

基于卷积注意力和对比学习的多视图聚类

倪团雄^{1,2}, 洪智勇^{1,2}, 余文华^{1,2}, 张 昕¹

(1. 五邑大学 智能制造学部, 广东 江门 529020;

2. 粤港澳工业大数据协同创新中心, 广东 江门 529020)

摘 要:多视图聚类能够综合不同视图的互补信息,往往能获得比单一视图更好的效果。然而,传统多视图聚类方法受限于线性及浅层的学习函数,难以表征数据的深层信息;现有的深度学习方法在表征多视图数据时,对多维度的细节特征关注度有所不足。针对这些问题,提出一种基于卷积注意力机制的编码器模型(AEMC),该模型根据不同视图的特定表征,在编码器中融入卷积注意力模块自适应学习各个视图的关键特征,此外,为了优化模型,根据编码器表征,通过对比学习策略构造正负样本,使正样本间的相似度增加,负样本的相似度减少,引导聚类过程从而使其更具鲁棒性。实证结果表明,模型优于当前大多数主流方法,并在E-MNIST、E-FMNIST、VOC和RGB-D数据集上聚类精度比基准模型分别提高了10.2%、8.1%、7.4%和4.9%,在手写数据集E-MNIST和E-FMNIST的聚类准确率分别高于目前最优的对比聚类方法(CoMVC)0.7%和1.3%,在VOC、RGB-D数据集上略低于对比聚类方法(CoMVC)。

关键词:编码器;多视图聚类;卷积注意力;对比学习;深度学习

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2023)08-0059-07

doi:10.3969/j.issn.1673-629X.2023.08.009

Multi-view Clustering Based on Convolution Attention and Contrast Learning

NI Tuan-xiong^{1,2}, HONG Zhi-yong^{1,2}, YU Wen-hua^{1,2}, ZHANG Xin¹

(1. Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China

2. Guangdong Hong Kong Macao Industrial Big Data Collaborative Innovation Center, Jiangmen 529020, China)

Abstract: Multi-view clustering can synthesize complementary information from different views, and often results better than a single view. However, the traditional multi-view clustering method is limited by linear and shallow learning functions, which is difficult to characterize the deep information of the data. Existing deep learning methods pay insufficient attention to multi-dimensional detailed features when characterizing multi-view data. In order to solve these problems, an encoder model based on convolutional attention mechanism (AEMC) is proposed, which integrates the convolutional attention module into the encoder according to the specific representation of different views to adaptively learn the key features of each view. In addition, in order to optimize the model, according to the encoder representation, positive and negative samples are constructed through a comparative learning strategy, so that the similarity between positive samples increases and the similarity of negative samples decreases, guiding the clustering process to make it more robust. The empirical results show that the model is better than most current mainstream methods, and its clustering accuracy on the E-MNIST, E-FMNIST, VOC and RGB-D datasets is improved by 10.2%, 8.1%, 7.4% and 4.9% compared with the benchmark model, respectively, and the clustering accuracy of E-MNIST and E-FMNIST datasets is higher than that of the current optimal Comparative Clustering Method (CoMVC) by 0.7% and 1.3%, respectively, slightly lower than that of the Contrasting Clustering Method (CoMVC) on the RGB-D dataset.

Key words: encoder; multi-view clustering; convolutional attention; comparative learning; deep learning

收稿日期:2022-10-31

修回日期:2023-03-02

基金项目:广东省基础与应用基础研究基金(2020A151011468);广东省普通高校特色创新类项目(2019KTSCX189);五邑大学港澳联合研发基金(2019WGALH21)

作者简介:倪团雄(1996-),男,硕士研究生,CCF会员(L8384G),研究方向为深度多视图聚类算法;通讯作者:洪智勇(1978-),男,教授,博导,博士,研究方向为大数据技术、数据挖掘;余文华(1960-),男,教授,博导,博士,研究方向为高性能计算、深度学习;张 昕(1965-),男,教授,博导,博士,研究方向为人工智能。

0 引言

随着信息社会的发展,数据的表示形式更加丰富,单一数据形式往往难以充分表达数据的完整信息,而多视图数据可以弥补样本特征单一的缺点,更加全面地表征数据。多视图数据是指从同一数据源提取的不同特征集或者多个来源的数据^[1]。例如,传感器信号的时域和频域特征、图像的颜色和纹理等特征、三维物体不同角度的成像特征等等。然而,实际应用中很难获取经过有效标注的数据样本,而无监督学习只需要关注数据本身,数据之间的映射关系存在很大的研究空间。聚类作为无监督学习重要的算法之一,可解释性强,学习参数少,因此无监督的多视图聚类学习吸引了学者们的广泛关注。多视图聚类根据数据本身的特征,综合不同视图的互补信息,通过学习共识函数,将具有相似结构的数据划分到不同的类簇中,已经在数据分析^[2]、生物信息^[3]、自然语言处理^[4]、社交网络^[5]等领域得到关注和应用。因此,多视图聚类学习具有重要的研究意义和应用场景。

