(\mathbf{V}^a, \mathbf{V}^b) \\ & + \beta \sum_{i=1}^m \| \mathbf{V}^i \|_F^2 \\ & \text{s. t. } 1 \leq a, b \leq m, \mathbf{U}^i, \mathbf{V}^a, \mathbf{V}^b, \vartheta \geq 0 \end{aligned} \quad (13)$$

其中, ϑ 是一个用来控制 DIVE 项权重的平衡参数, $\| \mathbf{V}^i \|_F^2$ 是一种平滑项, 为了避免视图过拟合, 使用 β 权重因子来控制平滑项大小。Tan 等人从稀疏因式分解中提取隐藏的结构信息, 并从误差矩阵中提取有价值的多样性判别信息^[31]。

使用矩阵分解的方法可能会增加视图间的冗余

信息和噪声, 为了解决这个问题, Ma 等人将 l_{2-1} 范数引入到模型中提出了一种具有判别表示的矩阵分解 (DMSMF) 方法, 由于 l_{2-1} 范数可以对特征维度进行稀疏化, 这使得它可以去除冗余特征, 降低噪声干扰, 提高模型鲁棒性^[32]。 l_{2-1} 范数还可以使得异常特征维度整体权重缩小, 从而降低异常值对模型的影响。但这种方法并不适用于随机噪声 (高斯噪声), 因为随机噪声通常分布在所有维度上, 使用 l_{2-1} 会影响整个数据集。另外, Huang 等人通过引入软加权正则项 $f(\mathbf{w}_{i-fea}^v, \lambda^v)$ 和 $f(\mathbf{w}_{i-sam}^v, \lambda_{sam}^v)$, 使得特征损失值较高的样本在训练过程中的贡献越来越小, 从而大大降低异常值和不相关特征的干扰^[33]。但是这种方法不适用于包含缺失数据的样本。

为了有效捕捉数据之间的内在关系, 保留数据的局部结构是必要的。常见的保留样本的局部流行结构的方法是采用流行正则^[34-36], 它的数学表达式如公式 (9) 所示。流行正则基于一种流行假设, 流行假设指的是如果两个数据点 x_i 和 x_j 在原始空间上相近, 那么这两个数据点的潜在表示也应该是相近的。与传统流行正则方法不同, Ma 等人利用伪标签形式, 通过对几个子问题的迭代优化来捕捉视图的内在几何结构^[32]。

然而, 传统的 NMF 主要是为无监督学习设计的, 这限制了它在半监督学习环境中的适用性。为了解决这个问题, Liu 等人提出一种有约束条件的非负矩阵分解 (CNMF) 方法, 它将标签信息作为额外的硬约束, 确保同一类别中的数据点被合并到一个统一的表示中^[37]。因此, CNMF 需要一个标签约束矩阵 \mathbf{A} , 其中,

$$\mathbf{A} = \begin{pmatrix} \mathbf{C}_{l \times c} & 0 \\ 0 & \mathbf{I}_{n-l} \end{pmatrix} \quad (14)$$

$\mathbf{C}_{l \times c}$ 是一种指示矩阵, 它表示有 l 个有标签的样本数据分别属于 c 个聚簇, 如果 $c_{ij}=1$ 表示第 i 个有标签的数据属于第 j 个聚簇, 否则 $c_{ij}=0$, $\mathbf{I}_{n-l} \in \mathbb{R}^{(n-l) \times (n-l)}$ 是一个单位矩阵, 表示剩下的 $n-l$ 个样本是没有标签的。为了将标注约束矩阵合并到矩阵 \mathbf{V} 中, 需要一个辅助矩阵 \mathbf{Y} , 这样 $\mathbf{V} = \mathbf{A}\mathbf{Y}$, 因此 CNMF 的目标函数表达式为

$$\mathbf{O} = \| \mathbf{X} - \mathbf{U}\mathbf{Y}^T \mathbf{A}^T \|_F^2 \quad (15)$$

为了增强聚类的判别能力, Cai 等人通过为每个视图引入一个辅助矩阵来学习表示, 并构建一个所有视图共享的标签约束矩阵, 确保标签信息能够有效地融合到每个视图的新表示中, 从而提高了聚

类的判别能力^[38]。Cui 等人提出了一种基于聚类簇间变异性信息和不同视图配准信息的融合正则化方法,该方法增强了特征的簇间变异性^[39]。Zhao 等人在模型中加入了半监督回归项,通过有效利用标签信息来捕捉数据的复杂流形结构,从而提高了低维特征的判别能力^[40]。Cai 还将标签信息纳入每个视图的聚类原型,通过正交项约束实现了最优表示^[41]。

与此同时,数据处理的计算复杂性也对 NMF 提出了巨大挑战。为了解决这一问题,Li 等人在 NMF 框架中引入了两个图正则化项,以实现多视图数据的高质量重建,进而减轻复杂数据对聚类性能的负面影响^[42]。Chen 等人提出了一种基于多样性损失的矩阵分解新型框架,旨在减少特征冗余的同时增强特征的判别能力^[43]。Yang 等人开发了一种快速多视图聚类(FMCNOF)方法,显著提高了大规模多视图数据的聚类效率,并有效降低了与矩阵因式分解相关的计算成本和整体复杂度^[44]。对称非负矩阵因式分解的提出也进一步减少了矩阵运算的计算量^[45]。

表 2 总结了部分相关算法的优缺点。

表 2 基于 NMF 的 MVC 方法

| 方法 | 优点 | 缺点 |
|---------------------------|-----------------------|---------------------|
| Multi-NMF ^[28] | 其聚类结果与多种视图兼容 | 不适合大规模数据 |
| SNMF ^[26] | 适用于真实数据集 | 计算成本高 |
| GNMF ^[27] | 具有更好的语义结构表示 | 模型的性能受以下参数的影响 |
| DMSMF ^[32] | 模型的稳健性 | 聚类性能对参数选择很敏感 |
| DSMVC ^[33] | 模型的稳健性 | 不适合大规模数据 |
| CMVNMf ^[29] | 处理未映射多视图数据的能力 | 不考虑各视图之间的高阶关系 |
| DiNMF ^[30] | 增强聚类多样性 | 不显示隐藏的视图语义信息 |
| LP-DiNMF ^[30] | 保留局部几何形状 | 不适用于高维数据 |
| RLMDOM ^[31] | 考虑到不同视图中信息的多样性,收敛速度很快 | 仅适用于二维矩阵,不适用于高阶张量 |
| MVCC ^[34] | 保留局部空间结构 | 不适用于大规模数据集 |
| ECNMF ^[35] | 获取关于各种观点的多样性和一致性的信息 | 未考虑缺失的现实数据 |
| DIMMA ^[36] | 捕捉不同对象之间的相互关系 | 不适用于高维数据 |
| CNMF ^[37] | 适合半监督聚类 | 不适用于大规模数据集 |
| MVCNMF ^[38] | 适用于半监督聚类 | 对参数敏感 |
| FRSMNMF ^[39] | 超参数少,判别能力强 | 不同的约束条件导致模型稳健性的巨大差异 |

(续表)

| 方法 | 优点 | 缺点 |
|--------------------------|--------------------|---------------|
| DHPS-NMF ^[40] | 保留数据的高阶关系 | 构建超图需要大量时间 |
| MVOC-NMF ^[41] | 增强视图之间的信息互补 | 不适用于数据流环境 |
| CCNMF ^[42] | 同时学习不同观点的一致性和互补性信息 | 对参数敏感 |
| DDMF ^[43] | 减少特征冗余,提高判别能力 | 未考虑各视图的一致性信息 |
| FMCNOF ^[44] | 降低计算复杂性 | 锚点的选择对聚类效应有影响 |

3.1.2 多视图子空间聚类

基于 NMF 的多视图聚类方法通过将数据分解到低维空间中表示可以处理一些高维数据,但是它难以利用不同视图间的互补信息产生具有一定判别表示的聚类结果。相较于 NMF 的多视图聚类方法,多视图子空间聚类方法通过自表示矩阵,可以充分利用不同视图的互补信息来揭示数据的潜在结构。其中每个数据点都可以用数据点本身的线性组合来表示,一般公式可以表示为

$$\min_{\mathbf{Z}} L_r(\mathbf{X}, \mathbf{XZ}) + \partial \text{Reg}(\mathbf{Z}) \quad (16)$$

其中, \mathbf{Z} 是数据集 \mathbf{X} 的自表达矩阵,根据多视图信息融合的不同阶段,现有的多视图子空间聚类主要分为前融合和后融合两类^[46]。

基于前融合的多视图子空间聚类主要学习多个视图数据潜在的,全面的表示,旨在增强下游聚类任务。这种方式允许来自不同视图的观察结果可以通过它们各自的重构模型 $\{P^1, \dots, P^m\}$ 和共享的潜在表示 $\mathbf{H} = \{h_i\}_{i=1}^n$ 来重建,不同视图下样本的重构表示可以描述为 $x_i^v = P^v h_i$, 由于存在一定的噪声,因此有

$$x_i^v = P^v h_i + e_i^v \quad (17)$$

其中, e_i^v 属于第 v 个视图的重构误差。基于前融合的多视图潜在表示的目标函数为

$$\min_{P, H} L_h(\mathbf{X}, \mathbf{PH}) \quad (18)$$

其中, $L_h(\cdot)$ 表示原始数据和用共享潜在表示重构数据之间的损失函数;另外,我们可以得到共享的潜在表示 \mathbf{H} 的子空间聚类目标函数:

$$\min_{\mathbf{Z}} L_r(\mathbf{H}, \mathbf{HZ}) + \partial \text{Reg}(\mathbf{Z}) \quad (19)$$

其中, ∂ 是用来调整正则项的平衡参数, $L_r(\cdot)$ 表示共享的潜在表示和它基于自表达矩阵的重构表示之间的损失函数。因此,基于前融合多视图子空间聚类的总体目标函数为

$$\min_{P, H, Z} L_h(\mathbf{X}, \mathbf{PH}) + \tau_1 L_r(\mathbf{H}, \mathbf{HZ}) + \tau_2 \text{Reg}(\mathbf{Z}) \quad (20)$$

其中, τ_1, τ_2 是一对平衡参数。Zhang 等人在此基础上使用 l_{2-1} 范数作为前两项的损失函数, 它相比于 Frobenius 范数, 对异常值更加稳健, 从而提升了模型的鲁棒性^[20]。在真实世界的应用中, 不同视图的数据特征和潜在表示之间的关系可能是非线性以及高维的, 使用传统核方法会存在泛化问题, 因此 Zhang 等人采用基于神经网络的方法来处理共享的潜在表示和视图原始特征的复杂关系^[47]。

大部分基于前融合的子空间聚类主要关注跨视图共享的潜在一致性表示, 而忽略了每个视图可能包含的独特知识, 这对利用多个视图的互补信息是一个挑战, 为了解决这个问题, Zhou 等人首先利用特定于视图的信息来封装每个视图的唯一属性, 然后将特定于视图的属性和共享的潜在一致性表示拼接在一起, 用于重建数据点, 其目标函数可以表示为

$$\min_{\mathbf{P}, \mathbf{H}_i, \mathbf{E}_i} \sum_{i=1}^m \|\mathbf{H}_i; \mathbf{H}_i\| - \mathbf{P}^i \mathbf{X}^i - \mathbf{E}_i \|_F^2 \quad (21)$$

其中, \mathbf{H}_i 表示特定于第 i 个视图的属性, \mathbf{P}^i 表示第 i 个视图的重构模型, \mathbf{E}_i 表示第 i 个视图下的噪声集合, 通过优化以上目标, 跨视图的互补信息得到有效利用^[48]。Si 等人基于共享表示和特定视图表示构建了一个具有一致性和互补性表示的亲矩阵 \mathbf{Q} , 它可以表示为

$$\mathbf{Q} = \frac{|\mathbf{H}| + |\mathbf{H}|^T}{2} + \frac{1}{m} \sum_{i=1}^m \frac{|\mathbf{H}_i| + |\mathbf{H}_i|^T}{2} \quad (22)$$

然后在 \mathbf{Q} 上进行谱聚类, 获得具有一致性和互补性的聚类结果^[49]。

相比于前融合多视图子空间聚类的方法, 基于后融合的多视图子空间聚类更强调将多个视图子空间表示 \mathbf{Z}^i 融合为统一的空间 \mathbf{Z} , 然后基于 \mathbf{Z} 可以进行聚类操作。

为了有效利用跨视图的互补信息, Luo 等人将统一空间 \mathbf{Z} 分解为一个一致性矩阵 \mathbf{J} 和特定于视图的矩阵 \mathbf{D}^i , 通过对 \mathbf{J} 和 \mathbf{D}^i 施加相应的约束, 达到同时获取跨视图一致性和互补性信息^[50]。这个过程中跨视图多样性信息和噪声是分开的, 可以用公式描述为

$$\mathbf{X}^i = \mathbf{X}^i (\mathbf{J} + \mathbf{D}^i) + \mathbf{E}_i \quad (23)$$

与 Luo 等人的方法不同, Cai 等人提出一种联合平滑的多视图子空间聚类 (JSMC) 方法, 该方法把不同视图数据的重构偏差看作跨视图多样性的一部分, 然后对特定视图的信息 \mathbf{D}^i 进行低秩约束, 从

而消除噪声的干扰^[51]。但现有大部分方法都是直接从原始空间获得子空间表示, 这可能会存在噪声无法保证子空间的干净。为了解决这个问题, Chen 等人提出一种综合的多视图自表达聚类 (CMSR) 方法, 该方法首先利用自表达矩阵将原始数据投影到低维空间中, 然后对不同视图下的低维表示矩阵进行分解, 得到一个共识一致性矩阵 \mathbf{J} , 最后对矩阵 \mathbf{J} 施加 Schatten p 范数来捕捉具有鲁棒性的一致性信息^[52]。Zhang 提出了共识一步多视图子空间聚类 (COMVSC) 方法, 该方法首先获取不同视图下的自表达矩阵, 然后对不同自表达矩阵进行聚类比如 K -均值, 获得不同视图下的聚簇, 最后将它们融合, 获得统一聚簇下的聚类结果。它解决了视图融合和聚类过程分开造成的次优聚类问题^[53]。

大多数基于后融合的多视图子空间聚类方法过于追求跨视图一致性, 从而忽略不同视图之间的差异性和多样性。为了更好地探索不同视图间的多样性, Cao 等人定义了一个基于希尔伯特正则化的多样性约束, 记为: $HSIC(\mathbf{Z}_i, \mathbf{Z}_j)$, 它通过最小化特征表示 \mathbf{Z}^i 和 \mathbf{Z}^j 之间的相关性以此来最大化视图之间的多样性^[54]。在此基础上, Lan 等人提出一种双层相关视图多视图聚类 (DVMSC) 方法, 通过引入不同多样性项来探索不同视图之间的多样性^[55]。使用 $HSIC$ 作为约束的缺点是 $HSIC$ 是一种数值感知标准 (即 \mathbf{Z}^i 和 \mathbf{Z}^j 之间的多样性与它们的数值有关) 由于 \mathbf{Z}^i 的尺度问题, 数值感知正则化可能会大大降低性能, 因此 Wang 等人提出一种基于位置感知的多样性项, 它可以表示为

$$\mathbf{H}(\mathbf{Z}^i, \mathbf{Z}^j) = \|\mathbf{Z}^i \odot \mathbf{Z}^j\|_0 \quad (24)$$

其中 \odot 表示 Hadamard 乘积, $\|\cdot\|_0$ 表示 l_0 范数, 这个多样性项鼓励两个矩阵尽可能多样化^[56]。另外, Huang 等人则为每个视图构建了相似性图, 随后确定其内在的一致性和多样性, 从而有效地利用了视图中国存在的互补信息^[57]。

为了有效保留数据的局部流行结构, Liu 等人提出一种自适应局部一致图正则化的多视图子空间聚类 (CGMSC) 方法, 该方法保留了数据的流形结构, 并通过加入局部图正则化项来对齐潜在子空间表示^[58]。Su 等人将图正则项集成到共享系数矩阵中, 保留局部流行结构同时还有效减少了噪声对子空间的影响^[59]。Gu 等人通过定义一个超图正则化约束保护了子空间表示数据的固有结构^[60]。

基于后融合的多视图子空间聚类方法主要通过在一个公共或多个私有子空间中对各个视图的数据

进行投影,从而发现数据的低维结构。然而,这类方法通常难以捕捉视图间的高阶相关性,解决这类问题的主要方法是通过将不同视图对应的自表达矩阵堆叠成一个张量形式,用处理张量的方法,比如,张量分解,可以更好地捕捉视图间高阶复杂关系。Zhang 等人提出一种基于张量的多视图子空间聚类方法(LT-MS),该方法首先使用张量核范数形式,将 2 阶张量推广到多阶张量,即

$$\|\mathbf{X}\|_* = \sum_{i=1}^m \xi_i \|\mathbf{Z}_i\|_* \quad (25)$$

其中, $\|\cdot\|_*$ 表示张量核范数,它强制要求张量在低秩约束下。 ξ_i 满足 $\xi_i > 0$ 并且 $\sum_{i=1}^m \xi_i = 1$ 。 $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_m}$ 是一个 m 阶的张量, \mathbf{Z}_i 是沿第 i 个模式展开的张量矩阵。由此带有低秩张量约束的 LT-MS 的目标函数为

$$\begin{aligned} \min_{\mathbf{Z}_i, \mathbf{E}_i} & \|\mathbf{X}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s. t. } & \mathbf{X}^i = \mathbf{X}^i \mathbf{Z}_i + \mathbf{E}_i, i = 1, 2, \dots, m, \\ & \mathbf{X} = \phi(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m), \\ & \mathbf{E} = [\mathbf{E}_1; \mathbf{E}_2; \dots; \mathbf{E}_m] \end{aligned} \quad (26)$$

其中, $\phi(\cdot)$ 表示将不同视图下子空间表示堆叠成一个三阶张量, \mathbf{E} 是将每个视图的误差列垂直连接在一起的矩阵。具有低秩性的张量很好地探索了跨视图间复杂的高阶关系^[61]。在此基础上, Xiao 等人通过将张量学习集成到子空间聚类中,探索了一种高阶相关结构,同时他们引入了位置感知排他性项,以捕捉每个视图的独特特征并增强特征互补性^[62]。由于多视图子空间的张量分解存在高度不平衡或旋转感, Long 等人利用多尺度纠缠重正则解析(MERA)解决了这个问题^[63]。另外, Deng 等人介绍了一种多视图聚类算法,重点是信息融合和增强。该方法通过纳入数据的联合视图来学习有利于信息融合的子空间,从而增强潜在信息并探索视图间的高阶相关性^[64]。

由于后融合产生的公共子空间可能存在矩阵分解或者张量分解,这增加了计算复杂度,为了降低时间开销,一些基于锚点学习的多视图子空间方法被提出^[59,65]。其中 Zhou 引入了一种新颖的锚选择方法,可以降低数据维度,从而最大限度地减少存储需求和时间开销^[65]。Qu 等人通过在公共子空间内应用均匀采样策略,将时间复杂度降低到线性时间内^[66]。大多数基于启发式的锚点采样和聚类过程是相互独立的,这导致区分锚点能力较弱同时也没

有很好地利用多视图间的互补信息, Sun 等人提出一种具有统一锚点的可伸缩多视图子空间聚类(SMVSC)方法,该方法将锚点学习和图重构放在一个优化框架中,这使得锚点可以更准确表示实际底层的数据分布,获得更具判别性的聚类结构^[67]。

子空间聚类是基于自表达矩阵对原始数据点的重构,因此在这个过程中会产生重构偏差(噪声、缺失项、异常值)。为了解决这个问题, Liu 等人提出一种低秩鲁棒性表示(LRR)方法,它通过优化以下目标函数实现去噪^[68]:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s. t. } & \mathbf{X} = \mathbf{XZ} + \mathbf{E} \end{aligned} \quad (27)$$

