

基于变分自编码器的深度解耦多视图聚类算法

李顺勇^{1,2}, 李青辉¹, 赵兴旺^{3,4}, 陈存⁵, 杨梦瑶¹

1. 山西大学数学与统计学院, 山西 太原 030006;
2. 山西大学复杂系统与数据科学教育部重点实验室, 山西 太原 030006;
3. 山西大学计算机与信息技术学院, 山西 太原 030006;
4. 山西大学计算智能与中文信息处理教育部重点实验室, 山西 太原 030006;
5. 郑州大学数学与统计学院, 河南 郑州 450001

摘要

多视图聚类是无监督学习领域中的一个重要研究方向。现有的多视图聚类方法在处理视图间的信息耦合及高维数据时仍存在诸多的局限性。为此, 提出一种基于变分自编码器的深度解耦多视图聚类算法——DDMVCVA。该算法采用VGG16作为特征提取器, 并在变分自编码器框架中引入了离散潜在变量和连续潜在变量, 以实现对多视图数据的有效解耦和融合。其中, 离散潜在变量负责提取各视图共有的聚类特征, 而连续潜在变量则用于捕捉每个视图特有的信息。通过控制互信息容量, DDMVCVA能够在保持视图间共享信息的同时, 有效地分离出视图特有的表示。基于解耦后的表示, 使用K-means算法进行聚类, 生成最终的聚类标签。在6个真实数据集上的实验结果显示, DDMVCVA的聚类性能优于对比的10种多视图聚类方法, 验证了其有效性。

关键词

多视图聚类; 变分自编码器; 深度聚类; 解耦表示学习; VGG

中图分类号: TP181

文献标志码: A

doi:10.11959/j.issn.2096-0271.2026030

Deep decoupling multi-view clustering algorithm based on variational autoencoder

Li Shunyong^{1,2}, Li Qinghui¹, Zhao Xingwang^{3,4}, Chen Cun⁵, Yang Mengyao¹

1. School of Mathematics and Statistics, Shanxi University, Taiyuan 030006, China
2. Key Laboratory of Complex Systems and Data Science of Ministry of Education, Shanxi University, Taiyuan 030006, China
3. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
4. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, China
5. School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China

Abstract

Multi-view clustering represents a significant research direction in the field of unsupervised learning. However, existing multi-view clustering methods were found to have several limitations when handling information coupled across views and high-dimensional data. To address these issues, a deep disentangled multi-view clustering algorithm based on

variational autoencoder, named DDMVCVA, was proposed. VGG16 was employed as the feature extractor, while both discrete and continuous latent variables were incorporated into the variational autoencoder framework to achieve effective disentanglement and fusion of multi-view data. Specifically, the discrete latent variables were designed to capture clustering features common to all views, whereas the continuous latent variables were used to extract view-specific information. By controlling the mutual information capacity, DDMVCVA was able to preserve shared information among views while effectively separating view-specific representations. Finally, based on the disentangled representations, the K-means algorithm was applied to generate the final cluster labels. Experimental results on six real-world datasets demonstrated that DDMVCVA outperformed ten comparative multi-view clustering methods, confirming its effectiveness.

Key words

multi-view clustering, variational autoencoder, deep clustering, decoupling representation learning, VGG

0 引言

随着数据采集与存储技术的飞速发展,数据获取渠道日益多元,研究对象的描述逐渐从单一维度向多维度延伸。例如,通过多机位部署实现物体的多角度拍摄,或借助多种特征提取算法获取图像的差异化特征信息。此类源自异构域、聚焦同一对象的数据集,被定义为多视图数据。如何有效挖掘多视图数据的内在结构,已成为无监督学习领域的核心挑战之一,而多视图聚类作为应对这一挑战的关键技术,其核心在于融合不同视图的异构信息以揭示底层数据分布规律,从而实现相较于单视图聚类更精准的样本分类。

依据技术原理,现有多视图聚类方法大致分为5类:基于非负矩阵分解的方法^[1]、基于子空间学习的方法^[2]、基于多核学习的方法^[3]、基于图学习的方法^[4-5]和基于深度学习的方法^[6-7],前4类统称为传统方法,其核心瓶颈是较高的算法复杂度导致高维数据处理效率偏低。近年来,深度聚类^[8-9]利用深度神经网络自动学习数据的高层特征表示,显著提升了聚类的性能,但现有深度方法多通过挖掘跨视图共享表

示或融合多源特征学习聚类特征。此类方法通常假设跨视图一致性与单视图特异性共同促进聚类效果,却将两类约束施加在同一特征空间,忽视了多视图表示纠缠性对聚类性能的潜在负面影响。例如,Zhang等^[10]提出通过多分支共享特征聚合模块将视图特征压缩至共享空间,并利用双重对比学习对齐特征与语义信息。尽管该方法在特征融合与一致性学习上取得进展,但其核心设计仍将视图特异性过滤与跨视图对齐约束作用于同一潜在空间,导致不同视图的私有信息与共享信息在特征重构中存在隐性耦合。这种表示空间的纠缠性使视图特异性信息与共享特征解耦不充分,共享特征的判别性受残留噪声持续影响。Huang等^[11]虽通过自适应加权模块动态调整视图贡献度,并采用动态聚类扩散增强簇间分离度,但其全局特征空间仍通过线性加权融合视图私有信息与共享语义,低质量视图噪声经权重分配对全局表征构建产生持续干扰。

解耦表示学习^[12-16]通过分解数据潜在变量为独立因子以表征不同属性特征,有效增强模型可解释性与泛化能力,对分离多视图共享与特有信息具有重要意义。构建解耦异构表示空间以分别建模多视图共性与单视图特性^[17-18],已成为深度多视图

聚类引入解耦表示学习的关键切入点。

综上，现有方法在处理视图间信息耦合问题及高维数据时存在显著的局限性，急需一种兼具高效性与解耦能力的多视图聚类算法。

针对上述问题，本文提出一种基于变分自编码器的深度解耦多视图聚类算法（deep decoupling multi-view clustering algorithm based on variational auto-encoder，DDMVCVA）。该算法依托VAE^[19]（变分自编码器）的生成与解耦特性，通过学习解耦的视图表示实现多视图数据的高效聚类。具体而言，DDMVCVA采用VGG16作为特征提取器，在变分自编码器框架中引入离散潜在变量和连续潜在变量，实现多视图数据的有效解耦与融合。其中，离散潜在变量用于提取各视图共有的聚类特征，连续潜在变量用于捕捉每个视图特有的信息。通过控制互信息容量，DDMVCVA能够在保持视图间共享信息的同时，有效分离视图特有的表示。最后，基于解耦后的表示，采用K-means算法生成最终聚类标签。具体算法流程将在第1

节结合模型结构进行详细介绍。

本文的主要贡献如下。

- （1）提出一种基于变分自编码器的深度解耦多视图聚类算法，通过解耦视图表示实现多视图数据的高效聚类。
- （2）构建基于VGG16的特征提取框架，在变分自编码器中引入离散与连续潜在变量，实现多视图数据的解耦与融合。
- （3）设计互信息容量控制方法，在保留视图共享信息的同时有效分离视图特有表示。
- （4）通过在6个真实数据集上的实验，验证DDMVCVA的聚类性能优于对比算法，表明其处理复杂多视图数据的优越性。

