

基于自注意力对抗的深度子空间聚类

尹明¹ 吴浩杨¹ 谢胜利¹ 杨其宇¹

摘要 子空间聚类 (Subspace clustering) 是一种当前较为流行的基于谱聚类的高维数据聚类框架. 近年来, 由于深度神经网络能够有效地挖掘出数据深层特征, 其研究倍受各国学者的关注. 深度子空间聚类旨在通过深度网络学习原始数据的低维特征表示, 计算出数据集的相似度矩阵, 然后利用谱聚类获得数据的最终聚类结果. 然而, 现实数据存在维度过高、数据结构复杂等问题, 如何获得更鲁棒的数据表示, 改善聚类性能, 仍是一个挑战. 因此, 本文提出基于自注意力对抗的深度子空间聚类算法 (SAADSC). 利用自注意力对抗网络在自动编码器的特征学习中施加一个先验分布约束, 引导所学习的特征表示更具有鲁棒性, 从而提高聚类精度. 通过在多个数据集上的实验, 结果表明本文算法在精确率 (ACC)、标准互信息 (NMI) 等指标上都优于目前最好的方法.

关键词 子空间聚类, 生成对抗网络, 自注意力模型, 深度学习

引用格式 尹明, 吴浩杨, 谢胜利, 杨其宇. 基于自注意力对抗的深度子空间聚类. 自动化学报, 2022, 48(1): 271–281

DOI 10.16383/j.aas.c200302

Self-attention Adversarial Based Deep Subspace Clustering

YIN Ming¹ WU Hao-Yang¹ XIE Sheng-Li¹ YANG Qi-Yu¹

Abstract Subspace clustering is a popular spectral clustering framework for high-dimensional data. In recent years, deep neural networks are witnessed to effectively mine the features of data, which has attracted the attention from various fields. Deep subspace clustering aims to learn the low-dimensional feature representation of data via the deep network, and subsequently calculate the similarity matrix fed into spectral clustering framework for final clustering result. However, real data are often with too high dimensions and complex data structure. How to obtain a robust data representation and improve clustering performance remains a challenge. Therefore, this paper proposes a deep subspace clustering algorithm (SAADSC) based on the self-attention adversarial mechanism. Specifically, the self-attention adversarial network is developed to impose a priori distribution constraint in the feature learning of the autoencoder to guide the learned feature representation to be more robust, thereby improving clustering accuracy. Through the experiments on multiple datasets, the results show that the proposed algorithm in this paper is superior to the state-of-the-art methods in terms of accuracy rate (ACC) and standard mutual information (NMI).

Key words Subspace clustering, Generative adversarial networks, self-attention model, deep learning

Citation Yin Ming, Wu Hao-Yang, Xie Sheng-Li, Yang Qi-Yu. Self-attention adversarial based deep subspace clustering. *Acta Automatica Sinica*, 2022, 48(1): 271–281

聚类作为无监督学习的技术^[1], 是许多领域中常用的统计分析技术, 如图像分割、人脸识别、文本分析等. 给定一组数据, 聚类算法旨在将数据分成若干簇, 同一簇内的数据具有相似特征, 而不

同簇的数据具有较大差异的特征, 通常衡量数据相似性可采用某种距离函数, 如欧氏距离、闵可夫斯基距离、信息熵等. 目前较流行的聚类方法有 k 均值 (k -means) 聚类^[2]、层次聚类^[3]、谱聚类^[4]等. 然而现实生活中存在高维数据, 单独使用以上方法聚类的效率极低, 并且在数据存在噪声干扰时结果也不够鲁棒.

近年来各国学者发现, 虽然高维数据的结构在整个数据空间很难聚类, 但高维数据的内在结构通常小于实际维度, 并且簇结构可能在某个子空间很容易被观测到^[5]. 因此, 为了聚类高维数据, 子空间聚类 (Subspace clustering, SC)^[6] 假定高维空间可分成若干个低维子空间, 然后将这些低维子空间中提取的数据点分割成不同的簇^[7]. 子空间聚类目前主要有 4 大类: 迭代法^[8]、代数法^[9]、统计法^[10]、基于

收稿日期 2020-05-12 录用日期 2020-09-14
Manuscript received May 12, 2020; accepted September 14, 2020

国家自然科学基金 (U1911401, 61973087, 61876042), 广东省自然科学基金 (2020A1515011493), 流程工业综合自动化国家重点实验室开放课题基金项目 (2020-KF-21-02) 资助

Supported by National Natural Science Foundation of China (U1911401, 61973087, 61876042), Guangdong Basic and Applied Basic Research Foundation (2020A1515011493), and State Key Laboratory of Synthetical Automation for Process Industries (2020-KF-21-02)

本文责任编辑 金连文

Recommended by Associate Editor JIN Lian-Wen

1. 广东工业大学自动化学院 广州 510006

1. School of Automation, Guangdong University of Technology, Guangzhou 510006

谱聚类的方法^[11-16]. 其中基于谱聚类的子空间聚类一经提出就受到了广泛的关注, 其基本思想是首先计算数据点间的相似性来构建相似度矩阵, 然后再采用谱聚类算法^[17] 获得最终聚类结果. 其中最成功的两种子空间聚类算法为: 稀疏子空间聚类 (Sparse subspace clustering, SSC)^[11] 通过 ℓ_0 范数正则化迫使每个数据由同一子空间的其他数据点尽可能稀疏地表示, 再利用表示系数构建相似度矩阵, 所得的相似度矩阵可捕捉到数据的局部结构; 低秩子空间聚类 (Low-rank representation, LRR)^[12] 通过核范数正则化来获得数据的最低秩表示, 这样获得的相似度矩阵具有数据的全局结构信息. 这两种算法都采用了数据“自表示”机制, 有效地刻画出数据的子空间结构.

