基于分布对齐变分自编码器的深度多视图聚类

谢胜利[1),4)] 陈泓达[1)] 高军礼[1),5)] 彭 玺[2)] 尹 明[1),3)]

[1)](广东工业大学自动化学院 广州 510006)

[2)](四川大学计算机学院 成都 610065)

[3)](华南师范大学半导体科学与技术学院 广东 佛山 528225)

[4)](物联网智能信息处理及系统集成教育部重点实验室 广州 510006)

[5)](粤港澳离散制造智能化联合实验室 广州 510006)

摘 要 多视图聚类(Multi-View Clustering,MVC)旨在利用不同视图间的一致性和互补性来高效处理多视图数据,是大数据分析中重要的研究方向之一.然而,现有方法无法有效学习到多视图信息间的潜在联系,且缺乏考虑视图重要性差异问题.针对上述这些问题,本文提出了一种基于分布对齐变分自编码器的深度多视图聚类方法(Deep Multi-View Clustering based on Distribution Aligned Variational Autoencoder,DMVCDA).首先,针对特定视图我们利用多个变分自编码器从不同视图中提取潜在特征,并对特征的分布进行对齐,以挖掘包含基本信息的潜在特征;然后,引入视图权重参数,获取共享的潜在特征;最后,在潜在特征上建立面向聚类的损失目标,使得学习到的潜在特征更适合聚类任务,从而提高聚类精度.在五个公共多视图数据集上的实验结果表明,我们的模型在精确度(*ACC*)、标准互信息(*NMI*)和纯度(*Purity*)等多个聚类评价指标上均表现出优异的性能.

关键词 多视图聚类;深度学习;变分自编码器;加权融合;对齐

中图法分类号 TP301 **DOI 号** 10.11897/SP.J.1016.2023.00945

Deep Multi-View Clustering Based on Distribution Aligned Variational Autoencoder

XIE Sheng-Li[1),4)] CHEN Hong-Da[1)] GAO Jun-Li[1),5)] PENG Xi[2)] YIN Ming[1),3)]

[1)](*School of Automation, Guangdong University of Technology, Guangzhou* 510006)

[2)](*College of Computer Science, Sichuan University, Chengdu* 610065)

[3)](*School of Semiconductor Science and Technology, South China Normal University, Foshan, Guangdong* 528225)

[4)](*The Key Laboratory of Intelligent Information Processing and System Integration of IoT, Ministry of Education of the P.R.C., Guangzhou* 510006)

[5)](*The Guangdong-HongKong-Macao Joint Laboratory for Smart Discrete Manufacturing, Guangzhou* 510006)

Abstract Multi-view data means that the same object can be described by multiple different data sources or features, and each data source or feature can be viewed as a specific view. Multi-view clustering aims to exploit the consistency and complementary information among different views to efficiently process multi-view data, which is one of the most important research topics in big data analysis. Recently, a surge of multi-view clustering methods have been developed and achieved promising success, which has attracted considerable attentions in machine learning and data mining community. However, most of the existing methods neither effectively learn the latent relationship among multiple views nor consider the different importance of each view. By such, the solution to multi-view clustering is often sub-optimal. In order to address these problems,

收稿日期:2021-12-07;在线发布日期:2022-10-13.本课题得到国家自然科学基金(61876042)、广东省自然科学基金(2020A1515011493)资助.**谢胜利**,博士,教授,主要研究领域为智能信号处理、生物医学信号处理、人工智能.E-mail:shlxie@gdut.edu.cn.**陈泓达**,硕士研究生,主要研究方向为数据挖掘、深度学习.**高军礼**,博士,副教授,主要研究方向为机器人、数据挖掘.**彭 玺**,博士,教授,主要研究领域为计算机视觉、数据挖掘.**尹 明**(通信作者),博士,教授,主要研究领域为计算机视觉、多视图学习.E-mail:yiming@gdut.edu.cn.

this paper proposes a deep multi-view clustering based on distribution aligned variational autoencoder so as to improve performance of multi-view clustering. First, we use view-specific variational autoencoder to extract latent features from different views, and align the learned view data distribution to further mine latent features containing basic information. Then, we introduce view weight parameters to fuse the view-specific features into a shared latent one. Finally, the loss function for clustering is established on the latent features, so that the learned latent features are more suitable for clustering tasks, thereby improving the clustering accuracy. The experimental results on five common multi-view datasets show that our model has achieved excellent performance in terms of multiple clustering evaluation metrics, such as accuracy (*ACC*), normalized mutual information (*NMI*) and purity (*Purity*), which validates the effectiveness of our model.

Keywords multi-view clustering; deep learning; variational autoencoder; weighted fusion; alignment

1 引 言

聚类作为一种无监督的学习方法,通过对无标签样本的学习来揭示数据内在性质及规律,并根据数据样本间的相似性度量划分为不同的簇,是大数据分析中最基础的研究任务之一.随着互联网和传感器技术的迅速发展,许多领域的数据都在海量性地增长,在一些实际应用中,数据通常是从不同领域和不同的传感器采集的,从而诞生了多视图(Multi-view)数据.具体来说,多视图数据是指同一个对象可以由来自多个不同数据源或多种不同特征所描述,而各个数据源或特征可以视作一个特定的视图,可以独立的用于聚类分析,并且不同视图之间既存在内在联系又存在差异,表明了数据表示的多样性.现如今多视图数据已变得越来越普遍,从单一视图获取的数据已不能满足描述需求,人们转而越来越关注多视图的数据描述.由此产生了多视图聚类问题,即如何高效地从多个视图中联合利用这些信息,为多视图数据实现可靠的聚类分析.

多视图聚类(Multi-View Clustering,MVC)是指以无监督方式学习多视图数据之间的互补信息,并且跨视图挖掘其一致性,最终将其合理地划分到不同的簇.近几十年来,学者们提出了许多单视图聚类算法并取得一定的成功,但无法推广到多视图聚类当中.最朴素的想法是将多视图数据直接拼接起来作为单视图信息进行聚类,然而,这种方法往往容易引起"维度灾难",并且忽略了各个视图不同的统计特性,导致聚类性能不尽人意.从信息的角度来看,多视图数据描述的是同一对象,不同视图中学习到的信息有一部分语义是共享的、一致的或者有关联的,并且每个视图还存在一些信息是该视图所特有的互补信息,这也是多视图数据相比于单视图数据的优势所在.因此,多视图聚类需要考虑不同视图的一致性和互补性信息,从而合理地利用多视图数据提升聚类效果.

学者们对多视图聚类的研究源于 21 世纪初,Bickel 等人[1]在 2004 年拓展了单视图期望最大化(Expectation Maximization,EM)算法,提出适用于两个视图数据的 EM 算法.在 2005 年,de Sa[2]提出了多视图场景下的谱聚类方法,通过最小化差异原则为多视图数据创建二分图,再采用拉普拉斯图划分方法实现聚类,但这只是一种针对两个视图的方法.此后,Huang 等人[3]提出的 AASC(Affinity Aggregation for Spectral Clustering)将谱聚类扩展到具有多个可用亲和力矩阵,以寻求它们的最佳组合,RMSC(Robust Multiview Spectral Clustering via low-rank and sparse decomposition)[4]则通过对各个视图构造转移概率矩阵,利用这些矩阵恢复得到共享的低秩转移概率矩阵.为了考虑视图的重要性差异,Zong 等人[5]遵循具有相似聚类结果的视图应被赋予相似的权重的原则,提出的 WMSC(Weighted Multi-view Spectral Clustering based on spectral perturbation)算法利用频谱扰动对视图权重进行建模.文献[6]提出了一种双重加权的多视角聚类方法,利用互信息自动学习视角权重,并设计新的优化方法保证模型收敛到局部最优解.由于图学习的兴起,出现了许多基于图学习的多视图聚类方法,其目

的是在所有视图中找到一个融合图，然后在融合图上应用图形切割或者其它聚类算法以得到聚类结果，如 Zhan 等人[7]提出的 MVGL(Graph Learning for Multiview clustering)对不同视图学习到的图进行优化并集成到全局图中，聚类结果可直接从全局图中获得，Zhan 等人[8]的 MCGC(Multiview Consensus Graph Clustering)则通过在拉普拉斯矩阵上施加秩约束，学习具有 K 个连通分量的共识图，最后从共识图中获取聚类标签，Liang 等人[9]为学习到多视图的一致性和不一致性信息，提出两个方法 SGF(Similarity Graph Fusion)和 DGF(Dissimilarity Graph Fusion)，前者是将多个相似图进行融合，后者是直接从多个视图的距离求相异图，其认为距离的融合可能比相似图的融合能更好地保留节点之间的关系.

虽然上述方法在一定程度上取得成功，但多视图聚类领域还存在着许多亟待解决的问题，如传统算法无法处理高维数据，无法平衡多视图数据一致性和互补性学习，缺乏考虑视图重要性差异等问题. 随着深度神经网络的快速崛起，深度学习已经渗透到各个领域，将深度学习与多视图聚类相结合为解决上述问题提供了新的思路. 由于神经网络超强的特征提取能力，将数据映射到低维的潜在子空间可以摆脱传统聚类算法无法解决高维数据等问题的束缚，更高效地挖掘多视图数据的潜在信息以提高聚类精度. 为此，本文结合深度学习模型与多视图聚类开展研究，探索更为高效合理的多视图聚类算法.