1 DDMVCVA

本节详细介绍本文提出的DDMVCVA，DDMVCVA整体框架如图1所示：该框架以多视图数据作为输入，首先利用VGG16对各视图进行特征提取；随后通过编码器分别学习离散潜在变量（用

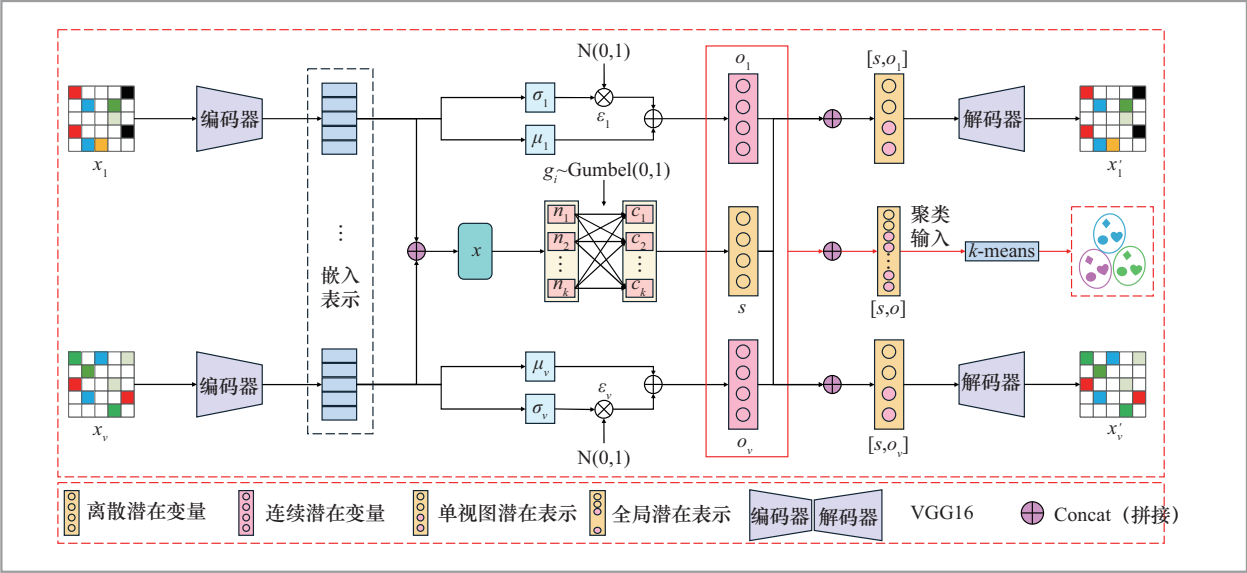


图1 DDMVCVA整体框架

于捕获视图间共享的聚类特征)和连续潜在变量(用于表达各视图特有信息);两类变量经拼接后输入解码器进行视图重构,并通过互信息容量控制机制实现有效解耦;最终基于解耦后的联合表示执行K-means聚类,输出聚类结果。整个流程在VAE的框架下,实现了多视图数据的深度解耦与聚类。

现将问题描述如下:给定具有 V 个视图的图像数据集 $\{x_1^1, x_1^2, \dots, x_1^V\}_{i=1}^N$,其中, $x_i^j \in \mathbb{R}^{d_v \times N}$ 表示第 i 个样本视图维度为 d_v 的第 j 个视图, N 为样本总数。DDMVCVA的目标是在VAE框架下,学习解耦的多视图潜在表示,以融合视图间的共性与特性,并据此将样本划分至 K 个簇。

本研究选取VGG16作为特征提取器,主要因其链式结构天然适配特征解耦且层级特征表达透明。相较于ResNet等多分支网络,其连续卷积层有效规避了跳跃连接引发的特征耦合,促进视图私有信息(浅层细节特征)与共享信息(深层语义特征)的分离。研究表明^[20],此类无分支虽FLOPs不占优,但凭借高内存访问效率和算子简化,实际推理更快,更适用于解耦任务。图像融合结果亦显示^[21],VGG16层级特征可清晰映射至不同物理属性,如结构、纹理,便于解耦分析;而多分支网络的复合特征易模糊特征来源,增加解释难度。综上,VGG16的结构简洁性、特征层级性及高效工程优势,使其成为解耦表示学习及多视图大数据处理的理想选择。

1.1 方法和模型

为实现多视图数据的解耦表示,我们定义联合生成模型:

$$p(x_v, o_v, s) = p(x_v | o_v, s) p(o_v) p(s) \quad (1)$$

其中, s 是所有视图共享的离散潜在变量,

编码视图间共通信息; o_v 是仅与第 v 个视图相关的连续潜在变量,编码该视图的特有信息。给定所有视图数据 $\{x_v\}$ 。因此, s 和 o_v 的后验分布分别为 $p(s|\{x_v\})$ 和 $p(o_v|x_v)$ 。考虑在VAE中计算后验分布的积分是不可行的,故本文使用参数化的 $q_\phi(s|\{x_v\})$ 和 $q_{\phi_v}(o_v|x_v)$ 来近似真实的后验分布。

在离散潜在变量的推理过程中,首先,利用VGG16作为特征提取器,获取各个视图的嵌入表示;其次,通过Concat操作融合这些视图的嵌入,旨在学习其共通信息;最后,利用全连接层(设置 K 个神经元记为 $n=\{n_1, n_2, \dots, n_K\}$)进一步处理融合后的嵌入,以提取离散潜在变量 s 。此外,为便于表示聚类分配并解决离散变量不可微问题,本文选择了独立均匀Gumbel-Softmax分布作为 s 的先验分布。令:

$$p(s) = p(s_1) p(s_2) \cdots p(s_K) \quad (2)$$

其中, $p(s_k) \sim \text{Gumbel}(0, 1)$ 。因此,近似后验 $q_\phi(s|\{x_v\})$ 可以表示为:

$$q_\phi(s|\{x_v\}) = \prod_{k=1}^K q_{\phi_k}(s_k|\{x_v\}) \quad (3)$$

根据Gumbel-Max重参数化技巧^[22],可以进一步得到以下表达式:

$$q_{\phi_k}(s_k|\{x_v\}) = G(n) = \frac{\exp((\log n_k + g_k)/\tau)}{\sum_{i=1}^K \exp((\log n_i + g_i)/\tau)} \quad (4)$$

其中, $g_k \sim \text{Gumbel}(0, 1)$, τ 是控制缩放的温度参数。

在连续潜在变量的推理过程中,本文假设不同视图的信息是连续的,因此选择标准正态分布作为连续潜在变量的先验分布,即 $p(\alpha_v) \sim N(0, 1)$ 。 $q_{\phi_v}(o_v|x_v)$ 由高斯分布参数化有:

$$q_{\phi_v}(o_v|x_v) = \prod_{i=1}^{O_v} q_{\phi_{vi}}(o_{vi}|x_v) \quad (5)$$

根据重参数化技巧^[23],可以得到以下

等式:

$$q_{\phi_v}(o_{iv}|x_v) = N(\mu_{vi}, (\sigma_{iv})^2) = \mu_{vi} + \sigma_{iv} \epsilon_{vi} \quad (6)$$