现有的多视图聚类方法主要分为两类:传统机器学习方法和深度学习方法。在传统算法中,一些机器学习方法^[6-7]将特征表示和聚类过程互相分离,这种方式不利于后续的聚类优化。为了改进这种缺陷,基于子空间学习的方法^[8-14]假设所有的视图数据共享一个潜在的公共子空间,通过探索多视图的互补信息将多视图数据投影得到共同的低维映射。如 Zhang 等^[12]直接通过多视图学习子空间表示,分别沿着交替方向最小化和广义拉格朗日乘子优化模型,使得每个视图的信息更加互补。基于协同训练的聚类算法^[15-16]则根据先验知识充分地学习促进不同视图的信息,达到最大限度的共识聚类。如 Kumar 等^[5]较先提出具有协同训练思维的多视图谱聚类,比如成对共正则谱聚类。算法假设一对数据点之间的关系在所有视图中保持一致的类簇,使得两个点在所有视图内要么都出现在同一簇中,要么都属于不同的簇。受到相同启发,Zhao 等人^[16]基于协同训练框架,提出了联合 K-均值聚类和线性判别分析^[17]的多视图聚类算法,利用在一个视图学习的标签引导其他视图的判别子空

间。其他主流算法如基于图^[18-22]和多核学习方法^[23-27],它们都假设多视图数据可以用一个或多个核矩阵构成特征空间。以上这些传统方法虽然解释性较强,但受限于浅层和线性的学习函数,难以表达数据特征的深层信息,从而影响聚类效果。基于深度学习方法的聚类工作原理与上述方法类似,只是表征学习采用深度神经网络。例如,基于深度图学习的方法^[28-31]使用图神经网络结合亲和度矩阵对多视图数据进行聚类。类似地,深度子空间方法^[32-33]与上述都有相同的子空间假设,但需要计算深度神经网络中间表示的自表示矩阵。另外,对抗性方法^[34-35]通过生成器和判别器对齐不同视图中隐藏特征表示的分布。如 Li 等^[34]通过深度自动编码器学习各个视图共享的潜在表示,同时利用对抗训练来进一步捕获数据分布和分解潜在的空间。这些深度模型虽然能学习到视图的深层信息,但卷积运算主要作用在图像的局部近邻上,往往会丢失全局的关键信息,难以合理分配特定视图的权重。

针对这些问题,笔者认为,基于深度学习的多视图聚类方法关键在于如何提取关键的视图信息及如何融合不同视图特征从而引导聚类过程。受卷积注意力算法^[36]的启发,该文提出一种基于卷积注意力和对比学习的深度多视图聚类网络(AEMC)。主要贡献在于:首先,该模型结合卷积层和卷积注意力模块,构造编码器网络,通过两个维度的注意力模块,提取每个视图的高层次细节特征,其次,为了引导聚类的优化方向,使用对比学习策略,通过构造正负样本间接引导模型的聚类效果。

1 基于卷积注意力的深度多视图聚类

1.1 卷积注意力模块

卷积注意力模块(CBAM)^[36]是一种沿着两个维度(通道和空间)计算注意力图的深度网络模块。假设给定一个特征图 F ,CBAM模块将依次通过通道和空间注意力模块,分别在两个维度上学习关键信息和位置,然后将注意力权重和特征图相乘,输出特征图的维度保持不变。卷积注意力模块的具体结构如图1,具体实现步骤如下:

