

基于重构误差的深度聚类方法

邓祥¹, 俞璐¹, 谢钧¹, 吕昊远¹, 姚昌华²

(1. 陆军工程大学, 江苏南京 210001;

2. 南京信息工程大学, 江苏南京 210044)

摘要:聚类是机器学习的核心任务之一,通常是在无标签条件下,依靠发掘数据潜在的结构进行聚类。近年来,数据的复杂度越来越高,数据隐空间中存在各种冗余复杂的空间结构,传统聚类算法难以从中分离出不同簇的数据。深度学习具有强大的特征表示和非线性逼近能力,在无监督聚类领域也显现出优越性,基于深度学习的聚类模型有效提高了各类复杂数据的聚类结果。该文提出了一种新的端到端深度聚类模型,在自编码器框架下,构建多个不同的聚类子空间,并利用高维样本在多个子空间的低维特征重构原始样本,同时增加一个对样本进行簇预测的网络,利用预测的概率向量对不同簇的解码样本进行加权融合,通过最小化融合样本与原始样本之间的重构误差并对子空间加以约束,最终实现对高维样本的聚类。模型同时兼顾聚类簇的子空间结构和不同簇之间的重构误差,在标准数据集上取得了较好的聚类效果。

关键词:聚类;深度聚类;深度学习;自编码器;模式识别

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2022)11-0030-07

doi:10.3969/j.issn.1673-629X.2022.11.005

Deep Clustering Method Based on Reconstruction Error

DENG Xiang¹, YU Lu¹, XIE Jun¹, LYU Hao-yuan¹, YAO Chang-hua²

(1. Army Engineering University of PLA, Nanjing 210001, China;

2. Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Clustering, usually without label knowledge, is one of the core tasks of machine learning, which divides data by discovering the potential structure of data. In recent years, the complexity of data is getting higher and higher, and there are various redundant and complex spatial structures in the data hidden space. It is difficult for traditional clustering algorithms to separate different clusters of data. Deep learning has strong feature representation and nonlinear approximation ability, and also shows superiority in the field of unsupervised clustering. The clustering model based on deep learning effectively improves the clustering results of different kinds of complex data. We propose an end-to-end deep clustering model. Under autoencoder framework, the model constructs multiple different clustering subspaces, and reconstructs the original samples by using the low-dimensional features of high-dimensional samples in multiple subspaces. At the same time, the model adds a cluster prediction network, uses the predicted probability vector to weighted fuse the decoded samples of different clusters. By minimizing the reconstruction error between the fused samples and the original samples and constraining the subspace, the model can cluster high-dimensional samples. The model takes the subspace structure of clusters and the reconstruction error between different clusters into account, and achieves ideal clustering results on standard data sets.

Key words: clustering; deep clustering; deep learning; autoencoder; pattern recognition

0 引言

聚类本质上是一个物以类聚的过程,通常在无监督条件下,依靠发掘数据潜在的结构、性质和规律实现聚类目的。现有的聚类模型以是否使用深度学习可以划分为传统聚类和深度聚类两类。传统聚类可以划分为原型聚类、密度聚类、图聚类、层次聚类等。原型聚

类算法用原型作为每个簇的代表,通过度量样本与原型之间的距离确定样本的归属,典型的算法有 k-means^[1]、高斯混合模型(GMM)^[2]等。密度聚类的主要目标是寻找被低密度区域分离的高密度区域,典型的算法是 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)^[3]。从图论中演化

收稿日期:2021-09-09

修回日期:2022-01-11

基金项目:国家自然科学基金面上项目(61971439);江苏省自然科学基金面上项目(BK20191329);中国博士后科学基金项目(2019T120987)

作者简介:邓祥(1994-),男,硕士研究生,通讯作者,研究方向为模式识别与深度学习;俞璐,博士,副教授,研究方向为机器学习、多媒体信息处理、图像处理等。

而来的图聚类模型通过构建一个基于相似度的无向加权图把聚类问题转化为图的最优划分问题,典型的算法是 spectral clustering (SC)^[4]。层次聚类通过度量数据节点之间的相似性或者距离构建一个树形结构在不同层面上对数据集进行划分,典型的算法是 AGglomerative NESTing (AGNES)。传统聚类算法从不同的角度出发挖掘数据簇的聚类规律,针对这个规律设计明确的聚类过程实现聚类簇的划分。传统聚类算法独立地进行特征提取与数据聚类,聚类算法的输入特征完全依赖于特征预处理,而多数特征提取算法是非目标导向算法,提取的特征未必是利于聚类的特征,这些特征提取算法包括线性的 PCA^[5]、非线性的核方法(kernel methods)^[6]等。近些年深度学习因其强大的非线性拟合能力和特征表示能力可以很好地缓解传统聚类算法的缺陷,将深度学习引入无监督聚类领域形成的深度聚类模型已经显示出了优越性。在深度聚类领域尚在探索阶段的今天,该文提出一种全新的深度聚类模型,该模型从不同的角度出发探索深度聚类的可行性方案。

