

# 结构化深度判别嵌入编码网络图像聚类

付兴武<sup>1</sup>, 吕明明<sup>1,2\*</sup>, 刘万军<sup>1</sup>, 魏宪<sup>2</sup>

<sup>1</sup> 辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105;

<sup>2</sup> 中国科学院海西研究院泉州装备制造研究所, 福建 泉州 362200

**摘要** 大部分现有深度聚类方法都试图最小化重构损失,然而深层特征的判别能力与重构损失并没有必然联系,并且这些深度聚类方法通常只关注从样本自身提取的有用特征,很少考虑样本背后的结构信息。为解决这些问题,提出一种新的结构化深度判别嵌入编码网络聚类(SDDECC)算法,用于无监督图像聚类。首先,在多层卷积自编码器网络中引入最大化互信息与最小化先验分布约束,然后使用传递算子将深度判别嵌入编码网络(DDECN)模块学习到的特征表示融入到图卷积神经网络(GCN)模块中,最后利用 Kullback-Leibler(K-L)散度联合双网络结构产生的潜在特征分布端到端地完成聚类训练。实验结果表明,SDDECC 算法能够有效提取更多有鉴别性的深层特征,并且由于在 GCN 中融合了样本的属性信息和结构信息,最终该模型取得了良好的聚类效果。

**关键词** 图像处理; 深度聚类; 图卷积神经网络; 无监督学习; 三元组互信息

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.0610016

## Structured Deep Discriminant Embedded Coding Network for Image Clustering

Fu Xingwu<sup>1</sup>, Lü Mingming<sup>1,2\*</sup>, Liu Wanjun<sup>1</sup>, Wei Xian<sup>2</sup>

<sup>1</sup> College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;

<sup>2</sup> Quanzhou Institute of Equipment Manufacturing Haixi Institutes, Chinese Academy of Sciences, Quanzhou, Fujian 362200, China

**Abstract** Most existing deep clustering methods are employed to minimize the reconstruction loss. However, the identification ability of potential representation is not necessarily related to the reconstruction loss. Moreover, these deep clustering methods focus only on extracting useful features from the sample itself and seldom consider the structure information behind the sample. To resolve these problems, a new structured deep discriminant embedded coding network clustering (SDDECC) algorithm is proposed for unsupervised image clustering. First, the maximum mutual information and minimum prior distribution constraints are embedded in a multilayer convolutional autoencoder network. Then, the feature representation learned by the deep discriminant error correction network (DDECN) module is integrated into a graph convolutional neural network (GCN) module by the transfer operator. Finally, Kullback-Leibler (K-L) divergence is used in combination with the potential feature distribution generated by the dual network structure and is trained end-to-end to guide the clustering learning. The experimental results show that SDDECC algorithm can effectively extract more discriminative deep features than those obtained using traditional methods. Moreover, because the attribute information of the sample itself and the structural information between the samples are integrated in the GCN, the model shows good clustering.

**Key words** image processing; deep clustering; graph convolutional neural network; unsupervised learning; triple mutual information

收稿日期: 2020-08-11; 修回日期: 2020-08-31; 录用日期: 2020-09-07

基金项目: 国家自然科学基金(61806186)

\* E-mail: 329010672@qq.com

OCIS codes 100.2000; 100.3008; 100.4996; 150.1135

## 1 引言

图像聚类是机器学习和计算机视觉领域中一项关键而又具有挑战性的任务。从技术上讲,图像聚类是在不使用任何监督信息的情况下将图像分组的过程,以便在同一簇内保留相似的图像。为了充分利用这些未标记数据并研究它们之间的相关性,许多聚类算法已经被提出来,并成功应用于各种实际应用中,如图像分割<sup>[1-2]</sup>、目标检测<sup>[3]</sup>和图像分类<sup>[4-5]</sup>。其中,传统的聚类方法,如 K 均值聚类(K-means)算法<sup>[6]</sup>、谱聚类(SC)算法<sup>[7-8]</sup>及非负矩阵因式分解聚类(NMF)算法<sup>[9]</sup>基于原始数据空间中距离的概念来捕获相似性,因此它们被认为是浅层模型。虽然浅层模型在多种场景中得到了成功的应用,但在原始数据空间中计算基于距离的度量仅适用于描述数据空间中的局部关系,对表示输入之间的潜在依赖性具有局限性,不足以发现语义相似性。

随着深度学习的蓬勃发展,许多研究者将注意力转移到深度无监督特征学习和聚类中<sup>[10-11]</sup>,由此产生了一种新的聚类策略,称为深度聚类。在处理大型、高语义和高维数据时,基于深度神经网络无监督表示学习的多层体系结构已经成为自然选择。此外,深度聚类将先验知识和聚类结合起来,可以得到最优的嵌入子空间来进行聚类。与传统的聚类方法相比,深度聚类方法能够有效地模拟输入的分布,捕捉输入的非线性特性,因此它能够很好地解决浅层模型的局限性,更适用于实际的聚类场景。深度聚类方法将聚类目标融入到深度学习的强大表示能力中,因此,学习有效的特征表示直接决定聚类质量。为使潜在表示更具有鉴别性,大部分现有的深度聚类方法都试图利用最小化重构损失。例如,Xie 等<sup>[12-13]</sup>利用聚类损失帮助自编码器学习具有高聚类内聚性的数据表示,Rousseeuw 等<sup>[14-15]</sup>使用变分自编码器学习用于聚类的更好的数据表示。

尽管深度学习在许多重要任务上取得了巨大成功,但利用深度神经网络执行聚类任务时仍然存在几个问题。首先,许多作者试图将成熟的聚类算法与深度学习结合起来,例如将网络训练与 K-means 目标结合起来<sup>[16-17]</sup>。然而,将聚类和表示学习方法简单地结合起来往往会出现退化解。其次,自编码

器被广泛应用于深度聚类时只考虑重构的特征表示,缺乏判别能力。最理想的方法是用对抗网络训练一个判别器,但是这样会进一步增加任务的难度<sup>[18]</sup>。为了学习更有鉴别性的深层特征,何果财等<sup>[19]</sup>挖掘了图像三元组蕴含的相似性信息,Hjelm 等<sup>[20]</sup>最大化特征之间的互信息。最后,大多数深度聚类在学习表示时通常只关注数据本身的特性,很少考虑数据之间的结构信息,往往这种结构能够揭示样本之间潜在的相似性,从而为学习表征提供有价值的指导。Bo 等<sup>[21]</sup>通过传递算子将堆栈式自编码器与图卷积神经网络(GCN)联系起来,并使用自监督机制来优化特征提取和聚类训练过程。虽然结构信息在数据表示学习中起着至关重要的作用,但却很少应用于深度聚类中。

