

近邻一致性策略下的图像深度聚类算法研究

钱宇华^{1,2*},程占文¹,李飞江¹,王婕婷¹

(1.山西大学 大数据科学与产业研究院,山西 太原 030006;
2.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)

摘要:图像聚类是计算机视觉领域的一项重要任务,尽管已经提出许多方法来解决图像聚类问题,但目前基于表示的深度聚类方法主要依赖于样本的判别性表示能力。然而,除了判别性表示外,还应该考虑同类样本分配概率的一致性,使得学习到的样本表示空间同时具备判别性、稳定性和一致性。基于此,本文使用近邻一致性策略对图像深度聚类算法进行研究,该方法由两阶段组成,第一阶段为表示学习阶段,训练了一个卷积自编码器,构建初始特征空间;第二阶段为聚类阶段,使用样本的近邻一致性作为聚类约束,并引入样本稳定性增强近邻一致性,通过迭代优化聚类损失,在第一阶段的基础上对卷积编码器的网络参数进行微调,获得最终的聚类分布。该方法在聚类时主要考虑每一样本与其最近邻样本分配概率的一致性,充分挖掘同类样本间的相似关系,获得紧密的样本分布。实验结果表明,在五个图像数据集上,本文提出的方法优于典型的聚类算法。

关键词:深度聚类;图像;自编码器

中图分类号:TP391 **文献标志码:**A **文章编号:**0253-2395(2025)06-1161-10

Research on Image Deep Clustering Algorithm Based on Near Neighbor Consistency Strategy

QIAN Yuhua^{1,2*}, CHENG Zhanwen¹, LI Feijiang¹, WANG Jieting¹

(1. Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China;
2. Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University, Taiyuan 030006, China)

Abstract: Image clustering is an important task in the field of computer vision. Although many methods have been proposed to solve the image clustering task, current deep clustering methods based on representation mainly rely on the discriminative representation capability of samples. However, in addition to discriminative representation, the consistency of sample distribution probabilities with their nearest neighbor samples should also be considered so that the learned sample representation space possesses discriminative, stable, and consistent properties. Based on this, this paper conducts research on image deep clustering algorithms using the strategy of nearest neighbor consistency. The method consists of two stages: the representation learning stage, where a convolutional autoencoder is trained to construct the initial feature space, and the clustering stage, where nearest neighbor consistency is used as a constraint and sample stability is incorporated as an enhancement to fine-tune the network based on the first stage, obtaining the final clustering distribution. This method primarily considers the consistency of probability assignment between each sample and its nearest neighbors during clustering, fully exploring the similarity relationships among samples of the same class to achieve a compact sample distribution.

收稿日期:2024-02-04;**接受日期:**2024-03-27

基金项目:国家自然科学基金(62106132;62306170);山西省科技重大专项(202201020101006);山西省基础研究计划(20210302124271;202103021223026);山西省科技创新人才团队专项资助(202304051001001)

* **通信作者:**钱宇华(1976-),男,山西晋城人,博士,教授,主要研究方向为机器学习、复杂网络、粒计算和人工智能。
E-mail:jinchengqyh@126.com

引文格式:钱宇华,程占文,李飞江,等.近邻一致性策略下的图像深度聚类算法研究[J].山西大学学报(自然科学版),2025,48(6):1161-1170. DOI:10.13451/j.sxu.ns.2024055.

Experimental results demonstrate that the proposed method outperforms typical clustering algorithms on five image datasets.

Key words: deep clustering; image; autoencoder

0 引言

作为计算机视觉领域的一项重要任务,图像聚类^[1-2]已经受到越来越多的关注。聚类的目标是根据相似性度量方法(如欧氏距离、余弦相似度)将相似的样本分配到同一个类簇,而将不相似的样本分配到不同的类簇中,从而在没有任何监督信息的情况下实现对数据集的分组。由于聚类能够挖掘数据样本的全局特征,因此在异常检测^[3-4]、域自适应^[5-6]、社区检测^[7-8]和判别表示学习^[9-10]等领域中都得到了广泛的应用。然而,由于缺乏先验知识,仅根据样本之间的相似关系对数据进行聚类仍然是一项非常具有挑战性的任务。

目前已有许多优秀的聚类算法被提出。K-Means^[11]、谱聚类^[12]、高斯混合模型^[13]、层次聚类^[14]等传统聚类算法利用手工提取的特征,根据数据的内在相似性将数据划分为不同的类簇,但随着数据量的爆炸性增长以及数据维度的不断提高,手工提取特征越来越无法满足实际需求。近年来,深度神经网络在提取非结构化数据和高维数据的特征表示方面取得了巨大的成功,深度聚类算法利用深度神经网络将输入数据映射到低维嵌入空间,并利用嵌入空间中数据样本的嵌入表示进行聚类,在大数据集上取得了较好的聚类结果^[2, 15-16]。例如,基于关系匹配的深度聚类算法^[17-18]认为,当利用深度神经网络将样本映射到 D 维嵌入空间和 k 维标签空间时,样本之间的关系应该是一致的,通过约束嵌入空间和标签空间中样本间关系的一致性,可以弥合表示学习和聚类之间的语义差距,保证在聚类过程中尽量不丢失语义信息,从而得到好的聚类效果。其次,基于伪标签的深度聚类算法^[19-20]引入半监督学习的思想,首先过滤出具有高置信度的样本子集,然后以监督的方式约束高置信度样本的预测标签的一致性。此外,自训练深度聚类算法^[15-16]根据聚类分配分布构造辅助分布,并通过最小化分配分布和辅助分布之间的KL散度(Kullback-

