

基于原型回放和动态更新的类增量学习方法

张禹, 曹熙卿, 钮赛赛, 许鑫磊, 张倩, 王喆

引用本文

张禹, 曹熙卿, 钮赛赛, 许鑫磊, 张倩, 王喆. [基于原型回放和动态更新的类增量学习方法](#)[J]. 计算机科学, 2023, 50(11A): 230300012-7.

ZHANG Yu, CAO Xiqing, NIU Saisai, XU Xinlei, ZHANG Qian, WANG Zhe. [Incremental Class Learning Approach Based on Prototype Replay and Dynamic Update](#) [J]. Computer Science, 2023, 50(11A): 230300012-7.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于知识蒸馏和高效通道注意力的异常检测](#)

Novelty Detection Method Based on Knowledge Distillation and Efficient Channel Attention

计算机科学, 2023, 50(11A): 220900034-10. <https://doi.org/10.11896/jsjcx.220900034>

[基于相似一致性的模型自蒸馏方法](#)

Similarity and Consistency by Self-distillation Method

计算机科学, 2023, 50(11): 259-268. <https://doi.org/10.11896/jsjcx.221000009>

[基于视频多帧融合的医学超声图像超分辨率重建方法](#)

Medical Ultrasound Image Super-resolution Reconstruction Based on Video Multi-frame Fusion

计算机科学, 2023, 50(7): 143-151. <https://doi.org/10.11896/jsjcx.220700232>

[开源软件中社区文档应用与维护的实证研究](#)

Empirical Study on Application and Maintenance of OSS Community Profile Documentation

计算机科学, 2023, 50(6A): 220600221-8. <https://doi.org/10.11896/jsjcx.220600221>

[基于知识蒸馏的抽取式自动摘要模型](#)

Extractive Automatic Summarization Model Based on Knowledge Distillation

计算机科学, 2023, 50(6A): 210300179-7. <https://doi.org/10.11896/jsjcx.210300179>

基于原型回放和动态更新的类增量学习方法

张禹¹ 曹熙卿^{2,3} 钮赛赛^{2,3} 许鑫磊¹ 张倩¹ 王喆¹

¹ 华东理工大学信息科学与工程学院 上海 200237

² 上海航天控制技术研究所 上海 201109

³ 中国航天科技集团公司红外探测技术研发中心 上海 201109

(1061941314@qq.com)

摘要 灾难性遗忘问题在增量学习场景中普遍存在,而对旧知识的遗忘会严重影响模型在整个任务序列上的平均性能。因此,针对在增量学习过程中原型偏移引起的旧知识遗忘问题,提出了一种基于原型回放和动态更新的类增量学习方法。该方法在原型更新阶段保留新类的原型后,进一步采用动态更新策略对旧类的原型进行实时更新。具体地,在学习新任务后,该策略基于当前可访问数据的已知偏移,来实现在旧类原型中存在的未知偏移的近似估计,并最终完成对旧类原型的更新,从而缓解原始的旧类原型与当前的特征映射间的不匹配。在 CIFAR-100 和 Tiny-ImageNet 数据集上的实验结果表明,所提出的基于原型回放和动态更新的类增量学习方法能够有效地减少对旧知识的灾难性遗忘,提高模型在类增量学习场景中的分类性能。

关键词: 类增量学习;原型更新;知识蒸馏;原型回放;灾难性遗忘

中图法分类号 TP391

Incremental Class Learning Approach Based on Prototype Replay and Dynamic Update

ZHANG Yu¹, CAO Xiqing^{2,3}, NIU Saisai^{2,3}, XU Xinlei¹, ZHANG Qian¹ and WANG Zhe¹

¹ School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

² Shanghai Aerospace Control Technology Institute, Shanghai 201109, China

³ Research and Development Center of Infrared Detection Technology, China Aerospace Science and Technology Corporation, Shanghai 201109, China

Abstract The problem of catastrophic forgetting is prevalent in incremental learning scenarios, and forgetting of old knowledge can severely affect the average performance of the model over the entire task sequence. Therefore, a class incremental learning approach based on prototype replay and dynamic update is proposed to address the problem of old knowledge forgetting caused by prototype offset in the incremental learning process. This method further updates the prototypes of the old classes in real time using a dynamic update strategy after retaining the prototypes of the new classes in the prototype update phase. Specifically, after learning the new task, the strategy achieves an approximate estimation of the unknown bias present in the old-class prototypes based on the known bias of the currently accessible data, and finally completes the update of the old-class prototypes, thus being able to alleviate the mismatch between the original old-class prototypes and the current feature mapping. Experimental results on CIFAR-100 and Tiny-ImageNet datasets show that the proposed class incremental learning approach based on prototype replay and dynamic update is effective in reducing catastrophic forgetting of old knowledge, thus improving the classification performance of the model in class incremental learning scenarios.