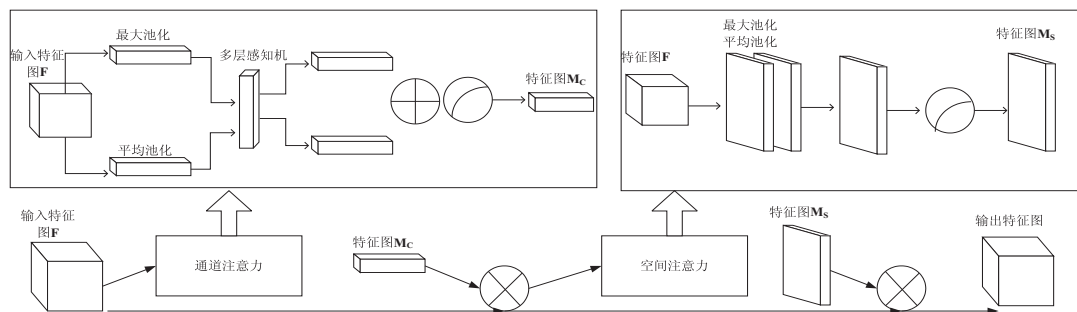


图1 卷积注意力模块

(1)将输入的特征图 $F(H \times W \times C)$ 分别经过基于宽和高的全局池化层和平均池化层,压缩得到两个一维矢量 F_{avg}^c 和 $F_{\text{max}}^c (1 \times 1 \times C)$,接着送到一个共享的多层感知器网络,再将各元素求和合并,经过激活函数后产生通道注意力图 M_c :

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(\varphi_0(\varphi_1(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (1)$$

其中, σ 、 φ_0 、 φ_1 分别是激活函数和多层感知器的参数。

(2)将通道注意力中得到的特征图和原始特征图对应元素逐一相乘得到空间注意力模块的输入特征 F' ,同理将 F' 通过平均池化和最大池化转化成特征图 F_{avg}^s 和 F_{max}^s ,经过激活函数后形成空间注意力图 M_s :

$$M_s(F) = \sigma(f^{(7 \times 7)}([\text{AvgPool}(F'); \text{MaxPool}(F')])) = \sigma(f^{(7 \times 7)}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (2)$$

将该模块融入设计的深度多视图聚类模型中,通过提取通道和空间多维度的特征信息,平衡全局和局部注意力特征,从而影响后续的多视图聚类效果。

1.2 模型结构

假设给定样本总数为 n 的多视图数据集 $X = \{X^{(1)} \dots X^{(m)}\}$, $X_i^{(v)} \in R^{(n \times d_v)}$ 表示视图 v 的维度为 d_v 的第 i 个数据样本,多视图聚类模型的目标是综合不同视图的数据样本 $\{x_i^{(1)}, \dots, x_i^{(v)}\}$,将其聚类到其中一个 k 种类簇中。如图2,提出的多视图聚类模型主要分成两部分:

(1)编码器网络 $e(i)$ 。在编码器模型结构中,将卷积注意力模块分别嵌入3层卷积神经网络中,构成编码器网络,每一层的结构如图3,其中卷积层采用 3×3 的卷积核,激活函数使用 Relu,池化层选择窗口为 2×2 ,步长为2的最大池化。

(2)聚类模块。该模型采用深度发散的聚类(DDC)^[37]方法作为基准聚类模块,结合视图编码器特征,将编码器学习得到的 k 维向量 a_i 用于聚类。受对比学习思想的启发,将经过数据增强的视图样本和融合后的样本作为正样本,其他的样本作为负样本,同时优化深度发散聚类和对比损失,更新网络参数从而优化聚类模型,具体见1.3节。编码器模块对每个数据样本 $x_i^{(v)}$ 提取特征,得到特征表示 $z_i^{(v)}$:

$$z_i^{(v)} = e(x_i^{(v)}) \quad (3)$$

接着,将不同视图的特征进行融合,得到公共表示 z_i :

$$z_i = \sum_{v=1}^V w_i z_i^{(v)} \quad (4)$$

其中, w_i 是初始化视图的权重。

最后,将公共表示 z_i 通过带 softmax 激活函数的全连接层进行降维,从而得到 k 维的输出向量 a_i 。

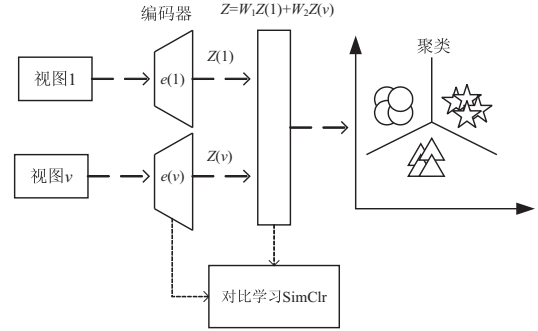


图2 聚类模型(主要由编码器模块和聚类模块组成:编码器用于提取视图的关键信息,聚类模块通过对比方法和深度发散聚类引导聚类过程)