另一方面, 随着神经网络的发展, 自动编码器 (Autoencoders, AEs)^[18] 成为流行的特征学习方法. 其通过编码器将原始数据编码成一个低维的编码, 然后再通过解码器把低维的编码重构回原始数据, 这个低维的编码数据可近视为原始数据的低维表示. 2006 年, Hinton 等对自动编码器进行改进提出深度自动编码器 (Deep autoencoder, DAE)^[19], 相较于自动编码器, 由于加深了网络深度的 DAE 可获得更鲁棒的数据表示. 之后, Vincent 等提出了去噪自动编码器 (Denoising autoencoders, DAEs)^[20] 通过在数据中加入噪声来进一步提高鲁棒性. 为了去掉数据的冗余信息, 获得稀疏的数据表示, Bengio 等提出稀疏自动编码器 (Sparse autoencoders, SAE)^[21]. Masci 等将编码器和解码器的全连接层网络替换为卷积神经网络提出堆叠卷积自动编码器 (Stacked convolutional autoencoders, CAE)^[22] 从而减少网络的参数量. 基于自动编码器网络在特征学习上的优势, 有研究者将其与聚类^[23-24] 算法相结合: 例如, 深度嵌入聚类 (Deep embedded Clustering, DEC)^[25], 同时进行深度学习与聚类算法 (Simultaneous deep learning and clustering, DCN)^[23], 深度连续聚类 (Deep continuous clustering, DCC)^[24] 以及 Ren 等提出基于深度密度的图像聚类算法 (Deep density-based image clustering, DDC)^[26] 和半监督深度嵌入聚类 (Semi-supervised deep embedded clustering, SDEC)^[27] 等.

在数据表示学习时, 我们可以加深网络深度学习更深层的表示^[28], 但通常来说网络并不是越深越好, 由于 AE 网络深度过长导致一些信息丢失, 尤其是某个特定特征^[29-30]. 为了解决这一问题, 注意力模型 (Attention model)^[30] 被提出来. 其基本思想是模仿人类的注意力机制, 即人类会根据内部经验、

外部感觉从一个庞大的信息快速聚焦于局部信息. 其计算可分为两步: 首先对输入信息计算注意力分布, 然后根据注意力分布计算输入信息的加权平均. 采用了自注意力机制的网络相较于其他特征学习网络, 会忽视无关的背景信息. 目前注意力模型主要包括软注意力模型 (Soft attention model)^[30-31]、硬注意力模型 (Hard attention model)^[32]、自注意力模型 (Self-attention model)^[33]、局部注意力模型 (Local attention model) 和全局注意力模型 (Global attention model)^[34] 等.

深度学习另一个突破的进展为 2014 年 Goodfellow 等提出生成对抗网络 (Generative adversarial networks, GAN)^[35], 由一个生成网络与一个判别网络组成, 通过让两个神经网络相互博弈的方式进行学习. 首先, 生成网络从潜在空间 (Latent space) 中随机采样作为输入, 其输出结果需要尽量模仿训练集里的真实样本. 而判别网络的输入则为真实样本或生成网络的输出, 其目的是将生成网络的输出从真实样本中尽可能分辨出来. 最终生成器生成的样本近似于真实样本. 近些年有学者发现生成对抗模型也可以在聚类分析上起到作用, Chen 等提出的 Info-GAN^[36] 在原本的输入增加一个新的潜在编码 (Latent code), 来控制生成结果, 当这个编码为离散编码时, 该算法具有聚类的作用. Mukherjee 等提出 ClusterGAN^[37], 在 Info-GAN 的基础上增加一个编码器来对生成器生成的图像再进行编码来对输入的潜在编码进行约束, 从而获得更好的聚类性能. 值得注意的是, Makhzani 在 2016 年将自动编码器和生成对抗网络相结合的对抗自动编码器 (Adversarial autoencoders, AAE)^[38], 该算法在半监督分类和无监督聚类均有效果.

尽管上述方法在某种程度上提升了聚类精度, 但如何有效地挖掘数据内蕴的子空间结构, 获得更鲁棒的数据表示仍待进一步研究. 因此, 本文拟提出一种基于自注意力对抗机制的深度子空间聚类方法. 在包含“自表示”网络层的深度自动编码器网络中, 我们引入自注意力机制以捕捉重要特征信息, 而且利用对抗网络增强特征学习的鲁棒性. 由此, 学习到更鲁棒的数据子空间结构, 获得更优的聚类结果. 归纳而言, 本文的主要贡献为:

- 1) 提出一种利用对抗机制提升子空间聚类的算法, 使得编码器学习到的特征表示更具有鲁棒性;
- 2) 引入自注意力模型来解决聚类分析中特征学习的长距离依赖问题.

本文章节安排如下: 第 1 节描述了和基于自注意力对抗的子空间聚类相关的算法, 第 2 节描述基

于自注意力对抗的子空间聚类的网络结构以及原理, 第3节通过在MNIST, Fashion-MNIST等数据集实验并分析, 第4节总结全文, 并提出进一步的研究方向.

1 相关工作

本文方法是在深度子空间聚类框架中引入自注意力模型和对抗训练机制的方法, 因此, 本节将围绕子空间聚类、自注意力模型和生成对抗网络进行简要介绍.

1.1 深度子空间聚类

给定一数据集 $X = \{x_1, x_2, \dots, x_n\} \in \mathbf{R}^{d \times n}$, 假设这组数据集属于 N 个线性子空间 $\{\mathcal{S}_i\}_{i=1}^N$, 子空间维度分别为 $\{d_i\}_{i=1}^N$. 假设属于某线性子空间 \mathcal{S}_i 的样本足够多, 且张成整个子空间 \mathcal{S}_i , 则 \mathcal{S}_i 中的任意一样本 x 均能表示为 X 中除去 x 的线性组合, 即数据集的“自表示”特性^[7], 则有如下子空间学习模型:

$$\min \frac{1}{2} \|X - XC\|_F^2 + \lambda \|C\|_p \quad (1)$$

$C \in \mathbf{R}^{n \times n}$ 为输入数据 X 的自表示系数矩阵, 其中 C_i 为第 i 个数据 X_i 由其他数据表示的系数向量. $\|C\|_p$ 为正则化项, $\|\cdot\|_p$ 为任意矩阵范数, 如稀疏子空间聚类的 1-范数 $\|C\|_1$ ^[41], 低秩子空间聚类核范数 $\|C\|_*$ ^[12] 和 F-范数^[39]. 然后使用谱聚类算法对由自表示系数矩阵构建的相似度矩阵 $A = \frac{1}{2} |C + C^T|$ 聚类, 获得最终聚类结果.

学者们发现基于自表示方法利用不同的正则化项可以处理受损数据, 例如, 包含噪声和异常值的数据, 而且自表示系数矩阵呈现出块对角化的结构, 这非常有利于后续的谱聚类^[7]处理. 因此, 如何获得鲁棒的自表示系数矩阵是基于谱聚类的子空间聚类算法的关键问题.