针对上述问题,本文提出一种结构化深度判别嵌入编码网络图像聚类(SDDECC)算法。该算法在卷积自编码器中嵌入互信息估算网络和最小化先验分布约束,然后将深度自编码网络学习到的样本自身属性信息融入到图卷积神经网络中,实现了样本自身属性信息和结构信息的协同学习,并端到端地完成聚类任务,有效提高特征判别能力的同时保留了更多可用信息。最后,通过在三个经典图像数据集上进行实验,验证所提算法的有效性。

## 2 深度聚类相关算法

### 2.1 基于自编码器的深度聚类

基于自编码器(AE)的聚类方法依赖于联合执行表示学习和聚类两个目标函数的线性组合。联合优化过程描述为

$$L = L_{\text{res}} + \gamma L_c, \quad (1)$$

式中: $L_{\text{res}}$ 为重构损失函数; $L_c$ 为嵌入聚类损失; $\gamma$ 为超参数,是控制嵌入空间扭曲程度的系数。基于 AE 的深度聚类算法的一般网络体系结构如图 1 所示,其中  $\mathbf{X}$  为输入图像, $\hat{\mathbf{X}}$  为重构图像。

### 2.2 基于变分自编码器的深度聚类

与传统聚类方法相比,基于 AE 的深度聚类方法有了显著的改进。然而,它们是专门为聚类而设计的,并没有揭示样本真正的底层结构。此外,基于降维技术的假设通常独立于聚类技术的假设,因此没有理论保证网络能够学习到可行的表示。而变分自编码器(VAE)作为一种深度生成模型,可以被认为是 AE 的一个生成变体,它对潜在特征表示施

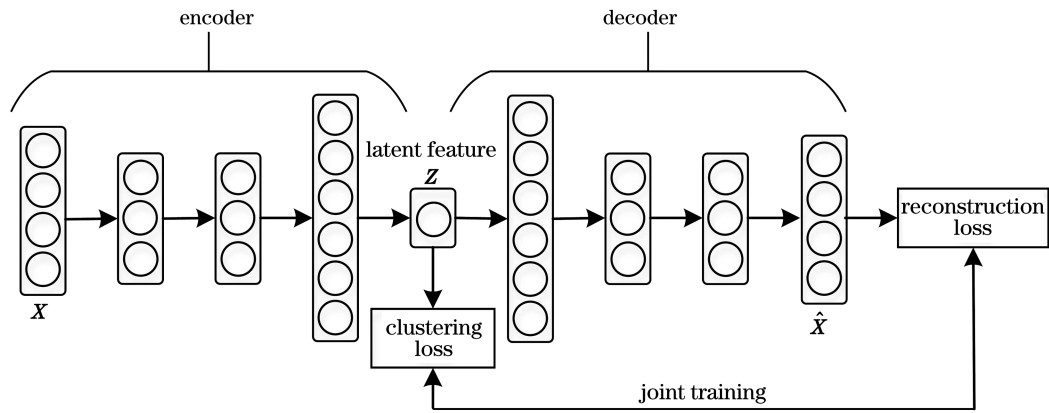


图 1 基于 AE 深度聚类算法的一般网络体系结构

Fig. 1 General network architecture of deep clustering algorithm based on AE

加一种概率先验分布,将变分贝叶斯方法与神经网络的灵活性和可扩展性结合,使用变分下界的重参数化得到一个可微下界无偏估计量。基于变分自编码器的深度聚类算法的目标函数表示为

$$L(\theta, \varphi; \mathbf{X}) = \sum_i^N \left\{ -D_{\text{KL}} [p_\theta(\mathbf{z} | \mathbf{x}_i) \| p(\mathbf{z})] + E_{p_\theta(\mathbf{z} | \mathbf{x}_i)} [\log_2 q_\varphi(\mathbf{x}_i | \mathbf{z})] \right\}, \quad (2)$$

式中:  $p(\mathbf{z})$  为整个潜在特征空间的先验分布;  $p_\theta(\mathbf{z} | \mathbf{x}_i)$  为条件后验分布;  $q_\varphi(\mathbf{x}_i | \mathbf{z})$  为似然函数;  $N$  为样本总数;  $D_{\text{KL}}(\cdot)$  为条件后验分布  $p_\theta(\mathbf{z} | \mathbf{x}_i)$  与整个潜在特征空间的先验分布  $p(\mathbf{z})$  之间的 Kullback-Leibler(K-L)散度;  $E(\cdot)$  为函数的期望。基于 VAE 的深度聚类算法的一般网络体系结构如图 2 所示,其中  $N(\mu_c, \sigma_c)$  为一种正态分布。

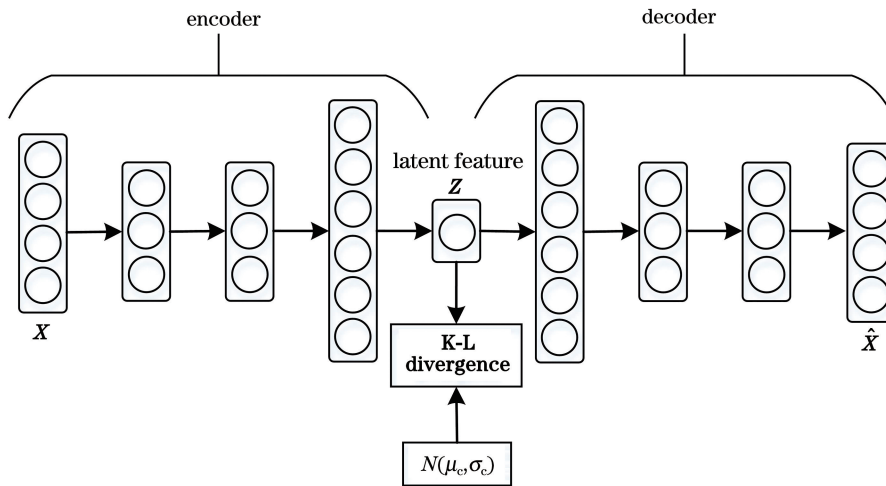


图 2 基于 VAE 深度聚类算法的一般网络体系结构

Fig. 2 General network architecture of deep clustering algorithm based on VAE

### 2.3 基于图卷积神经网络的深度聚类

这些深度聚类方法一般都只关注于从样本自身学习的数据表示,而学习表征的另一个重要信息,即数据的结构信息很少被考虑进来。为了处理数据背后的结构信息,基于 GCN<sup>[22]</sup> 的聚类方法得到了广泛的应用。其中, Kipf 等<sup>[23]</sup> 提出了图自编码器(GAE)和图变分自编码器(VGAE),它们利用图卷积作为编码器,将图结构集成到节点特征中,学习节点的嵌入。然而绝大多数基于 GCN 的聚类方法都依赖于重构邻接矩阵,这些方法只

能从图结构中学习数据表示,而忽略了数据本身的特点。

### 3 结构化深度判别嵌入编码网络聚类算法

SDDECC 算法的整体网络架构主要由两部分构成:深度判别嵌入编码网络(DDECN)模块和 GCN 模块。DDECN 模块中,在一个多层卷积自编码器中嵌入互信息估算网络和最小化先验分布约束,以学习到更多具有鉴别性的深层特征,同时