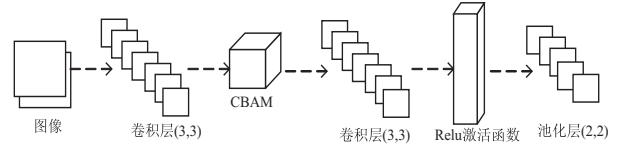


图3 编码器结构

1.3 损失函数

损失函数用于衡量模型的预测值和真实值的差异程度。该文选择基于深度发散的聚类(DDC)^[37]损失作为基准聚类损失,该聚类损失由三部分组成:多密度泛化的柯西—施瓦茨发散项,其公式为:

$$L_1 = \sum_{i=1}^{k-1} \sum_{j=i+1}^k \frac{(k_2)^{-1} \sum_{a=1}^n \sum_{b=1}^n a_{ai} k_{ab} a_{bj}}{\sqrt{\sum_{a=1}^n \sum_{b=1}^n a_{ai} k_{ab} a_{bi} \sum_{a=1}^n \sum_{b=1}^n a_{aj} k_{ab} a_{bj}}} \quad (5)$$

其中, k 表示聚类簇的数量, k_{ab} 为高斯核函数,表示为:

$$k_{ab} = e^{-\frac{\|z_a - z_b\|_2^2}{2\theta^2}} \quad (6)$$

其中, θ 是超参数。

第二项为不同数据样本之间的正交聚类分配:

$$L_2 = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j>i}^n a_i^T a_j \quad (7)$$

最后一项则将聚类分配向量逼近标准单纯形:

$$L_3 = \sum_{i=1}^{k-1} \sum_{j=i+1}^k \frac{\binom{k}{2}^{-1} \sum_{a=1}^n \sum_{b=1}^n m_{ai} k_{ab} m_{bj}}{\sqrt{\sum_{a=1}^n \sum_{b=1}^n m_{ai} k_{ab} m_{bi} \sum_{a=1}^n \sum_{b=1}^n m_{aj} k_{ab} m_{bj}}} \quad (8)$$

其中, $m_{bj} = \exp(-\|a_b - e_j\|^2)$,总的聚类损失 L_c 由式(5)(7)(8)组成,即:

$$L_c = L_1 + L_2 + L_3 \quad (9)$$

然而聚类过程还缺乏一定的引导方向,同时为避免模型过拟合以及优化模型,该文在聚类损失中融入对比损失函数,共同优化总的损失函数。受对比学习

方法 SimClr^[38] 的启发,将其重构并融入模型的聚类模块中。文中模型将 SimClr 的损失函数重构表示如下:

$$L_s = -\log \frac{\exp(\text{sim}(p(z), p(z_i))/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{k \neq i} \exp(\text{sim}(p(z), p(z_k))/\tau)} \quad (10)$$

其中, $\exp(\cdot)$ 函数表示以自然数 e 为底的指数函数, $p(\cdot)$ 采用 1 层隐藏层的多层感知器,将其映射到对比空间, $\text{sim}(\cdot)$ 表示余弦相似度, τ 为温度超参数。在该模型中,选择经过 dropout^[39] 数据增强后的视图样本和融合后的样本作为正样本,其他的样本作为负样本,对比损失目标在于增加正样本间的相似度,减少和负样本的相似度。因此,最后模型总的优化损失函数 L 表示为:

$$L = L_c + \rho L_s \quad (11)$$

其中, ρ 为对比损失参数。

2 实验

2.1 数据集

在四组公共多视图数据集上评估所提出的模型(AEMC),并将其与其他主流模型进行比较。四个数据集分别是:

(1) E-MNIST: E-MNIST 是一组公共基准数据集,由 28×28 像素的 60 000 个手写数字图像(10 个类别)组成。在实验中使用由文献[26]提供的版本(包含 60 000 个样本),包含原始的灰色图像视图和数字边缘视图。

(2) E-FMNIST: 这是手写数据集 MNIST 的另一个版本,数据视图同样包括原始数字和边缘检测双视图。

(3) PASCAL VOC2007 (VOC): 采用由文献[40]提供的数据集,总共 20 种类别,数据集包含人工标注的自然图像 GIST 特征和词频计数特征。

(4) SentencesNYU v2 (RGB-D), 由 1 449 张经过标注的室内场景的图像以及它们的相关描述信息组成。具体如表 1 所示。

表 1 数据集概述

数据集	样本数	类别	视图	维度
E-MNIST	60 000	10	2	28×28
E-FMNIST	60 000	10	2	28×28
VOC	5 649	20	2	51,399
RGB-D	1 449	13	2	2 048,300

