

# 基于 IMIU 的在线类增量对比学习\*

刘雨薇, 陈松灿

(南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106)

通信作者: 陈松灿, E-mail: [s.chen@nuaa.edu.cn](mailto:s.chen@nuaa.edu.cn)



**摘要:** 在线类增量连续学习旨在数据流场景下进行有效的新类学习, 并保证模型满足小缓存和小批次约束. 然而由于数据流的单趟 (one-pass) 特性, 小批次内的类别信息难以如离线学习那样被多趟探索利用. 为缓解该问题, 目前常采用数据多重增广并借助对比回放建模. 但考虑到小缓存和小批次限制, 现有随机选择和保存数据的策略不利于获取多样性的负样本, 制约了模型判别性. 已有研究表明困难负样本是提升对比学习性能的关键, 但这鲜少在在线学习场景被探索. Universum 学习提出的概念含混 (conduced) 数据恰好提供一种生成困难负样本的简单直观思路. 据此先前用特定系数插值混合 (mixup) 诱导出的 Universum 数据 (mixup-induced Universum, MIU) 已有效提升了离线对比学习的性能. 受此启发, 尝试将其引入在线场景. 但不同于先前静态生成的 Universum, 数据流场景面临着某些额外挑战. 首先随类数的动态增加, 相对基于全局给定类生成的静态 Universum 不再适用, 需重新加以定义和动态生成. 为此提出仅利用当前数据 (局部) 递归生成相对已见类熵为最大的 MIU (称为增量 MIU, IMIU), 并为其提供额外的小缓存从总体上满足内存限制; 其次将生成的 IMIU 和小批次内的正样本再次插值混合出多样且高质量的困难负样本. 最后综合上述各步, 发展出基于 IMIU 的在线类增量对比学习 (incrementally mixup-induced Universum based online class-increment contrastive learning, IUCL) 学习算法. 在标准数据集 CIFAR-10、CIFAR-100 和 Mini-ImageNet 上的对比实验验证所提算法一致的有效性.

**关键词:** 机器学习; 在线类增量学习; 对比学习; 插值混合; Universum

**中图法分类号:** TP18

中文引用格式: 刘雨薇, 陈松灿. 基于IMIU的在线类增量对比学习. 软件学报, 2024, 35(12): 5544–5557. <http://www.jos.org.cn/1000-9825/7094.htm>

英文引用格式: Liu YW, Chen SC. Online Class Incremental Contrastive Learning Based on Incremental Mixup-induced Universum. Ruan Jian Xue Bao/Journal of Software, 2024, 35(12): 5544–5557 (in Chinese). <http://www.jos.org.cn/1000-9825/7094.htm>

## Online Class Incremental Contrastive Learning Based on Incremental Mixup-induced Universum

LIU Yu-Wei, CHEN Song-Can

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** Online class-increment learning aims to learn new classes effectively under data stream scenarios and guarantee that the model meets the small cache and small batch constraints. However, due to the one-pass nature of data streams, it is difficult for the category information in small batches like offline learning to be exploited by multiple explorations. To alleviate this problem, current studies adopt multiple data augmentation combined with contrastive learning for model training. Nevertheless, considering the limitations of small cache and small batches, existing methods of selecting and storing data randomly are not conducive to obtaining diverse negative samples, which restricts the model discriminability. Previous studies have shown that hard negative samples are the key to improving contrastive learning performance, but this is rarely explored in online learning scenarios. The conduced data proposed in traditional Universum learning provides a simple yet intuitive strategy using hard negative samples. Specifically, this study has proposed mixup-induced Universum (MIU) with

\* 基金项目: 国家自然科学基金 (62076124); 南京航空航天大学研究生科研与实践创新计划 (xcxjh20221601)

收稿时间: 2023-07-25; 修改时间: 2023-09-16; 采用时间: 2023-11-22; jos 在线出版时间: 2024-06-12

CNKI 网络首发时间: 2024-06-14

certain coefficients previously, which effectively improves the performance of offline contrastive learning. Inspired by this, it tries to introduce MIU to online scenes, which is different from the previously statically generated Universum, and data stream scenarios face some additional challenges. Firstly, due to the increasing number of classes, the conventional approach of generating Universum based on globally given classes statically becomes inapplicable, necessitating redefinition and dynamic generation. Therefore, this study proposes to recursively generate MIU with the maximum entropy (incremental MIU, IMIU) relative to the seen (local) class and provides it with an additional small cache to meet the memory limit generally. Secondly, the generated IMIU and positive samples in small batches are mixed up together again to produce diverse and high-quality hard negative samples. Finally, by combining the above steps, the IMIU-based contrastive learning (IUCL) algorithm is developed. Meanwhile, comparison experiments on the standard datasets CIFAR-10, CIFAR-100, and Mini-ImageNet verify the validity of the proposed algorithm.

**Key words:** machine learning; online class incremental learning; contrastive learning; interpolation mixup; Universum

近年来,以人工神经网络为代表的机器学习方法在诸多领域取得了令人瞩目的进展<sup>[1,2]</sup>,尤其在训练数据整体给定的前提下.如今随着物联网传感器的增多,产生了一类永无止境并随时间连续变化的数据流,对其的学习打破了传统的离线学习要求,连续学习(continual learning, CL)<sup>[3]</sup>或终身学习<sup>[4]</sup>、顺序学习<sup>[5]</sup>的概念应运而生. CL旨在从数据流中积累新知识同时巩固已学旧知识,不断提升学习模型的性能.根据在训练和测试阶段是否提供任务标签,CL可划分为任务增量学习、域增量学习和类增量学习(class incremental learning, CIL)<sup>[6,7]</sup>,其中最具有现实价值的当属类增量学习<sup>[8]</sup>.目前CIL所面临的主要挑战为灾难性遗忘(catastrophic forgetting),对其的处理方法分为3种:1)参数隔离法,为不同任务分配对应参数;2)正则化方法,为网络参数更新施加约束;3)重放方法(replay),在缓存保留模型见过的部分旧知识,并在后继训练中重放旧数据.其中重放是一种相对简单而有效的学习策略<sup>[8]</sup>,因而获得了持续关注,并逐渐演变出了更有效的经验重放(experience replay, ER)<sup>[9]</sup>、最大干扰回放(maximally interfered retrieval, MIR)<sup>[10]</sup>、对抗 Shapley 值的经验重放(adversarial Shapley value experience replay, ASER)<sup>[11]</sup>等.我们也跟进该学习策略.

虽然基于重放的CIL在学新任务时不能获取旧任务的全部训练数据,但仍可利用它们的部分保留数据和当前任务的完整数据联合反复训练.这种反复训练难以适应现实中仅能单趟扫描的数据流,为应对这种场景,研究者提出了相应的在线CIL(online class incremental learning, OCIL)<sup>[12]</sup>,试图满足单趟扫描、有限内存和小批次处理的额外限制.这将面临如下的3个不充分:1)单趟扫描导致了模型对见过的训练数据挖掘不充分;2)有限内存导致了对先前知识保留的不充分;3)小批次导致了批次内训练样本多样性的不充分.这些不充分使得OCIL面临比CIL更棘手的灾难性遗忘<sup>[8]</sup>,目前的缓解策略可分为基于代理的重放和基于对比学习的重放,而结合类均值分类器(NCM)的后一种重放方法凭借其简单有效颇受关注<sup>[13-16]</sup>.尽管该策略缓解了灾难性遗忘,但数据流的在线特性带来的3个不充分仍制约了判别性的提高.具体来说,首先现有结合对比学习和数据流的重放方法大多采用随机存取旧样本的策略,导致可能缓存了判别性不足的样本,浪费有限的资源;其次,另外一些诸如MIR<sup>[10]</sup>的方法尽管通过数据筛选改进了ER,但仍不能保证为单趟扫描的新任务缓存所有已见类信息;最后,现有结合对比学习的方法仅在负样本信息丰富的大批次训练中效果相对较好,而一旦限于小批次,性能显著下降<sup>[17]</sup>.近来针对离线学习的已有研究表明困难负样本是有效提升对比学习判别性的关键<sup>[17,18]</sup>,但目前针对在线学习生成困难负样本的探究较少,本文工作尝试从该角度出发.