使编码空间更加规整,且有利于解耦特征。在 GCN 模块中,把结构信息融合到深度聚类中,能够同时学习样本自身的属性信息和样本之间的结构信息,使网络预测更加强大。为引入聚类信息并指导网络模型学习更有利于聚类的深层特

征,对 K-L 散度和双网络结构联合产生的潜在特征分布端到端地进行聚类训练并迭代更新网络参数。SDDECC 算法的整体网络框架如图 3 所示,其中  $\mathbf{H}$  为图卷积神经网络各层学习到的特征表示。

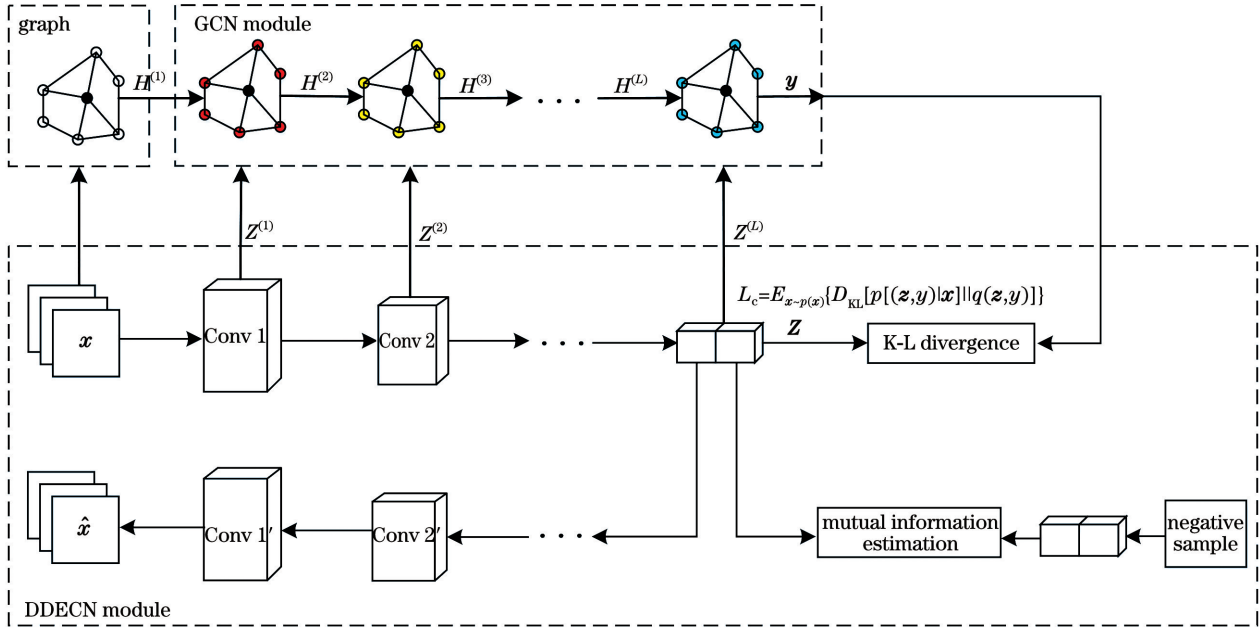


图 3 SDDECC 算法的整体网络框架

Fig. 3 Overall network framework of the SDDECC algorithm

### 3.1 深度判别嵌入编码网络模块

特征提取是无监督学习中很重要的一项任务,同时也是聚类的主要步骤。传统的自编码器是由多层感知机组成的。为了处理具有空间不变性的图像样本,采用由多个卷积层组成的卷积自编码网络,相比完全连接的自编码器,它具有更少的参数量,因此更易于训练。设原始图像的集合  $\mathbf{X} = \{x_i\}_{i=1}^N$ , 其中  $x_i$  表示第  $i$  张图像;卷积自编码网络产生的编码向量集合  $\mathbf{Z} = \{z_i\}_{i=1}^N$ , 其中  $z_i = f_{\theta_e}(x_i) \in \mathbf{R}^d$  是学习一个由  $\theta_e$  参数化的深度编码网络的映射函数  $f, d$  是特征维度。为确保所学习的表示  $z$  包含来自输入数据有意义的信息,将编码特征  $z_1^{(L)}, z_2^{(L)}, \dots, z_m^{(L)}$  馈入解码器中来重构图像  $\hat{x}$ , 其中  $\hat{x}_{z_i} = g(z_i; \theta_d)$  表示通过参数  $\theta_d$  重构的原始图像。

学习一个好的特征可以有效地提高聚类性能,为了能够从整个数据集中明确区分样本,设计了一个 DDECN 模块作为特征提取器,即在卷积自编码器中引入互信息最大化这一概念,以辨别出该样本最独特的信息。互信息定义为

$$I(x, z) = \iint p(z|x)\tilde{p}(x)\log_2 \frac{p(z|x)}{p(z)} dx dz = D_{KL}[p(z|x)\tilde{p}(x) || p(z)\tilde{p}(x)], \quad (3)$$

式中:  $p(z|x)$  为  $x$  所产生的编码向量的分布;  $\tilde{p}(x)$  为原始样本的分布; 整个潜在特征空间的分布  $p(z) = \int p(z|x)\tilde{p}(x) dx$ 。互信息最大化的本质是尽量拉大两种概率分布  $p(z|x)\tilde{p}(x)$  和  $p(z)\tilde{p}(x)$  之间的距离,那么  $x$  和  $z$  的相关性就越高。当  $p(z|x)$  远大于  $p(z)$  时,则意味着编码器找到了专属于  $x$  的潜在特征向量  $z$ 。因此,一个好的特征编码器的优化目标就是最大化互信息,即

$$p(z|x) = \arg [\max_{\theta_e} I(x; z)]. \quad (4)$$

此外,对于编码器学到样本  $x_i$  的潜在特征向量  $z_i$ ,其潜在特征空间分布  $p(z)$  应该服从标准正态分布的先验分布,这可以使编码空间更规整,有利于解耦特征,便于后续特征学习。该约束区别于 Hjelm 等<sup>[20]</sup>通过对抗加入的约束,因为对抗是一个最大化最小化的过程,需要交替训练,不够稳定。因此在 DDECN 模块中构建一个符合标准正态分布的先验分布  $q(z)$ , 训练过程中只需要最小化  $p(z)$  与先验



分布  $q(z)$  之间的 K-L 散度, 即

$$D_{KL} [p(z) \parallel q(z)] = \int p(z) \log_2 \frac{p(z)}{q(z)} dz. \quad (5) \quad \text{标为}$$