2.2 实验设置

实验系统环境为 Red Hat 4.8.5,硬件环境:GPU 为 16 GB 显存的 NVIDIA Quadro P500, CPU 为 Intel Xeon 的 GOLD 5118 处理器,软件环境:cuda10.02 和

python3.7,采用 Pytorch 深度学习框架搭建模型。

在公开的基准数据集上验证模型的效果,主要的超参数设置如下:实验的训练批次 epoch 设置为 100,使用默认参数的 ADAM 优化器优化模型,CBAM 模块使用文献[31]中的默认参数,超参数 θ 设置为 0.2,多视图权重 w_i 初始化为均值 0.5,温度超参数 τ 采用文献[33]推荐的 $\tau = 0.1$,学习率 lr 设置为 0.01,对比损失参数 ρ 设置为 0.05。为了降低实验结果的随机性干扰,在每组数据集重复训练 5 次并取其平均值作为实验评估结果。并和当前几种经典深度模型进行对比,同时,将深度发散聚类模型(DDC)作为独立参考模型,并通过消融实验对模型的组成部分进行分析对比。

评估指标:

为更精准地评价聚类算法,采用聚类准确率(ACC)和标准互信息(NMI)来评价深度聚类算法的性能。其中 ACC 表示正确聚类的样本占总样本的比例,NMI 则是基于信息论思想,用于度量两组样本之间的相似度,两组评价指标的值越大代表聚类效果越好。假设 y 和 y' 分别表示模型的聚类标签和数据集的真实标签,ACC 计算公式如下:

$$\text{ACC} = \frac{\sum_{i=1}^n \delta(y', \text{map}(y))}{n} \quad (12)$$

其中, $\delta(y', \text{map}(y)) = \begin{cases} 1, & \text{当 } y' = \text{map}(y) \\ 0, & \text{其他} \end{cases}$, $\text{map}(y)$ 为

排列映射函数。

假设 n 个数据样本聚类得到的类别 $C = c_1, c_2, \dots, c_k$, 真实聚类为 $L = l_1, l_2, \dots, l_j$, NMI 表达式为:

$$\text{NMI}(C, L) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^j \frac{I(l_j, c_k)}{\max(H(l_j), H(c_k))} \quad (13)$$

其中, $I(l, c) = H(l) - H(l|c)$ 表示互信息, $H(X)$ 为信息熵。

2.3 对比模型

为了评估模型的性能,将提出的基于卷积注意力编码器的多视图聚类模型(AEMC)同当前主流的多视图聚类模型进行对比,对比模型包括:

- (1) 深度多模态子空间聚类(DMSC)^[32];
- (2) 基于端到端对抗性注意力网络的多模态聚类(EAMC)^[35];
- (3) 深度典型相关性分析(DCCA)^[41];
- (4) 深度对抗多视图聚类(DAMC)^[34];
- (5) 对比多视图聚类(CoMVC)^[42]。

另外,为了检验改进后的模型整体结构相对基准聚类是否提高聚类效果,从而证明模型的卷积注意力编码器模型和对比学习策略的优势,将基准深度发散聚类模型(DDC)作为独立对比组。

2.4 结果与分析

2.4.1 模型对比结果分析

定量研究了 E-MNIST、E-FMNIST、VOC、RGB-D 四组数据集在卷积注意力编码器模型上的表现,结果见表 2。从表中对比结果可知,文中模型在手写数据集 E-MNIST 和 E-FMNIST 上的聚类准确率,标准互信息优于大多数模型,聚类准确率分别领先文献[42]中的对比聚类方法(CoMVC)0.7 百分点和 1.3 百分点,在 E-MNIST 上的标准互信息高于对比聚类方法(CoMVC)1.2 百分点;在 VOC 和 RGB-D 数据集上则优于大多数主流模型,但略低于对比聚类方法。

另外,在独立对比实验中,进一步探索了模型(AEMC)和深度发散聚类基准模型(DDC)在 4 组数据集上的聚类效果,具体见表 3。从表中数据可知,该模型在 4 组数据集上较深度发散基准模型都有明显的改善,特别是在 E-MNIST 公共数据集上,聚类准确率

和标准互信息分别提高了 10.2 百分点和 8.1 百分点。

综合以上分析,将卷积注意力模块嵌入编码器网络中以及聚类优化中使用对比学习策略,能够提取视图的更多关键特征,引导模型的聚类方向,从而提高聚类效果。聚类结果表明,多视图融合表征的质量有助于对比学习策略的使用,两者又相互促进聚类的效果。因此说明该模型具备一定的有效性和鲁棒性。