我们提出了一种基于分布对齐变分自编码器的深度多视图聚类(DMVCDA)方法. 该模型首先利用深度变分自编码器将原始视图数据映射到服从特定分布的隐变量，并在该过程中引入分布对齐策略来学习多视图的一致性和互补性信息；其次，考虑到不同视图信息对聚类任务贡献程度存在差异，引入一组自适应权重向量进行视图融合以得到共享的潜在表示；最后，与深度嵌入式聚类损失进行联合优化学习. 总而言之，我们的贡献如下：

(1) 基于多视图数据的一致性和互补性原则，提出一种分布对齐的深度变分自编码器网络，该模型通过对齐视图潜在分布进行多视图数据一致性学习，同时利用编码器重构损失保留视图特有的特征信息，以学习多视图的互补信息.

(2) 为充分考虑视图重要性差异问题，我们通过引入一组自适应权重向量，学习不同视图对全局共享信息的贡献. 这些权重将与分布对齐变分自编码器网络，以及聚类损失在统一的框架中一起学习优化.

(3) 我们采用不同规模的五个公共多视图数据集进行广泛的对比实验，实验结果证明了本文提出方法的有效性.

接下来，本文章节安排如下：第 2 节简要介绍深度聚类方法和变分自编码器的相关工作；第 3 节介绍本文提出算法的网络结构；第 4 节进行实验结果展示和分析；第 5 节对本文工作进行总结.

2　相关工作

2.1　深度聚类方法

深度聚类旨在通过深度神经网络学习到高质量的特征提高聚类性能，同时希望聚类结果可以引导网络学习更好的特征表示，其一般范式可表示为

$$\min \mathcal{L}=\alpha \mathcal{L}_n+\beta \mathcal{L}_c,\ \alpha \geq 0,\ \beta>0 \tag{1}$$

其中，$\mathcal{L}_n$ 为网络损失，$\mathcal{L}_c$ 为聚类损失.

较早的深度聚类方法通常分离的两部分工作：首先，训练深度特征提取器，将数据映射到低维的子空间；然后，使用传统聚类算法(K-means 或谱聚类)对特征向量进行分组. 虽然分离学习策略可以获得很好的结果，但忽视特征学习与聚类之间的关系仍可能限制聚类性能. 因此，联合学习方法逐渐占据主流地位，其关键特点是可以同时优化特征和聚类分配，以获得更好的聚类性能. 较为经典深度聚类算法是 Xie 等人[10]提出深度嵌入式聚类(Deep Embedding for Clustering，DEC)，这是一种联合优化特征学习和聚类任务的深度聚类方法. DEC 通过学习从数据空间到低维特征空间的映射，并在该映射中迭代地优化基于 KL 散度的聚类损失，可同时更新网络和聚类中心的参数. 此后，由于 DEC 的成功，Guo 等人[11]将 DEC 中的自编码器替换为卷积自编码器，可在学习到的特征空间中充分保留数据生成分布的局部结构，进一步提升网络学习能力和聚类效果.

由于深度聚类在单视图数据上表现出巨大的潜力，研究者们先后开展了基于深度学习的多视图聚类研究. 研究初期，Ngiam 等人[12]通过训练双峰深度自动编码器来提取两个视图的共享表示；而后 Wang 等人[13]受典型相关分析(Canonical Correlation Analysis，CCA)的启发，引入了自动编码器正则化

项并提出了一种新颖的深度规范相关自动编码器(Deep Canonically Correlation Autoencoders,DCCAE).然而,上述的两种算法仅针对两种视图情况,不能处理两个以上的多视图数据.此后,为将深度聚类方法扩展到两个及以上的视图,Zhao 等人[14]提出了基于深度矩阵分解的多视图聚类方法(Multi-View Clustering via Deep Matrix Factorization,DMFMVC),采用半非负矩阵分解以分层方式可以学习两个及以上视图数据的层次语义,并引入图规则化器以耦合深层结构的输出表示.而 Sun 等人[15]将深度矩阵分解和稀疏子空间学习集成在一个统一的框架中,提出深度连续多视图任务学习,其首先采用深度矩阵分解技术来捕获这个新的多视图任务的隐藏和分层表示,然后对每层提取的因子采用稀疏子空间学习模型,并通过自我表达约束进一步探索跨视图相关性.Zheng 等人[16]提出了一种具有局部和全局图信息的多视图子空间聚类网络,通过将局部多视图信息融合并集成到自表达层中得到共享的多视图子空间表示,而后使用谱聚类算法获得聚类结果.由于大多数多视图聚类方法仅考虑所有视图的全局结构,而忽略了每个视图之间的局部几何结构,导致无法学习不同视图的不同聚类的判别特征.Wang 等人[17]为此提出了一种新颖的具有统一和区分性学习的深度多视图子空间聚类,对不同视图之间使用判别约束,使得同一集群的样本具有较大的权重,而不同集群的样本具有较小的权重,为多视图聚类学习更好的共享连接矩阵.此外,为遵循多视图数据中的一致性和互补性原则,Xu 等人[18]提出了深度多视图概念学习,按层次对数据执行非负因式分解来捕获多视图数据中的语义结构.而 Yin 等人[19]提出了一种新颖的深度多视图聚类方法 MVCVAE(Shared Generative Latent Representation Learning for Multi-View Clustering),该方法通过采用深度生成模型变分自编码器学习服从高斯混合分布的共享生成潜在表示来实现多视图聚类,具有捕获所有视图之间相关性的能力.

2.2 变分自编码器(VAE)

变分自编码器[20]是源于变分贝叶斯(Variational Bayes,VB)推断框架的生成式网络,是深度无监督学习复杂分布的最流行方法之一.与自编码器不同的是,VAE 能够使隐变量服从特定的分布来描述原始数据,再通过学习到的分布进行采样重构出原始数据,这种方法可以使得学习到的特征表示更真实地描述原始数据.

VAE 的网络结构如图 1 所示,主要分为编码(推断)和解码(生成)两个过程,其中 z 是隐变量(latent variables),x 是被观测到的原始数据,$\hat{x}$ 表示重构数据.一般,假设数据服从高斯分布,首先,原始数据输入编码器网络 $q_\phi(z|x)$ 后得到的两个输出 μ 和 Σ,分别表示高斯分布的均值和方差,然后,利用重新参数化技巧[20]生成潜在变量 z.最后,z 通过解码器网络 $p_\theta(x|z)$ 重构出原始数据,ϕ、θ 分别表示编码和解码网络的参数.

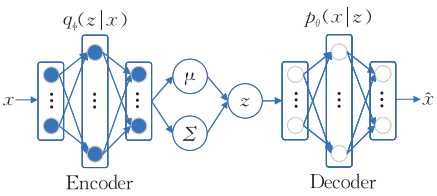


图 1 VAE 网络结构

为使隐变量服从特定的分布,VAE 引入 KL 散度来衡量两个分布之间的距离,其运用到变分推理的思想.在变分推理中,通常我们关心的分布是概率模型隐变量的后验分布,其核心想法是用一个简单且容易得到的分布 $q(z)$ 来近似 $p(z|x)$.在 VAE 中旨在找到潜在变量上的真实条件概率分布 $p_\theta(z|x)$,可以通过找到其最接近的代理后验 $q_\phi(z|x)$ 来对其进行近似估算.VAE 在优化模型参数 ϕ 和 θ 的过程中引入了变分推理的思想:

$$\log p_\theta(x)=D_{\mathrm{KL}}(q_\phi(z|x)\|p_\theta(z|x))+\mathcal{L}(\theta,\phi;x) \tag{2}$$

其中,$\log p_\theta(x)$ 是样本数据出现的概率对数.第一项的 KL 散度表示两个分布 $p_\theta(z|x)$ 和 $q_\phi(z|x)$ 的逼近程度,是我们需要最小化的对象,第二项被称为证据(变分)下限(Evidence Lower Bound,ELBO).在左边常数项不变的情况下,通过最大化 ELBO 可以使 KL 散度项取得最小.VAE 的最终目标函数是最小化重构损失和 KL 散度之和,经过变换推导后可以表示为

$$\mathcal{L}=\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]-D_{\mathrm{KL}}(q_\phi(z|x)\|p_\theta(z)) \tag{3}$$

其中,第一项是重构误差,第二项是推理网络 $q(z|x)$ 和先验分布 $p(z)$ 之间的 KL 散度.

3 基于分布对齐变分自编码器的深度多视图聚类

本节提出一种基于分布对齐变分自编码器的深

度多视图聚类算法，该模型包含三个模块，即分布对齐变分自编码器模块、加权融合模块和聚类模块，图 2 为该模型的网络结构，我们将对三个模块进行详细介绍.