Vapnik最先提出了Universum作为额外引入并与训练样本非同类(或矛盾)的同域数据<sup>[19]</sup>,并用其设计了Universum-支持向量机(简称U-SVM)验证了在适当选用Universum前提下可提升SVM性能.自此,后续研究将该策略分别扩展至回归问题<sup>[20]</sup>、半监督学习<sup>[21]</sup>和对比学习<sup>[22]</sup>等.它们的一个共同发现是由于满足矛盾性质的数据极度丰富,不同构建或生成Universum的方式使性能变化差异显著,因此如何构建或生成Universum极为重要.直觉上密切相关于目标任务的Universum相对更合理<sup>[22]</sup>.近年来从给定任务的域内样本生成Universum获得了关注<sup>[22,23]</sup>,尤其用特定系数mixup诱导出的同域Universum(mixup-induced Universum, MIU)质量相对高,已在离线对比学习中获得了有效性验证<sup>[22]</sup>.然而据我们所知,至目前还未有针对在线学习开展的相关研究.我们尝试走出第1步,但并不平凡,这将面临随新类增加相应生成Universum的新挑战.为克服该挑战,我们首先发展出了增量MIU(IMIU),然后据此提出了在线类增量对比学习算法(IUCL).具体而言,我们设计了一种随类增加动态递归生

成并更新 Universum 的方法 (incremental mixup induced Universum, IMIU), 使其仅严格保证对已见类的矛盾性质. 值得注意的是, IMIU 作为一个具有一般性的独立策略, 不仅能用于 OCIL 提升对比学习性能, 而且还能辅助 CIL 和 DIL 等学习. 随后我们提出基于 IMIU 策略的 IUCL 方法, 同时实现其他功能, 包括: 1) 在内存限制下为其提供额外小缓存; 2) 提出再 mixup 策略 (remixup), 将 IMIU 和正样本再次混合增广生成多样且高质样本, 在小批次限制下保证样本的多样性以实现有效的对比学习. 作为与全体已见类矛盾的数据, Universum 如图 1 所示. 图 1 中, 左侧展示了已见类、Universum 和困难负样本在区域内分布示意图, 其中车、船、狗、飞机为已见类, Universum 为类外最含混样本, 困难负样本为再混合生成的更靠近类边界数据; 中间方框展示了生成 Universum 的过程; 右侧方框说明了困难负样本区别于 Universum, 由 Universum 和正样本再次 mixup 而成. 由图 1 可见, 车、船、狗和飞机为已见类, 相对这些类生成的含混数据 Universum 填充了类外空间, 其中每个 Universum 都是从每个已见类样本中任取其一加以平均而成. 随后再混合正样本和 Universum 获得期望的困难负样本, 它们在特征空间中更能趋近类边界, 如“+”字符所示. IUCL 就是利用了这些困难负样本辅助提升对比学习的判别性.

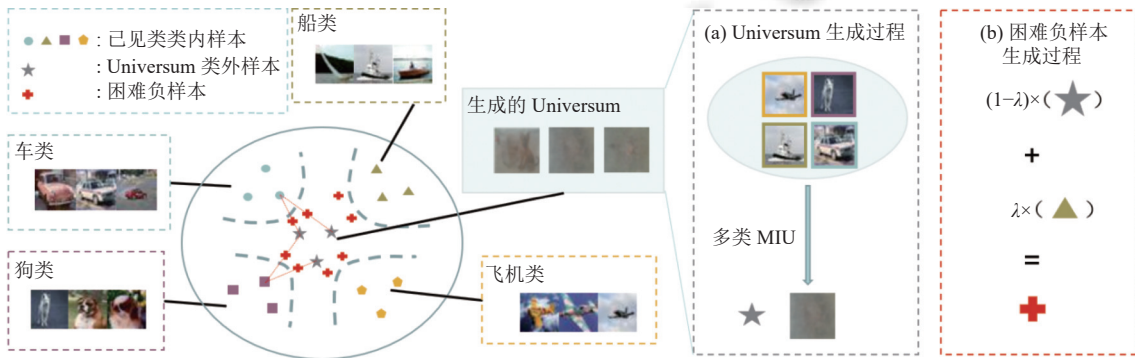


图 1 Universum 生成及其所在区域示意图

本文的主要贡献如下.

- (1) 提出了动态 IMIU 生成策略适应连续的 OCIL. 将原先相对于整体给定类定义的静态 Universum 改造为动态生成的局部 Universum (IMIU) 策略, 拓展了 Universum 的应用场景, 其生成策略具有独立可应用性.
- (2) 设计了基于 IMIU 的在线类增量对比学习算法 (IUCL). 其中通过将 Universum 和小批次中的正样本再混合生成多样而高质的困难负样本, 既满足缓存限制和也保证单趟扫描的模型判别性.
- (3) 实验验证了所提出的 IUCL 在多个基准数据集上的性能优越性.

## 1 相关工作

### 1.1 连续学习

CL<sup>[7,9,23-28]</sup>旨在按序学习以积累所获知识, 要求模型仅可利用当前任务的训练数据. 该过程会导致对旧任务训练数据的灾难性遗忘, 目前缓解该遗忘的策略主要分为 3 种: 1) 基于正则化的方法<sup>[7,24-27]</sup>为损失函数添加额外的正则项损失以巩固旧知识. 例如文献<sup>[24,25]</sup>等模型借助知识蒸馏减少模型对旧任务的特征漂移. 2) 基于参数隔离的方法<sup>[7,26-28]</sup>为每个任务分配不同的参数. 根据模型体系是否固定分为两类: 文献<sup>[7,27]</sup>等固定体系方法仅激活每个任务的相关参数但不改变体系. 而文献<sup>[5,27]</sup>等保持旧参数不变的同时, 通过引入新参数动态改变体系以适应新任务学习. 3) 基于回放的方法<sup>[9,28]</sup>为当前任务保存旧类的部分样本, 并在新任务训练时回放缓存数据. 基于该思想, Chaudhry 等人<sup>[9]</sup>最先提出了经验回放方法 (ER), 其通过从缓存中随机检索样本实现回放; 随后的 MER<sup>[28]</sup>利用元学习对齐改进回放策略, 进一步提升了性能.

由于回放策略简单高效, 因此通常被用来处理面对更棘手灾难性遗忘的 OCIL 场景<sup>[8]</sup>, 并通过改进 CIL 的重放方法以满足前文所述 OCIL 面临的 3 个限制. 针对检索策略的一些改进方法, 如 MIR<sup>[10]</sup>和 ASER<sup>[10]</sup>从缓存中选

择更重要的样本用于训练; 针对更新缓存数据策略的一些改进的方法筛选出更有效样本保存<sup>[29-31]</sup>; 针对改进模型的方法则通过调整模型更新策略提高学习效率<sup>[14,22]</sup>, 其中基于对比重演的 SCR 方法<sup>[13]</sup>用 NCM 分类器实现分类, 达到了优越性能. 为此我们也延用该模型, 但其仍面临小批次处理的挑战. 因为现有基于回放的对比学习仅限于在大批次中有效性验证, 而在小批次中性能显著下降. 对此 Gu 等人<sup>[32]</sup>提出了最大化双视图图像对间的互信息充分挖掘出语义信息而无需大量负样本, 因而在小批次内取得部分成功.

受到基于缓存数据更新策略的启发, 我们提出的基于回放对比学习的 IUCL 生成了矛盾于所有已见类的 Universum 数据, 这些数据包含了所有这些类信息, 可和任意正样本结合生成大量困难负样本辅助学习, 因而可作为富有高信息量的数据. 保存这些数据既可高效利用有限内存资源, 又缓解了小批次内负样本多样性的不足.

## 1.2 对比学习

近来对比学习为无监督表示学习提供了一个新策略, 并已在计算机视觉<sup>[33]</sup>、自然语言处理<sup>[34]</sup>、多模态<sup>[35]</sup>等领域取得成功. 其主要思想很直观: 拉近特征空间中相似样本而推远不同样本<sup>[35,36]</sup>. 典型的如 Chen 等人<sup>[35]</sup>首先设计了一个简单的 SimCLR 框架, 仅将同一样本增广视为正样本对; 随后 MoCo<sup>[36]</sup>提出了维护一个负样本队列并利用孪生网络提高队列一致性. 这两个框架已作为对比学习范式被后续工作广泛使用. 受到无监督对比学习成功的启发, SupCon<sup>[25]</sup>将对比学习引入监督学习之中, 使得同类正样本对相似性尽可能大, 异类样本对相似性尽可能小, 获得了比无对比的监督学习更优的性能. 最近对比学习也被引入至 OCIL 场景, 并结合简单高效的经典最近邻(类)均值分类器(NCM)实现了数据流的有效分类<sup>[13]</sup>. 但大多数(无论是面向静态或动态数据的)对比学习在学习时要求训练样本的大批次和多样性, 难以满足数据流限定的小批次. 尽管有少量针对静态场景的小批次对比学习研究, 典型的如 SwAV<sup>[37]</sup>采用多重增广策略初步表现出了有效性; Han 等人<sup>[22]</sup>通过引入 mixup 增广的(类含混)困难负样本(MIU)在显著缓解了大批次学习要求的同时提升了学习性能. 然而据我们所知, 目前尚未有针对 OCIL 场景解决该问题的研究. 对此我们提出了 IUCL 方法, 首先增量式动态生成高信息量 IMIU 适应数据流类增量的动态要求, 然后通过再混合正样本和 IMIU 增广高质且多样的困难负样本缓解小批次限制, 提高了模型判别性.

