

深度聚类算法研究综述

姬 强, 孙艳丰, 胡永利, 尹宝才

(北京工业大学信息学部多媒体与智能软件技术北京市重点实验室, 北京 100124)

摘要: 聚类分析是挖掘数据内在结构的关键技术,在大数据时代,人们面对的数据通常具有规模大、维度高、结构复杂等特点,直接应用传统聚类算法往往会失效. 深度学习凭借层次化非线性映射能力使得大规模深度特征提取成为可能,因此基于深度学习的聚类(深度聚类)算法迅速成为无监督学习领域的研究热点. 该文旨在对深度聚类的研究现状进行归纳和总结. 首先,从神经网络结构、聚类损失和网络辅助损失3个角度介绍深度聚类的相关概念;然后,根据网络的结构特点对现有的深度聚类算法进行分类,并分别对每类方法的优势和劣势进行分析和阐述;最后,提出好的深度聚类算法应具备的三要素:模型的可扩展性、损失函数的鲁棒性和特征空间的平滑性,并从这3个方面分别阐述未来可能的研究方向.

关键词: 聚类; 深度学习; 特征学习; 神经网络结构; 聚类损失; 网络辅助损失

中图分类号: TP 391

文献标志码: A

文章编号: 0254-0037(2021)08-0912-13

doi: 10.11936/bjtxb2021010013

Review of Clustering With Deep Learning

Ji Qiang, SUN Yanfeng, HU Yongli, YIN Baocai

(Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology,
Beijing University of Technology, Beijing 100124, China)

Abstract: Cluster analysis is a key technology to explore the intrinsic structure of data. In the era of big data, the data we face usually has the characteristics of large scale, high dimensionality, and complex structure. Traditional clustering algorithms often fail to process such data. Deep learning makes large-scale deep feature extraction possible with powerful hierarchical nonlinear mapping capabilities. Therefore, clustering algorithms based on deep learning (deep clustering) have quickly become a research focus in unsupervised learning. This paper aims to summarize the research status of deep clustering. First, the related concepts of deep clustering was introduced from the perspectives of neural network structure, clustering loss and ancillary network loss. Then the existing deep clustering algorithms were classified according to the structural characteristics of the network, and the advantages and disadvantages of each type of methods were analyzed. Finally, three elements of a good deep clustering algorithm were proposed, namely scalability of the model, robustness of the loss function and smoothness of the feature space, and the possible future research opportunities from each of these three aspects were illustrated.

收稿日期: 2021-01-05

基金项目: 国家自然科学基金资助项目(61772048)

作者简介: 姬 强(1989—), 男, 博士研究生, 主要从事深度学习和模式识别方面的研究, E-mail: jiqiang1989@emails.bjut.edu.cn

通信作者: 孙艳丰(1964—), 女, 教授, 博士生导师, 主要从事模式识别、优化算法方面的研究, E-mail: yfsun@bjut.edu.cn

Key words: clustering; deep learning; feature learning; neural network architecture; clustering loss; network-assisted loss

当今社会是一个信息爆炸的时代,随着 5G 技术的日趋成熟,数据的采集量和传输量呈指数增长。各个领域每时每刻都在产生海量的图像、文本、视频、音频等数据。这些数据往往具有规模大、实时性强、维度高、结构复杂等特点。这无疑给人工标注带来严峻的挑战。因此,如何能以无监督的方式从海量复杂数据中挖掘出有效信息是业界亟待解决的难点。

聚类作为无监督学习领域中一项关键技术,其通过发现数据间内在结构的相关性,将数据进行自动的分组。传统的聚类方法大致分为 5 类,分别是:以 K-means^[1] 等为代表的划分聚类、以 DBSCAN^[2] 为代表的密度聚类、以 Chameleon^[3] 为代表的层次聚类、以谱聚类^[4] 为代表的图聚类和以高斯混合聚类(Gaussian mixture model, GMM)^[5] 为代表的模型聚类。虽然上述的传统方法有其各自擅长的聚类情形,但是当面对大规模高维复杂的数据时,直接应用传统方法往往会失效。其原因在于:1) 传统欧式距离度量难以准确评估样本之间的相似性。2) 高维空间数据分布稀疏,簇类结构不清晰。3) 规模大、维度高所带来的高时间和空间复杂度难题,限制了算法的应用。针对上述问题,一种结合降维技术和聚类算法的解决方案被广泛认可,即先采用降维技术,如主成分分析^[6]、局部线性嵌入^[7]或核方法^[8]等对数据进行降维,再对降维后的特征进行聚类。该方案虽然在一定程度上降低了高维空间的聚类难度,但由于数据降维是独立于聚类任务的,这意味着提取的特征往往并不具备簇类结构。

子空间方法则提供另一种很好的思路。该方法假设高维数据分布于多个低维子空间的组合,通过将高维数据分割到各自所属的本征低维子空间,从而完成聚类。子空间聚类方法大致分为 5 类,即基于矩阵分解的方法^[9-10]、基于代数的方法^[11-13]、基于迭代的方法^[14-15]、基于统计的方法^[16-18]和基于谱聚类的方法^[19-21]。其中以稀疏子空间聚类(sparse subspaces clustering, SSC)^[20]和低秩子空间聚类(low-rank subspaces clustering, LRSC)^[21]为代表的谱聚类方法,凭借其完备的理论与突出的性能在同类方法中脱颖而出,成为近年来的研究热点。然而上述 2 种方案均采用的是浅层模型,由于浅层结构

的表达能力有限,导致其在处理结构复杂的数据时无法有效地提取更深层次的内在特征。

近年来,得益于深度学习的蓬勃发展,深度神经网络凭借其层次化结构和非线性映射能力使得大规模深层次特征提取成为了可能。因此,如何利用深度学习来提高聚类性能迅速成为无监督学习领域的研究热点。本文通过对现有的深度聚类工作进行归纳和梳理,总结出深度聚类的 3 个重要组成部分:神经网络结构、聚类损失和网络辅助损失,并全面地阐述和分析深度聚类的研究进展、面临的挑战以及未来可能的研究方向。

1 相关概念介绍

本章的主要内容是对深度聚类的 3 个重要组成部分:神经网络结构、聚类损失和网络辅助损失所涉及的相关概念进行介绍。

1.1 神经网络结构

神经网络主干可以看作深度聚类的骨骼,其主要作用为将原始输入经过层次化的非线性映射,得到新的特征表示。不同的神经网络结构其特征学习的方式以及所学特征的特点也不尽相同。主要采用的神经网络结构有以下几种。

1) 多层全连接网络

多层全连接网络(多层感知机)是一种基础的前馈结构网络,其中第一层为输入层,最后一层为输出层,中间的多层为隐藏层,并且每一层的神经元均与下一层神经元全连接。除了输入层,每层的神经元均可由非线性激活函数映射,网络的训练由反向传播算法完成。

2) 卷积神经网络

卷积神经网络(convolutional neural network, CNN)^[22]可以概括为一类具有卷积计算操作的前馈神经网络。不同于全连接网络,卷积神经网络能够利用输入数据的二维结构,避免输入向量化丢失空间信息。典型的卷积神经网络主要由 3 个部分组成:卷积层、池化层和全连接层。卷积层负责提取原始输入的局部特征;池化层用来大幅降低网络参数量级(降维);全连接层类似传统神经网络的部分,用来输出最终的特征。目前卷积神经网络已经成为深度学习主流的网络结构。

3) 深度信念网络

深度信念网络(deep belief networks, DBN)^[23]是一种由多个堆叠受限玻尔兹曼机(restricted Boltzmann machines, RBM)^[24]构成的概率有向图模型,其图结构由多层的节点构成. 每层节点的内部没有连接,相邻两层的节点之间为全连接. 网络的最底层为可观测的变量,其他层节点均为隐变量. 最顶部的两层间的连接是无向的,其他层之间的连接是有向的. 深度信念网络的训练分为2个阶段,首先是对RBM层自底向上逐层进行预训练,最小化每层的重构误差,然后再根据相应的目标函数对网络整体进行微调.

