

引用格式:王宏辉,殷进勇,杨建.基于记忆提炼的对比度量增强在线类增量学习图像分类方法[J].指挥控制与仿真,2024,46(1):44-54.WANG H H, YIN J Y, YANG J. Contrast metric enhancement based on memory extraction for online class-incremental learning in image classification[J]. Command Control & Simulation, 2024, 46(1): 44-54.

基于记忆提炼的对比度量增强 在线类增量学习图像分类方法

王宏辉, 殷进勇, 杨建

(江苏自动化研究所, 江苏 连云港 222061)

摘要:图像分类中类增量学习具有知识灾难性遗忘现象,现有的基于经验回放方法着重考虑的是记忆库的更新和采样方式,忽略了新旧样本之间的特征关系。为此,提出了一种基于记忆提炼的对比度量增强在线类增量学习图像分类方法(cME²),设计了两种新的正负样本对,对旧样本信息进行了加强重复再利用,强化了模型对冗余特征和共性特征的表达能力,基于最近邻均值分类器改善了嵌入空间中的样本分布合理性。最后,通过对比实验和消融实验验证了所提方法的有效性和高效性。

关键词:在线类增量学习; 灾难性遗忘; 对比学习; 经验回放

中图分类号:TP18

文献标志码:A

DOI:10.3969/j.issn.1673-3819.2024.01.006

Contrast metric enhancement based on memory extraction for online class-incremental learning in image classification

WANG Honghui, YIN Jinyong, YANG Jian

(Jiangsu Automation Research Institute, Lianyungang 222061, China)

Abstract: In view of the catastrophic forgetting of previous knowledge in class incremental learning for image classification, existing replay-based methods focus on memory updating and sampling, while overlooking the feature relationships between old and new samples. To this end, the paper proposes a method called contrast metric enhancement based on memory extraction (cME²) for Online Class-incremental Learning in Image Classification, which designs two new types of positive and negative sample-pairs, enhances the reuse of old sample feature information, and strengthens the ability of model to express redundant features and common features. It improves the distribution of samples in embedding space based on the nearest class mean classifier. Finally, the effectiveness and efficiency of the proposed method are verified by comparison experiment and ablation experiment.

Key words: online class-incremental learning; catastrophic forgetting; contrastive learning; experience replay

深度神经网络作为数据驱动模型取得了巨大的成果,随着近年来各种数据的爆炸式增长,模型的迭代更新需求也得到了相应的提升。然而,深度神经网络在接受新任务训练时,对旧任务性能的表现通常会迅速降低,这一现象被称为灾难性遗忘^[1]。为了使神经网络模型能够持续学习新知识而不会遗忘已有知识,在线类增量学习^[2]问题得到了广泛的研究。

目前,解决灾难性遗忘问题的增量学习方法主要可以分为正则化^[1,3-8]、参数隔离^[9-10]、重演回放^[11-16]三大类^[2]。正则化技术主要是在训练时施加惩罚限制以引导模型的更新;参数隔离方法的典型特征为不断增加单独的模型参数以适应新类别;重演回放方法则研究如何

更有效地保存和提取部分原始样本或伪样本以进行联合训练。由于重演回放方法保留了更多的历史数据,其综合分类性能往往最佳^[12]。为了进一步解决灾难性遗忘问题,可以对储存的历史数据进行增强利用,Mai Z等^[17]提出有监督对比回放(SCR)方法充分发挥最近邻均值分类器(NCM)的聚类性能,Zhang等^[18]提出重复增强回放(RAR)方法讨论了回放样本时模型欠拟合—过拟合问题,Lin H等^[19]提出基于代理的对比回放(PCR)方法来解决样本不平衡问题以缓解灾难性遗忘。基于这类思想,本文提出了一种基于记忆提炼的对比度量增强(Contrast Metric Enhancement Based on Memory Extraction, cME²)在线类增量学习图像分类方法,重新审视了对比学习中的对比关系,提出了两类新的正负样本对,改进了增量学习过程中的目标损失函数。同时,通过多种对比实验与消融实验验证了所提方法的有效性及其高效性,并分析了不同样本对之间的权重关系,不同采样批次大小对模型训练精度与时间的影响。

收稿日期: 2023-07-04

修回日期: 2023-09-01

作者简介: 王宏辉(1999—),男,硕士研究生,研究方向为增量学习。

殷进勇(1978—),男,博士,研究员。

1 相关工作

1.1 基于记忆提炼的方法

目前,大多数基于重演的方法仅关注旧样本的存储和采样,而未对旧样本进行特殊处理,只是简单地与新样本混合训练。为解决这一问题,TEM^[12]方法通过重复使用过去任务的信息来解决重演回放过程中模型拟合旧任务不足的情况。尽管对同一数据集重复训练容易导致过拟合^[20],但该方法指出,在增量学习中的新任务样本可以作为数据依赖的正则化因子,在此基础上融合旧任务样本进行联合训练,既可以重复利用旧任务信息,又不损失模型泛化能力。

基于这种重复利用的思想,本文提出了一种记忆提炼(Memory Extraction, ME)方法,将每批次训练样本已计算好的对比特征相似度信息进行提炼,通过反复利用现有的计算结果,在记忆区容量有限的情况下,增强了旧知识的利用率,从而降低模型学习新任务时对旧知识的遗忘。

1.2 基于对比度量增强的方法

基于正则化的方法主要通过向学习目标中添加正则项来约束模型学习。一些方法在学习新任务时,将梯度投影到子空间,从而约束梯度的更新方向,减少对先前任务的干扰^[6-8],这需要足够的先验知识且不具有通用性。另一类方法是通过计算参数重要性,限制与旧任务高度相关的参数更新,来缓解旧知识的遗忘^[1,4-5],但这样盲目计算所有参数会使计算量过大。虽然一些方法通过注意力机制分别关注新旧不同任务,学习表示注意力分布的关注度向量,对损失函数添加约束项,同时优化新旧任务损失,提高模型的泛化能力^[13,21]。然而,这些正则化增量学习方法在类增量学习设定下通常不能很好地工作^[22]。

因此,SCR方法提出使用NCM分类器替换Softmax分类器,避免了对新类的全连接层结构变化,并基于对比学习思想正则化约束模型训练,在在线类增量学习场景下达到了较好的性能^[17]。然而,该方法没有充分考虑所有样本间的对比关系,基于此,本文提出了一种对比度量增强(Contrast Metric Enhancement, cME)方法,在有监督对比学习^[23]的基础上添加了两个具有不同意义的正则化约束项,改进了目标损失函数,进一步将“相似”样本聚类,将“不同”样本分散,优化了样本在嵌入空间中的分布方式,提高了模型的泛化能力。