其中, $\epsilon_{vi} \sim N(0, 1)$ 。 μ_{vi} 和 σ_{iv} 由神经网络参数化, 其输入是第 v 个视图的嵌入。

在视图重构过程中, 每个视图的隐变量由离散潜在变量 s 和连续潜在变量 o_v 组成, 这两部分通过 Concat 操作融合, 以重构相应的视图。第 v 个视图的解码器进一步可以表示为:

$$\hat{x}_v = p_{\theta_v}(x_v|o_v, s) \quad (7)$$

其中, 参数 $\phi\{\phi_1, \phi_2, \dots, \phi_V\}$ 和 $\theta\{\theta_1, \theta_2, \dots, \theta_V\}$ 与解码器的参数共享。

2.2 变分下界

本文提出模型的对数似然可以表示为:

$$\sum_{v=1}^V \log p(x_v) = \sum_{v=1}^V \log \int_{o_v, s} p(x_v, o_v, s) d o_v \quad (8)$$

根据数学期望的定义, VAE 的证据下界可以表示为:

$$\sum_{v=1}^V \mathcal{L}_{\text{ELBO}}(x_v) = \sum_{v=1}^V \mathbb{E}_{q(o_v, s|\{x_v\})} \left[\log \frac{p(x_v, o_v, s)}{q(o_v, s|\{x_v\})} \right] \quad (9)$$

其中, $\mathcal{L}_{\text{ELBO}}(x_v)$ 是第 v 个视图的证据下界。此时, 最大化似然函数等价于最大化 ELBO。给定 $p(x_v, o_v, s) = p(x_v|o_v, s)p(o_v, s)$, 每个视图的 ELBO 可以写成:

$$\mathcal{L}_{\text{ELBO}}(x_v) = \mathbb{E}_{q(o_v, s|\{x_v\})} [\log p(x_v|o_v, s)] - D_{\text{KL}}(q(o_v, s|\{x_v\}) \| p(o_v, s)) \quad (10)$$

假设离散潜在变量和连续潜在变量是条件独立的, 即 $q(o_v, s|\{x_v\}) = q(o_v|x_v)q(s|\{x_v\})$, 且先验 $p(o_v, s) = p(o_v)p(s)$ 。则 KL 散度可以分解为两部分:

$$D_{\text{KL}}(q(o_v, s|\{x_v\}) \| p(o_v, s)) = D_{\text{KL}}(q(o_v|x_v) \| p(o_v)) + D_{\text{KL}}(q(s|\{x_v\}) \| p(s)) \quad (11)$$

通过上述方式, 离散潜在变量和连续潜在变量的 KL 散度项得以分离, 从而实现了视图共享表示和视图私有表示的解耦。对于第 v 个视图, 最大化目标变为:

$$\mathcal{L}_{\text{ELBO}}(x_v) = \mathbb{E}_{q(o_v, s|\{x_v\})} [\log p(x_v|o_v, s)] - D_{\text{KL}}(q(o_v|x_v) \| p(o_v)) - D_{\text{KL}}(q(s|\{x_v\}) \| p(s)) \quad (12)$$

1.3 解耦表示学习

本文分别为离散潜在变量和连续潜在变量的 KL 散度项定义了受控容量 C_s 和 C_o 。第 v 个视图的 ELBO 可以表示为:

$$\mathcal{L}_{\text{ELBO}}(x_v) = \mathbb{E}_{q(o_v, s|\{x_v\})} [\log p(x_v|o_v, s)] - \beta |D_{\text{KL}}(q(s|\{x_v\}) \| p(s)) - C_s| - \beta |D_{\text{KL}}(q(o_v|x_v) \| p(o_v)) - C_o| \quad (13)$$

其中, β 是权衡系数。特别地, 当 $p(s)$ 是一致分类分布时, 视图离散潜在变量的 KL 散度是有界的:

$$D_{\text{KL}}(q(s|\{x_v\}) \| p(s)) = \sum_{i=1}^K q_i \log \frac{q_i}{1/K} = -H(q) + \log K \leq \log K \quad (14)$$

其中, H 表示熵。基于此, 本文设定 $C_s = \log K$, 以控制离散潜在变量 s 中编码的变分信息的最大容量。考虑到不同视图的数据重构损失规模不同, 本文进一步在 β 上引入权重来平衡所有视图的解耦。对于第 v 个视图, 权重计算如下:

$$\beta_v = \beta \frac{\mathbb{E}_{q(o_v, s|\{x_v\})} [\log p(x_v|o_v, s)]}{\max_v \mathbb{E}_{q(o_v, s|\{x_v\})} [\log p(x_v|o_v, s)]} \quad (15)$$

最终, 总损失函数包含 3 个部分:

$$\begin{aligned} \mathcal{L}_{\text{loss}} = & - \sum_{v=1}^V \mathcal{L}_{\text{ELBO}}(x_v) = \\ & \sum_{v=1}^V \beta_v |D_{\text{KL}}(q(s|\{x_v\}) \| p(s)) - C_s| + \\ & \sum_{v=1}^V \beta_v |D_{\text{KL}}(q(o_v|x_v) \| p(o_v)) - C_o| - \\ & \sum_{v=1}^V \mathbb{E}_{q(o_v, s|\{x_v\})} [\log p(x_v|o_v, s)] \end{aligned} \quad (16)$$

其中,第1项和第2项分别优化以学习解耦的视图共享表示和视图私有表示。第3项为似然项,优化该项以保持VAE的重建质量。DDMVCVA如算法1所示。

算法1: DDMVCVA

输入: 多视图数据 $\{x_1, x_2, \dots, x_V\}_{i=1}^N$, 簇的数量 K , 权衡系数 β , 迭代轮次 epoch, 最大容量 $C_s = \log K$

输出: 聚类结果

Step 1: 随机初始化参数 ϕ 和 $\{\phi_v, \theta_v\}_{v=1}^V$;

Step 2: While $i \leq \text{epoch}$ do

Step 3: 根据编码器计算 $\{t_k\}_{k=1}^K$ 和 $\{\mu_v, \sigma_v\}_{v=1}^V$;

Step 4: 根据式(4)更新 $q(s|x_v)$;

Step 5: 根据式(6)更新 $q(o_v|x_v)$;

Step 6: 根据解码器生成 $\{p_{\theta_v}(x_v|o_v, s)\}_{v=1}^V$;

Step 7: 根据式(16)更新参数 ϕ 和 $\{\phi_v, \theta_v\}_{v=1}^V$;

Step 8: End While

Step 9: 将 s 和 $\{o_v\}_{v=1}^V$ 拼接得到全局潜在表示 $[s, o]$;

Step 10: 利用 K-means 对 $[s, o]$ 聚类得到聚类结果。

2 结果与分析

本节通过实验对DDMVCVA进行了系统性评估。在6个多视图数据集上,与10种多视图聚类方法进行对比,并采用准确率(ACC)、标准化互信息(NMI)、调整后兰德指数(ARI)3个指标来评价聚类性能。