4) 自编码器

自编码器(autoencoder, AE)是一种用于无监督特征学习的神经网络,其结构包含2大模块:编码器 f_θ 和解码器 g_ϕ . 通过编码器将原始输入 x 映射到特征空间 z ,即编码过程,然后再通过解码器将嵌入特征 z 映射回原始输入空间得到重构样本 x' ,即解码过程. 其优化目标则是通过最小化原始输入样本与重构样本之间的误差损失来同时训练编码器和解码器,从而学习得到样本输入的特征表示. 其目标函数为

$$\mathcal{L}(\theta, \phi) = \sum_i \|x_i - g_\phi(f_\theta(x_i))\|^2 \quad (1)$$

式中 θ 和 ϕ 分别表示编码器和解码器的参数. 自编码器可以大致分为标准自编码器(standard autoencoder, SAE)^[25]和变分自编码器(variational autoencoder, VAE)^[26]. 区分的依据在于是否对特征空间施加特定分布的约束.

5) 生成对抗网络

生成对抗网络(generating adversarial networks, GAN)^[27]是一种深度生成模型. 该模型框架包含两个相互竞争的神经网络结构:判别器 D 和生成器 G . 其中,生成器可以理解为一个伪造者,试图通过构造假样本骗过判别器的甄别;判别器则是尽可能甄别样本是来自于真实样本还是生成器构造的假样本. 二者通过对抗学习提升各自的能力,即生成模型生成更逼真的假样本骗过判别模型,而判别模型能更准确地识别真假样本. 整个的优化目标函数为

$$\min_G \max_D V(D, G) = E_{x \sim p_d} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(z))] \quad (2)$$

1.2 损失函数

目前深度聚类模型的损失函数普遍采用聚类损失与网络辅助损失组合的方式,即

$$\mathcal{L} = \lambda \mathcal{L}_R + (1 - \lambda) \mathcal{L}_C \quad (3)$$

式中: \mathcal{L}_C 和 \mathcal{L}_R 分别为聚类损失和网络辅助损失; $\lambda \in [0, 1]$ 为调节二者权重比例的超参数.

1.2.1 聚类损失

聚类损失是帮助深度聚类模型发现数据内在簇类结构并提取聚类友好型特征的关键部分. 聚类损失对于模型的聚类性能起决定性作用. 针对聚类损失函数的设计,目前较为普遍的思路是将传统聚类算法的损失函数迁移到神经网络结构中,通过交替优化或联合优化的方式实现聚类. Yang等^[28]将K-means的损失函数定义在自编码器的特征空间,通过网络参数和聚类中心的交替优化实现特征学习和聚类. JULE^[29]将层次聚类算法与卷积神经网络相结合,通过簇类合并与特征学习的交替优化,从而实现特征聚类. SpectralNet^[30]将谱聚类的思想引入到深度学习中,首先通过孪生网络学习特征间的相似度矩阵,然后根据谱聚类目标函数进行特征学习,最后在特征空间执行K-means来获得聚类分配. 变分深度嵌入(variational deep embedding, VaDE)^[31]将高斯混合模型聚类的思想引入变分自编码器中,通过对特征空间的分布约束,实现特征学习和聚类分配的联合优化.

另一种思路则是根据所期望的聚类假设,直接设计特定的聚类损失函数. 深度嵌入聚类(unsupervised deep embedding for clustering analysis, DEC)^[32]定义了2种概率分布:软聚类分配概率分布和辅助聚类分配概率分布,通过最小化2种分布的KL散度来实现聚类. 深度自适应聚类(deep adaptive image clustering, DAC)^[33]则是将多分类问题转化为二分类问题,基于自步学习的思想不断产生高置信度的正负样本对,并将其作为监督信息指导模型的训练,最终输出聚类结果. 自增强信息聚类(information maximizing self-augmented training, IMSAT)^[34]和不变信息聚类(invariant information clustering, IIC)^[35]都是基于同样的假设,即对数据的简单变换不会改变其内在的语义信息,通过最大化原始样本与其增强样本的信息熵来实现聚类.

1.2.2 网络辅助损失

由于聚类问题没有监督信息作为指导,单独使用聚类损失作为目标函数很容易导致模型退化. 网络辅助损失则可以理解为模型的正则项,虽然优化该损失函数不能直接获得聚类结果,但是通过对模型参数进行相应的约束,可以帮助模型稳定地训练,提升特征学习的有效性,避免出现退化解. 网络辅

助损失中一类是由于网络模型的结构特点自带的损失项,例如自编码器中的重构误差损失项、生成对抗网络中生成器和判别器的对抗损失项. 另一类则是基于对网络模型的约束期望而精心设计的. 例如局部保持损失^[36]其目的是期望神经网络提取的特征能够满足设定的近邻关系. 组稀疏损失^[36]将网络隐藏层的神经元分组,其目的是希望同类样本只激活相同组的神经元,不同类样本激活不同组的神经元,从而实现神经元结构化激活,获取更具判别性的特征表示.

2 深度聚类算法分类

近年来,深度学习在聚类分析领域取得长足的发展,越来越多优秀的深度聚类算法应运而生. 深度聚类能够实现特征学习和聚类的联合优化,逐渐成为处理大规模复杂数据聚类问题的首选方案. 为了更清晰地了解深度聚类算法之间的区别和联系,本文根据网络结构的特点,将已有的深度聚类模型大致分为 3 个类别. 第 1 类是以自编码器为模型框架的聚类算法,该类方法在训练过程中,通常会希望学习低维的中间层特征,并同时基于中间层特征构建相应的聚类损失. 第 2 类是以深度前馈神经网络(全连接或卷积)为模型框架的聚类算法,该类方法通常会基于网络的输出层特征构建聚类损失,形成一种端到端的聚类模式. 第 3 类是以生成对抗网络为模型框架的聚类算法,该类方法的关键在于让输入的噪声分布满足分类性质,使得模型不仅能够生成逼真的样本,还能够生成指定类别的样本. 基于上述的分类模式,本文将具有代表性的深度聚类算法进行梳理和总结.