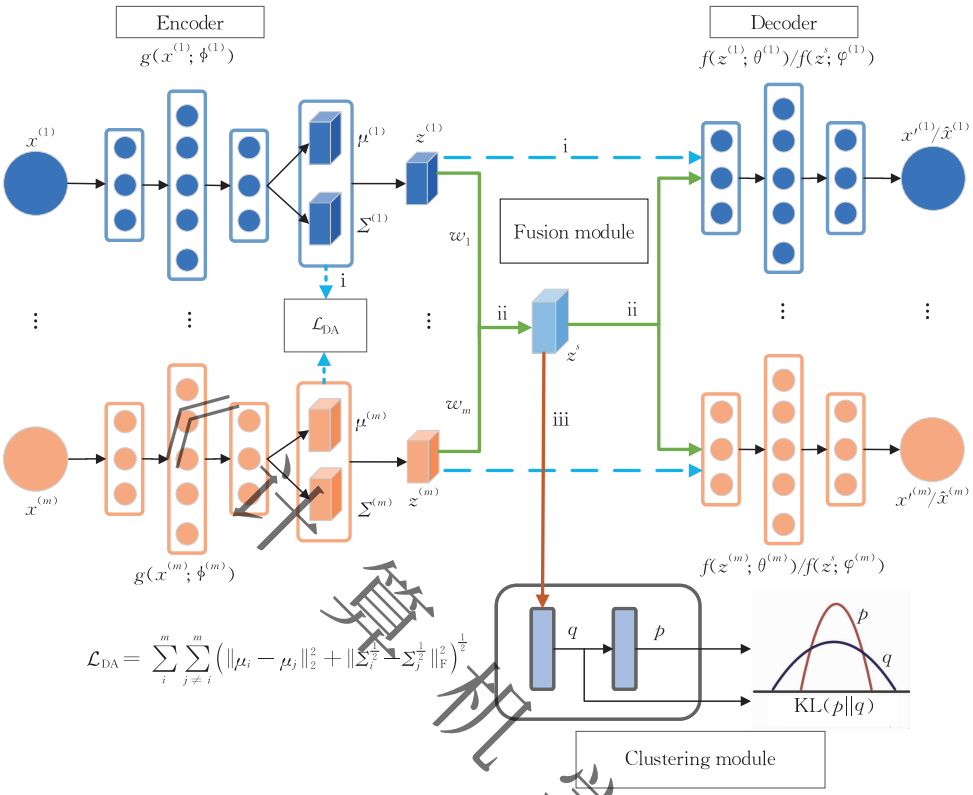


图 2　本文提出的 DMVCDA 的网络结构，主要由三个模块组成：(i) 分布对齐 VAE 模块：该模块可以学习到不同视图的潜在信息，即多视图的一致和互补信息；(ii) 加权融合模块：考虑到不同视图的重要性差异，引入自适应权重得到多视图的共享潜在表示；(iii) 聚类模块：最小化引入辅助目标分布 p 与软分配分布 q 的 KL 散度，与网络联合进行学习

3.1　分布对齐变分自编码器

为遵循多视图数据的一致性原则，我们对变分自编码器在学习潜在分布的过程中进行明确的对齐约束，以挖掘多视图数据的一致性信息，称为分布对齐(Distribution-Alignment，DA). 在变分自编码器中引入分布对齐策略的动机主要有两点：一是各自视图在学习的过程中有其他视图信息的加入，可以使网络学习到多视图的一致性信息；二是利用解码器重构损失的约束，使得视图在学习到多视图一致性信息的同时，保留了自身视图表达性强的特征信息，有利于挖掘多视图数据的一致性信息和互补性信息.

首先，我们基于深度生成模型 VAE 进行改进，使其能捕获视图特定的潜在特征，以便有效地保留各类判别信息；进而显式约束各视图的概率分布，进行对齐学习，获取各视图间的一致性信息. 假设网络结构包含 m 个编码器，m 表示视图的个数. 为使网络在学习多个视图的一致性信息同时，能够最大限度地减少视图自身信息的丢失，需要利用解码器网络重建原始数据，所以我们模型的基本 VAE 损失是 m 个视图的 VAE 损失之和，即

$$\mathcal{L}_{VAE}=\sum_{i}^{m}(\mathbb{E}_{q_\phi(z^{(i)}|x^{(i)})}[\log p_\theta(x^{(i)}|z^{(i)})]-\delta D_{KL}(q_\phi(z^{(i)}|x^{(i)})\|p_\theta(z^{(i)}))) \tag{4}$$

其中，δ 是 KL 散度的加权惩罚因子[21].

分布对齐有多种可行的方法，如通过最小化两个潜在分布之间的最大平均差异[22]或者最小化平方损失互信息[23]等. 然而，目前主流的分布距离度量方法是 KL 散度，这是因为 KL 散度的计算方式简单并且计算成本低，但其存在较多缺点，如无界、不对称等，并且在分布间无重合时，KL 散度值将会趋近无穷. 由于多视图数据大多数是异构的数据，视图间可能存在无重合的情况，使用 KL 散度得到的结果无意义，而使用 Wasserstein 距离[24]的值却仍然能提供有用信息，且更为平滑，这有助于梯度下降法的参数更新. 另外，最大平均差异方法主要用来解决域适应问题，常用于迁移学习当中，适用于两个及以上不同域的样本之间的分布关系，而平方损失互

信息本质上是随机变量联合分布和独立组合随机变量边缘分布之间的 KL 散度,因此,在本文中,我们采用 Wasserstein 距离来对齐视图间潜在变量 z 的高斯分布.两个多元高斯分布 i 和 j 之间的二阶 Wasserstein 距离 Y_{ij} 可表示为

$$Y_{ij}=[\|\mu_i-\mu_j\|_2^2+Tr(\Sigma_i)+Tr(\Sigma_j)-2Tr((\Sigma_i^{\frac{1}{2}}\Sigma_j\Sigma_i^{\frac{1}{2}})^{\frac{1}{2}})]^{\frac{1}{2}} \tag{5}$$

其中,$Tr(\cdot)$表示矩阵的迹,μ、Σ 分别是多元高斯分布的均值和方差.

由 $\Sigma_i\Sigma_j=\Sigma_j\Sigma_i$,$Tr(\Sigma_i^{\frac{1}{2}}\Sigma_j\Sigma_i^{\frac{1}{2}})=Tr(\Sigma_i\Sigma_j)$,则有

$$\begin{aligned}&Tr(\Sigma_i)+Tr(\Sigma_j)+2Tr((\Sigma_i\Sigma_j)^{\frac{1}{2}})\\&=Tr((\Sigma_i^{\frac{1}{2}}-\Sigma_j^{\frac{1}{2}})^2)\\&=\|\Sigma_i^{\frac{1}{2}}-\Sigma_j^{\frac{1}{2}}\|_{\text{Frobenius}}^2\end{aligned} \tag{6}$$

故式(5)可简化为[24]

$$Y_{ij}=(\|\mu_i-\mu_j\|_2^2+\|\Sigma_i^{\frac{1}{2}}-\Sigma_j^{\frac{1}{2}}\|_{\text{Frobenius}}^2)^{\frac{1}{2}} \tag{7}$$

对于 m 组多视图数据,其分布对齐的损失(DA loss)可以写为以下形式:

$$\mathcal{L}_{\text{DA}}=\sum_i^m\sum_{j\neq i}^m Y_{ij} \tag{8}$$

3.2 生成多视图共享潜在表示

多视图聚类研究的另一个难点在于如何融合多个视图的有效信息.前期研究表明,通过相似性矩阵相加或直接进行特征拼接组合多个视图的基本思想无助于有效改善聚类性能.这是因为易受干扰的相似性矩阵可能导致次优结果,而直接拼接往往会导致"维度灾难"的发生.因此,我们在重视多视图数据一致性和互补性学习的同时,也应考虑到不同视图对聚类任务的贡献存在差异性.对此,我们通过引入一组视图权重向量 $\boldsymbol{W}=[w_1,w_2,\cdots,w_m]^{\mathrm{T}}(w_i\geq 0,\sum w_i=1)$,对多个视图的潜在表示进行加权融合,得到一个共享的潜在表示 z^s,并作为后续聚类模块的输入来实现聚类优化,z^s可以表示为

$$\begin{cases}z^s=\sum_i^m w_i z^{(i)}\\\sum_i^m w_i=1\end{cases} \tag{9}$$

如图 2 所示,编码器网络通过加权融合后得到共享潜在表示 z^s,再经过解码器网络 $f(\cdot)$重构出视图的原始数据.利用神经网络反向传播求导机制[25]使网络可以自动学习权重 $\boldsymbol{W}$,加权融合模块的重构损失可以表示为如下形式:

$$\mathcal{L}_{\text{Fusion}}=\sum_i^m\|f(z^s;\varphi^{(i)})-x^{(i)}\|_2^2 \tag{10}$$

其中,$f(z^s;\varphi^{(i)})$表示由共享潜在表示 z^s 生成的第 i 个视图的重构数据 $\hat{x}^{(i)}$,$x^{(i)}$为第 i 个视图原始数据.我们通过对视图的潜在表示施加自适应权重而获得多视图的共享潜在表示,使得网络可以自主学习到视图差异信息对全局的重要性.