2 基于融合对比损失的类增量在线学习

2.1 问题描述

在线类增量学习旨在从流数据 $x_i \in X$ 中持续学习

不同类 $y_i \in Y$ 。为了研究方便,我们假设输入数据 x_i 具有确定性类 $y_i = h(x)$ 标签,其中, $y_i \in Y = \{1, 2, \dots, K\}$ 。通过将所有数据流划分为一系列的增量学习任务 $c_i \in C_{task}$,则对于含有 K 类标签的数据集,其特定学习任务的类数量应为 $k = |Y|/|C_{task}|$,且不同任务 c_i 在类中没有重叠部分。假设增量任务序列为 $\{c_1, c_2, \dots, c_n\}$,其中 $n = |C_{task}|$,在特定 c_i 任务阶段所接收到的新任务数据集为 $D_i = \{(X_i \times Y_i)_{i=1}^n, c_i\}$,由于不同任务阶段数据集之间所含类别互斥,故 D_i 中只含有 c_i 任务阶段的 n 个带标签样本 $(X_i \times Y_i)_{i=1}^n$ 。此时,保留在经验回放区中的旧任务数据集为前 $i-1$ 个任务阶段储存的少部分样本子集 $D_{old} \subseteq \{D_1, D_2, \dots, D_{i-1}\}$ 。故在整个类增量学习的过程中,当前任务模型的训练阶段只能使用数据集 $D_{cur} = D_i \cup D_{old}$,且 D_i 中的流数据 (x_i, y_i) 只能看到一次,除非该样本被保留在 D_{old} 中进行经验回放。

2.2 总体框图

为充分利用在线类增量学习过程中每个稍纵即逝的样本,本文提出了一种基于记忆提炼的对比度量增强(cME²)在线类增量学习图像分类方法,如图1所示。本文基于对比监督重放的类增量学习基础上,对存放在记忆缓冲区和当前任务数据集中的数据进行了多次特征增强利用,通过提炼训练阶段可用样本所涵盖的重要特征信息,改进了现有的有监督对比学习方法,提出了特定正负样本对的设置方法,在模型训练过程中有意识地扩充同类但异体、个体和增强样本之间的度量距离,改进了样本特征在嵌入空间中的分布方式,有效地提高了模型的整体性能。

2.3 基于记忆提炼的方法

由于在线类增量学习在整个学习过程中无法提供所有类别的、大量的训练数据集,导致模型通过当前任务数据集训练出的参数信息会覆盖之前已掌握的旧类别分类信息,从而造成灾难性遗忘问题。基于重演回放的方法则通过保留旧类别的训练数据集,让模型尽可能地获取所有类别的数据集信息,从而减少对旧知识的遗忘。因此,在基于重演回放的各种不同方法之间,对旧样本信息的利用方式在很大程度上影响了增量学习算法的性能。为此,本文提出了记忆提炼(Memory Extraction, ME)的方法,实现在有限的内存条件下,对记忆区存放的样本特征相似度信息进行了加强重复再利用,在没有额外增加大量计算量和训练时间的前提下,提高了增量学习算法的精度。

为了方便表示训练数据集中的不同正负样本对,可以将 D_{cur} 中存在的类别定义为 $\{A, B, C, \dots\}$,每个类

别中的个体则定义为 $A = \{a_1, a_2, \dots\}$ 、 $B = \{b_1, b_2, \dots\}$ 等,此时 $D_{cur} = \{a_1, a_2, \dots, b_1, b_2, \dots\}$ 。通过使用裁剪缩放、翻转、颜色扰动和灰度转换等数据增强方法,对每个样本特征进行扩充,可以得到对应的增强

样本 $D_{cur}^+ = \{a_1^+, a_2^+, \dots, b_1^+, b_2^+, \dots\}$ 。对样本 x_i 而言,假设其对应的类别标签为 A 类,则该样本可表示为 a_i ,相对于此样本的其他类则为 $\bar{A} = \{A, B, C, \dots\} / A$ 。

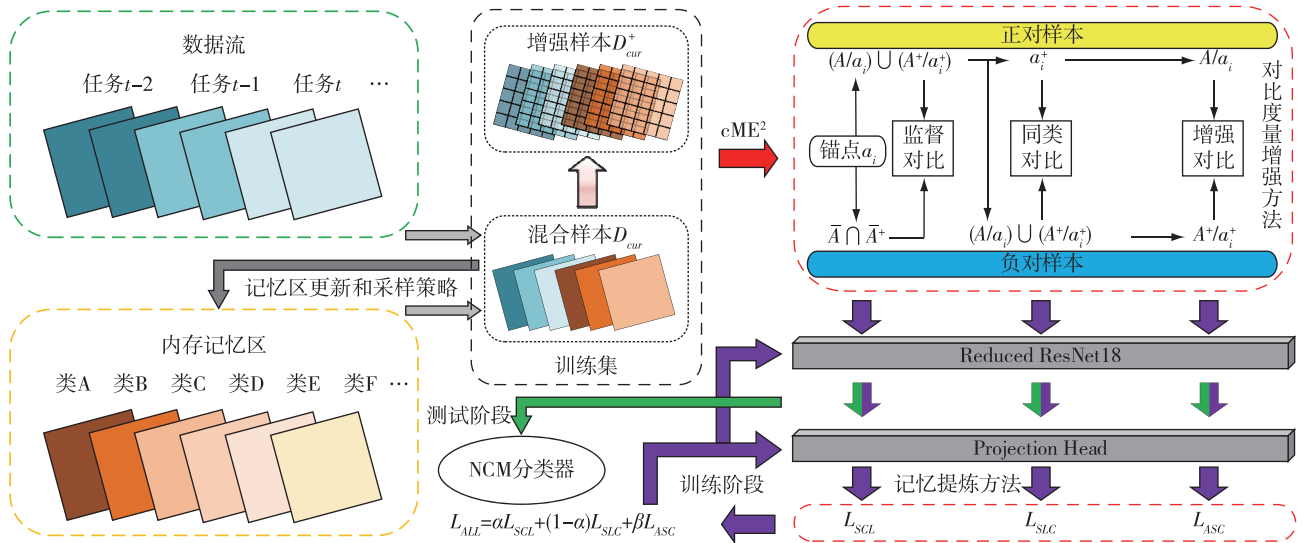


图 1 基于记忆提炼的对比度量增强的在线类增量学习框架

Fig. 1 Online class incremental learning framework of the cME²

表 1 三种不同对比方式的正负样本对设置

Tab. 1 Three different comparison methods of positive and negative sample pairs

对比方式	锚点	正对集合	负对集合
监督对比 L_{SCL}	a_i	$(A/a_i) \cup (A^+ / a_i^+)$	$\bar{A} \cap \bar{A}^+$
同类对比 L_{SLC}	a_i	a_i^+	$(A/a_i) \cup (A^+ / a_i^+)$
增强对比 L_{ASC}	a_i	A/a_i	A^+ / a_i^+