2.1 基于自编码器的聚类算法

自编码器作为一种经典的无监督学习方法,得益于自身良好的特征学习能力,使其成为深度聚类最普遍采用的网络结构. 该类方法的核心在于如何基于中间层特征构建好的聚类损失和辅助聚类损失,帮助模型发现数据内在的簇类结构. 基于标准自编码器的聚类算法总体模型框架如图 1 所示. 基于不同的聚类损失衍生出不同的算法.

Yang 等^[28]将自编码器与 K-means 聚类相结合,首先预训练自编码器来提高网络参数的初始化质量,然后将基于中间层特征构建的 K-means 聚类损失引入到模型中,通过网络参数与聚类中心之间的交替优化,完成特征学习与聚类任务. 由于 K-means 聚类操作是不可导的,因此 DCN 采取的是交

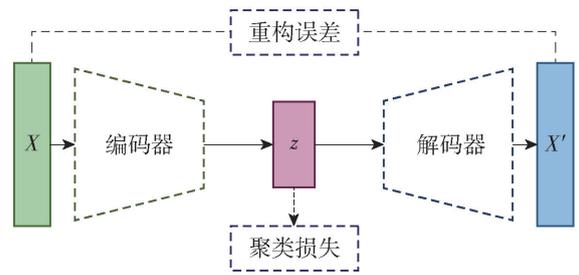


Fig. 1 Clustering algorithm framework based on standard autoencoder

替优化的方式. Fard 等^[37]基于软最大化函数 (softmax) 构建光滑可导的 K-means 聚类损失,从而实现聚类中心和网络参数的同步优化.

Ren 等^[38]通过将密度聚类与自编码器相结合,提出一种深度密度聚类框架. 该算法完成聚类需要 2 个阶段:第 1 阶段通过训练深度卷积自编码器学习低维的特征表示,并使用随机近邻嵌入算法 (t-distributed stochastic neighbor embedding, t-SNE)^[39]将提取的特征进一步降维到二维空间;第 2 阶段使用密度峰值算法^[40]对二维空间的特征进行聚类. 该算法继承了密度峰值聚类的优点,无须人为指定聚类数目并能有效处理任意形状的聚类情形.

Chen 等^[41]基于同类样本来自相同流形结构的假设,提出一种深度流形聚类算法. 该算法针对自编码器的中间层特征设计 2 种损失函数,其中局部保持损失用来挖掘数据内在的流形结构,聚类导向损失则会让在流形上距离相近的特征更加紧凑,并形成簇类. Huang 等^[36]不仅考虑到特征的局部特性,还将组稀疏的思想引入到模型中,实现特征层神经元的结构化激活,从而获得更具可分性的特征表示.

目前很多深度聚类模型通过聚类伪标签与特征表示的交替或联合优化的方式完成训练. 但在训练的初始阶段,由于特征的随机性,聚类标签会极为不可靠,二者的相互作用很可能导致模型的恶性循环. 针对此问题,Mrabah 等^[42]提出一种模型训练方式. 在预训练阶段,通过引入数据增强和对抗插值技术^[43],以自监督的方式学习可靠的特征表示. 在聚类学习阶段,不仅引入聚类损失,而且基于原始输入与重构输出构建判别器网络,最终使得模型在完成聚类的同时,学到更平滑的特征空间.

针对传统子空间聚类模型特征学习能力的不足,诸多基于自编码器的深度子空间聚类算法应运而生. Peng 等^[44]首先将基于原始数据构建的稀疏自表达式作为先验信息,然后约束中间层特征满

足先验信息的自表达性质,从而学到更具判别性的特征. 然而该算法的自表达系数在特征学习过程中是固定不变的,这会制约特征的表达能力. 针对此问题, Ji 等^[45]直接将自表达系数定义为网络的可学习参数,并基于自编码器的中间层特征构建自表达层,实现自表达系数与特征的联合优化,提出深度子空间聚类模型(deep subspace clustering networks, DSC-Nets). Zhou 等^[46]认为 DSC-Nets 模型在训练时缺少有效信息的监督,导致其聚类性能不稳定,提出深度对抗子空间聚类算法(deep adversarial subspace clustering, DASC). 该算法将生成对抗网络的思想引入到 DSC-Nets 模型中,并基于如下假设,即相同子空间数据的线性组合仍处于该子空间;反之不同子空间数据的线性组合处于不同子空间,生成新的正负样本,并通过对抗训练的方式获得高质量的自表达系数和特征表示. Zhang 等^[47]针对子空间聚类无法处理大规模数据的弊端,提出神经协同

子空间聚类算法,该算法扩展了 DSC-Nets 的模型框架,构建分类表示与自表达系数双模块,通过二者的协同学习解决大规模数据的子空间聚类问题.

以单一自编码器为网络框架构建聚类模型,虽然较传统聚类方法已有显著优势,但是网络结构自身仍有很大的探索空间. Dizaji 等^[48]提出共享的双编码器结构:噪声编码器和干净编码器,基于双编码器构建相对熵聚类损失,实现聚类与特征学习的联合优化. Yang 等^[49]针对低维特征的重构模糊化问题,提出双自编码器网络结构,通过引入相对重构损失和互信息估计来学习更具判别性的特征表示,并联合优化双自编码器和深度谱聚类获得最佳聚类结果.

不同于标准自编码器,变分自编码器作为生成模型,不仅能够完成输入的压缩和重构,还可以生成新的有意义的输出. 其关键在于对特征空间施加特定分布的约束. 基于变分自编码器的聚类算法模型框架如图 2 所示.

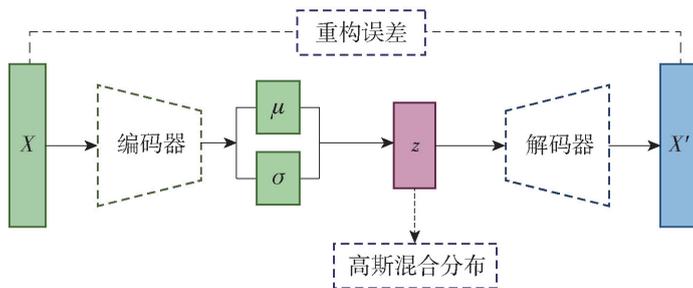


图 2 基于变分自编码器的聚类算法框架

Fig. 2 Clustering algorithm framework based on variational autoencoder

VaDE 是基于变分自编码器的生成式聚类模型. 该算法通过引入高斯混合模型来建模数据生成的过程. 具体建模过程分为 3 步:1) 由高斯混合模型选择一个簇类;2) 根据选择的簇类生成潜在的嵌入特征;3) 用解码器将嵌入特征重构输入. VaDE 的优化遵循变分自编码器的优化方式,即使用随机梯度变分贝叶斯(stochastic gradient variational Bayes, SGVB)和重参数技术优化证据下界(evidence lower bound, ELBO). GMVAE^[50]采用了与 VaDE 相似的策略,对特征空间施加高斯混合分布的约束. 同时针对 VAE 存在的过度规范化(over-regularization)的问题,GMVAE^[50]通过最小化信息约束使得模型在训练初期避免陷入局部解. Yang 等^[51]认为捕获数据的局部结构是对 VaDE 模型的必要补充,因此将图嵌入技术引入到模型中,使得嵌入特征的分布不再是独立学习,而是基于图信息关联特征所属的分布距离,从而帮助模型处理复杂数

据的聚类问题.

