计算机科学
COMPUTER SCIENCE

一种结构关系一致的对比聚类方法

许洁，王立松

引用本文

许洁，王立松. 一种结构关系一致的对比聚类方法[J]. 计算机科学，2023，50(9)：123-129.

XU Jie，WANG Lisong. Contrastive Clustering with Consistent Structural Relations[J]. Computer Science，2023，50(9)：123-129.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220700288

一种结构关系一致的对比聚类方法

许 洁 王立松
南京航空航天大学计算机科学与技术学院/人工智能学院/软件学院 南京 211106
(xujie85@nuaa.edu.cn)

摘 要 作为一项基本的无监督学习任务,聚类旨在将无标签的、混杂的图像数据划分成语义相似的类。最近的一些方法通过引入数据增强,利用对比学习方法学习特征表示和聚类分配,关注模型区分不同语义类的能力,可能导致来自同一语义类样本的特征嵌入被分离的情况。针对以上问题,提出一种结构关系一致的对比聚类方法(Contrastive Clustering with Consistent Structural Relations,CCR),在实例级和聚类级执行对比学习,并且增加关系级别的一致性约束,让模型学习更多来自结构关系的"正数据对"信息,从而减小聚类嵌入被分离所带来的影响。实验结果表明,CCR方法在图像基准数据集上得到了比近年来的无监督聚类方法更优异的结果。模型在CIFAR-10和STL-10数据集上的平均准确度比相同实验设置下的最好方法提升了1.7%,在CIFAR-100数据集上提升了1.9%。

关键词:无监督学习;聚类;对比学习;数据增强;过度聚类
中图法分类号 TP183

Contrastive Clustering with Consistent Structural Relations

XU Jie and WANG Lisong
College of Computer Science and Technology/College of Artificial Intelligence/College of Software,Nanjing University of Aeronautics and Astronautics,Nanjing 211106,China

Abstract As a basic unsupervised learning task,clustering aims to divide unlabeled and mixed images into semantically similar classes. Some recent approaches focus on the ability of the model to discriminate between different semantic classes by introducing data augmentation,using contrastive learning methods to learn feature representations and cluster assignments,which may lead to situations that feature embeddings from samples with the same semantic class are separated. Aiming at the above problems,a comparative clustering method with consistent structural relations(CCR) is proposed,which performs comparative learning at the instance level and cluster level,and adds consistency constraints at the relationship level. So that the model can learn more information of 'positive data pair' and reduce the impact of cluster embedding being separated. Experimental results show that CCR obtains better results than the unsupervised clustering methods in recent years on the image benchmark dataset. The average accuracy on the CIFAR-10 and STL-10 datasets improves by 1.7% compared to the best methods in the same experimental settings and improves by 1.9% on the CIFAR-100 dataset.

Keywords Unsupervised learning,Clustering,Contrastive learning,Data Augmentation,Over clustering

1 引言

近年来,在社交媒体平台、医学图像等领域产生了大量的视觉内容,其中大多数是没有标记的。手动标记这些数据非常耗时,超高的成本必然会给这些数据的共享和使用带来巨大的挑战,同时也导致人们对以无监督的方式有效地管理和使用如此大的数据量的需求增加。

聚类是一项基本的无监督学习方法。传统的聚类方法,如K-Means[1]、谱聚类[2]、非负矩阵分解聚类[3]等,只关注或过多地关注局部的、像素级的信息,忽略了图像更高层次的语义信息,因而性能有限。深度学习在近年来发展势头非常迅猛,越来越多的研究者将深度学习应用到聚类工作中[4-7]。IIC[4]使用图像及其随机增强后的图像组成数据对来训练模型学习聚类结果一致性;PICA[5]通过最大化分区置信度来学习语义上最可信的聚类解决方案;CC[6]创造性地提出"标签作为表示"的思想,显式地执行实例级和聚类级的对比学习;同样,DCDC[7]也注意到只从一个角度进行对比学习而忽略另一个角度,会导致性能较差,提出了特征级与聚类级相结合的方法。这类方法将图像样本数据看作实例,每个实例分别对应一个类,使用数据增强构建数据对,利用最大化互信息的

到稿日期:2022-07-29 返修日期:2022-12-05
基金项目:基础加强计划重点项目(2019JCJQZD33800)
This work was supported by the Key Projects of Foundation Strengthening Plan(2019JCJQZD33800).
通信作者:王立松(wangls@nuaa.edu.cn)

方式从中学习实例表示一致性和聚类表示一致性。实例表示一致性通过最大化实例表示特征与其增强之间的互信息来实现，有助于减少类内方差。聚类表示一致性通过最大化原始图像集群分配分布与其增强之间的互信息来实现，有助于增加类间方差，实现更具区分性的集群分配。尽管学习不同图像之间的区别有助于模型区分来自不同语义类的图片，但此类方法可能会导致同一类的实例被分离的情况，例如：同一类的实例被实例级损失函数认为是不同的图片而分离，导致聚类嵌入的类内方差较大，违背了“良好的聚类嵌入应该具有较小的类内方差和较大的类间方差”的初衷；同一类的图片被错误地分类，被聚类级损失函数认为是不同类别的图片而分离，这样的错误会给模型带来不稳定性，造成误差的积累。