然而, 上述子空间模型学习到的自表示结构仅适用于线性子空间. 另一方面, 现实数据常常具有高维的非线性结构, 传统子空间学习受到限制. 可

喜的是, 深度自动编码器可将数据转换至一个潜在的低维子空间, 捕获数据的非线性结构, 从而获得数据的低维特征表示. 因此与深度神经网络结合的子空间学习旨在低维特征上学习数据的自表示系数, 如深度子空间聚类算法 (Deep subspace clustering, DSC)^[40]. DSC 首先采用深度自动编码器学习原数据的低维特征表示, 然后利用一个由全连接网络构成的自表示层来学习数据的相邻关系, 该自表示层将神经元连接的权重视为同一子空间中数据样本之间的相似度. 其目标函数表示如下:

$$\min \frac{1}{2} \|X - \hat{X}\|_F^2 + \frac{\lambda_1}{2} \|Z - ZC\|_F^2 + \lambda_2 \|C\|_p \quad (2)$$

其中, $\hat{X} = f_d(Z) \in \mathbf{R}^{d \times n}$ 为 X 的重构数据, 式 (2) 中第一项表示编解码的保真度, 尽可能少地损失重要信息. 第二项中 $Z = f_e(X) \in \mathbf{R}^{k \times n}$ ($k \ll d$) 为特征表示矩阵, 结合“自表示”层网络学习自表示系数. $f_e(\cdot)$ 为编码器网络, $f_d(\cdot)$ 表示解码器网络, λ_1 和 λ_2 为权重参数用于调节各项平衡, 其网络结构如图 1 所示.

1.2 生成对抗网络

生成对抗网络 (GAN) 通过生成网络和判别网络相互博弈以达到纳什均衡^[35-41], 用 G 表示为生成器, D 为判别器, 其训练目标函数为:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (3)$$

式中, $E(\cdot)$ 表示分布函数的期望值, $p_{data}(x)$ 为真样本分布, $p_z(z)$ 定义低维噪声分布, $\log(\cdot)$ 表示对数运算. $V(D, G)$ 为真、假样本的差异程度, $\max_D V(D, G)$ 表示当生成器固定时, 使判别器最大化地判别出样本来自于真实数据还是生成的数据, $\min_G V(D, G)$ 表示当判别器固定时, 期望生成器最小化真样本与假样本的差异.

在生成对抗模型中, 首先由随机向量或噪声 z 通过生成器生成一个假样本 $G(z)$, 然后判别器对真

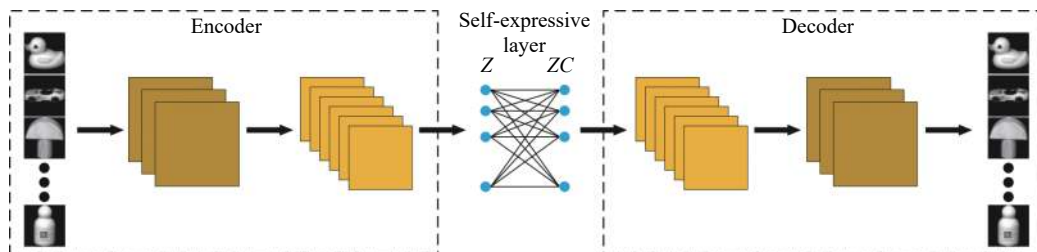


图 1 深度子空间聚类网络结构图

Fig.1 The framework of Deep Subspace Clustering

样本 x 和假样本 $G(z)$ 进行真假判断. 当判别器 D 对真样本 x 甄别越严格, $D(x)$ 值也越接近于 1, 此时 $\log D(x)$ 值也就越接近于 0. 其标准网络结构如图 2 所示, 通常式 (3) 中判别器损失采用交叉熵损失函数, 即通过交叉熵来判别两个分布的相似性. 当目标函数收敛时, 生成分布将拟合于真实分布.

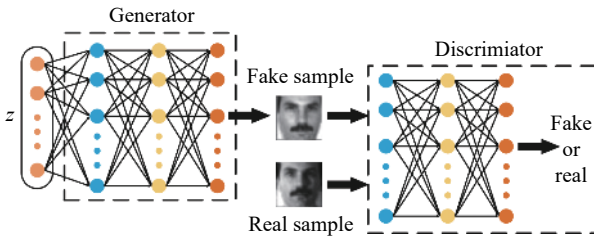


图 2 生成对抗网络结构图

Fig. 2 The framework of generative adversarial networks

由于生成对抗网络可以将一个生成样本分布拟合于真实分布, 使得其不仅局限于样本生成, 而且任何数据分布均可采用生成对抗网络来拟合, 例如, 对抗自动编码器 (AAE)^[38] 将特征表示分布拟合于标准高斯分布, 获得与变分自动编码器 (Variational autoencoders, VAE)^[42] 相似的效果.

1.3 自注意力模型

目前大多数注意力模型嵌入于编-解码器 (Encoder-decoder) 框架. 通常一个高维的数据通过编码输出一个低维的特征表示时, 会损失大量的信息, 注意力模型可以对不同的数据信息加权平均, 因此含有注意力模型的编-解码器框架会编码出一个低维且信息损失较少的特征表示. 注意力模型的数学表达如下:

$$\text{Attention}(Q, K, V) = s(Q, K^T) V \quad (4)$$

其中, $s(\cdot)$ 为计算 Q 和 K^T 的相似度函数, $K = V = f_e(\cdot)$, $Q = f_d(\cdot)$. 一般, 注意力模型可以抽象为计算输出信息 Q 与输入信息 $\langle K, V \rangle$ 的关联性.

自注意力模型是注意力模型家族中最为广泛应用的一种. 在该模型中 K, V 与普通注意力模型一样来自于输入信息, 另一方面为了能直接捕捉到输入数据矩阵中任意两个向量的关联性, Q 也源于输入信息. 但注意力模型不局限于编-解码器框架, 例如, 自注意力生成对抗网络 (Self-attention generative adversarial networks, SAGAN)^[43] 中将自注意力模型引入生成对抗网络, 不仅解决卷积结构带来的感受野大小限制, 也使得在生成图像时, 每个局部区域的生成会与全局细节相协调.

2 基于自注意力对抗的深度子空间聚类算法

尽管现有 DSC 算法在一定程度上改善了聚类性能, 但其网络结构的感受野受到限制^[43], 即过大的通道数导致卷积运算难以捕捉数据不同局部间结构, 而且所学习到的数据隐特征分布无法重构具有判别性的样本. 因此, 本节提出一种基于自注意力对抗机制的深度子空间聚类算法, 在 DSC 网络中引入自注意力模块, 并约束数据特征分布近似于任意的先验概率.