2.2 基于深度前馈神经网络的聚类算法

深度前馈神经网络包含众多目前主流的结构,比如 LeNet^[22]、AlexNet^[52]、VGGNet^[53]、GoogleNet^[54]、ResNet^[55]等. 这些模型凭借其强大的特征学习能力被广泛应用于诸多领域. 但不同于自编码器等无监督学习模型,深度前馈神经网络通常需要监督信息的指导来完成有效的特征学习. 这无疑提高了该类网络模型在聚类学习领域的使用门槛. 鉴于此,如何设计可靠的聚类损失函数来指导模型完成聚类任务迅速成为学者们的研究热点. 相较于基于自编码器的聚类算法,该类算法对聚类损失的鲁棒性要求更高,其关键在于如何产生可靠的确定性信息来辅助模型训练. 其算法框架如图 3 所示. 根据产生确定性信息的方式,本文将该类聚类算法分为 2 个子类:基于聚类进化的方法和基于自监督学习的方法.

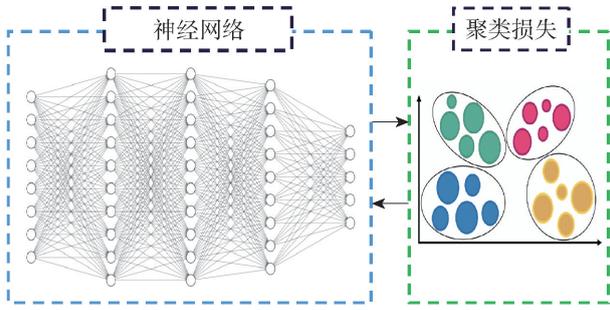


图 3 基于深度前馈神经网络的聚类算法框架

Fig. 3 Clustering algorithm framework based on deep feedforward neural network

2.2.1 基于聚类进化的方法

该类方法的主要特点在于通过优化聚类损失不断地产生或强化高置信度信息,并以此作为监督信息来指导模型训练。

DEC^[32]是具有代表性的深度聚类算法,该算法定义了 2 种概率分布:软聚类分配分布 Q 和辅助聚类分配分布 P 。模型具体定义为

$$\begin{aligned} \mathcal{L} &= \text{KL}(P \parallel Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \\ q_{ij} &= \frac{(1 + \|z_i - u_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_j (1 + \|z_i - u_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}} \\ p_{ij} &= \frac{q_{ij}^2/f_j}{\sum_j q_{ij}^2/f_j} \end{aligned} \quad (4)$$

式中:分布 Q 为基于学生 t 分布来建立特征 z 和聚类中心 u 的距离关系; q_{ij} 表示样本 i 属于第 j 类的概率; $f_j = \sum_i q_{ij}$ 可以理解为簇类 j 包含的样本数;分布 P 则可以看作对分布 Q 中的高置信度关系进行强化预测。通过最小化聚类分布 Q 与聚类强化分布 P 的 KL 散度来实现最终的聚类分配。Li 等^[56]借鉴了 DEC 的聚类思想,并通过使用卷积自编码器提升了算法的性能。Guo 等^[57]认为单一优化 DEC 定义的聚类损失具有破坏特征空间的风险,通过采用聚类损失和重构损失的集成方式来提升模型的鲁棒性。

Yang 等^[29]将模型的训练分为 2 个过程:特征空间的聚类过程和有监督的特征学习过程。首先模型采用自底向上的层次聚类算法对特征进行簇类合并,并选择高置信度的聚类结果作为监督信息,然后根据得到的监督信息进行特征学习。2 个过程用同一个优化问题表示,使二者目标函数一致,避免了分部带来的误差积累。聚类过程与特征学习过程交替

优化,最终完成整个模型的训练。

Chang 等^[33]提出的深度自适应聚类算法将聚类问题转化为二元判别问题,即判断两个样本是否属于同一类。首先模型将深度卷积网络的输出层特征定义为标签特征。然后计算标签特征之间的余弦相似度,并根据设置的阈值选取高置信度的正负样本对作为标记样本,DAC 以监督的方式利用标记样本训练模型。基于自步学习^[58]的思想,通过降低阈值要求不断增加监督信息,直到全部样本都参与训练。

Shaham 等^[30]针对谱聚类算法可扩展性差、无法样本外测试等不足,提出一种深度谱聚类网络。该算法首先根据设置的阈值通过孪生网络^[59]学习特征之间的相似度矩阵,并以此作为监督信息。然后在强约束特征正交性同时优化谱聚类损失函数,从而学习深度谱嵌入特征。最终通过对特征执行 K-means 聚类来获得聚类分配。

Peng 等^[60]从距离度量的角度提出一个聚类假设,即属于同一簇类的样本具有距离度量的不变性。该算法首先采用预训练的方式初始化网络参数,然后基于 2 种不同距离度量方式来设计聚类分配分布,最终通过最小化不同聚类分配分布的 KL 散度来实现聚类。

2.2.2 基于自监督学习的方法

自监督学习主要是利用辅助任务 (pretext) 从大规模的无监督数据中挖掘自身的监督信息,通过这种构造的监督信息对网络进行训练,从而可以学习到对下游任务有价值的表征。该类方法的主要特点是通过数据增强等技术为模型变向增加监督信息,以自监督学习的方式来提升模型的鲁棒性和有效性。

IMSAT^[34]是一种无监督的离散表示学习算法,其任务是通过训练深度神经网络将输入数据映射为离散的特征表示。该算法借鉴正则化信息最大化 (regularized information maximization, RIM) 模型的聚类思想,首先利用分类器网络输出聚类分配概率,然后通过最大化输入和离散聚类分配之间的互信息来实现聚类输出。同时,IMSAT 提出自我增强训练,通过最大化输入样本与其增强样本之间的互信息,显著增强模型的聚类性能。

Ji 等^[35]认为缺乏语义过滤,将传统聚类算法与深度学习强行结合并不能保证模型学到有聚类意义的特征,进而提出不变信息聚类算法。该算法回归网络结构本身,构建网络双输入,即原始样本输入和增强样本输入,根据信息瓶颈理论^[61],通过直接最

大化原始样本与增强样本分类概率输出之间的互信息,能够有效地过滤冗余信息,捕捉到样本之间的不变信息. Nina 等^[62]同样从网络结构入手,构建网络的三元输入结构:原始样本、增强正样本和随机负样本,并根据三者的语义类别设计三元聚类损失,提出一种基于三元关系挖掘的深度聚类模型.