## 2 基于 IMIU 的在线类增量对比学习

OCIL 场景除了需要解决类增量问题外, 还要额外满足单趟扫描、小批次处理、小缓存限制 3 个特性. 现有 OCIL 方法多采用基于重放的方法缓存部分旧数据缓解灾难性遗忘, 其中对比重放凭借其简单有效<sup>[13]</sup>被我们提出的 IUCL 方法所采纳. 但目前的对比重放方法采用的随机回放策略不利于高质且多样的数据存储从而损害模型性能, 具体体现为: 1) 为有限的内存资源随机选取数据保存, 导致信息量少的数据不利于判别性提高; 2) 批次内样本多样性匮乏不利于对比学习训练, 因此我们考虑设计一种为随机回放方法生成高质且多样性数据的方法, 而基于 Universum 的研究恰好为此提供了思路. 已有研究表明, 作为矛盾于所有类的数据, 适当选取或生成 Universum 构建正则化项可有效提升模型判别性<sup>[19]</sup>, 且少量 Universum 即可增广出大量多样负样本辅助对比学习, 因此 Universum 可被看作高信息量样本. 但如前所述, 目前满足矛盾性质的 Universum 非常丰富, 其中相对静态全局生成的 MIU 在对比学习中已有突出表现. 受此启发, 我们将其拓展为相对局部数据生成的动态 MIU (IMIU) 以适应数据流类增量要求. 其构建方式为从每个已见类样本中任取其一加以平均而成, 等概率属于各个类, 即也可被看作不属于任何类. 这种生成方式的好处有两点: 1) 可在统计上证明生成的数据为相对已见类熵为最大的含混样本, 即蕴含所有已见类信息还能和任意正样本结合生成对应困难负样本有效辅助学习. 2) 简洁的混合系数形式便利了后续 IMIU 的动态递归更新. 随后我们为类增量设计 IMIU 动态更新策略, 使其在新类出现时递归修改混合系数并仅利用新类样本(局部)更新 Universum 缓冲区, 满足 Universum 实时地与所有已见类矛盾的要求. 值得注意的是, 尽管我们首先在 OCIL 场景下发展出 IMIU, 其也可作为一个独立策略结合 CIL 的其他方法构造正则化项辅助学习.

根据基于缓存数据更新策略相关研究, 保存这样的 Universum 数据在相同资源代价下能更有效提高模型的判别性. 但其作为区别于旧数据的额外数据信息, 需要被独立保存和回放, 因此我们基于 IMIU 提出的 IUCL 方法将缓存分为记忆缓冲区和 Universum 缓冲区. 记忆缓冲区保存已见类的部分样本, 而 Universum 缓冲区保存 Univer-

sum 数据. 然后我们利用 Universum 与正样本再混合, 将含混数据拉至类边界附近并作为困难负样本辅助对比学习. 综合上述各步骤构成了我们的 IUCL 方法, 其整体框架如图 2 所示. 其中流数据不断到来, 每次选取一个批次样本训练. 批次内的这些样本不仅作为正样本的一部分进行对比学习, 同时还用来更新记忆和 Universum 缓冲区. 此外参与正样本构成的还包括回放的记忆缓冲区样本. 对于负样本, IUCL 回放 Universum 并与所有正样本再混合生成对应的困难负样本. 随后以对比项添加至 SupCon 损失诱导出了我们基于 IMIU 的对比损失. 其中 Universum 的更新为: 首先判断新训练批次内样本是否属于新类, 如是则用新混合系数更新; 否则则将其和记忆缓冲区数据生成新 Universum 并更新.

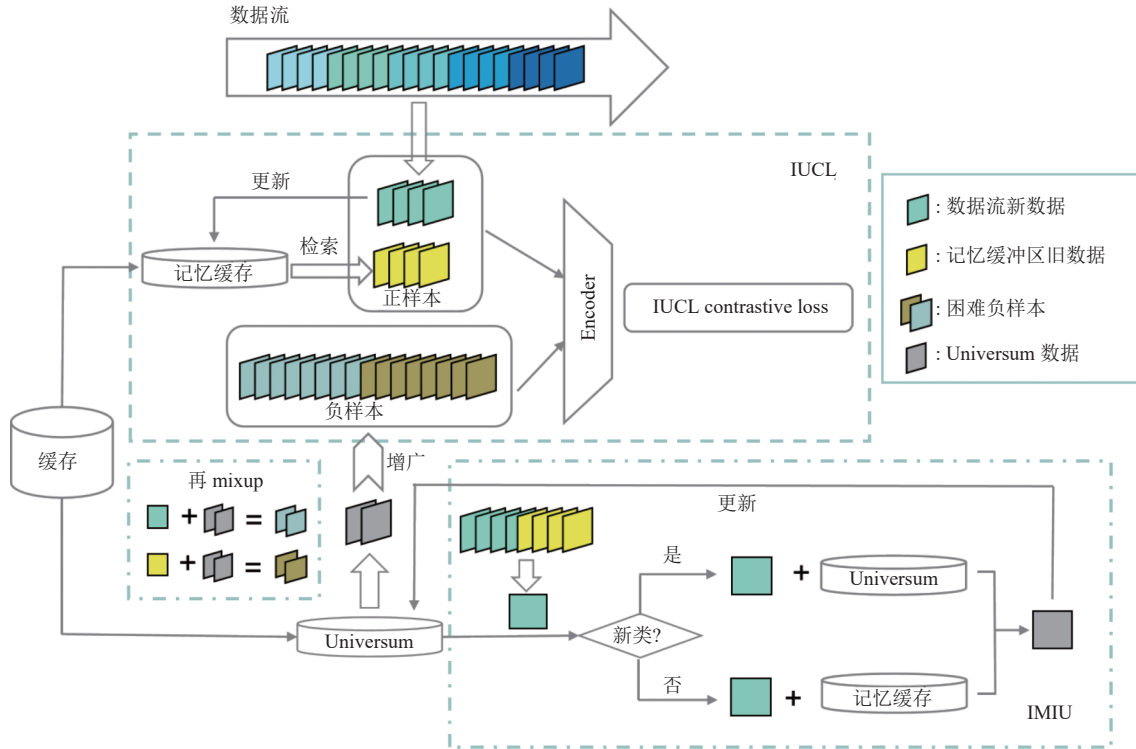


图 2 IUCL 算法框架

## 2.1 增量式生成 Universum (IMIU)

在为数据流动态生成 Universum 的实现中, 类随机增量出现导致各训练批次类先验不完整, 因而在批次内生成 Universum 的方法无法利用至目前出现的完整类信息. 为此我们实时更新记忆缓冲区并借助其保存的缺失类信息生成模糊于已见类的 Universum, 并为类增量设计了一种实时动态更新 Universum 缓冲区的方法, 即 IMIU, 使其满足单趟扫描限制. 其策略如后文图 3 所示, Universum 缓冲区旧数据由当前所有已见类样本任取其一平均而成, 保持了与所有已见类的矛盾性. 每次仅利用当前新批次训练数据更新 Universum 缓冲区, 保证了其能随数据流动态生成并更新. 具体而言针对批次内出现的新类, 用其和旧 Universum 数据以新系数混合递归更新缓冲区; 针对批次内出现的旧类则将其和记忆缓冲区样本混合生成新 Universum. 这种更新方法使 Universum 缓冲区动态实时保持对全局已见类的模糊性.

### 2.1.1 多类 MIU 混合

对于给定的输入及其标签对样本  $(x_i, y_i)$  和  $(x_j, y_j)$ , 其 mixup 由它们的线性凸组合生成如公式 (1) 所示:

$$\begin{cases} \tilde{x} = \lambda x_i + (1 - \lambda) x_j \\ \tilde{y} = \lambda y_i + (1 - \lambda) y_j \end{cases} \quad (1)$$

其中,  $\lambda \sim \text{Beta}(\alpha, \alpha)$ , 即采样于参数为  $\alpha$  的  $\text{Beta}$  分布, 其能被视为  $(x_i, y_i)$  出现的概率或所占比率. 对应  $\lambda$  为 1/2 的  $\text{mixup}$  形成了给定两类间的最模糊样本, 我们将其称为 MIU 样本. 到多类的 MIU 推广是直接的, 具体而言, 给定的一组多类输入标签样本对  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , 对于所有  $i \neq j$  存在  $y_i \neq y_j$ ,  $n$  表示类数. 它们的  $\text{mixup}$  为  $x = \sum_{i=1}^n \lambda_i x_i, y = \sum_{i=1}^n \lambda_i y_i$ , 其中  $\sum_{i=1}^n \lambda_i = 1$ . 定理 1 证明了当  $\lambda_1 = \lambda_2 = \dots = \lambda_n = 1/n$  时的特定  $\text{mixup}$ :  $u = \frac{1}{n} \cdot \sum_{i=1}^n x_i$  生成了多类 MIU.