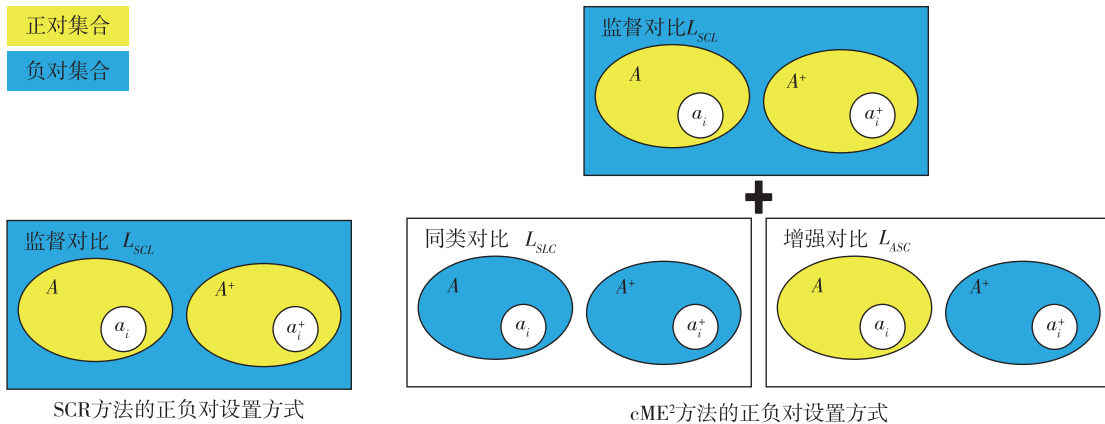


图 2 不同对比方式的正负对集合维恩图

Fig. 2 Positive and negative pair set venn diagrams with different comparison methods

基于上述定义,通过表 1 表述的三种不同正负样本对,用监督对比损失^[23] (Supervised Contrastive Learning Loss, L_{SCL}) 将相似的样本点(正对)推到一起,并拉开不同的样本点(负对)来产生数据的特征表示。在此基础上,根据类增量学习需要学习一系列任务的

特殊问题设定,基于记忆提炼的对比度量增强在线类增量学习图像分类方法基于信息熵对比估计^[24] (InfoNCE Loss, L_{NCE}) 的思想,在现有特征信息的基础上,进一步考虑其他特殊正负样本对之间的对比关系:
①同类间对比(Same Label Comparison Loss, L_{SLC}):通过

同一个类别间的不同样本对比,训练模型可以关注到聚类簇中的每个个体样本,避免模型的特征提取器只能获得某一个类别的粗糙特征表示,增强模型的泛化能力;②本体与增强样本之间的对比(Augmentation Sample Comparison Loss, L_{ASC}):由于本体样本与增强样本的高级语义特征基本相同,但二者之间在低级特征如颜色、纹理、形状等与原数据仍存在一定差异,会产生新的特征表示,故通过 L_{ASC} 可以引导模型认知本体样本与增强样本之间的微小差异,增强模型的鲁棒性。

图2所示的集合维恩图,可以通过可视化的方法清晰地发现这三种正负对设置没有产生额外的计算量,深入地挖掘了每个样本之间的特征信息,实现了对旧信息的重复增强再利用。

2.4 一种改进的对比损失函数

在使用 NCM 分类器聚类方法的情况下,需要着重考虑样本在嵌入空间中的分布情况,这直接关系到样本特征编码聚类后的分类结果,而且基于度量学习的嵌入空间网络能够较好地保持已有知识^[25]。因此,研究一种适合增量学习过程的特征提取方式,合理调整样本的聚类分布,可以提高模型的分类能力并降低模型在增量学习时的遗忘率。受有监督对比回放^[23]方法的启发,通过强调不同类别样本之间的差异性,可以让样本在嵌入空间中获得较好的特征分布。为了改善

这种特征分布方式,合理调整样本间差异对模型参数的影响,本文提出一种对比度量增强(Contrast Metric Enhancement, cME)方法,通过表1中额外提供的同类对比和增强对比两种对比方式,可以优化模型的训练方式,提取样本间的冗余特征,完善样本特征编码聚类方式,改善每个样本在嵌入空间中的分布情况,让模型在增量学习的过程中得到性能提升。

为了方便表示改进之后的对比损失函数,定义如下:给定当前任务的某个训练批次含有 $b \in D_{cur}$ 个样本,加上每个样本对应的增强视角样本 $b^+ \in D_{cur}^+$,批次 B_l 共有 $2b$ 个训练样本,其中 l 作为该批次所有样本的索引集。每个样本 x_i 通过编码器得到 $\text{Enc}(x_i)$ 用于下游 NCM 分类任务,再将其输入投影网络可得到 $z_i = \text{Proj}(\text{Enc}(x_i))$,故此时特征投影输出为 $Z_l = \{z_i\}_{i \in l} = \{\text{Proj}(\text{Enc}(x_i))\}_{i \in l}$ 。对于锚点 x_i 而言,与其标签一致的本体样本和增强样本可以分别表示为 $P(i) \equiv \{p \in D_{cur} \setminus x_i : y_p = y_i\}$, $P^+(i) \equiv \{p^+ \in D_{cur}^+ \setminus x_i^+ : y_{p^+} = y_i\}$;同理,与锚点的类别不一致的本体样本为 $Q(i) \equiv \{q \in D_{cur} : y_q \neq y_i\}$,对应的增强样本为 $Q^+(i) \equiv \{q^+ \in D_{cur}^+ : y_{q^+} \neq y_i\}$ 。最后,定义 $\delta(x_i, x_j) = \exp((z_i \cdot z_j)/\tau)$ 表示两个样本之间的相似度或差异,其中超参数 τ 是对比学习中用来控制正负对分离程度的温度系数。

$$L_{SCL}(Z_l) = \sum_{i \in l} \frac{-1}{|P(i)| + |P^+(i)|} \sum_{\{p \in P(i)\} \cup \{p^+ \in P^+(i)\}} \log \frac{\delta(x_i, x_p + x_p^+)}{\delta(x_i, x_p + x_p^+) + \sum_{\{x_q \in Q(i)\} \cup \{x_q^+ \in Q^+(i)\}} \delta(x_i, x_q + x_q^+)} \quad (1)$$

$$L_{SLC}(Z_l) = -\log \frac{\delta(x_i, x_i^+)}{\sum_{\{p \in P(i)\} \cup \{p^+ \in P^+(i)\}} \delta(x_i, x_p + x_p^+)} \quad (2)$$

$$L_{ASC}(Z_l) = \sum_{i \in l} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\delta(x_i, x_p)}{\sum_{x_p^+ \in P^+(i)} \delta(x_i, x_p^+)} \quad (3)$$

$$L_{ALL} = \alpha L_{SCL} + (1 - \alpha) L_{SLC} + \beta L_{ASC} \quad (4)$$

通过上述符号定义,有监督对比学习损失^[23]如公式(1)所示。 L_{SCL} 训练的模型可以有效地区分同类与异类样本,但没有深入地探讨每个聚类簇中的样本分布情况。为了让模型更好地适应在线类增量学习过程,本文进一步研究了不同样本对表1之间的距离分布关系对模型性能的影响,调整了不同样本对在嵌入空间中的距离位置。公式(2)通过 L_{SLC} 表示自身样本与同类其他样本之间的差异大小,公式(3)通过 L_{ASC} 表示了自身样本与增强样本之间的低级特征区别。

结合上述三种对比损失函数,通过加权和的方法可以得到本文训练模型时的损失函数 L_{ALL} 如公式(4)所示。本文分析了三种对比方式的不同目的,明确了

模型首先要区分的对象为不同类之间,其次是类间个体,最后才是增强个体。因此,超参数系数 $\alpha \in [0.8, 1]$, $\beta \in [0.0001, 0.04]$ 时,模型的性能表现较好。