1 相关工作

用于聚类的深度模型有多种选择,较为常用的包括自编码器^[7]、图神经网络和生成模型等。近年来,图卷积神经网络提取特征考虑到数据的结构,能够提取利于聚类的特征,故用图卷积神经网络完成聚类任务成为新兴的研究方向之一,基于图神经网络的深度聚类模型中较成功的有 Embedding Graph Auto-Encoder (EGAE)^[8]、Deep Attentional Embedded Graph Clustering (DAEGC)^[9]、Structural Deep Clustering Network (SDCN)^[10]、Adaptive Graph Convolution (AGC)^[11]、Adversarial Graph Auto-Encoders (AGAE)^[12]。其中 AGAE 的输入是相似性矩阵,利用图卷积神经网络处理相似性矩阵提取特征,并把它与特征先验分布中得到的特征一同进行对抗性训练,得到一个能完成聚类的概率分布。生成模型因其能够使用神经网络捕获数据分布并产生真实的未见样本而受到人们广泛关注,其中 GAN^[13]和 VAE^[14]是最成功的生成模型。将生成模型用于聚类任务是深度聚类的研究方向之一。基于生成模型的聚类分为:基于变分自编码器(VAE)的聚类和基于生成对抗模型(GAN)的聚类。变分自编码器假设样本服从高斯分布,通过变分推断法求出样本后验概率。将变分自编码器引入聚类中成功的模型有 Variational Deep Embedding (VaDE)^[15]、Gaussian Mixture Variational AutoEncoder (GMVAE)^[16]。VaDE 在变分自编码器的基础上加上 Mixture-of-Gaussians,使用变分推断法推断出样本所

属簇的后验概率,进而对数据样本进行簇推断。基于生成对抗网络的深度聚类模型借助 GAN 网络捕获数据分布的能力寻找样本的后验概率。这方面突出的模型有 ClusteringGAN^[17]、Categorical Generative Adversarial Networks (CatGAN)^[18]、InfoGAN^[19]等。InfoGAN 利用变分法推断出生成器输入与判别器输出之间的互信息,用于 GAN 的对抗性训练中。当训练完成后,判别器可以输出输入样本属于各簇的概率,从而完成聚类任务。

自编码器重构样本时可以产生表征数据本质的非线性映射特征,基于自编码器模型的聚类方案有 K-means - friendly Spaces (DCN)^[20]、Adversarial AutoEncoders (AAE)^[21]、Deep Embedded Clustering (DEC)^[22]等,其中 DEC 模型在自编码器的基础上提出使用神经网络同时学习特征表示和聚类分配。它在潜在的特征空间中设计了一个软分布和一个辅助概率分布,通过让两个分布的 KL 散度最小化实现聚类目的。DCN 模型通过联合优化特征空间和 k-means 算法,同步完成了聚类任务和寻找适合 k-means 聚类的特征子空间。文献[23]提出的模型在 DSC-NET 的基础上引入二元分类模块用于学习更抽象更具有区分度的特征,它同时与 DSC-NET 的自表示层进行协同训练达到聚类目的。针对无监督聚类领域经常出现的错误预测和过置信现象,文献[24]借鉴半监督聚类的思想,提出了基于鲁棒学习的改进无监督图像聚类模型 (Robust learning for Unsupervised Clustering, RUC),有效地解决了上述问题。

目前基于自编码器的深度聚类模型更关注自编码器的数据压缩能力,并基于中间特征(编码器输出)设计聚类的损失函数。该文关注自编码器的重构能力,提出了基于重构误差(解码器输出)设计损失函数,并对中间特征加以约束的端到端深度聚类方案,对自编码器应用于聚类任务做出了新的探索。

受生成模型和自编码器的启发,该文提出一种新的基于自编码器的深度聚类模型:基于重构误差的深度聚类方法 (Deep Clustering Method Based on Reconstruction Error, DCMBORE)。该模型在自编码器框架下,通过构建 K 个不同的聚类子空间,根据样本在各子空间的映射获得 K 个生成样本,并利用 K 个生成样本的概率加权和样本重构原始样本,通过最小化重构误差同时对子空间加以约束,最终实现对样本的聚类。主要贡献如下:

- 探索了以重构误差为聚类目标的深度聚类方案。
- 该模型能对同一簇下不同的小簇 (MiniCluster) 进行区分。

2 基于重构误差的深度聚类方法

2.1 重构差异的思想

聚类的目标是通过对无标记样本的学习来揭示和挖掘数据内在的结构、性质和规律。对于数据内蕴规律不同的先验和假设将导出不同的聚类模型。假设高维空间数据样本呈现出 K 个不同的簇,其本质原因在于数据采样于高维空间中的 K 个低维子空间,分别对应 K 个簇。高维样本空间和低维子空间之间的非线性映射揭示了数据的内在规律和本质,学习这些映射就是聚类的关键。该文提出一种端到端的深度聚类模型,借助神经网络强大的非线性逼近能力,通过学习高维样本空间和 K 个低维子空间之间的映射完成聚类任务。

借鉴自编码器的思想,采用重构误差最小化作为损失函数。如果已学习到 K 个低维子空间和高维空间样本之间的映射,那么将第 i 簇的样本 y 映射到 K 个子空间分别得到 K 个低维特征,显然由第 i 个特征重构的样本 \hat{y}_i 将更接近样本 y ,则可以通过比较 K 个重构误差确定样本 y 的归属。由于没有监督信息,训练过程中无法获知样本 y 的归属,该文提出用 K 个重构样本的概率加权和 $\hat{y} = \sum_{i=1}^K z_i \hat{y}_i$ (融合样本)重构原始样本 y 。在学习映射关系的同时学习概率分布 z 。重构误差最小化将引导概率分布 z 向 one-hot 向量逼近,并获得高维空间与子空间的映射关系。

2.2 子空间约束

为保证聚类效果,希望不同簇的子空间尽可能远离,同时每个簇的子空间尽可能收缩在一定范围内。假定有 K 个簇, μ_i 是簇 i 对应的子空间内的特征中心, x_{ni} 是样本 n 映射到子空间 i 的低维特征。 S_w 表示子空间内特征的散度, S_b 表示子空间之间的距离。该文对子空间做如下约束:

$$S_w = \sum_{n=1}^N \sum_{i=1}^K [\|x_{ni} - \mu_i\|_2 - d_1]_+ \quad (1)$$

$$S_b = \sum_{\substack{i \neq j \\ i, j=1}}^K d_2 - [\|\mu_i - \mu_j\|_2]_+ \quad (2)$$

其中, d_1 、 d_2 是超参数, $[x]_+ = \max(0, x)$ 表示正部函数。最小化 S_w 将使每个子空间内的低维特征都收缩在半径为 d_1 的超球内,以此保证同一簇内高维空间样本具有一定程度的多样性。最小化 S_b 将使任意两个子空间的特征中心的距离超过 d_2 ,以此保证不同簇子空间的分离性。

2.3 基于重构误差的深度聚类方法

该模型将一个样本 y_n 通过神经网络分别映射到 K 个子空间中,得到 K 个簇的低维特征 $\{x_{n1}, x_{n2}, \dots, x_{nk}\}$,再分别用 K 个低维特征重构出 K 个新样本 $\{\hat{y}_{n1},$

$\hat{y}_{n2}, \dots, \hat{y}_{nk}\}$ 。由于训练过程中无法获知样本 y_n 的归属,该文提出把 K 个重构样本按离散概率分布 $z_{ni} = [z_{n1}, z_{n2}, \dots, z_{nk}]$ 进行概率加权得到一个融合样本,用融合样本 \hat{y}_n 重构原始样本 y_n 最小化重构误差 R ,完整的损失函数如公式 3,最小化该损失函数在训练的过程中会将不同类别的样本聚类到不同簇中。

$$\begin{aligned} L = R + S_w + S_b = & \sum_{n=1}^N \left\| \sum_{i=1}^K \hat{y}_{ni} * z_{ni}^\alpha - y_n \right\|_2 + \\ & \sum_{n=1}^N \sum_{i=1}^K [\|x_{ni} - \mu_i\|_2 - d_1]_+ + \\ & \sum_{\substack{i \neq j \\ i, j=1}}^K [d_2 - \|\mu_i - \mu_j\|_2]_+ \\ \text{s. t. } & \sum_{i=1}^K z_{ni} = 1, z_{ni} \in [0, 1] \quad n = \{1, 2, \dots, N\} \\ & \mu_i = \frac{1}{N} \sum_{n=1}^N x_{ni} \\ & \alpha \geq 1 \end{aligned} \quad (3)$$