λ 是用于平衡低秩表示和噪声分离的超参数。这种方法通过对子空间施加核范数约束,保证了子空间的低秩性,从而恢复干净数据,同时对噪声施加 l_{2-1} 范数,捕获列稀疏性,使得异常点影响集中到少数列中,而不影响全局结构。由于 LRR 是基于噪声是无结构的假设,因此它并不适用于一些结构化噪声,比如条纹噪声,周期性干扰和局部区域遮挡。在此基础上 Wang 等人提出了基于结构化多路径网络的多视图子空间聚类(SMpNet)方法,该方法对子空间和高级结构化特征之间的连接矩阵应用低秩约束,从而将噪声的影响降至最低^[69]。对于一些有结构的高频噪声比如电磁干扰等,可以采用平滑的方式处理, Huang 采用低通滤波器从多个视图中得出平滑表示,提高了后续聚类任务的性能^[70]。Wei 等人设计了一种全新的自适应图滤波器,通过平滑数据增强了多视图融合过程的一致性,也降低了噪声的干扰^[71]。但是基于平滑数据的方式无法明显改善一些随机噪声,比如高斯噪声,因为它在数据上面的分布比较均匀。处理这类噪声通常使用均值滤波或者高斯滤波的方法。

我们在表 3 总结了部分基于子空间聚类算法的优缺点。

3.1.3 基于图学习的多视图聚类

虽然基于子空间的多视图聚类方法可以很好地利用跨视图间的互补信息,但是它却很难解释不同样本之间的关系,这不利于我们学习样本之间隐含的更深层次信息。基于图学习的方法通过相似性矩阵(亲和矩阵)来建模样本之间的关系,这种方法有利于学习数据的局部信息。

基于图学习的多视图聚类方法主要是根据不同的图构造方法来获取不同视图下的相似图(亲和图),然后通过图融合方法获得跨视图下的共识亲和

表 3 基于子空间学习的 MVC 方法

| 方法 | 优点 | 缺点 |
|--------------------------------------|------------------------------|----------------------|
| LRR ^[68] | 数据损坏时可恢复数据 | 在数据污染的情况下难以选择参数 |
| LT-MSC ^[61] | 捕捉数据中的高阶相关性 | 对参数敏感 |
| DIMSC ^[54] | 通过补充信息增强视图 | 参数对聚类效果的影响更大 |
| LMSC ^[20] | 具有更好的鲁棒性 | 不适用于大规模数据 |
| ILMSC, GLMSC ^[47] | 学习子空间的复杂结构 | 计算成本高 |
| DSS-MSC ^[48] | 捕捉描述每个独立视图特定属性的视图特定细节 | 未考虑不同观点之间的复杂相关性 |
| COMVSC ^[53] | 稳健性和更好的判别能力 | 计算复杂度高 |
| CGMSC ^[58] | 确保视图局部流形结构和视图一致性 | 随着数据量的增加,无法应对数据的不确定性 |
| CDMSC ^{[2][49]} | 同时学习一致性和多样性表示矩阵以寻找聚类结构 | 不适合真实世界的的数据 |
| SMpNet ^[69] | 探索多视图综合互补结构,减少噪音影响 | 不适用于大规模数据集 |
| MSC ² D ^[57] | 获更清晰的聚类结构图和快速收敛的算法 | 不适用于大规模数据集 |
| SMVC ^[67] | 降低大规模数据集的时间复杂性 | 忽略锚点选择对聚类的影响 |
| ECMSC ^[56] | 利用跨视图的表征完整性 | 不同数据集的聚类性能差异很大 |
| JSMC ^[51] | 各视图信息的一致性和不一致性都要考虑 | 不适合大规模应用 |
| AMVSCGL ^[59] | 解决大规模数据问题 | 锚点的选择对聚类的影响相对较大 |
| MERA-MSC ^[63] | 有效利用视图间和视图内的相关性 | 获得的基本聚类结构不准确 |
| FEMV ^[64] | 增强潜在信息,提高信息可用性 | 计算大规模数据的成本高昂 |
| DV-MSC ^[55] | 扩展基础空间多样性,探索高级聚类结构 | 计算时间成本高 |
| MVSC-HFD ^[66] | 探索视图间的可变性以提高聚类精度 | 不适用于大规模数据集 |
| CDTMS ^[62] | 探索多视图的多样性,获得出色的聚类结果 | 实验迭代次数越少,误差越大 |
| CMSR ^[52] | 提供可靠的视图一致性信息 | 不适用于大规模数据集 |
| CSMSC ^[50] | 适用于真实数据集 | 未考虑不同观点对聚类的影响 |
| NOODLE ^[46] | 捕获了多个视图存在的高阶相关性 | 涉及张量分解,计算复杂度比较高 |
| STONE ^[60] | 降低了计算复杂度并提高了弹性,还保留了数据的局部几何结构 | 不同的锚点选择可能会影响聚类的性能 |
| MVSC ² GF ^[71] | 降低了噪声对聚类的干扰 | 计算复杂度比较高 |

图,最后在共识亲和图的基础上进行聚类,获得最终聚类结果。我们根据学习共识亲和图的不同策略,可以将基于图学习的多视图聚类方法分为多视图谱聚类,基于自适应邻居图学习(ANGL)的多视图聚类,以及基于锚点图学习的多视图聚类。

多视图谱聚类:单个视图的谱聚类算法是通过归一化的图拉普拉斯矩阵解决以下优化问题:

$$\begin{aligned} \max_{\mathbf{C}^v \in \mathbf{R}^{n \times k}} \text{tr}(\mathbf{C}^{(v)\top} \mathbf{L}^v \mathbf{C}^v) \\ \text{s. t. } \mathbf{C}^{(v)\top} \mathbf{C}^v = \mathbf{I} \end{aligned} \quad (28)$$

其中, \mathbf{C}^v 表示第 v 个视图的聚类指示矩阵, $\mathbf{L}^v = \mathbf{D}^{(v)-1/2} \mathbf{S}^v \mathbf{D}^{(v)-1/2}$ 表示归一化的图拉普拉斯矩阵, \mathbf{D} 是度矩阵, \mathbf{S}^v 表示第 v 个视图的相似矩阵或者核矩阵,如果相似矩阵 \mathbf{S}^v 中所有的元素都是非负的,那么它的拉普拉斯矩阵 \mathbf{L}^v 有以下性质:0 作为 \mathbf{L}^v 的特征值,它的多重性数量 k 等于 \mathbf{S}^v 的连通分量数,该性质表明,如果满足约束条件 $\text{rank}(\mathbf{L}^v) = n - k$,则 \mathbf{S}^v 是理想的邻域分配,数据点已被划分为 k 个簇。我们把式子(28)引入到多视图数据中,就可以获得多视图下谱聚类的目标函数:

$$\max_{\mathbf{C}^1, \mathbf{C}^2, \dots, \mathbf{C}^m \in \mathbf{R}^{n \times k}} \sum_{i=1}^m \text{tr}(\mathbf{C}^i \mathbf{L}^i \mathbf{C}^{(i)\top}) \quad (29)$$

通常这些不同视图承认相同的数据基础聚类,因此可以通过查找跨视图一致的聚类来解决这个问题。Kumar 等人基于两种不同的假设(即:(1)不同视图下的相应样本具有同一聚类资格;(2)不同视图下的聚类存在一个兼容聚类。)提出了两种正则化的方法:(1)给成对视图施加一个约束,强迫成对视图的聚类指示矩阵的特征向量具有很高的相似性。(2)给不同视图下的聚类指示矩阵的特征向量施加一个正则约束,使得它们具有一个共识的聚类指示矩阵的特征向量。通过以上 2 种正则化方法获得具有跨视图一致的聚类结构^[72]。在真实场景中,不同视图数据可能存在很大的噪声,这通常会降低聚类性能。为了解决这个问题,Xia 等人提出一种具有鲁棒性的多视图谱聚类(RMSC)的方法,该方法首先从每个视图中构造一个转移概率矩阵 \mathbf{M}^v ,然后利用低秩约束将这些矩阵恢复为一个共享的转移矩阵 \mathbf{M} ,最后基于共享转移矩阵 \mathbf{M} 进行谱聚类。RMSC 的构造转移矩阵问题可以表述为

$$\begin{aligned} \min_{\mathbf{P}, \mathbf{E}_i} \text{rank}(\mathbf{M}) + \lambda \sum_{i=1}^m \|\mathbf{E}_i\|_0 \\ \text{s. t. } i = 1, 2, \dots, m, \mathbf{M}^i = \mathbf{M} + \mathbf{E}_i, \mathbf{M} > 0, \mathbf{M}\mathbf{1} = \mathbf{1} \end{aligned} \quad (30)$$

其中, $rank(M)$ 是矩阵 M 的秩, E_i 表示视图 i 的噪声, $\mathbf{1}$ 表示值全为 1 的向量。Xia 等人分别使用核范数 $\|M\|_*$ 和 l_1 范数 $\|E_i\|_1$ 来优化问题(30)以此来降低噪声对聚类的干扰^[73]。在此基础上, Wang 等人提出一种具有鲁棒性的低秩张量近似的多视图聚类(ELRTA)方法,该方法将不同视图下的转移矩阵堆叠成一个张量的形式,然后利用张量分解的方法得到一个共识转移矩阵,然后对共识转移矩阵施加一个张量核范数约束,既降低了噪声的干扰又很好地探索了视图间的高阶关系^[74]。

基于 ANGL 的多视图聚类:通常,ANGL 自动将每个样本的概率分配为另一个样本的邻值来学习亲和图,ANGL 的公式可以表述为

$$\min_S \left(\sum_{i=1}^n \sum_{j=1}^n D(\mathbf{x}_i, \mathbf{x}_j) s_{ij} \right) + \gamma \text{Reg}(\mathbf{S}) \quad (31)$$

$$\text{s. t. } \mathbf{S}^T \mathbf{1} = \mathbf{1}, 0 \leq s_{ij} \leq 1$$

其中,距离标准 $D(\cdot)$ 通常替换为任意两个样本之间的欧几里得距离,即 $\|\mathbf{x}_i - \mathbf{x}_j\|_2^2$, 距离越小就表示两个样本之间的相似性 s_{ij} 就越大, γ 是关于亲和图 \mathbf{S} 的正则化系数。通过公式(4)可以有效地捕捉数据点之间的局部流行结构。根据不同的融合方式(前融合与后融合),我们可以得到基于 ANGL 的多视图融合的 2 种公式:

$$\min_S \left(\sum_{v=1}^m \sum_{i=1}^n \sum_{j=1}^n D(\mathbf{x}_i^v - \mathbf{x}_j^v) s_{ij}^v \right) + \gamma \text{Reg}(\mathbf{S}) \quad (32)$$

$$\text{s. t. } \mathbf{S}^T \mathbf{1} = \mathbf{1}, 0 \leq s_{ij} \leq 1$$

$$\min_{S, S^v} \sum_{v=1}^m \sum_{i=1}^n \sum_{j=1}^n D(\mathbf{x}_i^v - \mathbf{x}_j^v) s_{ij}^v$$

$$+ \beta \varphi(\mathbf{S}^v, \mathbf{S}) + \gamma \text{Reg}(\mathbf{S}) \quad (33)$$

$$\text{s. t. } \mathbf{S}^T \mathbf{1} = \mathbf{1}, 0 \leq s_{ij} \leq 1$$

其中, $\varphi(\cdot)$ 表示将学习到的不同视图下的亲和图 \mathbf{S}^v 通过语义空间中的一些互补和一致性探索技术产生共识图 \mathbf{S} 。基于公式(32)和公式(33)出现了很多基于 ANGL 的多视图聚类方法, Peng 等人提出了无参数选择多视图聚类(COMIC)方法,该方法首先学习投影空间中的连接图,然后最小化来自不同视图的成对连接图之间的差异,从而探索跨视图的一致信息^[75]。Zhang 等人提出了可以联合嵌入学习的多视图张量积分(CELT)方法,该方法将所有视图内的亲和图堆叠成一个张量,以学习捕捉视图间一致性的共识亲和图^[76]。如果把直接从原始矩阵学习到的亲和图作为输入会存在一定的噪声,因

此很多方法都将多视图数据投影到一个低维的共享嵌入空间,然后在这个空间中学习到干净的亲和图^[77-79]。Sun 等人提出一种可以特征加权的多视图图聚类(FWMGC)方法,该方法通过自表达的方法获得了特定视图下的特征矩阵 \mathbf{W}^v , 然后对特征矩阵施加低秩约束,从而降低冗余特征,噪声和异常值的干扰,但是与传统使用核范数 $\|\cdot\|_*$ 的方法不同,该方法通过将 \mathbf{W}^v 分解成两个低秩矩阵 \mathbf{T}^v 和 \mathbf{Q}^v 的形式降低了使用核范数可能带来的计算复杂度^[80]。

基于锚点图学习的多视图聚类:为了减少图聚类昂贵的计算和存储支出,基于锚点图学习的方法选择 r 个锚点 $A^i \in \mathbb{R}^{r \times d_i}$ 作为整个数据的代表,然后使用小尺寸锚点图 $G^i \in \mathbb{R}^{n \times r}$ 。这一步涉及为每个视图构建单独锚点图,可以通过以下问题实现:

$$\min_{G^i} \|\mathbf{X}^i - \mathbf{G}^i \mathbf{A}^i\|_F^2 + \Omega(\mathbf{G}^i) \quad (34)$$

$$\text{s. t. } \mathbf{G}^i \geq 0, \mathbf{G}^i \mathbf{1}_m = \mathbf{1}_n$$

\mathbf{G}^i 是在目标函数有约束条件下的优化,条件 $\mathbf{G}^i \mathbf{1}_r = \mathbf{1}_n$ 保证每个数据点与所有锚点的总相似度被标准化为统一。根据构建的特定视图下的锚点图 $\{\mathbf{G}^i\}_{i=1}^m$, 传统锚点图融合的公式如下:

$$\min_{\partial \geq 0, \mathbf{G}} \left\| \sum_{i=1}^m \partial_i \mathbf{G}^i - \mathbf{G} \right\|_F^2, \quad (35)$$

$$\text{s. t. } \mathbf{G} \mathbf{1}_m = \mathbf{1}_n, \partial^T \mathbf{1}_m = 1$$

其中, $\partial \in \mathbb{R}^m$ 是视图的权重, \mathbf{G} 是融合之后的锚点图,是在其目标函数下的优化,通常情况下公式(35)可以通过迭代优化的方法来解决。我们可以通过锚点图 \mathbf{G} 来进行最终聚类。在此基础上, Zhang 等人提出了一种新颖的二元多视图聚类框架(BMVC),该模型适用于大规模多视图图像数据集^[81]。Zhao 将动态滤波器与学习到的锚点图信息相结合,在每次迭代时对数据进行自适应更新,从而实现了更灵活的数据处理^[82]。Lao 引入了一种基于锚点的子空间表示法来学习锚点图,并对其进行有效分割以生成基簇。这些基簇随后会被重构为统一的双簇图,用于最终聚类^[83]。Han 建议将由锚点和样本组成的双向图堆叠成一个张量,利用锚点和样本作为最终聚类的基础,通过结合低秩约束,有效捕捉样本间和视图间的相关性^[84]。

由于数据噪声可能会带来锚点偏移的现象,使得偏离锚点的潜在表征,导致性能下降, Li 等人提出一种噪声过滤机制的方法,通过压低噪声特征解

决了锚点偏移的现象,使得学习到的锚点有很强的表示能力^[85]。为了有效处理不同视图之间的不一致信息和噪声,Liu 等人提出一种具有鲁棒性的一致锚点图学习的多视图聚类(RCAGL)方法,该方法将特定视图下的锚点图 G^i 分解为一个共识亲和图 J 和特定视图下的噪声矩阵 E_i ,该矩阵包含不同视图间存在的不一致信息,然后通过引入一个正则项使得可以通过搜索适当的参数来调整不一致信息的稀疏程度,降低噪声干扰^[86]。在大规模场景中存在多视图锚点对齐问题(不同视图产生的锚点会存在语义不一致问题),为了解决这个问题,Wang 等人提出一种具有交叉视图对应锚对齐的可伸缩多视图图聚类(MV-CAGAF)的框架,该框架通过将不同视图下的锚点统一到 n 维空间中,然后定义了一种成对锚点不一致目标函数,探索每个视图锚点和样本之间的关系,成对不一致性越大,锚点未对齐的概率越大,通过最小化成对视图不一致目标函数,有效对齐了不同视图下的锚点^[87]。

为了获得一步聚类结果,而无需后处理,一些方法通过给共识亲和图施加一个图拉普拉斯秩约束 $rank(L) = n - k$,其中, L 是共识亲和和锚点图 G 的拉普拉斯矩阵。以此达到一步聚类的目的^[88-89]。此外,Zhao 等人提出了基于锚点图的一步多视图聚类特征选择(AGFS-OMVC)方法,该方法将锚图投影到标签空间,直接获得聚类标签,无需后处理^[90]。

近些年来,一些基于图学习和其他技术相结合的多视图聚类方法被相继提出,比如:Dai 等人利用强化学习通过图一致性奖励机制来推断未知的聚簇数量 k 。具体来说,在每次迭代期间评估聚类指标矩阵的多样性 k 。然后将未知 k 的推理过程表述为一种简洁的强化学习范式,其中奖励是通过集群间连接性来衡量的。因此,这种能够独立生成最终的聚类结果,而无需输入预定义的聚类编号^[91]。除此以外,Tan 等人通过将特定视图的亲亲和矩阵构建为一个张量形式,然后旋转该张量后切开,得到第 i 个前额切片,它对应第 i 个样本实例,最后通过将视图在样本级上融合,以视图的方式揭露了不同视图间的局部结构^[92]。

表 4 介绍了几种相关算法的优缺点。

图 9 展示了基于图学习的多视图聚类的工作流程。

3.1.4 基于张量学习的多视图聚类

尽管基于图学习的方法能有效捕获局部数据之间的关系,但在处理高维数据和探索视图间的高阶

表 4 基于图学习的 MVC 方法

| 方法 | 优点 | 缺点 |
|------------------------------|-----------------------------|------------------------------|
| BMVC ^[81] | 适用于大规模数据集 | 对某些参数的初始化敏感 |
| RMSC ^[73] | 具有更好的鲁棒性 | 不适合大规模数据集 |
| COMIC ^[75] | 在各种观点之间达成共识 | 不适合半监督学习 |
| MLEE ^[78] | 处理高维数据和干扰数据 | 参数选择对聚类有影响 |
| LLPGL ^[79] | 有效整合补充信息 | 不适合高维数据 |
| FWMGC ^[80] | 提高基本特征的重要性可降低计算复杂性 | 不适合大规模数据集 |
| MVBGC-NF ^[85] | 解决了由噪声特征引起的锚点偏差问题 | 锚点选择会对聚类结果产生影响 |
| SMCMB ^[83] | 适用于大规模数据集 | 不适用于复杂数据和不完整数据 |
| RCAGL ^[86] | 捕捉视图之间的共通点,过滤视图特有的噪音 | 不适用于不完整的多视图数据 |
| OMVFC-LI-CAG ^[88] | 模型具有可扩展性,可直接获得聚类结果 | 算法不稳定 |
| MLAN ^[89] | 直接形成集群结构,无需后期处理 | 某些真实数据集不可用 |
| AGFS-OMVC ^[90] | 坚固耐用,无需后期处理 | 忽略聚类效应的锚点选择 |
| CELT ^[76] | 增强观点的共识亲和力 | 没有考虑不同视角的重要性对聚类的影响 |
| ELRTA ^[74] | 利用马尔可夫转移矩阵得到一个具有鲁棒性的聚类结果 | 涉及张量分解,计算复杂度比较大 |
| MV-CAGAF ^[87] | 解决锚点未对齐问题,同时该方法可以拓展到大规模数据集中 | 在大规模不完整视图场景下,无法捕捉交叉视图锚点的对应关系 |