2.1 网络模型与结构

基于自注意力对抗的深度子空间聚类网络框架如图 4 所示, 为了保证在特征学习过程中长距离依赖, 我们在编码模块的最后一层卷积网络后添加一个自注意力模块. 其中自注意力模块结构如图 3 所示, 对前一层网络的数据通过 1×1 的卷积网络获得 K, V, Q . 将 K 转置与 V 相乘并经过 softmax 归一化得到注意力映射, 再与 Q 点积得到最终的自注意力特征映射. 在判别网络中, 倒数第二层卷积网络中输出的通道数为 1000, 由于过大的通道数会导致卷积运算很难处理不同局部间的关系, 因此我们在此处也增加一个自注意力模块.

在训练生成对抗网络时, 由于判别器过于强大很容易导致输出梯度值为 0, 即梯度消失, 导致没有足够的梯度信息去更新生成器. 针对生成对抗网络的梯度消失问题, 不少学者对其损失函数进行改进. Arjovsky 等认为式 (3) 交叉熵的选择无论是 KL-散度还是 JS-散度都有其局限, 提出 Wasserstein GAN (WGAN)^[44] 采用 Earth-mover (EM) 距离衡量两个分布距离, 并去掉了对数 \log 运算, 其生成损失和判别损失函数分开表述如下:

$$\mathcal{L}_{gen} = -\mathbb{E}_{x \sim p_{data}(x)} [D(x)],$$

$$\mathcal{L}_{dis} = \mathbb{E}_{z \sim p(z)} [D(G(z))] - \mathbb{E}_{x \sim p_{data}(x)} [D(x)] \quad (5)$$

但是, 上式中每次更新判别器需要将梯度信息绝对值截断, 不超过某个固定常数来保证判别器的稳定性. 因此, WGAN-GP (Improved training of Wasserstein GANs)^[45] 将梯度截断作为一个梯度惩罚项加入式 (7) 中, 但这需要满足 Lipschitz 条件才能保证梯度惩罚起作用, 而 Lipschitz 条件又使得截断的梯度值趋向于设定的固定常数的负值边界或者正值边界. 为解决这个问题, Wu 等采用 Wasserstein 散度的概念, 提出 Wasserstein divergence for gans (WGAN-div)^[46] 摆脱了 WGAN-GP 对 Lipschitz 条件约束的依赖.

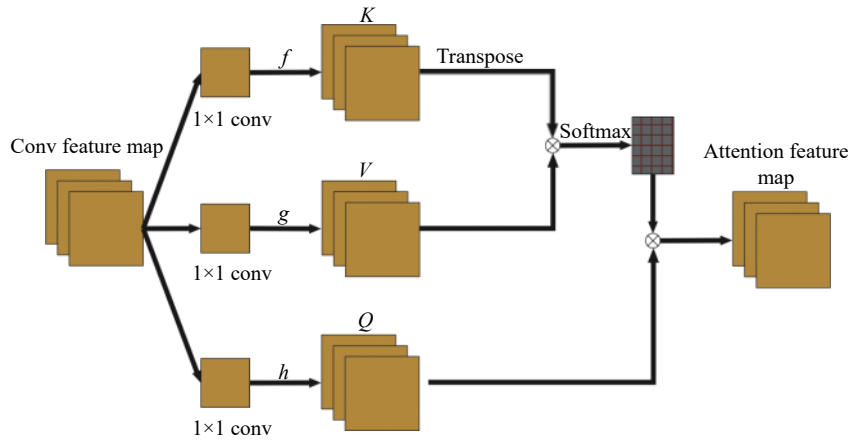


图 3 自注意力模块

Fig.3 Self-attention module

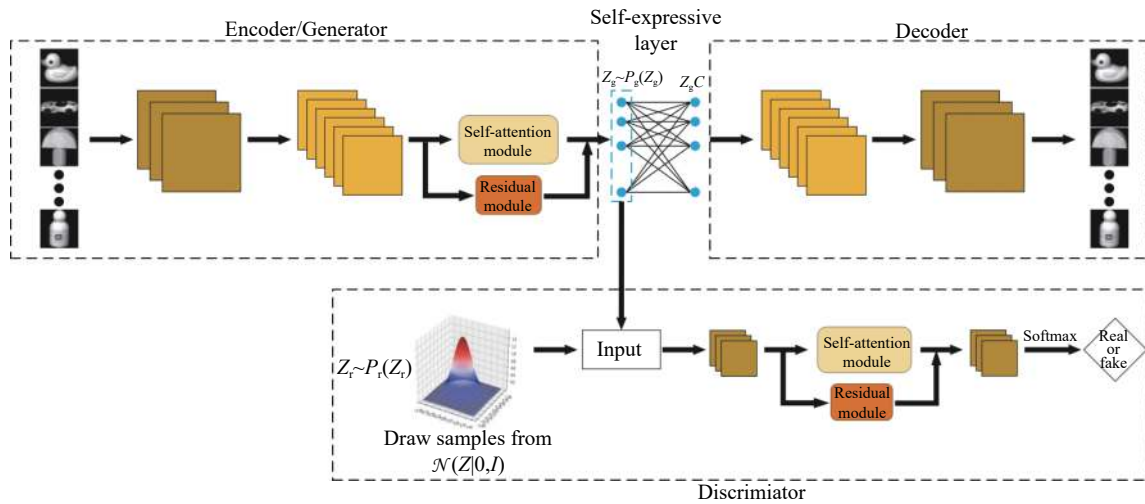


图 4 基于自注意力对抗的深度子空间聚类网络框架

Fig.4 The framework of self-attention adversarial network based deep subspace clustering

综合前人研究的 GAN 及其变种, 我们网络中生成对抗网络部分的损失函数^[46]构造如下:

$$\mathcal{L}_{gen} = -E_{Z_g \sim P_g(Z_g)} [f_D(Z_g)] \quad (6)$$

$$\mathcal{L}_{dis} = E_{Z_g \sim P_g(Z_g)} [f_D(Z_g)] - E_{Z_r \sim P_r(Z_r)} [f_D(Z_r)] + \lambda_3 E_{\hat{Z} \sim P(\hat{Z})} [\|f_D(\nabla \hat{Z})\|^3] \quad (7)$$

其中, $Z_g = f_e(X)$ 为输入数据的特征表示. 如图 4 所示, 在生成对抗网络结构中, 编码模块可以视为一个生成模块, 用于生成特征表示.