联合(4)、(5)式, 可以得到编码器的最小化总目

$$p(z|x) = \min_{\theta_e} \{-I(x, z) + D_{KL} [p(z) \parallel q(z)]\} = \min_{\theta_e} \left[ -\iint p(z|x) \tilde{p}(x) \log_2 \frac{p(z|x)}{p(z)} dx dz + \lambda \int p(z) \log_2 \frac{p(z)}{q(z)} dz \right]. \quad (6)$$

将(6)式改写为

$$p(z|x) = \min_{\theta_e} \left\{ -\iint p(z|x) \tilde{p}(x) \left[ -(1+\lambda) \log_2 \frac{p(z|x)}{p(z)} + \lambda \log_2 \frac{p(z|x)}{q(z)} \right] dz dx \right\}. \quad (7)$$

由(3)、(7)式, 整理可得

$$p(z|x) = \min_{\theta_e} \{-\lambda_1 \times I(x, z) + \lambda_2 \times E_{x \sim \tilde{p}(x)} \{D_{KL} [p(z|x) \parallel q(z)]\}\}, \quad (8)$$

式中: 平衡不同项的贡献系数  $\lambda_1, \lambda_2 > 0$ 。由于 K-L 散度理论上是无上界的, 最大化一个无上界的量, 极可能会得到无穷大的结果。为了更有效地进行优化, 使用一个有上界的 Jensen-Shannon(J-S)散度的变分推断作为度量。它被定义为

$$D_{JS} [p(x) \parallel q(x)] = \max_T \{E_{x \sim p(x)} \{\log_2 \sigma [T(x)]\} + E_{x \sim q(x)} \{\log_2 \{1 - \sigma [T(x)]\}\}\}, \quad (9)$$

式中: 判别器  $T(x) = \log_2 \frac{2p(x)}{p(x)+q(x)}$ ;  $\sigma(\cdot)$  为 Sigmoid 激活函数。根据  $I(x, z)$  的定义, 得到 J-S 散度的互信息为

$$D_{JS} [p(z|x) \tilde{p}(x) \parallel p(z) \tilde{p}(x)] = \max_T \{E_{(x,z) \sim p(z|x) \tilde{p}(x)} \{\log_2 \sigma [T(x, z)]\}, E_{(x,z) \sim p(z) \tilde{p}(x)} \{\log_2 \{1 - \sigma [T(x, z)]\}\}\}, \quad (10)$$

式中: 真实样本来自于  $p(z|x) \tilde{p}(x)$ ; 负样本来自于  $p(z) \tilde{p}(x)$ 。此时, (10)式对应的优化目标为

$$p(z|x) = \min_{\theta_e, T} \{-\lambda_1 \{E_{x \sim p(z|x) \tilde{p}(x)} \{\log_2 \sigma [T(x, z)]\} - E_{x \sim p(z) \tilde{p}(x)} \log_2 \{1 - \sigma [T(x, z)]\}\}, \lambda_2 E_{x \sim \tilde{p}(x)} \{D_{KL} [p(z|x) \parallel q(z)]\}\}. \quad (11)$$

(11)式表示输入  $x$  与  $z$  的全局互信息, 但无监督图像聚类应该是一个从整体到局部的过程。因此, 从卷积神经网络的中间层提取特征图, 构造特征图与潜在特征表示之间的关系, 即局部互信息。局部互信息估算网络把中间层特征与编码特征拼接起来得到一个更大的特征图, 然后对这个

特征图使用 3 个  $1 \times 1$  卷积层作为估算网络的方法。局部互信息估算网络如图 4 所示。对于局部互信息估计, 记中间层特征为  $\{C_{ij}(x) \mid i=1, 2, \dots, N; j=1, 2, \dots, N\}$ , 一共有  $M \times M$  个向量与编码向量  $z$  的互信息。因此, 编码器网络需要优化的损失函数为

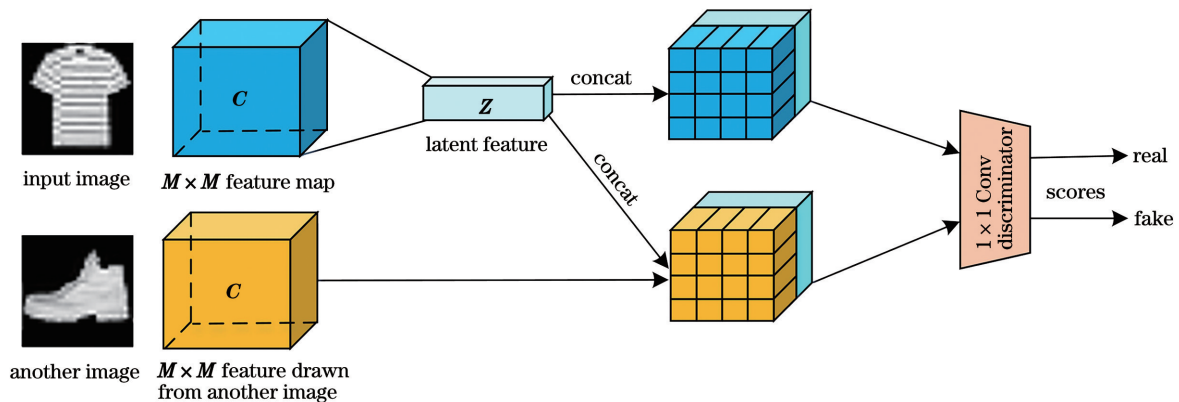


图 4 局部互信息估算网络

Fig. 4 Local mutual information estimation network

$$L_e = -\lambda_1 \{E_{x \sim p(z|x)\tilde{p}(x)} \{\log_2 \sigma [T_1(\mathbf{x}, \mathbf{z})]\} + E_{x \sim p(z)\tilde{p}(x)} \{\log_2 \{1 - \sigma [T_1(\mathbf{x}, \mathbf{z})]\}\}\} - \frac{\lambda_1}{M \times M} \sum_{i,j} \{E_{x \sim p(z|x)\tilde{p}(x)} \{\log_2 \sigma [T_2(\mathbf{C}_{ij}, \mathbf{z})]\} + E_{x \sim p(z)\tilde{p}(x)} \{\log_2 \{1 - [T_2(\mathbf{C}_{ij}, \mathbf{z})]\}\}\} + \lambda_2 E_{x \sim \tilde{p}(x)} \{D_{\text{KL}} [p(\mathbf{z} | \mathbf{x}) \| q(\mathbf{z})]\}, \quad (12)$$

式中:  $T_1(\cdot)$  为全局特征判别器;  $T_2(\cdot)$  为局部特征判别器。

为了确保所学习的编码向量  $\mathbf{z}$  包含来自输入数据的有意义的信息, 将特征映射  $\mathbf{z}_1^{(L)}, \mathbf{z}_2^{(L)}, \dots, \mathbf{z}_m^{(L)}$  馈入解码器网络以重构图像  $\hat{\mathbf{x}}$ 。因此, 此深度自编码网络的重构损失函数为

$$L_r = \frac{1}{2N} \sum_{j=1}^N \|\mathbf{x}_j - \hat{\mathbf{x}}_j\|_2^2 = \frac{1}{2N} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2, \quad (13)$$