Leibler Divergence, KLD)来优化聚类分配,达到聚类目的。

尽管上述方法已经取得了较好的聚类结果,但本文认为通过考虑样本与其最近邻样本分配概率的一致性关系,能够获得表示空间中同类样本更加紧密的样本分布,从而将不同类簇的边界划分的更加清晰。近邻一致性假设是机器学习领域的基本假设,它认为在特征空间中距离相近的样本,其标签具有一致性。分配概率一致性是常用的建模原理与思想,试图通过最大化样本分布之间的一致性来提高分类的准确性,比如对比学习方法通过构建正负样本对比较样本之间的相似性,对正样本对构建相似的分配概率,对负样本对构建不同的分配概率,学习具有判别性的特征表示;流形学习通过将高维数据映射到低维流形空间中,保留数据的局部结构和一致性,捕捉数据的内在特征。因此,本文提出一种同时增强同类样本之间的相似关系和不同类样本之间的差异关系的图像深度聚类方法,通过近邻样本分配概率的一致性来学习样本的判别表示。具体来说,本文提出的方法首先训练一个卷积自编码器,然后将其编码器部分作为图像特征提取器,将图像从输入空间映射到低维嵌入空间,最后使用样本与其最近邻样本之间分配概率的一致性来约束样本,并通过样本稳定性对其增强实现聚类。

本文的主要贡献可总结如下:

1) 本文提出一种使用近邻一致性策略约束样本分配概率相似性的图像深度聚类方法。

2) 本文主要使用近邻一致性约束样本的分布关系,并使用样本稳定性增强近邻一致性,在嵌入空间中学习样本的紧密分布。

3) 本文在5个数据集上对所提算法进行测试,实验结果表明,本算法得到了较好的结果。

1 相关工作

深度聚类^[2]旨在联合优化深度表示学习模块和聚类模块以实现聚类任务,其基本思想

是,良好的特征表示有助于产生良好的聚类结果,而良好的聚类结果可以反过来指导深度神经网络学习更好的特征表示,通过重复执行这两个过程,可以学习到有利于产生更好聚类性能的判别表示,进而实现卓越的聚类性能。

深度聚类中的表示学习模块将原始数据作为输入,并将输入数据的低维表示作为输出,在整个深度聚类模型中作为特征提取器,为聚类模块提取样本的特征表示。深度聚类中常用的表示学习模块如:基于自编码器的表示学习方法^[21]使用编码器和解码器两个神经网络模块线性堆叠成一个深度神经网络学习样本表示,基于生成模型的表示学习方法^[22]假设输入样本是由样本嵌入表示生成的,然后从输入数据中重构出其样本表示,基于互信息最大化的表示学习方法^[23]通过最大化神经网络不同层或不同部分的样本表示之间的互信息来学习样本表示,基于对比的表示学习方法^[24]利用将正样本对拉近而将负样本对推远的过程学习样本表示,子空间表示学习方法^[25]假设数据样本能够用相同子空间中的其他所有样本的线性组合表示,在将样本映射到低维子空间的过程中学习样本表示。本文采用基于自编码器的表示学习方法,设计了一个卷积自编码器作为表示学习模块。

深度聚类中的聚类模块将表示学习模块学到的低维表示作为输入,并输出用于硬聚类的硬标签或用于软聚类的聚类分配概率。目前,一种较为常用的方法是将样本输入前馈神经网络,将样本的维数降低到类簇个数 k ,并在最后一层应用softmax等方法建立分配分布,然后通过一些聚类约束使得该分配分布能够作为软标签,代表聚类分布。本文设计的聚类方法将自编码器中编码器最后一层的维度设置为类簇个数,在嵌入空间使用学生t分布建立聚类分配分布,并设计了一种样本近邻一致性策略作为聚类约束。

2 方法

2.1 整体框架

假设给定一个包含 N 个样本的数据集,定义为 $X = \{x_i\}_{i=1}^N, X \in R^d$,本文的目标是在不使用真实标签的情况下,利用卷积自编码器把数据 X 映射到聚类空间 $Y \in R^k$,将数据划分为 k 个类。

本文设计的聚类模型如图1所示。本模型由表示学习和聚类两个阶段组成。在表示学习阶段,本文设计了一个卷积自编码器,通过不断优化重构损失 L_r 来学习一组网络权重,使得卷积编码器学习到的特征表示能够准确地表示输入样本。在聚类阶段,丢弃解码器,只保留编码器,并对编码器学习到的每个样本的嵌入表示建立关于所有类别的分配分布,最后设计了一种近邻一致性策略,作为约束聚类的损失函数 L_c ,使得编码器学习以聚类为导向的样本表示,从而得到最终的聚类结果。

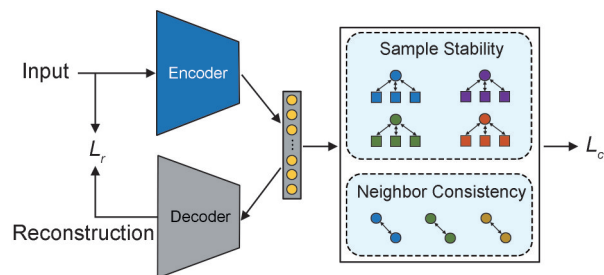


图1 模型整体框架图

Fig. 1 Overall framework of the model

2.2 表示学习模块

为了获得良好的特征表示,本文构建了一个卷积自编码器作为表示学习模块。具体来说,将图像数据 $X = \{x_i\}_{i=1}^N$ 输入到卷积自编码器中,卷积编码器 $z_i = f_{\theta_e}(x_i)$ 首先将输入样本 x_i 映射到嵌入空间 $Z \in R^k$,获得输入样本的嵌入表示 z_i ,然后,卷积解码器 $\hat{x}_i = g_{\theta_d}(z_i)$ 将嵌入空间 $Z \in R^k$ 中的数据重构到输入数据空间 $\hat{X} \in R^d$ 。通过迭代优化输入数据和重构数据之间的重构损失 L_r ,使样本的重构表示 \hat{x}_i 尽可能接近输入空间中的样本表示 x_i ,训练卷积自编码器 $g_{\theta_d}(f_{\theta_e}(x_i))$: $x_i \rightarrow z_i \rightarrow \hat{x}_i$ 。重构损失 L_r 可以写作:

$$L_r = \frac{1}{N} \sum_{i=1}^N \|g_{\theta_d}(f_{\theta_e}(x_i)) - x_i\|^2, \quad (1)$$