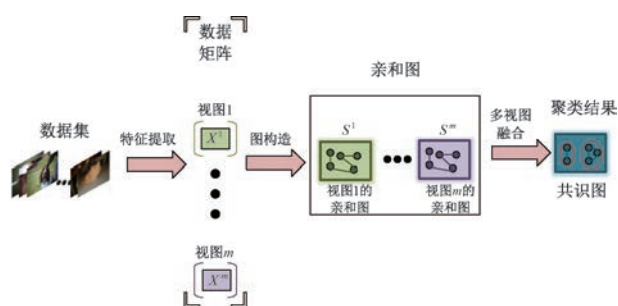


图 9 利用图形学习进行多视图聚类的一般过程

关系还存在不足,基于张量学习的多视图聚类因其能够有效处理视图信息融合和高维数据而受到了广泛关注。这种方法通过将多视图数据表示为张量结构,并应用张量分解等技术来进行聚类。传统的张

量分解方法包括 CP 分解^[24]、Tucker 分解^[93]、HOSVD 分解^[94]以及张量奇异值分解(T-SVD)。其中,T-SVD 已被证明在多视图数据的降维和降噪方面尤为有效。图 10 展示了 T-SVD 的分解过程。对于张量 \mathbf{X} ,其奇异值分解形式为

$$\mathbf{X} = \mathbf{U} * \mathbf{A} * \mathbf{V} \quad (36)$$

其中, $\mathbf{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ 、 $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ 是正交张量, \mathbf{A} 是 f 对角张量。Liu 等人提出了具有双约束的增强张量多视图聚类(ETMC-DC)方法,该方法将相似性矩阵堆叠成一个张量,并使用 T-SVD 方法提取大奇异值中编码的结构信息^[95]。Gao 等人将 T-SVD 方法应用于多视图子空间聚类,有效捕捉了跨视图的高阶互补信息^[96]。Qin 等人提出了一种基于马尔可夫链的柔性张量学习(FTLMCM)方法,该方法利用带有低阶约束的 T-SVD 来描述张量相似性矩阵的局部结构^[97]。

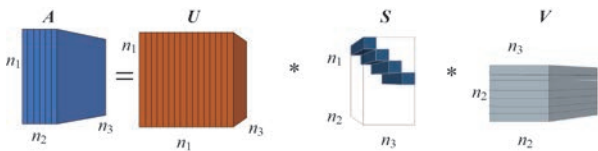


图 10 张量学习中的 T-SVD 分解过程

为了在多个视图获得具有互补性和一致性的信息,Xie 等人提出了张量多秩最小化聚类(T-SVD-MSD)方法,该方法先将不同视图下的自表示系数矩阵构建为一个张量形式,其次旋转张量,将自表示系数(模式 1)转换为模式 3,这使得沿着傅里叶域(模式 3)的每个前额切片都考虑了不同样本和不同视图信息,然后利用张量核范数(T-NN)对张量施加低秩约束,它可以更彻底、更有效地在所有视图之间探索和传播互补信息^[98]。但是自表达属性是对原始空间数据的编码,通常不能揭示真实数据下的非线性结构,因此 Chen 等人利用核方法,通过定义一个核函数 $\phi^v: \mathbb{R}^{d_v} \rightarrow \mathbf{H}^v$,将原始空间的数据映射到核空间中,然后将核空间下的数据表示矩阵堆叠成张量,利用张量学习的方法探索跨视图的互补性和一致性信息^[99]。另外,Pu 等人将不同视图下的亲和图堆叠成张量形式,然后利用非凸近似 TNN 的方法对张量进行约束,从而获得具有鲁棒性的低阶张量,然后把张量分解为具有一致性信息的共识亲和图^[100]。Chen 等人利用低阶张量近似法确保学习到的表示张量有效捕捉数据点之间的相似性^[101]。

张量核规范在低阶近似中非常重要,它被定义

为张量所有正切片的奇异值之和,其表达式为

$$\|\mathbf{X}\|_* = \sum_{k=1}^{n_3} \|\mathbf{X}_f^k\|_* = \sum_{i=1}^{\min(n_1, n_2)} \sum_{k=1}^{n_3} |\mathbf{X}_f^k| \quad (37)$$

其中, \mathbf{X}_f^k 是张量 \mathbf{X} 在三维空间中通过傅立叶变换得到的第 k 个矩阵, $\mathbf{X}_f^k(i, i)$ 是张量 \mathbf{A} 在三维空间中通过傅立叶变换得到的第 k 个正切片对应的第 i 个特征值。

为了有效捕捉视图间的相关性,Wu 等人提出了基于张量学习的多视图谱聚类(ETLMSC)方法,该方法通过张量核规范学习具有低阶约束的基本张量,从而探索这些相关性^[102]。Chen 等人利用 T-SVD 中的张量核规范来编码张量的低阶属性,以探索视图间的高阶关系^[103]。Ma 等人将不同视图的矩阵堆叠成一个张量,应用旋转操作,然后使用基于 T-SVD 的张量核规范来探索视图之间的高阶关系^[104]。Fu 等人提出了统一低秩张量学习和多视图子空间聚类的谱嵌入(ULTLSE)方法,该方法通过使用 T-SVD 张量核规范恢复基本分量,从而捕捉视图信息的互补性和一致性^[105]。

为了有效学习视图内和跨多个视图的嵌入特征的相关性,Long 等人提出一种可扩展的张量多视图聚类(S^2 MVTC)方法,该方法构建了一个张量低频近似器(TLFA),将不同视图的相似性矩阵整合到嵌入的特征学习中,有效实现了不同视图的特征平滑表示,捕获了跨视图相似性信息。同时施加一个共识约束,保证了视图内的语义一致性^[106]。但是大部分基于张量的方法很难描述相似性矩阵的局部结构,为了解决这个问题,Qin 等人将不同视图的相似性矩阵堆叠成一个张量,然后对张量的所有正面切片进行低秩和对称性正则化,即:

$$\begin{aligned} \|\mathbf{X}\|_f &= \sum_{i=1}^m \text{rank}(\mathbf{X}_f^i) \\ \text{s. t. } \mathbf{X}_f^i &= \mathbf{X}_f^{(i)\top} \end{aligned} \quad (38)$$

这种方法不仅从全局角度描述了张量不同条目之间的关系,而且从不同角度描述了相似性矩阵的局部结构^[97]。

大部分基于张量学习的多视图聚类方法,平等地对待不同视图,从而忽略了不同观点对聚类的重要性,为了解决这个问题,Xie 采用了一种加权张量核规范方法来约束旋转张量^[107],而 Pan 则提出了一种非凸张量核规范方法,该方法可以自适应地分配权重,并逼近张量的真实秩,从而减轻核规范^[108]

带来的秩误差。在此以后,Zhang 则以 Schetten p 范数作为张量约束,也充分挖掘了视图间的互补信息^[109]。

表 5 介绍了几种相关算法的优缺点。此外,为了帮助读者更清晰地区分上述不同的启发式 MVC 方法,我们对它们的关键差异进行了如下总结:

表 5 基于张量学习的 MVC 方法

| 方法 | 优点 | 缺点 |
|--------------------------------------|------------------|----------------------|
| T-SVD-MS ^[98] | 稳健性更好 | 不适合大规模数据集 |
| ETLMS ^[102] | 探索观点的高阶相关性 | 计算复杂度高,不适合大规模应用 |
| GLTA ^[103] | 亲和矩阵和张量表征可同时学习 | 不适合真实数据集 |
| SETKMC ^[101] | 非线性问题迎刃而解 | 不同数据集的聚类效果大相径庭 |
| OMLT ^[104] | 探索多视图高阶相关性 | 音频数据集的聚类效果不佳 |
| ULTLSE ^[105] | 区分各种观点的重要性 | 计算复杂度高,不适合大规模应用 |
| TLRR ^[107] | 探索视图高阶相关性和群组内在结构 | 不适合大规模数据集 |
| RLGMC ^[100] | 更强的鲁棒性 | 实际应用中确定非凸函数参数的值具有挑战性 |
| FTLMCM ^[97] | 探索当地结构 | 参数在某些数据集上很容易达到局部最优化 |
| ETMC-DC ^[95] | 获取了更高阶的信息 | 计算复杂度高 |
| GLMC-TBG ^[109] | 挖掘数据的互补性和高阶相关性 | 忽略锚点选择对聚类的影响 |
| S ² MVTC ^[106] | 捕获了跨视图间的相似信息 | 基于张量学习的方法,存在较高的计算复杂度 |

(1)数据生成假设的不同:这四种方法对数据的生成表示做出了不同的假设。矩阵分解假设样本数据可以用两个或多个低秩矩阵表示,以获得紧凑的特征表示。图学习假设数据可以用图结构建模,其中数据点对应图的节点,不同数据点之间的相似性由边权衡量。子空间聚类假设高维数据分布在多个低维子空间,每个聚类对应一个子空间。张量学习假设数据样本可以用一个低秩核心张量和一组因子矩阵表示,从而建模不同视图间的高阶关系。(2)优化目标的不同:矩阵分解旨在寻找低秩矩阵的最优分解,使得分解后的数据能够最小化与原始数据之间的重构误差。图学习通过构造图拉普拉斯正则化项,优化数据的低维嵌入表示,以保持同一聚类内数据点的紧密性。子空间聚类通过学习一个具有自表达属性的矩阵,使得每个数据点能够尽可能由同一子空间内的其他数据点进行线性重构。张量学习采

用张量分解(如 Tucker 分解、CP 分解),学习数据的低维表示,以最优方式保留多视图数据的结构信息。(3)适用场景的不同:矩阵分解适用于结构化数据(如文本数据),因为它能通过低秩分解提取跨视图共享特征。然而,它在处理非线性数据时表现较弱。子空间聚类适用于高维数据(如视频数据),能够在子空间结构中进行有效的聚类。图学习依赖于图结构建模,适用于结构复杂的数据(如社交网络数据),能够利用图结构学习数据的全局信息。张量学习适用于高阶数据(如多模态数据,包括视频和图像),能够建模不同视图间的复杂交互关系。

然而,启发式 MVC 通常涉及矩阵分解、图建模和张量分解等计算过程,这导致其计算复杂度较高,优化过程较为困难。在大规模数据场景下,这些方法可能面临高计算成本和存储开销的问题,限制了其在大规模数据分析任务中的应用。

3.2 基于判别的多视图聚类

与上小一节讨论的启发式多视图聚类方法相比,基于判别的多视图聚类方法能够更高效地挖掘多视图数据中的互补信息与一致性特征以及处理复杂的高维数据,从而显著提升聚类的判别能力。本节主要介绍两类基于判别的多视图聚类方法:深度嵌入方法和对比学习方法。我们还总结了这两类方法的代表性实验技术,详细信息如表 6 和表 7。

3.2.1 基于深度嵌入的多视图聚类

深度嵌入方法主要包括深度矩阵分解和深度自动编码器两种形式。其中,深度矩阵分解的过程可以用下面式子描述:

$$\begin{aligned} \mathbf{X} &\approx \mathbf{U}_1 \mathbf{V}_1 \\ \mathbf{X} &\approx \mathbf{U}_1 \mathbf{U}_2 \mathbf{V}_2 \\ &\vdots \\ \mathbf{X} &\approx \mathbf{U}_1 \cdots \mathbf{U}_m \mathbf{V}_m \end{aligned} \quad (39)$$

其中, \mathbf{U}_i 是第 i 层基矩阵, \mathbf{V}_i 是第 i 层数据指示矩阵。图 11 展示了深度非负矩阵分解模型的结构。

为了更有效地利用各视图之间的互补信息和一致性信息,Zhao 等人提出一种用于多视图聚类的框架,它采用半非负矩阵分解的方法来逐层学习多视图数据的分层语义,该模型可以描述为

$$\begin{aligned} \min_{\substack{\mathbf{U}_i^v, \mathbf{V}_i^v, \mathbf{V}_m^v \\ \mathbf{V}_m^v, \partial^v}} \sum_{v=1}^m (\partial^v)^{\gamma} (\|\mathbf{X}^v - \mathbf{U}_1^v \mathbf{U}_2^v \cdots \mathbf{U}_j^v \mathbf{V}_j^v\|_F^2) \\ \text{s. t. } \mathbf{V}_i^v \geq 0, \mathbf{V}_m^v \geq 0, \sum_{v=1}^m \partial^v = 1, \partial^v \geq 0 \end{aligned} \quad (40)$$

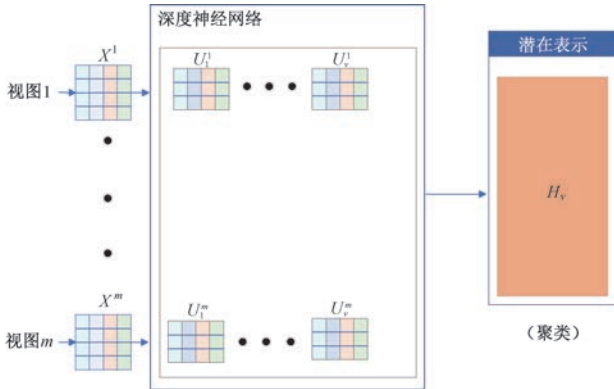


图 11 深度 NMF 的模型框架结构

其中, ∂^v 表示第 v 个视图的权重系数, γ 是用于控制权重分布的参数, U_i^v 是第 v 个视图的第 i 个映射层, j 是隐藏层的数量, V_j 是所有视图的共识表示。通过强制不同视图最后一层的特征表示都为 V_j , 最大限度利用了每个视图的互补信息^[110]。在此基础上, Li 等人提出一种深度图正则化非负矩阵多视图聚类(MVDGNMF)方法, 该方法通过增加约束项 $\sum_{i=1}^m \|V_i - V^*\|_F^2$ 将不同视图的最后一层特征表示推向共识, 该方法还加入图正则项保证了跨视图数据的局部结构^[111]。Gunawardena 开发了一种深度架构, NMF 与 CNMF 结合, 以提取视图间的互补信息和共识信息^[112]。

为改进聚类表示, Trigeorgis 等人提出了一种基于迭代半非负矩阵分解的深度模型, 该模型依赖于三个基本假设, 并根据各层的属性选择方差最小的层进行聚类, 从而增强隐藏表示的聚类能力^[113]。Wang 等人介绍了一种基于深度半 NMF 的多视图聚类算法(MCDS), 通过使用 SGD 算法迭代更新参数, 解决了由于缺乏学习率指导而导致聚类表现不佳的问题^[114]。在此基础上, Wang 等人通过矩阵更新方法构建一种深度网络参数, 并结合多种激活函数开发了一系列算法, 从而显著提升了聚类表示的效果^[115]。

另一方面, 基于深度自动编码器的多视图聚类方法也得到广泛应用, 这类方法有效捕捉跨视图之间的一致性和互补性信息。Du 等人使用自动编码器逐步捕捉每个视图的非线性结构, 将局部不变性与视图间的一致性和互补性信息相结合^[116]。Wang 等人提出了一种具有三重约束的协同学习框架(TCMGC), 该框架通过三元运算强制执行自聚类约束, 并实现视图间的独立性和互补性^[117]。Fang 等人通过堆叠多个显式可微分块网络, 学习具有解

释性的共识表示^[118]。为进一步提高聚类判别性能, Xu 将深度自动编码器与协同训练相结合, 以提升聚类的判别能力^[119]。此外, Xu 等人还采用了一种自监督方法, 通过利用全局视图的判别信息创建统一目标分布, 以学习更具代表性的特征并确保跨视图预测的一致性^[120]。

自编码器不同的内部结构可以处理不同的问题, 比如, Xu 引入了一种结合全连接层的深度自动编码器, 以学习特定视图的多层次特征, 并在编码器和解码器之间添加多个自表达层, 从而提取多层次子空间表征^[121]。Xie 等人采用了多种类型的自动编码器, 包括堆叠自动编码器(SAE)、卷积自动编码器(CAE)和卷积变分自动编码器(Conv-VAE), 以有效捕捉多视图特征^[122]。Wang 等人构建了一个由多个模块组成的神经网络, 用于学习稀疏正则性^[123]。Huang 等人提出了具有部分视图对齐(PVC)的方法, 结合可微分学习和表征学习模块, 以解决多视图部分视图配准问题^[124]。Zhou 等人则提出一种端到端的多视图对抗网络(EAMC), 通过对抗学习和注意力机制, 分别对潜在特征分布和模态量化进行配准^[125]。Qin 等人通过潜在表征的指导, 学习多个视图中配对数据的相关性^[126]。

最近一些基于注意力机制和图卷积网络(GCN)的方法来应用多视图聚类的方法也被提出, Wang 等人提出一种具有一致性多视图嵌入(GMGEC)的模型, 该模型利用基于注意力机制的图融合编码器对不同视图的互补信息进行编码, 然后引入多视图互信息最大模块来指导学习的共同表示, 保持每个视图相邻特征的相似性, 最后通过图融合网络(GFN)得到了具有共识一致性的图信息^[127]。由于不同视图对表示学习的贡献不同, Lu 等人通过引入注意力机制来推导每个视图的动态权重, 然后根据不同视图的权重进行图融合操作, 从而得到一个更具判别性的聚类结果^[128]。Xiao 等人利用特定视图下 GCN 生成每个视图下的一个隐藏表示, 然后采用自适应融合方法分别在特征级和结构级上进行融合, 最后将融合的矩阵作为输入传送到跨视图模块中, 有效探索了跨视图共识信息^[129]。一些方法用单视图的节点属性来描述多视图, 这会造成多视图节点表示学习不足的问题, 因此, Xia 等人首先利用欧拉变换来增强视图节点属性, 作为非欧几里得结构数据的新视图描述符, 然后利用 GCN 来获取不同视图的潜在特征 H^v , 通过自表达矩阵获取具有跨视图一致性的自表达系数矩阵 Z , 然后

对其施加 l_{2-1} 范数测量的块对角约束,很好地探索了集群结构^[130]。此外,Ou 等人提出一种基于 transformer 的交互式共享转换器,这个模块利用了两层 transformer 来描述视图和标签之间的相互作用,有效利用了视图间的互补信息,以及标签的潜力^[131]。

基于深度嵌入的多视图聚类方法的优缺点总结如表 6 所示。

表 6 基于深度嵌入式 MVC 方法

| 方法 | 优点 | 缺点 |
|-------------------------------|-------------------------|-----------------------|
| DeepSemi-NMF ^[113] | 聚类结果具有良好的可解释性 | 计算复杂度更高 |
| MVDGNMF ^[111] | 学习多重视图子结构 | 不适用于信号处理 |
| DMJC ^[122] | 充分挖掘多视图互补信息 | 不适合图像数据集以外的数据集 |
| PVC ^[124] | 处理部分不对齐视图的能力 | 无法处理完全未对齐或丢失数据的视图 |
| EAMC ^[125] | 生成简洁的聚类结构 | 不适合大规模数据集 |
| CMGEC ^[127] | 增强多视图的共识表示 | 不适合高维数据 |
| MVC-MAE ^[116] | 适用于大规模数据集 | 不考虑其他层对聚类的影响 |
| TCMGC ^[117] | 考虑每种观点的独立信息 | 计算效率低下 |
| DEMVC ^[119] | 降低大规模数据集的计算复杂度 | 缺乏访问高维复杂数据结构的能力 |
| DSRL ^[123] | 可以解决优化问题 | 不学习多重正则 |
| MVGC ^[130] | 模型极强的判别能力 | 不适合动态图形 |
| MCDS ^[114] | 适用于大规模数据集 | 坚固性差 |
| DNRSF ^[115] | 卓越的聚类特性 | 参数存在顺序更新问题 |
| SDMVC ^[120] | 克服低质量景观对集群的影响 | 缺乏稳健性 |
| ESCC ^[126] | 探索观点之间的高阶相关性 | 不适合大规模数据集 |
| DFP-GNN ^[129] | 卓越的模型辨别力 | 算法收敛缓慢 |
| DBO-Net ^[118] | 可处理复杂的多视图数据 | 网络结构缺乏可解释性 |
| DCCNMF ^[113] | 同时学习补充信息和共识信息 | 聚类效应对参数 |
| CRMSRL ^[122] | 具有更好的稳健性 | 未考虑不同视角对聚类效应的重要性 |
| AMVDSN ^[129] | 考虑不同视图权重对聚类的影响,具有一定的判别性 | 该方法不适用于一些缺失数据集和不完整数据集 |