式中:  $\|\cdot\|_F^2$  为 Frobenius 范数。

### 3.2 图卷积神经网络模块

自编码网络能够从样本自身学习有用的表示, 例如  $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(L)}$ , 但忽略了样本之间的结构信息。和 Bo 等<sup>[21]</sup> 一样, 为获取输入样本的结构信息, 对原始样本构建一个  $K$  最近邻(KNN)图, 然后使用传递算子  $\delta$  将 DDECN 模块学习到的各层特征表示融入到 GCN 模块对应的各个层中, 此时 GCN 模块可以同时学习到样本自身的属性信息及样本之间的结构信息。对于权重矩阵  $\mathbf{W}$ , 通过卷积运算, 可以获得 GCN 模块第  $l$  层的表示学习  $\mathbf{H}^{(l)}$ 。标准图卷积层传播公式定义为

$$\mathbf{H}^{(l)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l-1)} \mathbf{W}^{(l-1)}), \quad (14)$$

式中:  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ ;  $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$ ;  $\mathbf{I}$  为每个节点邻接矩阵  $\mathbf{A}$  的单位对角矩阵;  $\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$  为标准化邻接矩阵。

为了同时学习 GCN 模块中样本的结构信息和 DDECN 模块中样本自身的属性信息, 使用传递算子  $\delta$  对两个表征进行加权求和, 即

$$\tilde{\mathbf{H}}^{(l-1)} = (1 - \delta) \mathbf{H}^{(l-1)} + \delta \mathbf{Z}^{(l-1)}. \quad (15)$$

通过这种方式将 DDECN 模块与 GCN 模块逐

层连接起来。联合(14)、(15)式得到

$$\mathbf{H}^{(l)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{H}}^{(l-1)} \mathbf{W}^{(l-1)}). \quad (16)$$

这样, 深度自编码网络上一层学习到的表征将融入到 GCN 模块对应层中进行传播。在 GCN 模块的最后一层使用具有 Softmax 功能的多分类层输出  $\mathbf{Y}$ , 从而预测样本的分布情况。

$$\mathbf{Y} = \text{Softmax}(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(L)} \mathbf{W}^{(L)}), \quad (17)$$

式中:  $\mathbf{Y}$  为一种概率分布。

### 3.3 联合优化模型

图卷积神经网络通常用于半监督学习场景, 而不是专门为深度聚类设计的。因此, 为了使图卷积神经网络同样能应用于聚类问题中, SDDECC 算法利用 K-L 散度将 DDECN 模块输出的连续隐变量  $\mathbf{z}$  (代表编码向量) 和 GCN 模块输出的离散隐变量  $\mathbf{y}$  (代表类别信息) 联合起来, 这样双网络结构被统一到一个端到端可训练的聚类优化目标中, 同时指导 DDECN 模块和 GCN 模块学习更有利于聚类的特征表示并迭代更新参数。因此, (12) 式最后一项可以改写为

$$L_c = E_{x \sim \tilde{p}(x)} \{D_{\text{KL}} [p(\mathbf{z}, \mathbf{y}) | \mathbf{x}] \| q(\mathbf{z}, \mathbf{y})\}, \quad (18)$$

式中:  $p[(\mathbf{z}, \mathbf{y}) | \mathbf{x}] = p(\mathbf{y} | \mathbf{z}) p(\mathbf{z} | \mathbf{x})$ ,  $q(\mathbf{z}, \mathbf{y}) = q(\mathbf{z} | \mathbf{y}) q(\mathbf{y})$ ;  $p(\mathbf{z} | \mathbf{x})$  为得到的编码特征  $\mathbf{z}$ ;  $p(\mathbf{y} | \mathbf{z})$  表示对编码特征进行分类;  $q(\mathbf{z} | \mathbf{y})$  为一个均值为  $\mu_y$ 、方差为 1 的正态分布;  $q(\mathbf{y})$  为一个均匀分布, 以使每个类的分布尽量均衡。该目标函数可以使 DDECN 模块和 GCN 模块统一到一个端到端可以训练的联合优化框架中, 实现最优的聚类结果。因此, SDDECC 算法的总损失函数可以改写为

$$\min_{\theta_c, \theta_d, \theta_e} L = \min_{\theta_c, \theta_d, \theta_e} \left\{ -\lambda_1 \{E_{x \sim p(z|x)\tilde{p}(x)} \{\log_2 \sigma [T_1(\mathbf{x}, \mathbf{z})]\} + E_{x \sim p(z)\tilde{p}(x)} \{\log_2 \{1 - \sigma [T_1(\mathbf{x}, \mathbf{z})]\}\}\} - \frac{\lambda_1}{M \times M} \sum_{i,j} \{E_{x \sim p(z|x)\tilde{p}(x)} \{\log_2 \sigma [T_2(\mathbf{C}_{ij}, \mathbf{z})]\} + E_{x \sim p(z)\tilde{p}(x)} \{\log_2 \{1 - \sigma [T_2(\mathbf{C}_{ij}, \mathbf{z})]\}\}\} + \lambda_2 E_{x \sim \tilde{p}(x)} \{D_{\text{KL}} [p(\mathbf{z}, \mathbf{y}) | \mathbf{x}] \| q(\mathbf{z}, \mathbf{y})\} + \frac{1}{2N} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 \right\}. \quad (19)$$

经过训练, 直到达到最大迭代次数时, SDDECC 算法得到稳定收敛。分布  $\mathbf{y}$  中的软分配作为最终

的聚类结果,即分配给样本  $i$  的标签为

$$v_i = \arg \max_j y_{ij}, \quad (20)$$

式中: $v_i$  为最优预测标签; $y_{ij}$  由(17)式计算得到。

## 4 实验结果与分析

把实验分为三个部分。首先在 USPS、MNIST 和 Fashion-MNIST 三个经典数据集上对所提算法和其他 6 种聚类算法进行比较,采用聚类准确率 (ACC) 和标准化互信息 (NMI) 两种量化指标来评价 SDDECC 算法的聚类性能。然后通过控制几个影响因素,进行一系列消融实验来验证所提算法的有效性。最后还讨论了 SDDECC 算法参数  $\lambda_1$ 、 $\lambda_2$  的不同取值对聚类性能的影响。

### 4.1 数据集

为了证明 SDDECC 算法能够更好地处理各种类型的数据集,选择 USPS、MNIST 和 Fashion-MNIST 三个经典图像数据集进行实验。由于聚类任务是完全无监督的,因此实验中将训练样本与测试样本拼接起来。这些数据集的统计信息如表 1 所示。