其中, N 代表数据样本个数, K 代表聚类簇个数, y_n 代表第 n 个样本, x_{ni} 代表第 n 个样本在第 i 个簇的子空间映射特征, \hat{y}_{ni} 代表 x_{ni} 子空间特征的重构样本, z_{ni} 代表样本 y_n 属于第 i 簇的概率,超参数 α 控制着融合样本与原始样本的重构速度。

在模型中引入加权指数用来控制融合样本与原始样本的重构速度,增加聚类模型的灵活性和自由度。

3 实验结果与讨论

3.1 数据集

(1) MNIST 数据集^[25]包含了 70 000 张手写数字的图像,其中 60 000 个训练样本,10 000 个测试样本,该图像的像素大小为 28×28 。该数据集包含了 10 个大类别(0~9),几百个小类别(每个数字不同形态的写法)。

(2) USPS 数据集与 MNIST 类似,包含了 10 个类别的手写数字,训练集共 7 291 个样本,测试集共 2 007 个样本,每个样本都是 16×16 的灰度图像。

(3) Fashion-MNIST 数据集是 10 类不同风格衣服的数据集,该数据集包含 60 000 个训练样本,10 000 个测试样本,每个样本都是 28×28 的灰度图像。

3.2 评估指标

选用互信息(NMI)^[26]和 Homogeneity^[27]指数来评估不同聚类簇 K 下的聚类性能。其中 NMI 能够评估一个变量包含另一个变量的信息量。NMI 取值范围是 $[0, 1]$,越接近 1 说明变量 Y 包含变量 X 的信息量越小,变量 X 确定性越大。

$$\text{NMI}(X; Y) = \frac{2I(X; Y)}{H(X) + H(Y)} \quad (4)$$

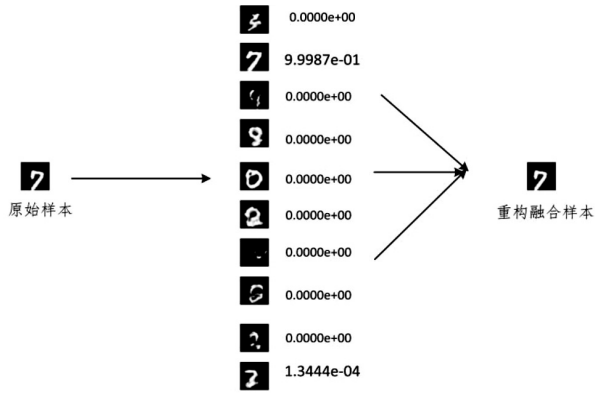
Homogeneity 指数用来评估聚类簇的同质性,即一个聚类簇中的样本是否只属于某一个单独类。Homogeneity 的取值范围为 $[0,1]$,其值越接近1则表明聚类簇中同类样本越多。假设真实样本标签为 $C = \{C_i | i = 1, 2, \dots, n\}$,预测簇标签为 $K = \{K_j | j = 1, 2, \dots, m\}$,则 Homogeneity 计算方法如公式5。该公式表明预测簇标记 K 的条件信息熵越小, K 的不确定性越小, Homogeneity 的数值越接近于1,预测簇 K 的同质性越强。

$$h = \begin{cases} 1 & \text{if } H(C, K) = 0 \\ 1 - \frac{H(C|K)}{H(C)} & \text{else} \end{cases} \quad (5)$$

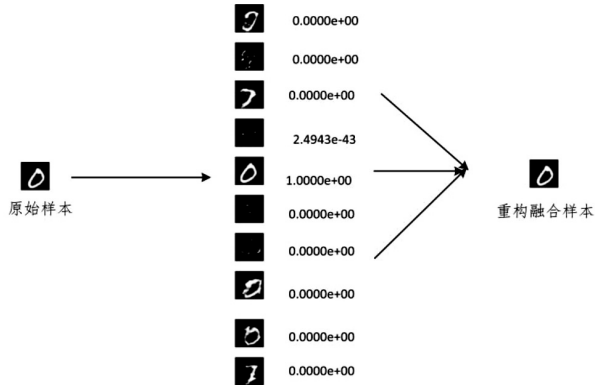
由于该模型能够聚类不同的模态,故在选择 NMI 指标的同时选择了 Homogeneity 指标来评估模型的细致划分能力。