2.4.2 损失函数研究

为更直观地观察基准损失和模型损失函数的变化,以数据集 VOC 为例,将数据集的基准聚类损失和总的损失函数进行可视化。模型在训练过程中 Loss 曲线随迭代批次 epoch 的变化如图 4 所示,其中,依次选择每 10 次迭代的 Loss 值构成散点。从图中曲线变化得知,模型经 100 次训练后逐渐收敛,基准模型和文中模型在训练 100 批次后损失函数也趋于稳定,因此 epoch 的设置合理,模型训练足够充分。

表 2 不同数据集上不同模型的对比结果 %

模型	E-MNIST		VOC		E-FMNIST		RGB-D	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
DCCA	47.6	44.3	39.7	42.5				
DMSC	65.3	61.4	54.1	53.8				
DAMC	64.6	59.4	56.0	55.2				
EAMC					55.2	62.5	32.3	20.7
CoMVC	95.5	90.7	61.9	67.5	65.3	61.4	41.3	40.5
AEMC	96.2	91.9	60.7	64.6	67.6	60.1	38.5	34.6

表 3 不同数据集上 AEMC 和 DDC 模型的对比结果 %

模型	E-MNIST		VOC		E-FMNIST		RGB-D	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
DDC	85.0	83.0	53.3	47.6	59.5	56.4	33.6	31.5
AEMC	96.2	91.9	60.7	64.6	67.6	60.1	38.5	34.6

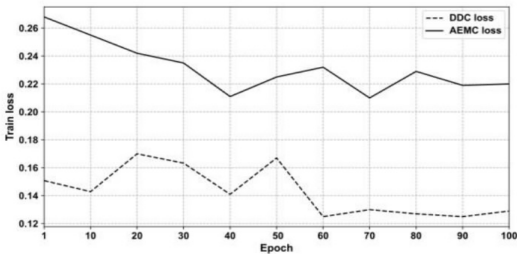


图 4 模型 Loss 变化曲线

2.4.3 消融实验

前面的实验初步证明模型具有一定的效果,然而模型的具体结构对聚类结果的影响程度还不清楚。因此,为了进一步探索模型具体结构的实际聚类效果,对模型不同结构组合的聚类结果进行消融对比实验。如表 4,实验分别设置无 CBAM 和正负对比样本,有 CBAM 模块但无正负对比样本,无 CBAM 模块但有正

负对比样本三组对比参照组。

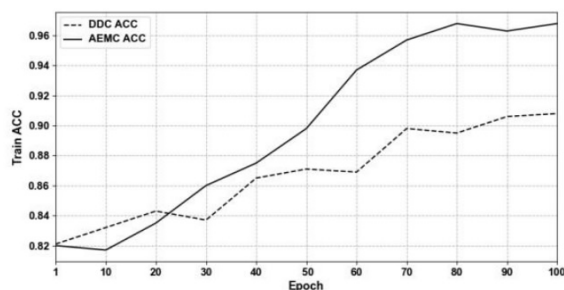
表 4 在 E-MNIST、VOC 数据集上的消融结果 %

数据集	CBAM	L_s	ACC	NMI
E-MNIST	×	×	89.7	90.2
	✓	×	92.3	91.2
	×	✓	87.3	89.5
	✓	✓	96.2	91.9
VOC	×	×	53.7	52.0
	✓	×	56.2	57.5
	×	✓	55.5	49.6
	✓	✓	60.7	64.6

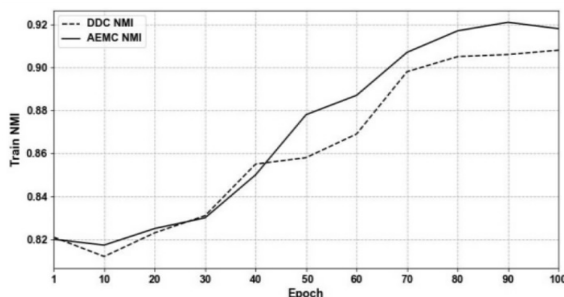
从表中结果可以发现,当模型缺少卷积注意力模块(CBAM)和对比损失优化函数 L_s 后,该模型在 E-MNIST、VOC 数据集上的聚类精度分别下降了 6.5 百

分点、7 百分点,同理,无卷积注意力模块(CBAM)和对比损失优化函数 L_s ,标准互信息值分别下降了 1.7 百分点和 12.6 百分点,由此可知卷积注意力模块和对比策略相对模型有所改进。