3.2.2 基于对比学习的多视图聚类

深度嵌入的方法虽然能有效捕捉非线性结构的

数据,但是它不适用于大规模数据,模型难以拓展,对比学习(Contrastive Learning)是一种在自监督学习中被广泛采用的技术,它以小批量和负采样的训练方式可以高效地拓展到大规模数据集。其核心原理是通过数据增强技术生成同一数据样本的多个增强版本,这些增强版本被视为正样本对,而不同数据样本的表示则被视为负样本对。目标是在同一嵌入空间中最大化正样本对的相似性,同时最小化负样本对的相似性。多视图聚类中对比学习过程如图 12 所示。

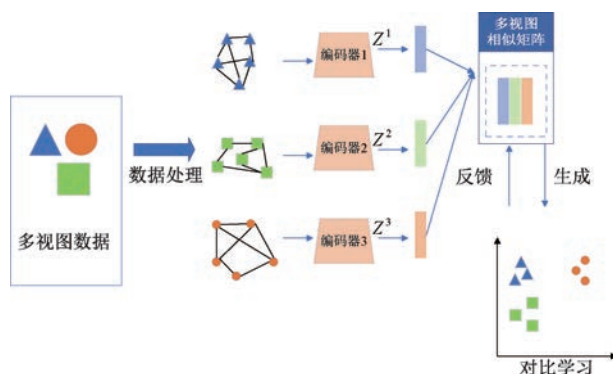


图 12 基于对比学习的多视图聚类模型框架结构示意图

为更好地捕捉跨视图的一致性和互补信息,Wang 等人将对比学习和注意力机制结合提出一种多视图聚类(CAGL)方法,通过对比学习最大化视图对之间的相似性,确保多视图数据的一致性^[132]。Liu 等人利用跨视图对比学习获取实例级的非冗余一致表示^[133]。Wang 等人结合对比重构损失,实现了重构图与原始图在样本级的近似,从而促进网络共识图的学习^[134]。Zhang 等人通过对比学习生成稳健的样本对,从不同视图中提取互补信息^[135]。Yang 等人在视图的节点和簇级别均采用对比学习,以确保不同视图间的信息一致性^[136]。

大多数深度聚类方法按视图聚合门控方式从多个视图中学习共识表示或视图特定表示,但它们都忽略了所有样本的结构关系,为了解决这个问题,Yan 等人提出一种融合跨视图全局特征(GCFAgg)的方法,该方法利用以结构为导向的对比学习模块将共识表示和特定视图表示对齐,这使得来自具有高结构关系的不同样本的视图特定表示相似,GCFAgg 利用余弦距离计算第 i 个样本的跨视图共识表示 \mathbf{H}_i 和特定视图表示 \mathbf{H}_i^v 之间的相似性,即

$$C(\mathbf{H}_i, \mathbf{H}_i^v) = \frac{\mathbf{H}_i^T \mathbf{H}_i^v}{\|\mathbf{H}_i\| \|\mathbf{H}_i^v\|} \quad (41)$$

然后以结构为引导的多视图对比学习可以描

述为

$$L_c = -\frac{1}{2N} \sum_{i=1}^n \sum_{v=1}^m \log \frac{C(\mathbf{H}_i, \mathbf{H}_i^v)/\tau}{\sum_{j=1}^N e^{(1-s_{ij})C(\mathbf{H}_j, \mathbf{H}_j^v)/\tau} - e^{1/\tau}} \quad (42)$$

其中, s_{ij} 表示样本 i 和样本 j 之间的结构关系, s_{ij} 越小, $C(\mathbf{H}_i, \mathbf{H}_j^v)$ 就越大, 当样本 i 和样本 j 之间关系越小, 对应的表示就越不一致, 否则, 它们的关系就越一致^[137]。在此基础上, Wu 等人提出一种具有自加权的深度多视图聚类 (SCMVC) 方法。该方法通过最大化视图一致性表示和全局表示之间的一致性来实现多视图对比融合, 此外, 该方法通过测量成对表示之间的差异, 设计一种自加权方法, 它可以自适应加强特征融合有用视图, 削弱不可靠视图, 减轻了视图表示退化问题^[138]。

此前, 深度多视图聚类方法通常设计非共享编码器来探索多视图中的集群信息, 但很难为单个视图制定编码器, 这容易增加信息丢失。为此, Luo 等人提出一种数据级融合方法来融合多视图信息, 该方法用融合数据替代所有视图, 避免了为不同视图指定网络^[139]。Zou 等人利用自动编码器独特的重建机制迫使模型专注于学习对重建输入数据至关重要的特征, 从而减少了冗余特征, 降低噪声的干扰^[140]。

为解决多视图聚类中的语义问题, Xu 提出了一种多级特征学习框架, 用于对比多视图聚类。该框架结合对比学习和语义标签, 通过实现较低特征层的重构与较高特征层的一致性, 有效提升了聚类性能^[141]。Shu 在特征和聚类层面引入了结构引导的对比学习策略, 充分捕捉不同视图中高层次特征与语义特征的相似聚类结构^[142]。Du 采用复合注意结构学习潜在表征, 不仅保留了视图内的重要特征, 还将共同语义纳入聚类分配^[143]。

此外, Bao 通过在嵌入空间中引入特征选择与对比学习, 增强了自表达系数矩阵对数据点关系的捕捉能力, 同时在原始空间中将对比学习作为修正项, 以保持数据点的空间分布^[144]。Pan 等人提出了一种对比正则化方法, 用以提升图的质量, 从而获得更优的聚类性能^[145]。Trosten 等人 and Lin 等人分别给出利用对比学习解决不同视图表示对齐的方法^[146-147]。Deng 等人提出一种增强对比 (SACC) 方法, 利用对比学习强化深度聚类表现^[148]。Du 等人借助对比学习训练 GCN, 从而整合拓扑和邻近节点的特征^[149]。Fang 等人设计了一种基于上下文感知

的对比图学习 (CA-CGL) 框架, 通过对比学习解决图像数据集中的一致性问题^[150]。

基于对比学习的多视图聚类 (MVC) 方法的优缺点总结如表 7 所示。

表 7 基于对比学习的 MVC 方法

| 方法 | 优点 | 缺点 |
|---------------------------|-------------------|--------------------------------------|
| MCGC ^[145] | 更好的鲁棒性和鉴别能力 | 不适合大规模数据 |
| SiMVC ^[146] | 考虑到表征排列对聚类的重要性 | 简单的排列分布会影响模型学习良好聚类能力 |
| CoMVC ^[146] | 可选对齐有助于模型学习更好的聚类 | 抗噪声能力较弱 |
| CAGL ^[132] | 高度解释性模型, 具有出色判别能力 | 不适用于大规模数据 |
| MC-TSL ^[133] | 获得了非冗余一致的表示 | 聚类效应对参数很敏感 |
| SACC ^[148] | 有更好的辨别能力 | 不适用于某些数据集 |
| NMVC-GCN ^[149] | 处理大规模数据集的能力 | 参数选择在很大程度上取决于数据类型 |
| CA-CGL ^[150] | 更强的鲁棒性 | 无法解决样本外问题, 对不平衡数据集敏感 |
| SGFCC ^[142] | 聚类结构紧凑, 辨别能力强 | 不适合高维数据集 |
| CAMVC ^[143] | 保留潜在的共同语义 | 对脏污样本的鲁棒性较低 |
| MCSC ^[144] | 卓越的模型识别能力 | 计算复杂度更高 |
| GCFAgg ^[137] | 对齐了不同视图的表示 | 忽略了真实世界的数据集可能是不完整或缺失的 |
| SCMVC ^[138] | 解决了视图退化问题 | 虽然考虑了不同视图权重对聚类的影响, 但是并没有获得聚类友好型的聚类表示 |
| LUCE-CMC ^[139] | 减少了视图冗余特征对聚类的影响 | 计算复杂度比较高 |

3.3 基于生成的多视图聚类

上一节中介绍的深度嵌入式和对比学习多视图聚类尽管可以有效提高聚类的判别能力, 但是在处理包含缺失数据的样本集时聚类效果不佳。生成式多视图聚类将生成式建模与多视图聚类相结合, 通过重新生成缺失数据来处理不完整视图数据和学习复杂的数据分布。在本节中, 我们将重点介绍两类主要方法: 基于变分自动编码器 (VAE) 的多视图聚类方法和基于生成对抗网络 (GAN) 的多视图聚类方法。

3.3.1 基于变分自动编码器的多视图聚类

变分自动编码器 (VAE) 将变分推理与自动编

码器相结合,基于 VAE 的多视图聚类方法通过分析样本的潜在概率分布,有效地捕捉数据的内在结构。各种编码器结构的如图 13 所示。

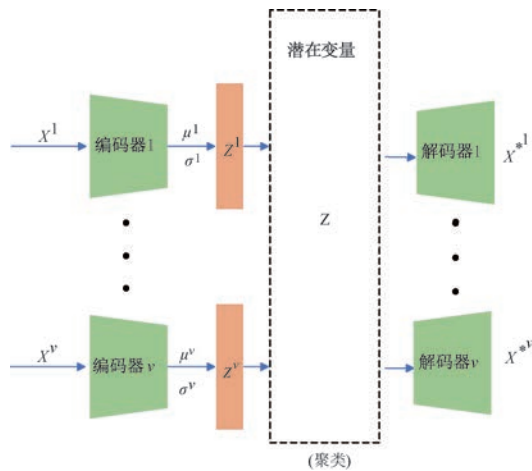


图 13 基于 VAE 的多视图聚类模型结构

原始的 VAE 设计并不适用于连续任务,Yin 在 VAE 框架内提出了共享潜表征机制,使多视图权重和深度自动编码器网络能够在统一框架内共同学习,从而促进最终的聚类分配^[151]。为了解决基于 VAE 的多视图聚类中,不同视图之间共享信息可能引发的干扰和依赖问题,Xu 等人提出了多重变分自编码器(Multi-VAE)方法,该方法通过学习视图的解耦表示,有效提升了聚类性能^[152]。

为更好地捕捉跨视图的一致性和互补性信息,Xu 等人利用 VAE 结合特定视图和通用表征的解耦信息,显著增强了聚类的一致性与互补性^[153]。Cui 等人通过 VAE 生成一致性信息,减少跨视图的冗余信息^[154]。

此外,Huang 等人将信息瓶颈理论扩展至无监督多视图学习场景,基于互信息定义了多视图表示的三个关键要求^[155]。Bai 等人提出了 HVAE 方法,通过分层 VAE 将文档的局部主题信息与全局主题信息关联起来,从而实现更精细的特征表达^[156]。

基于 VAE 的多视图聚类(MVC)方法的优缺点总结如表 8 所示。

3.3.2 基于生成式对抗网络的多视图聚类

生成式对抗网络(GAN)通过生成器和判别器之间的对抗训练,生成接近真实数据分布的样本。在应对不完整视图数据的挑战时,GAN 表现出极高的有效性。图 14 展示了生成式对抗网络在多视图场景中的应用结构。

为了解决缺失数据的问题,Wang 等人提出了

表 8 基于 VAE 的 MVC 方法

| 方法 | 优点 | 缺点 |
|----------------------------|---------------------|--------------------------|
| DMVCVAE ^[151] | 捕捉跨视图的复杂关系 | 不适合大规模数据集 |
| Multi-VAE ^[152] | 计算复杂度低,鲁棒性好 | 难以平衡视图共同变量和视图特殊变量以优化聚类效果 |
| UNTIE ^[153] | 生成样本以管理缺失数据 | 聚类对参数很敏感 |
| SCMVC ^[154] | 确保视图之间的一致性 | 视图之间存在冗余信息 |
| SUMVC ^[154] | 通过信息瓶颈原理减少视图之间的冗余信息 | 聚类对参数很敏感 |
| IMC ^[155] | 生成样本以克服数据缺失问题 | 模型仅限于参数的选择 |
| HAVE ^[156] | 具有更好的稳健性 | 每个视图的收敛问题不一致 |

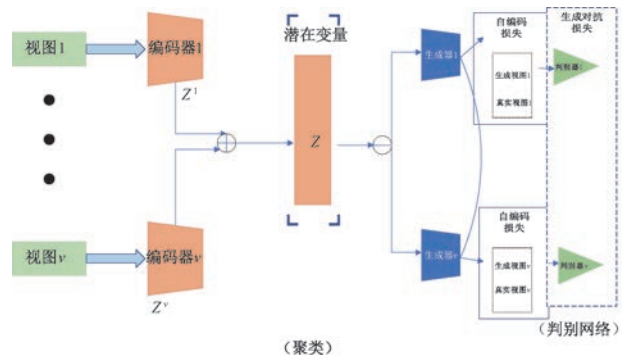


图 14 基于 GAN 的多视图聚类模型的结构框架

一种针对缺失数据的部分多视图聚类方法。该方法首先获取视图编码表征,然后通过对抗网络生成缺失数据^[157],Wang 等人还提出一种自适应融合和周期一致性的生成部分多视图聚类(GP-MVC)方法,利用 GAN 生成的公共子空间来填补不完整视图,从而提高聚类性能^[158]。Xu 等人则提出了对抗性 IMVC 框架(AIMC),通过挖掘多视图数据的共同潜在表示,实现原始数据的重建并推断缺失数据^[159]。

为获取视图间的一致性信息,Wang 等人提出了一种特定视图的生成式对抗网络(GAN),通过其他视图的共享表征生成目标视图的缺失数据,从而进一步探索视图间的一致性^[160]。Wang 等人将对比学习与生成式对抗网络相结合,通过挖掘不同子空间特征,共同学习数据恢复和一致性表征^[161]。他们还提出了一种对抗性多视图聚类(AMVC)网络,从所有视图中提取潜在描述,以确保多视图自动编码器中信息的一致性^[162]。

Li 提出了一种深度对抗多视图聚类(DAMC)方法,该方法通过对抗训练引入判别器,有效指导编

码器网络的训练,捕捉各视图的数据分布,并进一步实现潜在空间的解耦^[163]。

基于 GAN 的多视图聚类(MVC)方法的优缺点总结如表 9 所示。

表 9 基于 GAN 的 MVC 方法

| 方法 | 优点 | 缺点 |
|---------------------------------|-----------------------------|---------------------|
| Consistent-GAN ^[157] | 对不完整的视图数据进行插值,可进一步研究一致的公共结构 | 不适合大规模数据集 |
| GP-MVC ^[158] | 生成缺失数据,研究具有互补不完整视图的公共结构。 | 不适用于大规模数据 |
| AMVC ^[162] | 特征更具鉴别力 | 大规模数据集性能不佳 |
| MRL_CAL ^[161] | 可同时学习一致的表征和标签 | 在表征学习中没有明确利用特征关系和结构 |
| AAIMC ^[159] | 生成样本以获取缺失数据 | 聚类性能受模型训练和缺失数据影响 |
| GAIMC ^[159] | 直接搜索多视图互补信息,推断缺失数据信息 | 不适用于大规模数据 |
| DAMC ^[163] | 去耦电位空间 | 大规模数据集性能不佳 |

4 数据集与实验对比

4.1 多视图数据集

多视图聚类的数据集主要分为四类:文本数据、图像数据、图像—文本数据和视频数据集。

文本数据集主要由最新数据集(如 3sources、BBC Sport)、多语言文档数据集(如 Reuters)、引文数据集(如 CiteSeer)、网页数据集(如 WebKB)和一些大规模数据集(如 20NewsgroupS、Multi30k)组成。有关文本数据集的详细信息,请参见表 10。

表 10 文本数据集基准

| 数据集 | 实例 | 类别 | 视图 | 特 征 |
|----------------------------|-------|----|----|--|
| 3sources ^[164] | 169 | 6 | 3 | 3sources 数据集由三个不同来源的数据组成,因此适合多模态学习每个样本都可以有多个标签,因此非常适合多标签分类任务 |
| BBC Sport ^[165] | 544 | 5 | 2 | 该数据集包含丰富的文本,非常适合文章分类和信息检索等任务 |
| Reuters ^[166] | 18758 | 6 | 5 | 该数据集样本可以有一个或多个标签,因此适用于文本分类任务 |
| CiteSeer ^[166] | 3327 | 6 | 3 | 该数据集通常被视为单视角数据集,但与其他数据源结合后,可扩展为多视图方案 |
| WebKB ^[164] | 203 | 4 | 3 | 该数据集广泛应用于信息检索、网络分类 |
| 20Newsgroups | 20000 | 20 | 2 | 这个经典的文本分类数据集由 20 个新闻组的文章组成,通常将文章的内容和类别标作为不同的视图来构建 |
| Multi30k | 30000 | 5 | 2 | 该数据集包含来自不同语言的文本对(如英语和德语),用于多语言学习任务 |

图像数据集主要包括面部图像数据集(如 ORL、Yale、Notting-Hill)、手写数字数据集(如 HandWritten、UCI、MNIST)、物体图像数据集(如

MSRC-V1)、植物图像数据集(如 Flower17)以及自然场景数据集(如 Scene-15)。有关图像数据集的详细信息,请参见表 11。

表 11 基准图像数据集

| 数据集 | 实例 | 类别 | 视图 | 特 征 |
|-------------------------------|------|----|----|--|
| ORL ^[54] | 400 | 40 | 3 | 每张图像大小为 92□112 像素,包含表情、姿势和照明条件的变化 |
| Yale ^[167] | 165 | 15 | 2 | 每张图像的大小为 320□243 像素,图像中的表情、光照条件和姿势都有变化 |
| UCI ^[52] | 2000 | 10 | 3 | UCI 数据集涵盖多个领域,包含不同类型的特征,适合进行多样化的机器学习 |
| HandWritten ^[90] | 345 | 10 | 4 | 手写数据集图像大小为 28□28 像素,每幅图像都有一个相应的标签 |
| MSRC-V1 ^[168] | 210 | 7 | 5 | MSRC-V1 数据集包含不同分辨率和丰富色彩的图像,适合研究不同物体的特征提取和多视图聚类问题 |
| Notting-Hill ^[169] | 4660 | 5 | 2 | 提供不同视角下的行人行为图,数据集也适用于复杂场景 |
| Flower17 ^[64] | 1360 | 17 | 7 | 每幅花卉图像可能有不同的光线、角度和背景,每幅图像也有特定的花卉类别 |
| MNIST ^[170] | 4000 | 4 | 3 | 每幅图像的大小为 28□28 像素,适合输入图像处理和深度学习模型 |
| Scene-15 ^[76] | 4485 | 15 | 3 | 该数据集包含不同的场景,适用于复杂场景中的多视图聚类 |

图像—文本数据集包括果蝇胚胎数据集 BDGP、NBA-NASCAR sports (NNSpt) 数据集、Pascal VOC 数据集以及对象数据集(NUS-WIDE-