因此从生成对抗网络的角度看, Z_g 为假样本, 真样本 Z_r 来自于一个先验分布的采样, 该先验可以服从标准高斯分布、混合高斯分布等. 将真假特征同时输入至判别器网络, 通过博弈训练使得生成器生成的特征分布结构趋向于设定的先验分布的结

构, 导致解码器能够将采用自先验分布 $p(z)$ 的样本生成为观测数据, 从而提高特征学习的鲁棒性, 增加网络的抗干扰能力.

式 (7) 中 \hat{Z} 为真假两个样本之间随机插值, 定义为 $\hat{Z} = \alpha Z_r + (1 - \alpha) Z_g, \alpha \in U(0, 1)$, 并限制判别器的梯度值不会超过 $\|\nabla \hat{Z}\|^3$. 由于梯度惩罚的引入使得判别器稳定输出梯度信息来更新生成器, 从而保证整个网络的稳定性.

考虑引入了生成对抗网络, 我们对式 (2) 改写如下:

$$\mathcal{L}_c = \frac{1}{2} \|X - \hat{X}\|_F^2 + \frac{\lambda_1}{2} \|Z_g - Z_g C\|_F^2 + \lambda_2 \|C\|_F \quad (8)$$

相比于式 (2), 上式中的特征表示已经过生成对抗网络优化, 具有先验分布的结构特性.

此外, 除自注意力模块外, 我们还增加了一个

残差模块来加深网络的深度. 由于残差模块是深度神经网络内部的一部分, 网络的损失函数保持不变. 引入残差模块后, 自注意力模块的输出与残差模块的输出相加作为下一个神经网络单元的输入. 其中残差模块为两个卷积核大小为 3×3 的卷积网络, 卷积步长为 1.

2.2 网络训练

结合聚类和生成对抗机制, 网络的总体损失函数为

$$\mathcal{L}_{total} = \mathcal{L}_c + \mathcal{L}_{gen} + \mathcal{L}_{dis} \quad (9)$$

在训练生成对抗网络时, 通过生成损失和判别损失迭代训练来达成彼此互相更新. 因此, 首先最小化 (8) 的损失来获得特征表示, 然后再通过最小化判别损失和生成损失优化特征表示, 以 epoch 作为一个训练次数, 反复迭代训练三个损失函数达到稳定. 综上所述, 我们的网络训练步骤如下:

输入. 数据集 X , 学习率 η , 以及超参数 $\lambda_1, \lambda_2, \lambda_3$.

步骤 1. 依据式 (8) 求得特征表示 Z_g .

步骤 2. 依据式 (7) 判断 Z_g 与 Z_r 之间的相似性, 并由式 (6) 优化 Z_g .

步骤 3. 依据式 (8) 更新 Z_g .

步骤 4. 重复步骤 1 ~ 步骤 3, 直到达到最大 epoch 值为止.

输出. 自表示系数 C .

一旦完成网络训练, 将学习到数据的相邻关系, 最后利用谱聚类对 C 进行聚类, 获得数据的聚类结果.

3 实验结果与分析

本实验是基于 Python 编程语言进行仿真, 操作系统为 Ubuntu, 主要软件架构为 TensorFlow 1.0, 配置 CUDA 8.0 和 cuDNN 5.1, 使用 4 块英伟达 GPU GTX 1080Ti.

3.1 实验数据集与参数设置

为了验证所提算法的有效性, 我们在 5 个公开的数据集上进行了实验, 两个手写数字数据集 MNIST^[47] 和 USPS, 一个物品数据集 COIL-20^[48], 一个人脸数据集 Extended Yale B^[49] (下文统称 YaleB) 以及一个衣服数据集 Fashion-MNIST (下文统称 FMNIST). 数据集的详细信息见表 1 所示.

针对 5 个数据集, 我们的实验参数如表 2 所示. 其中, λ_1 和 λ_2 分别为对自表示项和正则化项贡献度的权重参数, 为了方便调参, 我们令 $\lambda_1 = 1$. 此

表 1 数据集信息

Table 1 Information of the datasets

数据集	类别	数量	大小
MNIST	10	1000	28×28
FMNIST	10	1000	28×28
COIL-20	20	1440	32×32
YaleB	38	2432	48×32
USPS	10	9298	16×16

表 2 参数设置

Table 2 Parameter setting

数据集	λ_1	λ_2	λ_3
MNIST	1	0.5	10
FMNIST	1	0.0001	100
COIL-20	1	30	10
YaleB	1	0.06	24
USPS	1	0.1	10

外, 在实验过程中, 我们发现式 (7) 中的参数 λ_3 对结果影响式微, 可能是生成对抗网络中只要存在梯度惩罚便可使网络稳定.

3.2 网络结构信息

实验中编码器网络结构为三层卷积网络, 解码器和编码器的网络结构对称, Fashion-MNIST 编码器为一层卷积网络和三个残差模块, 解码器也保持对称, COIL-20 仅用一层卷积网络, 卷积网络具体参数如表 3 所示.

表 3 网络结构参数

Table 3 Network structure parameter

数据集	卷积核大小	通道数
MNIST	[5, 3, 3]	[10, 20, 30]
FMNIST	[5, 3, 3, 3]	[10, 20, 30, 40]
COIL-20	[3]	[15]
YaleB	[5, 3, 3]	[64, 128, 256]
USPS	[5, 3, 3]	[10, 20, 30]

所有实验中判别器网络采用三层卷积网络, 均为 1×1 的卷积核, 以便增加通道的信息交互, 其通道数为 [1000, 1000, 1]. 同时, 我们将自注意力模块加在判别器网络的倒数第二层以增加 1000 个通道特征的长距离依赖. 对于算法的预训练, 大多数以自动编码器为架构的深度聚类算法均采用了自动编码器来进行预训练. 但由于本文算法引入了生成对抗网络, 为了避免判别器初始训练过于强大而干扰特征学习, 因此我们使用对抗自动编码器 (AAE) 进行预训练.

所有实验采用 Adam^[50] 优化, 在激活函数的选择上, 除 YaleB 采用 leaky relu 外, 其他几个数据集均采用 relu, 式 (8) 学习率设置为 0.0001, 动量因子为 0.9, 式 (6) 和 (7) 学习率为 0.0001, 动量因子为 0.9, batchsize 为表 1 中对应的样本数量.

