

联合数据增强的语义对比聚类

王气洪 贾洪杰 黄龙霞 毛启容

(江苏大学计算机科学与通信工程学院 江苏镇江 212013)

(firefruit_fight@163.com)

Semantic Contrastive Clustering with Federated Data Augmentation

Wang Qihong, Jia Hongjie, Huang Longxia, and Mao Qirong

(School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013)

Abstract Given the excellent performance of contrastive learning on downstream tasks, contrastive clustering has received much more attention recently. However, most approaches only utilize a simple kind of data augmentation. Although augmented views keep the majority of information from original samples, they also inherit a mixture of characteristic of features, including semantic and non-semantic features, which limits model's learning ability of semantic information under similar or identical view patterns. Even some approaches regard two different augmentation views being from the same sample and keeping similar view patterns as positive pairs, which results in sample pairs lacking of semantics. In this paper, we propose a semantic contrastive clustering method with federated data augmentation to solve these problems. Two different types of data augmentations, namely strong data augmentation and weak data augmentation, are introduced to produce two very different view patterns. These two view patterns are utilized to mitigate the disturbance of non-semantic information and improve the semantic awareness of the proposed approach. Moreover, a global k-nearest neighbor graph is used to bring global category information, which instructs the model to treat different samples from the same cluster as positive pairs. Extensive experiments on six commonly used and challenging image datasets show that the proposed method achieves the state-of-the-art performance and confirms the superiority and validity of it.

Key words strong data augmentation; weak data augmentation; contrastive learning; global category information; clustering

摘要 鉴于对比学习在下游任务中的优异表现,对比聚类的研究受到广泛关注。但是,大部分方法只采用一类简单数据增强技术,尽管增强后的视图保留了原始样本的大部分特征信息,但也继承了语义信息和非语义信息相融交织的特性,在相似或相同的视图模式下,该特性限制了模型对语义信息的学习。有些方法直接将来源于同一样本的具有相同视图模式的2个数据增强视图组成正样本对,导致样本对语义性不足。为解决上述问题,提出基于联合数据增强的语义对比聚类方法,基于一强一弱2类数据增强,利用视图间的差异降低非语义信息的干扰,增强模型对语义信息的感知能力。此外,基于全局k近邻图引入全局类别信息,由同一系列的不同样本形成正样本对。在6个通用的挑战性数据集上的实验结果表明该方

收稿日期: 2022-12-07; 修回日期: 2023-08-15

基金项目: 国家自然科学基金项目(61906077, 62102168, 62176106, U1836220); 江苏省自然科学基金项目(BK20190838, BK20200888); 中国博士后科学基金项目(2020T130257, 2020M671376); 江苏省博士后科学基金项目(2021K596C)

This work was supported by the National Natural Science Foundation of China (61906077, 62102168, 62176106, U1836220), the Natural Science Foundation of Jiangsu Province (BK20190838, BK20200888), the China Postdoctoral Science Foundation (2020T130257, 2020M671376), and the Postdoctoral Science Foundation of Jiangsu Province (2021K596C).

通信作者: 贾洪杰(jiahj@ujs.edu.cn)

法取得了最优的聚类性能,证实了所提方法的有效性和优越性.

关键词 强数据增强;弱数据增强;对比学习;全局类别信息;聚类

中图法分类号 TP181

聚类旨在发现数据的内部聚类结构,将相似的样本划分为一类,现已有许多高效的聚类算法被提出,如原型聚类、密度聚类、层次聚类等.聚类作为机器学习的核心技术之一,已在信息检索^[1]、医学诊断^[2]、社交网络分析^[3]、图像分割^[4]等多个领域中发挥着重要作用.

聚类属于无监督学习,对大量无标签样本进行划分,特别是外观相似、语义不同的样本,是一项重要挑战.经典聚类算法,如谱聚类^[5]、k-means^[6]等,由于特征提取能力缺陷导致特征区分性差的问题,已难以满足现有复杂场景上的任务.

深度聚类方法依靠深度学习^[7]强大的特征提取能力,并采用同步式训练策略来解决传统聚类算法的不足. DEC(deep embedding clustering)^[8]采用1步策略优化特征学习和聚类分析来解决错误传播.DFCN(deep fusion clustering)^[9]将自动编码器和图自动编码器学习的特征信息分别在属性和结构上融合来弥补单个的不足.但是,文献[8-9]方法学习的特征区分性不强,难以取得优异的聚类性能.

近些年,改进的对比学习技术^[10]在下游任务上取得优异表现,不断缩小无监督与有监督间的差距,并显现出超越有监督学习的潜力.大部分对比学习方法采用单一、简单的数据增强技术,如 MOCO (momentum contrast)^[11], SimCLR(simple framework for contrastive learning of visual representation)^[12]等,本文称之为弱数据增强,将同一样本的不同数据增强视图视作正样本对、其他为负样本对,利用对比损失将正样本对拉近、负样本对远离,有效提高特征的区分度.当对比学习和聚类任务结合时,以正负样本为纽带,特征模块和聚类模块呈现较强的相互促进、相互影响的关系.然而,从特征对比和聚类分析的角度,现有的对比聚类分析仍存在一些不足:1)单一的弱数据增强在保留原始样本大部分特征信息的同时,增强后的视图继承了语义信息和非语义信息相融交织的特性且具有单一的视图模式(本文将聚类主体相关信息视作语义信息,其他无关信息视为非语义信息,如背景信息、噪声信息等),这种特性和视图模式阻碍模型对语义信息的学习能力;2)同一样本的不同数据增强视图组成正样本对,正样本对本质上

具有等价的语义概念,降低特征对比的多样性和语义性,同时,这样的正样本对划分到同一类簇也让聚类分析缺乏语义性.

为提升特征对比、聚类分析的多样性和语义性,从而改善模型的语义感知能力,本文提出了联合数据增强的语义对比聚类方法(semantic contrastive clustering with federated data augmentation, SCCFA),在采用弱数据增强方法的基础上,再引入一类差异化的数据增强方法,称为强数据增强(更大的旋转角度、更综合的属性抖动和 CutOut),其增强后的样本称为强视图,弱数据增强生成的样本称为弱视图.由于强数据增强对图片结构破坏性大,强视图丢失原始样本的许多特征信息,形成与弱视图不同的视图模式,利用2种视图模式的差异降低非语义信息的干扰,提升模型的语义信息学习能力.此外,基于相邻样本属于同一类别的假设,借助k近邻(k-nearest neighbor, KNN)算法将同一类别的不同样本组成正样本对、其他为负样本对.如图1所示,基于KNN图提供的类别信息确定锚点样本的邻居样本,锚点样本与邻居样本构建正样本对,其余样本组成负样本对,正样本对靠近用“吸引”表示,负样本对远离用“排斥”表示.经过强和弱2种数据增强后,锚点样本和邻居样本分别拥有强和弱2种形态,并将两两形态用于对比学习而改善模型所学特征的一致性和语义性.锚点样本强视图与邻居样本弱视图的对比(“强 vs 弱”)等价于锚点样本弱视图与邻居样本强视图的对比(“弱 vs 强”,其他视图间对比类似),两者是一类视图间的对比,本文将前者以参数形式融合到后者.

在公开的6个挑战性数据集上的大量实验结果表明了SCCFA的有效性和优越性.

综上,本文的主要贡献包括5个方面:

- 1)提出了一种语义感知的对比聚类框架,利用联合数据增强技术和KNN算法,赋予特征对比和聚类分析语义性,改善模型的语义感知能力.
- 2)采用差异化的强和弱2类数据增强方法,在改善特征多样性和泛化性的同时,利用视图模式差异降低非语义信息的干扰,提高模型对语义信息的学习能力.
- 3)利用全局特征构造全局KNN图,将相同语义

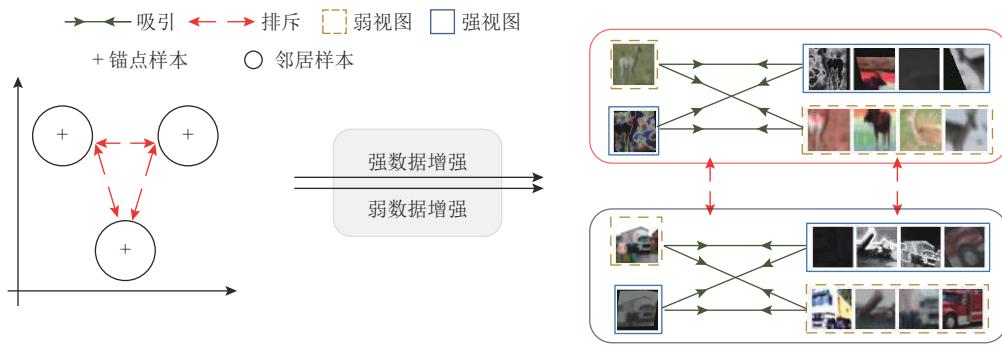


Fig. 1 Illustration of federated data augmentation technology

图1 联合数据增强技术示意图

的不同样本组成正样本对, 提升特征对比和聚类分析的多样性和语义性.