针对以上问题，本文提出了一种结构关系一致的对比聚类方法 CCR。具体来说，就是在实例级别和聚类级别的对比损失之外，增加一种新的损失函数来惩罚多个样本数据的结构关系之间的差异，目的是让模型学习更多的“正数据对”信息，减轻实例分离带来的影响。这种损失关注的是多个输出数据之间的结构关系一致性而不是单个数据对本身，如图 1 所示。可将多个数据样本之间的距离视为它们之间的结构关系，由此可以分别得到原始样本和增强样本的结构关系。通过约束原始样本关系与增强样本关系之间的差异，提高模型对同一批样本输出相似的关系矩阵的能力，可以提高模型的鲁棒性。并且，将关系表示损失与双重对比损失结合，可以获得更多的正向鉴别特征和更小的聚类嵌入类内方差，从而得到更好的聚类结果。

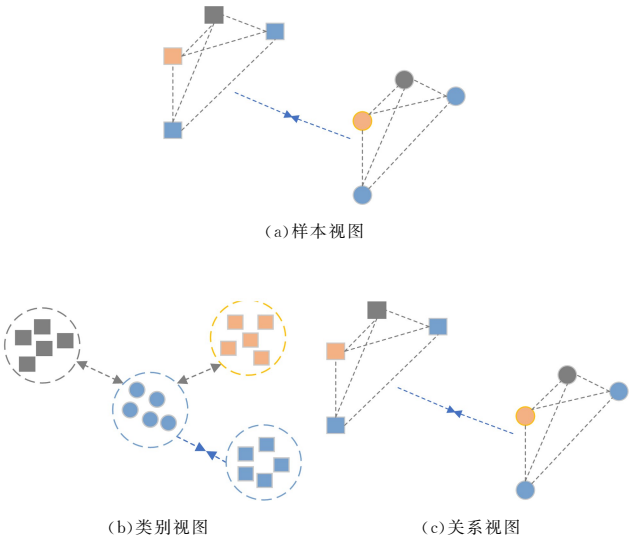

图 1 样本对、类别对及结构关系对之间的差别

Fig. 1 Differences between sample pairs, class pairs, and structural relationship pairs

2 相关工作

如何将不同的图片聚类到不同的簇中这一问题已经有了长期的研究和发展。本章将着重介绍两个方面的相关工作，即深度聚类和对比学习。

2.1 深度聚类

聚类的主要目的是将数据分成具有相似特征的数据点组。理论上，相似的样本会被划分到相同的类别，而不同的样本被划分到不同的类别。深度神经网络因为具有高度非线性转换的特性，被用于将数据转换为更适用于聚类的表示。深度嵌入聚类 DEC[8] 是第一个被广泛认可的深度聚类方法，Xie 等使用删除解码部分的自动编码器提取出的特征表示作为聚类模块的输入，并设计新的聚类损失对网络进行微调。Chang 等提出了一种基于单级网络的深度自适应聚类方法 DAC[9]，基于“成对图像之间的关系是二进制的”的假设，将图像聚类任务转化为一个判断图像对是否属于同一聚类的二分类问题。

另一方面，由于数据增强技术增强了模型的鲁棒性，研究人员也将其应用到无监督聚类工作中。DCCM[10] 全面探索了不同样本、几何变换的局部鲁棒性以及同一样本不同层特征之间的相互相关性，提出了特征间的三重互信息。针对大多研究中存在的只从单一视图角度学习聚类的问题，CC (Contrastive Clustering)[6] 使用双对比学习框架来学习实例级和聚类级相似性，同时学习鉴别特征并进行在线聚类。而 CRLC[11] 引入了一种新的批评家函数——“对数点积”来保证对比损失是最优的。在最新的研究中，SCAN[12] 与 SPICE[13] 取得了优异的结果，但此类多阶段图像聚类范式在实践中麻烦且不具备普遍适用性。因此，单阶段端到端的方法研究仍然是必要的。

2.2 对比学习

对比学习着重于学习同类样本之间的共同特征以区分非同类样本。与生成式学习相比，对比式学习可以忽略样本细节，只在抽象语义级别的特征空间上学习区分不同类别的样本，使模型具有更强的泛化能力。正对之间的相似度被最大化，负对之间的相似度被最小化，从而使正对相互靠近而负对相互远离。现有的方法如 DCDC[7] 和 DRC[14] 通常选择用样本的增强视图作为其正例，而将其他样本的增强视图作为负例。CC[6] 修改了正负数据对的选取规则，使用样本的不同增强视图组成正对，而将不同样本在同一增强及不同增强下的视图作为负对，从而得到更多的正负样本对信息，提高对比学习的性能。