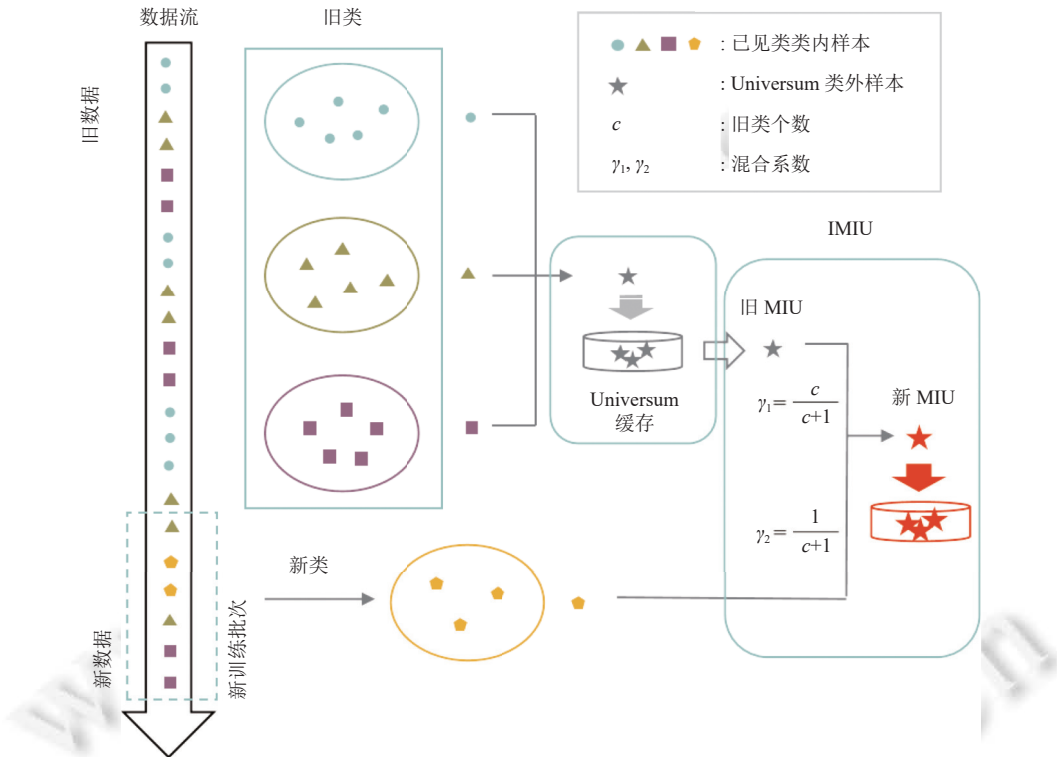


图 3 IMIU 流程图

**性质 1.** 给定  $n$  个已见类, 若从每个类选择一个样本混合得到相对各类熵为最大的含混数据, 那么对于每个类的混合系数都为  $1/n$ .

证明: 根据前文多类 MIU, 混合标签  $y = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$  满足  $\sum_{i=1}^n \lambda_i = 1, i \in n$ . 基于信息熵公式  $H(y) = -\sum_{i=1}^n \lambda_i \log(\lambda_i)$  使用拉格朗日法使  $H(y)$  在  $\sum_{i=1}^n \lambda_i = 1$  限制下值最大, 其中  $\xi$  是拉格朗日系数, 即:

$$f(\lambda_1, \lambda_2, \dots, \lambda_n, \xi) = -\sum_{i=1}^n \lambda_i \log(\lambda_i) + \xi \left( \sum_{i=1}^n \lambda_i - 1 \right) \quad (2)$$

求解公式 (2) 的极值, 分别对  $\lambda_i$  和  $\xi$  求导使其等于 0 并求解, 得到  $\lambda_1 = \lambda_2 = \dots = \lambda_n = 1/n$  时  $H(y)$  取最大, 此时数据的混合标签  $y$  具有熵为最大的含混性.

### 2.1.2 MIU 的增量生成

为数据流分类学习引入的 MIU 不仅要求满足多类形式, 还需要能够适应数据流动态生成. 由于其混合系数和类别数紧密关联, 而类数的增加打破了 Universum 仅限或相对已见类的含混性, 因此 MIU 需要随之更新. 为此我们特别设计了一种增量生成多类 MIU 策略, 使其模糊于所有旧类的同时还实时模糊于新增类.  $\mathcal{X}_i$  表示当前第  $i$  类

的数据, 从该类随机选取  $x_i \in \mathcal{X}_i$ .  $C$  表示某些类构成的下标集合,  $i \in C$  代表其中的第  $i$  类. 假设已见旧类集合为  $C_{old}$ , 新出现的类集合为  $C_{new}$ . 记  $n_{old} = |C_{old}|$ ,  $n_{new} = |C_{new}|$  分别表示类数. 我们的目标是基于相对  $C_{old}$  的 MIU 增量式生成相对  $C_{new} \cup C_{old}$  的新 MIU, 此时已见类总数为  $n_{old} + n_{new}$ . 由性质 1 可知, 当前 Universum 缓冲区中的数据可表示为:

$$u_{old} = \frac{1}{n_{old}} \sum_{i \in C_{old}} x_i, \quad B_{uni\_old} = \bigcup u_{old} \quad (3)$$

并基于定理 1 提出了增量生成 MIU (IMIU) 方法.

**定理 1.** 设当前相对  $C_{old}$  的 MIU 样本为  $u_{old}$ , 当一组新到的数据出现 (一个或多个) 新类  $C_{new}$  时, 可用公式 (4) 递推更新得相对  $C_{new} \cup C_{old}$  的 MIU, 即 IMIU.

证明: 根据性质 1 和多类 MIU 定义,  $u_{new}$  的混合系数  $\lambda_i = 1/(n_{old} + n_{new}), i \in n$ , 则有如下相对  $C_{new} \cup C_{old}$  的 MIU 表示:

$$\begin{aligned} u_{new} &= \frac{1}{n_{old} + n_{new}} \sum_{i \in C_{old} \cup C_{new}} x_i = \frac{1}{n_{old} + n_{new}} \sum_{i \in C_{old}} x_i + \frac{1}{n_{old} + n_{new}} \sum_{i \in C_{new}} x_i \\ &= \frac{n_{old}}{n_{old} + n_{new}} \cdot \frac{1}{n_{old}} \sum_{i \in C_{old}} x_i + \frac{1}{n_{old} + n_{new}} \sum_{i \in C_{new}} x_i = \frac{n_{old}}{n_{old} + n_{new}} u_{old} + \frac{1}{n_{old} + n_{new}} \sum_{i \in C_{new}} x_i \end{aligned} \quad (4)$$

更新的 MIU 相对  $C_{new} \cup C_{old}$  满足含混性, 简便了 MIU 的计算.

动态场景中我们需要根据是否出现新类采取不同策略更新 Universum 缓冲区.

(1) 若出现新类, 则根据定理 1 增量式生成 Universum (IMIU), 得到相对所有新旧类 MIU  $u_{new}$ , 此时更新缓冲区为:

$$B_{uni\_new} = \bigcup u_{new} \quad (5)$$

(2) 若无新类出现, 则基于性质 1 利用来自旧类的新数据和所存的部分旧数据共同生成同样相对于所有旧类的 MIU ( $u_{old}$ ), 并按照先进先出原则更新 Universum 缓冲区:

$$B_{uni\_new} = B_{uni\_old} \cup \bigcup u_{old} \quad (6)$$

这种 mixup 诱导 Universum 的独特策略也可不涉及标签的混合, 因此具有独立性, 可额外正则化模型进行辅助学习.