C5)。详细信息如表 12 所示。

视频数据集包括摄像机运动数据集(YouTube)、哥伦比亚消费者视频数据集(CCV)、行人视

频数据集 (Lab) 和人体运动序列数据集 (CMUMo-bo)。详细的多视图音频数据集信息见表 13。

表 12 基准图像—文本数据集

| 数据集 | 实例 | 类别 | 视图 | 特 征 |
|----------------------------|------|----|----|--------------------------------------|
| BDGP ^[5] | 5000 | 5 | 2 | 用于研究基因表达模式与胚胎发育阶段之间的关系, 适合多视图生物信息学学习 |
| NNSpt ^[5] | 840 | 2 | 2 | 多样化的图像数据和详细的文本数据适合图像和文本跨模态学习 |
| VOC ^[5] | 5649 | 20 | 2 | VOC 通常不被认为是多视图集, 但可以通过数据预处理生成不同的视图 |
| NUS-WIDE-C5 ^[5] | 4000 | 5 | 2 | 它具有丰富的多模态信息, 适合多视图学习 |

表 13 基准视频数据集

| 数据集 | 实例 | 类别 | 视图 | 特 征 |
|--------------------------|------|----|----|---------------------------------|
| YouTube ^[168] | 1596 | 11 | 3 | 包含 1596 个视频序列, 具有丰富的多模态内容和大规模数据 |
| CCV ^[5] | 6773 | 20 | 3 | 可用于研究如何在不同视图之间进行有效的特征融合和分类 |
| Lab ^[5] | 16 | 4 | 4 | 该数据集主要由面部图像组成, 适用于面部识别 |
| CMUMobo ^[5] | 96 | 24 | 4 | 包含 96 个动作序列, 可从多个视图捕捉动作 |

4.2 评估指标

内部指标不依赖于外部的真实标签, 一些著名的方法包括:

(1) 聚类内平方和 (WSS): 它测量同一聚类内各点之间的距离总和。WSS 越小, 表明聚类越好, 各点之间的距离越近。

(2) 轮廓系数: 它衡量聚类内部的紧密程度和聚类之间的分离程度。轮廓系数接近 1 表示集群分离良好、紧凑。

(3) 戴维斯—博尔丁指数 (DBI): 衡量簇内紧密度与簇间分离度的比率。DBI 越小, 表示聚类效果越好。

(4) Calinski-Harabasz 指数 (CH 指数): 平衡簇内紧密度和簇间分离度。CH 指数越大, 表示聚类效果越好。

(5) 邓恩 (dunn) 指数: 最小化聚类内最大距离, 最大化聚类间最小距离。邓恩指数越大, 表示聚类效果越好。

评估聚类性能的外部指标主要包括: 归一化互信息 (NMI)、准确率 (ACC)、调整兰德系数 (ARI)、纯度 (PUR) 和 F1 分数。本文主要将 NMI、ACC 和 PUR 作为评价聚类性能的外部指标。NMI 的计算公式为

$$NMI(U, V) = \frac{2 \times I(U, V)}{H(U) + H(V)} \quad (43)$$

其中, $I(U, V)$ 表示聚类结点 U 和 V 之间的互信息, $H(\cdot)$ 表示聚类结果的熵。ACC 通过计算聚类结果中的最优标签匹配来评价聚类效果, ACC 的计算公式如下:

$$ACC = \frac{1}{n} \sum_{i=1}^n \delta(y_i, \text{map}(c_i)) \quad (44)$$

其中, y_i 是第 i 个样本点对应的真实标签, c_i 是第 i

数据点的聚类标签, $\text{map}(c_i)$ 是将聚类标签 c_i 映射到真实标签的最优映射函数。 $\delta(x, y)$ 是一个指标函数, 当 $x = y$ 时, $\delta(x, y) = 1$, 否则 $\delta(x, y) = 0$ 。PUR 通过计算每个聚类中比例最大的真实标签类别的比例来评估聚类效果, PUR 的计算公式为

$$PUR = \frac{1}{n} \sum_{i=1}^k \max_j |c_i \cap L_j| \quad (45)$$

其中, k 是聚类个数, L_j 是真实标签为 j 的数据点集合, $\max_j |c_i \cap L_j|$ 表示聚类 c_i 中最多数据点所属的真实标签类别个数。

4.3 对比算法

本节主要引用和摘录了其他已发表论文中常用的 22 种多视图聚类 (MVC) 方法在 Reuters、MNIST、Handwritten、Yale、BBC Sport、MSRC-V1、ORL、UCI 和 Scene-15 等数据集上的实验结果, 以及 7 种不完全多视图聚类 (IMVC) 方法在 BBC Sport、3Sources 和 MSRC-V1 数据集上的运行结果。需要特别指出的是: 由于不同方法的实验设计存在差异, 包括数据预处理方式、特征提取维度、视图选择及参数设置等因素, 因此表中所列结果可能不可避免地存在一定偏差。为尽量减少理解上的歧义, 我们对不同 MVC 和 IMVC 方法的部分关键参数设置进行了归纳总结, 如表 14 和表 15 所示; 同时, 相关 MVC 方法在不同数据集上的实验结果汇总于表 16, 表 17 和表 18; 相关 IMVC 方法在不同数据集上的实验结果汇总于表 19, 表 20 和表 21。其中符号“—”表示相关文献中未公开的数据, 粗体用于突出标识在对应数据集上表现最佳的方法及其结果。需要说明的是, 以上数据和结果旨在展示不同方法在多视图聚类任务中的整体性能趋势, 而非在完全一致条件下的直接对比。

其中,常见的 22 种 MVC 方法包括:

(1) DIMSC:一种多视图子空间聚类方法利用希尔伯特-施密特独立性准则来增强子空间表征的学习^[54]。

(2) MLAN:能够同时进行聚类、半监督分类和局部结构学习的多视图学习模型^[89]。

(3) RMSC:一种稳健的多视图光谱聚类方法,通过低秩和稀疏分解减少噪声影响^[73]。

(4) LRR:将样本聚类到各自的子空间,同时去除异常值^[68]。

(5) ECMSC:一种利用位置感知排他性和一致性术语来捕捉多视图互补性和一致性的聚类方法^[56]。

(6) ETLMSC:基于马尔可夫链的多视图光谱聚类方法^[102]。

(7) ULTLSE:结合低阶张量学习和谱嵌入优化的多视图子空间聚类统一框架^[105]。

(8) CELT:一种双嵌入式多视图聚类方法,可从原始数据中学习视图内图形和低维空间^[76]。

(9) DAMC:DAMC 通过学习多视图数据中蕴含的内在结构,揭示了聚类的复杂结构^[163]。

(10) Multi-NMF:通过搜索因式分解提供多视图兼容聚类解决方案^[28]。

(11) SNMF:一种带符号的混合方法,允许混合数据矩阵,从而扩大了 NMF 的适用范围^[26]。

(12) DNSRF:基于深度网络的 SNMF 框架,用于解决 SNMF 在没有学习率引导的情况下导致的表示能力差的问题^[115]。

(13) FRSMNMF:统一约束多视图的配准和聚类间判别信息,并设计融合正则化^[39]。

(14) CDTMSC:这是一种在张量多视图子空间框架中同时利用多视图一致性和多样性信息的方法^[62]。

(15) SMCMB:SMCMB 为每个视图学习了多个不同的二叉图,显著提高了聚类性能,并保持了较高的竞争效率^[83]。

(16) MRL_CAL:采用自动编码器获取实例的低级视图特定表示,通过对抗学习估计原始不完整数据的分布,从而恢复数据^[161]。

(17) MVCNMF:通过引入辅助矩阵并增加正则化和稀疏性约束来实现信息互补的方法^[38]。

(18) FTMCM:它从全局角度描述张量中条目之间的关系,并利用低秩约束探索张量相似性矩阵的局部结构^[97]。

(19) MCDS:一种基于深度 SNMF 的多视图聚类算法,它以元素方法计算,可对大规模数据集进行有效聚类^[114]。

(20) Multi-VAE:它是一种聚类方法,能有效利用视图的共同表示法和视图特定表示法中的共同离散信息^[152]。

(21) DMVCVAE:它通过学习服从高斯混合分布的共享信息来生成潜在表征^[151]。

(22) AMVC:AMVC 直接搜索多视图综合表示法,并从同一集群中的实例中推断出缺失视图的信息^[159]。

为了模拟现实中的不完整数据,我们对缺失数据集中的数据进行随机化处理,概率为 p , 其中 $p \in \{0.1, 0.3, 0.5\}$ 代表数据缺失率。IMVC 方法的典型方法包括:

(1) IMG 利用一种新的图拉普拉斯正则项,将缺失样本和完整样本联系起来,通过这种方式使得缺失数据样本集也可以实施全局约束,从而将完整多视图聚类方法运用到不完整多视图数据集上^[171]。

(2) DAIMC 通过 semi-NMF 方法对数据样本进行重建,然后根据不同实例重建误差分配不同权重,针对缺失数据样本,将其权重置为 0,从而降低不完整数据对聚类影响^[172]。

(3) GIMC-FLS 将多个视图通过矩阵分解(MF)方法得到多个正交基矩阵,然后通过自加权正交基矩阵方法重建缺失数据^[173]。

(4) IMSC_AGL 通过用零元素来填充缺失样本数据,将不完整多视图转为完整多视图然后在完整多视图数据基础上进行聚类^[174]。

(5) UEAF 首先将原始数据中的缺失数据样本置为 0,然后利用相应视图的误差矩阵和权重矩阵对缺失数据进行二次填充,从而解决不完整数据问题,并达到对齐多视图特征效果^[175]。

(6) AGC_IMC:利用不同视图之间的互补性,对于含有缺失数据样本的视图,通过找到该样本对应于其他视图完整数据,然后通过其他视图数据完整信息推断出该样本的缺失数据信息^[176]。

(7) ARDGL:将缺失数据特征填充为其他数据特征的平均值,缓解了不完整数据对聚类影响^[177]。

4.4 实验评估

每个数据集的最佳聚类结果以粗体标出,表格中的"—"表示原始文章中没有提供该方法的聚类结果。通过对上表结果的分析,我们总结出以下结论:

表 14 MVC 方法的一些主要参数设置

| 对比实验 | 主要参数设置 |
|-----------|---|
| DIMSC | 从图像提取三种特征:强度,LBP,Gabor。选取正则项参数 λ_v 为0.01和0.03 |
| MLAN | 利用最近邻(k-NN)构造亲和图时 $k=9$,以确保样本局部关系。实验中标记样本数量占总样本数量的比例为: $\{0.1,0.2,0.3,0.4\}$,正则项参数 $p=1$ |
| RMSC | 正则项参数 $\lambda=0.0005$ |
| LRR | 惩罚参数 $\mu=10^{-6}$,平衡参数 $\lambda=4$ |
| ECMSC | 排他性项参数选择范围 $\{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}$,一致性选项参数选择为 $\{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}$,它们分别用于控制跨视图的互补性和一致性 |
| ETLMSC | 正则项的平衡参数为0.0008和0.01,相似度判别性参数 σ 和每个视图所有特征的欧氏距离(AED)保持相同 |
| ULTLSE | 正则项参数 λ 为0.01和2 |
| CELT | 软阈值嵌入项和低维嵌入项之间的权衡参数 λ 为 $\{0.0,0.2,0.4,0.6,0.8,1.0\}$,低秩约束的正则项系数 ρ 和控制变量 G 与 S 相似性程度参数 μ 的选择为 $\{0,0.01,0.1,1.0,10\}$ |
| DAMC | 学习率为0.001,迭代次数为30 确保模型收敛 |
| Multi-NMF | 正则项约束参数 λ_v 为0.01 |
| SNMF | 通过计算图 G 每列的平均值,测量实验图中的稀疏性,将所有低于平均值0.001设置为0 |
| DNSRF | 学习率为0.001,正则项参数 $\partial=\beta=0.01$ |
| FRSMNMF | 正则项参数 ∂,β 的取值范围分别为 $[1,e^5],[e^{-1},e^{-5}]$,标记样本数量占总样本的比例为 $\{0.1,0.2,0.3\}$ |
| CDTMSC | 多样性平衡因子 $\lambda_d=\{0.001,0.005,1.0,10\}$,误差项平衡因子 $\lambda_e=\{0.001,0.005,0.5,1.0\}$, Schatten-p 范数取值为 $\{0.2,0.4,0.8,0.9,1.0\}$ |
| SMCMB | 每个视图构造的二部图数量为5,每个视图的锚点数量为100,超参数 β 的值为0.0001 |
| MRL-CAL | 学习率为0.001,batch-size为256,最大迭代次数500,正则项参数 $\partial=\beta=0.01$ |
| MVCNMF | 最大迭代次数为200,超参数 λ_v 为0.01, λ_{vs} 为1或10 |
| FTMCM | 正则项平衡参数 λ_1,λ_2 取值为 $\{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}$,正则项参数 λ 取值为 $\{5,10,15,20,25,30,35,40,45,50\}$ |
| MCDS | 学习率为0.0005,正则项参数 $\partial=\beta=0.001$ |
| Multi-VAE | 学习率为0.0005,温度参数 τ 为0.67,最大迭代次数为500,确保模型收敛,平衡参数 β 为30 |
| DMVCVAE | 学习率为0.0001,优化器为Adam |
| AMVX | 学习率为0.001,最大迭代次数为30,优化器为Adam |

表 15 IMVC 方法的一些主要参数设置

| 对比实验 | 主要参数设置 |
|----------|--|
| IMG | 超参数 λ,β,γ 分别设置为 $1e-2,1e2,1e2$ |
| DAIMSC | 超参数 ∂ 设置为 $1e1$, β 参数选择范围为 $\{1e-1,1e0,1e1\}$ |
| GIMC-FLS | 平衡参数 λ_1 可选择范围 $\{0.001,0.01,0.1,1.0,10\}$,平衡参数 λ_2 可选择范围为 $\{1.0,10,100,1000,10000\}$,超参数 r 选择范围为 $[2,9]$ |
| IMSC_AGL | 超参数 $\lambda_1,\lambda_2,\lambda_3$ 分别设置为 $10^{-5},10^{-4},10^{-3}$ |
| UEAF | 平滑项参数 $r=3$, k -近邻构造亲和图 k 值为7,惩罚参数 $\lambda_1,\lambda_2,\lambda_3$ 的选择范围分别为: $[10^2,10^5],[10^{-4},10^5],[10^{-3},10^5]$ |
| AGC_IMC | 超参数 λ_1,λ_2 的选择范围为 $\{0.1,1.0,10,100\}$ |
| ARDGL | 由 k -近邻构造亲和图的 k 值为15,超参数为 $\lambda=10^{-5}$ |

表 16 MVC 方法在 ORL、UCI 和 MSRC-V1 数据集上的性能比较

| 衡量标准 | ORL | | | UCI | | | MSRC-V1 | | |
|---------------------------|--------------|--------------|--------------|-------|-------|-------|---------|-------|-------|
| | NMI | ACC | PUR | NMI | ACC | PUR | NMI | ACC | PUR |
| DIMSC ^[54] | 0.941 | 0.838 | 0.870 | 0.772 | 0.703 | 0.652 | 0.508 | 0.548 | 0.604 |
| MLAN ^[89] | 0.831 | 0.685 | 0.735 | 0.910 | 0.874 | 0.847 | 0.763 | 0.724 | 0.791 |
| RMSC ^[73] | 0.848 | 0.702 | 0.743 | 0.822 | 0.915 | 0.789 | 0.566 | 0.650 | 0.671 |
| LRR ^[68] | 0.930 | 0.830 | 0.863 | 0.768 | 0.871 | 0.736 | 0.556 | 0.679 | 0.681 |
| ECMSC ^[56] | 0.910 | 0.765 | 0.810 | 0.780 | 0.718 | 0.672 | — | — | — |
| ETLMSC ^[102] | 0.910 | 0.805 | 0.835 | 0.977 | 0.958 | 0.953 | 0.626 | 0.738 | 0.738 |
| ULTLSE ^[105] | 0.993 | 0.965 | 0.975 | — | — | — | — | — | — |
| CELT ^[76] | 0.998 | 0.998 | 0.998 | — | — | — | — | — | — |
| DAMC ^[163] | 0.460 | 0.162 | 0.162 | 0.290 | 0.316 | 0.316 | — | — | — |
| Multi-NMF ^[28] | 0.699 | 0.480 | 0.518 | — | — | — | — | — | — |

表 18 MVC 方法在数据集 Handwritten、Reuters、Scene-15 上的性能比较

(续表)

| 衡量标准 | Handwritten | | | Reuters | | | Scene-15 | | |
|----------------------------|--------------|-------|-------|---------|-------|-------|----------|-------|-------|
| | NMI | ACC | PUR | NMI | ACC | PUR | NMI | ACC | PUR |
| DNSRF ^[115] | — | — | — | — | — | — | — | — | — |
| FRSMNMF ^[39] | 0.947 | 0.933 | 0.929 | 0.607 | 0.586 | 0.586 | 0.901 | 0.874 | 0.874 |
| CDTMS ^[62] | 0.860 | 0.923 | — | — | — | — | — | — | — |
| SMCMB ^[83] | 0.897 | 0.947 | 0.947 | — | — | — | — | — | — |
| MRL_CAL ^[161] | — | — | — | — | — | — | 0.426 | 0.412 | — |
| MVCNMF ^[38] | — | — | 0.530 | 0.623 | — | — | — | — | — |
| FTMCM ^[97] | — | — | — | — | — | — | 0.932 | 0.889 | 0.889 |
| MCDS ^[114] | 0.911 | 0.944 | — | 0.667 | 0.749 | — | — | — | — |
| Multi-VAE ^[152] | 0.896 | 0.798 | 0.798 | 0.619 | 0.649 | 0.794 | 0.907 | 0.883 | 0.839 |
| DMVCVAE ^[151] | — | — | — | — | — | — | — | — | — |
| GAIMVC ^[159] | 0.793 | 0.871 | 0.871 | 0.375 | 0.477 | 0.477 | 0.748 | 0.767 | 0.761 |

表 19 IMVC 方法在 BBC Sport 的性能比较

| 指标/缺失率 | ACC | | | NMI | | | PUR | | |
|----------------------------|--------------|--------------|--------------|--------------|--------------|-------|--------------|--------------|--------------|
| | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| IMG ^[171] | 0.569 | 0.600 | 0.467 | 0.374 | 0.374 | 0.247 | 0.647 | 0.655 | 0.550 |
| DAIMC ^[172] | 0.740 | 0.685 | 0.540 | 0.613 | 0.547 | 0.345 | 0.807 | 0.759 | 0.612 |
| GIMC-FLSD ^[173] | 0.814 | 0.764 | 0.700 | 0.730 | 0.644 | 0.565 | 0.900 | 0.855 | 0.805 |
| IMSC-AGL ^[174] | 0.716 | 0.721 | 0.683 | 0.664 | 0.622 | 0.541 | 0.835 | 0.828 | 0.790 |
| UEAF ^[175] | 0.781 | 0.772 | 0.653 | 0.692 | 0.681 | 0.506 | 0.871 | 0.860 | 0.745 |
| AGC-IMC ^[176] | 0.836 | 0.809 | 0.704 | 0.731 | 0.679 | 0.533 | 0.867 | 0.835 | 0.762 |
| ARDGL ^[177] | 0.914 | 0.811 | 0.751 | 0.796 | 0.718 | 0.555 | 0.916 | 0.887 | 0.797 |