4) 在多个挑战性数据集上进行了大量实验, 实验结果表明 SCCFA 在其中大部分数据集上都取得明显的性能提升.

5) 大量消融实验表明联合数据增强技术能够缓解非语义信息的干扰, 改善模型对语义信息的学习能力; 也表明类别信息能有效改善特征对比和聚类分析的多样性和语义性.

1 相关工作

1.1 深度聚类

传统的聚类算法能成熟地应用于简单的低维数据集, 取得优异的聚类性能. 但这些算法由于特征提取能力缺陷, 无法处理复杂的高维数据集. 深度聚类依靠深度学习强大的特征提取能力, 在复杂数据集上不断拉大与传统聚类算法间的性能差距. DEC^[8] 通过特征嵌入空间的解码和编码进行表征学习. DCCM (deep comprehensive correlation mining)^[13] 利用多类特征间的相互关系来学习特征信息. 从训练策略划分, 这些方法属于交替式训练模型, 将特征学习和聚类分析分 2 步优化, 交替式训练模型由于存在错误传播, 只能学到次优解. 另一种训练策略称为同步式训练, 将特征学习和聚类分析进行 1 步优化, 由于解决了错误传播问题, 聚类效果更好. 比如, IIC (invariant information clustering)^[14] 在概率分布矩阵上提取类别的互信息. PICA (partition confidence maximization)^[15] 提出划分不确定性指标 (partition uncertainty index, PUI) 来衡量解的全局语义性, 以便找到语义最合理的一个类簇划分. 尽管文献 [14–15] 方法取得了不错的聚类效果, 却忽略了特征学习与聚类分析间的关系. 本文利用正负样本对为桥梁, 强化特征学习和聚

类分析间相互促进、相互影响的关系, 让学习的特征同时具备区分性和聚类友好性.

1.2 对比学习

对比学习作为一种自监督学习技术, 近年来受到广泛关注, 许多改进模型在下游任务上取得优异性能, 表现出巨大的潜能. MOCO (momentum contrast)^[11] 将对比学习视作字典查询, 以队列存储负样本, 以动量编码器保持迭代前后负样本的稳定性. MOCO v2 (improved baselines with momentum contrastive learning)^[16] 在 MOCO 的结构上添加一个多层感知机 (multilayer perceptron, MLP), 并将高斯模糊加入到原有数据增强方法中. SimCLR^[12] 简化网络结构, 以 big-batch 形式增加负样本数量. 这些方法虽然能学习到区分性的特征, 但如何与下游任务结合并取得理想性能仍充斥着巨大的挑战. 此外这些方法也存在 2 点不足: 1) 数据增强策略单一、简单, 导致语义信息学习效率低; 2) 正负样本对构建策略缺乏语义概念, 不利于聚类性能的进一步提升. 针对这 2 点不足, 本文引入差异化的强和弱 2 类数据增强方法, 利用联合数据增强技术降低非语义信息对模型学习语义信息的干扰. 此外借助 KNN 算法将同类的不同样本组成正样本对, 完善正负样本对构建策略.

2 本文模型

2.1 数据增强: 弱和强

数据增强是一种通过让有限的数据产生更多的等价数据来人工扩展数据集的技术. 在对比学习中, 数据增强不仅可以生成正样本, 也能改善对比学习在源领域与目标领域间的迁移能力^[17].

在传统对比学习方法中, 弱数据增强除拥有数据增强的基本特性, 如增加训练样本数量外, 主要用于生成锚点样本的正样本. 为使网络尽可能学习到

正样本对间的共有信息,弱数据增强方法应尽量保留锚点样本的特征信息,因此弱数据增强策略具备变换种类单一、变换力度小的特点,该策略一般由4种基本类型组成,分别是ColorJitter, ResizedCrop, HorizontalFlip, GaussianBlur。

文献[18]通过实验表明在全监督或半监督学习中,变换力度大的数据增强方法能提升网络性能。文献[19]通过对多种基本类型的数据增强方法随机组合,探索多种数据增强策略在自监督学习中对网络的影响。受文献[19]的启发,本文采用相似的数据增强随机组合策略,称之为强数据增强策略。强数据增强策略由14种基本类型组成,分别是Sharpness, Brightness, Rotate, TranslateX, TranslateY, Color, Solarize, Invert, Equalize, Contrast, Posterise, AutoContrast, ShearX, ShearY。强数据增强对样本的破坏程度高于弱数据增强,导致增强后的视图丢失原始样本的较多特征信息,产生与弱数据增强不同的

视图模式。为了与弱数据增强形成差异性以及限制强数据增强对样本的破坏程度,本文的强数据增强由5种基本类型的数据增强方法组成,共有 $C(14,5)$ 种形态应用在数据集上。强数据增强策略和弱数据增强策略对样本的影响可视化具体如图2所示。强数据增强样本称为强视图,如图2(b);弱数据增强样本称为弱视图,如图2(c)。强视图和弱视图明显地体现出强数据增强和弱数据增强的特性。另外从图2(b)中可以看到,某些随机组合的强数据增强对样本的特征破坏大,有些样本的内在结构甚至被完全破坏,如图2(b)右下角样本所示。如果只存在强数据增强,网络无法从强视图模式中学习到样本间的相似性和差异性,导致网络无法收敛,但是强数据增强和弱数据增强的联合能达到预期目标,即利用强视图和弱视图间的模式差异作为监督信息,引导模型学习嵌入在强视图中的特征信息,提高网络的学习能力。



Fig. 2 The comparison of strong data augmentation and weak data augmentation

图2 强数据增强和弱数据增强的比较

2.2 框架总览

SCCFA的网络结构如图3所示,输入样本由锚点样本和邻居样本组成。联合数据增强器(union data augmentor, UDA)对样本进行强和弱2种数据的增强。第1个MLP产生嵌入空间,用于特征学习;第2个MLP产生概率分布空间,用于聚类分析。联合特征对比器(federated feature comparator, FFC)和联合类别对比器(federated category comparator, FCC)分别对应特征对比和类别对比。

假设小批次为 n 的锚点样本集合为 $X = \{x_i\}$, $i \in \{1, 2, \dots, n\}$,邻居样本集合为 $Y = \{y_{ij}\}$, $i \in \{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, k\}$,其中 y_{ij} 表示锚点样本 x_i 的第 j 个邻居样本,邻居样本借助KNN图确定,则输入样本集合为 $X \cup Y$ 。集合 $X \cup Y$ 在UDA下进行强数据增强和弱数据增强,生成强视图 $X^s \cup Y^s$ 和弱视图 $X^w \cup Y^w$ (其中s和

w分别表示强和弱的含义)后输入卷积网络进行特征提取。为便于表达,锚点样本集合和邻居样本集合的强视图、弱视图分别用 X^a , $a \in \{s, w\}$ 和 Y^b , $b \in \{s, w\}$ 表示。特征矩阵生成后,FFC和FCC在KNN图的指导下分别进行特征学习和聚类分析,FFC和FCC这2个模块以正负样本对为纽带,彼此形成较强的相互促进、相辅相成的关系。KNN图的质量好坏直接影响网络性能的高低,因此在训练的过程中,利用迭代的全局特征库更新KNN图,使KNN图的质量随着迭代而改善。

2.3 联合特征对比器

假设锚点样本和邻居样本的视图集合 X^a 和 Y^b 的特征空间分别是 $\mathbf{F}^a \in \mathbb{R}^{n \times m}$, $a \in \{s, w\}$ 和 $\tilde{\mathbf{F}}^b \in \mathbb{R}^{n \times m}$, $b \in \{s, w\}$, m 为卷积网络所提特征的维度,在MLP的非线性变换下, \mathbf{F}^a 和 $\tilde{\mathbf{F}}^b$ 对应的嵌入矩阵分别是 $\mathbf{Z}^a \in \mathbb{R}^{n \times d}$, $a \in \{s, w\}$ 和