2.1 数据集

实验选取了多视图聚类领域中广泛使

用的6个真实公开数据集,分别为MNIST、Fashion-MNIST、USPS、ORL、COIL10和COIL20。各数据集信息具体说明如下。

(1) MNIST: 经典手写数字数据集,包含0~9共10类手写数字样本,数据来源为美国国家标准与技术研究院(NIST)的手写数字数据库整理所得,本文从中随机抽取样本以构建视图数为2的多视图数据。

(2) Fashion-MNIST: 用于替代MNIST的衣物图像数据集,涵盖10类常见衣物类别,由Zalando公司相关研究团队公开提供,本文构建视图数为3。

(3) USPS: 手写数字识别研究专用数据集,包含0~9类手写数字样本,来源于美国邮政服务局(USPS)收集的手写邮件地址数字图像,本文构建视图数为2。

(4) ORL: 人脸图像数据集,包含40位受试者的面部图像,由英国剑桥大学工程系机器视觉实验室采集构建,视图数为2。

(5) COIL10与COIL20: 物体图像数据集,分别包含10类和20类常见物体的多角度拍摄图像,由美国哥伦比亚大学计算机科学系公开发布。本文将不同拍摄角度视为不同视图,视图数设置为3。

上述数据集的概况信息汇总于表1。

2.2 实验细节

DDMVCVA在NVIDIA 3090 GPU和PyTorch 1.9.0的环境中运行,特征提取部分采用VGG16。编码器的网络结构参考了文献[24]的设计,最终映射得到10维的潜在变量。解码器与编码器相对称,网络配置相反,两者采用的温度参数 τ 均为0.67。模型训练中使用学习率为0.0001的Adam对模型进行200轮的训练。 β 设

为 30。离散潜在变量的最大控制容量设置为 5。最后，本文提出的算法在所有实验中均独立运行了 5 次，并报告了各个指标在每个数据集上的平均结果。

2.3 对比方法

为全面评估DDMVCVA的性能，本文选取 10 种方法开展对比分析，包括跨视图匹配聚类^[25] (COMIC, 2019)、均匀流形逼近与投影加高斯混合模型^[26] (UMAP+GMM, 2020)、基于协同训练的深度嵌入多视图聚类^[27] (DEMVC, 2021)、通过对比预测的不完全多视图聚类^[28] (Completer, 2021)、基于分布对齐的自适应特征投影深度不完全多视图聚类^[29] (APADC, 2023)、集成快速多视图聚类^[30] (FastMICE, 2023)、多视图深度嵌入聚类^[31] (MDEC, 2023)、分布一致性保持深度嵌入聚类^[32] (DCP-DEC, 2023)、跨视图锚图潜在信息引导单步多视图模糊聚类^[33] (OMVFC-LICAG, 2024) 及高阶随机游走解耦对比多视图聚类^[34] (DIVIDE, 2024)。其中，UMAP+

表1 数据集信息				
数据集	类别数	样本数	视图数	图像大小
MNIST	10	70 000	2	28像素×28像素
Fashion-MNIST	10	70 000	3	28像素×28像素
USPS	10	11 000	2	16像素×16像素
ORL	40	400	2	32像素×32像素
COIL10	10	720	3	不同尺寸
COIL20	20	1 440	3	不同尺寸

GMM 为传统单视图聚类方法；COMIC 和 OMVFC-LICAG 属于传统多视图聚类方法；其余方法均为深度学习多视图聚类方法。

2.4 实验结果

为验证 DDMVCVA 的有效性，本文在 6 个基准数据集上进行了实验，聚类结果比较见 表 2、表 3。整体而言，DDMVCVA 的优势显著，且对比方法的性能瓶颈进一步凸显了其解耦设计的必要性与优越性。

表2 在MNIST、USPS和Fashion-MNIST数据集上的聚类结果比较									
算法	MNIST			USPS			Fashion-MNIST		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
COMIC (2019)	0.581	0.745	0.547	0.547	0.721	0.523	0.578	0.642	0.436
UMAP+GMM (2020)	0.916	0.866	0.820	0.811	0.900	0.800	0.589	0.651	0.520
DEMVC (2021)	0.908	0.915	0.885	0.903	0.898	0.912	0.726	<u>0.864</u>	0.713
Completer (2021)	0.825	0.906	0.810	0.972	0.945	0.939	0.835	0.862	0.762
APADC (2023)	0.950	0.910	0.889	<u>0.984</u>	<u>0.959</u>	<u>0.964</u>	0.819	0.843	0.701
FastMICE (2023)	0.929	0.912	0.901	0.955	0.927	0.926	0.873	0.841	<u>0.763</u>
MDEC (2023)	<u>0.985</u>	0.908	0.895	0.980	0.809	0.893	0.705	0.677	0.655
DCP-DEC (2023)	—	—	—	0.888	0.810	0.784	0.610	0.580	0.550
OMVFC-LICAG (2024)	0.650	0.664	0.677	0.648	0.624	0.697	0.750	0.773	0.640
DIVIDE (2024)	0.979	<u>0.948</u>	<u>0.955</u>	0.962	0.919	0.920	0.838	0.859	0.715
DDMVCVA	0.999	0.997	0.998	0.999	0.998	0.995	<u>0.839</u>	0.866	0.772

注:最好的结果用粗体标注,次优结果用下划线标注。

表3 在COIL20、COIL10和ORL数据集上的聚类结果比较

算法	COIL20			COIL10			ORL		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
COMIC (2019)	0.495	0.769	0.309	0.736	0.769	0.648	0.560	0.743	0.399
UMAP+GMM (2020)	—	—	—	—	—	—	0.562	0.752	0.423
DEMVC (2021)	0.815	0.904	0.816	0.740	0.774	0.656	0.570	0.764	0.416
Completer (2021)	0.826	0.928	0.816	<u>0.976</u>	<u>0.961</u>	<u>0.948</u>	0.564	0.751	0.351
APADC (2023)	0.917	0.921	0.841	0.780	0.800	0.705	0.575	0.775	0.334
FastMICE (2023)	<u>0.947</u>	<u>0.976</u>	<u>0.946</u>	0.843	0.920	0.887	<u>0.730</u>	<u>0.870</u>	0.502
MDEC (2023)	—	—	—	0.932	0.933	0.924	0.686	0.832	<u>0.651</u>
DCP-DEC (2023)	—	—	—	—	—	—	0.707	0.833	0.413
OMVFC-LICAG (2024)	0.725	0.792	0.758	—	—	—	0.726	0.852	0.518
DIVIDE (2024)	0.884	0.925	0.847	—	—	—	0.723	0.846	0.569
DDMVCVA	0.999	0.998	0.998	0.997	0.994	0.993	0.800	0.903	0.727