Wu 等^[63]认为基于单一数据关联关系学到的特征表示缺乏识别能力,提出深度综合关联挖掘聚类算法 (deep comprehensive correlation mining, DCCM). 该算法全面总结了原始样本与特征之间的关系,包括:不同样本的相关性、不同特征的相关性、对于数据变换的局部鲁棒性以及同一样本在不同网络层之间的关系. 首先,针对不同样本之间的相关性,采用深度卷积网络生成输入的预测特征. 在适当的约束条件下,预测特征将趋向于 one-hot 向量;然后计算余弦相似度,构造相似度图. 基于相似度图和预测特征,通过设置阈值来获得高置信度的伪图和伪标签来指导特征学习. 其次,通过引入数据增强技术,提升模型的局部鲁棒性. 最后,通过最大化输入与其特征的互信息来提升特征的识别能力,结合这 3 个不同方面的损失函数,以端到端的方式完成关联挖掘和聚类输出.

Huang 等^[64]认为将属于相同语义类别的样本划分到不同类别会同时降低类内的紧密性和类间的多样性,提出深度语义聚类算法. 该算法将最大化间隔聚类^[65]的思想引入到深度学习中,不对特征之间的相关性做任何假设,通过优化全局划分置信度损失获得最终的聚类分配.

Li 等^[66]认为特征矩阵的行和列可分别视为样本的软标签特征和类表达特征,基于上述的发现,作者通过同时优化软标签特征和类表达特征的对比损失,即最大化正样本对的相似度,最小化负样本对的相似度,从而以端到端的方式实现特征学习和聚类分配的联合优化.

2.3 基于生成对抗网络的聚类算法

随着生成对抗网络在诸多领域的成功应用,越来越多的学者将研究工作聚焦在如何利用对抗训练的思想来解决聚类任务中的问题.

对抗自编码器^[67]包含 3 个模块——编码器、解码器、判别器,前两者构成一个普通的自编码器,判别器输入编码向量,判定它是来自一个真实的先验分布,还是来自编码器的输出. 判别器试图区分编码向量的真假,而编码器则扮演生成器的角色试图混淆判别器,二者通过对抗训练完成特征学习. 对

抗自编码器是一个通用框架,能够实现监督学习、半监督学习和无监督学习,针对聚类任务,通过对输入编码为连续特征和分类特征,并以重构损失和对抗损失双重目标训练网络,分类特征即为聚类分配输出. 深度对抗聚类^[68]是一种生成式聚类模型,该算法沿用了对抗自编码器网络框架,其差异在于约束编码特征服从高斯混合分布,从而使得最终学到的特征具有聚类结构.

CatGAN^[69]是一种从未标记或部分标记的数据中学习判别分类器的算法,该算法将生成对抗网络中判别器的二分类问题扩展为多分类问题,并通过多分类目标函数权衡观察到的示例及其预测的分类类别分布之间的相互信息以及分类器自身的鲁棒性.

GAN 通过生成器与判别器的对抗学习来生成逼真的输出,但是由于生成器的输入是随机的噪声信号,没有任何约束,导致输入信号维度的可解释性很差. 为了实现输入信号的解耦表示,InfoGAN^[70]将输入向量分为两部分:连续的噪声向量和潜在语义向量,其中 Info 代表生成数据与潜在语义向量之间的互信息. InfoGAN 在对抗训练的同时最大化二者的互信息,从而得到可分解的特征表示. 而语义向量则可以控制模型生成指定类别的样本,实现无监督聚类生成. 其具体模型框架如图 4 所示.

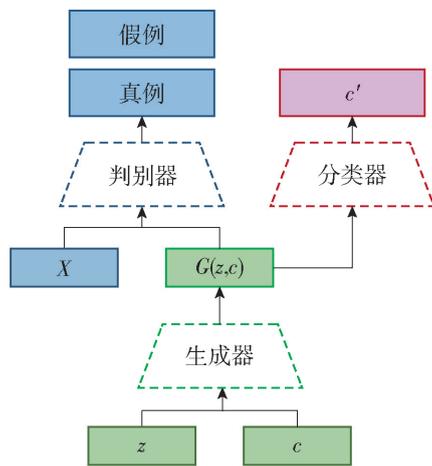


图 4 InfoGAN 模型框架

Fig. 4 Framework of InfoGAN

由于生成对抗网络的潜变量分布通常是平滑的,不存在可观测簇类结构,不利于聚类任务. 针对该问题, Mukherjee 等^[71]提出聚类生成对抗网络. 该算法利用混合的离散和连续潜变量 (one-hot 编码和高斯随机变量),以创建一个非光滑几何潜空间,其目的是既保证良好的类间插值 (高斯随机变量),又能实现更好的聚类效果 (one-hot 编码). 通过设

计一种能适应离散-连续混合情形的反向传播算法,并给出显式的逆映射网络来获取给定数据的潜在变量.将生成对抗网络与反映射网络联合训练,并以特定聚类损失来训练模型,从而实现聚类输出.

3 深度聚类算法的应用

得益于深度学习的快速发展,深度聚类算法凭借其层次化结构和强大的非线性映射能力使得大规模深度特征提取成为了可能,是目前机器学习、计算机视觉、图像处理、模式识别等领域的研究热点.深度聚类算法相较于传统聚类算法更擅长处理大规模高维非结构化数据,在图像聚类、人脸识别、图像分割等领域取得了成功的应用.

3.1 图像聚类

自然图像聚类是一项极具挑战性的任务,传统的聚类算法往往无法有效处理高维非结构化的图像数据,深度聚类算法凭借其强大的特征学习能力成功地突破了自然图像聚类的瓶颈.Chang等^[74-75]提出的深度自进化图像聚类算法(deep self-evolution clustering, DSEC)将聚类问题转化为二元判别问题,即判断2个样本是否属于同一类.首先模型将深度卷积网络的输出层特征定义为标签特征.然后计算标签特征之间的余弦相似度,并根据设置的阈值选取高置信度的正负样本对作为标记样本,DSEC以监督的方式利用标记样本训练模型.基于自步学习^[58]的思想,通过降低阈值要求不断增加监督信息,直到全部样本都参与训练.Zhao等^[76]提出了一种深度图像聚类框架,该算法旨在学习一种类别式的潜在表示,以实现类别信息与图像风格是解耦关系.为了实现这一目标,应用互信息最大化的方法将相关信息嵌入到潜在表示中.此外,还采用了增强不变量损失来将表征分解为类别部分和风格部分.最后对潜在表示施加了一个先验分布,以确保类别向量的元素可以作为聚类的概率.

3.2 人脸识别

人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术,是计算机视觉与模式识别领域的一个重要研究问题.Tapaswi等^[77]提出一种球簇学习(ball cluster learning, BCL),通过将嵌入空间刻画成大小相等的球簇,在模型训练过程中将学习到的球半径转化为迭代合并算法的停止标准,从而实现人脸聚类.由于图像在不同的照明和姿态条件下图像质量的差异化,导致训练(源域)和测试(目标域)数据之间的域差异,Wang等^[78]提出了一

种基于聚类的域适应方法,该方法是针对源域和目标域不共享的人脸识别任务而设计的.通过在全局范围内对齐特征域,有效地学习鉴别性的目标特征,同时在局部区分目标簇.通过最小化全局域的差异性来学习更可靠的聚类表示,以减少域的差距,最后应用简化的谱聚类方法在域不变的特征空间中生成聚类结果.Lin等^[79]提出一种深度密度聚类算法(deep density clustering, DDC),该算法基于测量特征空间中局部邻域之间的密度亲和力,通过学习每个邻域的最小覆盖球,封装底层结构的信息.该封装还能够定位邻域的高密度区域,有助于测量邻域的相似性.并且从理论上证明,封装渐进地收敛到Parzen窗口密度估计器.