3.3 假设验证

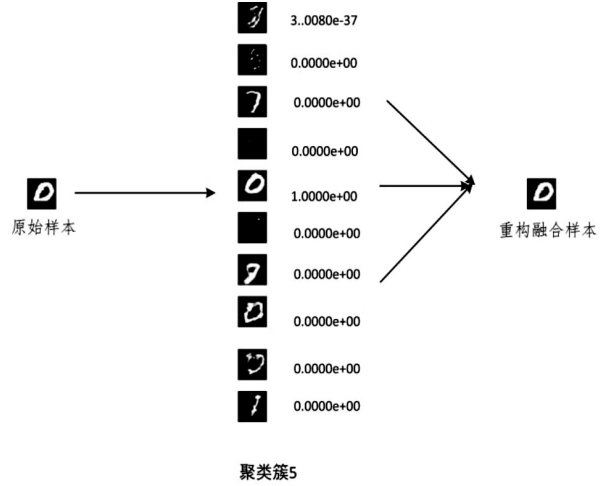
如前所述,该文的主要工作基于如下假设:原始样本在 K 个子空间的映射特征上所生成的 K 个样本中只有对应簇的生成样本接近给定的原始样本,且这 K 个生成样本在概率向量的加权融合下能够逼近原始样本。为说明假设的合理性,进行了实验验证,结果如图1所示。



(a) 簇2中样本7的验证图



(b) 簇5中样本0的验证图



(c) 簇5中不同风格的样本0的验证图

图1 验证图

图1在超参数 $\alpha = 2$, $K = 10$, $d_1 = 5$ 的模型上展示了三种不同的手写数字的融合过程。从图1(a)可以看出原始样本0在10个子空间上的投影特征所生成的样本中只有簇5对应的样本与原始样本接近,且原始样本对应的概率推断向量也逼近 one-hot 向量,其结果基本符合预定的假设。

3.4 模型实现

该文对神经网络的设计分为共享层网络F、子空间映射网络P、概率推断网络Q、生成网络G。网络结构如图2所示。采用图像数据集,故神经网络选用卷积神经网络和反卷积神经网络。网络F设计了3层卷积网络;概率推断网络Q设计了两层卷积网络加一个 Softmax 层,其输入来源于共享层网络F(或者单独设计一个网络Q,独立于共享层网络F,其输入为样本 y); K 个映射网络P与概率推断网络Q并列,用于提取具有区分度的特征,其输入来自于共享层网络F,每个映射网络P都设计两层卷积神经网络;生成网络G用于生成对应簇的样本,其输入特征来自映射网络P,该网络设计了四层反卷积神经网络加 Sigmoid 函数用于输出灰度图像。优化函数选着带动量的随机梯度下降法 Adam^[28],学习率 $lr = 0.001$,迭代周期 $epoch = 200$ 。该模型实际聚类结果随聚类簇个数 K ,类内距离 d_1 的变化而呈现不同的聚类现象。模型中超参数 d_2 控制着不同簇低维子空间的分离性, d_2 的取值只要能够保持子空间相互分离即可,在此条件下, d_2 取值的差异对聚类结果影响不显著。模型中包含如下超参数 K 、 d_1 、 d_2 、 α 。

3.5 实验结果

从生成网络G的生成图(图5)和聚类评估指标 Homogeneity、NMI 的数值变化(表1~表5)来看(实验结果取自多次实验的最优值),提出的模型具有把数据集进行簇划分的能力。