从公式(1)~(4)中可以发现,本文所提出来的对比损失函数无须额外计算其他的样本间相似度,只需要使用 L_{SCL} 训练过程中产生的计算结果进行增强再利用,实现了在没有增加大量内存消耗及计算量的同时,改进了训练模型的损失函数。

基于记忆提炼的对比度量增强(cME²)方法的系统框架如图1所示。为了学习一系列增量学习任务 C_{task} ,模型在训练阶段,通过当前任务的流数据 D_t 接收一个小批量数据 b_{cur} ,通过 b_{cur} 和重演回放区储存的数

据 D_{old} 提取另一个包含旧样本的小批量数据 b_{mem} 。将两批次数据混合得到当前任务的训练数据集 B_{cur} , 并通过数据增强手段 $Aug(\cdot)$ 扩充训练数据集后, 由编码器 $Enc_{\theta}(\cdot)$ 和投影网络 $Proj_{\phi}(\cdot)$ 得到所有的特征信息。利用 cME² 方法中对特征信息加强重复再利用的方法, 提出额外的两种正负样本对设置进行对比度量增强, 并结合有监督对比学习得到最终的加权对比损

失函数。在使用随机梯度下降方法训练编码器 $Enc_{\theta}(\cdot)$ 和投影网络 $Proj_{\phi}(\cdot)$ 中的模型参数后, 更新记忆区存放的样本。在推理测试阶段, 样本则只需要通过编码器 $Enc_{\theta}(\cdot)$ 得到对应的嵌入空间分布, 最后利用 NCM 分类器预测分类的结果。算法 1 总结了算法的整个训练和推理过程。

算法 1: 基于记忆提炼的对比度量增强 (cME²) 方法

```

初始化: 数据增强  $Aug(\cdot)$ ; 编码器  $Enc_{\theta}(\cdot)$ ; 投影网络  $Proj_{\phi}(\cdot)$ ;
以及对比样本集合设置  $P(\cdot), P^+(\cdot), Q(\cdot), Q^+(\cdot)$ , 及相似度函数  $\delta(\cdot, \cdot)$ 
1. for  $c_i \in C_{task} = \{c_1, c_2, \dots, c_n\}$  do
2.   训练阶段:
3.   for  $b_{cur} \sim D_i$  do
4.      $b_{mem} \leftarrow MemoryRetrieval(b_{cur}, D_{old})$ 
5.      $B_{cur} \leftarrow b_{cur} \cup b_{mem}$ 
6.      $B_I \leftarrow B_{cur} \cup Aug(B_{cur})$ 
7.      $Z_I \leftarrow Proj_{\phi}(Enc_{\theta}(B_I))$ 
8.     for  $z_i \sim Z_I$  do
9.        $\hat{Z}_I = P(i) \cup P^+(i) \cup Q(i) \cup Q^+(i) \cup \{z_i^+\} \leftarrow Z_I$ 
10.      for  $z_j \sim \hat{Z}_I$  do
11.         $\delta(x_i, x_j) \leftarrow Z_i, Z_j$ 
12.         $L_{SCL} \leftarrow \delta(x_i, x_j) \in P(i) \cup P^+(i) \cup Q(i) \cup Q^+(i) \cup \{z_i^+\}$ 
13.         $L_{SLC} \leftarrow \delta(x_i, x_j) \in P(i) \cup \{z_i^+\}$ 
14.         $L_{ASC} \leftarrow \delta(x_i, x_j) \in P(i) \cup P^+(i)$ 
15.         $L_{ALL} \leftarrow aL_{SCL} + (1-a)L_{SLC} + \beta L_{ASC}$ 
16.       $\theta, \phi \leftarrow SGD(L_{ALL}(Z_I), \theta, \phi)$ 
17.       $D_{old} \leftarrow MemoryUpdate(b_{cur}, D_{old})$ 
18.   推理阶段:
19.   for  $l \in \{1, 2, \dots, k * c_i\}$  do //l is number of observed classes
20.      $\mu_l = \frac{1}{n_l} \sum_i^{|\mathcal{M}|} Enc_{\theta}(x_i) \cdot 1\{y_i = l\}$ 
21.      $y^* = \operatorname{argmin}_{l=1, \dots, t} \| Enc_{\theta}(x) - \mu_l \|$ 

```

3 实验及分析

3.1 实验数据集以及场景划分

本文实验采用的数据集是广泛应用于图像分类问题的 CIFAR-10 和 CIFAR-100 经典数据集, 该数据集中的图像像素大小为 32×32 , 都是带有标签的彩色图像。其中, CIFAR-10 数据集包含 10 个类别, 每个类别包含 6 000 张图片, 有 5 000 张图片用于训练模型, 另外 1 000 张用于测试模型性能; CIFAR-100 数据集包含 100 个类别, 每个类别包含 600 张图像, 500 张作为训练集, 另外 100 张作为测试集。为进行增量学习实验, 可以将数据集平均分为若干个任务, 每个任务所包含的类别互不相交,

从而构建 Split CIFAR-10 和 Split CIFAR-100 数据集^[25]。

3.2 实验设置细节及评价指标

3.2.1 实验设置

本文实验在 Windows 10 系统上进行, GPU 为单块 RTX-3060Ti, 实验环境为 CUDA 11.1 + Python 3.8 + Torch2.1.0。主干模型设置为精简的 Resnet-18, 投影网络是一个具有三层网络的多层感知器。在模型的训练过程中, 对比学习中的超参数 τ 默认为 0.1, 经验区回放采用随机更新策略和水库采样方法, 优化器使用随机梯度下降方法。除此之外, 本文只有 cME² 和 SCR 方法中使用了 NCM 分类器, 其他方法均使用交叉熵损失进行训练, 并使用 Softmax 进行分类。

3.2.2 评价指标

如公式(5)所示,本文使用的评价标准为平均准确度(Average Accuracy, Acc_n),该指标衡量了模型在学习了总共 n 个任务后的综合性能。其中, $a_{j,i}$ 表示模型在顺序学习 j 个任务后,在第 i 个任务经过所有测试集得到的平均准确率。

$$Acc_n = \frac{1}{n} \sum_{i=1}^n a_{n,i} \quad (5)$$

3.3 对比模型简介

本文实验选择了几种最先进的增量学习算法进行了比较,包括基于正则化方法的 LwF^[3] 和 EWC++^[1] 方法,以及基于回放方法的 AGEM^[13]、GSS^[14]、iCaRL^[11]、ER^[12]、MIR^[15] 和 SCR^[17] 方法。其中,基于正则化的两种方法由于没有保存旧样本样例,减少遗忘率的效果没有基于回放方法的好,而基于样本回放方法的效果主要与回放记忆区的内存大小、记忆区样本构建和提