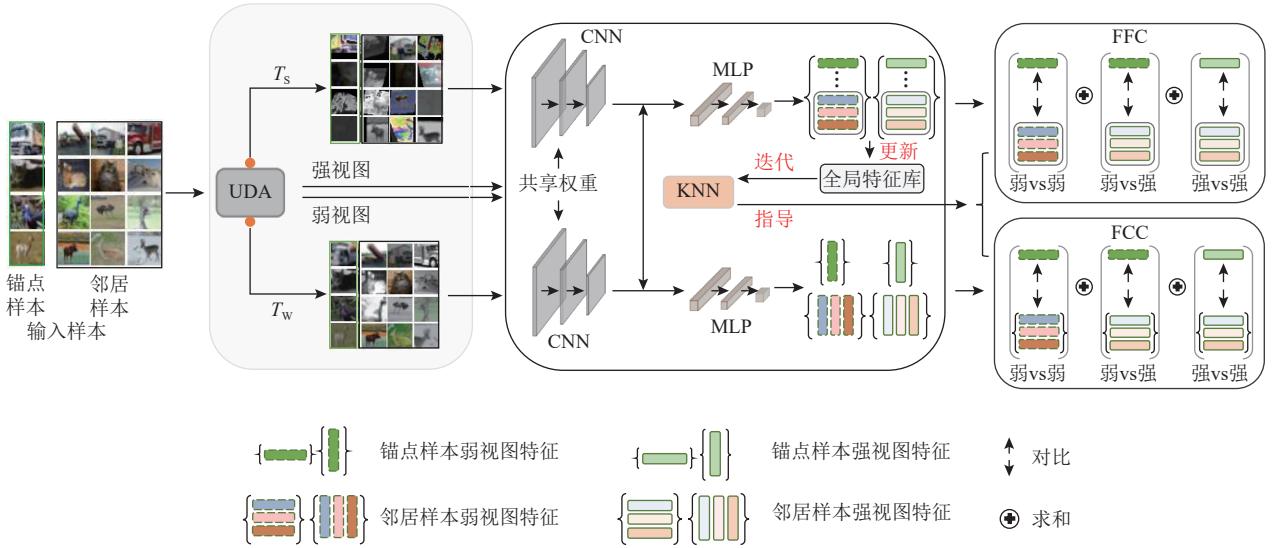


Fig. 3 The network framework of SCCFA

图 3 SCCFA 的网络框架

$\tilde{\mathbf{Z}}^b \in \mathbb{R}^{n \times d}$, $b \in \{s, w\}$, d 是特征维度 m 降维后的大小,

那么 \mathbf{Z}^a 和 $\tilde{\mathbf{Z}}^b$ 的矩阵形式为

$$\mathbf{Z}^a = \begin{pmatrix} \mathbf{z}_1^a \\ \mathbf{z}_2^a \\ \vdots \\ \mathbf{z}_n^a \end{pmatrix}_{n \times d} \text{ 和 } \tilde{\mathbf{Z}}^b = \begin{pmatrix} \tilde{\mathbf{z}}_1^b \\ \tilde{\mathbf{z}}_2^b \\ \vdots \\ \tilde{\mathbf{z}}_n^b \end{pmatrix}_{n \times d}.$$

本文采用余弦相似性度量特征间的相似程度, 相似性函数为:

$$\text{sim}(\mathbf{z}_i^a, \tilde{\mathbf{z}}_j^b) = \frac{\mathbf{z}_i^a \cdot \tilde{\mathbf{z}}_j^b}{\|\mathbf{z}_i^a\| \times \|\tilde{\mathbf{z}}_j^b\|}. \quad (1)$$

为改善聚类准确率而对特征进行归一化^[20], 有 $\|\mathbf{z}_i^a\| = 1$ 和 $\|\tilde{\mathbf{z}}_j^b\| = 1$, 则式(1)转化为

$$\text{sim}(\mathbf{z}_i^a, \tilde{\mathbf{z}}_j^b) = \mathbf{z}_i^a \cdot \tilde{\mathbf{z}}_j^b. \quad (2)$$

再引入温度超参 τ , 相似性函数的最终形式为

$$\text{sim}(\mathbf{z}_i^a, \tilde{\mathbf{z}}_j^b) = \frac{\mathbf{z}_i^a \cdot \tilde{\mathbf{z}}_j^b}{\tau}. \quad (3)$$

采用和 GCC(graph contrastive clustering)^[21]相似的损失函数形式, FFC 的特征对比损失函数定义为:

$$\mathcal{L}_{\text{ins}}^{a,b} = -\frac{1}{n} \sum_{i=1}^n \ln \frac{\sum_{r \in R(i)} e^{\text{sim}(\mathbf{z}_i^a, \tilde{\mathbf{z}}_r^b)}}{\sum_{g \in G(i)} e^{\text{sim}(\mathbf{z}_i^a, \tilde{\mathbf{z}}_g^b)} + \sum_{t \in T(i)} e^{\text{sim}(\mathbf{z}_i^a, \tilde{\mathbf{z}}_t^b)}}, \quad (4)$$

其中 $a, b \in \{s, w\}$, $R(i)$ 表示锚点样本 x_i 的邻居样本, $G(i)$ 表示锚点样本 x_i 所在集合下的负样本, $T(i)$ 表示邻居样本所在集合下的负样本. $R(i)$, $G(i)$, $T(i)$ 的数学表达为:

$$\left\{ \begin{array}{l} R(i) = \{y_j | j \in \{1, 2, \dots, k\}\}, \\ G(i) = \{x_t | t \neq i \wedge t \in \{1, 2, \dots, n\}\}, \\ T(i) = \{y_t | t \neq i \wedge t \in \{1, 2, \dots, n\} \wedge j \in \{1, 2, \dots, k\}\}. \end{array} \right.$$

那么锚点样本与邻居样本间“弱 vs 弱”的损失

函数为

$$\mathcal{L}_{\text{ins}}^{w,w} = -\frac{1}{n} \sum_{i=1}^n \ln \frac{\sum_{r \in R(i)} e^{\text{sim}(\mathbf{z}_i^w, \tilde{\mathbf{z}}_r^w)}}{\sum_{g \in G(i)} e^{\text{sim}(\mathbf{z}_i^w, \tilde{\mathbf{z}}_g^w)} + \sum_{t \in T(i)} e^{\text{sim}(\mathbf{z}_i^w, \tilde{\mathbf{z}}_t^w)}}, \quad (5)$$

同理, “弱 vs 强”和“强 vs 强”的损失分别为:

$$\mathcal{L}_{\text{ins}}^{w,s} = -\frac{1}{n} \sum_{i=1}^n \ln \frac{\sum_{r \in R(i)} e^{\text{sim}(\mathbf{z}_i^w, \tilde{\mathbf{z}}_r^s)}}{\sum_{g \in G(i)} e^{\text{sim}(\mathbf{z}_i^w, \tilde{\mathbf{z}}_g^s)} + \sum_{t \in T(i)} e^{\text{sim}(\mathbf{z}_i^w, \tilde{\mathbf{z}}_t^s)}}, \quad (6)$$

$$\mathcal{L}_{\text{ins}}^{s,s} = -\frac{1}{n} \sum_{i=1}^n \ln \frac{\sum_{r \in R(i)} e^{\text{sim}(\mathbf{z}_i^s, \tilde{\mathbf{z}}_r^s)}}{\sum_{g \in G(i)} e^{\text{sim}(\mathbf{z}_i^s, \tilde{\mathbf{z}}_g^s)} + \sum_{t \in T(i)} e^{\text{sim}(\mathbf{z}_i^s, \tilde{\mathbf{z}}_t^s)}}, \quad (7)$$

因此, FFC 的总的损失函数为

$$\mathcal{L}_{\text{FFC}} = \mathcal{L}_{\text{ins}}^{w,w} + 2 \times \mathcal{L}_{\text{ins}}^{w,s} + \mathcal{L}_{\text{ins}}^{s,s}. \quad (8)$$

2.4 联合类别对比器

假设锚点样本和邻居样本的特征空间 \mathbf{F}^a 和 $\tilde{\mathbf{F}}^b$ 对应的概率分布矩阵分别是 $\mathbf{P}^a \in \mathbb{R}^{n \times c}$, $a \in \{s, w\}$ 和 $\tilde{\mathbf{P}}^b \in \mathbb{R}^{n \times c}$, $b \in \{s, w\}$, 其中 c 是类别数量, \mathbf{P}^a 和 $\tilde{\mathbf{P}}^b$ 的矩阵表现形式为

$$\mathbf{P}^a = \begin{pmatrix} \mathbf{p}_1^a \\ \mathbf{p}_2^a \\ \vdots \\ \mathbf{p}_n^a \end{pmatrix}_{n \times c} \text{ 和 } \tilde{\mathbf{P}}^b = \begin{pmatrix} \tilde{\mathbf{p}}_1^b \\ \tilde{\mathbf{p}}_2^b \\ \vdots \\ \tilde{\mathbf{p}}_n^b \end{pmatrix}_{n \times c}.$$