注:最好的结果用粗体标注,次优结果用下划线标注。

传统方法的性能局限在 6 个数据集上表现突出。传统单视图聚类方法 UMAP+GMM 虽凭借降维建模在 MNIST (ACC=0.916)、USPS (ACC=0.811) 数据集上优于部分传统多视图聚类方法,但因完全舍弃视图互补信息,在 ORL (ACC=0.562) 人脸数据集上性能显著下降,暴露了单源数据建模的天然缺陷,也印证了多视图聚类挖掘跨视图协同信息的核心价值。在传统多视图聚类方法中,COMIC 依赖线性特征融合与传统图模型,无法有效建模高维数据的非线性耦合关系,在所有数据集上的聚类性能始终处于较低水平,与其他对比算法存在明显差距;OMVFC-LICAG 虽引入锚图潜在信息引导机制,但其通过线性加权混合视图信息构建全局特征空间,低质量视图噪声持续干扰聚类决策,在 MNIST (ACC=0.650)、ORL (ACC=0.726) 数据集上的表现显著低于 DDMVCVA 等深度学习方法,且 ORL 数据集上的性能进一步低于其在 Fashion-MNIST 数据集上的表现,反映出非深度学习框架在高维数据解耦与复杂特征建模上的局限性。

与传统方法相比,深度学习方法通过深度神经网络的特征抽象能力实现了性能提升,但在视图解耦层面仍存在共性短板。DEMVC 借助协同训练机制对齐跨视图特征,但其共享特征空间中视图特异性信息混杂,导致在 USPS (ACC=0.903)、COIL20 (ACC=0.815) 数据集上性能提升受限;Completer 通过对比预测增强特征一致性,却未显式分离视图私有信息,在 Fashion-MNIST (NMI=0.862)、ORL (ACC=0.564) 数据集上的核心指标显著落后于 DDMVCVA。近期提出的 MDEC、DIVIDE 等方法虽尝试通过分布对齐或高阶随机游走优化聚类结构,但因未在特征空间中构建相互独立的共性与特性子空间,表征的纠缠性制约了性能的进一步提升:MDEC 在 USPS (NMI=0.809)、COIL10 (ACC=0.932) 数据集上的指标与 DDMVCVA (USPS NMI=0.998、COIL10 ACC=0.997) 差距显著;DIVIDE 在 COIL20 (ACC=0.884)、ORL (ACC=0.723) 数据集上的 ACC 分别低于 DDMVCVA (COIL20 ACC=0.999、ORL ACC=0.800) 11.5、7.7 个百分点,反映

出共享特征中残留噪声对判别性的持续影响。此外，集成类方法 FastMICE 虽在 COIL20（ACC=0.947）、ORL（ACC=0.730）数据集上表现优于多数对比方法，但其未涉及视图解耦设计，在 COIL20 数据集上的 ACC 仍比 DDMVCVA 低 5.2 个百分点，未能突破表征耦合瓶颈。

DDMVCVA 的性能优势源于其解耦表示学习的核心设计：离散潜在变量（ s ）通过联合推理精准捕获跨视图共性聚类特征，连续潜在变量（ o ）有效过滤视图特异性噪声，辅以互信息容量调控机制动态平衡共享信息与私有信息，使其在不同特征分布的数据集上均展现强适配性。在类间边界清晰的 MNIST、USPS、COIL10、COIL20 数据集上，DDMVCVA 的 ACC 均接近理论上限，NMI、ARI 同步达到 0.99 以上，体现对强区分度特征的精准捕捉能力；在视图差异显著的 Fashion-MNIST（NMI=0.866）与人脸特征复杂的 ORL（ACC=0.800、NMI=0.903）数据集上，其核心指标仍领先所有对比方法，其中在

ORL 数据集上的 ACC 较次优方法 FastMICE（ACC=0.730）提升 7.0 个百分点，印证解耦设计对复杂数据分布的强适应性。

综上，对比实验表明：传统方法受限于线性建模与单源信息利用，深度学习方法则因视图解耦不充分难以突破表征纠缠瓶颈；而 DDMVCVA 通过离散与连续潜在变量的解耦设计，搭配互信息容量调控机制，有效解决了视图间信息耦合问题，在 6 个数据集上的表现充分验证了其有效性与泛化能力。

2.5 消融实验

为了进一步分析离散潜在变量和连续潜在变量对 DDMVCVA 的贡献，并全面说明本文方法设计的合理性和有效性，本节设计了多组消融实验。在 MNIST、Fashion-MNIST、USPS、ORL、COIL10 和 COIL20 数据集上损失函数的消融研究结果见表 4 和表 5，“√”表示 DDMVCVA

表 4 在 MNIST、Fashion-MNIST 和 USPS 数据集上损失函数的消融研究

方法	损失项			MNIST		Fashion-MNIST		USPS	
	L_s	L_o	L_s+L_o	ACC	NMI	ACC	NMI	ACC	NMI
A	—	—	—	0.407	0.406	0.611	0.673	0.861	0.854
B	√	—	—	0.440	0.480	0.611	0.673	0.861	0.854
C	—	√	—	0.826	0.765	0.715	0.727	0.865	0.820
D	—	—	√	0.999	0.997	0.839	0.866	0.999	0.998

表 5 在 ORL、COIL10 和 COIL20 数据集上损失函数的消融研究

方法	损失项			ORL		COIL10		COIL20	
	L_s	L_o	L_s+L_o	ACC	NMI	ACC	NMI	ACC	NMI
A	—	—	—	0.610	0.816	0.722	0.800	0.791	0.858
B	√	—	—	0.610	0.816	0.722	0.800	0.791	0.858
C	—	√	—	0.602	0.804	0.661	0.743	0.813	0.860
D	—	—	√	0.800	0.903	0.997	0.994	0.999	0.998

模型包含该模块。方法A代表仅使用基础VAE模块的基线模型，而方法D则代表使用了全部模块的完整DDMVCVA模型。方法B和方法C分别代表在基线模型的基础上，依次增加了离散潜在变量（ s ）和连续潜在变量（ o ）。实验结果验证了以下结论。

（1）离散潜在变量（ s ）在一致性学习方面做出了显著贡献：当仅引入离散潜在变量（ s ）时，如方法B所示，仅MNIST数据集ACC从0.407提升至0.440，其余无变化。因该变量旨在提取多视图共享聚类特征，MNIST视图共享信息一致、噪声少故作用凸显；其他数据集或特异性信息占比高、或共享特征复杂，增益未显，但验证了其特定场景有效性。

（2）连续潜在变量（ o ）在互补性学习方面起到了重要作用：当仅引入连续潜在变量（ o ）时，如方法C所示，ACC同样有所提高。例如，在MNIST数据集上，ACC从0.407提升至0.826，显示出连续潜在变量对于捕捉每个视图特有的信息至关重要，有助于区分不同视图的独特特征。

（3）整合离散和连续潜在变量的完整模型（方法D）在所有数据集上均取得最佳性能。特别地，在MNIST数据集上，相较于仅使用连续潜在变量的方法C，完整模型的ACC从0.826大幅提升至0.999。这一显著差异主要归因于2个因素。