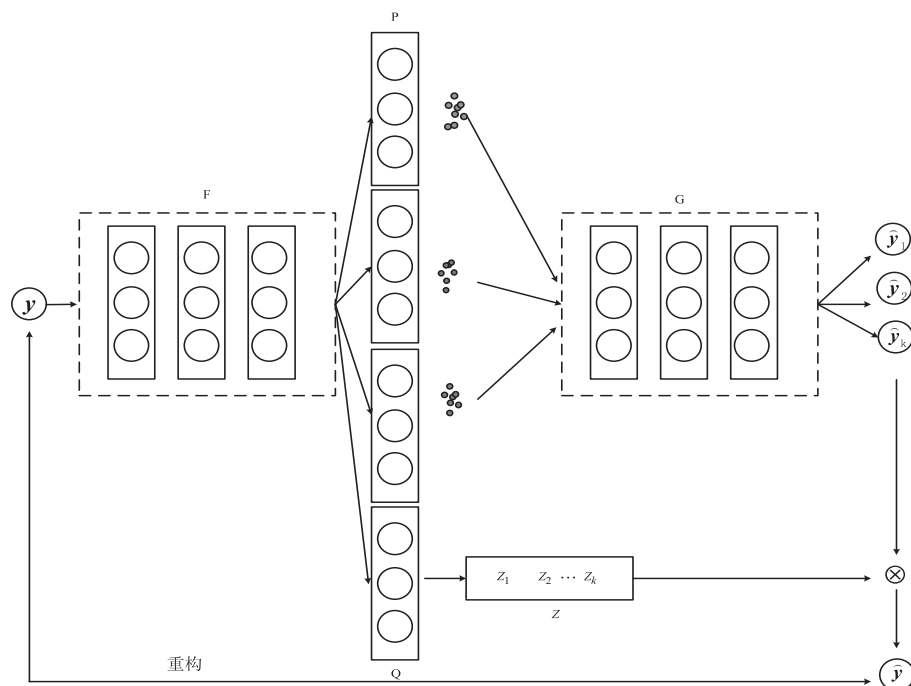


图 2 网络结构

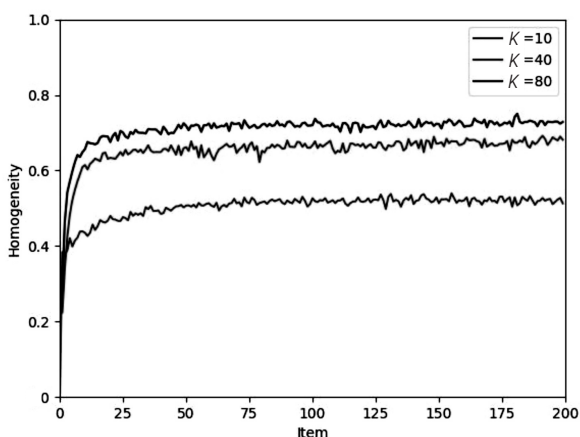


图 3 在超参数 $\alpha = 1, d_1 = 15, d_2 = 200$ 时
MNIST 数据集上不同 K 下模型的
Homogeneity 指标变化情况

表 1 DCMBORE 在超参数 $\alpha = 1$ 时
MNIST 数据集上的聚类指标

means	K	NMI	Homogeneity
DCMBORE ($d_1 = 15$)	10	0.611	0.609 8
	40	0.597	0.731 8
	80	0.551	0.762 1
DCMBORE ($d_1 = 0$)	10	0.541	0.533 1
	40	0.546	0.700 2
	80	0.535	0.750 1

在表 1 中可以看出, d_1 的变化对 NMI 指标影响不大, 排除 d_1 对 NMI 的影响之后, 对比表 1 和表 2, 不难看出随着 α 变大, 模型的 NMI 指标上升。这是因为超参数 α 变大, 融合样本与原始样本之间的重构速度变

快, 模型对样本间差异的区分力度变小, 因此模型将学得不同数字之间的差异, 而忽略同一数字下不同写法之间的差异。故评估聚类结果相似度的 NMI 指标上升, 而评估聚类簇样本同质性的 Homogeneity 指标略有下降。

表 2 DCMBORE 在超参数 $\alpha = 2$ 时
MNIST 数据集上的聚类指标

means	K	NMI	Homogeneity
DCMBORE ($d_1 = 5$)	10	0.73	0.67
	15	0.734	0.67
	30	0.801	0.69

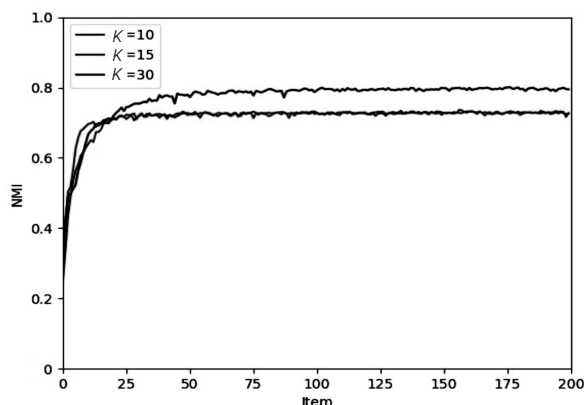


图 4 在超参数 $\alpha = 2, d_1 = 5, d_2 = 200$ 时 MNIST 数据
集上不同 K 下模型的 NMI 指标变化情况

超参数 α 控制着样本间差异的区分力度, 观察在 $\alpha = 1$ 时表 1 的 Homogeneity 指标和网络 G 的生成图 (图 5), 可以看出模型对样本差异的区分力度变大, 不同簇之间学得的差异是同一数字下不同写法之间的差

异,此时模型得到的每一簇分别表示同一数字不同形态的写法(如图 5(n)和(o)所示),可视为 0~9 十类别下的不同小簇(MiniCluster)。所以聚类簇数目 K 越大,模型对样本的划分越细致。