表 20 IMVC 方法在 3sources 的性能比较

| 指标/缺失率 | ACC | | | NMI | | | PUR | | |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| IMG ^[171] | 0.614 | 0.543 | 0.423 | 0.506 | 0.412 | 0.264 | 0.698 | 0.636 | 0.542 |
| DAIMC ^[172] | 0.516 | 0.489 | 0.507 | 0.501 | 0.435 | 0.416 | 0.696 | 0.659 | 0.652 |
| GIMC-FLSD ^[173] | 0.683 | 0.613 | 0.570 | 0.655 | 0.560 | 0.515 | 0.817 | 0.756 | 0.742 |
| IMSC-AGL ^[174] | 0.664 | 0.654 | 0.625 | 0.635 | 0.610 | 0.512 | 0.788 | 0.773 | 0.717 |
| UEAF ^[175] | 0.580 | 0.523 | 0.537 | 0.593 | 0.533 | 0.447 | 0.767 | 0.734 | 0.644 |
| AGC-IMC ^[176] | 0.774 | 0.723 | 0.680 | 0.681 | 0.624 | 0.520 | 0.832 | 0.774 | 0.743 |
| ARDGL ^[177] | 0.800 | 0.788 | 0.710 | 0.719 | 0.683 | 0.556 | 0.838 | 0.819 | 0.766 |

表 21 IMVC 方法在 MSRC-V1 上的性能比较

| 指标/缺失率 | ACC | | | NMI | | | PUR | | |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 | 0.1 | 0.3 | 0.5 |
| IMG ^[171] | 0.831 | 0.542 | 0.263 | 0.722 | 0.446 | 0.112 | 0.831 | 0.543 | 0.271 |
| DAIMC ^[172] | 0.750 | 0.750 | 0.624 | 0.671 | 0.637 | 0.489 | 0.760 | 0.764 | 0.635 |
| GIMC-FLSD ^[173] | 0.808 | 0.800 | 0.696 | 0.712 | 0.689 | 0.563 | 0.814 | 0.807 | 0.705 |
| IMSC-AGL ^[174] | 0.848 | 0.815 | 0.733 | 0.761 | 0.710 | 0.622 | 0.848 | 0.817 | 0.750 |
| UEAF ^[175] | 0.792 | 0.600 | 0.556 | 0.679 | 0.466 | 0.431 | 0.797 | 0.604 | 0.561 |
| AGC-IMC ^[176] | 0.751 | 0.744 | 0.662 | 0.753 | 0.693 | 0.563 | 0.784 | 0.769 | 0.668 |
| ARDGL ^[177] | 0.852 | 0.792 | 0.737 | 0.789 | 0.723 | 0.572 | 0.849 | 0.809 | 0.777 |

现有的聚类方法缺乏一定的可扩展性。例如,ECMSC 方法在 ORL 数据集上的聚类效果比在 BBC Sport 数据集上的表现更好,这表明同一种方法在不数据集上聚类效果差异性较大,不适合将该方法拓展到其他数据集上。

一些方法受制于数据集的规模。例如,DIMSC、MLAN 和 RMSC 等方法在路透社数据集上表

现不佳,表明它们在处理大规模数据方面存在一定的局限性。

最新的先进方法越来越注重原始数据质量对聚类的影响。在 MVC 中,CELT 利用张量微分的方法解决了数据嵌入带来的数据压缩难题。在 IMVC 中,ARDGL 通过应用动态降维和加权图学习,减轻了数据噪声和冗余的影响。

通过比较不同数据集的实验结果,我们总结了以下几点看法和分析:

在 MVC 中, DIMSC 在 MNIST 和 Scene-15 等图像数据集上的聚类性能较差。这可能是由于 DIMSC 未能考虑不同聚类的不同重要性,从而降低了整体聚类效果。

在 BBC 体育文本数据集中,聚类方法的性能差异很大。ETLMSC 性能最好, NMI 为 0.972, ACC 为 0.959, 而 ECMSC 性能低得多, NMI 为 0.04, ACC 为 0.299。两种方法之间的显著差异主要是由于 ECMSC 忽略了多视图之间的高阶相关性,降低了其聚类效果。

在 IMVC 中,同一方法的聚类性能受到缺失率的影响。例如,在 BBC 体育数据集上,当缺失率为 0.1 时,ARDGL 的 NMI 为 0.914,但当缺失率增加到 0.5 时, NMI 降为 0.7507。

另外,对于缺失数据的处理方法也十分影响聚类效果,具体的,一些方法直接忽略缺失数据的影响,得到的实验结果数据比较差,比如 IMG, DAIMC; 一些基于重构缺失数据的方法得到的实验结果相比于以上两种方法,聚类性能得到了较大的提升,比如, GIMC-FLS 方法在 BBC Sport 数据上的 ACC 指标相比于 IMG 有大约 24% 的提升,这说明直接忽略缺失数据将导致缺少充足的聚类判别信息,从而影响聚类效果。由表可知,基于特征平均值填充的方法要优于零元素填充的方法,造成这种次优聚类效果的原因是基于零元素填充的方法忽略了数据样本之间的关联性。

我们采用表格形式对比了不同类型的多视图聚类方法在 ORL, UCI 等八个不同标准数据集上的优缺点,详细结果如表 22 和表 23 所示。

表 22 不同类型的多视图聚类算法在 ORL, UCI, MSRC-V1, Yale 数据集上的优缺点

| | ORL | UCI | MSRC-V1 | Yale |
|------|---|--|---|---|
| NMF | 适用于 ORL 数据,能够保留局部空间结构,但计算复杂度较高 | 在 UCI 低维特征数据表现良好,但对于 UCI 高维数据,可能面临特征冗余问题。子空间学习可以很好地分离不同类别,提高聚类的区分度由于手写数字的变形较大,部分方法难以适应数据变化 | 适用 MSRC-V1 多个类别的图像数据,但计算复杂度较高 | 适用于 Yale 数据集,具有很好的解释性,但计算复杂度较高,不适用拓展版 Yale 数据集 |
| 子空间 | 能够很好地保留 ORL 图像的局部结构,但是可能存在视图间冗余信息,影响聚类结果 | | 子空间学习的方法可以捕捉不同视图间的互补信息,但是需要对不同视图进行特定预处理,以确保稳定性 | 能够有效提取人脸数据的局部特征,但是在极端光照变化下,部分方法的效果可能下降 |
| 图学习 | 通过图构造能够有效捕捉图像数据的局部结构,结合谱聚类方法,实现更精细的分类,但是计算复杂度较高,容易受异常数据点的影响 | 适用于 UCI 低维数据,能够较好地发现不同类别之间的拓扑结构,依赖参数调节,在 UCI 上的泛化能力较弱 | 适用于 MSRC-V1 数据集,通过图结构可以有效建模不同类别的特征关系,但是图学习方法中视图间的权重分配问题可能影响最终聚类效果 | 适用于光照、姿态变化较大的 Yale 数据集,能够增强分类能力,但是计算复杂度较高,不适用于拓展的 Yale 数据集 |
| 张量学习 | 通过张量建模能够有效捕捉 ORL 数据的高阶特征信息,但计算复杂度较高,不适用于大规模数据集 | 结合张量正则化方法可增强 UCI 数据结构信息,提高分类准确率但对噪声数据敏感,可能需要额外的去噪步骤 | 结合张量分解技术,可提取更具判别力的 MSRC-V1 数据的高阶特征。但需要调整张量秩和正则化参数,以优化聚类效果 | 通过张量低秩表示方法,可以更好地去除 Yale 数据的噪声,提高聚类准确性。但是参数调节较为复杂,需要优化超参数以达到最佳性能 |
| 深度嵌入 | 能够自动提取高阶特征,适应复杂的 ORL 数据,但是依赖大量数据进行训练,可能导致过拟合 | 能够自动学习 UCI 不同视图的特征,提高聚类的泛化能力。但是训练过程较慢,计算资源消耗较大 | 结合注意力机制或 GCN 方法,可提升聚类的稳健性,可能对特定视图的依赖较大,影响跨视图信息融合 | 能够学习 Yale 不同视图的人脸特征,提高聚类的一致性,训练成本较高,可能需要高性能计算资源 |
| 对比学习 | 通过对比损失函数增强视图间的一致性,能够学习 ORL 数据稳定嵌入,但是需要数据预处理 | 通过视图对比学习,可以提高不同手写数字视图之间的特征对齐能力,计算消耗大 | 适应 MSRC-V1 类别复杂、类别间相似度较高的情况,但依赖较大的训练样本规模,小样本情况下可能难以获得稳定结果 | 能够利用对比损失约束嵌入空间,使得聚类类内更加紧密,类间更加分离。但是需要额外的样本对选择策略,否则可能影响对比学习的效果 |
| 生成式 | 适用于处理 ORL 数据中的姿态、光照变化,提高鲁棒性。生成模型训练难度大 | 结合变分自编码器(VAE),可进行特征去噪,提高聚类稳定性。但是可能面临模式坍塌问题,导致视图信息丢失 | 结合 GAN 方法,可提升 MSRC-V1 特征分布的可变性,提高聚类效果。但是对某些特定视图存在过拟合风险 | 可以学习 Yale 不同视图下的脸部特征,提高跨视图一致性。但是训练成本高,可能需要 GPU 加速支持 |

表 23 不同类型的多视图聚类算法在 BBC Sport, MNIST, Handwritten, Reuters 数据集上的优缺点

| | BBC Sport | MNIST | Handwritten | Reuters |
|------|---|--|---|--|
| NMF | 适用于 BBC Sport 文本数据,能够提取主题特征,提高聚类可解释性,但对文本数据的非线性关系建模能力有限 | 适用 MNIST 低维数据,但是计算复杂性高,难以拓展到大规模数据 | 适用于手写字符数据,能够进行特征降维,提高聚类可分性。但是对于具有复杂笔迹变化的数据,可能难以捕捉非线性特征 | NMF 由于涉及大量的矩阵分解,因此并不适用于大规模文本数据集,但使用低秩分解可以有效降维,提高计算效率 |
| 子空间 | 能够保留局部结构信息,增强文本的可分性。但是不同视图间的信息融合可能受限,影响聚类效果 | 适用于 MNIST 手写数字数据,能够去除冗余特征,提高区分度。但是依赖参数调节,模型在 MNIST 上泛化能力弱 | 适用于手写字符数据,可学习视图间的共享子空间,提高信息融合能力,但是计算量大,训练时间长 | 虽然可以保留数据的局部结构信息,增强文本的可区分性,但是计算复杂度高,使得它并不适用 Reuters 较大规模数据集 |
| 图学习 | 有效学习文本数据结构之间的关系,但是图学习方法依赖邻接矩阵构造,影响聚类效果 | 适用 MNIST 数据集,结合图正则化方法可增强数据结构信息,提高分类稳定性。但可能难以处理某些复杂的非线性数据模式 | 适用于不同视图间信息较强的数据,计算资源消耗大,难以扩展到超大规模数据集 | 通过图构造能够学习文本数据之间的结构关系,提高聚类一致性。但是构造亲和图以及融合亲和图得到共识亲和图需要大量的矩阵运算,并不适用 Reuters 数据集 |
| 张量学习 | 能够利用张量分解学习文本的高阶特征,但是计算复杂度较高,训练过程耗时 | 适用于手写数据,能够提取跨视图的高阶信息,提高聚类精度依赖于张量秩的选择,参数敏感性较强 | 适用于字符数据,能够建模跨视图的非线性关系。但计算成本较大,尤其在数据量增加时难以扩展 | 张量学习的方法涉及张量分解,计算复杂度较高,不适用 Reuters 数据集,但是基于张量学习的方法可以很好捕捉数据之间的高阶关系 |
| 深度嵌入 | 通过深度神经网络自动提取文本的高层次语义特征,依赖大规模数据进行训练,否则可能会发生过拟合 | 通过 DAE 可以学习 MNIST 数据的潜在表示,提高了聚类精度,但是对参数敏感 | 适用于手写字符数据,能够捕捉更丰富的特征信息,提高聚类质量,但是可能对某些特定视图依赖较大,影响跨视图信息融合 | 能够有效建模多视图数据之间的非线性关系,提高信息融合能力。但是训练时间长需要大量的计算资源 |
| 生成式 | 能够生成高质量的文本特征,提高聚类效果,但是生成质量依赖于模型架构和超参数选择,调优难度大 | 结合变分自编码器(VAE),可进行特征去噪,提高鲁棒性。但是训练时间长,计算资源消耗大 | 适用于手写字符数据,能够生成特征增强数据,提高聚类泛化能力。可能对某些类别的视图存在过拟合风险 | 利用 VAE 或者 GAN 网络能够生成高质量的文本特征,提高聚类效果,但是生成质量依赖 GPU 和超参选择,很难调优 |

如表 22 和表 23 所示,基于 NMF 和张量学习的多视图聚类方法适用于一般规模数据集,比如 ORL 数据集和 Yale 数据集,但是不适用于较大规模数据集,比如 Reuters。这主要因为基于 NMF 和张量学习的多视图聚类方法涉及大量的矩阵和张量的运算,需要消耗巨大的计算资源。对于视图数量比较多的数据集,比如 MSRC-V1 数据集和 Reuters 数据集,基于对比学习和深度嵌入式的多视图聚类方法可以很好地捕捉视图间的复杂关系,和不同视图之间的非线性结构,更有利于提升聚类效果。虽然基于子空间和图学习的多视图聚类方法可以有效捕捉数据局部结构,但是这些方法需要构造自表达系数矩阵或者邻接矩阵,增加了计算的复杂度因此不适用于大规模数据,而且,不同的构造方法对聚类性能也有一定的影响,基于生成式的多视图聚类方法可以有效地弥补缺失数据,增强模型的鲁棒性,因

此它比较适用现实生活中的真实数据集,但是生成式模型太依赖超参的选择,很难调优。

4.5 复杂性分析

本节总结了上述实验方法的时间和空间复杂度,分析了每种方法的计算复杂度。表 24 显示了 MVC 方法的复杂度数据,表 25 显示了 IMVC 方法的复杂度数据。所有复杂度数据均来自已发表的文章,符号“-”表示原始论文未提供相应方法的复杂度信息。

根据表格分析和相关文献,我们发现某些涉及矩阵因式分解和张量分解的实验方法具有很高的计算复杂度。例如,ULTLSE 和 CELT 的时间复杂度均为 $O(n^2)$,而 ETLMSC、FTMCM 和 CDTMSC 涉及参数更新、张量快速傅里叶变换(FFT)及其求反,在某些情况下复杂度可达 $O(n^3)$ 。

锚点学习能有效降低时间复杂度。例如,

表 24 基于 MVC 实验方法的复杂性的总结

| 方法 | 时间复杂性 |
|----------------------------|--|
| DIMSC ^[54] | — |
| MLAN ^[89] | $O(n^2d + tcn^2)$ |
| RMSC ^[73] | — |
| LRR ^[68] | $O(n^2d) + O(n_s(dnr_A + nr_A^2 + r_A^3))$ |
| ECMSC ^[56] | — |
| ETLMSC ^[102] | $O(N^3 + KMN^2(M + \log(N)))$ |
| ULTLSE ^[105] | $O(tmn^2 \log(n))$ |
| CELT ^[76] | $O(T(\sum_{v=1}^V (d^v)^3 + VN^2(m + V + \log(N))))$ |
| DAMC ^[163] | — |
| Multi-NMF ^[28] | $O(t_{out}t_{in}n_vMNk)$ |
| SNMF ^[26] | — |
| DNSRF ^[115] | $O(LPQk)$ |
| FRSMNMF ^[39] | $O(tvm^vnd^v)$ |
| CDTMS ^[62] | $O(T(2VN^3 + 2N^2V\log(N)))$ |
| SMCMB ^[83] | $O(NM(q(dt + k)) + k^2t)$ |
| MRL_CAL ^[161] | — |
| MVCNMF ^[38] | $O(n_vtm_vnk)$ |
| FTMCM ^[97] | $O(VN^2 \log(N) + V^2N^2 + VN^3 + N^3)$ |
| MCDS ^[114] | $O(Tnd)$ |
| Multi-VAE ^[152] | $O(VNM^2)$ |
| DMVCVAE ^[151] | — |
| GAIMVC ^[159] | — |

表 25 基于 IMVC 实验方法的复杂性的总结

| 方法 | 时间复杂性 |
|----------------------------|--|
| IMG ^[171] | — |
| DAIMC ^[172] | $O(nvLd^3 + LMdkN)$ |
| GIMC-FLSD ^[173] | $O(\tau \sum_{K=1}^V m_k d^2)$ |
| IMSC-AGL ^[174] | $O(\tau(kn^3 + n^3 + \sum_v n_v^3))$ |
| UEAF ^[175] | $O(\tau(2n^3 + \sum_{v=1}^t m_v c^2))$ |
| AGC-IMC ^[176] | $O(\tau cn^2)$ |
| ARDGL ^[177] | $O(4nd^2 + d^3 + n^2)$ |

SMCMB 等方法与锚学习相结合后,复杂度达到了 $O(n)$ 。与传统的 NMF 方法相比,Multi-NMF 引入了一种新颖的矩阵更新策略,将时间复杂度降低到与样本大小成线性关系。

基于深度学习框架的 MVC 方法也能有效降低计算复杂度。例如,基于深度 SNMF 框架的 MCDS 的计算复杂度为 $O(n)$ 。

在 IMVC 方法中,计算复杂度与矩阵运算密切相关。例如,IMSC-AGL、UEAF、AGC-IMC 和 ARDGL 等方法由于涉及矩阵反演和特征值分解等复杂运算,其复杂度大于 $O(n^2)$ 。不过,通过实施矩

阵更新策略,如 DAIMC 和 GIMC-FLSD 中的策略,时间复杂度可以大大降低。

5 应用

在本节中,我们对多视图聚类在现实生活中的一些应用进行了简要概述,其中主要包括在计算机视觉领域中的三维(3D)形状理解与 3D 重建方面的应用,在生物信息方面的应用,以及在自然语言处理领域中文本聚类的应用。

5.1 3D 形状理解

当前最先进的 3D 模型的识别能力受到数据集的限制,这些数据集具有少量注释数据和一组预定义的类别。在二维(2D)对应物中,最新的进展表明,通过使用来自其他模式(例如语言)的知识可以显著缓解类似问题,Xue 等人提出一种来学习图像、文本和 3D 点云的统一表示(ULIP)方法,该方法利用来自三个视图关系的对象三元组进行模型预训练^[178]。由于现有的多视图 3D 表示学习方法中,图像视图的贡献不如语言。这归因于 2D 图像中的域偏移和每种模态的不同焦点。为了在预训练中更有效地利用这两种模式,Zhang 等人提出一种基于三个协同适配器的新型两阶段学习方法。首先,将 CLIP 图像适配器通过调整 CLIP 的视觉表示以用于合成图像—文本对,减轻了 3D 渲染图像和自然图像之间的主要差距。其次,将 3D 形状表示空间分为两个完整的心理子空间:一个专注于视觉属性,另一个用于语义理解,从而确保更全面、更有效的多模态预训练^[179]。

5.2 3D 重建

由于从 3D 场景中提取平面对于机器人和增强现实中的下游任务非常有用,Waston 等人将流行的聚类算法与 3D 几何估计的最新改进相结合,得到了一个极具竞争力的基线,解决了从姿势图像中估计场景的平面表面问题^[180]。从多视图图像进行逼真人脸渲染时,由于面部的复杂空间变化反射特性和几何特性,要有效恢复 3D 面部表示具有挑战性,Zheng 等人提出一种神经面部渲染(NeuFace)方法,该方法通过将基于 MLP 的神经网络整合到基于物理渲染中,以协作方式捕捉复杂的面部几何形状^[181]。学习特定的运动结构(sfm)是 3D 重建的重要内容,Brynte 提出一种新模型,该模型从多个视图中检测到的 2D 关键点作为输入,然后输出相应的相机姿势和 3D 关键点坐标,然后利用图神经网络

络学习特定于 sfm 的基元,快速推理新的和看不见的序列侦察结构^[182]。He 等人提出一种无检测器的 sfm 框架,它首先从量化模型中重建一个粗略的 sfm 模型,然后它通过新颖的迭代优化管道来优化模型,该管道在基于注意力的多视图匹配模块之间迭代以优化特征跟踪,并在几何优化模块之间迭代以提高重建精度^[183]。由于观察多个运动时,识别集群是一项非常具有挑战性的任务,Lochman 等人通过学习从轨迹到表示生成的运动的嵌入向量映射,获得的轨迹嵌入用于对多个观察到的运动子空间聚类,达到识别潜在运动目的^[184]。