3.3 深度嵌入聚类损失

一般,我们可假设$x^{(v)}$表示第 v 个视图的原始特征空间,则多视图数据的特征可以用$\{x^{(v)}\}:=\{x^{(1)},\cdots,x^{(v)},\cdots,x^{(m)}\}$来表示,并将 m 个视图数据点$\{x^{(1)},x^{(2)},\cdots,x^{(m)}\}$聚成 K 个簇.各个视图中的各个簇通过一个聚类中心 $\mu_j^{(m)}$ 表示,其中$j=1,2,\cdots,K$,K为聚类个数.由于在原始空间进行聚类会带来沉重的计算负担,并且为克服单独学习策略导致的聚类结果不理想,我们通常将多视图特征从原始空间映射到隐式嵌入特征空间(Latent embedded feature space),进而联合优化特征表示和聚类分配(cluster assignment).

在本文中,我们采用深度嵌入式聚类 DEC[10]的方法定义基于质心的软分配分布,通过最小化该分布与辅助目标分布之间的 KL 差异,并与网络联合进行优化学习.DEC 通过对网络进行编码得到的嵌入特征 z_i 进行 K-means 聚类,得到网络初始化参数和聚类中心 μ_j 的初始估计.然后,度量嵌入特征 z_i 与聚类中心 μ_j 之间的相似性:

$$q_{ij}=\frac{(1+\|z_i-\mu_j\|^2/\epsilon)^{-\frac{\epsilon+1}{2}}}{\sum_{j'}(1+\|z_i-\mu_{j'}\|^2/\epsilon)^{-\frac{\epsilon+1}{2}}} \tag{11}$$

其中,ϵ 表示学生分配的自由度.q_{ij} 被称为软分配,可以解释为将样本 i 分配给第 j 个聚类的概率.然而,由于无法在无监督方式下对验证集上的ϵ进行交叉验证,不失一般性,故将ϵ设置为 1.DEC 引入辅助目标分布 p,将目标函数定义为软分配 q_{ij} 和辅助分布 p_{ij} 之间的 KL 散度距离,迭代地优化了聚类目标.辅助分布 p 定义如下:

$$p_{ij}=\frac{q_{ij}^2/f_j}{\sum_{j'}q_{ij'}^2/f_{j'}} \tag{12}$$

其中 $f_j=\sum_i q_{ij}$ 是软簇的频率.上述辅助分布可以通过提高软分配的置信度得分来引导聚类分析,故聚类的损失函数为

$$\mathcal{L}_C=KL(P\|Q)=\sum_{i,j}p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{13}$$

针对上述聚类目标函数，本文利用随机梯度下降法(Stochastic Gradient Descent，SGD)来联合优化多视图共享潜在表示 z^s 和聚类分配任务学习.

3.4　网络训练

我们的模型联合学习多视图特征和进行聚类分析，总体目标函数为

$$\mathcal{L}_{\mathrm{DMVCDA}}=\mathcal{L}_{\mathrm{VAE}}+\alpha\mathcal{L}_{\mathrm{DA}}+\beta\mathcal{L}_{\mathrm{Fusion}}+\gamma\mathcal{L}_{C} \quad (14)$$

式中的 α,β,γ 为分布对齐损失，加权融合损失和聚类损失的平衡参数. 我们网络训练步骤如下.

算法 1.　DMVCDA 训练步骤.

输入：多视图数据集 X，簇的数量 k，z^s 维度 d

初始化：视图权重 w，参数 $\alpha,\beta,\gamma,\delta$

1. 使用式(4)对 VAE 网络进行预训练.
2. 加入式(8)和式(10)训练，得到多视图共享潜在表示的 z^s.
3. 进行微调阶段，加入式(13)去掉分布对齐损失式(8)，优化 z^s 和聚类损失.
4. 若相邻两次 epoch 的聚类结果 ACC 相差小于 0.001，网络停止训练；否则，达到最大 epoch 时停止训练.

输出：聚类最终结果指标

在我们的实验中，编码器和解码器网络采用全连接网络，网络的结构参数如 4.1 节表 2 所示. 该模型使用 Adam 优化器[26]进行训练，学习率设置为 0.000 15. 首先，在视图各自的 VAE 网络训练 5 个 epoch，使网络前期仅对各自视图特有信息进行编码，而后加入分布对齐损失和加权融合层损失，α 权重因子以一个时期 0.54 的速率在第 6 个到第 22 个 epoch 开始递增后保持不变，在 Caltech101 和 ALOI 数据集测试中 β 设置为 20，其余为 10 并保持不变，γ 设置为 10. 对于式(4)中 KL 散度中的加权惩罚系数 δ，我们参考退火方案[27]以一个时期 0.0026 的速率增加，直到第 90 个 epoch 保持不变. KL 散度退火方案的目的是让 VAE 在“平滑”之前学习“有用的”表示，否则其将是一个过强的正则子.

4　实验结果与分析

本文的实验是基于 Python3.6 编程语言，并使用深度学习框架 Pytorch1.1.0 搭建的环境，配置为 GTX 1080Ti 和 CUDA10.0. 本文实验代码已上传到 https://github.com/chenvvgood/DMVCDA 以供学习交流.

4.1　数据集与网络结构信息

为验证本文提出的 DMVCDA 算法的有效性，我们在五个公开的数据集上进行广泛的实验，表 1 中简要地介绍数据集的统计信息.

表 1　数据集信息

Datasets	Views	Classes	Instances
MSRC-v1	5	7	210
Yale	3	15	165
NUS-WIDE	5	31	2000
Caltech101	6	102	9144
ALOI	4	100	10 800

(1) MSRC-v1[28]. 这是一个图像数据集，由属于七个类的 210 个对象组成. 这七个类别包括树、建筑物、飞机、动物、面部、汽车和自行车. 在我们的实验中，MSRC-vl 数据集包含五个视图，它们是 CM 特征(24)、HOG 特征(576)、GIST 特征(512)、LBP 特征(256)和 GENT 特征(254).

(2) Yale. 一种广泛使用的人脸图像数据集，由 165 个灰度图像组成，这些图像具有 15 个不同的类别属性. 数据集的变化由中央光、戴眼镜、快乐、左光、不戴眼镜、正常、右光、悲伤、困倦、惊讶和眨眼组成. 在我们的实验中，使用三个视图，即 Intensity 特征、LBP 特征和 Gabor 特征，其尺寸分别为 4096、3304 和 6750.

(3) NUS-WIDE. NUS 是一种用于对象识别的数据集，该数据集包含从文献中[29]报告的 NUS-WIDEOBJECT 数据集的图像中提取的多视图特征. 我们使用五种类型的低级特征，包括 CH(65)、CM(226)、CORR(145)、EDH(74)和 WT(129)，共有 2000 张图像 31 类.

(4) Caltech101. Caltech101 是一个图像物体识别数据集，包含 101 类物体和背景杂波类的对象的图像，各个类别约 40 至 800 张图像，共 9144 张图像. 在我们的实验中，使用了为 Gabor、WM、CENT、HOG、GIST 和 LBP 特征作为六个视图，其尺寸分别为 48、40、254、1984、512 和 928.

(5) ALOI. ALOI(Amsterdam Library of Object Images)[30]图像库是在各种光照条件和旋转角度下拍摄的 110 250 个具有 1000 个小对象类别图像的集合. 由于对一些算法来说数据集太大，我们借鉴 Houle 等人[31]的观点使用一个子集，即 10 800 张图像共 100 个类. 在我们的实验中使用四个视图，分别为 CS(77)、HAR(13)、HSB(64)和 RGB(125)特征.

我们在模型中使用三层全连接神经网络，编码器(Encoder)和解码器(Decoder)的网络结构对称，

在表 2 中我们给出实验中各个数据集的网络结构参数，其中 Fusion layer 表示为网络中多视图共享潜在表示 z^s 的维度. 针对 z^s 维度的选择我们进行如下实验测试，如图 3 所示.

表 2 网络结构参数

Datasets	Encoder	Fusion layer	Decoder
MSRC-v1	500-500-1024	7	1024-500-500
Yale	2000-1024-500	15	500-1024-2000
NUS-WIDE	500-500-1024	31	1024-500-500
Caltech101	500-500-1024	10	1024-500-500
ALOI	500-500-1024	10	1024-500-500

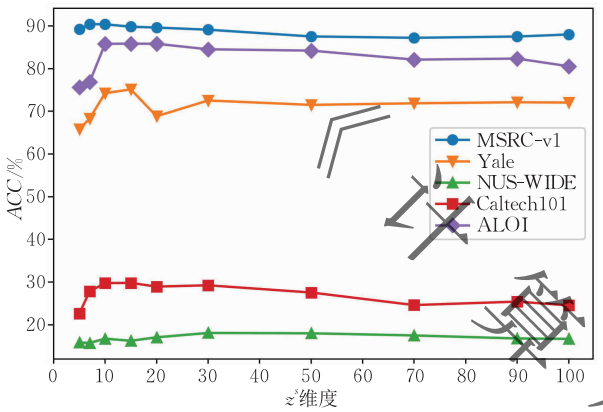


图 3 z^s 维度实验测试

从图 3 中分析可知，对共享潜在表示 z^s 维度的选择会对聚类结果产生影响，我们认为主要有两个因素，即数据集的样本规模与聚类类别数量. 结合表 1 分析可知，受限于样本数量的数据集 MSRC-v1、Yale 和 NUS-WIDE，z^s 维度的最优选择在其聚类类别数量附近，选择过小的维度，网络无法学习到足够有效的信息，而选择过大的维度则会增加冗余信息，都不利于最终的聚类结果；对于较大规模的 Caltech101 和 ALOI 数据集，尽管需要聚类类别数量很大，但由于训练样本量足够多，神经网络可以学习到表达性更强的特征信息，因此对 z^s 维度的最优选择在 10～30 之间. 实际上，z^s 维度的选择也从侧面反应网络模型对数据特征信息的学习能力，以更少的特征维度信息可以达到更好的聚类结果也正是我们所期望的. 因此根据经验，对于样本规模不大且聚类类别数量较少的数据集，我们以聚类类别数量作为 z^s 的维度，对于规模较大的数据集，z^s 维度选择为 10.