Keywords Class incremental learning, Prototype update, Knowledge distillation, Prototype replay, Catastrophic forgetting

1 引言

深度学习^[1]已经在许多计算机任务中取得巨大的成功,包括图像分类^[2]、目标检测^[3]、语音识别^[4]和自然语言处理^[5]等。近年来,增量学习逐渐成为深度学习领域的研究热点,其学习目标是在由多个子任务组成的任务序列上,模型能够

持续地学习新任务,同时保留在已经接触过的旧任务上的知识^[6]。与经典的静态训练方式相比,增量式训练方式的不同在于在学习新任务的阶段,模型并不能访问旧的数据。因此,在只接触新数据的极端场景中,模型不可避免地会更加擅长处理新任务,并导致旧任务的性能显著下降。这种问题被称为灾难性遗忘^[7],它是增量学习面临的主要挑战。

基金项目:上海市科技计划项目(21511100800,20511100600);国家自然科学基金项目(62076094);国防科技领域基金项目(2021-JCJQ-JJ-0041);中国航天科技集团公司第八研究院产学研合作基金资助项目(SAST2021-007)

This work was supported by the Shanghai Science and Technology Program(21511100800,20511100600), National Natural Science Foundation of China(62076094), Chinese Defense Program of Science and Technology(2021-JCJQ-JJ-0041) and China Aerospace Science and Technology Corporation Industry-University-Research Cooperation Foundation of the Eighth Research Institute(SAST2021-007).

通信作者:王喆(wangzhe@ecust.edu.cn)

增量学习领域已经探索各种策略来缓解灾难性遗忘问题。早期的增量学习方法^[8]主要从正则化^[9]、参数隔离^[10]和样例回放^[11]等方面进行优化。然而,这些方案都存在一些不可忽视的局限性。正则化方法过分关注于对模型参数的约束,而导致模型面向新任务的可塑性被严重削弱。参数隔离方法则是随着新任务的持续学习,模型将不得不面临参数急剧增加的问题。而样例回放方法存在天生的劣势,即无法应用于一些强调数据隐私和内存空间不足的实验场景。最近,抗遗忘性能优异的基于原型回放的方法获得了大量的关注。这类方法首先保留旧类的原型,并在新类的学习阶段回放旧类原型,以有效地巩固旧类的知识,进而达到新旧类之间稳定性和可塑性^[12]的平衡。然而,在增量学习过程中,模型的特征提取器会被持续地学习,从而导致原先保留的旧类原型不能很好地匹配最新的特征提取器,即旧类原型在当前特征空间中存在不同程度的偏移。

针对上述问题,本文将原型动态更新策略引入到原型回放阶段,实现从新类数据的已知偏移到老类原型的未知偏移的推断,从而有效地缓解旧类原型与新的特征映射间不匹配的困境。同时,本文专注于更困难的类增量学习场景,而融合动态更新的原型回放策略通过动态更新旧类原型来最小化对旧知识的遗忘,从而帮助类增量模型在可塑性和稳定性之间取得一个良性的平衡。

2 相关工作

增量学习的研究已经取得很大的进展,然而灾难性遗忘问题仍是制约增量算法性能的主要因素。主流的增量学习方法大多从以下 4 个方面缓解灾难性遗忘,分别是正则化方法、参数隔离方法、样例回放方法和原型回放方法。

1) 正则化方法。在模型学习过程中,正则化方法通过对模型参数的更新设置额外的约束来减少旧知识的遗忘。Kirkpatrick 等^[13]提出了弹性权重巩固(Elastic Weight Consolidation, EWC)的方法,其使用 Fisher 信息矩阵来计算给定任务中每个模型参数的重要性,并设计了一个二次惩罚代价来灵活约束模型参数。类似地,Aljundi 等^[14]提出的记忆感知突触(Memory Aware Synapses, MAS)也是计算神经网络参数的重要性,并在学习新任务时对重要参数的变化进行惩罚,防止与之前任务相关的重要信息被遗忘。Joseph 等^[15]提出了基于能量的潜在对齐器的方法(ELI),该方法引入了一个基于能量的正则化项,它通过学习潜在在表征的能量流形来缓解增量学习过程中发生的表征偏移。

2) 参数隔离方法。此类方法的出发点是根据新旧任务或任务相似性的准则,对模型参数进行划分隔离来缓解知识遗忘。这类方法主要有两种实现思路,一是为新任务扩展更多的参数。Yoon 等^[16]提出的动态可扩展网络(Dynamically Expandable Network, DEN)就是对上述的基础思路进行改进,即在学习每个新任务时只选择必要数量的单元,来动态地扩展模型参数。Yan 等^[17]提出动态可扩展表征(Dynamically Expandable Representation, DER)的方法,该方法在每个增量阶段中先冻结之前学习到的表征,再学习一些新的额外的特征维度来动态扩展旧的表征。另一种思路是引入掩码实现任务间的参数隔离。如, Mallya 等^[18]提出 Piggyback 方法,其为每个新任务生成一组二进制掩码,选择性确保重要参数

处于激活状态,从而在保留旧知识的同时提高新任务的性能。

3) 样例回放方法。与其他方法相比,样例回放方法基于松弛的增量设置,即每个类在学习阶段后都能保留少量的样例。在学习新任务时,新类和旧类的样例会混合在一起训练,从而防止旧类的知识被遗忘。具体地, Rebuffi 等^[19]提出的 iCaRL 方法在有限的内存大小中对最有代表性的样例进行采样,并学习了一个增量分类器和相应的数据表示。Shin 等^[20]提出了深度生成回放(Deep Generative Replay, DGR)的方法,该方法利用生成对抗网络生成旧任务的伪训练数据,然后使用新旧任务数据集联合训练网络模型。Belouadah 等^[21]提出的 IL2M 方法利用微调 and 双记忆的策略来减少图像识别中灾难性遗忘的负面影响,其中双内存分别存储了过去类的范例图像和在最初学习时获得的过去的类统计数据。