3.3 图像分割

图像分割是指将图像分成若干具有相似性质的区域的过程,是计算机视觉领域的一个重要的研究方向,是图像语义理解的重要一环.Khan等^[80]提出了一种基于深度学习的无监督图像分割方法,该方法包括基于全卷积密集网络的无监督深度表示为导向进行聚类,然后通过将浅层特征的高维区域合并,以产生最终的分割图像.Zhou等^[81]提出了一种基于新型深度图像聚类(deep image clustering, DIC)模型的无监督分割框架.DIC由一个特征变换子网络(feature transformation subnetwork, FTS)和一个可训练的深度聚类子网络(deep clustering subnetwork, DCS)组成,用于无监督的图像聚类.FTS建立在一个简洁高效的网络架构上.DCS可以通过迭代更新聚类关联和聚类中心来分配不同聚类号的像素.Saha等^[82]提出了一种基于语义引导的无监督卷积神经网络的图像分割方法,该方法不需要任何标记的训练数据,并且可以在单个图像输入上操作.该算法通过使用预先训练的网络来提取中层的深层特征,以捕获输入图像的语义,提取的深层特征被进一步反馈到可训练的卷积层.最后通过softmax分类层的进一步空间细化获得分割标签.获得的分割标签和可训练的卷积层权重在迭代中联合优化,从而实现将空间相邻的像素和相似特征的像素分配到同一个标签上.

4 深度聚类算法总结与分析

纵观上述3类深度聚类方法,都从不同的角度提升了聚类算法的性能,但同时也各自带来新的难题与挑战.本节将从这3类深度聚类方法切入,分别总结其各自的优势与不足.

凭借其独有的编码器与解码器结构,通过对输入样本进行重构,自编码器同主成分分析算法思想一样,能够有效地提取包含原始输入重要信息的特征表示,这一特点决定了基于自编码器的聚类方法普遍具有较好的鲁棒性.原因在于神经网络其庞大的参数量,使得整个网络模型的自由度非常高,而聚类问题又缺少确定性的先验信息,如果对网络参数缺乏合理的约束,极易导致模型训练无效而得到退化的特征.自编码器的重构误差损失作为正则项会使得模型在兼顾聚类损失目标的情况下,约束特征空间在一个合理的范围以确保其能够保留输入样本的重要信息,使提取的特征表示不会过于受到聚类损失的影响而导致特征空间的合理性被破坏,即聚类损失目标虽然被充分优化,但学到的特征却无实际应用价值.但也正是由于这种编码器与解码器对称的网络结构,使得模型需要计算资源和时间来训练解码器实现重构,这会导致在计算资源有限的情况下,编码器的深度被严重限制,从而使特征的表达能力也会显著降低.并且当模型训练完毕后,解码器也不会参与新样本的聚类预测,这更是一种对计算资源的浪费.因此如何权衡模型的鲁棒性和扩展性是该类方法亟待解决的难题.

重构误差虽然在一定程度上能够提升模型的鲁棒性,但并没有理论能够证明重构误差有助于帮助模型提取具有簇类结构的特征.不同于自编码器,深度前馈神经网络无须构建解码器对输入样本进行重构,使得编码器的可扩展性得到显著的提升,而且由于深度前馈神经网络发展速度飞快,拥有众多成熟并且前沿的网络结构,这无疑更加保障了其强大的特征表达能力.但另一方面,由于多数的深度前馈神经网络设计本身都是为有监督任务服务的,其结构并没有自带无监督网络损失来有效地约束模型参数,因此模型的训练需要依赖给定的监督信息或目标函数.这意味着相较于基于自编码器的聚类算法,基于深度前馈神经网络的聚类方法对聚类损失的鲁棒性有更高的要求,传统聚类算法的目标函数往往无法迁移到网络结构中直接使用,因为很容易出现聚类目标损失被充分优化但所学特征无实际意义的情况.因此如何设计鲁棒性好的聚类损失是突破该类方法瓶颈的关键.

上述2种方法都是基于判别式模型设计的.而基于变分自编码器的聚类算法和基于生成对抗网络的聚类算法则是生成式深度聚类模型.生成式聚类模型不仅能够实现聚类,而且还能产生新的样本,拓

宽了方法的应用领域.基于变分自编码器的聚类算法相较于基于传统自编码器的聚类算法的优势在于,前者通过对特征空间施加特定的分布约束(高斯混合分布等),使得特征空间更加的规整和平滑,拥有较为完备的理论保证,提升模型的鲁棒性.但是其同样继承了传统自编码器的网络结构所带来的弊端.相较于变分自编码器,生成对抗模型在先验分布的选择上有着更强的灵活性,其对抗训练的方式能够让模型生成更逼真、更丰富的样本.但是目前基于生成对抗网络的聚类方法无论在数量还是性能方面都逊色于前2类方法,本文认为其主要原因在于2个方面,其一是对抗训练方式收敛慢,容易出现模型坍塌,导致训练失败;其二是生成对抗网络需要生成新样本,计算成本高,难以形成以聚类目标为核心导向的生成模型.

5 未来的研究方向

基于对上述3类方法的总结与分析,本文认为一个好的深度聚类模型应该同时满足3个要素:模型的可扩展性、损失函数的鲁棒性和特征空间的平滑性.

1) 模型的可扩展性

模型的可扩展性对于大规模复杂数据学习是必不可少的.自编码器作为主流的无监督学习模型,其对称的结构特点使得编码器深度受限,虽然已经有理论证明最大化输入与其特征的互信息近似等价于最小化重构误差^[72],为模型不失鲁棒性的基础上摒弃解码器提供了理论支持,但目前的理论建立在严格的假设基础上,且连续互信息的估计方法仍然是不成熟的^[73].因此,对于自编码器可扩展性的理论探索是值得研究的方向.

2) 损失函数的鲁棒性

事实上,深度聚类所面临的多数挑战都可以归咎于确定性信息的缺失.因此如何从数据中产生尽可能多的、可靠的确定性信息,是提升模型鲁棒性和有效性的关键.目前主流的方式是通过数据增强技术,例如对图像进行剪切、旋转、平移、缩放等操作,生成多个增强图像.数据增强技术擅长产生可靠的正样本,但是对于负样本的生成目前并没有较好的方法,因此如何生成可靠的负样本是非常具有研究价值的.