4.2 比较方法

我们将本文提出的算法模型 DMVCDA 与近年来提出的多视图聚类方法进行比较，在我们的实验比较方法中，(1)和(2)是基于特征的传统聚类方法，(3)至(6)采用的是基于相似度矩阵的传统多视图聚类方法，(7)至(9)是基于图学习的多视图聚类方法，(10)和(11)是基于深度思想和深度模型的深度多视图聚类.

(1) FeatConcate. 将多视图数据中不同视图的同一类数据通过特征拼接的方法组成新的数据集，再通过 K-means 方法进行聚类.

(2) AWP[32]. Multiview clustering via adaptively Weighted Procrustes 方法，通过对 procrustes 均值(Procrustes Average，PA)进行加权改进使其考虑到不同视图聚类能力的差异，从而应用于多视图聚类学习.

(3) AASC[3]. Affinity Aggregation for Spectral Clustering 将谱聚类扩展到具有多个可用的相似性矩阵情况，以寻求相似性矩阵的最佳组合，并将其聚合为一个再执行谱聚类.

(4) RMSC[4]. Robust Multiview Spectral Clustering via Low-Rank and Sparse Decomposition 首先从各个单视图构造一个转移概率矩阵，然后使用这些矩阵来恢复共享的低秩转移概率矩阵，最后将该矩阵作为标准马尔可夫链聚类方法的关键输入.

(5) WMSC[5]. Weighted Multiview Spectral Clustering based on spectral perturbation 利用频谱扰动对视图的权重进行建模，遵循各个视图上的聚类结果应接近于共识聚类结果和具有相似聚类结果的视图应被赋予相似的权重的原则，使用最大规范角来衡量谱聚类结果之间的差异.

(6) MCLES[33]. Multi-view Clustering in Latent Embedding Space 称为潜在嵌入空间中的多视图聚类，该方法能够在学习到的潜在嵌入空间中聚类多视图数据，同时学习全局相似性矩阵和准确的聚类指标矩阵.

(7) MVGL[7]. Graph Learning for Multi-View clustering 是利用基于图学习的方法来提高图的质量. 从不同视图的数据点获知的图称为初始图，其在拉普拉斯矩阵上具有秩的约束可进一步得到优化，再将这些优化图集成到全局图中. 由于有秩的限制，聚类结果可直接从全局图获得.

(8) MCGC[8]. Multiview Consensus Graph Clustering 通过最小化不同视图之间的分歧并限制拉普拉斯矩阵的秩来学习共识图. 不同于大多数基于图的方法使用固定的亲和力矩阵，其在拉普拉斯矩阵上施加秩约束，以学习具有正好 k 个连通分量的共识图，k

是聚类的数量，聚类标签可直接从共识图获取.

(9) SGF、DFG[9]. Similarity Graph Fusion 和 Distance(Dissimilarity) Graph Fusion 是基于图学习的多视图聚类方法，在统一的目标函数中同时对多视图一致性和多视图不一致性进行显式建模. 相似度图融合(SGF)将多个相似度图融合为一个，相异图(DGF)直接从多个视图的距离求相异图，其认为距离的融合可能比相似性的融合更好地保留节点之间的关系.

(10) DMFMVC[14]. Multi-View Clustering via Deep Matrix Factorization 是一种用于多视图聚类的深度矩阵分解方法，采用半非负矩阵分解以分层方式学习多视图数据的层次语义，并且引入图规则化器以耦合深层结构的输出表示.

(11) MVCVAE[19]. Shared Generative Latent Representation Learning for Multi-View Clustering 通过学习服从高斯混合分布的共享生成潜在表示来实现多视图聚类，该方法采用深度生成模型变分自编码器，具有捕获所有视图之间相关性的能力.

4.3　评价指标

为验证本文提出的算法的真实性能，我们使用四种常见的聚类评估指标来评估这些多视图聚类方法的性能，分别是聚类准确性(*ACC*)、标准化互信息(*NMI*)、纯度(*Purity*)和调整兰德系数(*ARI*).

聚类准确性(*ACC*)用于测量通过算法获得的实际标签和预测标签之间的准确性. 其表达式为

$$ACC=\max_{m}\frac{\sum_{i=1}^{n}\mathbf{1}\{l_i=m(c_i)\}}{n} \tag{15}$$

其中，n 是样本数，l_i 是真实标签，c_i 是预测标签. $m(c_i)$是置换映射，该映射最大程度地匹配地面标签和预测标签. $\mathbf{1}\{l_i=m(c_i)\}$是判别函数，若相等其值为 1，否则为 0.

标准化互信息(*NMI*)是计算相同数据的两个标签之间的相似度的标准化度量，可以将其规范化为

$$NMI=\frac{\mathbf{I}(l;c)}{\max\{H(l),H(c)\}} \tag{16}$$

其中，$\mathbf{I}(l;c)$表示 l 和 c 之间的互信息，$H(\cdot)$表示熵.

纯度(*Purity*)是衡量一个簇包含一个类中数据的程度，可以通过式(17)计算其值：

$$Purity=\sum_{i=1}^{k}\frac{n_i}{n}P(\mathbf{S}_i),\ P(\mathbf{S}_i)=\frac{1}{n_i}\max_{j}P(n_i^j) \tag{17}$$

式中 S_i是大小为 i 的特定簇，n_i^j代表的是数据点，即第 i 个类别的数据点分配给第 j 个集群.

调整兰德系数(*ARI*)是兰德指数(*RI*)的修正概率版本，从广义的角度来讲，*ARI* 衡量的是两个数据分布的吻合程度. 修改后的兰德指数可以将其计算为

$$ARI=\frac{RI-E(RI)}{\max(RI)-E(RI)} \tag{18}$$

不同的评价指标侧重从聚类的不同角度进行度量，以上四个指标的值越大，说明聚类算法的性能越好.

4.4　实验结果与算法效率

在本文实验中，对比方法使用它们论文中提供最优参数设置进行实验，由于 MCLES 方法只适用于小数据集，故引用其论文中的 Yale 数据集实验数据和对 MSRC-v1 数据集进行实验测试. 所以方法运行 10 次实验并报告性能指标的平均值和标准差，如表 3 所示. 粗体数字突出显示最佳结果.

从表 3 中实验结果可知，我们的模型在不同规模和类别的数据集都取得优异的表现. 对于 MSRC-v1 和 Yale小样本数据集，我们提出的方法在*ACC*、*NMI*、*Purity* 和 *ARI* 上都取得最优值；在 NUS-WIDE 数据集上，我们比次优模型 DFG 在 *ACC*、*NMI* 和 *Purity* 上分别高出 1.27%、2.75%和 1.57%；在较大规模的 Caltech101 和 ALOI 数据集上同样有不错的表现，特别是在 Caltech101 上，*ACC*、*NMI* 和 *Purity* 指标比次优值分别高出 3.56%、2.93%和 3.64%.

通过分析，在与 FeatConcate 方法比较中可发现，相比于将多视图数据直接地进行特征拼接再执行聚类而言，我们的方法通过生成一个多视图的共享潜在表示再联合聚类优化，不但可以去除多视图数据存在的冗余信息，而且可以学习到多视图间的潜在联系以提升聚类效果. 不同于基于相似性矩阵的聚类方法 AASC、RMSC、WMSC 和 MCLES，旨在最大化所有视图聚类一致性的想法，即强调各个视图聚类结果的一致性，而忽略视图信息的差异性对全局的影响，也不同于 AWP、SGF 和 DFG 方法使用特定加权的方式来解决视图的重要性差异问题，我们的 DMVCDA 方法通过引入自适应加权的方式，利用深度神经网络自主学习视图权重，由网络通过自身学习来决定视图的重要性差异；再者，为摆脱 AASC、RMSC 和 MVCVAE 等方法在视图学习的过程无使用到除自身视图外其他视图信息的局限性，DMVCDA 在变分自编码中使用分布对齐策略进行视图一致性学习，使得视图在学习自身信息的同时可以学到其他视图的一致性信息. 总之，我们提出的 DMVCDA 方法可以更合理地挖掘多视图数据的一致性和互补性信息.