其中 N 表示样本数量, $\|\cdot\|^2$ 表示二范数。

2.3 聚类模块

在聚类阶段,本模型只保留卷积自编码器的编码器部分作为特征提取器,并使用聚类损失对模型进行约束,对编码器的网络参数进行微调,使其学习面向聚类任务的特征表示,建

立样本关于所有类簇的分配分布。

经过表示学习阶段的训练,可以获得所有样本在嵌入空间中的初始特征表示。本阶段,首先对嵌入空间中的初始样本表示执行一次k-means聚类,获取初始聚类质心 $m = \{m_1, m_2, \dots, m_k\} \in R^{k \times d}$,其中 k 和 d 分别表示质心向量的个数和维度。接下来,使用样本的邻近一致性约束网络执行聚类,具体过程如下:

首先,通过使用学生t分布计算每个样本的嵌入表示 z_i 关于所有质心 m_j 的相似性,建立每个样本关于所有质心的类簇分配概率,

$$q_{ij} = \frac{\left(1 + \|z_i - m_j\|^2 / \alpha\right)^{-\frac{\alpha+1}{2}}}{\sum_j \left(1 + \|z_i - m_j\|^2 / \alpha\right)^{-\frac{\alpha+1}{2}}}, \quad (2)$$

其中 $z_i = f_{\theta}(x_i) \in Z$ 表示样本 x_i 在嵌入空间中的嵌入表示, α 为学生t分布中的自由度, q_{ij} 可以表示第 i 个样本关于第 j 个类簇的分配概率。

获得所有样本的分配概率后,可以将其作为样本的软标签,揭示嵌入空间中所有样本的类簇分布情况。为了获得清晰的样本分布,让样本与其在输入空间中属于同一类的样本在聚类空间中更紧密地分配到同一类簇,同时与其他类中的样本清晰地分离开来,本文认为应该重点关注类内样本分布的一致性,让每一类簇内的所有样本具有相似的分配概率。如图2所示,假设输入空间中样本 x 属于类A,而在嵌入空间中,如果样本 x 关于类A的分配概率略小于其关于类B的分配概率,就会错误地将样本 x 划分到类B。因此,本文认为,如果能够约束样本 x 在嵌入空间中获得与类A中其他样本一致的分配概率,就能够将样本 x 正确地分配到类A。

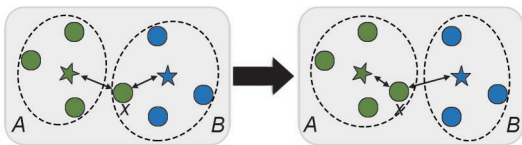


图2 类内样本分配概率

Fig. 2 Intra-class sample distribution probability

然而,在嵌入空间中,很难准确地将真实空间中所有属于同一类的样本确定为同一类,因此无法直接约束类内所有样本分配概率的一致性。不过,由于我们能够以较高的确定性保证

样本与其最近邻样本属于同一类簇,所以本文希望通过近邻一致性来约束每一样本与其最近邻样本的分配概率之间的一致性关系,防止误判嵌入空间中样本的分配概率。

如图3所示,在嵌入空间中,很容易找到每个样本的最近邻样本,如果约束每一样本与其最近邻样本获得一致的分配概率,从局部到整体约束样本的类簇分配概率,就很容易得到真实的样本分布情况。

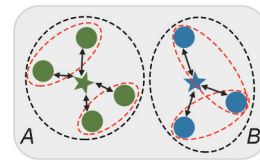


图3 近邻分配一致性

Fig. 3 Neighbor distribution consistency

给定样本 x_i ,令 $N(x_i)$ 表示样本 x_i 的最近邻样本,本文提出如下假设。

假设1 给定样本空间 $X \in R^d$ 中两个样本 x_i 和 x_j ,其在低维嵌入空间 $Z \in R^k$ 中的嵌入表示分别为 z_i 和 z_j ,如果在样本空间中 x_j 是 x_i 的最近邻,即 $x_j \in N(x_i)$,则在嵌入空间 $Z \in R^k$ 中, z_i 和 z_j 具有相似的分配概率。

根据假设1,可以通过约束嵌入空间中每一个样本与其最近邻样本分配概率的一致性,确保同一类的样本被划分到同一类簇,在嵌入空间中具有紧密的样本分布。具体来说,本文衡量每一个样本与其最近邻样本的分配分布的均方对数误差,通过迭代缩小样本与其最近邻样本分配分布的差距,提高样本与其最近邻样本软标签的一致性,进而使得每一类簇中的样本分布更加紧密。样本近邻一致性的损失函数设计如下:

$$L_{NC} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k \left(\log(q_{ij}) - \log(q_{N(x_i)j}) \right)^2, \quad (3)$$

其中 q_{ij} 表示第 i 个样本关于第 j 个类簇的分配概率, $q_{N(x_i)j}$ 表示第 i 个样本的最近邻关于第 j 个类簇的分配概率, N 表示样本个数, k 表示类簇个数。

此外,为了确保不同类样本的分布具有明显边界,使得近邻一致性的度量值更加准确,本文直接引入样本稳定性增强的深度嵌入聚类

(Deep Embedding Clustering Driven by Sample Stability, DECS)^[26]中的样本稳定性对近邻一致性进行增强。根据样本的分配概率,首先使用一个基于分配概率绝对值的确定性函数来度量每个样本 x_i 与所有质心 $\mathbf{m} = \{m_1, m_2, \dots, m_k\}$ 之间的确定性关系:

$$fq(q_{ij}) = \begin{cases} \left| \frac{q_{ij} - t}{t} \right|, & q_{ij} < t \\ \left| \frac{q_{ij} - t}{1 - t} \right|, & q_{ij} \geq t \end{cases}, \quad (4)$$