和 CC^[22] 的思想相似, 本文将标签信息视作一种特殊的特征信息, 为便于计算, 将 \mathbf{P}^a 和 $\tilde{\mathbf{P}}^b$ 转置, 得到 $\mathbf{Q}^a \in \mathbb{R}^{c \times n}$, $\tilde{\mathbf{Q}}^b \in \mathbb{R}^{c \times n}$, $a, b \in \{s, w\}$:

$$\mathbf{Q}^a = (\mathbf{q}_1^a, \mathbf{q}_2^a, \dots, \mathbf{q}_c^a)_{c \times n}^T, \\ \tilde{\mathbf{Q}}^b = (\tilde{\mathbf{q}}_1^b, \tilde{\mathbf{q}}_2^b, \dots, \tilde{\mathbf{q}}_c^b)_{c \times n}^T.$$

FCC 中类别对比损失函数定义为

$$\mathcal{L}_{\text{clus}}^{a,b} = -\frac{1}{c} \sum_{i=1}^c \ln \frac{e^{\text{sim}(q_i^a, \tilde{q}_i^b)}}{\sum_{j=1}^c e^{\text{sim}(q_i^a, \tilde{q}_j^b)}}, a, b \in \{s, w\}. \quad (9)$$

和 FFC 的思路类似, 样本间的“弱 vs 弱”“弱 vs 强”“强 vs 强”的类别对比损失分别依次为:

$$\mathcal{L}_{\text{clus}}^{w,w} = -\frac{1}{c} \sum_{i=1}^c \ln \frac{e^{\text{sim}(q_i^w, \tilde{q}_i^w)}}{\sum_{j=1}^c e^{\text{sim}(q_i^w, \tilde{q}_j^w)}}, \quad (10)$$

$$\mathcal{L}_{\text{clus}}^{w,s} = -\frac{1}{c} \sum_{i=1}^c \ln \frac{e^{\text{sim}(q_i^w, \tilde{q}_i^s)}}{\sum_{j=1}^c e^{\text{sim}(q_i^w, \tilde{q}_j^s)}}, \quad (11)$$

$$\mathcal{L}_{\text{clus}}^{s,s} = -\frac{1}{c} \sum_{i=1}^c \ln \frac{e^{\text{sim}(q_i^s, \tilde{q}_i^s)}}{\sum_{j=1}^c e^{\text{sim}(q_i^s, \tilde{q}_j^s)}}. \quad (12)$$

因此, 总的类别对比损失函数为

$$\mathcal{L}_{\text{clus}} = \mathcal{L}_{\text{clus}}^{w,w} + \mathcal{L}_{\text{clus}}^{w,s} + \mathcal{L}_{\text{clus}}^{s,s}. \quad (13)$$

同时, 为解决聚类分析中稳定性问题, 即避免所有样本划分到同一类别这种极端情况, 本文添加一个额外的聚类约束损失:

$$\mathcal{L}_{\text{cc}} = \ln c - H(\mathbf{I}), \quad (14)$$

其中 $H(\cdot)$ 为熵函数, \mathbf{I} 表示包含 c 个元素的分布. \mathbf{I} 中元素定义为:

$$I_j = \frac{\sum_{t=1}^n q_{jt}^w}{\sum_{i=1}^c \sum_{t=1}^n q_{it}^w}. \quad (15)$$

于是 FCC 的损失函数由类别对比损失和聚类约束损失 2 部分组成:

$$\mathcal{L}_{\text{FCC}} = \mathcal{L}_{\text{clus}} + \lambda \times \mathcal{L}_{\text{cc}}, \quad (16)$$

其中 λ 为平衡超参.

综合 FFC 和 FCC, SCCFA 的目标函数为

$$\mathcal{L} = \alpha \times \mathcal{L}_{\text{FFC}} + \mathcal{L}_{\text{FCC}}, \quad (17)$$

其中 α 为平衡超参.

2.5 伪代码

锚点样本的邻居样本准确率(正样本准确率)取决于 KNN 图的质量. 为创建一个质量较好的 KNN 图, 设计前置任务用于特征学习. 第 1 阶段进行特征对比学习, 同一样本的数据增强视图作为正样本对; 第 2 阶段引入类别对比学习, 构建 KNN 图并不断更新图结构. SCCFA 的训练过程具体如算法 1 所示.

算法 1. SCCFA 算法.

输入: 数据集 X , 最小批次 n , 第 1 阶段次数 t , 迭代次数 $epochs$, 类别数量 c , 特征对比中相似性度量的温度超参 τ_{ins} , 类别对比中相似性度量的温度超参

```

 $\tau_{\text{clus}}$ ;
输出: 逼近最优解的对比聚类模型  $\Phi$ .
/*训练阶段*/
① for epoch = 1 to epochs do
②   从数据集  $X$  中抽样  $n$  个随机样本  $X'$ ;
③   if epoch < t /*第 1 阶段*/
④      $X'$  进行强和弱数据增强处理;
⑤     通过式(8)计算损失  $\mathcal{L}_{\text{FFC}}$ ;
⑥      $\mathcal{L}_{\text{FCC}} = 0$ ;
⑦   else /*第 2 阶段*/
⑧     if epoch = t
⑨       初始化 KNN 图;
⑩   end if
⑪   根据 KNN, 抽样  $X'$  的邻居样本  $Y'$ ;
⑫    $X' \cup Y'$  进行强和弱数据增强;
⑬   通过式(8)计算损失值  $\mathcal{L}_{\text{FFC}}$ ;
⑭   通过式(16)计算损失值  $\mathcal{L}_{\text{FCC}}$ ;
⑮   end if
⑯   通过式(17)计算总体损失  $\mathcal{L}$ ;
⑰   SGD 算法最小化  $\mathcal{L}$  来更新网络参数  $\theta$ ;
⑱   重新计算 KNN 图;
⑲ end for
/*测试阶段*/
① for  $X'$  in  $X$  do
②    $X'$  进行数据增强得  $X''$ ; /*不同于训练
阶段的数据增强方法*/
③    $X''$  输入对比聚类模型  $\Phi$ , 计算软标签;
④ end for

```

2.6 时间复杂度分析

假设算法①第 1 阶段迭代次数为 t_1 , 第 2 阶段迭代次数为 t_2 , 数据集 X 和小批次 X' 大小分别为 N 和 n , 数据集 X 划分为 N/n 批, 近邻样本个数为 k , 类别数量为 c , 嵌入矩阵维度为 d , 且不考虑神经网络前向和反向传播的时间开销, 以及忽略 Faiss 工具包构建 KNN 图的时间开销.

1) 第 1 阶段. 小批次 X' 在每次迭代下, 数据增强的时间复杂度为 $O(n)$, 利用式(8)计算特征对比损失的时间复杂度为 $O(dn^2)$, 因为迭代次数为 t_1 , 每次迭代有 N/n 批, 所以第 1 阶段总的时间复杂度为 $O(t_1 dn N)$.

2) 第 2 阶段. 计算特征对比损失的总时间复杂度是 $O(t_2 kdnN)$, 通过式(16)计算类别对比损失的总时间复杂度为 $O(t_2 c^2 N)$, 所以第 2 阶段总时间复杂度为 $O(t_2 kdnN + t_2 c^2 N)$.

因此算法 1 的总时间复杂度为 $O((t_1 + kt_2)dnN + t_2c^2N)$.

3 实验与分析

3.1 数据集

为充分评估本文方法的有效性和先进性, 在 6 个广泛使用的基准数据集上进行实验以验证本文方法的聚类性能. 表 1 列出了这些数据集的统计信息.

Table 1 Statistics of Datasets

表 1 数据集统计

数据集	图片数量	图片尺寸	类别个数
CIFAR-10	6 万	3×32×32	10
CIFAR-100	6 万	3×32×32	20
STL-10	1.3 万	3×96×96	10
ImageNet-10	1.3 万	3×96×96	10
ImageNet-Dogs	1.95 万	3×96×96	15
Tiny-ImageNet	11 万	3×64×64	200

1) CIFAR-10^[23] 是一个由 10 个类别的共 6 万张 32×32 彩色图片组成的数据集, 每类有 0.6 万张图片.

2) CIFAR-100^[23] 包含 100 个类别, 可归纳为 20 个超类, 每类别含 600 张图片, 其中 500 张用于训练, 100 张用于测试. 本文将这 20 个超类作为该数据集的标签信息.