取方式、记忆样本的利用方式等因素有关。AGEM 方法是利用记忆库样本的历史梯度来修正模型参数,GSS 方法通过基于梯度的样本选择,构建了一种多样化梯度的样本记忆区,iCaRL 方法是通过记忆库中样本进行蒸馏学习来保留旧知识,MIR 方法通过检索使损失增加的内存本来引导模型训练,ER 方法和 SCR 方法思路比较简单,主要是通过将记忆区中样本与新样本混合在一起,让模型训练时可以获取更多的旧知识,但 SCR 方法通过一个对比学习表示空间和 NCM 分类器进一步提高了模型减少遗忘率的效果。

3.4 对比实验

本文所提 cME² 方法与上述对比模型算法的实验结果如表 2 和表 3 所示,表中展示了每个算法在 Split CIFAR-10 和 Split CIFAR-100 数据集上训练结束时的平均分类准确率以及增量学习过程中的每阶段分类准确率。

表 2 训练结束时的平均准确率

Tab. 2 Average accuracy at the end of training

增量学习方法	单位:%					
	Split CIFAR-10			Split CIFAR-100		
	Memory size=0.1k	Memory size=0.2k	Memory size=0.5k	Memory size=1k	Memory size=2k	Memory size=5k
LwF		16.7±1.2			12.8±0.7	
EWC++		17.2±1.2			5.4±0.5	
AGEM	17.9±1.0	17.4±1.0	17.7±1.0	5.7±0.5	5.3±0.4	5.6±0.3
GSS	19.3±1.0	20.2±1.0	25.4±1.1	10.5±0.5	13.0±0.9	17.7±0.8
ER	21.1±1.2	22.5±1.0	26.8±1.6	11.4±0.7	15.4±0.6	20.8±0.9
MIR	21.2±2.4	21.9±1.3	27.5±1.9	10.7±0.7	15.1±0.5	21.7±1.2
iCaRL	26.2±3.1	30.1±3.4	38.2±2.7	16.0±0.4	17.7±0.6	19.3±0.4
SCR	38.7±1.8	46.2±2.4	51.4±1.6	26.5±0.5	32.8±0.4	37.3±0.4
cME ²	41.3±1.5	49.7±1.4	57.3±1.4	27.1±0.4	33.1±0.5	37.5±0.3

表 3 每个任务阶段的平均准确率

Tab. 3 Average accuracy of each task stage

增量学习方法	单位:%				
	Split CIFAR-10(Memory size=0.5k)				
	task1	task2	task3	task4	task5
LwF	78.1	44.8	27.7	20.7	16.1
EWC++	78.1	44.8	30.0	20.7	16.9
AGEM	90.8	41.4	29.6	22.6	18.5
GSS	78.3	44.3	35.2	23.0	19.8
ER	82.4	52.2	34.6	24.6	21.2
MIR	73.7	54.8	45.3	35.5	29.1
iCaRL	85.0	63.5	52.7	45.7	38.5
SCR	83.4	69.6	61.5	57.6	52.8
cME ²	84.3	73.4	67.1	62.3	57.7

从表 2 中可知, cME^2 方法在两种数据集上的增量实验结果都优于文中对比的其他算法, 其中, cME^2 方法在 Split CIFAR-10 数据集上的分类准确率达到 57.3%, 在 Split CIFAR-100 数据集上的分类准确率达到 37.5%。与增量学习中最经典的 LwF 方法相比, 使用较小内存回放记忆区的 cME^2 方法在平均准确率上分别提升了 24.6% 和 14.3%, 这说明只要使用有限内存来保存少数旧样本带来的性能提升是巨大的。与其他基于样本回放的方法如 AGEM、GSS、ER、MIR 和 iCaRL 方法相比, cME^2 方法展现出了明显的高效性。与同样使用对比学习训练模型的 SCR 方法相比, 在 Split CIFAR-10 数据集上平均准确率提高了 5.9%, 在 Split CIFAR-100 数据集上也保持了一定的优势。除此之外, 基于重演回放的增量学习方法通过增加内存容量可以明显提高算法精度, 所以对于 Split CIFAR-10 数据集这种类别较少的任务, 在相同内存容量下可以储存更多类别的样本, 这使得基于回放的增量学习算法更适用于这种类别数量有限的增量学习设定。

表 3 展示了每个算法在增量学习过程中每个任务阶段的平均准确率, 更精准地展现了整个训练过程中模型的性能。图 3 中可以清晰地看出 cME^2 方法在每次学习新任务时, 平均准确率都处于最佳状态, 但对于初始训练模型而言, AGEM 和 iCaRL 方法的性能都比 cME^2 方法要好, 这进一步说明 cME^2 方法降低遗忘率的效果更加优秀。综上, cME^2 方法让模型在增量学习过程中具有较好的性能提升。

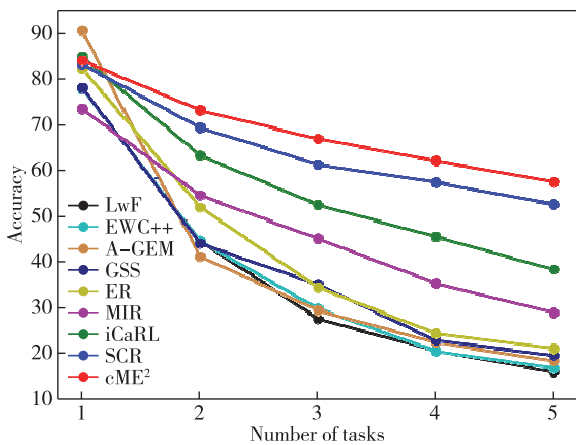


图 3 不同算法的平均准确率

Fig. 3 Average accuracy of different algorithms

3.5 消融实验

本文对所提基于记忆提炼的对比度量增强创新点在 Split CIFAR-10 数据集上进行消融实验, 实验结果如图 4 所示。紫色虚线展示了基准模型 SCR 实验结果, 绿色实线为仅添加有监督对比与本体和增强样本间对比 ($L_{SCL}+L_{ASC}$) 的实验结果, 蓝色实线展示仅使用有监

督对比与同类间对比 ($L_{SCL}+L_{SLC}$) 的实验结果, 红色为 cME^2 的实验结果。从图中可以看出, 仅添加本体和增强样本间对比或仅添加同类间对比的方法都可以提高只使用了有监督对比学习方法的性能, 而且添加同类间对比的方法与添加本体和增强样本间对比相比, 可以更大幅度地提升增量学习过程中的平均准确率, 这也证实了同类别个体间的特征差异比本体增强样本间之间的更加明显。因此, 通过 cME^2 可以充分挖掘所有样本中的细微差异, 让增量学习模型具有更好的鲁棒性与准确率。