3.3 评价指标

为了评估我们算法的优越性, 我们采用两个常用的度量方法: 聚类精度 (ACC) 和标准互信息 (NMI)^[51] 来作为聚类的效果.

$$\text{ACC\%} = \frac{\text{正确的聚类样本}}{\text{总样本}} \times 100\% \quad (10)$$

$$\text{NMI\%} = \frac{2I(A, B)}{H(A) + H(B)} \times 100\% \quad (11)$$

式 (11) 中, $I(\cdot)$ 表示为互信息, 即两个随机变量的相关程度, $H(\cdot)$ 为熵, A 与 B 分别表示聚类的标签和正确标签, 其中 $I(A, B) = H(A) + H(B) - H(A, B)$.

3.4 实验结果

在实验中, 本文采用了与所提出算法相关的深度聚类方法进行实验结果对比, 其中包括: StructAE^[52]、DASC^[53]、DCN、DSC 及 DEC. 实验结果如表 4 所示, 实验数据均为重复 30 次的平均值, 其中加粗数字表示最优值. 因 StructAE 与 DASC 作者没有开源代码, 其实验数据为引用自原论文. 此外, 由于其没有在 FMNIST 和 USPS 上进行实验, 因此在这两个数据集上无实验数据对比. DSC 在数据集 YaleB、COIL-20、MNIST 的实验结果引用其论文, 而在 FMNIST 和 USPS 的实验结果为我们测试所得. DEC 与 DCN 在 YaleB 和 COIL-20 上的实验结果为我们测试, 其余引用原论文结果. 其中 DEC 在 YaleB 的实验结果过于不合理, 我们多次调节网络参数均无明显改变, 因此 DEC 在 YaleB

无测试结果, 用 “*” 表示.

从表 4 可看出, 我们算法在 ACC 和 NMI 指标上均优于其他 6 个深度聚类算法, 通过 DSC-L1、DSC-L2 和我们算法结果对比可看出经过自注意力生成对抗网络学习到的特征表示在聚类上可以获得更好的结果, 例如, 在 MNIST 数据集上, 我们的算法相比于次优 DEC 的 ACC 和 NMI 分别提高了 0.1110 和 0.1281. DEC 在 YaleB 上的结果与 DCN 在 COIL-20 上结果欠佳是因为其没有自表示结构, 没有很好地捕捉数据之间的关联性. DSC、DASC 以及我们的算法均含有自表示结构, 因此在部分数据上性能要优于 DEC 和 DCN. 为了探讨我们生成对抗网络的稳定性, 因此我们对 MNIST 的训练损失可视化. 如图 5 所示, 可看出生成损失和判别损失存在对称性, 表明式 (5) 中的梯度惩罚项是的网络中生成器和判别器的相互博弈均衡, 不存在一方过于强大的现象. 此外, 本文算法的网络训练损失曲线呈现逐渐下降的走势, 且最终趋向平缓, 表明在自表示系数学习过程中, 网络稳定, 损失函数收敛, 并没有因生成对抗损失的引入导致存在不稳定现象. 此外 DSC 网络中相较于 SAADSC 没有自注意力模块和对抗网络, 我们选择了 DSC-L2 的结果进行对比, 可发现增加自注意力模块和对抗网络后, 算法在聚类精度和标准互信息上均有所提高.

为了探讨不同先验分布对结果的影响, 我们选择三种不同的分布在三个数据集上进行实验对比. 如表 5 所示, 高斯分布取得最优的结果, 这是因为当样本容量无限大的时候, 数据的样本分布趋向于高斯分布. 其次, 高斯分布的熵值很大, 当数据分布未知时选择熵值最大的模型效果会更好.

3.5 各模块作用分析 (Ablation study)

为了评估所提出网络中每个模块的作用, 我们分别对其进行测试, Test1 表示去掉自注意力模块及残差模块后的网络; Test2 为去掉自表示层的网

表 4 5 个数据集的实验结果
Table 4 Experimental results of five datasets

数据集	YaleB		COIL-20		MNIST		FMNIST		USPS	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
DSC-L1	0.9667	0.9687	0.9314	0.9395	0.7280	0.7217	0.5769	0.6151	0.6984	0.6765
DSC-L2	0.9733	0.9703	0.9368	0.9408	0.7500	0.7319	0.5814	0.6133	0.7288	0.6963
DEC	*	*	0.6284	0.7789	0.8430	0.8000	0.5900	0.6010	0.7529	0.7408
DCN	0.4300	0.6300	0.1889	0.3039	0.7500	0.7487	0.5867	0.5940	0.7380	0.7691
StructAE	0.9720	0.9734	0.9327	0.9566	0.6570	0.6898	-	-	-	-
DASC	0.9856	0.9801	0.9639	0.9686	0.8040	0.7800	-	-	-	-
SAADSC	0.9897	0.9856	0.9750	0.9745	0.9540	0.9281	0.6318	0.6246	0.7850	0.8134

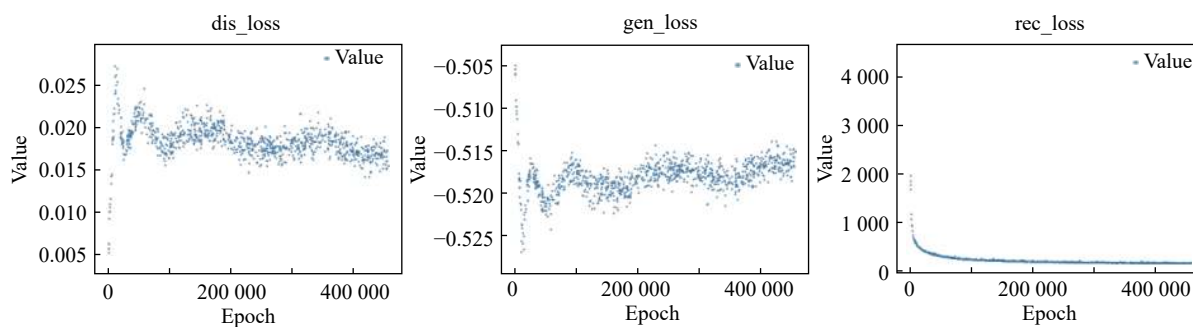


图 5 MNIST 的网络训练损失

Fig. 5 The loss function of SAADSC during training on MNIST

表 5 不同先验分布的实验结果

Table 5 Clustering results on different prior distributions

数据集	MNIST		FMNIST		USPS	
度量方法	ACC	NMI	ACC	NMI	ACC	NMI
高斯分布	0.9540	0.9281	0.6318	0.6246	0.7850	0.8134
伯努利分布	0.9320	0.9043	0.6080	0.5990	0.7755	0.7917
确定性分布	0.8670	0.8362	0.5580	0.5790	0.7796	0.7914

络, 通过谱聚类算法对 $C = Z_g^T Z_g$ 构成的邻接矩阵进行聚类; Test3 为排除自表示层后使用 K -means 对 Z_g 聚类; Test4 排除了网络中残差模块. 实验结果如表 6 所示, 我们可以看出残差模块对网络贡献力度最小, 自表示层作用最大, 其次是自注意力模块. 由于自表示层构建数据之间的线性表示, 通过自表示层获得的系数矩阵反映了类内数据的关联性和类间数据的不相关性.