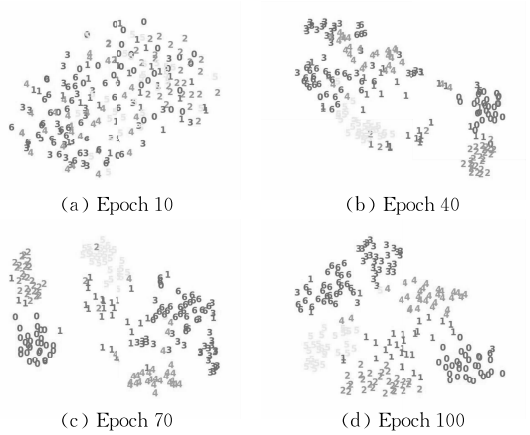
表 3 多视图聚类方法的聚类性能实验对比结果(平均值和标准偏差)("—"表示模型代码中未使用的评价指标)

Datasets	Methods	*ACC*	*NMI*	*Purity*	*ARI*
MSRC-v1	FeatConcate	81.27±4.45	72.49±3.378	2.90±3.78	66.18±3.03
	AWP	76.19±0.00	67.71±0.00	79.52±0.00	62.25±0.00
	AASC	77.33±0.25	69.78±0.14	77.33±0.25	59.90±0.18
	RMSC	71.05±1.70	68.87±1.90	76.14±1.80	55.03±2.40
	WMSC	76.52±0.45	72.11±0.40	81.14±0.25	65.02±0.55
	SGF	80.43±0.15	78.88±0.07	83.81±0.00	72.28±0.10
	DFG	87.14±0.00	80.98±0.00	87.14±0.00	75.35±0.00
	MVGL	68.10±0.00	65.58±0.00	72.86±0.00	49.67±0.00
	MCGC	84.76±0.00	71.80±0.00	84.76±0.00	68.09±0.00
	MCLES	87.48±0.74	77.91±1.16	87.48±0.74	—
	DMFMVC	65.33±0.20	51.84±0.13	65.24±0.18	—
	MVCVAE	75.24±1.46	68.50±1.34	79.05±1.06	60.96±2.00
	DMVCDA	**90.43±0.29**	**82.06±0.44**	**90.43±0.29**	**78.92±0.56**
Yale	FeatConcate	61.96±3.99	67.18±3.06	62.40±3.76	46.09±4.53
	AWP	67.27±0.00	69.42±0.00	67.88±0.00	49.31±0.00
	AASC	65.88±1.50	66.39±2.10	66.00±1.50	42.36±4.00
	RMSC	68.79±1.40	70.25±1.40	69.27±1.30	51.43±2.10
	WMSC	69.70±0.57	71.89±0.76	69.70±0.57	51.95±1.20
	SGF	70.79±0.26	73.88±0.34	70.79±0.26	54.82±0.53
	DFG	70.91±0.00	73.90±0.00	70.91±0.00	54.83±0.00
	MVGL	64.85±0.00	65.95±0.00	64.85±0.00	43.81±0.00
	MCGC	61.82±0.00	67.17±0.00	63.03±0.00	47.35±0.00
	MCLES	70.48±0.01	72.54±0.01	70.55±0.00	—
	DMFMVC	74.73±2.91	73.42±1.64	74.85±0.43	—
	MVCVAE	41.82±0.86	47.06±1.31	44.24±1.24	21.91±0.96
	DMVCDA	**75.13±0.98**	**73.96±0.83**	**75.13±0.97**	**54.89±1.65**
NUS-WIDE	FeatConcate	14.91±0.04	19.21±0.51	25.24±0.65	4.30±0.37
	AWP	14.60±0.00	15.96±0.00	22.85±0.00	3.75±0.00
	AASC	15.70±0.19	17.83±0.43	23.92±0.40	4.13±0.18
	RMSC	15.49±0.62	18.95±0.29	24.48±0.30	4.53±0.30
	WMSC	15.02±0.10	19.03±0.22	25.68±0.53	4.71±0.12
	SGF	16.22±0.45	19.69±0.12	25.86±0.43	4.97±0.21
	DFG	16.80±0.12	19.86±0.26	26.77±0.31	**5.96±0.20**
	MVGL	13.85±0.00	5.50±0.00	15.70±0.00	0.16±0.00
	MCGC	12.75±0.00	14.55±0.00	22.40±0.00	2.43±0.00
	DMFMVC	8.41±0.18	7.73±0.10	13.96±0.14	—
	MVCVAE	16.35±0.26	12.43±0.11	19.25±0.07	2.93±0.99
	DMVCDA	**18.07±0.18**	**22.61±0.17**	**28.34±0.18**	5.82±0.10
Caltech 101	FeatConcate	23.42±0.73	48.96±0.26	46.36±0.40	17.36±1.01
	AWP	26.22±0.00	44.52±0.00	42.79±0.00	15.28±0.00
	AASC	23.80±0.77	37.88±0.69	40.11±0.39	7.18±1.50
	RMSC	22.77±0.93	41.52±0.33	38.98±0.34	21.57±1.70
	WMSC	23.29±0.67	45.81±0.25	45.94±0.37	15.39±0.92
	SGF	24.04±0.45	46.39±0.15	46.26±0.21	14.71±0.65
	DFG	23.53±0.39	46.76±0.08	46.28±0.34	14.28±0.47
	MVGL	13.44±0.00	14.13±0.00	21.46±0.00	−0.55±0.00
	MCGC	23.00±0.00	41.97±0.00	43.12±0.00	13.84±0.00
	DMFMVC	21.28±0.55	41.67±0.15	40.90±0.25	—
	MVCVAE	25.45±0.23	44.63±2.32	40.31±2.14	**22.46±1.23**
	DMVCDA	**29.78±0.06**	**51.89±0.04**	**50.00±0.02**	22.08±0.03
ALOI	FeatConcate	71.30±1.47	83.87±0.39	74.16±1.16	58.51±1.94
	AWP	59.04±0.00	69.90±0.00	60.25±0.00	47.42±0.00
	AASC	15.90±0.41	35.68±0.50	18.40±0.36	6.39±0.34
	RMSC	77.04±1.50	82.45±0.68	78.44±2.30	65.61±1.50
	WMSC	78.22±0.59	84.22±1.70	79.78±0.47	68.03±0.52
	SGF	84.07±1.50	90.90±0.35	86.08±1.10	78.32±1.20
	DFG	84.51±1.00	**91.08±0.33**	86.41±0.79	78.84±1.20
	MVGL	42.47±0.00	46.94±0.00	44.98±0.00	2.48±0.00
	MCGC	56.62±0.00	69.75±0.01	60.39±0.00	41.61±0.01
	DMFMVC	60.14±1.13	76.47±0.24	63.43±0.78	—
	MVCVAE	45.58±2.88	56.00±2.34	47.32±1.94	26.45±3.01
	DMVCDA	**85.81±0.07**	90.84±0.07	**86.61±0.06**	**80.34±0.12**

为进一步探讨我们算法在生成多视图共享潜在表示的表现性能,我们使用 t-SNE[34] 可视化数据集 MSRC-v1 上的潜在隐空间,将其维数减少到二维空间,结果如图 4 所示. 具体而言,我们可以看出随着网络训练的不断进行,从多个视图信息学习到的共享潜在表示聚类变得越来越分离,并且簇与簇之间

也越来越容易区分，这表明我们的算法模型学习到更适用于聚类任务的多视图共享潜在表示，最终取得不错的聚类效果，也说明我们的算法模型可以有效地挖掘学习多视图数据的潜在信息.

在图 5 中我们对 MSRC-v1 数据集的训练损失函数曲线进行可视化，展示了 DMVCDA 网络中总体损失函数和各个模块的损失函数的收敛曲线. 从图 5(a)可以看到，本文提出的算法在 MSRC-v1 数据集上的网络训练损失函数曲线是收敛的，随着网络的训练，损失函数曲线逐渐下降后保持平缓稳定，并不会因为各个模块融合为一体进行训练而导致网络存在不稳定的现象. 从其他模块的损失函数曲线我们也可以发现，各个模块的损失函数总体呈现逐渐下降的趋势，最终都能达到收敛的效果. 总而言之，以上这些结果证明了我们的模型的有效性和稳定性.