3) 特征空间的平滑性

特征空间的平滑性主要体现在特征插值和分布约束.变分自编码器就是通过对特征空间施加特定

的先验分布,使得学到的特征更加规整,并通过特征插值生成新样本.平滑性约束对于特征学习非常必要,但目前的深度聚类模型缺少对特征平滑性约束的新探索,因此如何设计新颖的特征分布约束是深度聚类研究的重要挑战之一.

4) 新颖的聚类假设

目前深度聚类模型仍然是沿用传统算法的聚类假设,例如基于距离关系、基于类中心、基于高密度区域、基于信息熵等.这导致深度聚类模型同样继承了这些聚类假设的缺陷,甚至缺陷被放大.因此,如何抛开传统聚类的思路,以深度学习的视角建立新颖的聚类假设是突破聚类算法瓶颈的关键.

参考文献:

- [1] LLOYD S P. Least squares quantization in PCM [J]. *IEEE Transactions on Information Theory*, 1982, 28(2): 129-137.
- [2] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C] // *International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 1996: 226-231.
- [3] KARYPIS G, HAN E H, KUMAR V. CHAMELEON: a hierarchical clustering algorithm using dynamic modeling [J]. *Computer*, 1999, 32(8): 68-75.
- [4] NG A, JORDAN M, WEISS Y. On spectral clustering: analysis and an algorithm [C] // *Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2001: 849-856.
- [5] BISHOP C M. *Pattern recognition and machine learning* [M]. Berlin: Springer, 2006.
- [6] WOLD S, ESBENSEN K, GELADI P. Principal component analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1/2/3): 37-52.
- [7] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, 2000, 290(5500): 2323-2326.
- [8] CHANG Y W, HSIEH C J, CHANG K W, et al. Training and testing low-degree polynomial data mappings via linear SVM [J]. *Journal of Machine Learning Research*, 2010, 11(11): 1471-1490.
- [9] BOULT T E, BROWN L G. Factorization-based segmentation of motions [C] // *Proceedings of the IEEE Workshop on Visual Motion*. Piscataway: IEEE, 1991: 179-186.
- [10] WU Y, ZHANG Z Y, HUANG T S, et al. Multibody grouping via orthogonal subspace decomposition [C] // *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2001: 252-257.
- [11] VIDAL R, MA Y, SASTRY S. Generalized principal component analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(12): 1945-1959.
- [12] RAO S R, YANG A Y, SASTRY S, et al. Robust algebraic segmentation of mixed rigid-body and planar motions from two views [J]. *International Journal of Computer Vision*, 2010, 88(3): 425-446.
- [13] MA Y, YANG A Y, DERKSEN H, et al. Estimation of subspace arrangements with applications in modeling and segmenting mixed data [J]. *SIAM Review*, 2008, 50(3): 413-458.
- [14] HO J, YANG M H, LIM J, et al. Clustering appearances of objects under varying illumination conditions [C] // *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2003: 11-18.
- [15] BRADLEY P S, MANGASARIAN O L. K-plane clustering [J]. *Journal of Global Optimization*, 2000, 16(1): 23-32.
- [16] TIPPING M E, BISHOP C M. Mixtures of probabilistic principal component analyzers [J]. *Neural Computation*, 1999, 11(2): 443-482.
- [17] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [18] MA Y, DERKSEN H, HONG W, et al. Segmentation of multivariate mixed data via lossy data coding and compression [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(9): 1546-1562.
- [19] LAUER F, SCHNORR C. Spectral clustering of linear subspaces for motion segmentation [C] // *Proceedings of the 12th IEEE International Conference on Computer*. Piscataway: IEEE, 2009: 678-685.
- [20] ELHAMIFAR E, VIDAL R. Sparse subspace clustering: algorithm, theory, and applications [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2765-2781.
- [21] LIU G C, LIN Z C, YU Y. Robust subspace segmentation by low-rank representation [C] // *Proceedings of the 27th International Conference on Machine Learning*. New York: ACM, 2010: 663-670.
- [22] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*,

- 1998, 86(11): 2278-2324.
- [23] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. *Neural Computation*, 2006, 18(7): 1527-1554.
- [24] HINTON G E. A practical guide to training restricted Boltzmann machines [M]//MONTAVON G, ORR G B, MULLER K R. *Neural networks: tricks of the trade*. Berlin: Springer, 2012: 599-619.
- [25] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [26] KINGMA D P, WELING M. Auto-encoding variational bayes[J]. *ArXiv Preprint ArXiv*, 2013: 1312. 6114.
- [27] GOODFELLOW I J, POUGET A J, MIRZA M, et al. Generative adversarial networks [C] // *Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2014: 2672-2680.
- [28] YANG B, FU X, SIDIROPOULOS N D, et al. Towards k-means-friendly spaces: simultaneous deep learning and clustering [J]. *Proceedings of Machine Learning Research*, 2017, 70: 3861-3870.
- [29] 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*. Piscataway: IEEE, 2016: 5147-5156.
- [30] SHAHAM U, STANTON K, LI H, et al. SpectralNet: spectral clustering using deep neural networks [C] // *International Conference on Learning Representations*. Piscataway: IEEE, 2018: 267-280.
- [31] JIANG Z, ZHENG Y, TAN H, et al. Variational deep embedding: an unsupervised and generative approach to clustering [C] // *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann, 2017: 1965-1972.
- [32] XIE J Y, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis [C] // *International Conference on Machine Learning*. New York: ACM, 2016: 478-487.
- [33] CHANG J, WANG L, MENG G, et al. Deep adaptive image clustering [C]//*IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2017: 5879-5887.
- [34] HU W H, MIYATO T, TOKUI S, et al. Learning discrete representations via information maximizing self-augmented training [C] // *International Conference on Machine Learning*. New York: ACM, 2017: 1558-1567.
- [35] JI X, HENRIQUES J F, VEDALDI A. Invariant information clustering for unsupervised image classification and segmentation [C] // *Proceedings of the IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2018: 9865-9874.
- [36] HUANG P, HUANG Y, WANG W, et al. Deep embedding network for clustering [C] // *Proceedings of the IEEE 22nd International Conference on Pattern Recognition*. Piscataway: IEEE, 2014: 1532-1537.
- [37] FARD M M, THONET T, GAUSSIER E. Deep k-means: jointly clustering with k-means and learning representations [J]. *Pattern Recognition Letters*, 2020, 138: 185-192.
- [38] REN Y, WANG N, LI M, et al. Deep density-based image clustering [J]. *Knowledge-Based Systems*, 2020, 197: 105841.
- [39] LAURENS V D M, HINTON G E. Visualizing data using t-SNE [J]. *Journal of Machine Learning Research*, 2008, 9(11): 2579-2605.
- [40] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks [J]. *Science*, 2014, 344(6191): 1492.
- [41] CHEN D, LV J C, YI Z. Unsupervised multi-manifold clustering by learning deep representation [C] // *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2017: 385-391.
- [42] MRABAH N, KHAN N M, KSANTINI R, et al. Deep clustering with a dynamic autoencoder: from reconstruction towards centroids construction [J]. *Neural Networks*, 2020, 130: 206-228.
- [43] BERTHELOT D, RAFFEL C, ROY A, et al. Understanding and improving interpolation in autoencoders via an adversarial regularizer [C] // *International Conference on Learning Representations*. Piscataway: IEEE, 2019: 272-284.
- [44] PENG X, FENG J, ZHOU J T, et al. Deep subspace clustering [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 99: 1-13.
- [45] JI P, ZHANG T, LI H, et al. Deep subspace clustering networks [C] // *Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2017: 24-33.
- [46] ZHOU P, HOU Y, FENG J. Deep adversarial subspace clustering [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2018: 1596-1604.
- [47] ZHANG T, JI P, HARANDI M, et al. Neural collaborative subspace clustering [C] // *International Conference on Machine Learning*. New York: ACM, 2019: 7384-7393.