其中 t 表示样本关于质心的分配概率的最低值,本文使用大津法自适应地确定 t 的取值。其次,本文使用如下样本稳定性函数计算每个样本 x_i 的稳定性:

$$sq(x_i) = \frac{1}{k} \sum_{j=1}^k fq(q_{ij}) - \frac{\lambda}{k} \sum_{j=1}^k \left(fq(q_{ij}) - \frac{1}{k} \sum_{j=1}^k fq(q_{ij}) \right)^2, \quad (5)$$

其中 k 表示类簇个数, λ 是一个比例系数,公式第一项和第二项分别表示第 i 个样本关于所有类簇质心的确定性的均值和方差。基于此,样本的不稳定性便可作为损失进行迭代,即

$$L_{SS} = 1 - \sum_{i=1}^N sq(x_i), \quad (6)$$

其中 N 表示数据集中样本个数。

综上,联合近邻一致性和样本稳定性,本文设计的模型在聚类模块的总体目标函数如下:

$$L_C = L_{NC} + L_{SS}, \quad (7)$$

即,本文使用样本稳定性增强的近邻一致性约束编码器学习面向聚类的网络参数,进而获得聚类分配概率。

3 算法流程

算法 1 近邻一致性策略下的图像深度聚类

输入: 数据集 $X = \{x_i\}_{i=1}^N$, 类个数 k , 最大迭代次数 $MaxIter$

输出: 聚类质心 $\mathbf{m} = \{m_i\}_{i=1}^k$, 聚类分配 s

(1) 使用 X 通过迭代优化公式(1)初始化卷积自编码器;

(2) 使用k-means算法初始化 m 和 s ;

(3) **while** $iter \leq MaxIter$ **do**;

(4) 根据公式(2)计算样本关于每个类簇

的分配概率;

(5) 搜索每个样本的最近邻,获取最近邻的分配概率;

(6) 根据公式(3)计算样本的近邻一致性损失;

(7) 根据公式(6)计算样本稳定性损失;

(8) 迭代更新公式(7),更新编码器的参数和质心 \mathbf{m} ;

(9) **end while**;

(10) 返回:质心 $\mathbf{m} = \{m_i\}_{i=1}^k$, 聚类分配 s 。

4 实验

4.1 数据集

为了验证本文提出的聚类算法的性能,本文在5个常用的公开图像数据集上进行实验。由于聚类任务是完全无监督的,本文将所有数据集的训练集和测试集合并后作为训练样本一起训练模型。其中,MNIST数据集是手写数字识别数据集,由60000个训练样本和10000个测试样本组成,均为 $1 \times 28 \times 28$ 像素的灰度图像;MNIST-test数据集是MNIST数据集的测试集部分;USPS数据集由美国邮政服务信封上的手写数字图像组成,包括7291个训练样本和2007个测试样本,均为 $1 \times 16 \times 16$ 像素的灰度图像;Fashion数据集是一个包含各种不同类型的衣服、鞋子等的图像数据集,由60000个训练样本和10000个测试样本组成,均为 $1 \times 28 \times 28$ 像素的灰度图像;YTF数据集是来自YouTube的真实人脸数据集,本文使用其前41个类,共包含12183个样本,均为 $3 \times 55 \times 55$ 像素的彩色图像。以上数据集的详细介绍如表1所示。

表 1 数据集信息

Table 1 Information of datasets

Dateset	Samples	Classes	Dimensions
MNIST	70 000	10	$1 \times 28 \times 28$
MNIST-test	10 000	10	$1 \times 28 \times 28$
USPS	9 298	10	$1 \times 16 \times 16$
Fashion	70 000	10	$1 \times 28 \times 28$
YTF	12 183	41	$3 \times 55 \times 55$

4.2 评价指标

本文使用两个常用的无监督评价指标来验证所提模型在聚类中的表现,分别为聚类准确

度(ACC)和标准化互信息(NMI),这两个指标的值越高,表明模型的性能越优。

准确度(ACC)是一个用于度量预测标签和真实标签的相似性的常用指标,定义为:

$$ACC = \frac{\sum_{i=1}^N I(y_i, m(c_i))}{N}, \quad (8)$$

其中 $I(x, y) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise} \end{cases}$ 是一个指示函数, y_i

表示真实标签, c_i 表示模型预测的类簇分配概率, m 将软分配概率映射到硬标签, N 表示样本总量。

标准化互信息(NMI)是另一个常用的度量预测标签和真实标签相似性的指标,定义为:

$$NMI = \frac{2I(y, \hat{y})}{H(y) + H(\hat{y})}, \quad (9)$$

其中 y 和 \hat{y} 分别表示真实标签集和聚类结果集, $I(y, \hat{y})$ 表示 y 和 \hat{y} 的互信息, $H(y)$ 和 $H(\hat{y})$ 分别表示 y 和 \hat{y} 的熵。

4.3 实验细节

本文搭建的卷积自编码器中的编码器部分由4个卷积层构成,为了保持与对比算法参数的一致性,其通道大小分别设置为32、64、128和256,卷积核的大小为 3×3 ,步长为2,使用ReLU函数作为激活函数,并在每一层卷积层后面加一个归一化层和最大池化层。卷积自编码器的解码器部分采用与编码器对称的网络结构。

在训练时,为了提高模型的泛化性能,本文对输入的图像数据进行了数据增强操作,如随机旋转、平移和裁剪等。具体来说,给定一组数据样本 $X = \{x_i \in R^d\}_{i=1}^N$, 本文首先对每个样本 x_i 进行随机增强操作 T_{random} , 得到随机增强后的样本 $x' = T_{\text{random}}(x_i)$, 然后将 x' 作为输入空间的样本输入到卷积自编码器进行训练。