①协同效应：离散和连续潜在变量的结合不仅能够充分挖掘多视图数据中的共享信息，还能有效地分离视图特有特征，从而提供更全面的数据表示。

②互信息容量控制：通过精确调控互信息容量，完整模型能够在保持视图间共享信息的同时，最大化视图特有信息的利用效率，进一步优化聚类效果。

2.6 参数敏感性分析

本节在6个数据集上对DDMVCVA中的参数进行敏感性分析。具体来说，算法包含连续潜在变量的权衡系数 β 和最大控制容量 C_o 。这2个参数。为系统地评估这2个参数对算法性能的影响，令 β 的取值范围为{10, 20, 30, 40, 50}， C_o 的取值范围为{3, 4, 5, 6, 7}，共进行了25组对比实验，参数敏感性分析如图2所示。在6个数据集上，参数 β 和 C_o 的变化对算法性能影响较小，由图2可见DDMVCVA在取值范围内对参数不敏感，说明DDMVCVA的性能较为稳定。

2.7 收敛性分析

收敛性分析如图3所示，展示了DDMVCVA在6个不同类型数据集上的收敛曲线。从图中可以看出，随着迭代轮次的增加，模型的损失函数值逐渐下降并趋于稳定。尽管在初始迭代中存在一定的波动，但最终收敛且趋于稳定。具体来说，在MNIST数据集上，视图1和视图2的损失函数值在前30轮迭代内迅速降低，随后缓慢下降并在50轮迭代后趋于平稳。类似地，在其余5个数据集上，各视图的损失函数值也在训练初期迅速下降，最终均在大约150轮迭代处达到稳定状态。

2.8 聚类可视化分析

为了直观展示DDMVCVA在多视图聚类任务中的性能，本节采用t-SNE方法将其在COIL10和USPS数据集上的聚类结果进行了可视化，具体如图4和图5所示。图4(a)~图4(c)分别展示了在COIL10数据集上迭代20轮、100轮和200轮时的聚类效果，同理，图5分别展示了在USPS