此外,以 E-MNIST 数据集为例,分别可视化其在训练过程中模型和深度发散基准模型的曲线变化结果,如图 5 所示。其中图 5(a)是 E-MNIST 数据集训练聚类精度随训练迭代次数的变化曲线,模型的精度随着训练次数增加逐渐收敛,并在 95% 附近趋于稳定,且精度高于基准聚类模型。同理,图 5(b)为 E-MNIST 数据集标准互信息随训练批次的变化曲线,在 92% 附近逐渐平缓,同样高于基准聚类模型。因此,消融实验表明模型的卷积模块和对比学习模块促进聚类的效率,提高了聚类的准确率和标准互信息值。



(a) 聚类精度



(b) 标准互信息值

图 5 E-MNIST 数据集的消融实验

3 结束语

针对传统多视图聚类算法浅层学习的限制和深度学习多度特征学习的局限问题,提出一种基于卷积注意力机制的深度多视图聚类网络,将卷积注意力模块结合卷积模块构成编码器网络,提取每个视图的通道和空间关键特征,提高多度特征的关注度。此外,在优化聚类模型时使用对比学习策略,防止过拟合及引导模型的聚类方向。实验结果表明,模型的聚类准确率较高,聚类效果明显。

然而,该研究也存在诸多不足,比如数据集形式单一、模型训练参数较多、训练内存和时间消耗大等等。未来的研究将会关注复杂的多视图数据及网络改进,如图像和文本数据的结合,数据缺失的多视图数据的

处理,以及预训练网络,探索更多类型的数据集是否有利于注意力参数的优化和网络的兼容性。

参考文献:

- [1] 何雪梅. 多视图聚类算法综述[J]. 软件导刊, 2019, 18(4): 79-81.
- [2] ZHAN K, CHANG X, GUAN J, et al. Adaptive structure discovery for multimedia analysis using multiple features[J]. IEEE Transactions on Cybernetics, 2018, 49(5): 1826-1834.
- [3] WANG B, MEZLINI A M, DEMIR F, et al. Similarity network fusion for aggregating data types on a genomic scale[J]. Nature Methods, 2014, 11(3): 333-337.
- [4] KIM Y M, AMINI M R, GOUTTE C, et al. Multi-view clustering of multilingual documents[C]//Proceedings of the 33rd international ACM SIGIR conference on research and development in information. Geneva: ACM, 2010: 821-822.
- [5] LI J H, WANG C D, LI P Z, et al. Discriminative metric learning for multi-view graph partitioning[J]. Pattern Recognition, 2018, 75: 199-213.
- [6] LIVESCU K. Multi-view clustering via canonical correlation analysis[C]//Proceedings of the 26th annual international conference on machine learning. Quebec: ICML, 2009: 129-136.
- [7] BLASCHKO M B, LAMPERT C H. Correlational spectral clustering[C]//2008 IEEE conference on computer vision and pattern recognition. Anchorage: IEEE, 2008: 1-8.
- [8] CAO X, ZHANG C, FU H, et al. Diversity-induced multi-view subspace clustering[C]//2015 IEEE conference on computer vision and pattern recognition (CVPR). Boston: IEEE, 2015: 586-594.
- [9] LIU G, LIN Z, YAN S, et al. Robust recovery of subspace structures by low-rank representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171-184.
- [10] XIE D, ZHANG X, GAO Q, et al. Multiview clustering by joint latent representation and similarity learning[J]. IEEE Transactions on Cybernetics, 2020, 50(11): 4848-4854.
- [11] YANG Z, XU Q, ZHANG W, et al. Split multiplicative multi-view subspace clustering[J]. IEEE Transactions on Image Processing, 2019, 28(10): 5147-5160.
- [12] ZHANG C. Generalized latent multi-view subspace clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(1): 86-99.
- [13] 范瑞东, 侯臣平. 鲁棒自加权的多视图子空间聚类[J]. 计算机科学与探索, 2021, 15(6): 1062-1073.
- [14] 郭圣, 仲兆满, 李存华. 基于深度自编码的多视图子空间聚类网络[J]. 计算机工程与应用, 2020, 56(17): 60-68.
- [15] KUMAR A, DAUMÉ H. A co-training approach for multi-view spectral clustering[C]//Proceedings of the 28th international conference on machine learning. Bellevue: ICML,