- [48] DIZAJI K G, HERANDI A, HUANG H. Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 5736-5745.
- [49] YANG X, DENG C, ZHENG F, et al. Deep spectral clustering using dual autoencoder network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 4066-4075.
- [50] DILOKTHANAKUL N, MEDIANO P A M, GARNELO M, et al. Deep unsupervised clustering with gaussian mixture variational autoencoders [J]. ArXiv Preprint ArXiv, 2016: 1611.02648.
- [51] YANG L, CHEUNG N M, LI J, et al. Deep clustering by gaussian mixture variational autoencoders with graph embedding [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2019: 6440-6449.
- [52] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84-90.
- [53] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint ArXiv, 2014: 1409. 1556.
- [54] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1-9.
- [55] HE K M, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [56] LI F, QIAO H, ZHANG B. Discriminatively boosted image clustering with fully convolutional auto-encoders [J]. Pattern Recognition, 2018, 83: 161-173.
- [57] 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. San Francisco: Margan Kaufmann, 2017: 1753-1759.
- [58] TULLIS J G, BENJAMIN A S. On the effectiveness of self-paced learning [J]. Journal of Memory & Language, 2011, 64 (2): 109-118.
- [59] CHICCO D. Siamese neural networks: an overview [M] // CARTWRIGHT D H. Artificial neural networks. Berlin: Springer, 2021.
- [60] PENG X, ZHU H Y, FENG J S, et al. Deep clustering with sample-assignment invariance prior [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 99: 1-12.
- [61] SAXE A M, BANSAL Y, DAPELLO J, et al. On the information bottleneck theory of deep learning [J]. Journal of Statistical Mechanics: Theory and Experiment, 2019, 12: 124020.
- [62] NINA O, MOODY J, MILLIGAN C. A decoder-free approach for unsupervised clustering and manifold learning with random triplet mining [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2019: 207-218.
- [63] WU J L, LONG K Y, WANG F, et al. Deep comprehensive correlation mining for image clustering [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2019: 8150-8159.
- [64] HUANG J B, GONG S G, ZHU X T. Deep semantic clustering by partition confidence maximization [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 8849-8858.
- [65] XU L L, NEUFELD J, LARSON B, et al. Maximum margin clustering [C] // Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004: 1537-1544.
- [66] LI Y F, HU P, LIU Z T. Contrastive clustering [J]. ArXiv Preprint ArXiv, 2020: 2009. 09687.
- [67] MAKHZANI A, SHLENS J, JAITLEY N, et al. Adversarial autoencoders [C] // International Conference on Learning Representations. Piscataway: IEEE, 2016: 445-461.
- [68] HARCHAOUI W, MATTEI P A, BOUYEYRON C. Deep adversarial gaussian mixture auto-encoder for clustering [C] // International Conference on Learning Representations. Piscataway: IEEE, 2017: 1-5.
- [69] SPRINGENBERG J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks [J]. ArXiv Preprint ArXiv, 2015: 1511. 06390.
- [70] CHEN X, DUAN Y, HOUTHOOFT R, et al. Infogan: interpretable representation learning by information maximizing generative adversarial nets [C] // Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2016: 2172-2180.
- [71] MUKHERJEE S, ASNANI H, LIN E, et al. ClusterGAN: latent space clustering in generative adversarial networks [C] // Proceedings of the AAAI

- Conference on Artificial Intelligence. Menlo Park: AAAI, 2019: 4610-4617.
- [72] BELGHAZI M I, BARATIN A, RAJESHWAR S, et al. Mutual information neural estimation [C] // International Conference on Machine Learning. New York: ACM, 2018: 531-540.
- [73] TSCHANNEN M, DJOLONGA J, RUBENSTEIN P K. On mutual information maximization for representation learning [J]. ArXiv Preprint ArXiv, 2019: 1907.13625.
- [74] CHANG J L, MENG G F, WANG L F, et al. Deep self-evolution clustering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(4): 809-823.
- [75] CHANG J L, GUO Y W, WANG L F, et al. Deep discriminative clustering analysis [J]. ArXiv Preprint ArXiv, 2019: 1905.01681.
- [76] ZHAO J J, LU D H, MA K, et al. Deep image clustering with category-style representation [C] // European Conference on Computer Vision. Berlin: Springer, 2020: 54-70.
- [77] TAPASWI M, LAW M T, FIDLER S. Video face clustering with unknown number of clusters [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 5027-5036.
- [78] WANG M, DENG W H. Deep face recognition with clustering based domain adaptation [J]. Neurocomputing, 2020, 393: 1-14.
- [79] LIN W A, CHEN J C, CASTILLO C D, et al. Deep density clustering of unconstrained faces [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8128-8137.
- [80] KHAN Z, YANG J. Bottom-up unsupervised image segmentation using FC-Dense u-net based deep representation clustering and multidimensional feature fusion based region merging [J]. Image and Vision Computing, 2020, 94: 103871.
- [81] ZHOU L, WEI W Y F. DIC: deep image clustering for unsupervised image segmentation [J]. IEEE Access, 2020, 8: 34481-34491.
- [82] SAHA S, SUDHAKARAN S, BANERJEE B, et al. Semantic guided deep unsupervised image segmentation [C] // International Conference on Image Analysis and Processing. Berlin: Springer, 2019: 499-510.

(责任编辑 张 蕾)

(上接第 823 页)

- [16] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The kitti vision benchmark suite [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 3354-3361.
- [17] GEIGER A, ZIEGLER J, STILLER C. Stereoscan: dense 3d reconstruction in real-time [C] // 2011 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE, 2011: 963-968.
- [18] CIARFUGLIA T, COSTANTE G, VALIGI P, et al. Evaluation of non-geometric methods for visual odometry [J]. Robotics and Autonomous Systems, 2014, 62(12): 1717-1730.
- [19] ZHOU T, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1851-1858.
- [20] ZHAN H G, GARG R, CHAMARA S, et al. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 340-349.
- [21] RANJAN J, JAMPANI V, BALLE S, et al. Competitive collaboration: joint unsupervised learning of depth, camera motion, optical flow and motion segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12240-12249.
- [22] BIAN J W, LI Z C, WANG N Y, et al. Unsupervised scale-consistent depth and ego-motion learning from monocular video [J]. ArXiv Preprint ArXiv, 2019: 1908.10553.

(责任编辑 张 蕾)