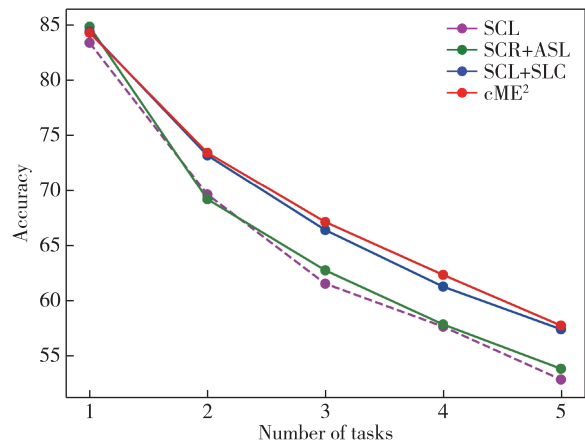


图 4 不同对比损失函数的平均准确率

Fig. 4 Average accuracy of different contrastive loss function

3.5.1 超参数 a, b 对算法的影响

1) 当 $a=1$ 时, 不同的 β 值对算法性能的影响

从图 5 中可以看出, L_{SCL} 由于 L_{ASC} 加入, 很大范围的 β 值都使得原有的对比学习方法性能得到了一定程度上的提高, 特别是当 β 的值接近于 0.0001 时所展现的效果更加明显, 这也证实了样本本身与增强样本之间存在的差异性原本就比较小。

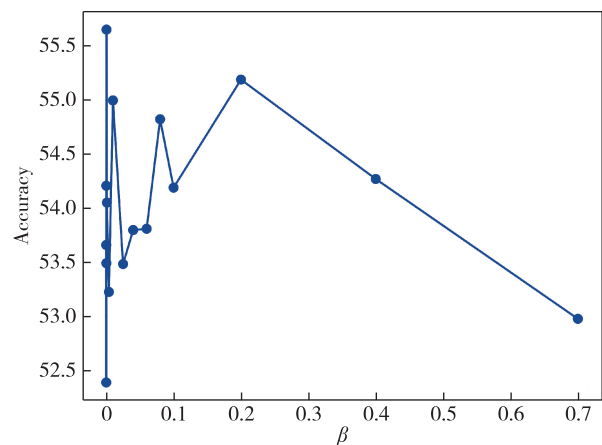


图 5 β 值对性能的影响

Fig. 5 Impact of β on performance

2) 当 $\beta=0.0001$ 时, 不同的 a 值对算法性能的影响

从图 6 中可以看出, α 值在区间 $[0.65, 1)$ 时, cME^2 方法减少遗忘率的效果都好于 SCR 方法, 但当 α 的值过低时, 由于区分不同类别的主导损失函数所占比重较少, 导致算法的性能迅速恶化。

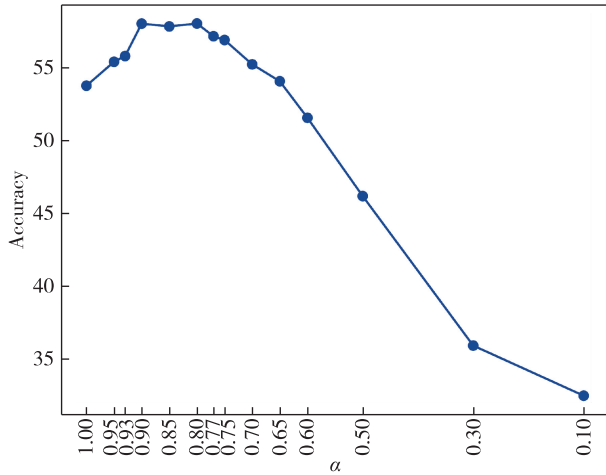


图 6 α 值对性能的影响

Fig. 6 Impact of α on performance

3.5.2 超参数 τ 对算法的影响

温度系数 τ 在对比学习中的作用通常是同于调节对困难样本的关注程度, 从图 7 中可以发现, 当 τ 的值过大或过小时都会影响算法的性能。在本文所提 cME^2 方法中, 温度系数 τ 的值在一个较大的区间 $(0.04, 0.4)$ 内都取得了较好的结果。

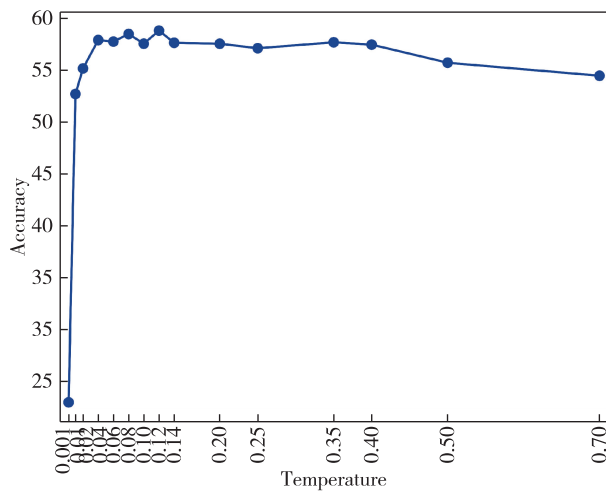


图 7 τ 值对性能的影响

Fig. 7 Impact of τ on performance

3.5.3 样本采样批量大小对算法的影响

基于重演回放方法的性能不仅与内存容量大小有关, 还会受到不同样本采样批量大小的影响, 批量太小容易导致过拟合当前任务, 难以巩固旧知识; 批量太大每次使用太多老数据进行训练, 模型难以适应当前任务。

对表 4 中的数据进行分析, 发现在一定的批量大

小条件下, AGEM, GSS, ER, MIR 和 iCaRL 方法对批量大小变化所产生的影响不大, 这是因为这些方法采用了特殊的样本选择机制, 这可以保证模型主动选择与算法最相关和有效的老样本进行重演回放, 这些样本总是会被充分利用, 使训练集对批量大小的变化在一定范围内不太敏感。而 SCR 方法和 cME^2 方法因为是完全随机的选择回放样本训练模型, 这就导致不同的批量大小设置对这两类算法的性能会产生较大的影响。从表中可以发现, 对于使用对比学习方式训练模型的方法通常在采样批量大小降低时性能会降低, 这可能是因为对比学习主要依靠的是不同类别之间的差异性来训练模型, 这就需要较大的采样批量设置来获取尽可能多的旧类别加入训练集以增强对比损失函数的作用。当样本采样批量大小设置从每批次 100 个样本降为 10 个样本时, 使用 Split CIFAR-10 数据集时的 SCR 方法的平均准确率精度会降低 7.6%, 而 cME^2 方法仅下降了 2.9%, 即使是在类别更多的 Split CIFAR-100 数据集中, cME^2 方法也比 SCR 方法好 0.5%。除此之外, cME^2 方法在每一种采样批量大小设置下的性能都比 SCR 方法表现得更优秀, 这说明 cME^2 方法训练的模型具有更好的鲁棒性。在在线类增量学习任务场景下, 模型的演化更新速度也是一个重要的性能指标, 因此, 本文进一步讨论了使用不同的记忆采样批量大小对模型训练阶段所耗时间产生的影响。如表 5 所示, 每批次采样的旧样本过多或过少都会降低模型的训练时间, 因此如何根据最终模型准确率和更新速度二者之间的重要性来选择合适的重演回放批次大小是非常重要的。除此之外, 如图 8 所示, 我们使用每个算法最短的耗时进行比较, 发现 cME^2 方法仅仅比 SCR 方法仅仅多用耗费了 1/10 的时间, 在不牺牲时间成本的前提下实现了更优秀的准确率。