(a) Epoch 10　(b) Epoch 40　(c) Epoch 70　(d) Epoch 100

图 4　可视化 MSRC-v1 数据集在 DMVCDA 训练 100 个 epochs 过程的共享潜在隐空间

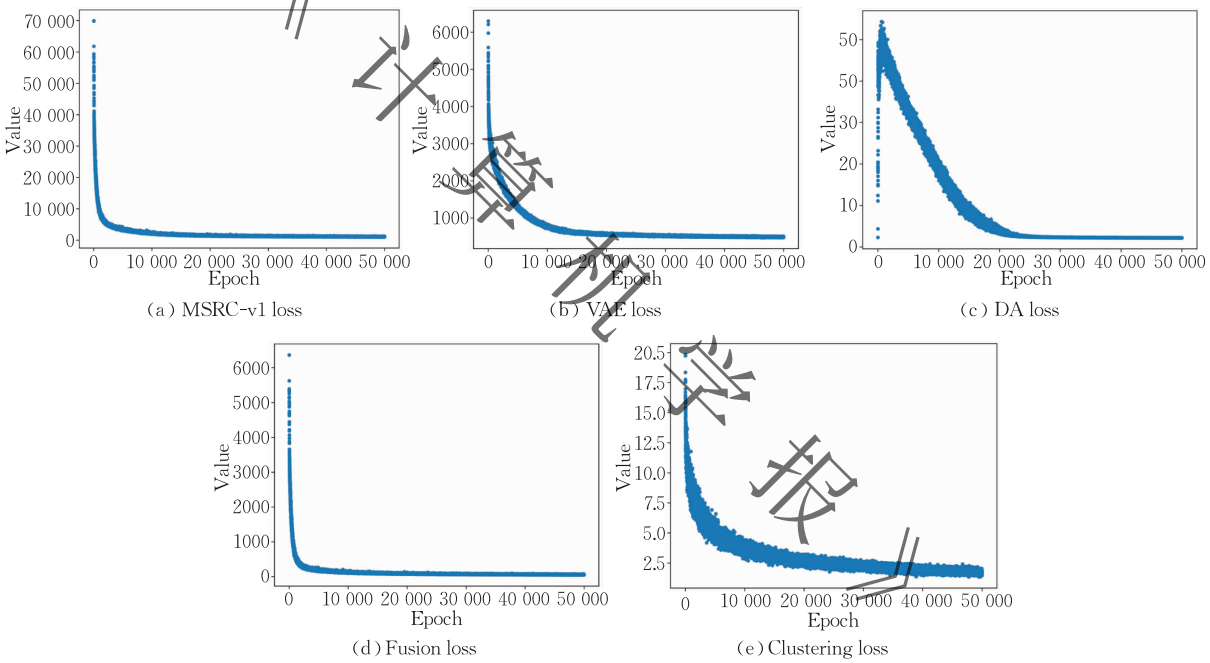


图 5　我们的 DMVCDA 算法在 MRSC-v1 数据集的总体损失和各模块损失的收敛图

我们在五个实验数据集中选取三个不同规模的数据集进行算法效率对比实验，运行时间如表 4 所示. 对于传统方法，我们记录从所有样本输入到输出聚类结果的时间；深度方法则记录将批尺寸(batch size)设为所有样本数，并执行一次 epoch 得到一次聚类结果的时间. 表 4 中数值为取 10 次结果给出平均值(以 s 为单位).

从表 4 实验结果中我们可以看到，在中小数据集 NUS-WIDE 和 Yale 上，我们的方法运行时间仅次于 AWP 算法，虽然在运行效率上并不突出，但结合表 3 结果来看，我们方法聚类效果更好. 这可能与我们网络结构设置的层数和神经元数量有关. 而在大规模数据集 ALOI 上，基于深度模型的方法在运行时间上优势较为明显，如 MVCVAE 与我们的 DMVC-DA，但我们方法比MVCVAE更快. 综上分析，能取得

表 4　算法运行时间对比　(单位：s)

Methods	Yale(165)	NUS-WIDE(2000)	ALOI(10 800)
AWP	**0.07**	**0.90**	36.10
AASC	0.14	2.47	39.40
RMSC	0.11	36.80	388.00
WMSC	0.09	1.24	71.70
SGF	0.11	1.17	14.30
DFG	0.10	1.17	14.40
MVGL	1.00	54.30	5330.00
MCGC	0.87	55.80	713.00
DMFMVC	0.82	42.50	325.50
MVCVAE	0.18	1.22	7.24
DMVCDA	0.15	1.05	**5.68**

优异的表现.

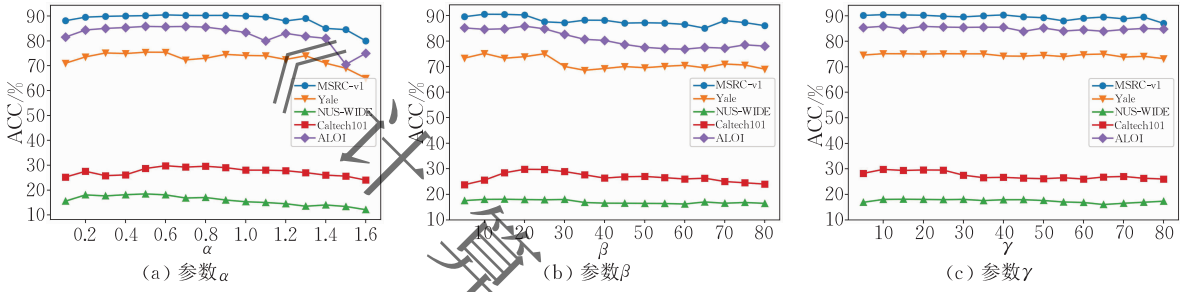
4.5 参数分析

本小节中,我们对模型总体目标函数即式(13)中的三个平衡参数 α、β 和 γ 进行分析. 在分析一个参数时,将其他两个参数设置为固定值,我们在比较实验中对三个模型参数的固定值设置如表 5 所示.

表 5 对比实验中模型参数设置

Datasets	α	β	γ
MSRC-v1	0.54	10	10
Yale	0.54	10	10
NUS-WIDE	0.54	10	10
Caltech101	0.54	20	10
ALOI	0.54	20	10

从图 6 分析可知,参数 α 控制视图分布对齐程度,其值过小给网络带来的影响并不明显,而过大则会带来强约束导致聚类性能下降. 但是,从图 6(a)中可以发现,对于各个数据集都有一个较宽的最优取值范围;对于参数 β,影响的是共享潜在表示 z^* 融合视图信息的能力,不同的数据集对 β 的敏感度不同,但都存在最优的取值区间;参数 γ 是聚类损失项的平衡参数,从图 6(c)中来看,模型对 γ 在取值范围内相对不敏感. 因此,为了提高模型的调参效率,我们在比较实验中对参数 α 在五个数据集的公共最优取值区间[0.5,0.6]取值,参数 β 在 Caltech101 和 ALOI 数据集取 20,其他为 10,而 γ 统一设置为 10.

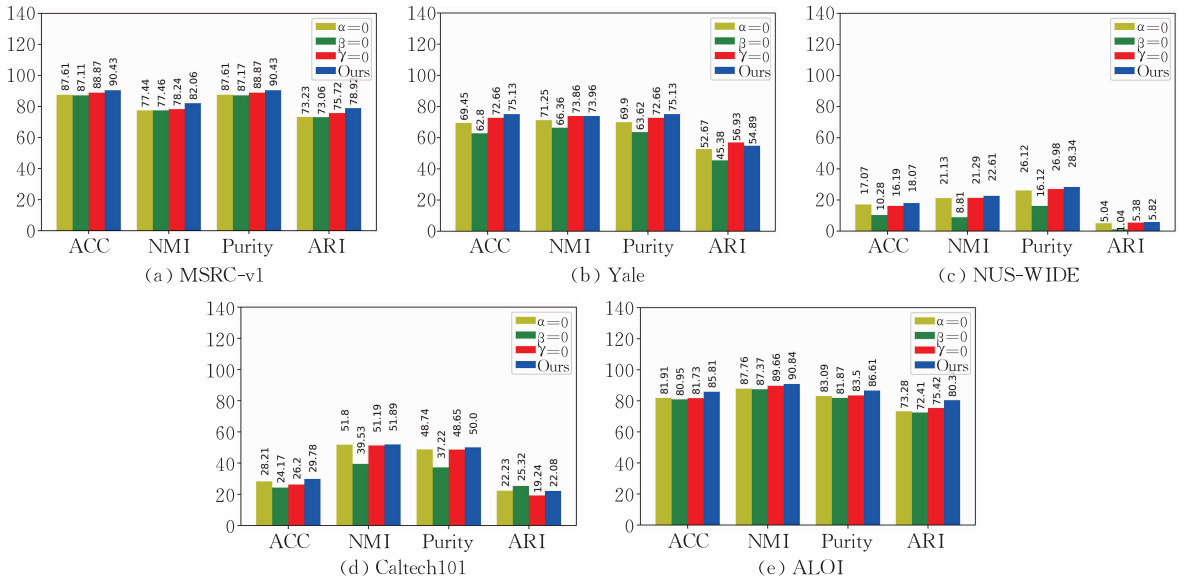


(a) 参数 α (b) 参数 β (c) 参数 γ

图 6 在所有五个数据集上对参数 α,β 和 γ 进行分析

4.6 消融实验

为了考虑不同模块对网络结构的合理性贡献,我们分别对网络中的各个模块进行消融实验,$\alpha=0$ 表示去掉自编码器中分布对齐模块,其余模块保持不变;$\beta=0$ 表示去掉加权融合模块,对多个视图的潜在表示使用拼接的方法得到共享的潜在表示;$\gamma=0$ 为去掉网络中的聚类损失模块,改为使用 K-means 聚类算法对得到的共享潜在表示进行聚类,Ours 则为我们 DMVCDA 模型,即把所有模块融合为一体,实验结果如图 7 所示,表中结果取 10 次平均值. 我们可以清楚地发现本文模型把各模块融为一体. 展示了最好的性能. 结合五个数据集上的表现来看,各个模块在不同数据集上发挥的作用有稍许不同,原因可能是由于数据集的大小不同和视图的维度不



(a) MSRC-v1 (b) Yale (c) NUS-WIDE

(d) Caltech101 (e) ALOI

图 7 消融对比实验

同，但不难看出，其中起最大作用的是加权融合模块，其次是分布对齐模块，通过这两个模块对多视图间信息的学习和融合，有效地学习到多个视图信息间的一致性和互补性信息，表明我们的模型能够从各种视图信息来源中高效地学习共享的潜在表示，由此证明了我们方法的有效性.