训练过程中,本文首先对卷积自编码器进行端到端训练500个epoch,使用Keras默认参数的Adam作为优化器。随后,使用聚类损失对卷积编码器进行微调,使用batch size为256的batch迭代20000次。此外,通过超参数敏感性实验可以发现本文所设计的模型对超参数不敏感,参数 α 和 λ 分别设置为3.0和0.8时,实验结果相对较优,因此本文在后续实验中将参数 α 和 λ 分别设置为3.0和0.8,同时,为了保持与对比算法学习率的一致,

本文将学习率设置为0.001。

4.4 对比方法

在对比实验中,本文首先与几个典型的传统聚类算法进行比较,包括k-means算法^[11]、高斯混合模型(Gaussian Mixture Model, GMM)^[13]、归一化割的谱聚类(Normalized-cut Spectral Clustering, SC-Ncut)^[27]和大规模谱聚类^[28]。其次,与几个较新的深度聚类算法进行比较,包括深度嵌入聚类(Deep Embedded Clustering, DEC)^[15]、改进的深度嵌入聚类(IDEC)^[16]、变分深度嵌入聚类(Variational Deep Embedding, VaDE)^[29]、联合深度表示学习和图像聚类的无监督学习(Joint Unsupervised Learning of Deep Representations and Image Clusters, JULE)^[30]、深度嵌入正则化聚类(Deep Embedded Regularized Clustering, DEPICT)^[31]、使用可靠样本改进的深度卷积嵌入聚类(Improved Deep Convolutional Embedded Clustering using Reliable Samples, IDCEC)^[32]、可解释的神经网络聚类(Terpretable Neural Clustering, TELL)^[33]、通过近似交替方向法正则化互信息的聚类(L_2 -regularized Mutual-information Clustering Optimized with an Approximate Alternating Direction Method, MI-ADM)^[34]、使用双通道自编码器的深度谱聚类(Deep Spectral Clustering using Dual Autoencoder Network, DSCDA)^[35]、使用动态自编码器的深度聚类(Deep Clustering with a Dynamic Auto-Encoder, DynAE)^[36]、基于Transformer和分布信息的深度嵌入图像聚类(Transformer-based Deep Embedded Image Clustering with Distribution Information, TDEC)^[37]、自适应确定类个数的深度非参数聚类(Deep Non-Parametric Clustering with Adaptive Cluster-number Estimation, DeepDPM)^[38]。

4.5 实验结果

为了验证所设计的聚类方法的性能,本文在5个数据集上进行实验,并将聚类结果与上述传统聚类算法和深度聚类算法进行对比,表2和表3分别展示了不同聚类方法在准确度和标准化互信息两个指标上的性能表现。

对于实验中使用的对比方法,本文使用原论文中公开的实验数据作为对比,如果该方法在某个数据集上没有进行实验,本文使用原论文中使用的参数运行代码,并在表格中将结果

用(*)标记,如果代码没有公布或者无法在该数据集上运行,在表格中用(-)标记,每个数据集上表现最好的结果用加粗字体显示。

表2 不同聚类方法的准确度

Table 2 Accuracy of different clustering methods

methods	MNIST	MNIST-Test	USPS	Fashion	YTF
k-means	0.532	0.542	0.668	0.474	0.601
GMM	0.433	0.540	0.551	0.556	0.348
SC-Ncut	0.656	0.660	0.649	0.508	0.510
SC-LS	0.714	0.740	0.746	0.496	0.544
DEC	0.863	0.856	0.762	0.518	0.371
IDEC	0.881	0.846*	0.761	0.529*	0.400*
VaDE	0.944	0.944*	0.566*	0.629	0.601*
JULE	0.964	0.961	0.950	0.563*	0.684
DEPICT	0.965*	0.963*	0.899	0.392	0.621
IDCEC	0.948	0.923	0.812	—	0.632
TELL	0.952	0.776*	0.865*	0.584*	—
MI-ADM	0.969	0.871	0.979	—	0.606
DSCDA	0.978	0.980	0.869	0.662	0.691
DynAE	0.987	0.987	0.981	0.591	—
DeepDPM	0.980	—	0.940	0.610	0.821
TDEC	0.985	0.975	0.976	0.645	0.950
Ours	0.991	0.990	0.993	0.646	0.883

注:*表示本文使用原文献中使用的参数运行代码,一表示如果代码没有公布或者无法在该数据集上运行,粗体表示数据集上表现最好的结果。

表3 不同聚类方法的标准化互信息

Table 3 NMI of different clustering methods

methods	MNIST	MNIST-Test	USPS	Fashion	YTF
k-means	0.499	0.500	0.626	0.512	0.776
GMM	0.366	0.593	0.530	0.557	0.411
SC-Ncut	0.731	0.704	0.794	0.575	0.701
SC-LS	0.706	0.756	0.755	0.497	0.759
DEC	0.834	0.830	0.767	0.546	0.446
IDEC	0.867	0.802*	0.785	0.557*	0.483*
VaDE	0.891	0.885*	0.512*	0.611	0.753*
JULE	0.913	0.915	0.913	0.608*	0.848
DEPICT	0.917*	0.915*	0.906	0.392	0.802
IDCEC	0.906	0.853	0.858	—	0.793
TELL	0.888	0.751*	0.786*	0.658*	—
MI-ADM	0.922	0.885	0.948	—	0.801
DSCDA	0.941	0.946	0.857	0.645	0.857
DynAE	0.964	0.963	0.948	0.642	—
DeepDPM	0.950	—	0.900	0.500	0.930
TDEC	0.957	0.935	0.935	0.693	0.980
Ours	0.974	0.973	0.978	0.707	0.913