3.6 鲁棒性实验

另外, 为了测试我们的特征表示相比于深度子空间聚类算法是否更具有鲁棒性, 针对 COIL-20 和 USPS 数据集我们进行了噪声测试. 具体而言, 将对应百分比的像素点替换为随机的高斯噪声, 采用算法 DSC-L1、DSC-L2 和 DASC 进行对比, 因 DASC 代码没开源, 因此其在 USPS 噪声实验中,

无对比数据. 且由于 DASC 采取柱状图进行对比, 因此表 7 中 COIL-20 的实验数据为对其论文柱状图的近似估计. 其余均为测试结果. 实验结果如表 7 和表 8 所示. 可看出, DSC-L1 和 DSC-L2 随着噪声越大结果下降越明显, 而 DASC 和 SAADSC 由于引入生成对抗网络使得算法具有抗干扰能力. 相较于 DASC, 我们的网络的自注意力机制会在生成对抗学习过程中起到积极作用, 从而提升了算法的抗干扰能力.

4 结论

针对现实数据结构复杂、如何获得更鲁棒的数据表示以便改善聚类性能问题, 本文提出一种基于对抗训练的深度子空间聚类, 利用对抗网络的博弈学习能力使得编码器网络学习到的特征表示服从预设定的先验分布特性, 而且引入自注意力机制和残

表 6 SAADSC 网络中不同模块的作用

Table 6 Ablation study on SAADSC

数据集	YaleB		COIL-20		MNIST		FMNIST		USPS	
度量方法	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Test1	0.9725	0.9672	0.9382	0.9493	0.8820	0.8604	0.6080	0.6110	0.7748	0.7838
Test2	0.0711	0.0961	0.4229	0.6263	0.6420	0.5940	0.5380	0.4917	0.6105	0.5510
Test3	0.0843	0.1222	0.6993	0.7855	0.6610	0.6763	0.6140	0.5922	0.3826	0.3851
Test4	0.9782	0.9702	0.9683	0.9741	0.9500	0.9275	0.6211	0.6143	0.7850	0.7986
DSC-L2	0.9733	0.9703	0.9368	0.9408	0.7500	0.7319	0.5814	0.6133	0.7288	0.6963
SAADSC	0.9897	0.9856	0.9750	0.9745	0.9540	0.9281	0.6318	0.6246	0.7850	0.8134

表 7 含有噪声的 COIL-20 聚类结果
Table 7 Clustering results on the noisy COIL-20

算法 度量方法	SAADSC		DSC-L1		DSC-L2		DASC	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
无噪声	0.9750	0.9745	0.9314	0.9353	0.9368	0.9408	0.9639	0.9686
10%噪声	0.9590	0.9706	0.8751	0.8976	0.8714	0.9107	0.9021	0.9392
20%噪声	0.9111	0.9593	0.8179	0.8736	0.8286	0.8857	0.8607	0.9193
30%噪声	0.8708	0.9638	0.7989	0.8571	0.8072	0.8784	0.8357	0.9143
40%噪声	0.8569	0.9272	0.6786	0.7857	0.7250	0.8187	0.7805	0.8753

表 8 含有噪声的 USPS 聚类结果
Table 8 Clustering results on the noisy USPS

算法 度量方法	SAADSC		DSC-L1		DSC-L2	
	ACC	NMI	ACC	NMI	ACC	NMI
无噪声	0.7850	0.8134	0.6984	0.6765	0.7288	0.6963
10%噪声	0.7778	0.7971	0.6704	0.6428	0.6562	0.6628
20%噪声	0.7757	0.7901	0.6667	0.6158	0.6530	0.6429
30%噪声	0.7719	0.7844	0.6386	0.5987	0.6454	0.6394
40%噪声	0.7674	0.7750	0.6042	0.5752	0.6351	0.6164

差网络模块, 来增强特征学习的鲁棒性, 从而提升聚类性能. 实验结果表明, 本文算法在精确率 (ACC) 和标准互信息 (NMI) 等指标上都优于目前最好的方法.

References

- Aggarwal C C. An introduction to data classification. *Data Classification: Algorithms and Applications*, 2014, **125**(3): 142–147
- RMacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 1967, 1(14): 281–297
- Johnson S C. Hierarchical clustering schemes. *Psychometrika*, 1967, **32**(3): 241–254
- Ng A Y, Jordan M I, Weiss Y. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 2002
- Basri R, Jacobs D W. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(2): 218–233
- Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data. *Data Mining and Knowledge Discovery*, 2005, **11**(1): 5–33
- Wang Wei-Wei, Li Xiao-Ping, Feng Xiang-Chu, Wang Si-Qi. A survey on sparse subspace clustering. *Acta Automatica Sinica*, 2015, **41**(8): 1373–1384
(王卫卫, 李小平, 冯象初, 王斯琪. 稀疏子空间聚类综述. 自动化学报, 2015, **41**(8): 1373–1384)
- Bradley P S, Mangasarian O L. K-plane clustering. *Journal of Global Optimization*, 2000, **16**(1): 23–32
- Gear C W. Multibody grouping from motion images. *International Journal of Computer Vision*, 1998, **29**(2): 133–150
- Yang A Y, Rao S R, Ma Y. Robust statistical estimation and segmentation of multiple subspaces. In: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). IEEE, 2006.
- Elhamifar E, Vidal R. Sparse subspace clustering. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2009: 2790–2797
- Liu G, Lin Z, Yu Y. Robust subspace segmentation by low-rank representation. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010: 663–670
- Luo D, Nie F, Ding C, Huang H. Multi-subspace representation and discovery. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Berlin, Heidelberg, 2011: 405–420
- Zhuang L, Gao H, Lin Z, Ma Y, Zhang X, Yu N. Non-negative low rank and sparse graph for semi-supervised learning. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2012: 2328–2335
- Kang Z, Zhao X, Peng C, Zhu H, Zhou J T, Peng X, et al. Partition level multiview subspace clustering. *Neural Networks*, 2020, **122**: 279–88
- Kang Z, Pan H, Hoi S C, Xu Z. Robust graph learning from noisy data. *IEEE Transaction on Cybernetics*, 2019: 1833–1843
- Zhou Lin, Ping Xi-Jian, Xu Sen Zhang Tao al. Cluster Ensemble Based on Spectral Clustering. *Acta Automatica Sinica*, 2012, **38**(8): 1335–1342
(周林, 平西建, 徐森, 张涛. 基于谱聚类的聚类集成算法. 自动化学报, 2012, **38**(8): 1335–1342)
- Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *Nature*, 1986, **323**(6088): 533–536
- Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527–1554
- Vincent P, Larochelle H, Bengio Y, Manzagol P A. Extracting