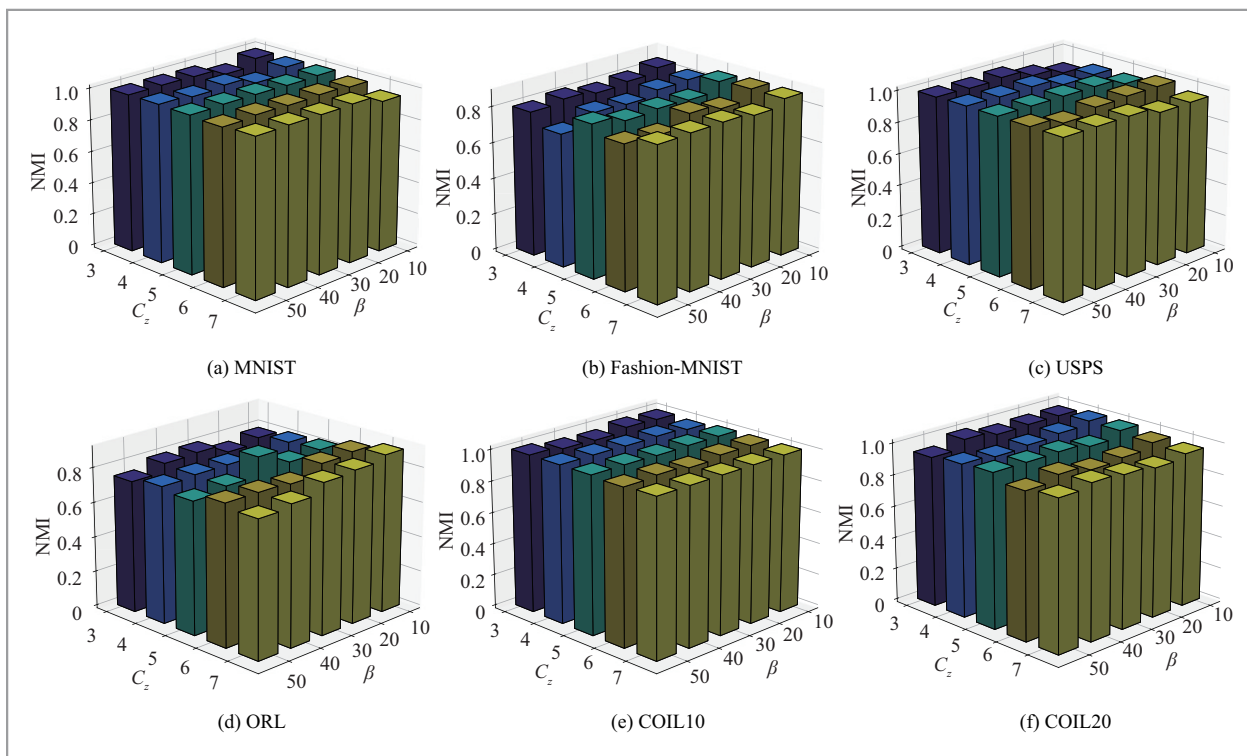


图2 参数敏感性分析

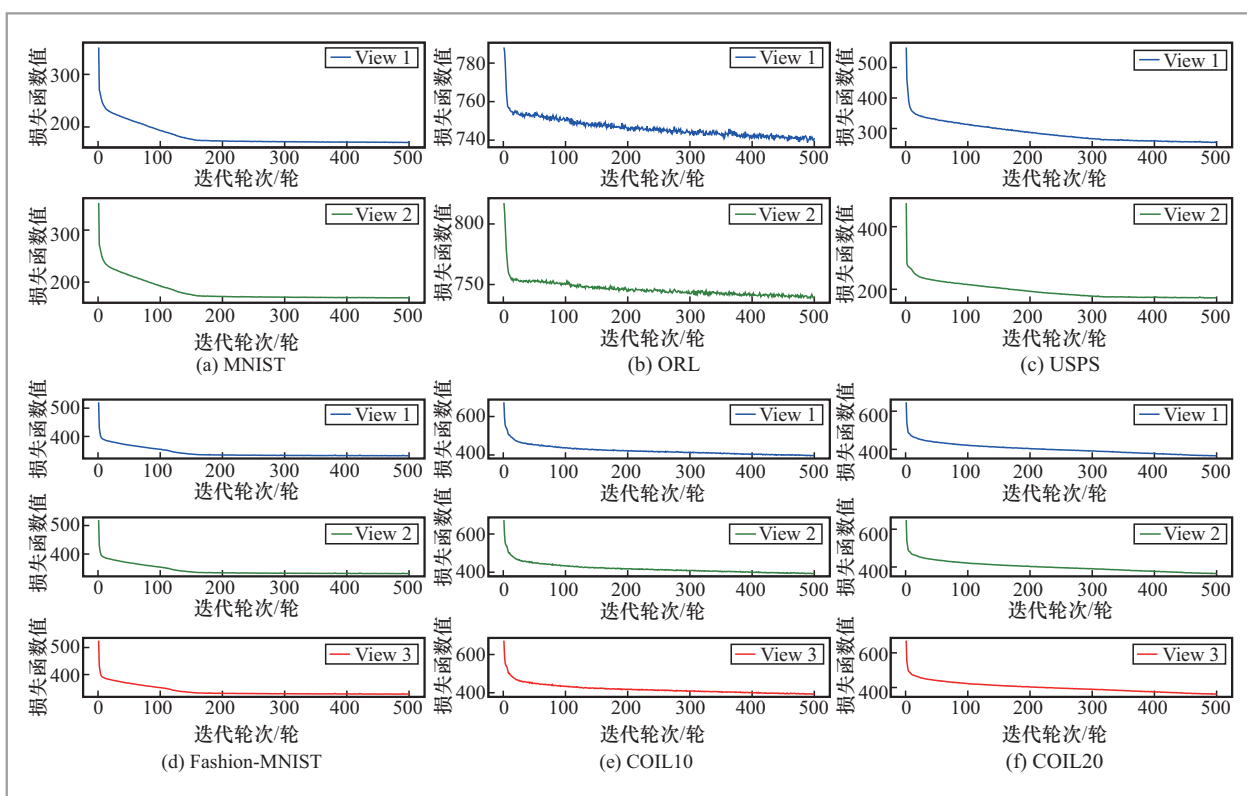


图3 收敛性分析

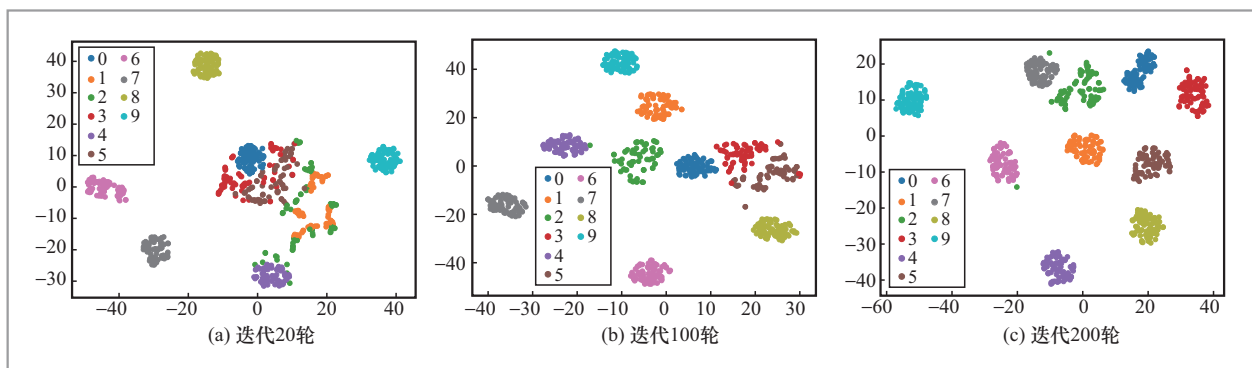


图4 在COIL10数据集上t-SNE可视化

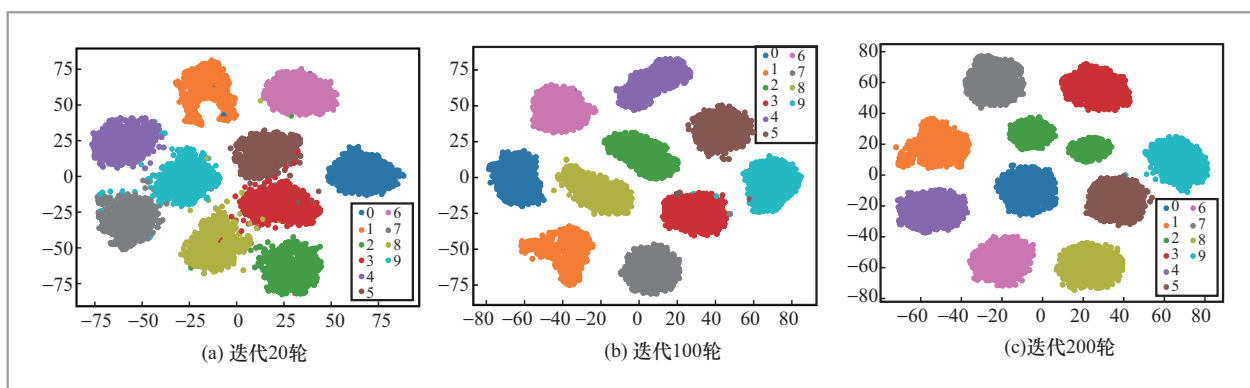


图5 在USPS数据集上t-SNE可视化

数据集上迭代相同轮次时的聚类效果。

在 COIL10 数据集上，迭代 20 轮时，DDMVCVA 初步区分了不同的物体类别，但聚类边界尚不清晰；迭代 100 轮时，聚类边界变得更为明确，不同类别间的分离度显著提高；迭代 200 轮时，聚类效果达到最优状态，各类别之间的分离最为清晰。在 USPS 数据集上，迭代 20 轮时，DDMVCVA 初步区分了不同的字母类别，但存在一定的重叠；迭代 100 轮时，聚类效果显著提升，不同字母类别的分隔更加清晰；迭代 200 轮时，聚类结果达到最佳分离状态，各类别之间的边界最为明确。

可视化分析表明，随着迭代轮次的增

加，DDMVCVA 的聚类性能逐步提升，并在迭代 200 轮时达到最优状态，验证了该方法的有效性和稳健性。

3 结束语

本文提出了一种基于变分自编码器的深度解耦多视图聚类算法——DDMVCVA。与现有方法不同，DDMVCVA 通过引入离散潜在变量和连续潜在变量，在 VAE 框架内实现了对多视图数据的有效解耦和融合。具体而言，离散潜在变量用于提取多个视图的共同聚类特征，而连续潜在变量则用于捕捉每个视图的独有特征。通过控制互

信息容量，该算法能够有效地解耦视图共享和视图私有的表示，从而分离多个视图的连续特征，并挖掘它们的共同离散聚类信息。最终，基于解耦表示，采用 K-means 算法产生聚类标签。

此外，VGG16 的层次化特征提取能力有助于视图特征的解耦，从而提升了模型性能。未来的研究将进一步探索其在更大规模和更多样化数据集上的应用潜力，以及在实际场景中的稳健性和泛化能力。

参考文献：

[1] Zhang C, Jia X Y, Li Z C, et al. Learning cluster-wise anchors for multi-view clustering[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(15): 16696-16704.

[2] Moujahid A, Dornaika F. Advanced unsupervised learning: a comprehensive overview of multi-view clustering techniques[J]. Artificial Intelligence Review, 2025, 58(8): 234.

[3] Wu T T, Feng S H, Yuan J Z. Low-rank kernel tensor learning for incomplete multi-view clustering[C]//Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. New York: ACM, 2024: 15952-15960.

[4] 刘鹏仪, 胡节, 王红军, 等. 基于对比共识图学习的多视图属性图聚类算法[J]. 计算机科学, 2024, 51(11): 73-80.

Liu P Y, Hu J, Wang H J, et al. Multi-view attributed graph clustering based on contrast consensus graph learning[J]. Computer Science, 2024, 51(11): 73-80.

[5] 仵匀政, 杜韬, 周劲, 等. 基于三阶张量的大规模数据谱聚类集成算法[J]. 大数据, 2024, 10(3): 133-148.