注:*表示本文使用原文献中使用的参数运行代码,一表示如果代码没有公布或者无法在该数据集上运行,粗体表示数据集上表现最好的结果。

从表2和表3可以看出,本文设计的方法在准确度和标准化互信息两个指标上都表现出了较好的性能,在对比的5个数据集中的4个数据集上都取得了最好的聚类结果。尤其在USPS数据集上,本文提出的方法在准确度和标准化互信息两个指标上分别比性能第二好的DynAE高1.2%和3%,并且在MNIST和USPS两个数据集上的聚类准确度均超过了99%。此外,得益于深度神经网络良好的特征提取能力,本文的方法明显超过了4个传统聚类算法。

为了证明本文设计的算法在聚类任务上的有效性,本文在USPS数据集上使用t分布-随机邻近嵌入(t-Distributed Stochastic Neighbor Embedding, t-SNE)算法对聚类过程进行了可视化分析,结果如图4所示。其中,图4(a)—图4(c)表示学习阶段样本分布的变化过程,图4(d)—图4(f)表示聚类阶段样本分布的变化过程。从图中可以观察到,本文设计的聚类算法在迭代过程中能够学习到紧密分布的样本判别表示,能够较好地将不同类的样本分离开来,而将同类的样本聚集到一起。

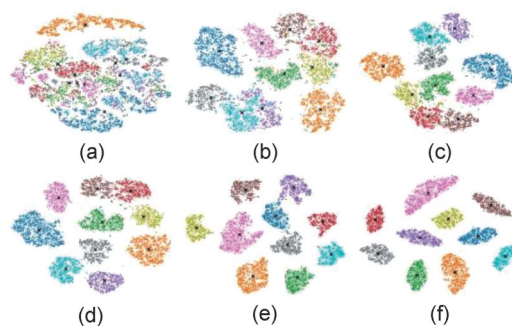


图4 聚类过程的t-SNE可视化

(a)—(c)为表示学习阶段样本分布的变化过程,(d)—(f)为聚类阶段样本分布的变化过程。

Fig. 4 t-SNE visualization in clustering processes

(a)-(c) is the process of sample distribution in the representation learning stage, and (d)-(f) is the process of sample distribution in the clustering stage.

4.6 超参数分析

为了验证本文提出的方法对不同超参数的敏感性,本文在USPS和MNIST两个数据集上对聚类损失 L_c 中的 α 和 λ 两个超参数的不同取值对模型性能的影响进行了分析,如下所述。

图5描述了在USPS数据集上参数 α 的不同取值对模型性能的影响程度,图5(a)和图5(b)分别展示了参数 α 取不同的值时,对应的准确度和标准化互信息的结果。从图中可以看出,当参数 α 在1.0到4.0之间变化时,模型在USPS数据集上的准确度介于0.991到0.993之间,变化幅度为0.002,标准化互信息介于0.973和0.978之间,变化幅度为0.005。

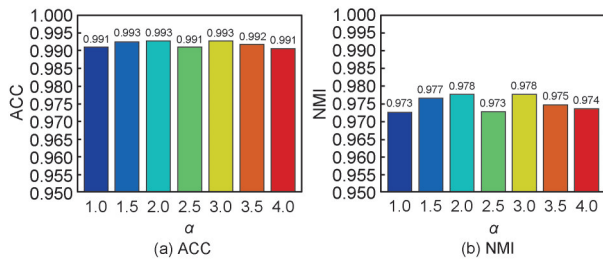


图5 USPS数据集上对参数 α 的分析

(a)和(b)分别表示参数 α 取不同值时的ACC和NMI。

Fig. 5 Analysis of parameter α on USPS data set

(a) and (b) represent the ACC and NMI for different values of the parameter α , respectively.

图6描述了在USPS数据集上参数 λ 的不同取值对模型性能的影响程度,图6(a)和图6(b)分别展示了参数 λ 取不同的值时,对应的准确度和标准化互信息的结果。从图中可以看出,当参数 λ 在0.2到1.0之间变化时,模型在USPS数据集上的准确度介于0.992到0.993之间,变化幅度为0.001,标准化互信息介于0.976和0.979之间,变化幅度为0.003。

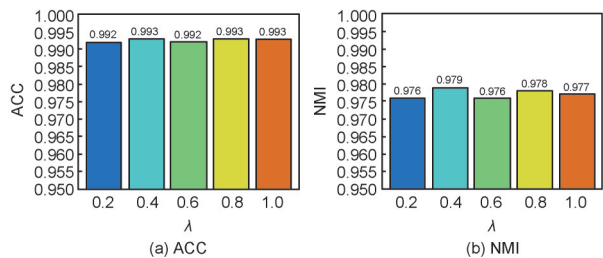


图6 USPS数据集上对参数 λ 的分析

(a)和(b)分别表示参数 λ 取不同值时的ACC和NMI。

Fig. 6 Analysis of parameter λ on USPS data set

(a) and (b) represent the ACC and NMI for different values of the parameter λ , respectively.

图7描述了在MNIST数据集上参数 α 的不同取值对模型性能的影响程度,图7(a)和图7(b)分别展示了参数 α 取不同的值时,对应的准

确度和标准化互信息的结果。从图中可以看出,当参数 α 在1.0到4.0之间变化时,模型在MNIST数据集上的准确度介于0.987到0.991之间,变化幅度为0.004,标准化互信息介于0.961和0.974之间,变化幅度为0.013。

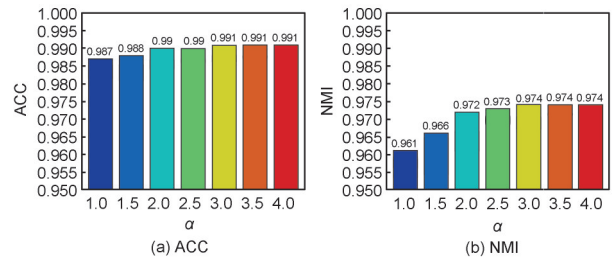


图7 MNIST数据集上 α 对参数的分析

(a)和(b)分别表示参数 α 取不同值时的ACC和NMI。

Fig. 7 Analysis of parameter α on MNIST data set

(a) and (b) represent the ACC and NMI for different values of the parameter α , respectively.