- and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. 2008: 1096–1103
- 21 Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems*, 2007: 153–160
- 22 Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: Proceedings of the International Conference on Artificial Neural Networks, pages 52–59. Springer, 2011
- 23 Yang B, Fu X, Sidiropoulos N D, Hong, M. Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 3861–3870
- 24 Shah S A, Koltun V. Deep continuous clustering [Online]. available: <https://arxiv.org/abs/1803.01449>, March 5, 2018.
- 25 Xie J, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis. In: Proceedings of the 36th International Conference on Machine Learning (ICML). 2016: 478–487
- 26 Ren Y, Wang N, Li M. and Xu Z. Deep density-based image clustering. *Knowledge-Based Systems*, 2020: 105841
- 27 Ren Y, Hu K, Dai X, Pan L, Hoi S C, Xu Z. Semi-supervised deep embedded clustering. *Neurocomputing*, 2019, **325**: 121–30
- 28 Yang J, Parikh D, Batra D. Joint unsupervised learning of deep representations and image clusters. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 5147–5156
- 29 Cho k, Van M B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder–decoder approaches. In: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pages 103–111, Doha, Qatar, October 2014. ACL.
- 30 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. In: Proceedings of SSST-8 Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014.
- 31 Xiao T, Xu Y, Yang K, Zhang J, Peng Y, Zhang Z. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR). 2015: 842–850
- 32 Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, et al. Show, attend and tell: Neural image caption generation with visual attention. In: Proceedings of the 35th International Conference on Machine Learning (ICML). 2015: 2048–2057
- 33 Cheng J, Dong L, Lapata M. Long short-term memory-networks for machine reading. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016
- 34 Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015.
- 35 Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the Neural Information Processing Systems (NIPS). 2014: 2672–2680
- 36 Chen X, Duan Y, Houthoofd R, Schulman J, Sutskever I, Abbeel P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Proceedings of the Neural Information Processing Systems (NIPS). 2016: 2172–2180
- 37 Mukherjee S, Asnani H, Lin E, Kannan S. Clustergan: Latent space clustering in generative adversarial networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019, **33**: 4610–4617
- 38 Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B. Adversarial autoencoders [Online]. available: <https://arxiv.org/abs/1511.05644>, May 25, 2015.
- 39 Ji P, Salzmann M, Li H. Efficient dense subspace clustering. *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2014: 461–468
- 40 Ji P, Zhang T, Li H, Salzmann M, Reid I. Deep subspace clustering networks. In: Proceedings of the Neural Information Processing Systems (NIPS). 2017: 24–33
- 41 Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative Adversarial Networks: The State of the Art and Beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络GAN的研究进展与展望. 自动化学报, 2017, **43**(3): 321–332)
- 42 Kingma D P, Welling M. Auto-encoding variational Bayes. In Proceedings of the International Conference on Learning Representations (ICLR), 2014
- 43 Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, California, USA: PMLR 97, 2019. 7354–7363
- 44 Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. In: Proceedings of the International Conference on Machine Learning (ICML), 2017. 214–223
- 45 Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A C. Improved training of wasserstein GANs. In: Proceedings of the Neural Information Processing Systems (NIPS). 2017: 5767–5777
- 46 Wu J, Huang Z, Thoma J, Acharya D, Van G L. Wasserstein divergence for gans. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018: 653–668
- 47 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, 1998, **86**(11): 2278–2324
- 48 Nene S A, Nayar S K, Murase H. Columbia object image library (coil-20). 1996.
- 49 Lee K C, Ho J, Kriegman D J. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(5): 684–698
- 50 Kingma D P, Ba J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR), 2015.
- 51 Xu W, Liu X, Gong Y. Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. 2003: 267–273
- 52 Peng X, Feng J, Xiao S, Yau W Y, Zhou J T, Yang S. Structured autoencoders for subspace clustering. *IEEE Transactions on Image Processing*, 2018, **27**(10): 5076–5086
- 53 Zhou P, Hou Y, Feng J. Deep adversarial subspace clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 1596–1604.



尹明 广东工业大学自动化学院教授. 主要研究方向为图像处理, 模式识别, 计算机视觉, 机器学习.

E-mail: yiming@gdut.edu.cn

(YIN Ming Professor at the School of Automation, Guangdong University of Technology. His research

interest covers image processing, pattern recognition, computer vision, and machine learning.)



谢胜利 广东工业大学自动化学院教授, IEEE Fellow. 主要研究方向为盲信号处理, 生物医学信号处理.

E-mail: shlxie@gdut.edu.cn

(XIE Sheng-Li Professor at the School of Automation, Guangdong University of Technology, IEEE

Fellow. His research interest covers blind signal processing, and biomedical signal processing.)



吴浩杨 广东工业大学自动化学院硕士研究生. 主要研究方向为子空间学习, 深度聚类.

E-mail: tarkovskiyfans@163.com

(WU Hao-Yang Master student at the School of Automation, Guangdong University of Technology. His

research interest covers subspace learning and deep clustering.)



杨其宇 广东工业大学自动化学院讲师. 主要研究方向为信号处理, 实时数据处理. 本文通信作者

E-mail: yangqiyu@gdut.edu.cn

(YANG Qi-Yu Lecturer at the School of Automation, Guangdong University of Technology. His re-

search interest covers signal processing and real time data processing. Corresponding author of this paper.)