5.3 生物信息

单细胞生物技术的进步已经产生了细胞水平基因表达谱的单细胞 RNA 测序(scRNA-seq),但由于 scRNA-seq 的异质性,高维性和噪声,在研究细胞类型分布仍然存在许多问题,为了解决这个问题,Wu 等人提出一种针对 scRNA-seq 数据的基于图学习的多视图聚类(MCGL),它构建了多个特征空间,并利用多视图学习从不同角度对 scRNA-seq 数据进行综合表示,它自适应地学习细胞的相似性图,克服了对固定相似性的依赖,将 scRNA-seq 分析转化为对聚类的分析^[185]。由于数据的固有特性,设计一种在各种类型的多组学数据中表现良好的聚类方法是一个挑战,Chen 等人提出一种用于单细胞聚类的图正则化多视图集成聚类(GRMEC-SC)模型,该方法可以自适应整合多个组学数据,并利用来自多个碱基聚类结果的见解,该模型在具有不同特征的不同组学数据中取得了有竞争力的性能^[186]。由于数据类型的多样性和有效集成异构信息的困难,现有对基因组重叠分组的方法在实际中面临挑战,Wang 等人提出一种基于对比多视图表示学习的分箱方法,该方法利用数据增强生成每个重叠群的多个片段,并通过对比学习获得异构特征的高质量嵌入^[187]。另外,Shi 等人提出一种多视图基于图的聚类(MRGC)的方法来有效识别癌症亚型,该方法首先从原始组中学习稳健的潜在表示,以减轻噪声影响,然后根据这些新表示自适应学习一组矩阵,最后利用图的共识结构得到全局相似图,这三部分可以通过迭代方法相互加强,在聚类方面取得很好效果^[188]。Shah 等人根据多视图数据包含相似性的固有的集群结构这一推论,提出一种基于多视图内核学习的疾病基因鉴定算法(DiGid),该方法旨在学习共识内核,有效捕捉视图异构信息,并描述底层固有的集群结构,同时对学习到的多视图内核施加低秩

约束,可以有效将它划分为 k 个集群,学习到的关键簇结构用于管理一组潜在的疾病基因^[189]。

5.4 文本聚类

多视图文档聚类(MVDC)是自然语言处理中的一种复杂方法,由于现有基于 MVDC 的方法通常取决于词袋和预训练语言模型的性能,往往会忽略上下文语义,为了解决这个问题,Du 等人提出一种利用增强语义嵌入和一致上下文语义的深度模型(MVDC),该模型使用语义嵌入通过考虑互补语义信息来解决高维挑战,同时它通过先进的嵌入技术捕捉了单词和文档之间错综复杂的语义关系,确保文本内容更丰富更细致表示^[190]。现有来自社交媒体的新闻文本具有很强的时间敏感性,在过去的新闻主题识别中没有考虑将时间建模和语义建模结合起来,Zhang 等人提出一种融合新闻数据集语义和时间信息的自适应最优主题编号确定方法以解决此问题,该方法首先将语义和时间作为两个不同视角进行提取,然后对多视图信息进行密度峰值聚类,聚类结果被用作最佳主题数^[191]。Saini 等人基于多视图聚类框架,同时考虑了科学文献的两种视图(语义和句法),然后使用差分进化算法作为底层优化策略,获得了对应于不同视图的改进分区,在获得最优分区后,研究各种句子级别的特征,如长度,句子位置等,以提取高分句子形成摘要^[192]。Li 等人提出一种多视角学者聚类模型(MSCT)该模型包含两个新颖方案,第一种是多视角整合,即 MSCT 协作跟踪学者时变主题分布。第二种是动态兴趣跟踪,即根据当前学者的兴趣分布,动态模拟每个学者的兴趣分布动态模型^[193]。

6 结论与展望

在大数据迅速发展的时代,多视图数据变得越来越重要。作为理解多视图数据结构的有力工具,多视图聚类(MVC)备受关注。尽管近年来对 MVC 进行了大量调查研究,但仍缺乏对 MVC 方法论的全面总结,对实验数据中不完全多视图聚类的探索仍然有限。为了弥补这些不足,本文为 MVC 提供了一个更清晰、更完整的路线图。

我们对现有的多视图聚类(MVC)方法进行了系统梳理,并将其归纳为三大类:基于启发式的多视图聚类方法、基于判别式的多视图聚类方法和基于生成式的多视图聚类方法,我们在本文的第 3 章对它们进行了具体分析,在本文第 4 章汇总了 22 种常

用完整多视图聚类方法在 9 个公开数据集上的聚类性能,以及 7 种不完整多视图聚类(IMVC)方法在 3 个数据集上的聚类结果,并对这些方法的基准性能进行了详细分析和比较。最后,我们在第 5 章归纳了目前多视图聚类的一些前沿应用。我们相信,这一系统性的分析与总结对于进一步推动 MVC 领域的研究与发展具有重要的参考价值和指导意义。最后,根据分析结果,本文确定了推进多视图聚类(MVC)和指导未来研究工作的几个关键研究方向:

(1)处理大规模和高维数据:当前的 MVC 方法通常在中小型数据集上表现出色,但在应用于大规模和高维数据时却面临巨大挑战。高维数据通常包含冗余特征,这会使特征表示复杂化,并对聚类性能产生负面影响。未来的 MVC 方法应侧重于提高更有效地处理此类大规模高维数据集的能力。

(2)处理增量性质和不确定性:大多数现有的 MVC 方法都假定视图数据是静态的,这就限制了它们对动态和增量数据集的适用性。未来的技术应支持增量更新并有效处理动态变化。并且可以采用模糊聚类技术来减少现实世界数据集中的噪音和不确定性,对于提高鲁棒性也至关重要。

(3)处理未映射的多视图数据:许多 MVC 方法假定各视图之间的样本对齐一致,但往往忽略了未映射样本,这对聚类性能产生了负面影响。开发有效的配准技术来处理未映射样本是进一步研究的关键领域。

(4)降低计算复杂度:计算复杂度高仍然是 MVC 的一大挑战,因为许多方法都依赖于密集的矩阵和张量分解,这增加了内存使用量和计算时间。未来的发展应优先简化这些操作,以降低复杂性并提高效率。

(5)确定群集数量:目前的许多 MVC 方法都假定有固定数量的簇,这在实际应用中可能并不可行。开发能自动确定簇数的方法,将大大提高 MVC 在各种情况下的灵活性和实际效果。

(6)增强解释力:虽然基于深度学习的 MVC 方法能够捕提高阶关系,但它们引入了额外的参数,往往缺乏理论上的可解释性。未来的研究应致力于优化这些网络,提高 MVC 的可解释性和解释力,使其更加实用和易懂。

参 考 文 献

[1] Askari S. Fuzzy C-Means clustering algorithm for data with

unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, 2021, 165: 113856

- [2] Haris M, Yusoff Y, Zain A M, et al. Breaking down multi-view clustering: A comprehensive review of multi-view approaches for complex data structures. *Engineering Applications of Artificial Intelligence*, 2024, 132: 107857
- [3] Zhao J, Xie X, Xu X, et al. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 2017, 38: 43-54
- [4] Elkahky A M, Song Y, He X. A multi-view deep learning approach for cross domain user modeling in recommendation systems//*Proceedings of the International Conference on World Wide Web*. Florence, Italy, 2015: 278-288
- [5] Zhou L, Du G, Lü K, et al. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 2024, 56(7): 1-38
- [6] Xu Y, Niu G. Research on multi-view clustering algorithm based on sequential three-way decision. *Applied Soft Computing*, 2024, 158: 111590
- [7] Sharma K K, Seal A. Outlier-robust multi-view clustering for uncertain data. *Knowledge-Based Systems*, 2021, 211: 106567
- [8] Li A, Xu H, Feng C, et al. Incomplete multi-view clustering via local and global bagging of anchor graphs. *Expert Systems with Applications*, 2024, 248: 123402
- [9] Liu W, Jing X, Jia X, et al. Low-rank representation induced missing-view recovery for incomplete multi-view clustering. *Neurocomputing*, 2024, 595: 127870
- [10] Hao W, Pang S, Bai X, et al. Tensor-based incomplete multi-view clustering with low-rank data reconstruction and consistency guidance. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(12): 7156-7169
- [11] Zhou P, Shen Y D, Du L, et al. Incremental multi-view spectral clustering. *Knowledge-Based Systems*, 2019, 174: 73-86
- [12] Yin H, Hu W, Zhang Z, et al. Incremental multi-view spectral clustering with sparse and connected graph learning. *Neural Networks*, 2021, 144: 260-270
- [13] Huang Y, Guo K, Yi X, et al. Incremental unsupervised feature selection for dynamic incomplete multi-view data. *Information Fusion*, 2023, 96: 312-327
- [14] Yang M, Li Y, Hu P, et al. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(1): 1055-1069
- [15] Cai B, Lu G F, Yao L, et al. Aligned multi-view clustering for unmapped data via weighted tensor nuclear norm and adaptive graph learning. *Neurocomputing*, 2024, 597: 128016
- [16] Tang J, Yi Q, Fu S, et al. Incomplete multi-view learning: Review, analysis, and prospects. *Applied Soft Computing*, 2024, 153: 111278

- [17] Chao G, Sun S, Bi J. A survey on multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2021, 2(2): 146-168
- [18] Yang Y, Wang H. Multi-view clustering: A survey. *Big data mining and analytics*, 2018, 1(2): 83-107
- [19] Fang U, Li M, Li J, et al. A comprehensive survey on multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(12): 12350-12368
- [20] Zhang C, Hu Q, Fu H, et al. Latent multi-view subspace clustering//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA, 2017: 4279-4287
- [21] Kang Z, Zhou W, Zhao Z, et al. Large-scale multi-view subspace clustering in linear time//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA, 2020: 4412-4419
- [22] Wang H, Yang Y, Liu B. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 32(6): 1116-1129
- [23] Kolda T G, Bader B W. Tensor decompositions and applications. *SIAM Review*, 2009, 51(3): 455-500
- [24] Carroll J D, Chang J J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 1970, 35(3): 283-319
- [25] Zong L, Zhang X, Zhao L, et al. Multi-view clustering via multi-manifold regularized non-negative matrix factorization. *Neural Networks*, 2017, 88: 74-89
- [26] Ding C H Q, Li T, Jordan M I. Convex and semi-nonnegative matrix factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 32(1): 45-55
- [27] Cai D, He X, Han J, et al. Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 33(8): 1548-1560
- [28] Liu J, Wang C, Gao J, et al. Multi-view clustering via joint nonnegative matrix factorization//*Proceedings of the SIAM International Conference on Data Mining*. Austin, USA, 2013: 252-260
- [29] Zong L, Zhang X, Liu X. Multi-view clustering on unmapped data via constrained non-negative matrix factorization. *Neural Networks*, 2018, 108: 155-171
- [30] Wang J, Tian F, Yu H, et al. Diverse non-negative matrix factorization for multiview data representation. *IEEE Transactions on Cybernetics*, 2017, 48(9): 2620-2632
- [31] Tan J, Yang Z, Ren J, et al. A novel robust low-rank multi-view diversity optimization model with adaptive-weighting based manifold learning. *Pattern Recognition*, 2022, 122: 108298
- [32] Ma J, Zhang Y, Zhang L. Discriminative subspace matrix factorization for multiview data clustering. *Pattern Recognition*, 2021, 111: 107676
- [33] Huang Z, Ren Y, Pu X, et al. Dual self-paced multi-view clustering. *Neural Networks*, 2021, 140: 184-192
- [34] Wang H, Yang Y, Li T. Multi-view clustering via concept factorization with local manifold regularization//*Proceedings of the IEEE 16th International Conference on Data Mining (ICDM)*. Barcelona, Spain, 2016: 1245-1250
- [35] Huang H, Zhou G, Zheng Y, et al. Exclusivity and consistency induced NMF for multi-view representation learning. *Knowledge-Based Systems*, 2023, 281: 111020
- [36] Luong K, Nayak R. Learning inter-and intra-manifolds for matrix factorization-based multi-aspect data clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(7): 3349-3362
- [37] Liu H, Wu Z, Li X, et al. Constrained nonnegative matrix factorization for image representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 34(7): 1299-1311
- [38] Cai H, Liu B, Xiao Y, et al. Semi-supervised multi-view clustering based on constrained nonnegative matrix factorization. *Knowledge-Based Systems*, 2019, 182: 104798
- [39] Cui G, Wang R, Wu D, et al. Semi-supervised Multi-view Clustering based on NMF with Fusion Regularization. *ACM Transactions on Knowledge Discovery from Data*, 2024, 18(6): 1-26
- [40] Zhang D P, Luo Y H, Yu Y Y, et al. Semi-supervised multi-view clustering with dual hypergraph regularized partially shared non-negative matrix factorization. *Science China Technological Sciences*, 2022, 65(6): 1349-1365
- [41] Cai H, Liu B, Xiao Y, et al. Semi-supervised multi-view clustering based on orthonormality-constrained nonnegative matrix factorization. *Information Sciences*, 2020, 536: 171-184
- [42] Li G, Song D, Bai W, et al. Consensus and complementary regularized non-negative matrix factorization for multi-view image clustering. *Information Sciences*, 2023, 623: 524-538
- [43] Chen Z, Lin P, Chen Z, et al. Diversity embedding deep matrix factorization for multi-view clustering. *Information Sciences*, 2022, 610: 114-125
- [44] Yang B, Zhang X, Nie F, et al. Fast multi-view clustering via nonnegative and orthogonal factorization. *IEEE Transactions on Image Processing*, 2020, 30: 2575-2586
- [45] Chen W S, Xie K, Liu R, et al. Symmetric nonnegative matrix factorization: A systematic review. *Neurocomputing*, 2023, 557: 126721
- [46] Gu Z, Feng S, Li Z, et al. NOODLE: Joint cross-view discrepancy discovery and high-order correlation detection for multi-view subspace clustering. *ACM Transactions on Knowledge Discovery from Data*, 2024, 18(6): 1-23
- [47] Zhang C, Fu H, Hu Q, et al. Generalized latent multi-view subspace clustering. *IEEE Transactions on Pattern Analysis*

- and Machine Intelligence, 2018, 42(1): 86-99
- [48] Zhou T, Zhang C, Peng X, et al. Dual shared-specific multi-view subspace clustering. *IEEE Transactions on Cybernetics*, 2019, 50(8): 3517-3530
- [49] Si X, Yin Q, Zhao X, et al. Consistent and diverse multi-view subspace clustering with structure constraint. *Pattern Recognition*, 2022, 121: 108196
- [50] Luo S, Zhang C, Zhang W, et al. Consistent and specific multi-view subspace clustering//*Proceedings of the AAAI Conference on Artificial Intelligence*. New Orleans, USA, 2018: 3730-3737
- [51] Cai X, Huang D, Zhang G Y, et al. Seeking commonness and inconsistencies: A jointly smoothed approach to multi-view subspace clustering. *Information Fusion*, 2023, 91: 364-375
- [52] Cheng Y, Song P, Mu J, et al. Comprehensive multi-view self-representations for clustering. *Expert Systems with Applications*, 2024, 251: 124103
- [53] Zhang P, Liu X, Xiong J, et al. Consensus one-step multi-view subspace clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(10): 4676-4689
- [54] Cao X, Zhang C, Fu H, et al. Diversity-induced multi-view subspace clustering//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 586-594
- [55] Lan S, Zheng Q, Yu Y. Double-level view-correlation multi-view subspace clustering. *Knowledge-Based Systems*, 2024, 284: 111271
- [56] Wang X, Guo X, Lei Z, et al. Exclusivity-consistency regularized multi-view subspace clustering//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA, 2017: 923-931
- [57] Huang S, Liu Y, Tsang I W, et al. Multi-view subspace clustering by joint measuring of consistency and diversity. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(8): 8270-8281
- [58] Liu X, Pan G, Xie M. Multi-view subspace clustering with adaptive locally consistent graph regularization. *Neural Computing and Applications*, 2021, 33(22): 15397-15412
- [59] Su C, Yuan H, Lai L L, et al. Anchor-based multi-view subspace clustering with graph learning. *Neurocomputing*, 2023, 547: 126320
- [60] Gu Z, Feng S. From dictionary to tensor: A scalable multi-view subspace clustering framework with triple information enhancement. *Advances in Neural Information Processing Systems*, 2024, 37: 103545-103573
- [61] Zhang C, Fu H, Liu S, et al. Low-rank tensor constrained multiview subspace clustering//*Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile, 2015: 1582-1590
- [62] Xiao C, Huang Y, Huang H, et al. Consistency and diversity induced tensorized multi-view subspace clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 9(1): 798-809
- [63] Long Z, Zhu C, Chen J, et al. Multi-view MERA subspace clustering. *IEEE Transactions on Multimedia*, 2023, 26: 3102-3112
- [64] Deng X, Zhang Y, Gu F. Fusion and enhancement of consensus matrix for multi-view subspace clustering. *Mathematics*, 2023, 11(6): 1509
- [65] Zhou S, Yang M, Wang X, et al. Anchor-based scalable multi-view subspace clustering. *Information Sciences*, 2024, 666: 120374
- [66] Ou Q, Wang S, Zhang P, et al. Anchor-based multi-view subspace clustering with hierarchical feature descent. *Information Fusion*, 2024, 106: 102225
- [67] Sun M, Zhang P, Wang S, et al. Scalable multi-view subspace clustering with unified anchors//*Proceedings of the 29th ACM International Conference on Multimedia*. Virtual, 2021: 3528-3536
- [68] Liu G, Lin Z, Yan S, et al. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 35(1): 171-184
- [69] Wang Q, Tao Z, Gao Q, et al. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 35(5): 7244-7250
- [70] Huang S, Liu Y, Ren Y, et al. Learning smooth representation for multi-view subspace clustering//*Proceedings of the 30th ACM International Conference on Multimedia*. New York, USA, 2022: 3421-3429
- [71] Wei L, Song S. Multi-view subspace clustering via an adaptive consensus graph filter//*Proceedings of the International Conference on Multimedia Retrieval*. New York, USA, 2024: 776-784
- [72] Kumar A, Rai P, Daume H. Co-regularized multi-view spectral clustering//*Proceedings of the Advances in Neural Information Processing Systems*. Granada, Spain, 2011: 1413-1421
- [73] Xia R, Pan Y, Du L, et al. Robust multi-view spectral clustering via low-rank and sparse decomposition//*Proceedings of the AAAI Conference on Artificial Intelligence*. Quebec, Canada, 2014: 2149-2155
- [74] Wang S, Chen Y, Jin Y, et al. Error-robust low-rank tensor approximation for multi-view clustering. *Knowledge-Based Systems*, 2021, 215: 106745
- [75] Peng X, Huang Z, Lv J, et al. COMIC: Multi-view clustering without parameter selection//*Proceedings of the International Conference on Machine Learning*. New York, USA, 2019: 5092-5101
- [76] Zhang Y, Sun X, Cai H, et al. Collaborative embedding