## 2.2 在线类增量对比学习算法 (IUCL)

### 2.2.1 问题设定

本文遵循监督型 OCIL 的问题设定<sup>[8,10,11]</sup>, 即在模型学习数据流中新类的同时, 样本仅被单趟扫描. 具体来说, 数据流中各任务类相互独立, 即  $\{D_i\} \cap \{D_j\} = \emptyset$ ,  $\{D_k\}$  表示第  $k$  个任务的训练数据集, 任务  $k$  的目标是在  $D_i$  上分类的同时保持在  $D_i, i < t$  上的分类能力. 模型由两部分组成: 一个将输出样本映射到  $d$  维特征向量的编码器  $f: X \mapsto \mathbb{R}^d$  和一个将嵌入映射到  $c$  类输出的分类器  $g: \mathbb{R}^d \mapsto \mathbb{R}^c$ .  $t$  时刻输入的小批次  $B_t$ ,  $t$  时刻的 Universum 缓存  $M_{t-1}^{umi}$  和  $t-1$  时刻记忆缓存  $M_{t-1}^m$  的部分数据生成 IMIU 并更新  $M_t^{umi}$ , 随后利用  $B_t, M_{t-1}^m, M_t^{umi}$  协同训练模型, 即  $A_t$ :

$$A_t: \begin{cases} \langle B_t, M_{t-1}^m, M_{t-1}^{umi} \rangle \rightarrow \langle M_t^{umi} \rangle \\ \langle (f, g)_{t-1}, B_t, M_{t-1}^m, M_t^{umi} \rangle \rightarrow \langle (f, g) \rangle \\ \langle B_t, M_{t-1}^m \rangle \rightarrow \langle M_t^m \rangle \end{cases}$$

### 2.2.2 问题设定

由于数据流场景下训练数据带标签信息, 因此我们着重关注监督对比学习 (SCL)<sup>[33,34,38]</sup>. 具体实现框架遵循以前的工作<sup>[38]</sup>由 3 部分组成: 1) 样本增广. 为每个样本  $x$  生成增广样本  $\tilde{x} = Aug(x)$ ; 2) 特征编码. 将样本映射到特征嵌入  $r = Enc(x) \in \mathbb{R}^{D_E}$ ; 3) 投影映射. 将特征  $r$  映射到投影向量  $z = Proj(r) \in \mathbb{R}^{D_P}$  随后  $\mathcal{L}_2$  正则化. 其中首先对长度为  $b$  的小批次  $B = \{x_k, y_k\}_{k=1, \dots, b}$  中每个样本生成对应增广样本作为正样本, 即:

$$B_t = B \cup \tilde{B} = \{x_k, y_k\}_{k=1, \dots, b} \cup \{\tilde{x}_k = Aug(x_k), y_k\}_{k=1, \dots, b} \quad (7)$$

SCL 的损失函数为:

$$\mathcal{L}_{SCL}(Z_I) = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{j \in A(i)} \exp(z_i \cdot z_j / \tau)} \quad (8)$$

其中,  $i$  为  $B_I$  的索引,  $A(i) = I \setminus \{i\}$  表示  $B_I$  中除了第  $i$  个样本外其他样本索引,  $P(i) \equiv \{p \in A(i) : y_p = y_i\}$  表示所有和第  $i$  个元素有相同标签的样本索引,  $|P(i)|$  是  $P(i)$  的基,  $Z_I = \{z_i\}_{i \in I} = \{Proj(Enc(x_i))\}_{i \in I}$ ,  $\tau \in \mathbb{R}^+$  是温度调节参数.

IUCL 在数据流场景中为 SCL 引入 IMIU, 通过再 mixup 生成困难负样本辅助训练. 借鉴监督对比回放<sup>[11]</sup>,  $t$  时刻回放的旧样本  $B_t^m \subset M_{t-1}^m$ ; 回放的 Universum 数据  $B_t^{umi} \subset M_t^{umi}$ . 当前批次训练数据分为:  $B_t$ 、 $B_t^m$ 、 $B_t^{umi}$ , 其中正样本如公式 (9) 所示:

$$B_t = B_t \cup B_t^m \cup \tilde{B}_t \cup \tilde{B}_t^m \quad (9)$$

已有研究表明靠近决策边界的困难负样本可有效提升对比学习性能<sup>[17,18]</sup>, 但由于将 Universum 直接作为所有类的困难负样本效果不好, 因此我们将所有  $B_t^{umi}$  中的 Universum 和正样本再 mixup 得到更靠近对应类决策边界的困难负样本辅助训练.

$$\bar{x}_i = \lambda \cdot x_i + (1 - \lambda) \cdot x_u \quad (10)$$

其中,  $x_i \in B_t$ ,  $x_u \in B_t^{umi}$ , 为  $\bar{x}_i$  赋  $x_i$  对应的标记  $y_i$ , 表示为  $y_i$  类的负样本. 每个正样本  $x_i$  都可以得到对应  $|B_t^{umi}|$  个困难负样本  $\bar{x}_i$ , 将其组成集合  $U_i = \cup \{\bar{x}_i, y_i\}$ . 所有  $x_i$  的  $U_i$  集合为  $U = \bigcup_{i=1}^{|B_t|} U_i$ . 结合 IMIU 的损失函数如公式 (11):

$$\begin{cases} \mathcal{L}_{IMIU}(Z_I) = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{j \in A(i)} \exp(z_i \cdot z_j / \tau) + \alpha \cdot \sum_{k \in U(i)} \exp(z_i \cdot z_k / \tau)} \\ U(i) = \{k \mid \{x_k, y_k\} \in U, y_k = y_i\} \end{cases} \quad (11)$$

其中, 相关符号定义和 SCL 损失函数一致,  $U(i)$  为  $U$  中样本  $x_i$  对应类的所有困难负样本下标集合. 这些样本可能非  $x_i$  和 Universum 混合而得, 但同属一类, 因此也可被视为  $x_i$  的困难负样本参与训练. 此外, 我们在  $\mathcal{L}_{IMIU}$  基础上加入交叉熵损失监督学习, 最后训练的损失函数为:

$$\mathcal{L}_{IUCL} = \beta \cdot \mathcal{L}_{IMIU} + \mathcal{L}_{CE} \quad (12)$$

### 2.2.3 NCM 分类器

文献 [13] 的研究表明由于数据流场景存在新旧类数据的不平衡, Softmax 分类器的 FC 层对新类权值比旧类会大得多, 因此导致对最近任务的预测偏差, 产生灾难性遗忘. NCM 分类器<sup>[39]</sup>基于样本与各类均值的相似性度量分类, 其因不涉及 FC 层且结构简单而广泛被连续学习相关研究采纳. 由此, 我们在训练好特征提取网络后用 NCM 分类器分类, 将样本的嵌入与原型进行比较并给最相似的原型分配类标签:

$$\mu_c = \frac{1}{n_c} \sum_i f(x_i) \cdot \mathbb{R}\{y_i = c\} \quad (13)$$

$$y^* = \arg \min_{c=1, \dots, J} \|f(x) - \mu_c\| \quad (14)$$

其中,  $n_c$  为  $c$  类样本数,  $\mathbb{R}\{y_i = c\}$  表示所有标签为  $c$  的样本下标. IUCL 的特征学习为 NCM 分类器提供了高质量的特征, 共同促进数据流分类学习. 算法 1 总结了设计的 IUCL 算法.

---

#### 算法 1. IUCL 算法.

---

输入:  $C$ : 类数;  $n_c$ : 缓冲区各类样本数;

输出: 分类结果  $y^*$ .

---

初始化: 缓存  $\mathcal{M} \leftarrow \{\} * M$ ;  $\mathcal{M}_u \leftarrow \{\} * M_u$ ;  $Aug(\cdot)$ ;  $Enc_\theta(\cdot)$ ;  $Proj_\phi(\cdot)$ ;

1. **for**  $n \in \{1, \dots, N\}$  **do**

2. 训练阶段

3. **for**  $B_n \sim D_n$  **do**

---



- 
4.  $B_M \leftarrow \text{MemoryRetrieval}(B_n, M)$
  5.  $B_{nM} \leftarrow B_n \cup B_M$
  6.  $B_I \leftarrow B_{nM} \cup \text{Aug}(B_{nM})$
  7.  $Z_I \leftarrow \text{Proj}_\phi(\text{Enc}_\theta(B_I))$
  8.  $M \leftarrow \text{MemoryUpdate}(B_n, M)$
  9.  $M_u \leftarrow \text{UniversumUpdate}(B_n, M_u)$  //参考公式 (5)、公式 (6)
  10.  $B_u \leftarrow \text{UniversumRetrieval}(M_u)$
  11.  $U(i) \leftarrow \text{CalculateHardNegative}(B_I, B_u)$  //公式 (10)
  12.  $Z_u \leftarrow \text{Proj}_\phi(\text{Enc}_\theta(U(i)))$
  13.  $\theta, \phi \leftarrow \text{SGD}(\mathcal{L}_{\text{UCL}}(Z_I, Z_u), \theta, \phi)$  //公式 (12)
  14. **end for**
  15. 训练阶段
  16. **for**  $c \in \{1, \dots, C\}$  **do**
  17.  $\mu_c = \frac{1}{n_c} \sum_i^{|M|} \text{Enc}_\theta(x_i) \cdot \mathbb{I}\{y_i = c\}$
  18. **end for**
  19.  $y^* = \underset{c=1, \dots, t}{\text{argmin}} \|\text{Enc}_\theta(x) - \mu_c\|$
  20. **end for**
- 

### 3 实验分析

在本节中, 我们给出了相关比较方法的基准测试数据集、度量标准以及基准工作, 并将我们的方法和这些方法比较, 总结和分析了实验结果验证了我们方法的有效性.