Wu Y Z, Du T, Zhou J, et al. Spectral clustering ensemble algorithm based on three-order tensor for large-scale data[J]. Big Data Research, 2024, 10(3): 133-148.

[6] 安瑞, 鲁进, 杨晶晶. 基于双分支小波卷积自编码器和数据增强的深度聚类方法[J]. 计算机科学, 2025, 52(4): 129-137.

An R, Lu J, Yang J J. Deep clustering method based on dual-branch wavelet convolutional autoencoder and data augmentation[J]. Computer Science, 2025, 52(4): 129-137.

[7] 朱云华, 孔兵, 周丽华, 等. 图对比学习引导的多视图聚类网络[J]. 计算机应用, 2024, 44(10): 3267-3274.

Zhu Y H, Kong B, Zhou L H, et al. Multi-view clustering network guided by graph contrastive learning[J]. Journal of Computer Applications, 2024, 44(10): 3267-3274.

[8] Wen H R, Guo W, Li X. A novel deep clustering network using multi-representation autoencoder and adversarial learning for large cross-domain fault diagnosis of rolling bearings[J]. Expert Systems with Applications, 2023, 225: 120066.

[9] Rong Z W, Liu Z L, Song J L, et al. MCluster-VAEs: an end-to-end variational deep learning-based clustering method for subtype discovery using multi-omics data[J]. Computers in Biology and Medicine, 2022, 150: 106085.

[10] Zhang Y Y, Yan W Q, Tang C, et al. Multi-branch Space Sharing Feature Aggregation for contrastive multi-view clustering[J]. Pattern Recognition, 2025, 167: 111704.

[11] Huang H J, Mei Y B, Wei X X, et al.

- Self-weighted dual contrastive multi-view clustering network[J]. Scientific Reports, 2025, 15: 16154.
- [12] Chartsias A, Joyce T, Papanastasiou G, et al. Disentangled representation learning in cardiac image analysis[J]. Medical Image Analysis, 2019, 58: 101535.
- [13] Hou Z Q, Ma J Y, Yu W S, et al. Multi-template global re-detection based on Gumbel-Softmax in long-term visual tracking[J]. Applied Intelligence, 2023, 53(18): 20874–20890. [LinkOut]
- [14] Burgess C P, Higgins I, Pal A, et al. Understanding disentangling in β -VAE[EB]. arXiv preprint, 2018, arXiv:1804.03599.
- [15] Chen R T Q, Li X C, Grosse R, et al. Isolating sources of disentanglement in VAEs[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2018: 2615–2625.
- [16] 梁吉业, 钱宇华, 李德玉, 等. 面向大数据的粒计算理论与方法研究进展[J]. 大数据, 2016, 2(4): 13–23.
- Liang J Y, Qian Y H, Li D Y, et al. Research development on granular computing theory and method for big data[J]. Big Data Research, 2016, 2(4): 13–23.
- [17] Wang X, Chen H, Tang S A, et al. Disentangled representation learning[EB]. V4. arXiv preprint, 2022, arXiv: 2211.11695.
- [18] Liu Y, Yu Z Y, Liu Z T, et al. Improving disentanglement in variational autoencoders via feature imbalance-informed dimension weighting[J]. Knowledge-Based Systems, 2024, 296: 111818.
- [19] Vahdat A, Kautz J. NVAE: A deep hierarchical variational autoencoder[C]//Advances in Neural Information Processing Systems. 2020, 33: 19667–19679.
- [20] Ding X H, Zhang X Y, Ma N N, et al. RepVGG: making VGG-style ConvNets great again[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 13728–13737..
- [21] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614–2623..
- [22] Huijben I A M, Kool W, Paulus M B, et al. A review of the gumbel-max trick and its extensions for discrete stochasticity in machine learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 1353–1371.
- [23] Rezende D J, Mohamed S, Wierstra D. Stochastic backpropagation and approximate inference in deep generative models[C]//Proceedings of the 31st International Conference on International Conference on Machine Learning – Volume 32. New York: ACM, 2014: II–1278–II–1286.
- [24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB]. arXiv preprint, 2014, arXiv:1409.1556.
- [25] Peng X, Huang Z, Lv J, et al. COMIC: Multi-view Clustering Without Parameter Selection [C]//Proceedings of the 36th International Conference on Machine Learning. New York: PMLR, 2019: 5092–5101.
- [26] Allaoui M, Kherfi M L, Cheriet A. Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study[C]//Image and Signal Processing. Cham:

Springer, 2020: 317–325.

[27] Xu J, Ren Y Z, Li G F, et al. Deep embedded multi-view clustering with collaborative training[J]. Information Sciences, 2021, 573: 279–290.

[28] Lin Y J, Gou Y B, Liu Z T, et al. COMPLETER: incomplete multi-view clustering via contrastive prediction[C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 11169–11178.

[29] Xu J, Li C, Peng L, et al. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering[J]. IEEE Transactions on Image Processing, 2023, 32: 1354–1366.

[30] Huang D, Wang C D, Lai J H. Fast multi-view clustering via ensembles: towards scalability, superiority, and simplicity[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(11): 11388–11402.

[31] Diallo B, Hu J, Li T R, et al. Auto-attention mechanism for multi-view deep embedding clustering[J]. Pattern Recognition, 2023, 143: 109764.

[32] Zheng Y M, Jia C Y, Yu J, et al. Deep embedded clustering with distribution consistency preservation for attributed networks[J]. Pattern Recognition, 2023, 139: 109469.

[33] Zhang C B, Chen L, Shi Z Y, et al. Latent information-guided one-step multi-view fuzzy clustering based on cross-view anchor graph[J]. Information Fusion, 2024, 102: 102025.

[34] Lu Y D, Lin Y J, Yang M X, et al. Decoupled contrastive multi-view clustering with high-order random walks[C]// Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. New York: ACM, 2024: 14193–14201.

作者简介



李顺勇（1975–），男，博士，山西大学数学与统计学院教授，CCF会员，主要研究方向为统计机器学习和数据挖掘。



李青辉（1999–），男，山西大学数学与统计学院硕士生，CCF会员，主要研究方向为统计机器学习。



赵兴旺（1984-），男，博士，山西大学计算机与信息技术学院教授，CCF 会员，主要研究方向为数据挖掘和机器学习。



陈存（1988-），女，博士，郑州大学数学与统计学院副教授，CCF 会员，主要研究方向为数据科学、机器学习、材料数学。



杨梦瑶（1999-），女，山西大学数学与统计学院硕士生，CCF 会员，主要研究方向为统计机器学习。

收稿日期: 2025-02-28

通信作者: 赵兴旺, zhaoxw@sxu.edu.cn

基金项目: 国家自然科学基金资助项目(No.82274360);山西省留学回国人员科技活动择优资助项目(No.20250001);山西省基础研究计划资助项目(No.202303021221054;202403021211086);山西省回国留学人员科研教研资助项目(No.2024-002)

Foundation Items: The National Natural Science Foundation of China (No.82274360), The Shanxi Provincial Funding for Selected Scientific Research Projects of Returned Overseas Scholars (No.20250001), Supported by Fundamental Research Program of Shanxi