4) 原型回放方法。相比样例回放方法,原型回放方法的优势在于其并不需要存储任何旧任务的数据,只要求为每个旧类保留一个原型。这类方法在保持严格的增量设置的前提下,具有优秀的抗遗忘性能,从而受到越来越多的关注。Zhu 等^[22]提出了 PASS 方法,该方法采用自监督学习来学习新任务中更通用的特征,同时在新任务的数据中结合旧类的原型进行训练,来保持旧任务的决策边界。Yu 等^[23]提出了 SDC 方法,该方法在嵌入网络中学习特征的语义漂移量,并根据空间变化信息对发生漂移的原型进行语义信息补偿。

本文方法从原型回放的思路出发,在完成新任务的增量学习后,额外引入了面向旧类原型的动态更新策略,来缓解由特征空间优化导致的原型偏移,从而保持旧任务的决策边界。相比现有主流的分类增量学习方法,本文方法不需要回放任何旧样例,这很大程度上减少了增量学习过程中的资源消耗。相反,所提方法以原型回放技术作为基础架构,并结合了原型动态更新的策略,来缓旧类原型的表征信息在新的特征空间中的滞后性。因此,尽管样例回放方法和原型回放方法在增量学习任务上表现出色,但是本文方法通过采用动态更新策略,仍然能够取得显著的性能优势。这意味着本文方法在应对增量学习任务时能够更好地适应新数据,并在长任务序列的性能上取得令人满意的结果。

3 模型框架

3.1 问题定义

在标准的类增量学习的设置中,模型在一个任务不可知的任务序列 $\{D^{(1)}, D^{(2)}, \dots, D^{(T)}\}$ 上持续地学习,并最终保留一个统一的分类器,来学习并分类所有不相交类别。具体地, $D^{(t)}$ 是第 t 个任务的训练集,且 $D^{(t)} = \{x_i^{(t)}, y_i^{(t)}\}_{i=1}^{N_t}$, 其中 $x_i^{(t)}$ 表示输入样本, $y_i^{(t)}$ 表示对应的类别标签, N_t 表示该任务的样本总数。一般情况下,每个任务都包含多个类,而这些不同任务的类之间并不相交。同时,我们执行严格的增量设置,即模型在新任务的训练过程中,不能访问来自旧任务的训练样本。最后,在类增量学习场景中的学习目标是,模型在最大限度地不遗忘旧知识的前提下持续完成对新类别的学习,即最终实现对所有类别的正确分类。

3.2 总体框架

本文在类增量学习中引入原型动态更新模块来实现旧类原型的实时更新。首先,在新任务的学习阶段,除了必须的分类损失,模型还包括了特征知识蒸馏和基于旋转自监督的

标签增强。前者一定程度上约束模型的特征提取器的参数变化过大,而后者则帮助模型学习到一些更为鲁棒和通用的特征。然后,模型进一步加入融合动态更新的原型回放损失,来巩固模型对旧类的分类知识。同时,在不接触旧数据的前提下,动态更新原型能够保证更新后的旧类原型更好地适配最新的特征提取器。所提方法的流程图如图1所示。

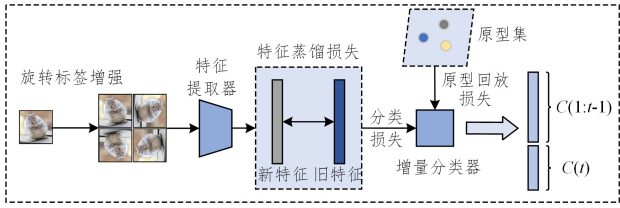


图1 基于原型回放和动态更新的类增量学习方法流程图

Fig. 1 Flow chart of class incremental learning approach based on prototype playback and dynamic update

3.3 融合动态更新的原型回放

在重视数据隐私的许多现实场景中,增量学习算法无法存储过去任务的旧数据,导致缺乏对旧知识的有效维护。同时,相比基于旧样例回放的方法,单纯地通过正则化或参数隔离等方式约束模型参数并不能很好地解决灾难性遗忘问题。因此,本文提出了一种基于原型回放的类增量学习算法。

在嵌入网络中,原型表示类特征的均值中心。具体地,特征提取器将某一类的数据映射到特定嵌入空间中,其中特征间的距离反映了数据间的差异。通过对同一类中所有数据的特征进行均值操作,可以得到该类的原型。原型的定义如式(1)所示:

$$\mu_c = \frac{1}{n_c} \sum_i [y_i = c] \cdot f_i \quad (1)$$

其中, n_c 表示类别为 c 类的训练样本总数, f 表示训练样本对应的特征, μ_c 代表第 c 类的原型。而 $[y_i = c]$ 则是一个条件判断式,当 y_i 等于 c 时,则该值为1,否则该值为0。