#### 3.1 实验设定

##### 3.1.1 数据集

(1) Split CIFAR-10: 将 CIFAR-10 数据集分割成 5 个部分<sup>[22]</sup>, 每个部分为一个任务并且每个任务之间包含 2 个不重叠的类.

(2) Split CIFAR-100: 将 CIFAR-100 数据集分割成 10 个部分<sup>[22]</sup>, 每个部分为一个任务并且每个任务间包含 10 个不重叠的类.

(3) Split Mini-ImageNet: 将 Mini-ImageNet 数据集分割成 10 个部分<sup>[22]</sup>, 每个部分为一个任务并且每个任务间包含 10 个不重叠的类.

##### 3.1.2 评估指标

由于连续学习建立的模型要求保证能够在之前的任务上保持较好的性能, 而减少对之前任务学习到知识的遗忘. 由此引入性能指标评估模型性能.

平均精度: 评估在所见任务的测试机上的整体性能.  $a_{i,j}$  表示模型从任务 1 训练到任务  $i$  后, 在第  $j$  个任务的保留测试集上的表现.

$$\text{Average Accuracy}(A_N) = \frac{1}{N} \sum_{j=1}^N a_{N,j}.$$

##### 3.1.3 对比方法

我们将 IMIU 和以下几个目前最先进的方法 fine-tune、iid-offline、EWC<sup>[40]</sup>、A-GEM<sup>[25]</sup>、ER<sup>[9]</sup>、MIR<sup>[10]</sup>、GSS<sup>[29]</sup>、DER++<sup>[41]</sup>、GDumb<sup>[42]</sup>、ASER<sup>[11]</sup>、SCR<sup>[13]</sup>、DVC<sup>[32]</sup>比较. 为保证对比公平, 将回放的批次大小都控制为 10.

fine-tune: 不采取任何遗忘预防措施, 仅按照数据到来顺序训练模型的持续学习方法.

iid-offline: 非连续学习方法, 但是连续学习方法的性能上限, 即在独立同分布的训练数据集上反复多次训练.

EWC (elastic weight consolidation): 通过限制对旧任务重要参数的更新保留对旧任务重要的信息, 由 fisher 信息矩阵度量.

GEM (averaged gradient episodic memory): 一种使用内存中的样本约束参数更新的方法.

GSS (gradient-based sample selection): 一种使重放存储器中样本的梯度多样化的重放方法.

DER++ (dark experience replay): 利用知识蒸馏来保留过去的经验.

GDumb (greedy sampler and dumb learner): 贪婪地更新内存缓冲区, 以保持类平衡分布. 每次只使用平衡的内存缓冲区从头开始训练一个模型.

### 3.1.4 实验配置

实验主干网络为简化的 ResNet18, 并采取随机梯度下降 (SGD) 方法优化网络, 学习率设为 0.1, 训练迭代次数 epoch 为 1 保证单趟扫描. 从流数据中获取训练样本的批次大小为 10, 从记忆缓冲区中获取就样本的批次大小为 10, 从 Universum 缓冲区中获取 Universum 样本的批次大小设为 5. 训练的投影网络为具有一个隐层 (ReLU) 的多层感知机 (MLP), 分类器为 NCM 分类器. 记忆和 Universum 缓冲区都通过随机选取进行样本采样.

## 3.2 实验结果

### 3.2.1 实验结果分析

我们在 CIFAR-10、CIFAR-100 和 Mini-ImageNet 这 3 个数据集上将我们的 IUCL 方法与对比方法进行了比较, 评估指标为最后一个任务训练结束后, 模型对之前所有任务测试集上准确率的平均值 (表 1). 实验结果显示, 相比于各个 OCIL 方法, 我们的方法在各个标准数据集上都有显著的提高. 具体来说, 在 CIFAR-100 数据集上我们的方法与 DVC 方法相比, 准确率平均提高了 6%; 在 CIFAR-10 数据集上平均提高了 5%; 在 Mini-ImageNet 数据集上平均提高了 3.1%. 从稳定性角度来看, 在 CIFAR-100 和 Mini-ImageNet 上, 我们的算法表现较 CIFAR-10 更为稳定. 在 CIFAR-10 上, 不同样本数量的性能提升差异较大, 样本数量为 200、500、1000 时分别提升了约 1%、5%、10%. 而在 CIFAR-100 数据集上, 样本数量分别为 1000、2000、5000 时的提升分别约为 5.5%、6%、6%, 相比之下稳定性较高. 这可能是因为 CIFAR-10 数据集包含的类别较少, 属于相对简单的数据集, 现有方法已经取得了优秀的效果, 因此改进方法只能有限地提高性能. 而对于较为复杂的 CIFAR-100 和 Mini-ImageNet 数据集, 我们的方法有更大的提升空间. 从内存角度来看, 随着内存增加, 准确率逐渐提高的结果符合直觉和目前 OCIL 的研究结果. 我们的方法在 CIFAR-10、CIFAR-100 和 Mini-ImageNet 数据集上同样保持了这一趋势, 且无论内存大小准确率都有所提升.

表 1 训练至最后一个任务的平均准确率 (%)

Method	Mini-ImageNet			CIFAR-100			CIFAR-10		
	M=1000	M=2000	M=5000	M=1000	M=2000	M=5000	M=200	M=500	M=1000
fine-tune	4.3±0.2	4.3±0.2	4.3±0.2	5.8±0.3	5.8±0.3	5.8±0.3	18.1±0.3	18.1±0.3	18.1±0.3
iid-offline	51.4±0.2	51.4±0.2	51.4±0.2	49.6±0.2	49.6±0.2	49.6±0.2	81.7±0.1	81.7±0.1	81.7±0.1
EWC	3.1±0.1	3.1±0.1	3.1±0.1	4.8±0.2	4.8±0.2	4.8±0.2	17.9±0.3	17.9±0.3	17.9±0.3
A-GEM	4.4±0.2	4.3±0.2	4.3±0.2	6.0±0.1	6.0±0.1	5.9±0.2	18.1±0.3	18.3±0.1	18.3±0.1
ER	10.2±0.5	12.9±0.8	16.4±0.9	11.6±0.5	15.0±0.5	20.5±0.8	23.2±1.0	31.2±1.4	39.7±1.3
GSS	9.3±0.8	14.1±1.1	15.0±1.1	9.7±0.2	12.4±0.6	16.8±0.8	23.0±0.9	28.3±1.7	37.1±1.6
MIR	10.1±0.6	14.2±0.9	18.5±1.0	11.3±0.3	15.1±0.3	22.2±0.7	24.6±0.6	32.5±1.5	42.8±1.4
GDumb	7.3±0.3	11.4±0.2	19.5±0.5	10.0±0.2	13.3±0.4	19.2±0.4	26.6±1.0	31.9±0.9	37.5±1.1
DER++	10.9±0.6	15.0±0.7	17.4±1.5	11.8±0.4	15.7±0.5	20.8±0.8	28.1±1.2	35.4±1.3	42.8±1.9
ASER	11.5±0.6	13.5±0.8	17.8±1.0	14.3±0.5	17.8±0.5	22.8±1.0	29.6±1.0	38.2±1.0	45.1±2.0
SCR	12.9±0.7	14.3±1.4	16.2±1.3	13.1±0.3	14.9±0.2	16.3±0.4	<b>48.2±5.0</b>	<u>55.1±2.3</u>	<u>57.8±1.1</u>
DVC	<u>15.4±0.7</u>	<u>17.2±0.8</u>	<u>19.1±0.9</u>	<u>19.7±0.7</u>	<u>22.1±0.9</u>	<u>24.1±0.8</u>	45.4±1.4	50.6±2.9	52.1±2.5
<b>Ours</b>	<b>17.7±1.0</b>	<b>20.4±1.3</b>	<b>22.8±1.8</b>	<b>25.4±1.1</b>	<b>28.6±1.5</b>	<b>30.2±2.2</b>	<u>46.2±2.8</u>	<b>55.4±3.2</b>	<b>60.7±2.9</b>

注: M表示内存大小; 加粗为目前最好的实验结果, 下划线为次好的实验结果

图 4、图 5 分别在 CIFAR-10、CIFAR-100 数据上比较了各 OCIL 方法在各个任务阶段的平均准确率, 比较了几个目前即经典又有效的方法: A-GEM、ER、GSS、MIR、DSR++、GDumb、ASER、DVC. 通过同一种方法在不同任务上的准确率, 我们可以得出该算法在各个任务的效果以及发展趋势, 从而体现算法性能. 图 4 展示了在 CIFAR-100 数据集下各算法的对比结果, 从图 4 中可以观察到, 在刚训练完第 1 个任务时, 我们的方法准确率可能不及 DER++ 和 ASER, 然而, OCIL 不能仅关注第 1 个任务, 还需要保证在后续任务上准确性. 从第 2 个任务开始 IUCL 相比对比算法展现出突出的优势, 并在后续任务中稳定保持这个优势. 类似的结果在 CIFAR-100 数据集上也得到验证. 有趣的是, MIR 在 CIFAR-100 数据集的前 4 个任务上表现优于我们的方法, 但是由于算法的不稳定性, 导致在后续的任务测试集上准确度迅速下降, 最终效果不如 IUCL. 这些对比结果表明, 我们的 IUCL 方法在数据流场景中表现出卓越的稳定性和准确性, 相比其他经典有效的 OCIL 方法具有更优越的性能.