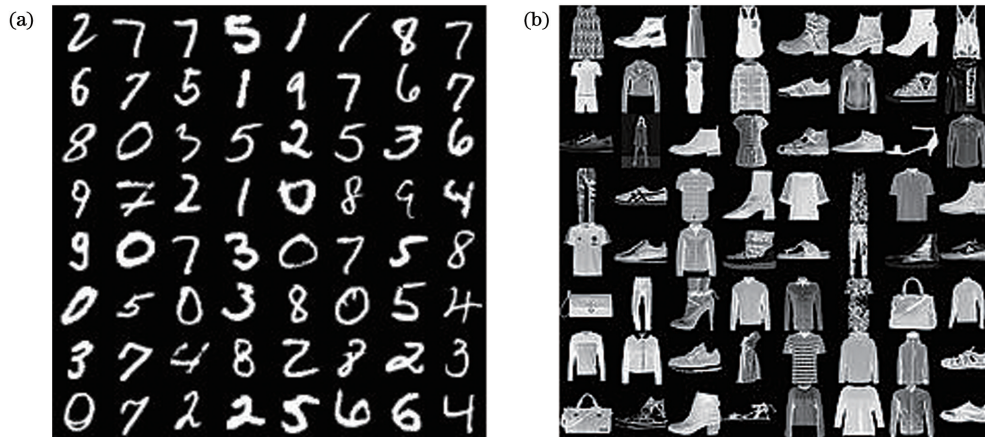


图 5 MNIST 和 Fashion-MNIST 数据集中的部分样本。(a) MNIST 数据集;(b) Fashion-MNIST 数据集

Fig. 5 Some samples in MNIST and Fashion-MNIST datasets. (a) MNIST dataset; (b) Fashion-MNIST dataset

### 4.2 实验设置

实验采用两种标准无监督评价指标来评估算法的聚类性能,即 ACC<sup>[24]</sup> 和 NMI<sup>[25]</sup>。这两个指标在聚类任务中有不同的特点,值越高表示聚类性能越好。实验软件环境为 Ubuntu16.04 系统,硬件环境为 i7-6700 处理器和 NVIDIA GeForce GTX1080 显卡,采用 Python 语言编程,并使用深度学习框架 Pytorch 实现所提算法。实验中,在每个 batch 内对图片进行随机打乱,然后按照随机打乱顺序选择负样本。为了减少超参数搜索次数,统一设置最近邻数  $k=3$ ,传递算子  $\delta=0.5$ 。为减小随机误差,在相同条件下进行

表 1 数据集介绍

Table 1 Dataset introduction

Dataset	Number of samples	Number of classes	Dimension
USPS	9298	10	1×16×16
MNIST	70000	10	1×28×28
Fashion-MNIST	70000	10	1×28×28

USPS 美国邮政服务局的手写数字数据集由 16×16 像素大小的灰度数字手写图像组成,一共包含 9298 张图片,10 个类别,其中 4649 张图片作为训练样本,4649 张图片作为测试样本。MNIST 手写数据集由 28×28 像素大小的灰度数字手写图像组成,一共包含 70000 张图片,分为 10 个类别,其中 60000 张图片作为训练集,10000 张图片作为测试集。Fashion-MNIST 数据集涵盖了来自 10 种类别的共 70000 张各种类型的时尚产品,而不是数字,其中 60000 张图片作为训练集,10000 张图片作为测试集。图 5(a)、(b)分别给出了 MNIST 和 Fashion-MNIST 数据集的部分样本示例。

了 10 次实验,取 10 次实验结果的平均值。自编码网络的通道数和核大小设置如表 2 所示。

表 2 自编码网络的通道数和核大小

Table 2 Number of channels and core size of autoencoder network

Dataset	Encoder-1/ Decoder-4	Encoder-2/ Decoder-3	Encoder-3/ Decoder-2	Encoder-4/ Decoder-1
USPS	3×3×16	3×3×32		
MNIST	3×3×16	3×3×16	3×3×32	3×3×32
Fashion-MNIST	3×3×16	3×3×16	3×3×32	3×3×32

### 4.3 实验结果

7 种算法在 MNIST、USPS 和 Fashion-MNIST 三个数据集上的 ACC 和 NMI 如表 3 所示。从表 3 可以看出,相比于其他对比算法,所提 SDDECC 算法在 3 个经典图像数据集上均取得了最高 ACC 和 NMI 值,表明所提算法可以有效提高聚类性

能,达到较好的实验效果。特别是在样本复杂的 Fashion-MNIST 数据集上,所提算法仍然能够获得最佳的实验结果。在 USPS、MNIST 和 Fashion-MNIST 三个图像数据集上,与次优聚类算法相比,SDDECC 算法聚类精度分别提高了 1.97%、2.16%和 3.18%。

表 3 不同聚类算法在三个数据集上的聚类结果

Table 3 Clustering results of different clustering algorithms on three datasets

Algorithm	USPS		MNIST		Fashion-MNIST	
	ACC	NMI	ACC	NMI	ACC	NMI
K-means	0.6682	0.6270	0.5322	0.5004	0.4742	0.5120
AE+K-means	0.6931	0.6620	0.8076	0.7303	0.5853	0.6142
DEC	0.7408	0.7529	0.8655	0.8372	0.5180	0.5462
IDEC	0.7605	0.7846	0.8806	0.8672	0.5291	0.5570
Deepcluster	0.5623	0.5403	0.7971	0.6615	0.5422	0.5100
SDCN	0.7789	0.7926	0.8530	0.8427	0.5780	0.6047
SDDECC	0.7986	0.8142	0.9022	0.8958	0.6171	0.6306

基于深度聚类的方法通常优于传统聚类方法,如 K-means 算法。这主要是因为与浅层聚类方法相比,深度神经网络具有降维能力,可以有效模拟输入分布,捕捉输入的非线性特性和学习良好的深层特征表示。因此,在处理高维非线性数据时,基于深度聚类模型的聚类性能大多优于浅层聚类模型的聚类性能。相较于其他基于深度聚类的方法,如 AE、DEC、IDEC 及 Deepcluster,SDCN 算法由于融合样本自身属性信息和结构信息,实现了样本自身属性信息和结构信息的协同学习,因此得到了更好的预测效果。与 SDCN 算法相比,所提算法在多层卷积自编码器中嵌入互信息估算网络和最小化先验分布约束,有效挖掘了更具有鉴别性样本的深层特征,并使编码空间更加规整,提高了无监督特征提取

的编码质量,进而提升了聚类性能。实验结果表明,SDDECC 算法在 USPS、MNIST 和 Fashion-MNIST 三个经典数据集上获得了比当前先进算法更好的聚类效果。

### 4.4 消融实验

对所提模型的 4 种不同训练策略进行了一系列实验,以验证 SDDECC 算法的有效性。1)只训练一个多层卷积自编码器(ConvAE);2)在卷积自编码器中嵌入互信息估算网络(ConvAE+MI)训练模型;3)卷积自编码器与图卷积神经网络(ConvAE+GCN)共同参与模型训练;4)联合以上三种策略(ConvAE+MI+GCN)协同参与模型训练。这 4 种训练策略对聚类性能的影响如表 4 所示。