在实现时，如何设计最优的对比损失是研究者需要解决的一个重要问题。很多方法使用噪声对比估计 (Noise Contrastive Estimation, NCE)[15] 作为对比损失，其核心思想是通过学习数据分布样本和噪声分布样本之间的区别来发现数据特性。NCE 将问题转换成二分类问题，区分数据样本和噪声样本，只适用于简单的分类问题。因此，更通用的对比损失 InfoNCE[16] 被推导出来，现有的对比学习方法大多使用 InfoNCE 作为损失函数。

近两年，对比学习在计算机视觉领域和自然语言处理领域都取得了很多成果，如 MoCo[17]，SimCLR[18]，ConSERT[19]，SimCSE[20] 等。在 CV 的一些任务上，基于对比学习思想的模型的表现甚至超过了有监督学习。本文同样使用对比学习来完成端到端的模型训练。

3 结构关系一致的对比聚类方法

对比学习的目标在于最大化正对之间的相似性而最小化

负对之间的相似性，其中一项非常重要的任务在于如何设计正负数据对来满足聚类任务的要求，即相似的样本相互靠近而不同的样本相互远离。

针对对比学习更加关注区分不同实例而忽略类内表现的问题，本文提出一种结构关系一致的对比聚类方法CCR，同时利用实例特征表示、类别表示和关系表示进行聚类，如图2所示。受到SimCLR[18]的启发，CCR使用数据增强来构建数据对作为输入。SimCLR[18]全面展示了不同的增强策略对下游任务性能的影响，本文选择随机裁剪、水平翻转、色彩抖动和灰度化这4种类型的数据增强方法。具体来说，给定一个原始数据，在数据增强方法的作用下，得到其对应的增强数据。神经网络作为深度聚类模型的骨干部分，主要作用是将输入图像数据经过层次化的非线性映射得到新的低维特征表示。为方便与其他已有的工作进行比较，本文采用ResNet34作为骨干网络。

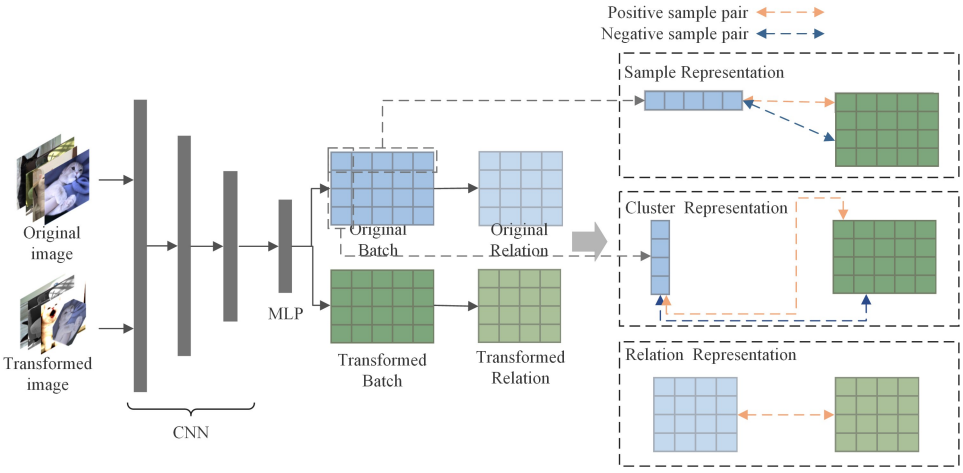

图2　关系结构一致的对比学习框架图

Fig. 2　Contrastive learning framework with consistent relational structure

神经网络输出图像的分配概率矩阵被视为特征矩阵。自然地，将分配概率矩阵的行（即每张图片的分配概率向量）看作是图像的特征表示，并且根据“标签作为表示”的思想[9]，矩阵的列作为数据分布情况被看作代表不同语义类的聚类表示。而原始图像和增强图像的结构关系矩阵可以通过相应的概率分配矩阵获得。

图2中虚线框部分展示了CCR所使用到的3种损失。在实例级别，最小化原始图像与增强图像之间的相似度，保证原始图像及其增强的特征表示一致性；在聚类级别，最小化原始类与增强类之间的相似度，保证原始图像及其增强的分配一致性（即原始类与增强类之间的聚类表示一致性）；在关系级别，最小化原始结构关系与增强结构关系之间的相似度，保证原始图像及其增强的关系表示一致性。3种损失共同训练，有助于形成良好的、更鲁棒的聚类。下面将详细介绍模型中涉及的几种损失函数。

3.1　实例表示损失

基于对比学习的思想，CCR方法将原始图像及其增强视为正对，而将原始图像与其他图像的增强视为负对。形式化来说，给定一批大小为 N 的原始样本 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_N\}$，其对应的 N 个增强样本为 $\boldsymbol{X}'=\{x_1',x_2',\cdots,x_N'\}$，那么对于任意样本 x_i，x_i 与 N 个增强样本总共可以组成 N 个数据对。本文将该样本与其对应的增强样本组成的数据对 (x_i,x_i') 视为正对，将该样本与其他 $N-1$ 个增强样本组成的数据对 (x_i,x_j') 视为负对。