- 2011;393–400.
- [16] ZHAO X, EVANS N, DUGELAY J L. A subspace co-training framework for multi-view clustering[J]. Pattern Recognition Letters, 2014, 41: 73–82.
 - [17] LUO D, DING C, HUANG H. Linear discriminant analysis: New formulations and overfit analysis [C]//Twenty-fifth AAAI conference on artificial intelligence. San Francisco: AAAI, 2011.
 - [18] CUI B, YU H, ZHANG T, et al. Self-weighted multi-view clustering with deep matrix factorization [C]//Proceedings of the eleventh Asian conference on machine learning. New York: ACML, 2019: 567–582.
 - [19] TAO Z, LIU H, LI S, et al. Marginalized multiview ensemble clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(2): 600–611.
 - [20] WANG B. Robust self-weighted multi-view projection clustering [C]//Proceedings of the thirty-fourth international joint conference on artificial intelligence. New York: AAAI, 2020: 6110–6117.
 - [21] ZHAN K, NIE F, WANG J, et al. Multiview consensus graph clustering [J]. IEEE Transactions on Image Processing, 2019, 28(3): 1261–1270.
 - [22] 黄奕轩, 杜世强, 余瑶, 等. 基于特征选择与鲁棒图学习的多视图聚类[J]. 计算机工程, 2022, 48(12): 95–103.
 - [23] DU L, ZHOU P, SHI L, et al. Robust multiple kernel K-means using L21-Norm [C]//Proceedings of the twenty-fourth international joint conference on artificial intelligence. Buenos Aires: IJCAI, 2015.
 - [24] GÖNEN M, MARGOLIN A A. Localized data fusion for kernel k-Means clustering with application to cancer biology [C]//Proceedings of the 27th international conference on neural information processing systems. Montreal: NIPS, 2014: 1305–1313.
 - [25] LI Miaomiao, LIU Xinwang, WANG Lei, et al. Multiple kernel clustering with local kernel alignment maximization [C]//Proceedings of the twenty-fifth international joint conference on artificial intelligence. New York: IJCAI, 2016: 1704–1710.
 - [26] LIU X, DOU Y, YIN J, et al. Multiple kernel k-Means clustering with matrix-induced regularization [C]//Proceedings of the thirtieth AAAI conference on artificial intelligence. [s. l.]: AAAI, 2016: 1888–1894.
 - [27] 夏冬雪, 杨燕, 王浩, 等. 基于邻域多核学习的后融合多视图聚类算法[J]. 计算机研究与发展, 2020, 57(8): 1627–1638.
 - [28] CHENG Jiafeng. Multi-view attribute graph convolution networks for clustering [C]//Proceedings of the twenty-ninth international joint conference on artificial intelligence. Yokohama: IJCAI, 2020: 2973–2979.
 - [29] HUANG Shudong. Auto-weighted multi-view clustering via deep matrix decomposition [J]. Pattern Recognition, 2020, 97(7): 107015.
 - [30] LI Jianqiang. Deep graph regularized non-negative matrix factorization for multi-view clustering [J]. Neurocomputing, 2020, 390: 108–116.
 - [31] TANG Wei. Clustering with multiple graphs [C]//2009 ninth IEEE international conference on data mining. Miami Beach: IEEE, 2009: 1016–1021.
 - [32] ABAVISANI M, PATEL V M. Deep multimodal subspace clustering networks [J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12: 1601–1614.
 - [33] ARAUJO A F R, ANTONINO V O, PONCE-GUEVARA K L. Self-organizing subspace clustering for high-dimensional and multi-view data [J]. Neural Networks, 2020, 130: 253–268.
 - [34] LI Zhaoyang. Deep adversarial multi-view clustering network [C]//Proceedings of the 28th international joint conference on artificial intelligence. Macao: IJCAI, 2019: 2952–2958.
 - [35] ZHOU Runwu, SHEN Yidong. End-to-end adversarial-attention network for multi-modal clustering [C]//2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle: IEEE, 2020: 14607–14616.
 - [36] WOO S. CBAM: convolutional block attention module [C]//European conference on computer vision (2018). Munich: Springer, 2018: 3–19.
 - [37] KAMPFFMEYER M, LØKSE S, BIANCHI F M, et al. Deep divergence-based approach to clustering [J]. Neural Networks, 2019, 113: 91–101.
 - [38] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations [C]//Proceedings of the 37th international conference on machine learning. Vienna: IJCAI, 2020: 1597–1607.
 - [39] KRIZHEVSKY A. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2012, 60: 84–90.
 - [40] EVERINGHAM M, VAN GOOLL, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303–338.
 - [41] ANDREW G. Deep canonical correlation analysis [C]//Proceedings of the 30th international conference on machine learning. Atlanta: ICML, 2013: 1247–1255.
 - [42] TROSTEN D J. Reconsidering representation alignment for multi-view clustering [C]//2021 IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE, 2021: 1255–1265.