在增量学习过程中,学习每个类别后,我们会额外保留该类别的原型。然后,在学习新任务时,我们提取所有旧类别的原型,并将其输入到分类器中进行分类,来保持分类器对于旧任务的决策经验。因此,在任务 t 的增量学习阶段,基于原型回放(Prototype Replay, PR)的损失项的定义为:

$$L_{t,pr} = \sum_{i=1}^{t-1} L(C(\mu_i; \delta_t), y_i) \quad (2)$$

其中, $L(\cdot)$ 表示模型的损失函数, $C(\cdot)$ 表示模型的分类器, δ_t 则表示学习任务 t 时分类器的参数。

然而,随着对任务序列的不断学习,原先保留的旧原型不能很好地匹配最新的特征提取器,也就无法准确地代表旧类别。因此,本文提出了一种原型动态更新策略来缓解原型中表征信息的损耗。其核心思路是在学习新任务后,模型基于当前可访问数据的已知偏移,来估计已保留旧原型的未知偏移,并动态更新滞后的旧原型。

首先,本文定义符号 $\mu_{c_k}^t$ 表示在学习任务 t 后类 c_k 的真实原型,其中任务 t 相对任务 k 在任务序列中的位置更加靠后,而类 c_k 泛指任务 k 中所有类的集合。因此,原先保留的原型和真实的原型之间存在误差,我们将其称为原型的偏移,并定义为:

$$\Delta_{c_k}^{t-1} = \mu_{c_k}^t - \mu_{c_k}^{t-1} \quad (3)$$

然而,在任务 t 的学习过程中,模型不能直接访问类 c_k 的

数据,进而也就无法测量原型 $\mu_{c_k}^t$ 的偏移量。相反,模型能够直接计算出当前任务中样本点在特征空间中的最新偏移量。

$$\partial_i^{t-1} = f_i^t - f_i^{t-1}, y_i \in c_t \quad (4)$$

其中, f_i^{t-1} 和 f_i^t 分别表示类 c_t 中的样本点 i 在学习任务 $t-1$ 和任务 t 后的特征值。进一步,我们将单个样本点的偏移 ∂_i^{t-1} 扩展到原型的偏移 $\Delta_{c_k}^{t-1}$,并给出原型偏移的计算式:

$$\Delta_{c_k}^{t-1} = \frac{\sum_{(i, y_i \in c_t)} \alpha_i \cdot \partial_i^{t-1}}{\sum_{(i, y_i \in c_t)} \alpha_i} \quad (5)$$

$$\alpha_i = e^{-\frac{\|f_i^{t-1} - \mu_{c_k}^{t-1}\|^2}{2\sigma^2}} \quad (6)$$

其中, σ 是高斯核的标准差。在得到旧类原型的邻近偏移 $\Delta_{c_k}^{t-1}$ 后,模型能够进一步迭代计算旧类原型的全局偏移 $\Delta_{c_k}^{k \rightarrow t}$,并最终使用式(3)近似估计在增量阶段 t 时的真实原型 $\mu_{c_k}^t$ 。原型动态更新策略的定义如下:

$$\mu_{c_k}^t = \underbrace{\mu_{c_k}^k + \Delta_{c_k}^{k \rightarrow k+1} + \dots + \Delta_{c_k}^{t-1}}_{\mu_{c_k}^{k+1}} + \Delta_{c_k}^{t-1} \quad (7)$$

原型动态更新策略的流程如图2所示。

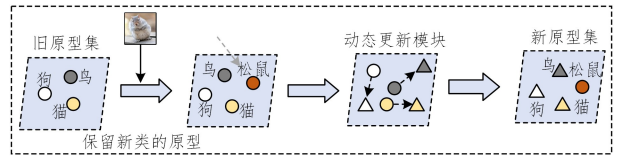


图2 动态更新模块的实现流程

Fig. 2 Implementation flow of dynamic update module

此外,受PASS^[22]的启发,本文在使用旧类原型进行原型回放时会引入高斯噪声,来对原始的原型 μ 进行表征信息的增强。增强后的原型如下:

$$\phi = \mu + u \cdot r \quad (8)$$

其中, $u \sim N(0, 1)$ 是随机生成的高斯噪声,且与原型 μ 的尺寸保持一致。而 r 是控制高斯噪声不确定性的参数。通常情况下, r 是预先设定好的,且一般设定为该类别中数据特征表示的方差值。

最后,基于增强原型回放(Enhanced Prototype Replay, EPR)的损失项的定义如下:

$$L_{t,epr} = \sum_{i=1}^{t-1} L(C(\phi_i; \delta_t), y_i) \quad (9)$$

3.4 模型的优化目标

首先,在训练新任务时,模型会学习一个分类损失。然而,与原始的分类问题相比,我们进一步将旋转自监督^[24]的思想引入到本文的分类问题中。具体地,对于新任务中的每个类,我们会将训练数据分别旋转 90° , 180° 和 270° ,并相应地扩充标签。训练数据的转换表达式如下:

$$\left\{ \begin{array}{l} \{Rot(x, 90\eta)\} \\ \{4 \cdot y + \eta\} \end{array} \right., \eta \in \{0, 1, 2, 3\} \quad (10)$$