3) STL-10^[24] 是一个数量为 1.3 万的数据集, 包含额外 10 万张无标签信息的图片, 挑战性更大. 本文将无标签信息的图片舍弃, 只选择有标签信息的 1.3 万张图片用于网络训练.

4) ImageNet-10^[25] 是从 ImageNet^[26] 中挑选 10 个类别的图片组成的数据集, 共包含 1.3 万张图片.

5) ImageNet-Dogs^[25] 与 ImageNet-10 同出于 ImageNet, 共 1.95 万张图片.

6) Tiny-ImageNet^[27] 有 200 个类别, 共 12 万张图片, 其中含有 10 万张训练图片、1 万张验证图片以及 1 万张测试图片, 由于测试图片无标签信息被舍弃, 只有 11 万张图片用于本文方法的训练和测试.

3.2 评价指标

从公平性和全面性的角度出发, 本文采用广泛使用的聚类有效性指标: ACC (accuracy), NMI (normalized mutual information), ARI (adjusted rand index) 评价本文方法在 6 个数据集上的性能.

1) ACC 指模型预测的准确率, 用于衡量预测标签与真实标签之间一对一关系的正确程度, 计算式为:

$$ACC(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n \delta(y_i, \hat{y}_i), \quad (18)$$

其中 Y 表示真实标签, \hat{Y} 是预测标签, $\delta(\cdot)$ 是指示函数, 当 $y_i = \hat{y}_i$ 时, $\delta(\cdot) = 1$, 否则 $\delta(\cdot) = 0$.

2) NMI 用于度量预测标签与真实标签之间的相近程度. 给定数据集 X , 假设真实标签为 Y , 预测标签为 \hat{Y} , NMI 的计算式为

$$NMI(Y, \hat{Y}) = \frac{2 \times MI(Y, \hat{Y})}{H(Y) + H(\hat{Y})}, \quad (19)$$

其中 $H(\cdot)$ 表示熵函数, $H(Y)$ 和 $H(\hat{Y})$ 的定义为:

$$H(Y) = - \sum_{i=1}^{|Y|} P(i) \times \text{lb}(P(i)), \quad (20)$$

$$H(\hat{Y}) = - \sum_{j=1}^{|Y|} P(j) \times \text{lb}(P(j)), \quad (21)$$

其中 $P(\cdot)$ 表示联合概率, 此外, 式(19)中的 $MI(\cdot)$ 指预测值与真实值间的互信息, 定义为

$$MI(Y, \hat{Y}) = \sum_{i=1}^{|Y|} \sum_{j=1}^{|Y|} P(i, j) \times \text{lb} \left(\frac{P(i, j)}{P(i) \times P(j)} \right). \quad (22)$$

3) ARI 为调整兰德指数, 用于评价预测标签与真实标签间的匹配程度, 定义为

$$ARI = \frac{RI - E(RI)}{\max(RI) - E(RI)}, \quad (23)$$

$$RI = \sum_{ij} C_{n_{ij}}^2, \quad (24)$$

其中, $E(\cdot)$ 表示期望.

3.3 实现细节

本文方法采用 Python 语言和 Pytorch 框架实现, 在 Nvidia Geforce RTX 3090 24 GB 服务器上运行.

将残差网络 ResNet-18 用于特征提取, 特征学习和聚类分析 MLP 的深度分别是 2 层和 1 层. 此外, 模型采用 SGD 优化算法, 其学习率、权值衰减以及动量系数分别为 0.4, 1E-4, 0.9. 考虑收敛的稳定性, 学习率以 0.1 的衰减率进行余弦衰减. 模型的超参共有特征对比中相似性度量的温度超参 τ_{ins} 、类别对比中相似性度量的温度超参 τ_{clus} 以及平衡超参 α 和 λ , 它们的初始值分别为 0.1, 1.0, 2.0, 3.0. 开源库 Faiss 计算 6 个数据集的 KNN 图 ($k=5$), 鉴于 Faiss 高效的特性, KNN 图的计算开销可以忽略不计. 由于数据集的图片尺寸、学习难度等各有差异, 再加上物理限制, 第 1 阶段次数 t 、小批次 n 根据数据集的实际情况而定. 最后, 所有数据集上的训练次数为 800.

3.4 实验结果分析

在相同的 6 个数据集上,本文方法分别从 *ACC*, *NMI*, *ARI* 这 3 个方面与其他 17 个聚类方法进行对比,实验结果记录在表 2~4 中. 其他方法数据来源于 GCC^[21].

综合表 2~4 可知,传统聚类算法在高维数据集上的性能远低于深度聚类方法. 此外,基于互信息的方法比基于重构的方法取得了更优的聚类性能,比如 IIC^[14] 和 AE(autoencoder)^[30]. 对比聚类方法学习到强区分性的特征,并通过正负样本强化了特征学习和聚类分析间相互促进的关系,进一步提升了聚类效果,如 CC^[22] 比 IIC 有接近 10% 的提升. 同时, GCC 与 CC 的对比说明对比学习应用在下游任务时,样本对构建策略应与下游任务实际结合. 更重要的是,相较于已有的对比聚类方法,如 PICA^[15], GCC, SCCFA 在除了 Tiny-ImageNet 数据集外的其他 5 个基准数据集上都取得了明显的性能提升,在 ImageNet-10 数据集上 *ACC* 和 *ARI* 的提升分别有 6 个百分点和 10 个百分点,且所有数据集上的训练次数缩减了 20 个百分点. 由此可见,联合数据增强技术和 KNN 算法有效提升了模型的学习能力和学习效率. 在 Tiny-ImageNet 数据集上只取得次优结果的可能原因有:

1) Tiny-ImageNet 数据集有 200 个类别,相对其他数据集拥有更多的语义类别,此时,假负样本对模型性能的影响更大^[36],干扰模型学习不同类簇之间的差异性,然而本文方法未对假负样本作进一步处理.

2) Tiny-ImageNet 数据集的正样本对准确率相对其他数据集较低,说明假正样本较多,假正样本会干扰模型对同类样本间一致性的学习.

3.5 联合数据增强影响分析

关于联合数据增强的消融实验结果如表 5 所示.“弱 vs 弱”实验作为对照实验,用以说明弱数据增强尽管保留了原始样本的大部分特征信息,但也继承了语义信息与非语义信息消融交织的特性,在单一的视图模式下,该特性阻碍了模型对语义信息的学习. 从“弱 vs 强”实验可知,由于强数据增强对样本破坏大,生成与弱数据增强不同的视图模式,利用 2 类差异性的视图模式间的对比缓解“弱 vs 弱”存在的不足.“强 vs 强”实验说明在单一的强视图模式下,对比学习范式无法从破坏严重的样本中学习到有效的共有信息. 最后,“弱/强 vs 强/弱”实验是本文方法采用的联合数据增强策略,可以有效解决单个实验的不足,既利用“弱 vs 强”解决“弱 vs 弱”中语义学习能力不足的问题,又利用“弱 vs 弱”缓解“弱 vs 强”

Table 2 ACC Comparison of Different Methods on Six Datasets

表 2 不同方法在 6 个数据集上的 *ACC* 比较

方法	CIFAR-10	CIFAR-100	STL-10	ImageNet-10	ImageNet-Dogs	Tiny-ImageNet
k-means ^[6]	0.229	0.130	0.192	0.241	0.105	0.025
SC ^[5]	0.247	0.136	0.159	0.274	0.111	0.022
AC ^[28]	0.228	0.138	0.332	0.242	0.139	0.027
NMF ^[29]	0.190	0.118	0.180	0.230	0.118	0.029
AE ^[30]	0.314	0.165	0.303	0.317	0.185	0.041
DAE ^[31]	0.297	0.151	0.302	0.304	0.190	0.039
GAN ^[32]	0.315	0.151	0.298	0.346	0.174	0.041
DeCNN ^[33]	0.282	0.133	0.299	0.313	0.175	0.035
VAE ^[34]	0.291	0.152	0.282	0.334	0.179	0.036
JULE ^[35]	0.272	0.137	0.277	0.300	0.138	0.033
DEC ^[8]	0.301	0.185	0.359	0.381	0.195	0.037
DAC ^[25]	0.522	0.238	0.470	0.527	0.275	0.066
DCCM ^[13]	0.623	0.327	0.482	0.710	0.383	0.108
IIC ^[14]	0.617	0.257	0.610			
PICA ^[15]	0.696	0.337	0.713	0.870	0.352	0.098
CC ^[22]	0.766	0.426	0.747	0.895	0.342	0.140
GCC ^[21]	0.856	0.472	0.788	0.901	0.526	0.138
SCCFA (本文)	0.882	0.506	0.802	0.963	0.543	0.132

注: 最好的结果用粗体数字标注.