表 4 不同训练策略对聚类性能的影响

Table 4 Impact of different strategies on clustering performance

Method	USPS		MNIST		Fashion-MNIST	
	ACC	NMI	ACC	NMI	ACC	NMI
ConvAE	0.6981	0.6519	0.7762	0.7450	0.5462	0.5563
ConvAE+MI	0.7853	0.7442	0.8350	0.8023	0.5922	0.6091
ConvAE+GCN	0.7822	0.7875	0.8574	0.8449	0.5843	0.6167
ConvAE+MI+GCN	0.7986	0.8142	0.9022	0.8958	0.6171	0.6306

表 4 清楚地显示出,在多层卷积自编码器的基础上使用每一种训练策略均能有效提高聚类性能,尤其是在 ConvAE 中加入互信息估计网络和融合

样本的结构信息之后,聚类效果得到明显提升。在策略 2)中,多层卷积自编码器中嵌入互信息估算网络,由于同时考虑输入与潜在特征表示之间的全局



互信息和中间层特征与潜在特征表示之间的局部互信息,特别是局部互信息相当于将每个小局部看成一个样本,这样原来 1 个样本变成了  $1+M \times M$  个样本,大大增加了样本量,从而提高了无监督特征提取的编码质量。在 USPS 数据集上,策略 2) 的 ACC 和 NMI 相较于策略 1) 分别提升了 0.087 和 0.092; 在 MNIST 数据集上,ACC 和 NMI 分别提升了 0.059 和 0.057; 在 Fashion-MNIST 数据集上,ACC 和 NMI 分别提升了 0.046 和 0.053。在策略 3) 中,将 ConvAE 中不同层学习到的特征融入到 GCN 模块对应的层中,使模型能够同时学习样本自身的属性信息和样本之间的结构信息,因此 ConvAE 与 GCN 相结合的策略同样产生了优于仅使用 ConvAE 的效果。在 USPS 数据集上,ACC 和 NMI 相较于策略 1) 分别提升了 0.084 和 0.136; 在 MNIST 数据集上,ACC 和 NMI 分别提升了 0.081 和 0.100; 在 Fashion-MNIST 数据集上,ACC 和

NMI 分别提升了 0.038 和 0.060。在策略 4) 中,将以上三种策略联合起来端到端地联合优化特征提取和聚类分配过程,最终该模型产生了更强大的预测效果。在 USPS 数据集上,相较于策略 2) 和策略 3),策略 4) 的 ACC 分别提升了 0.013 和 0.016, NMI 分别提升了 0.07 和 0.027; 在 MNIST 数据集上,ACC 分别提升了 0.067 和 0.045, NMI 分别提升了 0.094 和 0.051; 在 Fashion-MNIST 数据集上,ACC 分别提升了 0.025 和 0.033, NMI 分别提升了 0.022 和 0.014。

通过使用 t-SNE 可视化方法,在 MNIST 数据集中直观显示了不同训练策略的聚类结果,如图 6 所示。图 6(a) 为 ConvAE 潜在子空间中数据点的分布情况,图 6(b)~(d) 为所提模型的不同策略在子空间中数据点的分布情况。从图 6 各个策略在潜在空间的分布情况来看,结合三者策略协同参与的模型训练得到的嵌入子空间中数据点具有更清晰的分布结构。

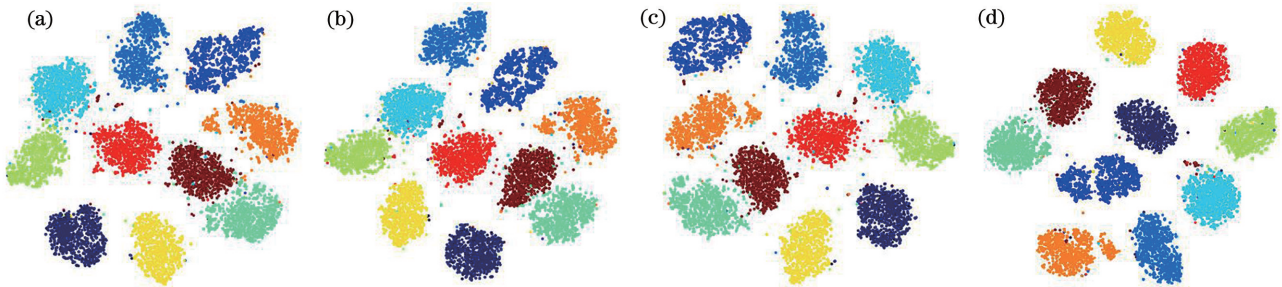


图 6 在 MNIST 数据集上可视化不同策略的嵌入子空间分布情况。(a) ConvAE; (b) ConvAE+MI; (c) ConvAE+GCN; (d) ConvAE+MI+GCN

Fig. 6 Distribution visualization of embedded subspaces of different strategies on the MNIST dataset. (a) ConvAE; (b) ConvAE+MI; (c) ConvAE+GCN; (d) ConvAE+MI+GCN

#### 4.5 算法参数评估实验

为了研究 SDDECC 算法对参数  $\lambda_1$  和  $\lambda_2$  的敏感性,使用 ACC 和 NMI 来评估不同参数对聚类性能的影响。图 7 显示了在 Fashion-MNIST 数据集上,参数  $\lambda_1$  在空间  $\{0.001, 0.005, 0.01, 0.05, 0.1\}$

和  $\lambda_2$  在空间  $\{1, 1.25, 1.5, 1.75, 2\}$  中搜索时 ACC 和 NMI 的变化情况。可以看出,参数  $\lambda_1$  和  $\lambda_2$  在不同组合下对 ACC 和 NMI 有不同程度的影响,但在大多数参数组合下,ACC 和 NMI 都能保持相对稳定的结果。

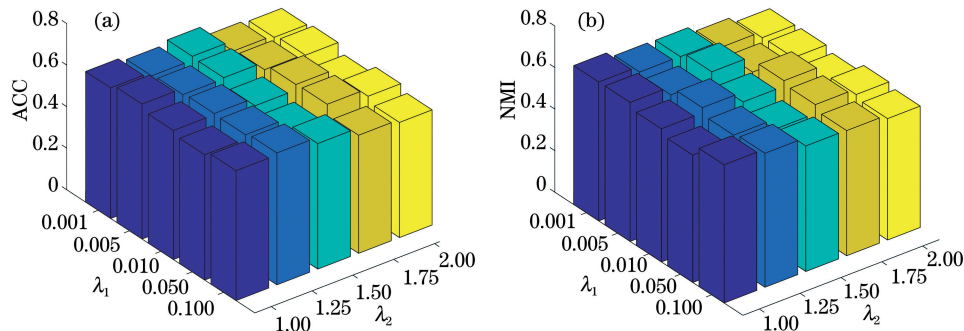


图 7 在 Fashion-MNIST 数据集上参数  $\lambda_1$  和  $\lambda_2$  的不同组合对 ACC 和 NMI 的影响。(a) 对 ACC 的影响; (b) 对 NMI 的影响

Fig. 7 Effect of different combinations of parameter  $\lambda_1$  and  $\lambda_2$  on the ACC and NMI on the Fashion-MNIST dataset.