因此,原始的 c 分类问题被人为地扩展为 $4c$ 分类问题,训练样本 x 和对应的标签 y 也分别地由式(9)转换操作后的 \tilde{x} 和 \tilde{y} 替代。对原始分类任务施加自监督约束,能够有效地帮助模型学习在多个旋转视角下更为鲁棒的特征,从而缓解增量学习中的灾难性遗忘。最后,在学习新任务 t 时,采用交叉熵损失(Cross-entropy Loss, CE)作为分类损失,表示如下:

$$L_{t,ce} = L_t(C(F(\tilde{x}; \theta_t); \delta_t), \tilde{y}_t) \quad (11)$$

此外,为了缓解特征层面的知识遗忘,我们采用特征知识蒸馏对模型的特征提取器施加一个正则化约束。这种约束很好地保证了旧模型和新模型之间的特征表示的稳定性,同时也在一定程度上减少了旧原型中由特征提取器过度学习导致的偏移。具体地,特征知识蒸馏使用 L_2 范数实现。在学习任务 t 时,特征知识蒸馏 (Feature Knowledge Distillation, FKD) 的定义如下:

$$L_{t, fkd} = \| F_t(\tilde{x}_t; \theta_t) - F_{t-1}(\tilde{x}_t; \theta_{t-1}) \|_2^2 \quad (12)$$

最后,我们得到了本文方法的总损失函数,它分别由式(9)、式(11)和式(12)这3个子项组成。具体地,在任务 t 的训练阶段,模型的总损失定义如下:

$$L_{t, final} = L_{t, ce} + \lambda_1 \cdot L_{t, fkd} + \lambda_2 \cdot L_{t, epr} \quad (13)$$

其中, λ_1 和 λ_2 是两个损失系数,且在实验中都设置为 10。

4 实验及结果分析

4.1 数据集介绍

本文在 CIFAR-100 和 Tiny-ImageNet 数据集上分别进行了实验。其中 CIFAR-100 数据集有 100 个类,每个类包含 600 张图片,而这些图片会细分为 500 张训练图片和 100 张测试图片。对于 CIFAR-100 数据集,本文选取了两种类增量设置。两种设置中的基类都是前 50 个类,而只有增量步骤存在差异,分别为 5 个和 10 个。在每个增量步骤后,模型会对现阶段学习过的所有类别进行测试。而 Tiny-ImageNet 数据集是 ImageNet 数据集^[25]的缩减版本,它由 200 个类组成。每个类都包括 500 张训练图片、50 张验证图片和 50 张测试图片。相比 CIFAR-100 数据集, Tiny-ImageNet 数据集的类别更多,训练和测试样本更少,对模型分类性能的要求也就越高。同时, Tiny-ImageNet 数据集也提供了两种类增量设置,除了基类中的类别增加到前 100 个类,其他都与 CIFAR-100 数据集保持一致。上述两个数据集的示例图如图 3 所示。



图 3 数据集样例展示图

Fig. 3 Presentation diagram of dataset samples

4.2 实验设置

对于 CIFAR-100 和 Tiny-ImageNet 数据集,本文方法都是在 ResNet-18^[26]网络架构上进行实验。在实验中,我们设置批大小为 64,初始学习率为 0.001 以及训练轮数为 100。同时,我们使用 Adam 优化器进行模型的训练,且在经过 45 和 90 个训练轮数后,自动地将学习率乘以 0.1。本文中的实验结果都是经过 3 次实验的平均值,避免了实验数据的偶然性。此外,我们还选取了一些主流的基于样例回放的方法 (iCaRL^[19], EEIL^[27], UCIR^[28]) 作为对比方法。对于这些样

例回放方法的回放缓冲区场景,我们选择了通用的内存设置,其中模型为每个旧类别都保留恒定的 20 个样例。

4.3 对比实验分析

所有方法在 CIFAR-100 和 Tiny-ImageNet 数据集上的平均分类准确率结果如表 1 所列。同时每个数据集都定义有两种类增量设置,因此最终得到了 4 列对比结果。以中间的横线为分隔,其中前 4 种方法是基于无样例回放的方法,而后 4 种则属于有样例回放的范畴。由表 1 可知,本文方法在所有实验设置中都显著优于其基模型 PASS,其中在 10 阶段的 CIFAR-100 上的差距达到了 2.1%,而在性能最接近的 5 阶段的 CIFAR-100 上的提升也有 1.4%。这种提升受益于原型动态更新策略,其有效地缓解了类增量过程中的原型偏移,从而巩固旧类的知识。

同时,与主流的基于样例回放的方法相比,本文方法也具有明显的优势。这充分说明了结合高效更新策略的原型回放方法完全能够在不使用旧样例时,取得超过样例回放方法的性能。具体地,相比性能第二的方法 PASS 和 UCIR,本文方法在 CIFAR-100 数据集上的性能分别提高了 1.4% 和 1.0%。而在 Tiny-ImageNet 数据集上的优势进一步达到了 1.7% 和 0.3%。

表 1 CIFAR-100 和 Tiny-ImageNet 数据集的准确率结果
Table 1 Accuracy results for CIFAR-100 and Tiny-ImageNet datasets

method	(单位: %)			
	CIFAR-100		Tiny-ImageNet	
	5 steps	10 steps	5 steps	10 steps
LwF	32.5	17.8	22.3	17.7
MUC	49.5	28.4	32.6	26.8
iCaRL-CNN	51.3	48.5	34.9	31.1
iCaRL-NCM	58.8	53.8	46.1	43.4
EEIL	62.6	60.8	47.4	45.2
UCIR	63.8	62.0	49.3	48.9
PASS	65.5	60.9	49.5	47.3
Ours	66.9	63.0	51.2	49.2