Table 3 NMI Comparison of Different Methods on Six Datasets
表 3 不同方法在 6 个数据集上的 NMI 比较

方法	CIFAR-10	CIFAR-100	STL-10	ImageNet-10	ImageNet-Dogs	Tiny-ImageNet
k-means ^[6]	0.087	0.084	0.125	0.119	0.055	0.065
SC ^[5]	0.103	0.090	0.098	0.151	0.038	0.063
AC ^[28]	0.105	0.098	0.239	0.138	0.037	0.069
NMF ^[29]	0.081	0.079	0.096	0.132	0.044	0.072
AE ^[30]	0.239	0.100	0.250	0.210	0.104	0.131
DAE ^[31]	0.251	0.111	0.224	0.206	0.104	0.127
GAN ^[32]	0.265	0.120	0.210	0.225	0.121	0.135
DeCNN ^[33]	0.240	0.092	0.227	0.186	0.098	0.111
VAE ^[34]	0.245	0.108	0.200	0.193	0.107	0.113
JULE ^[35]	0.192	0.103	0.182	0.175	0.054	0.102
DEC ^[8]	0.257	0.136	0.276	0.282	0.122	0.115
DAC ^[25]	0.396	0.185	0.366	0.394	0.219	0.190
DCCM ^[13]	0.496	0.285	0.376	0.608	0.321	0.224
PICA ^[15]	0.591	0.310	0.611	0.802	0.352	0.277
CC ^[22]	0.681	0.424	0.674	0.862	0.401	0.340
GCC ^[21]	0.764	0.472	0.684	0.842	0.490	0.347
SCCFA (本文)	0.808	0.511	0.733	0.910	0.525	0.343

注：最好的结果用粗体数字标注。

Table 4 ARI Comparison of Different Methods on Six Datasets
表 4 不同方法在 6 个数据集上的 ARI 比较

方法	CIFAR-10	CIFAR-100	STL-10	ImageNet-10	ImageNet-Dogs	Tiny-ImageNet
k-means ^[6]	0.049	0.028	0.061	0.057	0.020	0.005
SC ^[5]	0.085	0.022	0.048	0.076	0.013	0.004
AC ^[28]	0.065	0.034	0.140	0.067	0.021	0.005
NMF ^[29]	0.034	0.026	0.046	0.065	0.016	0.005
AE ^[30]	0.169	0.048	0.161	0.152	0.073	0.007
DAE ^[31]	0.163	0.046	0.152	0.138	0.078	0.007
GAN ^[32]	0.176	0.045	0.139	0.157	0.078	0.007
DeCNN ^[33]	0.174	0.038	0.162	0.142	0.073	0.006
VAE ^[34]	0.167	0.040	0.146	0.168	0.079	0.006
JULE ^[35]	0.138	0.033	0.164	0.138	0.028	0.006
DEC ^[8]	0.161	0.050	0.186	0.203	0.079	0.007
DAC ^[25]	0.306	0.088	0.257	0.302	0.111	0.017
DCCM ^[13]	0.408	0.173	0.262	0.555	0.182	0.038
PICA ^[15]	0.512	0.171	0.531	0.761	0.201	0.040
CC ^[22]	0.606	0.282	0.606	0.825	0.225	0.071
GCC ^[21]	0.728	0.305	0.631	0.822	0.362	0.075
SCCFA (本文)	0.777	0.347	0.680	0.920	0.368	0.062

注：最好的结果用粗体数字标注。

中非语义信息考虑不足的问题，改善模型的泛化性。

“强 vs 强”的引入是为了和“弱 vs 弱”“弱 vs 强”相

互配合，在保持模型能继续学习的前提下，从多种视

图模式对比中捕获更多的语义信息，改善特征一致性。

Table 5 Different Combinations Between Strong Data Augmentation and Weak Data Augmentation**表 5 强数据增强和弱数据增强间的不同联合**

策略	符号表示	ACC	NMI	ARI
弱 vs 弱	(w, w)	0.841	0.743	0.704
弱 vs 强	(w, s)	0.855	0.766	0.733
强 vs 强	(s, s)	-	-	-
弱/强 vs 强/弱	(w, w) + (w, s) + (s, s)	0.882	0.808	0.777

注: 最好的结果用粗体数字标注; “vs”的2个对象表示数据增强的类型; “-”表示模型无法学习, 聚类失败。

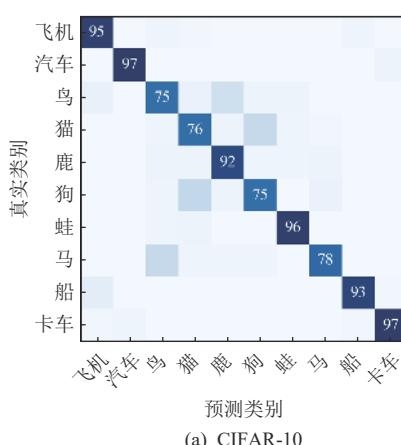
3.6 类别信息影响分析

围绕类别信息展开的消融实验结果如表6所示。第1行实验在特征学习和聚类分析2个模块上都没有引入类别信息, 代表传统的正样本对构建策略, 即将同一样本的不同数据增强视图作为正样本对。FFC没有引入类别信息时, 仅学习语义等价的样本对间的共有信息。FCC没有引入类别信息时, 仅具有将同一样本的不同数据增强视图划分为一类的能力。当FFC和FCC都引入类别信息时, 即在2个模块上将同一类的不同样本作为正样本对。从ACC数据可以看出, 这明显改善了模型的语义感知能力, 同时, ACC的增长幅度反映了FFC和FCC间相互促进、相辅相成的关系。

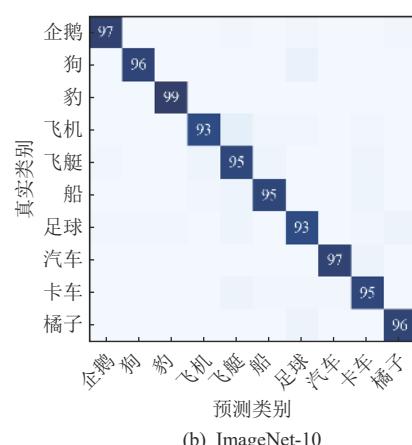
Table 6 Effect of the Global Category Information**表 6 全局类别信息的影响**

FFC	FCC	ACC	NMI	ARI
-	-	0.852	0.768	0.726
√	-	0.863	0.785	0.744
-	√	0.861	0.777	0.735
√	√	0.882	0.808	0.777

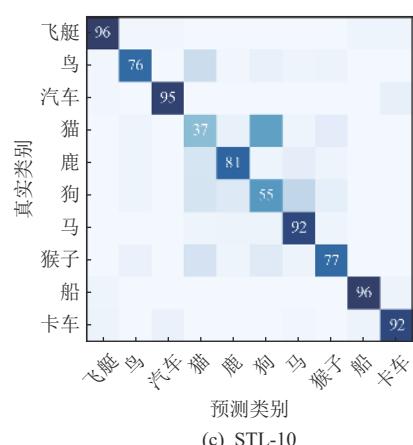
注: 最好的结果用粗体数字标注; “√”表示引入类别信息, “-”表示没有引入类别信息。



(a) CIFAR-10



(b) ImageNet-10



(c) STL-10

Fig. 4 Confusion matrices of three datasets**图 4 3个数据集的混淆矩阵**

4 定性研究

4.1 案例分析

为了知晓本文方法在不同类别图像上的预测能力, 本文引入聚类结果的混淆矩阵, 也通过具体的案例以更加直观的方式分析本文方法在各个类别上的不足。图4是CIFAR-10数据集、ImageNet-10数据集以及STL-10数据集上聚类结果的混淆矩阵, 横坐标和纵坐标分别表示预测类别和真实类别, 矩阵元素数字表示预测准确度。图5是关于CIFAR-10数据集和ImageNet-10数据集上的正样本对案例图, 其中“真正样本”表示与锚点样本同类的样本, “假正样本”表示与锚点样本异类的样本。本文方法借助全局KNN图实现正样本配对, 在标签信息缺失的情况下, 正样本配对存在一定误差, 而正样本对准确与否严重影响本文方法聚类性能的好坏。综合图4和图5可知, 在CIFAR-10数据集上, 狗的预测准确率较低, 而本文方法错误地将猫视作其邻居样本。另外, 本文方法在STL-10数据集上的差异性较大, 猫的准确率只有37%, 而飞机的准确率高达96%, 两者间相差59个百分点。最后, 在ImageNet-10数据集上, 尽管差异性不大, 却错误地将轮船当作飞机的邻居样本、将企鹅作为足球的邻居样本。从ImageNet-10数据集可见, 本文方法利用联合数据增强技术尽管缓解了非语义信息的干扰, 改善模型对语义信息的感知, 仍无法避免将语义信息占比少且非语义信息极其相似的样本划分为一类的现象, 这也是聚类任务中普遍存在的一个问题。另外, 从CIFAR-10和STL-10数据集中猫和狗2个类别的数据可见, 仅从特征相似性匹配正样



Fig. 5 Case analysis of positive pairs on two datasets

图 5 在 2 个数据集上的正样本对案例分析

本可能导致将特征相似、语义不同的 2 个样本错误匹配, 这在一定程度上限制了本文方法的性能.