图8描述了在MNIST数据集上参数 λ 的不同取值对模型性能的影响程度,图8(a)和图8(b)分别展示了参数 λ 取不同的值时,对应的准确度和标准化互信息的结果。从图中可以看出,当参数 λ 在0.2到1.0之间变化时,模型在USPS数据集上的准确度介于0.990到0.991之间,变化幅度为0.001,标准化互信息介于0.972和0.974之间,变化幅度为0.002。

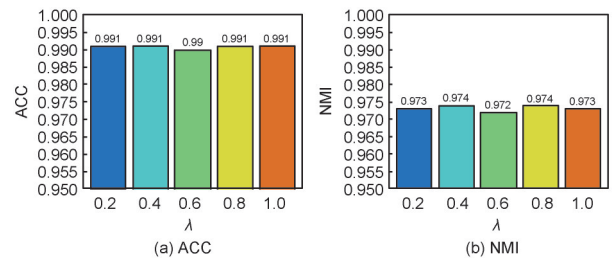


图8 MNIST数据集上对参数 λ 的分析

(a)和(b)分别表示参数 λ 取不同值时的ACC和NMI。

Fig. 8 Analysis of parameter λ on MNIST data set

(a) and (b) represent the ACC and NMI for different values of the parameter λ , respectively.

综上所述,本文设计的深度聚类算法对超参数的设置不敏感,超参数设置不同的值时,算法在聚类准确度和标准化互信息两个指标上的变化非常小,即本文设计的模型基本不受超参数变化的影响。

5 结论

为了增强同类样本之间的相似性和不同类样本之间的差异性,使同一类样本聚得更加紧密,而不同类样本相互分离,本文在近邻一致性策略下对图像深度聚类算法进行研究。该算法由一个卷积自编码器和一个聚类层构成,在卷积编码器构造的特征空间中,本文使用样本与其最近邻样本之间分配概率的一致性约束特征空间中的样本表示,并引入样本稳定性方法对近邻一致性进行增强,使嵌入空间中的样本分布不断接近输入空间中的样本分布,进而获得所有样本的类簇分布。实验结果表明,本文提出的算法性能优于所对比的传统聚类算法和深度聚类算法,证明了该算法的有效性。

未来,本算法将使用表示学习能力更强的表示学习方法作为特征提取器,提取更加有助于聚类的样本表示来提高模型的聚类性能。同时,在有监督或半监督场景下,可以设计双通道模型,同时利用监督信息和样本之间的一致性关系来约束模型以取得更好的性能。此外,从理论上分析近邻一致性对聚类泛化性能的影响将是有趣的研究问题。

参考文献:

- [1] XU R, WUNSCH D. Survey of Clustering Algorithms [J]. *IEEE Trans Neural Netw*, 2005, **16**(3): 645-678. DOI: 10.1109/TNN.2005.845141.
- [2] SARKER I H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions[J]. *SN Comput Sci*, 2021, **2**(6): 420. DOI: 10.1007/s42979-021-00815-1.
- [3] AYTEKIN C, NI X Y, CRICRI F, et al. Clustering and Unsupervised Anomaly Detection with L2 Normalized Deep Auto-encoder Representations[C]//2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018: 1-6. DOI: 10.1109/IJCNN.2018.8489068.
- [4] SOHN K, YOON J, LI C L, et al. Anomaly Clustering: Grouping Images into Coherent Clusters of Anomaly Types [C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). New York: IEEE, 2023: 5468-5479. DOI: 10.1109/WACV56688.2023.00544.
- [5] ZHOU Q, ZHOU W A, WANG S R. Cluster Adaptation Networks for Unsupervised Domain Adaptation[J]. *Image Vis Comput*, 2021, **108**: 104137. DOI: 10.1016/j.imavis.2021.104137.
- [6] LI J C, LI G B, SHI Y M, et al. Cross-domain Adaptive Clustering for Semi-supervised Domain Adaptation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2021: 2505-2514. DOI: 10.1109/CVPR46437.2021.00253.
- [7] MALLIAROS F D, VAZIRGIANNIS M. Clustering and Community Detection in Directed Networks: A Survey [J]. *Phys Rep*, 2013, **533**(4): 95-142. DOI: 10.1016/j.physrep.2013.08.002.
- [8] SU X, XUE S, LIU F Z, et al. A Comprehensive Survey on Community Detection with Deep Learning[J]. *IEEE Trans Neural Netw Learn Syst*, 2022, PP(99): 1-21. DOI: 10.1109/TNNLS.2021.3137396.
- [9] JIA X D, JING X Y, ZHU X K, et al. Semi-supervised Multi-view Deep Discriminant Representation Learning [J]. *IEEE Trans Pattern Anal Mach Intell*, 2021, **43**(7): 2496-2509. DOI: 10.1109/TPAMI.2020.2973634.
- [10] REZAEI M, DORIGATTI E, RUEGAMER D, et al. Joint Debaised Representation Learning and Imbalanced Data Clustering[C]//2022 IEEE International Conference on Data Mining Workshops. Orlando: IEEE, 2022: 55-62. DOI:10.1109/ICDMW58036.2022.00017.
- [11] MACQUEEN J. Some Methods for Classification and Analysis of Multivariate Observations[J]. *Berkeley Symp Math Stat Probab*, 1967: 281-297.
- [12] NG A, JORDAN M, WEISS Y. On Spectral Clustering: Analysis and An Algorithm[J]. *Adv Neural Inf Process Syst*, 2001, 14. DOI: 10.5555/2980539.2980649.
- [13] BISHOP C. Pattern Recognition and Machine Learning [J]. *Springer Google Schola*, 2006, **2**: 531-537. DOI: 10.1198/jasa.2008.s236.
- [14] SNEATH P H A, SOKAL R R. Numerical Taxonomy[J]. *Nature*, 1962, **193**: 855-860. DOI: 10.1038/193855a0.
- [15] XIE J Y, GIRSHICK R, FARHADI A. Unsupervised Deep Embedding for Clustering Analysis[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48. New York: ACM, 2016: 478-487. DOI: 10.5555/3045390.3045442.
- [16] GUO X F, GAO L, LIU X W, et al. Improved Deep Embedded Clustering with Local Structure Preservation [C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York: ACM, 2017: 1753-1759. DOI: 10.5555/3172077.3172131.
- [17] VAN GANSBEKE W, VANDENHENDE S, GEORGOULIS S, et al. Learning to Classify Images without Labels[C]//Proc. of the 16th European Conference on Computer Vision. Cham: Springer, 2020: 268-285. DOI: 10.1007/978-3-030-58580-8_16.