为了进一步比较方法的性能,我们以 CIFAR-100 数据集的两种类增量设置为例,分别展示在各个增量步骤时的准确率,如图 4 和图 5 所示。观察实验结果,本文方法在所有阶段的性能都显著优于其他方法,包括性能最好的基于样例回放的方法 UCIR。此外,同样基于原型回放的 PASS 在绝大多数阶段也与 UCIR 性能相当,甚至取得了更好的分类准确率。这些有力地证明了基于原型回放的方法能够在不保留任何旧类样例的情况下有效地缓解了灾难性遗忘问题。进一步地,通过图 4 和图 5 可知,本文方法在前两个增量阶段上的性能和基模型 PASS 相当,其优势在于从三阶段开始的后期增量阶段上的性能提升。这是因为第一个阶段是模型在基类上进行学习,此时的数据都是新类,也就不存在旧类的原型。而第二个阶段虽然在保留新类的原型后会旧类的原型进行动态更新,但是模型学习过程中的原型回放损失是基于原始的旧类原型计算的,并没有涉及更新后的旧类原型。因此,原型动态更新策略也就无法优化模型在前两个阶段上的性能。同时,得益于对旧类原型的动态更新,本文方法能够利用更新后的适配于当前特征空间的原型,来实现在后期增量阶段上性能的有效提升。

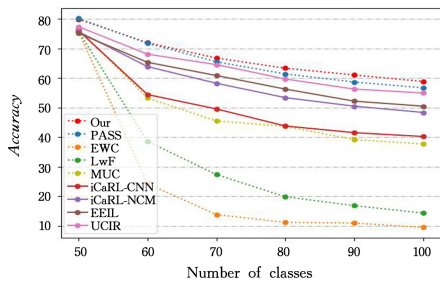


图4 CIFAR-100数据集(5个增量步骤)上的实验结果
Fig.4 Experimental results on CIFAR-100 dataset (5 incremental steps)

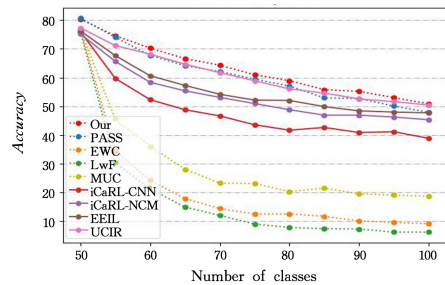


图5 CIFAR-100数据集(10个增量步骤)上的实验结果
Fig.5 Experimental results on CIFAR-100 dataset (10 incremental steps)

4.4 消融实验分析

为了进一步探究所提出的原型动态更新(Prototype Dynamic Update, PDU)策略的作用,我们进行了在 CIFAR-100 和 Tiny-ImageNet 数据集的多种类增量场景中的消融实验,相应的实验结果如表 2 所列。可以发现,结合原型动态更新策略的模型在所有实验场景中都达到了最优的分类准确率,相比基模型分别有 1.41%, 2.70%, 1.74% 和 1.93% 的性能提升。这有力地证明了原型动态更新策略能够高效地实时修正旧类原型中滞后的表征信息,从而进一步提升模型的性能。

表2 CIFAR-100 和 Tiny-ImageNet 数据集的消融实验

Table 2 Ablation study on CIFAR-100 and Tiny-ImageNet datasets (单位: %)

method	CIFAR-100		Tiny-ImageNet	
	5 steps	10 steps	5 steps	10 steps
baseline	65.46	60.32	49.47	47.25
+PDU	66.87	63.02	51.21	49.18

4.5 可视化实验分析

为了更形象地呈现原型动态更新模块的效果,以 5 阶段的 CIFAR-100 数据集为例,本文随机选取 3 个类别的数据,并对其更新前后的原型进行 t-SNE^[29] 可视化展示,相应的结果分别如图 6 和图 7 所示。其中图 6 展示的是原始的旧类原型,即这些原型都是最新保留的,并不存在特征空间上的偏移。图 6 中的圆形即是每一个类别的原型。然后,图 7 进一步展示后一个增量阶段的更新后的旧类原型。我们发现即使只相差一个增量阶段,旧类数据的特征及原型在嵌入空间中仍然存在偏移。随着模型的特征提取器不断学习并更新,这些偏移也会产生越来越大的差值,进而加剧了旧任务的灾难性遗忘问题。

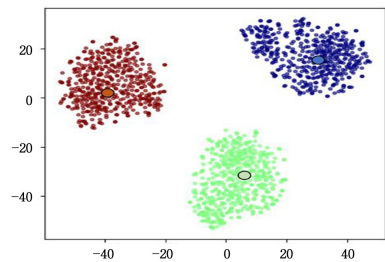


图6 原始的旧类原型

Fig.6 Original old-class prototypes

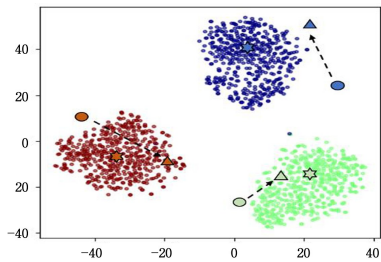


图7 更新后的旧类原型

Fig.7 Updated old-class prototypes

具体地,在图 7 中,圆形表示上一个增量步骤保留的原型,三角形表示动态更新后的新原型,虚线箭头表示原型动态更新时的方向向量,而六角星表示当前增量步骤的真实原型,它在严格的增量实验设置下是无法计算获得的。观察图 7 可以发现,更新后的旧类原型显著改善了原始原型的位置,使其更加接近于当前增量阶段的真实旧类原型的位置。因此,在学习新任务时,原型动态更新模块能够有效减少原型偏移导致的误差,从而很大程度上避免旧知识的灾难性遗忘,提高整体任务上的分类性能。