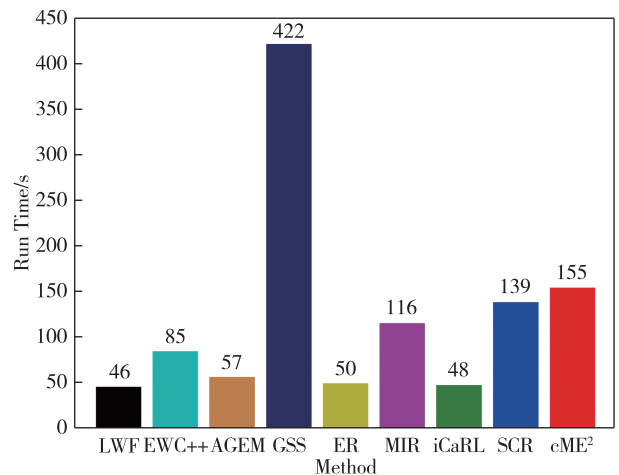


图 8 不同算法的运行时间比较

Fig. 8 Runtime of different algorithms

表 4 不同采样批量大小时的平均准确率

Tab. 4 Average accuracy for different sampling batch sizes

增量学习方法	单位: %					
	Split CIFAR-10(Memory size=0. 5k)			Split CIFAR-100(Memory size=2k)		
	Memory batch=10	Memory batch=50	Memory batch=100	Memory batch=10	Memory batch=50	Memory batch=100
AGEM	17. 7±1. 0	17. 5±1. 1	17. 4±1. 5	5. 3±0. 4	5. 2±0. 5	5. 4±0. 5
GSS	25. 4±1. 1	24. 5±1. 5	20. 8±1. 2	13. 0±0. 9	13. 2±0. 8	13. 2±0. 8
ER	26. 8±1. 6	25. 6±2. 0	24. 6±2. 1	15. 4±0. 6	14. 0±0. 9	12. 1±1. 5
MIR	27. 5±1. 9	28. 1±1. 6	28. 1±1. 6	15. 1±0. 5	14. 9±0. 8	14. 9±0. 8
iCaRL	38. 2±2. 7	38. 2±2. 7	38. 2±2. 7	17. 7±0. 6	17. 7±0. 6	17. 7±0. 6
SCR	51. 4±1. 6	57. 1±1. 0	59. 0±0. 9	14. 7±1. 0	28. 6±0. 6	32. 8±0. 4
cME ²	57. 3±1. 4	59. 2±1. 1	60. 2±1. 1	15. 5±0. 6	29. 1±0. 8	33. 1±0. 5

表 5 不同采样批量大小时的训练时间

Tab. 5 Training time for different sampling batch sizes

增量学习方法	单位: s		
	Split CIFAR-10(Memory size=0. 5k)		
	Memory batch=10	Memory batch=50	Memory batch=100
LWF		46	
EWC++		85	
AGEM	57	133	88
GSS	449	483	422
ER	50	66	96
MIR	116	131	141
iCaRL	48	48	48
SCR	188	139	176
cME ²	202	155	187

3.6 特征空间可视化

由于模型推理测试阶段使用的是最近邻类均值分类器对样本进行聚类,因此样本的特征嵌入质量会直接影响模型的性能。为了验证本文所提方法 cME²对模型特征提取能力的提升并有效调整样本在分布空间中的位置,本文通过 tsne 可视化工具对 CIFAR-10 的样例特征空间进行可视化展示并与其他使用 NCM 分类器的模型结果进行了对比。图 9a)、图 9b)、图 9c)、图 9d)分别为 iCaRL 方法在学习完增量任务 1 和增量任务 2 时的特征分布和使用 SCR 方法的特征分布;同样,图 9e)为 cME²方法在只学习了第一个任务后对该任务中存在的两个类别进行测试的分布情况,图 9f)则是模型增量学习第二个任务后对第一个任务中两个旧类别的测试样本特征降维分布。从图 9 中可以看出,虽然 SCR 方法对 iCaRL 方法中使用二进制交叉熵损失函数训练的模型进行了一定程度的改进,但使用该方法的特征空间分布仍然存在许多不同类别的样本之间互相重合或边界模糊的情况。然而,使用 cME²方法可以让模型首次学习任务中的类别样例分布更加

均匀,并具有明确的分类界限,即使在模型增量学习第二个任务后,其测试样本特征空间分布仍能减少异类样本的混合,更大程度上保留了已学习过的旧知识,减少了遗忘率。因此,样例在通过 cME²方法后可以充分发挥 NCM 分类器的性能,从而得到更加清晰的特征分布空间。

3.7 收敛性分析

本节内容展示了模型在 Split CIFAR-100 数据集上使用 cME²方法增量学习 100 个任务和 10 个任务时的损失函数收敛过程。如图 10 所示,蓝色曲线表示有监督对比学习的损失函数 L_{SCL} ,橙色曲线表示同类间对比损失函数 L_{SIC} ,绿色曲线表示本体与增强样本之间的对比 L_{ASC} ,而 cME²方法的加权损失函数则由红色曲线表示。由于有监督对比损失的作用是区分不同类别的样本,在整个损失函数中占主导作用,同类间对比损失作为辨别样本在同类中展现的个体独特性,起到了辅助作用,而增强样本对比损失则对对比学习中使用到的增强方式起到了正则化作用,具有一定的效果。从整