4.2 特征可视化

为了了解本文方法的特征提取和聚类分析在训练过程中的演变过程, 本文在 CIFAR-10 数据集上借助 t-SNE^[37] 实现特征可视化, 具体为随机从每个类别中选择 500 个样本, 用颜色或数字表示类别信息及其对应的特征, 详见图 6. 从图 6 可见, 类别 1~9 中每个类别的类内紧凑度以及类间的区分度在训练过程中不断改善. 具体而言, 类簇 2, 3, 5 在所有类簇中最难以清楚划分. 从图 6(d)可知, 在训练结束时, 类簇 3, 5 的类内紧凑度很好, 并且类簇 2, 3, 5 之间的区分性

也很明显, 说明本文方法能有效处理高维的数据集.

4.3 收敛性分析

为了分析本文方法结构的合理性, 进行了收敛性分析. 在 4 个数据集上的损失值收敛过程如图 7 所示, 其与图 8 相对应. 图 8 展示了这 4 个数据集上聚类准确率和正样本对准确率的变化过程. 在标签信息缺失情况下, 由于借助 KNN 图寻找正样本存在一定偏误, 本文用正样本对准确率描述该偏误, 正样本对准确率是 top-5 准确率, 即锚点样本的前 5 个伪正样本中真正正样本所占的比例. 从图 7、图 8 可知, 损失值在所有数据集上有效、平稳地收敛, 聚类准确率和正样本对准确率有效地上升; 此外, 整体损失值、

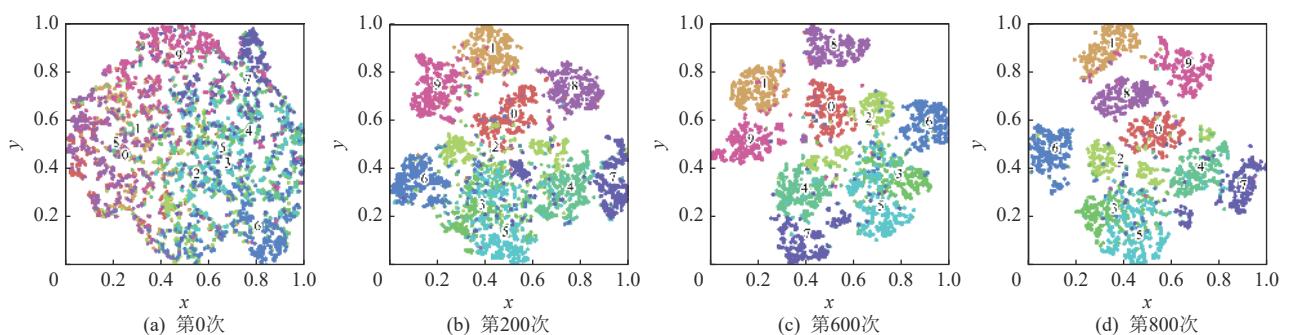


Fig. 6 The evolution of instance features and clustering analysis on CIFAR-10 dataset during iteration process

图 6 迭代过程中在 CIFAR-10 数据集上实例特征和聚类分析的演变

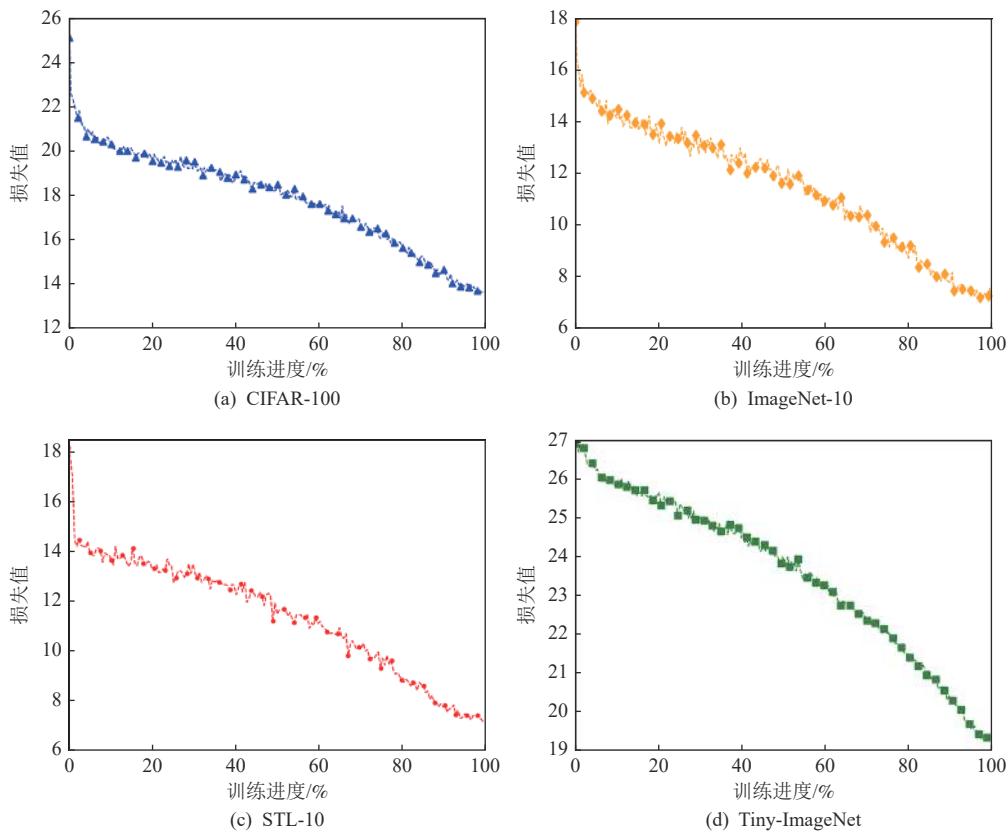


Fig. 7 Loss convergence process of four datasets

图 7 4 个数据集的损失收敛过程

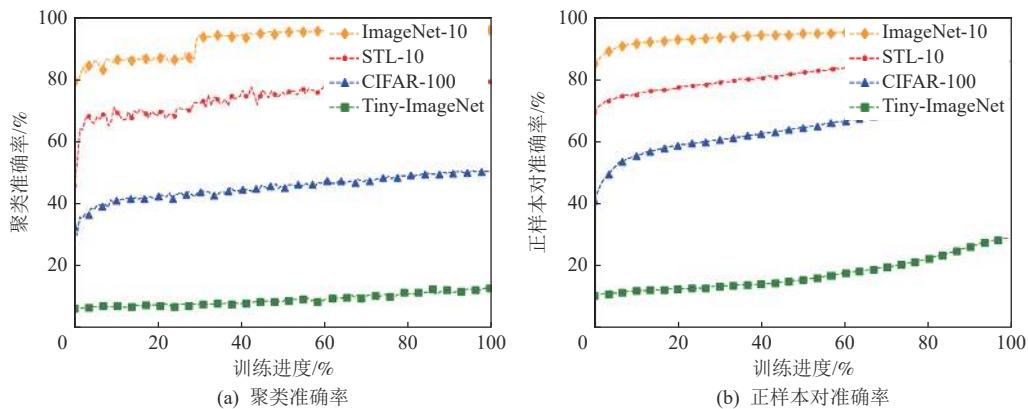


Fig. 8 Variance process of clustering accuracy and positive pairs accuracy on four datasets

图 8 4 个数据集的聚类准确率和正样本对准确率的变化过程

聚类准确率和正样本对准确率三者间具有相互影响的关系。以上分析说明本文方法的结构设计合理，训练稳定；也说明了正样本对的准确率在语义对比学习框架中起着重要的作用。

5 结 论

本文提出了一种新颖的语义对比聚类方法，即联合数据增强的语义对比聚类(SCCFA)。与已有的

方法相比，本文方法在全局 KNN 图的指导下，基于近邻样本属于同类样本的假设，寻找锚点样本的 top- k 近邻样本作为其正样本，改善了特征对比和类别对比的多样性和语义性。另外，本文采用差异化的强和弱 2 类数据增强方法，在提升特征的多样性和泛化性的同时，借助强数据增强和弱数据增强生成的模式差异缓解非语义信息对模型的干扰，提升模型对语义信息的学习效率和学习能力。最后，在 6 个基准数据集上与已有的多个聚类方法进行性能对比，证