为了减少对比学习带来的信息损失，本文没有直接使用神经网络 $f_\theta(\cdot)$ 输出的特征，而是使用非线性 $MLPg(\cdot)$ 将其映射到概率分配空间中，得到的概率分配被视为实例的特征表示 $u=g(f_\theta(x))$，$u'=g(f_\theta(x'))$。原始样本和增强样本本质是同一实例，应当具有相同的类分配概率。为方便起见，本文选择余弦相似度作为评价正样本对的分配概率是否保持一致性的指标，公式定义为：

$$\cos(u,u')=\frac{u^{\mathrm{T}}u'}{\|u\|_2\|u'\|_2} \tag{1}$$

其中，$\|\cdot\|_2$ 代表 L_2 归一化。

根据InfoNCE[16]，实例级别的损失可以定义为：

$$\mathscr{L}_{\mathrm{sam}}=-\mathbb{E}\left[\log\frac{\exp(\cos(u_i,u_i'))/\tau}{\sum_{j=1}^{N}\exp(\cos(u_i,u_j'))/\tau}\right] \tag{2}$$

其中，$\tau>0$ 是温度参数。

3.2　聚类表示损失

当将某一数据样本投影到维数等于聚类数 C 的空间时，其特征的第 j 个元素可以解释为该样本属于第 j 个类别的概率，特征向量相应地表示其软标签。形式上，与样本级别类似，设原始图像和增强图像分别对应输出概率分配矩阵 $\boldsymbol{V}=[v_1,v_2,\cdots,v_C]_{N\times C}$ 和 $\boldsymbol{V}'=[v_1',v_2',\cdots,v_C']_{N\times C}$。理想情况下的软标签往往是one-hot编码，那么 $\boldsymbol{V}$ 和 $\boldsymbol{V}'$ 的列空间即 v_j 与 v_j' 可以说明哪些图像被分配给聚类 j，也就是说 $\boldsymbol{V}$ 和 $\boldsymbol{V}'$ 的第 j 个列可以被看作是第 j 个聚类的表示。因此，被归为同一类的聚类可以被看作是正类对，例如，v_j' 实际上是 v_j 的增强，应当是同属一类的，可以将其视为正类对，其聚类表示应当是一致的。同样，这里使用余弦距离来衡量聚类表示对之间的相似性，即

$$\cos(v,v')=\frac{v^{\mathrm{T}}v'}{\|v\|_2\|v'\|_2} \tag{3}$$

相应地，对于温度参数 τ，聚类级别的损失就可以定义为：

$$\mathscr{L}_{\mathrm{clu}}=-\mathbb{E}\left[\log\frac{\exp(\cos(v_i,v_i'))/\tau}{\sum_{j=1}^{C}\exp(\cos(v_i,v_j'))/\tau}\right],\tau>0 \tag{4}$$

3.3 关系表示损失

本文所说的关系是指不同样本之间的结构关系。当高维的数据投影到不同低维空间中时，样本之间的结构关系应该保持一致。结构关系的表示方法可以有多种，如距离、角度等。为方便起见，本文使用空间中的欧氏距离作为两图像之间的关系表示：

$$\phi(u_i,u_j)=\frac{1}{\mu}\parallel u_i-u_j\parallel_2 \tag{5}$$

其中，μ 是距离的标准化因子。

为了关注其他样本对之间的相对距离，将 μ 设置为每个 batch 的数据对集合 B 中所有数据对之间的平均距离，采用式(6)计算 μ 的取值：

$$\mu=\frac{1}{|B|}\sum_{(u_i,u_j)\in B}\parallel u_i-u_j\parallel_2 \tag{6}$$

不同增强下的同一批图像，其数据点的距离结构关系应该是一致的。基于此，设计新的损失函数：

$$L_{\mathrm{rd}}=\sum_{(u_i,u_j)\in n}l_\sigma(\phi(u_i,u_j),\phi(u_i',u_j')) \tag{7}$$

其中，l_σ 为均方误差。

那么，综合以上 3 种损失的总损失函数可以写成：

$$\mathscr{L}_{\mathrm{total}}=\mathscr{L}_{\mathrm{sam}}+\mathscr{L}_{\mathrm{clu}}+\alpha\mathscr{L}_{\mathrm{rd}} \tag{8}$$

其中，α 为权重参数。

4 实验

4.1 数据集与评价指标

本文在 4 个被广泛使用的基准数据集上进行实验。对于 CIFAR-10，CIFAR-100 和 STL-10 数据集，实验时同时使用它们的训练集和测试集，对于 Tiny-ImageNet 数据集，实验中只使用训练集。下面将详细介绍这些数据集的特征。