在超参数 $\alpha = 1$ 固定时,观察表 1 在 d_1 取值较大时 Homogeneity 指标略有提升,这是因为 d_1 取值较大,不同簇的低维子空间变大,对应的高维生成空间样本多样性变大,可以包含更多变化的同质性样本。因此超参数 d_1 控制着高维生成空间的样本多样性。

该模型聚类效果对参数 α , K , d_1 的变化较敏感。当 K 的数值越大, α 、 d_1 数值越小时,聚类划分越细致。同时需要注意,在加权指数 $\alpha = 1$ 且 d_1 和 K 都变小时,高维样本多样性减少,类别划分得细致,而 K 的数目较小,则现有的聚类簇数目难以包含所有类别,所以会出现难以区分不同类别的样本(如图 5(d)和(e)所示)。此外在 K 值较小时,实验结果受随机性影响较大。

就实验结果而言,相较于传统聚类算法,该模型的

聚类效果有显著的提高。相比较于最先进的深度聚类算法,该模型具有可调节的簇区分能力,功能更灵活。模型的参数规模随着聚类簇的增加而变大,因此,该模型的时间复杂度较高。同时由于该模型是在自编码器的框架下设计的聚类算法,因此该模型的泛化能力较弱。

表 3 DCMBORE 在 $\alpha = 1.6$ 、 $d_1 = 13$ 、 $d_2 = 200$ 、 $K = 15$ 时 USPS 数据集上的 Homogeneity、NMI 指标

dataset	NMI	Homogeneity
USPS	0.77	0.774

表 4 DCMBORE 在 $\alpha = 1.4$ 、 $d_1 = 10$ 、 $d_2 = 200$ 、 $K = 10$ 时 Fashion-MNIST 数据集上的 Homogeneity、NMI 指标

dataset	NMI	Homogeneity
Fashion-MNIST	0.590	0.53

表 5 DCMBORE 与不同聚类模型在给定数据集上的 NMI 聚类指标

methods	DCMBORE	Traditional Clustering			Deep Clustering				
		K-means	Fuzzy-c-means	NMF	DCN	DEC	InfoGAN	AEC	IEC
MNIST	0.801	0.49	0.51	0.45	0.63	0.80	0.86	0.669	0.542
Fashion-MNIST	0.590	0.512 4	0.513 4	0.51	0.594	0.601	0.59	-	-
USPS	0.77	0.531	0.56	0.53	-	-	-	0.651	0.641

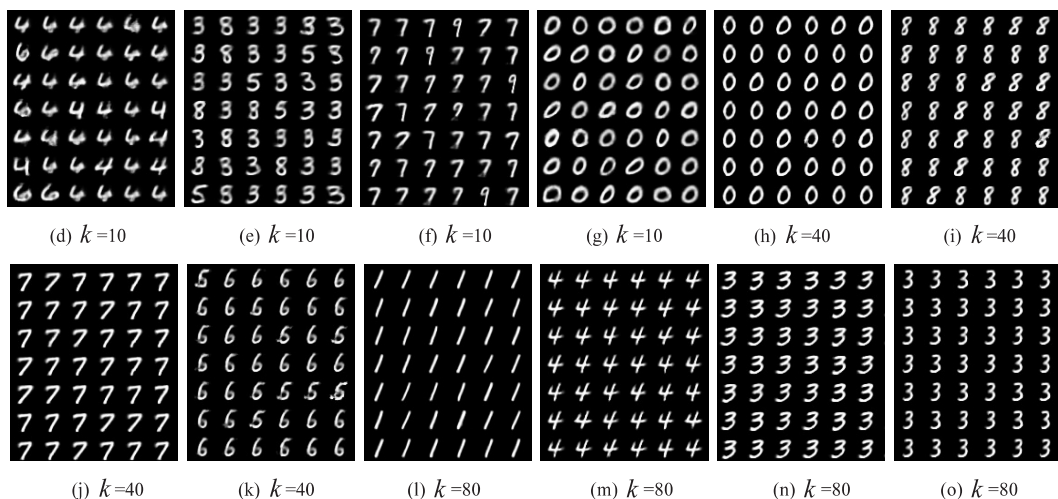


图 5 聚类生成图($\alpha = 1, d_1 = 0, d_2 = 200$)

(图中每一个图片代表一个子空间簇的生成样本,可以看出随着 K 的增加,对数据集的划分也将越来越细致)