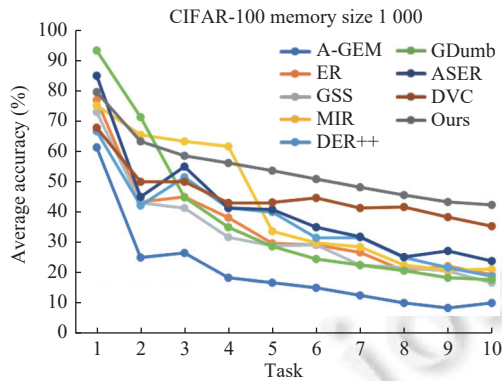


图 4 CIFAR-100 数据集上各阶段结果对比

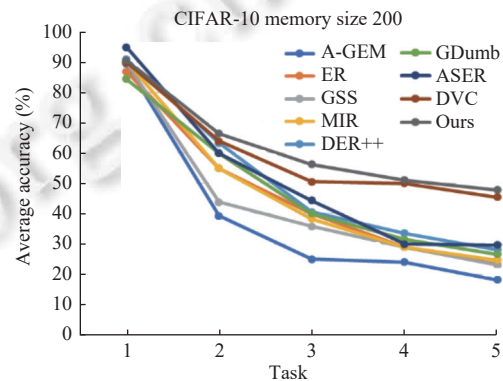


图 5 CIFAR-10 数据集上各阶段结果对比

### 3.2.2 消融实验

在本节中我们探究引入 IMIU 对于基线算法的影响, 并选取 SCR 算法作为基线. 传统的 SCR 方法要求使用大批次训练对比学习模型, 因此记忆缓冲区回放的批次大小为 100. 我们的方法克服了这一限制, 并确保在缓存回放大小为 10 时仍有效. 为确保公平性, 我们将 SCR 的记忆缓冲区回放批次大小与我们的 IUCL 保持一致, 都设置为 10, 并将此修改后的 SCR 作为基线方法. 在表 2 中, 我们展示了引入 IMIU 策略后, 我们的方法和基线方法在 CIFAR-100 数据集上的准确率有明显提升. 具体而言, 当内存大小分别为 1000、2000、5000 时, 我们的方法分别提升了 12%、14% 和 14% 的准确率. 然而, 在 CIFAR-10 数据集上的实验结果对比发现, 随着内存缓冲区大小的减小, Universum 作为困难负样本对对比学习性能的提高效果变得较小, 甚至在内存缓冲区为最小的情况下, 为 200 时, 最后一个任务的平均准确率甚至不如基线方法. 这可能是由于在数据集相对简单且类数目较少的情况下, 有监督对比学习已经能够学习到较好的特征用于解决简单任务, 而生成的 Universum 没有展现足够的多样性, 因此不能体现 Universum 对结果的正向影响.

表 2 CIFAR-100 和 CIFAR-10 数据集上的 IMIU 消融实验 (%)

Method	CIFAR-100			CIFAR-10		
	$M=1000$	$M=2000$	$M=5000$	$M=200$	$M=500$	$M=1000$
Baseline	13.1±0.3	14.9±0.2	16.3±0.4	48.2±5.0	55.1±2.3	57.8±1.1
Baseline+IMIU	25.4±1.1	28.6±1.5	30.2±2.2	46.2±2.8	55.4±3.2	60.7±2.9

IMIU 作为一种在类增量场景下动态递归生成 Universum 的策略具有广泛性, 可作为一种独立方法, 广泛和其他 CL 算法结合构成正则化项辅助学习. 本文仅选取 OCIL 的对比重放模型作为验证 IMIU 有效性的一个实现场景.

## 4 总结与展望

在 OCIL 场景中, 对比回放作为一种缓解灾难性遗忘方法已经取得了一定的成果, 我们进一步阐明了引入困难负样本的简单有效性. 由于 OCIL 小批次、小缓存和单趟扫描的限制, 目前现有的静态生成困难负样本的方法无法满足需求. 幸运的是, 我们发现 Universum 可以作为生成大量困难负样本的基础, 因为它是相对于已见类矛盾的数据. 在众多 Universum 选取或生成方法中, 通过 mixup 生成 Universum 的方法由于其简单有效以及在小批次中的突出优势被采纳. 然而, 目前现有的静态方法不能直接应用到数据流动态场景. 因此, 我们在此基础上拓展并提出了增量式生成 Universum 的策略 (IMIU). IMIU 不仅解决了动态场景中的类增量困境, 同时其生成的高质量含混数据可构建正则化项, 有效辅助对比学习. 基于此, 我们为 OCIL 的对比回放引入 IMIU, 并通过其与正样本的再混合生成多样且高质困难负样本. 这一策略不仅考虑了数据流的缓存限制, 还能够满足为小批次样本生成多样困难负样本辅助学习的需求, 恰好解决了在线带来的困难. 最终, 通过实验结果我们发现, 使用 IMIU 为对比学习生成困难负样本的 IUCL 方法相较于现有的 CL 和 OCIL 方法具有显著优势, 训练完新任务的模型能够更好地保持旧任务的效果, 并且针对 OCIL 场景进行更有效的学习. 最后, 我们阐明了 IMIU 作为一种独立于所有算法的方法, 能够提升 CIL、DIL 领域其他模型的性能. 下一步的工作将面临更大的挑战, 针对本文场景和提出的算法面临的鲁棒性不足的问题进行进一步分析研究.

## References:

- [1] Hall T, Beecham S, Bowes D, Gray D, Counsell S. A systematic literature review on fault prediction performance in software engineering. *IEEE Trans. on Software Engineering*, 2012, 38(6): 1276–1304. [doi: [10.1109/TSE.2011.103](https://doi.org/10.1109/TSE.2011.103)]
- [2] Yang K, Liu D, Chen ZB, Wu F, Li WP. Spatiotemporal generative adversarial network-based dynamic texture synthesis for surveillance video coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 2022, 32(1): 359–373. [doi: [10.1109/TCSVT.2021.3061153](https://doi.org/10.1109/TCSVT.2021.3061153)]
- [3] Ring MB. Continual learning in reinforcement environments [Ph.D. Thesis]. Austin: University of Texas at Austin, 1994.
- [4] Zhai MY, Chen L, Mori G. Hyper-lifelongGAN: Scalable lifelong learning for image conditioned generation. In: *Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 2246–2255. [doi: [10.1109/CVPR46437.2021.00228](https://doi.org/10.1109/CVPR46437.2021.00228)]
- [5] Lesort T, Stoian A, Filliat D. Regularization shortcomings for continual learning. arXiv:1912.03049, 2021.
- [6] Aljundi R, Babiloni F, Elhoseiny M, Rohrbach M, Tuytelaars T. Memory aware synapses: Learning what (not) to forget. In: *Proc. of the 15th European Conf. on Computer Vision*. Munich: Springer, 2018. 144–161. [doi: [10.1007/978-3-030-01219-9\\_9](https://doi.org/10.1007/978-3-030-01219-9_9)]
- [7] Lee S, Ha J, Zhang DS, Kim G. A neural dirichlet process mixture model for task-free continual learning. In: *Proc. of the 8th Int'l Conf. on Learning Representations*. Addis Ababa: OpenReview.net, 2020.
- [8] Mai ZD, Li RW, Jeong J, Quispe D, Kim H, Sanner S. Online continual learning in image classification: An empirical survey. *Neurocomputing*, 2022, 469: 28–51. [doi: [10.1016/j.neucom.2021.10.021](https://doi.org/10.1016/j.neucom.2021.10.021)]
- [9] Chaudhry A, Rohrbach M, Elhoseiny M, Ajanthan T, Dokania PK, Torr PHS, Ranzato M. On tiny episodic memories in continual learning. arXiv:1902.10486, 2019.
- [10] Aljundi R, Caccia L, Belilovsky E, Caccia M. Online continual learning with maximally interfered retrieval. In: *Proc. of the 33rd Conf. on Neural Information Processing Systems*. Vancouver: NeurIPS, 2019.
- [11] Shim D, Mai ZD, Jeong J, Sanner S, Kim H, Jang J. Online class-incremental continual learning with adversarial Shapley value. In: *Proc. of the 35th AAAI Conf. on Artificial Intelligence*. AAAI, 2021. 9630–9638. [doi: [10.1609/aaai.v35i11.17159](https://doi.org/10.1609/aaai.v35i11.17159)]
- [12] Hu QH, Gao YC, Cao B. Curiosity-driven class-incremental learning via adaptive sample selection. *IEEE Trans. on Circuits and Systems for Video Technology*, 2022, 32(12): 8660–8673. [doi: [10.1109/TCSVT.2022.3196092](https://doi.org/10.1109/TCSVT.2022.3196092)]
- [13] Mai ZD, Li RW, Kim H, Sanner S. Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning. In: *Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*. Nashville: IEEE, 2021. 3584–3594. [doi: [10.1109/CVPRW53098.2021.00398](https://doi.org/10.1109/CVPRW53098.2021.00398)]
- [14] Guo YD, Liu B, Zhao DY. Online continual learning through mutual information maximization. In: *Proc. of the 39th Int'l Conf. on Machine Learning*. Baltimore: PMLR, 2022. 8109–8126.
- [15] Cha H, Lee J, Shin J. Co2L: Contrastive continual learning. In: *Proc. of the 2021 IEEE/CVF Int'l Conf. on Computer Vision*. Montreal: IEEE, 2021. 9496–9505. [doi: [10.1109/ICCV48922.2021.00938](https://doi.org/10.1109/ICCV48922.2021.00938)]
- [16] Lin HW, Zhang BQ, Feng SS, Li XT, Ye YM. PCR: Proxy-based contrastive replay for online class-incremental continual learning. In:

- Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 24246–24255. [doi: [10.1109/CVPR52729.2023.02322](https://doi.org/10.1109/CVPR52729.2023.02322)]
- [17] Cai TT, Frankle J, Schwab DJ, Morcos AS. Are all negatives created equal in contrastive instance discrimination? arXiv:2010.06682, 2020.
- [18] Robinson JD, Chuang CY, Sra S, Jegelka S. Contrastive learning with hard negative samples. In: Proc. of the 9th Int'l Conf. on Learning Representations. ICLR, 2021.
- [19] Vapnik V. Estimation of Dependences Based on Empirical Data. New York: Springer, 2006.
- [20] Dhar S, Cherkassky V. Universum learning for SVM regression. In: Proc. of the 2017 Int'l Joint Conf. on Neural Networks. Anchorage: IEEE, 2017. 3641–3648. [doi: [10.1109/IJCNN.2017.7966314](https://doi.org/10.1109/IJCNN.2017.7966314)]
- [21] Zhang D, Wang JD, Wang F, Zhang CS. Semi-supervised classification with Universum. In: Proc. of the 2008 SIAM Int'l Conf. on Data Mining. Philadelphia: Society for Industrial and Applied Mathematics, 2008. 323–333. [doi: [10.1137/1.9781611972788.29](https://doi.org/10.1137/1.9781611972788.29)]
- [22] Han AY, Geng CX, Chen SC. Universum-inspired supervised contrastive learning. IEEE Trans. on Image Processing, 2023, 32: 4275–4286. [doi: [10.1109/TIP.2023.3290514](https://doi.org/10.1109/TIP.2023.3290514)]
- [23] Zhang EH, Geng CX, Chen SC. Class-aware Universum inspired re-balance learning for long-tailed recognition. arXiv:2207.12808, 2022.
- [24] Rebuffi SA, Kolesnikov A, Sperl G, Lampert CH. iCaRL: Incremental classifier and representation learning. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5533–5542. [doi: [10.1109/CVPR.2017.587](https://doi.org/10.1109/CVPR.2017.587)]
- [25] Chaudhry A, Ranzato MA, Rohrbach M, Elhoseiny M. Efficient lifelong learning with A-GEM. In: Proc. of the 7th Int'l Conf. on Learning Representations. New Orleans: OpenReview.net, 2019.
- [26] Fernando C, Banarse D, Blundell C, Zwols Y, Ha D, Rusu AA, Pritzel A, Wierstra D. PathNet: Evolution channels gradient descent in super neural networks. arXiv:1701.08734, 2017.
- [27] Yoon J, Yang E, Lee J, Hwang SJ. Lifelong learning with dynamically expandable networks. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018.
- [28] Riemer M, Cases I, Ajemian R, Liu M, Rish I, Tu YH, Tesauro G. Learning to learn without forgetting by maximizing transfer and minimizing interference. In: Proc. of the 7th Int'l Conf. on Learning Representations. New Orleans: OpenReview.net, 2019.
- [29] Aljundi R, Lin M, Goujaud B, Bengio Y. Gradient based sample selection for online continual learning. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 1058. [doi: [10.5555/3454287.3455345](https://doi.org/10.5555/3454287.3455345)]
- [30] He JP, Zhu FQ. Online continual learning for visual food classification. In: Proc. of the 2021 IEEE/CVF Int'l Conf. on Computer Vision Workshops. Montreal: IEEE, 2021. 2337–2346. [doi: [10.1109/ICCVW54120.2021.00265](https://doi.org/10.1109/ICCVW54120.2021.00265)]
- [31] Jin XS, Sadhu A, Du JY, Ren X. Gradient-based editing of memory examples for online task-free continual learning. In: Proc. of the 35th Conf. on Neural Information Processing Systems. NeurIPS, 2021. 29193–29205.
- [32] Gu YN, Yang X, Wei K, Deng C. Not just selection, but exploration: Online class-incremental continual learning via dual view consistency. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 7432–7441. [doi: [10.1109/CVPR52688.2022.00729](https://doi.org/10.1109/CVPR52688.2022.00729)]
- [33] Khosla P, Teterwak P, Wang C, Sarna A, Tian YL, Isola P, Maschiot A, Liu C, Krishnan D. Supervised contrastive learning. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 1567. [doi: [10.5555/3495724.3497291](https://doi.org/10.5555/3495724.3497291)]
- [34] Giorgi J, Nitski O, Wang B, Bader G. DeCLUTR: Deep contrastive learning for unsupervised textual representations. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing. Association for Computational Linguistics, 2020. 879–895. [doi: [10.18653/v1/2021.acl-long.72](https://doi.org/10.18653/v1/2021.acl-long.72)]
- [35] Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: Proc. of the 37th Int'l Conf. on Machine Learning. JMLR.org, 2020. 149.
- [36] He KM, Fan HQ, Wu YX, Xie SN, Girshick R. Momentum contrast for unsupervised visual representation learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 9726–9735. [doi: [10.1109/CVPR42600.2020.00975](https://doi.org/10.1109/CVPR42600.2020.00975)]
- [37] Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 831. [doi: [10.5555/3495724.3496555](https://doi.org/10.5555/3495724.3496555)]
- [38] Tian YL, Krishnan D, Isola P. Contrastive multiview coding. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 776–794. [doi: [10.1007/978-3-030-58621-8\\_45](https://doi.org/10.1007/978-3-030-58621-8_45)]
- [39] Mensink T, Verbeek J, Perronnin F, Csurka G. Distance-based image classification: Generalizing to new classes at near-zero cost. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2624–2637. [doi: [10.1109/TPAMI.2013.83](https://doi.org/10.1109/TPAMI.2013.83)]
- [40] Chaudhry A, Dokania PK, Ajanthan T, Torr PHS. Riemannian walk for incremental learning: Understanding forgetting and intransigence.

- In: Proc. of the 15th European Conf. on Computer Vision. Munich: Springer, 2018. 556–572. [doi: [10.1007/978-3-030-01252-6\\_33](https://doi.org/10.1007/978-3-030-01252-6_33)]
- [41] Buzzega P, Boschini M, Porrello A, Abati D, Calderara S. Dark experience for general continual learning: A strong, simple baseline. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 1335. [doi: [10.5555/3495724.3497059](https://doi.org/10.5555/3495724.3497059)]
- [42] Prabhu A, Torr PHS, Dokania PK. GDumb: A simple approach that questions our progress in continual learning. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 524–540. [doi: [10.1007/978-3-030-58536-5\\_31](https://doi.org/10.1007/978-3-030-58536-5_31)]



刘雨薇(1999—), 女, 硕士生, 主要研究领域为机器学习, 模式识别.



陈松灿(1962—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习, 模式识别, 神经计算.

www.jos.org.cn

www.jos.org.cn