1)CIFAR-10/100：CIFAR-10/CIFAR-100 数据集分别包含 10 个类和 20 个超类，由 60000 张 32×32×3 的图像组成，其中 50000 张用于训练，10000 张用于测试。

2)STL-10：包含 10 类物品，每类 1300 张，其中 500 张用于训练，800 张用于测试，每张图像大小为 96×96×3。除此之外，STL-10 还包含 100000 张无类别信息的图片样本，用于训练。

3)Tiny-ImageNet：它是 ImageNet 的一个子集，是一个具有挑战性的图像数据集。Tiny-ImageNet 包含 200 个类，每个类有 500 个训练样本、50 个验证样本、50 个测试样本，每个样本的大小为 64×64×3。

在实验评估阶段，本文使用了 3 种流行的聚类评价指标：准确性(ACC)、归一化互信息(NMI)和调整后的兰德指数(ARI)。这些指标的值越大，表明聚类性能越好。

4.2 实现细节

本文使用 PyTorch1.4 来完成所有的实验，并用 Adam 进行优化，设置学习速率为固定值 0.003。为方便与其他方法进行公平比较，使用了与大部分方法相同的神经网络 ResNet34 作为骨干网络进行训练。与 PICA[5] 等方法一致，本文使用了额外的过度聚类头来增加学习到的特征表示的表达性。对于过度聚类头，本文为 Tiny-ImageNet 设置了 700 个集群，为其他集群设置了 128 个集群。对于模型中涉及的超参数，本文将其设置为固定值，即温度参数=0.5，权重参数=0.004。需要注意的是，针对不同的数据集，本文选取了不同的批处理大小，即对于 CIFAR-10 数据集，批处理大小被设置为 50，CIFAR-100 数据集的批处理大小被设置为 200，STL-10 数据集的批处理大小为 50，Tiny-ImageNet 数据集的批处理大小为 350。在实验中，每个批次的样本被重复 3 次，并使用相同的数据增强方式。实验使用 Nvidia TITAN RTX 24G 将模型从头开始训练 200 个 epoch，与 DCDC[7] 等方法一致。其结果将在后面展示。

4.3 实验结果及分析

本文采用了 3 个不同的评价指标，在 4 个被广泛使用且具有挑战性的数据集上进行了实验，并与包括传统的聚类方法和深度聚类方法在内的 15 种具有代表性的聚类方法进行了比较，如表 1 所列。这些聚类方法包括 K-means[1]、谱聚类(SC)[2]、凝聚聚类(AC)[21]、基于非负矩阵分解的聚类(NMF)[3]、自动编码器(AE)[22]、去噪自动编码器(DAE)[23]、反卷积网络(DeCNN)[24]、变分自编码(VAE)[25]、联合无监督学习(JULE)[26]、深度嵌入聚类(DEsC)[27]、深度自适应图像聚类(DAC)[9]、不变信息聚类(IIC)[4]、深度综合相关挖掘(DCCM)[10]、分区置信度最大化(PICA)[5] 和双重对比学习(DCDC)[7]。

表 1 不同聚类方法在 4 个基准数据集上的聚类性能

Table 1 Clustering performance of different clustering methods on 4 baseline datasets

Methods	CIFAR-10			CIFAR-100			STL-10			Tiny-ImageNet		
	ACC	*NMI*	*ARI*	*ACC*	*NMI*	*ARI*	*ACC*	*NMI*	*ARI*	ACC	NMI	ARI
K-means	0.229	0.087	0.049	0.130	0.084	0.028	0.192	0.125	0.061	0.025	0.065	0.005
SC	0.247	0.103	0.085	0.136	0.090	0.022	0.159	0.098	0.048	0.022	0.063	0.004
AC	0.228	0.105	0.065	0.138	0.098	0.034	0.332	0.239	0.140	0.027	0.069	0.005
NMF	0.190	0.081	0.034	0.118	0.079	0.026	0.180	0.096	0.046	0.029	0.072	0.005
AE	0.314	0.239	0.169	0.165	0.100	0.048	0.303	0.250	0.161	0.041	0.131	0.007
DAE	0.297	0.251	0.163	0.151	0.111	0.046	0.302	0.224	0.152	0.039	0.127	0.007
DeCNN	0.282	0.240	0.174	0.133	0.092	0.038	0.299	0.227	0.162	0.035	0.111	0.006
VAE	0.291	0.245	0.167	0.152	0.108	0.040	0.282	0.200	0.146	0.036	0.113	0.006
JUNE	0.272	0.192	0.138	0.137	0.103	0.033	0.277	0.182	0.164	0.033	0.102	0.006
DEC	0.301	0.257	0.161	0.185	0.136	0.050	0.359	0.276	0.186	0.037	0.115	0.007
DAC	0.522	0.396	0.306	0.238	0.185	0.088	0.470	0.366	0.257	0.066	0.190	0.007
DCCM	0.623	0.496	0.408	0.327	0.285	0.173	0.482	0.376	0.262	0.108	0.224	0.017
IIC	0.617	—	—	0.257	—	—	0.610	—	—	—	—	—
PICA	0.696	0.591	0.512	0.337	0.310	0.171	0.713	0.611	0.531	0.098	0.277	0.038
DCDC	0.699	0.585	0.506	0.349	0.310	0.179	0.734	0.621	0.547	0.164	0.323	**0.073**
Ours(CCR)	**0.716**	**0.613**	**0.538**	**0.368**	**0.339**	**0.204**	**0.751**	**0.643**	**0.575**	**0.167**	**0.341**	0.072