4.7　大规模多视图数据实验

随着信息技术的迅速发展，各种数据的数量也随之剧增，如何有效解决大规模多视图数据的聚类分析，也给研究者们带来了巨大的挑战. 由于传统的聚类算法需要求解约束优化问题，如二次规划，或需要求解多个矩阵特征值和特征向量，如谱聚类等，它们的计算复杂度和空间复杂度较高，无法处理大规模高维的多视图数据. 考虑到本文方法基于深度学习模型，可扩展应用于处理大规模高维数据，因此我们将采用大规模多视图数据集 Large-ALOI 进一步测试我们的方法. Large-ALOI 是 ALOI 数据集的大规模数据版本，包含 110 250 个样本，4 个视图，共 1000 个类别. 由于前述提到的对比方法大部分无法处理该数据集或无开源代码供测试，故本文只与 MVCVAE 方法进行比较，实验结果取 10 次平均值，如表 6 所示. 由表所知，本文方法在四项指标中均优于 MVCVAE 方法，也验证了本文方法在处理大规模多视图数据的优势.

表 6　在 Large-ALOI 数据集上聚类性能对比

Methods	ACC	NMI	Purity	ARI
MVCVAE	30.11±0.25	61.52±0.74	36.44±0.81	18.53±0.52
DMVCDA	**34.62±0.12**	**68.79±0.53**	**38.08±0.77**	**21.79±0.55**

5　结　论

为更合理地利用多视图数据提升聚类性能，我们在遵循多视图学习的一致性和互补性的原则下，提出了基于分布对齐变分自编码器的深度多视图聚类 DMVCDA 方法. 具体来说，通过最小化各个视图的 Wasserstein 距离对齐潜在空间分布以进行一致性学习，并利用变分自编码器重构损失的约束保留视图特有的特征信息，学习多视图的互补信息. 为考虑到不同视图信息对全局的重要性存在差异，引入一组自适应权重融合视图信息得到多视图共享潜在表示，最后联合深度嵌入聚类损失与网络一同进行优化，获得最终聚类结果. 实验结果表明，本文提出的方法与其他较新的多视图聚类方法相比，在多项指标上表现优异，并且我们的方法可以应用于更大规模多视图数据上.

致　谢　在此，对各位评审专家表示由衷的感谢！

参　考　文　献

[1] Bickel S, Scheffer T. Multi-view clustering//Proceedings of the 4th IEEE International Conference on Data Mining. Brighton, UK, 2004: 19-26

[2] de Sa V R. Spectral clustering with two views//Proceedings of the ICML Workshop on Learning with Multiple Views. Bonn, Germany, 2005: 20-27

[3] Huang H C, Chuang Y Y, Chen C S. Affinity aggregation for spectral clustering//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 773-780

[4] Xia R, Pan Y, et al. Robust multi-view spectral clustering via low-rank and sparse decomposition//Proceedings of the AAAI Conference on Artificial Intelligence. Québec City, Canada, 2014: 2149-2155

[5] Zong L, Zhang X, Liu X, et al. Weighted multi-view spectral clustering based on spectral perturbation//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 4621-4628

[6] Hu Shi-Zhe, Lou Zheng-Zheng, Wang Ruo-Bin, et al. A dual-weighted multi-view clustering method. Chinese Journal of Computers, 2020, 43(9): 1708-1720(in Chinese)
(胡世哲，娄铮铮，王若彬等. 一种双重加权的多视角聚类方法. 计算机学报，2020，43(9)：1708-1720)

[7] Zhan K, Zhang C, Guan J, et al. Graph learning for multiview clustering. IEEE Transactions on Cybernetics, 2017, 48(10): 2887-2895

[8] Zhan K, Nie F, Wang J, et al. Multiview consensus graph clustering. IEEE Transactions on Image Processing, 2018, 28(3): 1261-1270

[9] Liang Y, Huang D, Wang C D, et al. Multi-view graph learning by joint modeling of consistency and inconsistency. arXiv preprint arXiv:2008.10208, 2020

[10] Xie J, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis//Proceedings of the 33rd International Conference on Machine Learning. New York, USA, 2016: 478-487

[11] Guo X, Liu X, Zhu E, et al. Deep clustering with convolutional autoencoders//Proceedings of the International Conference on Neural Information Processing. Guangzhou, China, 2017: 373-382

[12] Ngiam J, Khosla A, Kim M, et al. Multimodal deep learning//Proceedings of the 28th International Conference on Machine Learning. Bellevue, USA, 2011: 689-696

[13] Wang W, Arora R, Livescu K, et al. On deep multi-view representation learning//Proceedings of the 32nd International Conference on Machine Learning. Lille, France, 2015: 1083-1092

[14] 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

[15] Sun G, Cong Y, Zhang Y, et al. Continual multiview task learning via deep matrix factorization. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 139-150

[16] Zheng Q, Zhu J, Ma Y, et al. Multi-view subspace clustering networks with local and global graph information. Neurocomputing, 2021, 449: 15-23

[17] Wang Q, Cheng J, Gao Q, et al. Deep multi-view subspace clustering with unified and discriminative learning. IEEE Transactions on Multimedia, 2020, 23: 3483-3493

[18] Xu C, Guan Z, Zhao W, et al. Deep multi-view concept learning//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018: 2898-2904

[19] 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, 34(4): 6688-6695

[20] Kingma D P, Welling M. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114, 2013

[21] Higgins I, Matthey L, Pal A, et al. beta-VAE: Learning basic visual concepts with a constrained variational framework //Proceedings of the 5th International Conference on Learning Representations(ICLR). Toulon, France, 2017

[22] Gretton A, Borgwardt K M, Rasch M J, et al. A kernel two-sample test. The Journal of Machine Learning Research, 2012, 13(1): 723-773

[23] Suzuki T, Sugiyama M. Sufficient dimension reduction via squared-loss mutual information estimation//Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, Chia Laguna Resort. Sardinia, Italy, 2010: 804-811

[24] Givens C R, Shortt R M. A class of Wasserstein metrics for probability distributions. The Michigan Mathematical Journal, 1984, 31(2): 231-240

[25] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533-536

[26] Kingman D P, Ba J. Adam: A method for stochastic optimization //Proceedings of the International Conference for Learning Representations(ICLR). San Diego, USA, 2015

[27] Bowman S R, Vilnis L, Vinyals O, et al. Generating sentences from a continuous space. arXiv preprint arXiv:1511.06349, 2015

[28] Winn J, Jojic N. LOCUS: Learning object classes with unsupervised segmentation//Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05), Volume 01. Beijing, China, 2005: 756-763

[29] Chua T S, Tang J, Hong R, et al. Nus-wide: A real-world web image database from national university of Singapore// Proceedings of the ACM International Conference on Image and Video Retrieval. Santorini Island, Greece, 2009: 1-9

[30] Geusebroek J M, Burghouts G, Smeulders A. The amsterdam library of object images. International Journal of Computer Vision, 2005, 61(1): 103-112

[31] Houle M E, Oria V, Satoh S, et al. Knowledge propagation in large image databases using neighborhood information// Proceedings of the 19th ACM International Conference on Multimedia. New York, USA, 2011: 1033-1036

[32] Nie F, Tian L, Li X. Multiview clustering via adaptively weighted Procrustes//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA, 2018: 2022-2030

[33] 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, 34(4): 3513-3520

[34] Van Der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008, 9(86): 2579-2605

XIE Sheng-Li, Ph. D., professor, Ph. D. supervisor. His research interests include intelligent signal processing, biomedical signal processing, and artificial intelligence.

CHEN Hong-Da, M. S. candidate. His research interests include data mining and deep learning.

GAO Jun-Li, Ph. D., associate professor. His research interests cover robot control and data mining.

PENG Xi, Ph. D., professor. His research interests cover computer vision and data mining.

YIN Ming, Ph. D., professor. His research interests cover computer vision and multi-view learning.

Background

With the rapid development of the Internet and sensor technology, the data in many fields are growing massively. In the practical applications, data are usually collected from different fields and sensors, resulting in multi-view data. Multi-view data means that the same object can be described by multiple different data sources or features, and each data source or feature can be viewed as a specific view. Multi-view clustering is an important approach for pattern discovery on complex data. However, most of these existing methods neither effectively learn the latent relationship among multiple views nor consider the different importance of each view.

In this work, a deep multi-view clustering is proposed based on distribution aligned variational autoencoder, so as to improve performance of multi-view clustering. First, we use view-specific variational autoencoder to extract latent features from different views, and align the learned view data distribution to further mine latent features containing basic information. Then, we introduce view weight parameters to fuse the view-specific features into a shared latent one. Finally, the loss function for clustering is established on the latent features, so that the learned latent features are more suitable for clustering tasks, thereby improving the clustering accuracy. The experimental results on five common multi-view datasets show that our model has achieved excellent performance in terms of multiple clustering evaluation metrics, such as accuracy (*ACC*), normalized mutual information (*NMI*) and purity (*Purity*), which validates the effectiveness of our model.

This work has been supported by the National Natural Science Foundation of China (No. 61876042), the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515-011493).