4 结束语

该文以一种新的方式将自编码器引入聚类领域,借助自编码器的特征表示和样本重构能力,通过设置基于重构误差的聚类损失项,并加以子空间约束项,实

现有效聚类。模型可通过调节聚类簇个数,得到不同粒度的聚类结果,以满足数据分析的不同需要。虽然该模型的聚类性能相比前沿的深度聚类模型优势并不明显,但却对自编码器用于深度聚类任务做出了一个全新的探索。

参考文献:

- [1] HARTIGAN J A, WONG M A. Algorithm AS 136: a k-means clustering algorithm[J]. Journal of the Royal Statistical Society; Series C (Applied Statistics), 1979, 28(1): 100–108.
- [2] REYNOLDS D A. Gaussian mixture models[J]. Encyclopedia of Biometrics, 2009, 741: 659–663.
- [3] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of the 2nd international conference on knowledge discovery and data mining. Portland: AAAI Press, 1996: 226–231.
- [4] LUXBURG U V. A tutorial on spectral clustering[J]. Statistics and Computing, 2004, 17(4): 395–416.
- [5] WOLD S, ESBENSEN K, GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1–3): 37–52.
- [6] HOFMANN T, SCHÖLKOPF B, SMOLA A J. Kernel methods in machine learning[J]. The Annals of Statistics, 2008, 36(3): 1171–1220.
- [7] NG A. Sparse autoencoder [J]. CS294A Lecture Notes, 2011, 72: 1–19.
- [8] ZHANG H, ZHANG R, LI X. Embedding graph auto-encoder for graph clustering[J]. arXiv: 2002.08643v2, 2020.
- [9] WANG C, PAN S, HU R, et al. Attributed graph clustering: a deep attentional embedding approach [J]. arXiv: 1906.06532, 2019.
- [10] BO D, WANG X, SHI C, et al. Structural deep clustering network[C]//Proceedings of the web conference 2020. New York: Association for Computing Machinery, 2020: 1400–1410.
- [11] ZHANG X, LIU H, LI Q, et al. Attributed graph clustering via adaptive graph convolution [J]. arXiv: 1906.01210, 2019.
- [12] TAO Z, LIU H, LI J, et al. Adversarial graph embedding for ensemble clustering[C]//Twenty-eighth international joint conference on artificial intelligence. Macau, China: AAAI Press, 2019: 3562–3568.
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 27: 2672–2680.
- [14] KINGMA D P, WELING M. Auto-encoding variational bayes[J]. arXiv: 1312.6114, 2013.
- [15] 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. Melbourne: AAAI Press, 2017: 1965–1972.
- [16] DILOKTHANAKUL N, MEDIANO P A M, GARNELO M, et al. Deep unsupervised clustering with Gaussian mixture variational autoencoders[J]. arXiv: 1611.02648, 2016.
- [17] 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. [s. l.]: AAAI Press, 2019: 4610–4617.
- [18] SPRINGENBERG J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks [J]. arXiv: 1511.06390, 2015.
- [19] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[C]//Proceedings of the 30th international conference on neural information processing systems. Red Hook: Curran Associates Inc, 2016: 2180–2188.
- [20] YANG B, FU X, SIDIROPOULOS N D, et al. Towards k-means-friendly spaces: simultaneous deep learning and clustering [C]//International conference on machine learning. Sydney: JMLR. org, 2017: 3861–3870.
- [21] MAKHZANI A, SHLENS J, JAITLEY N, et al. Adversarial autoencoders[J]. arXiv: 1511.05644, 2015.
- [22] XIE J, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis[C]//International conference on machine learning. New York: JMLR. org, 2016: 478–487.
- [23] ZHANG T, JI P, HARANDI M, et al. Neural collaborative subspace clustering [C]//International conference on machine learning. Long Beach: JMLR. org, 2019: 7384–7393.
- [24] PARK S, HAN S, KIM S, et al. Improving unsupervised image clustering with robust learning [C]//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Kuala Lumpur: IEEE, 2021: 12278–12287.
- [25] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [26] ESTÉVEZ P A, TESMER M, PEREZ C A, et al. Normalized mutual information feature selection [J]. IEEE Transactions on Neural Networks, 2009, 20(2): 189–201.
- [27] ROSENBERG A, HIRSCHBERG J. V-measure: a conditional entropy-based external cluster evaluation measure [C]//EMNLP-CoNLL. Prague Czech Republic: Association for Computational Linguistics, 2007: 410–420.
- [28] DIEDERIK K, JIMMY B. Adam: a method for stochastic optimization[J]. arXiv: 1412.6980, 2014.