从表 1 的结果来看，CCR 方法始终优于其他的先进方法，特别是本文的灵感来源于 DCDC[7]，这表明了本文方法的有效性。具体来说，以聚类平均准确度(ACC)为例，本文的方法在 CIFAR-10 和 STL-10 数据集上均提升了 1.7%，在 CIFAR-100 数据集上提升了 1.9%。在归一化互信息(NMI)方面，本文方法在 CIFAR-10 数据集上提升了 2.8%，在 STL-10 数据集上提高了 2.2%，在 CIFAR-100 数据集上提升了 2.9%。以上结果可以很好地证明本文方法在无监督聚类方面的有效性。

4.4　定性研究

4.4.1　集群分配可视化

为了便于理解无监督聚类的过程以及更好地说明含有关系表示的双重对比深度聚类方法在无监督聚类工作上的有效性，本文利用 t-SNE 编码对 CIFAR-10 数据集的测试集中的 10 000 张图像在训练过程中不同 epoch 下的特征表示分布进行了可视化，如图 3(a)－图 3(c)所示。可以看到，在训练未开始时，所有类别的图像混杂在一起，随着模型训练的进行，同一类别的图像彼此逐渐靠近，形成不同的簇，同时不同的簇也彼此逐渐远离。图 3(d)展示了在 150 epoch 时的分配概率，从图中可以看出，虽然仍有少部分图像被错误地分配到了其他类，但总体而言，不同类别的图像被有效地分离。

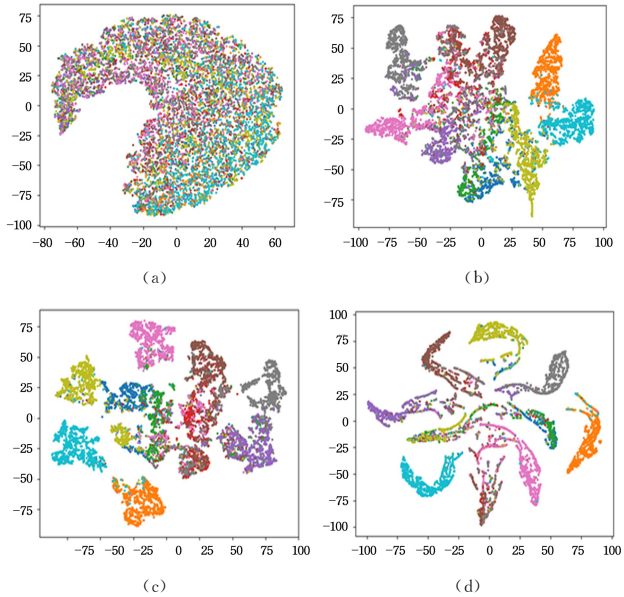

图 3　CIFAR-10 数据集中的图像在训练过程中的特征分布

Fig. 3　Feature distribution of images during training on CIFAR-10 dataset

4.4.2　成功与失败案例研究

为了更好地了解模型的性能，本文可视化了 CIFAR10 数据集中一些成功和失败的案例，如图 4 所示。左边线框中的图片代表模型成功预测其所属集群的案例，右边线框中的图片则代表对应集群中出现的失败案例。可以看出，模型能够将来自同一或不同语义类的具有相似形状或背景的图像聚集到同一集群中。实验中的错误案例往往来自于与正确案例相似的类，比如猫和狗、马和鹿等类别图像间的混淆。例如，第三行中，与鹿具有相似形状的马被错误地聚类到“鹿”集群中，而前两行中，“猫”集群当中的失败案例大多来自于“狗”集群，同样的情况也出现在“狗”集群中。

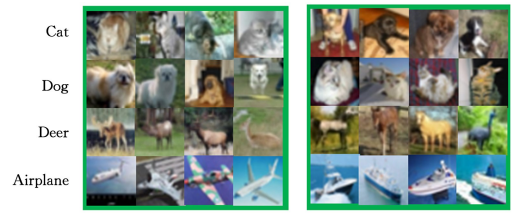

图 4　CIFAR-10 案例研究

Fig. 4　Cases studies on CIFAR-10

4.5　消融研究

为了进一步了解本文模型设置及实验设置中的不同选择对实验结果的影响，本文进行了如下消融研究。

4.5.1　3 种损失的影响

CCR 方法共设置了 3 种不同的损失函数。表 2 列出了这 3 种损失函数对模型的影响。其中 IRL 表示实例表示损失，CRL 表示聚类表示损失，RRL 表示关系表示损失。由于关系表示损失是实例表示损失与聚类表示损失的补充，因此不探讨单独使用 RRL 的情况。