- learning via tensor integration for multi-view clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 8(2): 1841-1852
- [77] Gao Q, Wan Z, Liang Y, et al. Multi-view projected clustering with graph learning. *Neural Networks*, 2020, 126: 335-346
- [78] Ren Z, Li X, Mukherjee M, et al. Robust multi-view graph clustering in latent energy-preserving embedding space. *Information Sciences*, 2021, 569: 582-595
- [79] Dai J, Ren Z, Luo Y, et al. Multi-view clustering with latent low-rank proxy graph learning. *Cognitive Computation*, 2021, 13(4): 1049-1060
- [80] Sun Y, Ren Z, Cui Z, et al. Feature weighted multi-view graph clustering. *IEEE Transactions on Consumer Electronics*, 2023, 70(1): 401-413
- [81] Zhang Z, Liu L, Shen F, et al. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 41(7): 1774-1782
- [82] Zhao X, Wang S, Liu X, et al. Multi-view clustering via dynamic unified bipartite graph learning. *Pattern Recognition*, 2024, 156: 110715
- [83] Lao J, Huang D, Wang C D, et al. Towards scalable multi-view clustering via joint learning of many bipartite graphs. *IEEE Transactions on Big Data*, 2023, 10(1): 77-91
- [84] Han X, Zhou F, Ren Z, et al. View-specific anchors coupled tensorial bipartite graph learning for incomplete multi-view clustering. *Information Sciences*, 2024, 664: 120335
- [85] Li L, Zhang J, Wang S, et al. Multi-view bipartite graph clustering with coupled noisy feature filter. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(12): 12842-12854
- [86] Liu S, Liao Q, Wang S, et al. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(8): 4207-4219
- [87] Wang S, Liu X, Liao Q, et al. Scalable Multi-View Graph Clustering With Cross-View Corresponding Anchor Alignment. *IEEE Transactions on Knowledge and Data Engineering*, 2025, 37(5): 2932-2945
- [88] Zhang C, Chen L, Shi Z, et al. Latent information-guided one-step multi-view fuzzy clustering based on cross-view anchor graph. *Information Fusion*, 2024, 102: 102025
- [89] Nie F, Cai G, Li J, et al. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, 2017, 27(3): 1501-1511
- [90] Zhao W, Li Q, Xu H, et al. Anchor Graph-Based Feature Selection for One-Step Multi-View Clustering. *IEEE Transactions on Multimedia*, 2024, 26: 7413-7425
- [91] Dai H, Liu Y, Su P, et al. Multi-view clustering by inter-cluster connectivity guided reward//*Proceedings of the international conference on machine learning*. Vienna, Austria, 2024: 9846-9855
- [92] Tan Y, Liu Y, Huang S, et al. Sample-level multi-view graph clustering//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada, 2023: 23966-23975
- [93] De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 2000, 21(4): 1253-1278
- [94] Tucker L R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966, 31(3): 279-311
- [95] Liu W, Liu L, Zhang Y, et al. Enhanced tensor multi-view clustering via dual constraints. *Engineering Applications of Artificial Intelligence*, 2023, 123: 106209
- [96] Gao Q, Xia W, Wan Z, et al. Tensor-SVD based graph learning for multi-view subspace clustering//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA, 2020: 3930-3937
- [97] Qin Y, Tang Z, Wu H, et al. Flexible tensor learning for multi-view clustering with Markov chain. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 36(4): 1552-1565
- [98] Xie Y, Tao D, Zhang W, et al. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 2018, 126: 1157-1179
- [99] Chen Y, Xiao X, Zhou Y. Jointly learning kernel representation tensor and affinity matrix for multi-view clustering. *IEEE Transactions on Multimedia*, 2019, 22(8): 1985-1997
- [100] Pu X, Pan B, Che H. Robust low-rank graph multi-view clustering via Cauchy norm minimization. *Mathematics*, 2023, 11(13): 2940
- [101] Chen Y, Wang S, Xiao X, et al. Self-paced enhanced low-rank tensor kernelized multi-view subspace clustering. *IEEE Transactions on Multimedia*, 2021, 24: 4054-4066
- [102] Wu J, Lin Z, Zha H. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 2019, 28(12): 5910-5922
- [103] Chen Y, Xiao X, Zhou Y. Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix. *Pattern Recognition*, 2020, 106: 107441
- [104] Ma S, Liu Y, Liu G, et al. Orthogonal multi-view tensor-based learning for clustering. *Neurocomputing*, 2022, 500: 592-603
- [105] Fu L, Chen Z, Chen Y, et al. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE Transactions on Multimedia*, 2022, 25: 4972-4985
- [106] Long Z, Wang Q, Ren Y, et al. S2mvtc: a simple yet efficient scalable multi-view tensor clustering//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 26213-26222
- [107] Xie D, Gao Q, Yang M. Enhanced tensor low-rank repre-

- sensation learning for multi-view clustering. *Neural Networks*, 2023, 161: 93-104
- [108] Pan B, Li C, Che H. Error-robust multi-view subspace clustering with nonconvex low-rank tensor approximation and hyper-Laplacian graph embedding. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108274
- [109] Zhang D, Huang H, Zhao Q, et al. Generalized latent multi-view clustering with tensorized bipartite graph. *Neural Networks*, 2024, 175: 106282
- [110] Zhao H, Ding Z, Fu Y. Multi-view clustering via deep matrix factorization//*Proceedings of the AAAI Conference on Artificial Intelligence*. San Francisco, USA, 2017: 2921-2927
- [111] Li J, Zhou G, Qiu Y, et al. Deep graph regularized non-negative matrix factorization for multi-view clustering. *Neurocomputing*, 2020, 390: 108-116
- [112] Gunawardena S, Luong K, Balasubramaniam T, et al. DC-CNMF: Deep complementary and consensus non-negative matrix factorization for multi-view clustering. *Knowledge-Based Systems*, 2024, 285: 111330
- [113] Trigeorgis G, Bousmalis K, Zafeiriou S, et al. A deep semi-nmf model for learning hidden representations //*Proceedings of the International Conference on Machine Learning*. Beijing, China, 2014: 3677-3688
- [114] Wang D, Li T, Huang W, et al. A multi-view clustering algorithm based on deep semi-NMF. *Information Fusion*, 2023, 99: 101884
- [115] Wang D, Li T, Deng P, et al. Dnsrf: deep network-based semi-nmf representation framework. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(5): 1-20
- [116] Du G, Zhou L, Yang Y, et al. Deep multiple auto-encoder-based multi-view clustering. *Data Science and Engineering*, 2021, 6(3): 323-338
- [117] Wang R, Li L, Tao X, et al. Trio-based collaborative multi-view graph clustering with multiple constraints. *Information Processing & Management*, 2021, 58(3): 102466
- [118] Fang Z, Du S, Lin X, et al. DBO-Net: Differentiable bi-level optimization network for multi-view clustering. *Information Sciences*, 2023, 626: 572-585
- [119] Xu J, Ren Y, Li G, et al. Deep embedded multi-view clustering with collaborative training. *Information Sciences*, 2021, 573: 279-290
- [120] Xu J, Ren Y, Tang H, et al. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(7): 7470-7482
- [121] Xu K, Tang K, Su Z, et al. Clean and robust multi-level subspace representations learning for deep multi-view subspace clustering. *Expert Systems with Applications*, 2024, 252: 124243
- [122] Xie Y, Lin B, Qu Y, et al. Joint deep multi-view learning for image clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 33(11): 3594-3606
- [123] Wang S, Chen Z, Du S, et al. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(9): 5042-5055
- [124] Huang Z, Hu P, Zhou J T, et al. Partially view-aligned clustering. *Advances in Neural Information Processing Systems*, 2020, 33: 2892-2902
- [125] Zhou R, Shen Y D. End-to-end adversarial-attention network for multi-modal clustering//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2020: 14619-14628
- [126] Qin Y, Pu N, Wu H. Elastic multi-view subspace clustering with pairwise and high-order correlations. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 36(2): 556-568
- [127] Wang Y, Chang D, Fu Z, et al. Consistent multiple graph embedding for multi-view clustering. *IEEE Transactions on Multimedia*, 2021, 25: 1008-1018
- [128] Lu R, Liu J, Zuo X. Attentive multi-view deep subspace clustering net. *Neurocomputing*, 2021, 435: 186-196
- [129] Xiao S, Du S, Chen Z, et al. Dual fusion-propagation graph neural network for multi-view clustering. *IEEE Transactions on Multimedia*, 2023, 25: 9203-9215
- [130] Xia W, Wang S, Yang M, et al. Multi-view graph embedding clustering network: Joint self-supervision and block diagonal representation. *Neural Networks*, 2022, 145: 1-9
- [131] Ou S, Xue Z, Li Y, et al. View-category interactive sharing transformer for incomplete multi-view multi-label learning//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 27467-27476
- [132] Wang R, Li L, Tao X, et al. Contrastive and attentive graph learning for multi-view clustering. *Information Processing & Management*, 2022, 59(4): 102967
- [133] Liu D, Peng S J, Liu X, et al. Inconsistency distillation for consistency: Enhancing multi-view clustering via mutual contrastive teacher-student leaning//*Proceedings of the IEEE International Conference on Data Mining (ICDM)*. Orlando, USA, 2022: 251-258
- [134] Wang S, Lin X, Fang Z, et al. Contrastive consensus graph learning for multi-view clustering. *IEEE/CAA Journal of Automatica Sinica*, 2022, 9(11): 2027-2030
- [135] Zhang J, Sun Y, Guo Y, et al. Multi-view subspace clustering with consensus graph contrastive learning//*Proceedings of the ICASSP 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Seoul, Republic of Korea, 2024: 6340-6344
- [136] Yang S, Liao Z, Chen R, et al. Multi-view fair-augmentation contrastive graph clustering with reliable pseudo-labels.

- Information Sciences, 2024, 674: 120739
- [137] Yan W, Zhang Y, Lv C, et al. Gcfagg: global and cross-view feature aggregation for multi-view clustering//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023: 19863-19872
- [138] Wu S, Zheng Y, Ren Y, et al. Self-weighted contrastive fusion for deep multi-view clustering. IEEE Transactions on Multimedia, 2024, 26: 9150-9162
- [139] Luo C, Xu J, Ren Y, et al. Simple contrastive multi-view clustering with data-level fusion//Proceedings of the International Joint Conference on Artificial Intelligence. Jeju, South Korea, 2024: 4697-4705
- [140] Zou G, Ye Y, Chen T, et al. Learning dual enhanced representation for contrastive multi-view clustering//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia, 2024: 8731-8739
- [141] Xu J, Tang H, Ren Y, et al. Multi-level feature learning for contrastive multi-view clustering//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 16051-16060
- [142] Shu Z, Li B, Mao C, et al. Structure-guided feature and cluster contrastive learning for multi-view clustering. Neurocomputing, 2024, 582: 127555
- [143] Du T, Zheng W, Xu X. Composite attention mechanism network for deep contrastive multi-view clustering. Neural Networks, 2024, 176: 106361
- [144] Bao Y, Zhao W, Zhao Q, et al. Dual contrastive learning for multi-view clustering. Neurocomputing, 2024, 599: 128093
- [145] Pan E, Kang Z. Multi-view contrastive graph clustering. Advances in Neural Information Processing Systems, 2021, 34: 2148-2159
- [146] Trosten D J, Lokse S, Jenssen R, et al. Reconsidering representation alignment for multi-view clustering //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 1255-1265
- [147] Lin F, Bai B, Bai K, et al. Contrastive multi-view hyperbolic hierarchical clustering. arXiv preprint arXiv: 2205.02618, 2022
- [148] Deng X, Huang D, Chen D H, et al. Strongly augmented contrastive clustering. Pattern Recognition, 2023, 139: 109470
- [149] Du G, Zhou L, Li Z, et al. Neighbor-aware deep multi-view clustering via graph convolutional network. Information Fusion, 2023, 93: 330-343
- [150] Fang U, Li J, Lu X, et al. Robust image clustering via context-aware contrastive graph learning. Pattern Recognition, 2023, 138: 109340
- [151] Yin M, Huang W, Gao J. Shared generative latent representation learning for multi-view clustering //Proceedings of the AAAI conference on artificial intelligence. New York, USA, 2020: 6688-6695
- [152] Xu J, Ren Y, Tang H, et al. Multi-VAE: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada, 2021: 9234-9243
- [153] Xu J, Ren Y, Shi X, et al. UNTIE: Clustering analysis with disentanglement in multi-view information fusion. Information Fusion, 2023, 100: 101937
- [154] Cui C, Ren Y, Pu J, et al. A novel approach for effective multi-view clustering with information-theoretic perspective. Advances in Neural Information Processing Systems, 2023, 36: 44847-44859
- [155] Huang W, Yang S, Cai H. Generalized information-theoretic multi-view clustering. Advances in neural information processing systems, 2023, 36: 58752-58764
- [156] Bai R, Huang R, Qin Y, et al. HVAE: A deep generative model via hierarchical variational auto-encoder for multi-view document modeling. Information Sciences, 2023, 623: 40-55
- [157] Wang Q, Ding Z, Tao Z, et al. Partial multi-view clustering via consistent GAN//Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM). Singapore, 2018: 1290-1295
- [158] Wang Q, Ding Z, Tao Z, et al. Generative partial multi-view clustering with adaptive fusion and cycle consistency. IEEE Transactions on Image Processing, 2021, 30: 1771-1783
- [159] Wang Q, Tao Z, Xia W, et al. Adversarial multiview clustering networks with adaptive fusion. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(10): 7635-7647
- [160] Wang Q, Ding Z, Tao Z, et al. Generative partial multi-view clustering. arXiv preprint arXiv:2003.13088, 2020
- [161] Wang H, Zhang W, Ma X. Contrastive and adversarial regularized multi-level representation learning for incomplete multi-view clustering. Neural Networks, 2024, 172: 106102
- [162] Xu C, Liu H, Guan Z, et al. Adversarial incomplete multi-view subspace clustering networks. IEEE Transactions on Cybernetics, 2021, 52(10): 10490-10503
- [163] Li Z, Wang Q, Tao Z, et al. Deep adversarial multi-view clustering network//Proceedings of the International Joint Conference on Artificial Intelligence. Macao, China, 2019: 2952-2958
- [164] Wu J S, Li Y, Gong J X, et al. Collaborative and Discriminative Subspace Learning for unsupervised multi-view feature selection. Engineering Applications of Artificial Intelligence, 2024, 133: 108145
- [165] Tang C, Chen J, Liu X, et al. Consensus learning guided

- multi-view unsupervised feature selection. *Knowledge-Based Systems*, 2018, 160: 49-60
- [166] Liu X, Liu L, Liao Q, et al. One pass late fusion multi-view clustering//*Proceedings of the International Conference on Machine Learning*. Vienna, Austria, 2021: 6850-6859
- [167] Chen M S, Huang L, Wang C D, et al. Multi-view clustering in latent embedding space//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA, 2020: 3513-3520
- [168] Dong X, Zhu L, Song X, et al. Adaptive collaborative similarity learning for unsupervised multi-view feature selection. *arXiv preprint arXiv:1904.11228*, 2019
- [169] Wu B, Zhang Y, Hu B G, et al. Constrained clustering and its application to face clustering in videos//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013: 3507-3514
- [170] Deng L. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 2012, 29(6): 141-142
- [171] Zhao H, Liu H, Fu Y. Incomplete multi-modal visual data grouping//*Proceedings of the International Joint Conference on Artificial Intelligence*. New York, USA, 2016: 2392-2398
- [172] Hu M, Chen S. Doubly aligned incomplete multi-view clustering. *arXiv preprint arXiv:1903.02785*, 2019
- [173] Wen J, Zhang Z, Zhang Z, et al. Generalized incomplete multiview clustering with flexible locality structure diffusion. *IEEE Transactions on Cybernetics*, 2020, 51(1): 101-114
- [174] Wen J, Xu Y, Liu H. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Transactions on Cybernetics*, 2018, 50(4): 1418-1429
- [175] Wen J, Zhang Z, Xu Y, et al. Unified embedding alignment with missing views inferring for incomplete multi-view clustering//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, USA, 2019: 5393-5400
- [176] Wen J, Yan K, Zhang Z, et al. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 2020, 23: 2493-2504
- [177] Yu Y, Sun D. Incomplete multi-view clustering based on dynamic dimensionality reduction weighted graph learning. *IEEE Access*, 2024, 12: 19087-19099
- [178] Xue L, Gao M, Xing C, et al. Ulip: Learning a unified representation of language, images, and point clouds for 3d understanding//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada, 2023: 1179-1189
- [179] Zhang Z, Cao S, Wang Y X. Tamm: Triadapter multi-modal learning for 3d shape understanding//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 21413-21423
- [180] Watson J, Aleotti F, Sayed M, et al. AirPlanes: Accurate plane estimation via 3D-consistent embeddings//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 5270-5280
- [181] Zheng M, Zhang H, Yang H, et al. Neuface: Realistic 3d neural face rendering from multi-view images//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada, 2023: 16868-16877
- [182] Brynte L, Iglesias J P, Olsson C, et al. Learning structure-from-motion with graph attention networks//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 4808-4817
- [183] He X, Sun J, Wang Y, et al. Detector-free structure from motion//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 21594-21603
- [184] Lochman Y, Olsson C, Zach C. Learned trajectory embedding for subspace clustering//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2024: 19092-19102
- [185] Wu W, Zhang W, Hou W, et al. Multi-view clustering with graph learning for scRNA-seq data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023, 20(6): 3535-3546
- [186] Chen F, Zou G, Wu Y, et al. Clustering single-cell multi-omics data via graph regularized multi-view ensemble learning. *Bioinformatics*, 2024, 40(4): btad169
- [187] Wang Z, You R, Han H, et al. Effective binning of metagenomic contigs using contrastive multi-view representation learning. *Nature Communications*, 2024, 15(1): 585
- [188] Shi X, Liang C, Wang H. Multiview robust graph-based clustering for cancer subtype identification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022, 20(1): 544-556
- [189] Shah E, Maji P. Multi-view kernel learning for identification of disease genes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023, 20(3): 2278-2290
- [190] Du Y, Sun H, Abdollahi M J. Toward deep multi-view document clustering using enhanced semantic embedding and consistent context semantics. *Knowledge and Information Systems*, 2025, 67(2): 1073-1100
- [191] Zheng M, Jiang K, Xu R, et al. An adaptive LDA optimal topic number selection method in news topic identification. *IEEE Access*, 2023, 11: 92273-92284
- [192] Saini N, Reddy S M, Saha S, et al. Multi-view multi-objective clustering-based framework for scientific document summarization using citation context. *Applied Intelligence*, 2023, 53(14): 18002-18026
- [193] Li A, Li Y, Shao Y, et al. Multi-view scholar clustering with dynamic interest tracking. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(9): 9671-9684



DING Shi-Fei, Ph. D. , professor.

His research interests include artificial intelligence, pattern recognition, machine learning, data mining.

WEI Xiu-Xi, Ph. D. candidate, associate professor. His research interests include deep learning, clustering analysis.

MEI Yan-Bin, Master candidate. His research interests include deep learning, clustering analysis.

HUANG Hua-Juan, Ph. D. , professor. Her research interests include machine learning, data mining.

Background

In recent years, several review articles have summarized multi-view clustering research from different perspectives, but these articles generally lack a summary of the latest methods. In this paper, we provide a comprehensive analysis and summary of existing MVC methods. First, we introduce common MVC data representation methods and analyze the primary problems and challenges faced by MVC, and list some complex data problems, such as missing data, incremental data, and unaligned data features. Then, according to the different technical mechanisms, we categorize the existing MVC methods into three main categories: (1) heuristic-based MVC methods, (2) discriminative-based MVC methods, and (3) generative-based MVC methods. In this section, we explore in detail the specific strategies and ef-

fectiveness of these approaches in solving common MVC problems. Subsequently, we summarize several representative MVC datasets and compile experimental results of current mainstream MVC approaches. Finally, this paper discusses future research directions worth exploring in the area of MVC. We hope that this paper can provide a clear research framework for researchers to help them better understand the field of MVC and advance its further development.

This work was supported by the National Natural Science Foundation of China General Program (Nos. 62276265, 62206297), the National Natural Science Foundation of China Regional Program (No. 62266007), and the Guangxi Natural Science Foundation of China (No. 2021GXNSFAA220068).