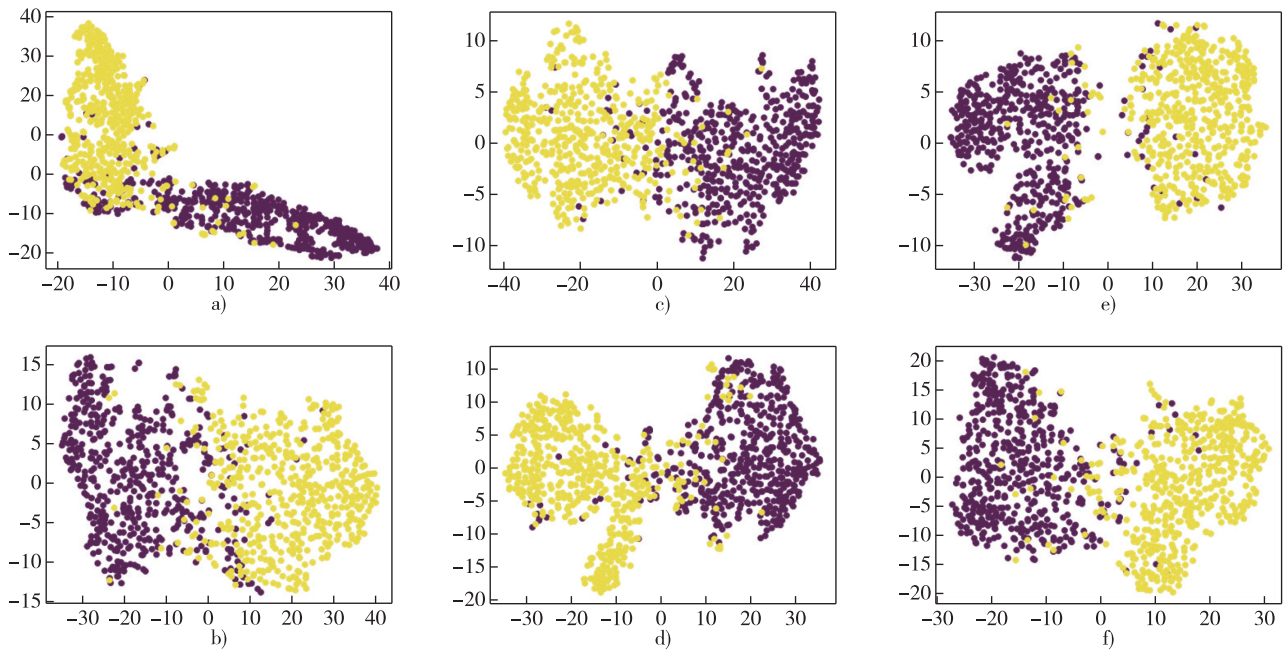


图 9 特征空间可视化图

Fig. 9 Visualization of feature space

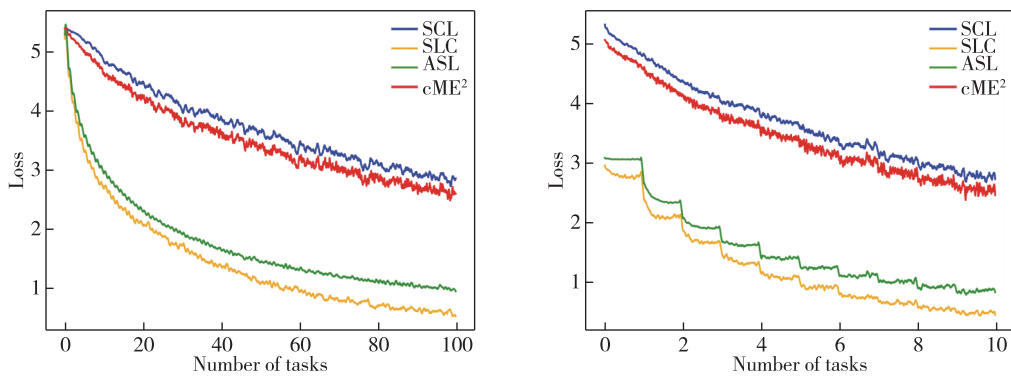


图 10 损失函数收敛曲线

Fig. 10 Convergence curve of loss function

体来看, cME^2 方法各项损失函数在每阶段增量任务中具有相对一致的收敛趋势, 说明各损失函数能够相互配合, 对模型的增量训练是有利的。综上, cME^2 方法具有良好的收敛性且能够保持较为稳定的状态。

4 结束语

为解决在线类增量学习任务场景中产生的灾难性遗忘问题, 本文提出了一种新的基于记忆提炼的对比度量增强 (cME^2) 方法。通过记忆提炼 (Memory Extraction, ME) 挖掘了内存有限的重演回放区中的样本间特征相似度距离, 实现了对旧样本信息的加强重复再利用, 充分巩固了旧任务知识。通过对比度量增强 (Contrast Metric Enhancement, cME) 提出了两种新的正负样本对, 完善了有监督对比学习提取样本间差异特征的方法, 有效地修正了特征提取器的参数, 让样本可以在嵌入空间中的位置获得更合理的分布。在 Split

CIFAR-10 和 Split CIFAR-100 数据集中, 本文所提方法相比主流类增量学习方法都取得了更好的增量性能表现。通过消融实验, 证明了本文所提的方法是有效的。

参考文献:

- [1] Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks [J]. Proceedings of the national academy of sciences, 2017, 114 (13): 3 521-3 526.
- [2] De Lange M, Aljundi R, Masana M, et al. A continual learning survey: Defying forgetting in classification tasks [J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(7): 3 366-3 385.
- [3] Li Z, Hoiem D. Learning without forgetting [J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(12): 2 935-2 947.

- [4] Chaudhry A, Dokania P K, Ajanthan T, et al. Riemannian walk for incremental learning: Understanding forgetting and intransigence[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 532-547.
- [5] Zenke F, Poole B, Ganguli S. Continual learning through synaptic intelligence[C]. International conference on machine learning. PMLR, 2017: 3 987-3 995.
- [6] Farajtabar M, Azizan N, Mott A, et al. Orthogonal gradient descent for continual learning[C]. International Conference on Artificial Intelligence and Statistics. PMLR, 2020: 3 762-3 773.
- [7] Wang S, Li X, Sun J, et al. Training networks in null space of feature covariance for continual learning[C]. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. 2021: 184-193.
- [8] Zeng G, Chen Y, Cui B, et al. Continual learning of context-dependent processing in neural networks[J]. Nature Machine Intelligence, 2019, 1(8): 364-372.
- [9] Rusu A A, Rabinowitz N C, Desjardins G, et al. Progressive neural networks[J]. arXiv preprint arXiv: 1606.04671, 2016.
- [10] Mallya A, Lazebnik S. Packnet: Adding multiple tasks to a single network by iterative pruning[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018: 7 765-7 773.
- [11] Rebuffi S-A, Kolesnikov A, Sperl G, et al. icarl: Incremental classifier and representation learning[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 2 001-2 010.
- [12] Chaudhry A, Rohrbach M, Elhoseiny M, et al. On tiny episodic memories in continual learning[J]. arXiv preprint arXiv:1902.10486, 2019.
- [13] Chaudhry A, Ranzato M A, Rohrbach M, et al. Efficient lifelong learning with a-gem[J]. arXiv preprint arXiv: 1812.00420, 2018.
- [14] Aljundi R, Lin M, Goujaud B, et al. Gradient based sample selection for online continual learning[C]. In NeurIPS. 2019: 11 816-11 825.
- [15] Aljundi R, Caccia L, Belilovsky E, et al. Online Continual Learning with Maximally Interfered Retrieval[J]. arXiv preprint arXiv:1908.04742, 2019.
- [16] Yu D, Zhang M, Li M, et al. Squeezing More Past Knowledge for Online Class-Incremental Continual Learning[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 722-736.
- [17] Mai Z, Li R, Kim H, et al. Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3 589-3 599.
- [18] Zhang Y, Pfahringer B, Frank E, et al. A simple but strong baseline for online continual learning: Repeated Augmented Rehearsal[J]. arXiv preprint arXiv: 2209.13917, 2022.
- [19] Lin H, Zhang B, Feng S, et al. PCR: Proxy-based Contrastive Replay for Online Class-Incremental Continual Learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 24 246-24 255.
- [20] Lopez-Paz D, Ranzato M A. Gradient episodic memory for continual learning[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6 470-6 479.
- [21] Serra J, Suris D, Miron M, et al. Overcoming catastrophic forgetting with hard attention to the task[C]. International Conference on Machine Learning. PMLR, 2018: 4 548-4 557.
- [22] Van de Ven G M, Tolias A S. Three scenarios for continual learning[J]. arXiv preprint arXiv: 1904.07734, 2019.
- [23] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning[C]. Advances in neural information processing systems. 2020: 18 661-18 673.
- [24] Oord A v d, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv:1807.03748, 2018.
- [25] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[J]. Handbook of Systemic Autoimmune Diseases, 2009, 1(4): 1-58.

(责任编辑:李楠)