表 2　关系表示损失的影响

Table 2　Effect on relationship representation loss

	CIFAR10			CIFAR100		
	ACC	*NMI*	*ARI*	*ACC*	*NMI*	*ARI*
IRL	0.606	—	—	0.307	—	—
CRC	0.458	—	—	0.209	—	—
IRL＋RRL	0.668	0.569	0.488	0.324	0.309	0.178
CRL＋RRL	0.613	0.520	0.423	0.317	0.308	0.163
IRL＋CRL(DCDC)	0.690	0.581	0.499	0.339	0.318	0.183
CCR	**0.716**	**0.613**	**0.538**	**0.368**	**0.339**	**0.204**

从表 2 可以看出，单独使用实例表示损失比单独使用聚类表示损失效果更好，这是由于实例级别的特征表示比聚类级别的特征表示携带了更多的对比性信息。而关系表示损失能够减小来自同一类别的样本嵌入被分离的影响，对于两种级别的对比学习都有一定的提升效果。同时使用聚类级别和实例级别的对比损失，比单独使用两者中的任一级别的效果更好，这是因为两种级别联合使用能够获得更多的鉴别信息。而最终的实验结果表明结构关系级别的对比学习同样能够在两项对比损失的基础上进一步优化模型，得到更优异的结果。

4.5.2　方差分析

关系表示损失的初衷在于为模型提供更多的正数据对信息，以达到减小类内聚类嵌入方差的目的。本文利用分配概率计算 CIFAR-10 数据集来自 10 个类别(分别编号为 1－10)中 60 000 个样本的聚类结果的类内方差。

结果如图 5 所示，可以看出，含有关系表示损失(RRL)的模型相比仅使用聚类表示损失和实例表示损失的模型(DCDC)获得的类内方差更小，聚类表现更优异，也说明了关系表示损失能够减轻同一类别的图像嵌入被分离的影响。

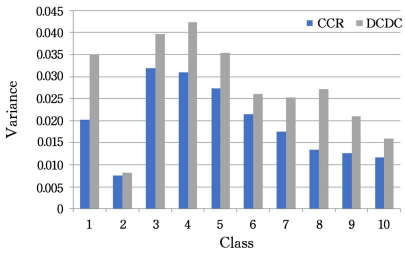

图 5 CIFAR-10 数据集的类内方差分析

Fig. 5 Intra-class variance analysis on CIFAR-10

4.5.3 批处理大小的影响

目前的对比学习方法认为，在训练时使用大批量的数据总是会获得更好的性能。为了评估批量大小(batch_size)对性能的影响，本文对其进行了不同的赋值，探究在 batch_size 取 100，200，300，500 时模型在 CIFAR100 数据集上的性能。实验结果如表 3 所列。

表 3 批处理大小对 CIFAR-100 数据集的影响

Table 3 Effect of batch_size on CIFAR100 dataset

batch_size	*ACC*	*NMI*	*ARI*
100	0.342	0.325	0.182
200	**0.368**	**0.339**	**0.204**
300	0.339	0.324	0.189
500	0.333	0.314	0.182

从结果中可以看出，对比学习受益于更大批量规模的认识可能并不适用于所有的深度聚类任务。对于类别数量多的数据集，如拥有 1000 个类别 ImageNet 数据集，应用对比学习时，大批量的数据包含更多类别的图像，给图像样本带来更多的负例表示，因此模型可以学习到样本更多更具鉴别性的特征，以此展现出更好的性能。而对于类别数量较小的数据集，仅仅简单地设置更大的批处理大小似乎并不奏效，如本文实验中展示的那样。因此，选取合适的批处理大小可以获得更好的模型性能。

4.5.4 过度聚类的影响

根据 IIC[4]，过度聚类头输出比真实的聚类数更多的聚类预测，能够获取更高维数的聚类和样本表示，增加特征表示的表达性。本文实验同样使用了过度聚类头来帮助模型学习，结果如表 4 所列，过度聚类(Over-clustering)对模型的性能有着较大的提升。

表 4 过度聚类头的影响

Table 4 Effect of over-clustering(OC)

	CIFAR10			CIFAR100		
	ACC	*NMI*	*ARI*	*ACC*	*NMI*	*ARI*
Without OC	0.654	0.553	0.471	0.269	0.256	0.133
CCR	**0.716**	**0.613**	**0.538**	**0.368**	**0.339**	**0.204**

结束语 为了弥补现有的基于对比学习的聚类方法的不足，本文提出了一种结构关系一致的对比聚类方法 CCR。与以往的方法不同，本文在实例表示一致性和聚类表示一致性的基础上增加了关系表示一致性的约束，认为同一批样本及其增强应当具有相似的结构关系表示。受益于这种新的约束，本文的方法在 4 个广泛使用的数据集上展现出了良好的性能。

但是，在通用数据集上进行的实验仅仅能够得出 3 种约束方式所对应的损失函数能够有效地提升模型的性能，而无法针对性地得出 3 种损失对应的适用场景及其原因。同样，本文所提出的方法仅仅减轻了聚类嵌入被分离所带来的影响，图 4 所示的错误聚类案例仍然没有得到根本的解决。这两个问题的解决方案超出了本文的范围，将在以后的工作中进一步完善。

参 考 文 献

[1] MACQUEEN J. Some methods for classification and analysis of multivariate observations[C]// Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability. 1967, 1(14): 281-297.