(a) Effect on ACC; (b) effect on NMI

## 5 结 论

为有效提高深层特征的判别能力,充分利用未标记样本之间的结构信息,联合优化样本的特征提取和聚类过程,提出一种结构化深度判别嵌入编码网络聚类算法。该算法在卷积自编码网络中嵌入互信息估算网络和最小化先验分布约束,同时考虑输入样本自身的属性信息和样本之间的结构信息,有效提高特征判别能力的同时保留了更多可用的结构信息。在图像聚类任务上,SDDECC 算法利用 K-L 散度联合双网络结构产生的潜在特征分布,建立一个特征学习和聚类相统一的联合优化框架,端到端地指导网络模型学习更有利于聚类的深层特征并迭代更新参数。实验结果表明,SDDECC 算法在三个经典图像数据集上的聚类精度均得到较显著的提升。特别是在样本复杂的 Fashion-MNIST 数据集上,相较于次优聚类算法,所提算法精度提高了 3.18%。然而,仅在规模较小的图像数据集上验证了 SDDECC 算法的有效性,如何在更多大数据集上有效提高聚类性能是下一步研究的重点。

### 参 考 文 献

- [1] Shi J B, Malik J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22 (8): 888-905.
- [2] Zhao Z M, Zhu Z L, Liu Y J, et al. Fuzzy C-means clustering algorithm for image segmentation insensitive to cluster size[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021001.  
赵战民, 朱占龙, 刘永军, 等. 对类大小不敏感的图像分割模糊 C 均值聚类方法[J]. 激光与光电子学进展, 2020, 57(2): 021001.
- [3] Pan Z H, Chen Y. Full-convolution object detection network based on clustering region generation [J]. Laser & Optoelectronics Progress, 2019, 56(15): 151001.  
潘志浩, 陈莹. 基于聚类式区域生成的全卷积目标检测网络[J]. 激光与光电子学进展, 2019, 56(15): 151001.
- [4] Yang M L, Deng C, Nie F P. Adaptive-weighting discriminative regression for multi-view classification [J]. Pattern Recognition, 2019, 88: 236-245.
- [5] Zeng C P, Ju L J, Zhang J C. Hyperspectral image classification based on clustering dimensionality reduction and visual attention mechanism [J]. Laser & Optoelectronics Progress, 2019, 56(21): 212802.  
曾朝平, 琚丽君, 张建辰. 基于聚类降维和视觉注意机制的高光谱影像分类[J]. 激光与光电子学进展, 2019, 56(21): 212802.
- [6] MacQueen J. Some methods for classification and analysis of multivariate observations [C] // Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, June 21-July 18, 1967, Berkeley. [S.l.:s.n.], 1967: 281-297.
- [7] Ng A Y, Jordan M I, Weiss Y. On spectral clustering: analysis and an algorithm [C] // Proceedings of the 14th International Conference on Neural Information Processing Systems, December 3-8, 2001, Vancouver, British Columbia, Canada: ACM Press, 2001: 849-856.
- [8] Jiang W H, Chung F L. Transfer spectral clustering [M] // Flach P A, Bie T D, Cristianini N. Machine learning and knowledge discovery in databases. Lecture notes in computer science. Heidelberg: Springer, 2012, 7524: 789-803.
- [9] Xu W, Liu X, Gong Y H. Document clustering based on non-negative matrix factorization [C] // Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, July 28-August 1, 2003, Toronto, Canada. New York: ACM Press, 2003: 267-273.
- [10] Ji P, Zhang T, Li H D, et al. Deep subspace clustering networks [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. New York: Curran Associates, 2017: 23-32.
- [11] Chang J L, Wang L F, Meng G F, et al. Deep adaptive image clustering [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5880-5888.
- [12] Xie J Y, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis [EB/OL]. (2016-05-24) [2020-08-10]. <https://arxiv.org/abs/1511.06335>.
- [13] 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, August 19-25, 2017, Melbourne, Australia. California: IJCAI, 2017: 1753-1759.
- [14] Rousseeuw P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis [J]. Journal of Computational and Applied Mathematics,

- 1987, 20: 53-65.
- [15] Jiang Z X, Zheng Y, Tan H C, et al. Variational deep embedding: an unsupervised and generative approach to clustering [C] // Proceedings of the 26th International Joint Conference on Artificial Intelligence, August 19-26, 2017, Melbourne, Australia. California: IJCAI, 2017: 1965-1972.
- [16] Haeusser P, Plapp J, Golkov V, et al. Associative deep clustering: training a classification network with no labels [M] // Brox T, Bruhn A, Fritz M. Pattern recognition. Lecture notes in computer science. Cham: Springer, 2019, 11269: 18-32.
- [17] Caron M, Bojanowski P, Joulin A, et al. Deep clustering for unsupervised learning of visual features [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11216: 139-156.
- [18] Su J L. Mutual information in deep learning: unsupervised feature extraction [EB/OL]. (2018-11-03) [2020-4-18]. <https://kexue.fm/archives/6024>.  
苏剑林. 深度学习的互信息: 无监督提取特征 [EB/OL]. (2018-11-03) [2020-4-18]. <https://kexue.fm/archives/6024>.
- [19] He G C, Liu X B. Unsupervised visual representation learning with image triplets mining [J]. Chinese Journal of Computers, 2018, 41 (12): 2787-2803.
- 何果财, 刘峡壁. 基于图像三元组挖掘的无监督视觉表示学习 [J]. 计算机学报, 2018, 41 (12): 2787-2803.
- [20] Hjelm R D, Fedorov A, Lavoie-Marchildon S, et al. Learning deep representations by mutual information estimation and maximization [EB/OL]. (2019-02-22) [2020-04-18]. <https://arxiv.org/abs/1808.06670>.
- [21] Bo D Y, Wang X, Shi C, et al. Structural deep clustering network [C] // Proceedings of the Web Conference 2020, April 20-24, 2020, Taipei, Taiwan. New York: ACM, 2020: 1400-1410.
- [22] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [EB/OL]. (2017-02-22) [2020-04-18]. <https://arxiv.org/abs/1609.02907>.
- [23] Kipf T N, Welling M. Variational graph auto-encoders [EB/OL]. (2016-11-21) [2020-04-18]. <https://arxiv.org/abs/1611.07308>.
- [24] Cai D, He X F, Han J W. Locally consistent concept factorization for document clustering [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(6): 902-913.
- [25] Strehl A, Ghosh J. Cluster ensembles: a knowledge reuse framework for combining multiple partitions [J]. Journal of Machine Learning Research, 2002, 3 (12): 583-617.