此外,以 5 阶段的 CIFAR-100 数据集为例,本文还从初始基类中随机选取了 5 个类别的数据,并在完成所有的增量任务后,将这些类别的数据对应的特征分布进行 t-SNE 可视化展示,不同方法的效果可视化结果如图 8 和图 9 所示。通过对比分析发现,本文方法在学习全部增量任务后得到相同类别的特征会呈现出更加紧凑的分布,同时不同类别的特征分布仍然保持着良好的距离。相比之下, PASS 最终得到的特征分布则不太理想,其每个类别的特征分布过于松散,且不同类别间的特征分布距离也太过狭小。综上所述,即本文方法在增量任务后学习到的特征分布类内距离更小,类间距离更大,进而从根本上保证了模型能够获得一个更高的分类准确率。最后,值得注意的是,相比 PASS,本文方法学习的特征分布中的离群点更少,分类边界更加清晰,从而为模型营造出了一个更好的分类条件。

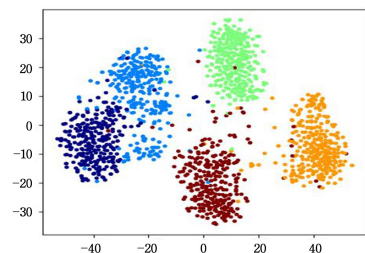


图8 PASS 的 t-SNE 可视化结果(特征分布)

Fig.8 t-SNE visualization results of PASS(feature distribution)

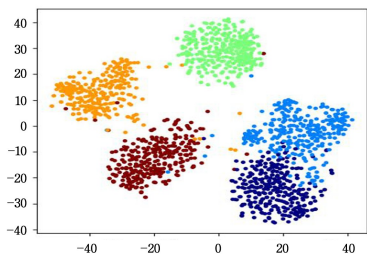


图9 所提方法的 t-SNE 可视化结果(特征分布)

Fig. 9 t-SNE visualization results of the proposed method (feature distribution)

4.6 复杂度实验分析

最后,本文比较并分析了所有方法的时间复杂度(Complexity)、计算量(Flops)和参数量(Params),结果如表3所列。其中, n 表示输入样本数, m 表示回放样本数, p 表示回放原型数, d 表示分布外数据集大小,而 t 表示模型参数量。可以发现,无样例回放方法 LwF, EWC 和 MUC 的时间复杂度分别为 $O(n)$, $O(n)+O(t)$ 和 $O(n)+O(d)+O(t)$,这是因为 EWC 和 MUC 需要计算每个模型参数对旧任务的重要性,而 MUC 进一步利用分布外数据集实现了子分类器集的差异最大化。同时,样例回放方法 iCaRL, EEIL 和 UCIR 都在学习过程中回放旧样例来缓解知识遗忘,因此时间复杂度均为 $O(n)+O(m)$ 。最后,作为原型回放方法, PASS 和本文方法都取得了 $O(n)+O(p)$ 的复杂度。相比 PASS,本文方法引入了原型动态更新的过程,而该更新策略的复杂度为 $O(p)$,因此时间复杂度为 $O(n)+2O(p)$ ($O(n)+O(p)$ 级别)。

进一步,由于回放原型数往往远小于回放样例数,因此对比样例回放方法,本文方法能够在复杂度更低的同时取得更高的准确率。相比无样例回放方法,本文方法在复杂度上的性能基本相当,然而在准确率上的表现远远领先前者。同时,表3中的方法都是基于正则化约束或回放策略的类增量学习方法,因此其计算量和参数量主要受到骨干网络的影响。值得注意的是,这些方法都采用了 ResNet-18 网络作为基础实现,因此它们的计算量和参数量基本一致。此外,可以发现 MUC 的参数量更大,这是因为其需要引入额外的子分类器集合来有差异性地学习增量任务。

表3 模型复杂度分析

Table 3 Analysis of model complexity

method	Complexity	Flops	Params
LwF	$O(n)$	27.3×10^9	11.41×10^6
EWC	$O(n)+O(t)$	27.3×10^9	11.41×10^6
MUC	$O(n)+O(d)+O(t)$	27.3×10^9	11.65×10^6
iCaRL	$O(n)+O(m)$	27.3×10^9	11.41×10^6
EEIL	$O(n)+O(m)$	27.3×10^9	11.41×10^6
UCIR	$O(n)+O(m)$	27.3×10^9	11.41×10^6
PASS	$O(n)+O(p)$	27.3×10^9	11.41×10^6
Ours	$O(n)+O(p)$	27.3×10^9	11.41×10^6