- [18] LONG M S, CAO Y, WANG J M, *et al.* Learning Transferable Features with Deep Adaptation Networks[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37. New York: ACM, 2015: 97–105. DOI: 10.5555/3045118.3045130.
- [19] CARON M, BOJANOWSKI P, JOULIN A, *et al.* Deep Clustering for Unsupervised Learning of Visual Features [C]//Computer Vision-ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIV. New York: ACM, 2018: 139–156. DOI: 10.1007/978-3-030-01264-9_9.
- [20] NIU C, SHAN H M, WANG G. SPICE: Semantic Pseudo-labeling for Image Clustering[J]. *IEEE Trans Image Process*, 2022, **31**: 7264–7278. DOI: 10.1109/TIP.2022.3221290.
- [21] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Internal Representations by Error Propagation [M]//Readings in Cognitive Science. Amsterdam: Elsevier, 1988: 399–421. DOI: 10.1016/b978-1-4832-1446-7.50035-2.
- [22] KINGMA D P, WELING M. Auto-encoding Variational Bayes[EB/OL]. (2013-12-21)[2024-02-01]. <https://arxiv.org/abs/1312.6114>.
- [23] KINNEY J B, ATWAL G S. Equitability, Mutual Information, and the Maximal Information Coefficient[J]. *Proc Natl Acad Sci USA*, 2014, **111**(9): 3354–3359. DOI: 10.1073/pnas.1309933111.
- [24] VAN DEN OORD A, LI Y Z, VINYALS O. Representation Learning with Contrastive Predictive Coding[EB/OL]. (2018-07-11)[2024-02-01]. <https://arxiv.org/abs/1807.03748>.
- [25] ELHAMIFAR E, VIDAL R. Sparse Subspace Clustering: Algorithm, Theory, and Applications[J]. *IEEE Trans Pattern Anal Mach Intell*, 2013, **35**(11): 2765–2781. DOI: 10.1109/TPAMI.2013.57.
- [26] CHENG Z W, LI F J, WANG J T, *et al.* Deep Embedding Clustering Driven by Sample Stability[C]//Proc. of the 33rd International Joint Conference on Artificial Intelligence. Jeju Island: IJCAI. 2024: 426–434. DOI: 10.24963/ijcai.2024/426.
- [27] SHI J B, MALIK J. Normalized Cuts and Image Segmentation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2000, **22**(8): 888–905. DOI: 10.1109/34.868688.
- [28] CHEN X L, CAI D. Large Scale Spectral Clustering with Landmark-based Representation[C]//Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. New York: ACM, 2011: 313–318. DOI: 10.5555/2900423.2900.
- [29] JIANG Z X, ZHENG Y, TAN H C, *et al.* Variational Deep Embedding: an Unsupervised and Generative Approach to Clustering[C]//Proc. of the 26th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI. 2017: 1965–1972.
- [30] YANG J W, PARIKH D, BATRA D. Joint Unsupervised Learning of Deep Representations and Image Clusters [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 5147–5156. DOI: 10.1109/CVPR.2016.556.
- [31] DIZAJI K G, HERANDI A, DENG C, *et al.* Deep Clustering via Joint Convolutional Autoencoder Embedding and Relative Entropy Minimization[C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2017: 5747–5756. DOI: 10.1109/ICCV.2017.612.
- [32] LU H, CHEN C, WEI H, *et al.* Improved Deep Convolutional Embedded Clustering with re-selectable Sample Training[J]. *Pattern Recognit*, 2022, **127**: 108611. DOI: 10.1016/j.patcog.2022.108611.
- [33] PENG X, LI Y F, TSANG I, *et al.* XAI beyond Classification: Interpretable Neural Clustering[J]. *J Mach Learn Res*, 2018, **23**: 6: 1–6: 28.
- [34] JABI M, PEDERSOLI M, MITICHE A, *et al.* Deep Clustering: On the Link between Discriminative Models and K-means[J]. *IEEE Trans Pattern Anal Mach Intell*, 2021, **43** (6): 1887–1896. DOI: 10.1109/TPAMI.2019.2962683.
- [35] YANG X, DENG C, ZHENG F, *et al.* Deep Spectral Clustering Using Dual Autoencoder Network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 4061–4070. DOI: 10.1109/CVPR.2019.00419.
- [36] MRABAH N, KHAN N M, KSANTINI R, *et al.* Deep Clustering with a Dynamic Autoencoder: From Reconstruction towards Centroids Construction[J]. *Neural Netw*, 2020, **130**: 206–228. DOI: 10.1016/j.neunet.2020.07.005.
- [37] ZHANG R L, ZHENG H Y, WANG H P. TDEC: Deep Embedded Image Clustering with Transformer and Distribution Information[C]//Proceedings of the 2023 ACM International Conference on Multimedia Retrieval. New York: ACM, 2023: 280–288. DOI: 10.1145/3591106.3592268.
- [38] RONEN M, FINDER S E, FREIFELD O. DeepDPM: Deep Clustering with an Unknown Number of Clusters [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 9851–9860. DOI: 10.1109/CVPR52688.2022.00963.