明了本文方法的有效性和先进性。

聚类作为计算机视觉研究中的关键技术，本文所提方法尽管解决了对比聚类中的多个不足，但基于KNN图寻找正样本仍存在一些误差，未来将在这方面进行更深入的研究。

作者贡献声明：王气洪负责算法模型的提出和实现、实验结果的整理与分析、论文的撰写和修改；贾洪杰提出研究方向，把握论文的创新性并指导和参与论文的修订；黄龙霞和毛启容讨论算法的改进，设计实验方案，并指导论文修改。

参 考 文 献

- [1] Zhang Xuesong, Jia Caiyan. A new documents clustering method based on frequent item-sets[J]. *Journal of Computer Research and Development*, 2018, 55(1): 102–112(in Chinese)
(张雪松, 贾彩燕. 一种基于频繁词集表示的新文本聚类方法[J]. 计算机研究与发展, 2018, 55(1): 102–112)
- [2] Sun Shenshen, Guo Yang, Ren Huizhi, et al. Juxta-vascular nodule segmentation based on the flowing entropy and geodesic distance feature[J]. *SCIENTIA SINICA Informationis*, 2013, 43(9): 1136–1146(in Chinese)
(孙申申, 郭阳, 任会之, 等. 基于流向特征熵和测地线距离的粘连血管型肺结节聚类分割[J]. *中国科学: 信息科学*, 2013, 43(9): 1136–1146)
- [3] Guo Hongyi, Liu Gongshen, Su Bo, et al. Collaborative filtering recommendation algorithm combining community structure and interest clusters[J]. *Journal of Computer Research and Development*, 2016, 53(8): 1664–1672(in Chinese)
(郭弘毅, 刘功申, 苏波, 等. 融合社区结构和兴趣聚类的协同过滤推荐算法[J]. 计算机研究与发展, 2016, 53(8): 1664–1672)
- [4] Li Hua, Jia Yuheng, Cong Runmin, et al. Superpixel segmentation based on spatially constrained subspace clustering[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(11): 7501–7512
- [5] Zelnik-Manor L, Perona P. Self-tuning spectral clustering [C] //Proc of the 17th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2004: 1601–1608
- [6] MacQueen J. Some methods for classification and analysis of multivariate observations [C] //Proc of the 5th Berkeley Symp on Mathematical Statistics and Probability. Oakland, CA: University of California, 1967: 281–297
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444
- [8] Xie Junyuan, Girshick R, Farhadi A. Unsupervised deep embedding for clustering analysis [C] //Proc of the 33rd Int Conf on Machine Learning. New York: ACM, 2016: 740–749
- [9] Tu Wenxuan, Zhou Sihang, Liu Xinwang, et al. Deep fusion clustering network [C] //Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 9978–9987
- [10] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping [C] //Proc of the 24th IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 1735–1742
- [11] He Kaiming, Fan Haoqi, Wu Yuxin, et al. Momentum contrast for unsupervised visual representation learning [C] //Proc of the 38th IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2020: 9726–9735
- [12] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations [C] //Proc of the 37th Int Conf on Machine Learning. New York: ACM, 2020: 1575–1585
- [13] Wu Jianlong, Long Keyu, Wang Fei, et al. Deep comprehensive correlation mining for image clustering [C] //Proc of the 17th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 8149–8158
- [14] Ji Xu, Vedaldi A, Henriques J. Invariant information clustering for unsupervised image classification and segmentation [C] //Proc of the 17th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 9864–9873
- [15] Huang Jiabo, Gong Shaogang, Zhu Xiatian. Deep semantic clustering by partition confidence maximisation [C] //Proc of the 38th IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2020: 8846–8855
- [16] Chen Xinlei, Fan Haoqi, Girshick R, et al. Improved baselines with momentum contrastive learning [J]. arXiv preprint, arXiv: 2003.04297, 2020
- [17] Zhao Xuyang, Du Tianqi, Wang Yisen, et al. ArCL: Enhancing contrastive learning with augmentation-robust representations [C/OL] // Proc of the 11th Int Conf on Learning Representations. San Juan, CA: ICLR, 2023[2023-05-25].<https://arxiv.org/abs/2303.01092>
- [18] Cubuk E, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space [C] //Proc of the 38th IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2020: 3008–3017
- [19] Wang Xiao, Qi Guojun. Contrastive learning with stronger augmentations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(5): 5549–5560
- [20] Khosla P, Teterwak P, Wang Chen, et al. Supervised contrastive learning [C] //Proc of the 34th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2020: 18661–18673
- [21] Zhong Huasong, Wu Jianlong, Chen Chong, et al. Graph contrastive clustering [C] //Proc of the 18th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 9204–9213
- [22] Li Yunfan, Hu Peng, Liu Zitao, et al. Contrastive clustering [C] //Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 8547–8555
- [23] Tay Y, Dehghani M, Bahri D, et al. Efficient transformers: A survey[J]. *ACM Computing Surveys*, 2022, 55(6): 1–28
- [24] Coates A, Lee H, Ng A. An analysis of single-layer networks in unsupervised feature learning [C] //Proc of the 14th Int Conf on Artificial Intelligence and Statistics. Brookline, MA: Microtome Publishing, 2011: 215–223

- [25] Chang Jianlong, Wang Lingfeng, Meng Gaofeng, et al. Deep adaptive image clustering [C] //Proc of the 16th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 5880–5888
- [26] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks [J] Communications of the ACM, 2017, 60(6): 84–90
- [27] Geng Chuanxing, Huang Shengjun, Chen Songcan. Recent advances in open set recognition: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3614–3631
- [28] Gowda K, Krishna G. Agglomerative clustering using the concept of mutual nearest neighbourhood[J]. *Pattern Recognition*, 1978, 10(2): 105–112
- [29] Cai Deng, He Xiaofei, Wang Xuanhui, et al. Locality preserving nonnegative matrix factorization [C] //Proc of the 21st Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2009: 1010–1015
- [30] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks [C] //Proc of the 19th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2006: 153–160
- [31] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. *Journal of Machine Learning Research*, 2010, 11(12): 3371–3408
- [32] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [C/OL] //Proc of the 4th Int Conf on Learning Representations. San Juan, CA: ICLR, 2016 [2023-05-25].<https://arxiv.org/abs/1511.06434>
- [33] Zeiler M, Krishnan D, Taylor G, et al. Deconvolutional networks [C] // Proc of the 28th IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2010: 2528–2535
- [34] Kingma D, Welling M. Auto-encoding variational Bayes [C/OL] //Proc of the 2nd Int Conf on Learning Representations. San Juan, CA: ICLR, 2014 [2023-05-25].<https://arxiv.org/abs/1312.6114>
- [35] Yang Jianwei, Parikh D, Batra D. Joint unsupervised learning of deep representations and image clusters [C] //Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2016: 5147–5156
- [36] Chen Tsaishien, Hung Weichih, Tseng Hungyu, et al. Incremental false negative detection for contrastive learning [C/OL] //Proc of the 10th Int Conf on Learning Representations. San Juan, CA: ICLR, 2022 [2023-05-25].<https://arxiv.org/abs/2106.03719>
- [37] Van Der Maaten L, Hinton G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9(86): 2579–2605



Wang Qihong, born in 1999. Master candidate. His main research interest includes deep clustering. 王气洪, 1999年生. 硕士研究生. 主要研究方向为深度聚类.



Jia Hongjie, born in 1988. PhD, associate professor. His main research interests include data clustering, deep learning, and pattern recognition. 贾洪杰, 1988年生. 博士, 副教授. 主要研究方向为数据聚类、深度学习、模式识别.



Huang Longxia, born in 1991. PhD, associate professor. Her main research interests include privacy protection, data sharing, and cloud computing. (hlxia@ujs.edu.cn) 黄龙霞, 1991年生. 博士, 副教授. 主要研究方向为隐私保护、数据共享、云计算.



Mao Qirong, born in 1975. PhD, professor, PhD supervisor. Member of IEEE and ACM. Her main research interests include affective computing, pattern recognition, and multimedia analysis. (mao_qr@ujs.edu.cn) 毛启容, 1975年生. 博士, 教授, 博士生导师. IEEE 和 ACM 会员. 主要研究方向为情感计算、模式识别、多媒体分析.