结束语 针对增量学习过程中原型偏移引起的旧知识遗忘问题,本文提出了一种基于原型回放和动态更新的类增量学习方法。该方法首先计算得到当前可访问数据的已知偏移,然后根据该偏移进一步估计旧类原型的未知偏移,并最终实现所有旧类原型的动态更新,从而极大地减少新模型对旧知识的遗忘。在 CIFAR-100 和 Tiny-ImageNet 数据集上的实验结果表明,与目前主流的增量学习方法相比,本文方法

能够有效地缓解类增量学习场景中的灾难性遗忘问题,并在所有实验设置上均取得了最好的性能。

在未来的工作中,我们会进一步探讨对原型动态更新策略的优化,使其能够适应在长任务序列上的类增量学习场景。

参考文献

- [1] AKHTAR N, MIAN A. Threat of adversarial attacks on deep learning in computer vision: A survey[J]. IEEE Access, 2018, 6: 14410-14430.
- [2] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:558-567.
- [3] BAEK S H, HEIDE F. Polka lines: Learning structured illumination and reconstruction for active stereo[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:5757-5767.
- [4] FAYEK H M, LECH M, CAVEDON L. Evaluating deep learning architectures for speech emotion recognition[J]. Neural Networks, 2017, 92: 60-68.
- [5] LIPPI M, MONTEMURRO M A, DEGLI ESPOSTI M, et al. Natural language statistical features of LSTM-generated texts [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3326-3337.
- [6] DABBAGH N, CASTANEDA L. The PLE as a framework for developing agency in lifelong learning[J]. Educational Technology Research and Development, 2020, 68(6): 3041-3055.
- [7] ZHANG T, WANG X, LIANG B, et al. Catastrophic interference in reinforcement learning: A solution based on context division and knowledge distillation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022: 1-15.
- [8] MASANA M, LIU X, TWARDOWSKI B, et al. Class-incremental learning: survey and performance evaluation on image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5513-5533.
- [9] CHEN H, WANG Y, HU Q. Multi-Granularity Regularized Re-Balancing for Class Incremental Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7263-7277.
- [10] LI K, WAN J, YU S. CKDF: Cascaded knowledge distillation framework for robust incremental learning[J]. IEEE Transactions on Image Processing, 2022, 31: 3825-3837.
- [11] LIN H, FENG S, LI X, et al. Anchor Assisted Experience Replay for Online Class-Incremental Learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(5): 2217-2232.
- [12] LIN G, CHU H, LAI H. Towards better plasticity-stability trade-off in incremental learning: a simple linear connector [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 89-98.
- [13] 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): 3521-3526.
- [14] ALJUNDI R, BABILONI F, ELHOSEINY M, et al. Memory aware synapses: Learning what (not) to forget[C]// Proceed-

- ings of the European Conference on Computer Vision. 2018:139-154.
- [15] JOSEPH K J, KHAN S, KHAN F S, et al. Energy-based Latent Aligner for Incremental Learning [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:7452-7461.
- [16] YOON J, YANG E, LEE J, et al. Lifelong Learning with Dynamically Expandable Networks[C]// International Conference on Learning Representations. 2018.
- [17] YAN S, XIE J, HE X. Der: Dynamically expandable representation for class incremental learning [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:3014-3023.
- [18] MALLYA A, DAVIS D, LAZEBNIK S. Piggyback: Adapting a single network to multiple tasks by learning to mask weights [C]//Proceedings of the European Conference on Computer Vision. 2018:67-82.
- [19] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. icarl: Incremental classifier and representation learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2017:2001-2010.
- [20] SHIN H, LEE J K, KIM J, et al. Continual learning with deep generative replay[C]// Advances in Neural Information Processing Systems. 2017:1-10.
- [21] BELOUADAH E, POPESCU A. Il2m: Class incremental learning with dual memory[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:583-592.
- [22] ZHU F, ZHANG X Y, WANG C, et al. Prototype augmentation and self-supervision for incremental learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:5871-5880.
- [23] YU L, TWARDOWSKI B, LIU X, et al. Semantic drift compensation for class-incremental learning[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:6982-6991.
- [24] LEE H, HWANG S J, SHIN J. Self-supervised label augmentation via input transformations[C]// International Conference on Machine Learning. 2020:5714-5724.
- [25] YANG K, YAU J H, LI F F, et al. A study of face obfuscation in imagenet[C]// International Conference on Machine Learning. 2022:25313-25330.
- [26] ULLAH A, ELAHI H, SUN Z, et al. Comparative analysis of AlexNet, ResNet18 and SqueezeNet with diverse modification and arduous implementation[J]. Arabian Journal for Science and Engineering, 2022, 47: 2397-2417.
- [27] CASTRO F M, MARÍN-JIMÉNEZ M J, GUIL N, et al. End-to-end incremental learning [C] // Proceedings of the European Conference on Computer Vision. 2018:233-248.
- [28] HOU S, PAN X, LOY C C, et al. Learning a unified classifier incrementally via rebalancing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 831-839.
- [29] CHATZIMPAMPAS A, MARTINS R M, KERREN A. t-visne: Interactive assessment and interpretation of t-sne projections[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(8): 2696-2714.



ZHANG Yu, born in 1996, postgraduate. His main research interests include incremental learning and deep learning.



WANG Zhe, born in 1981, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include pattern recognition and image processing.