[2] ZELINIK-MANOR L, PERONA P. Self-Tuning Spectral Clustering[C]// Advances in Neural Information Processing Systems (NIPS). 2004.

[3] CAI D, HE X, WANG X, et al. Locality preserving nonnegative matrix factorization[C]// Twenty-first International Joint Conference on Artificial Intelligence. 2009.

[4] JI X, HENRIQUES J F, VEDALDI A. Invariant information clustering for unsupervised image classification and segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9865-9874.

[5] HUANG J, GONG S, ZHU X. Deep semantic clustering by partition confidence maximisation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8849-8858.

[6] LI Y, HU P, LIU Z, et al. Contrastive clustering [C]// 2021 AAAI Conference on Artificial Intelligence(AAAI). 2021.

[7] DANG Z, DENG C, YANG X, et al. Doubly contrastive deep clustering[J]. arXiv: 2103.05484, 2021.

[8] XIE J, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis[C]// International Conference on Machine Learning. PMLR, 2016: 478-487.

[9] CHANG J, WANG L, MENG G, et al. Deep adaptive image clustering[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 5879-5887.

[10] WU J, LONG K, WANG F, et al. Deep comprehensive correlation mining for image clustering[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8150-8159.

[11] DO K, TRAN T, VENKATESH S. Clustering by maximizing mutual information across views[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 9928-9938.

[12] VAN GANSBEKE W, VANDENHENDE S, GEORGOULIS S, et al. Scan: Learning to classify images without labels[C]// European Conference on Computer Vision. Cham: Springer, 2020: 268-285.

[13] NIU C, SHAN H, WANG G. Spice: Semantic pseudo-labeling for image clustering[J]. arXiv: 2103.09382, 2021.

[14] ZHONG H, CHEN C, JIN Z, et al. Deep robust clustering by contrastive learning[J]. arXiv: 2008.03030, 2020.

[15] GUTMANN M U，HYVARINEN A. Noise-Contrastive Estimation of Unnormalized Statistical Models，with Applications to Natural Image Statistics[J]. Journal of machine learning research，2012，13(2)：307-361.

[16] VAN DEN OORD A，LI Y，VINYALS O. Representation learning with contrastive predictive coding[J]. arXiv：1807. 03748，2018.

[17] HE K，FAN H，WU Y，et al. Momentum contrast for unsupervised visual representation learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：9729-9738.

[18] CHEN T，KORNBLITH S，NOROUZI M，et al. A simple framework for contrastive learning of visual representations[C]// International Conference on Machine Learning. PMLR，2020：1597-1607.

[19] YAN Y，LI R，WANG S，et al. Consert：A contrastive framework for self-supervised sentence representation transfer[J]. arXiv：2105. 11741，2021.

[20] GAO T，YAO X，CHEN D. Simcse：Simple contrastive learning of sentence embeddings[J]. arXiv：2104. 08821，2021.

[21] GOWDA K C，KRISHNA G. Agglomerative clustering using the concept of mutual nearest neighbourhood[J]. Pattern recognition，1978，10(2)：105-112.

[22] BENGIO Y，LAMBLIN P，POPOVICI D，et al. Greedy layer-wise training of deep networks[J]. Advances in Neural Information Processing Systems，2006，19：153-160.

[23] VINCENT P，LAROCHELLE H，LAJOIE I，et al. Stacked denoising autoencoders：Learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research，2010，11(12)：3371-3408.

[24] ZEILER M D，KRISHNAN D，TAYLOR G W，et al. Deconvolutional networks[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE，2010：2528-2535.

[25] KINGMA D P，WELLING M. Auto-encoding variational bayes[J]. arXiv：1312. 6114，2013.

[26] YANG J，PARIKH D，BATRA D. Joint unsupervised learning of deep representations and image clusters[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：5147-5156.

[27] XIE J，GIRSHICK R，FARHADI A. Unsupervised deep embedding for clustering analysis[C]// International Conference on Machine Learning. PMLR，2016：478-487.

XU Jie，born in 1998，master. Her main research interest is image clustering and retrieval.

WANG Lisong，born in 1969，Ph.D，professor，is a member of China Computer Federation. His main research interests include natural language processing and formal method.

（责任编辑：何杨）