增量学习

- 概念
  - 增量学习的能力就是能够不断地处理现实世界中连续的信息流，在吸收新知识的同时保留甚至整合、优化旧知识的能力。
    - 任务增量学习(Task-incremental Learning)
      - 如需要引入加密恶意流量的恶意样本，相当于在已知类中添加样本，使模型"认识"以前的恶意样本
    - 类增量学习(Class-incremental Learning)
      - 已经认识m个类，如何扩充至认识(m+n)个类，是难点
- 目的
  - 解决"灾难性遗忘"问题
    - 传统模型假设数据分布是固定或平稳的，训练样本是独立同分布的
  - 预期结果
    - 必须表现出从新数据中整合新知识和提炼已有知识的能力(可塑性)
    - 必须防止新输入对已有知识的显著干扰(稳定性)
- 与加密恶意流量检测的关系
  - 二分类
    - 只需要使用增量训练方法来改进相关模型，即可取得不错的效果 已解决√
      - sklearn中的增量训练（我的方法）
  - 多分类
    - 如何改进模型的泛化性能
      - 继续使用增量训练方法，并对比改进
        - 知道如何做，未测试
      - 可以参考的方法
        - sklearn中的增量训练（我的方法）
        - lightGBM的增量训练方法（张晓雨方法）
        - 基于已有的机器学习模型
        - keras的增量训练方法（深度学习）
        - 需重新编写深度学习模块
    - 如何让已有模型认识更多的类（难）
      - 类增量学习
        - 参考论文
          - Learning without Forgetting (ECCV 2016)
            - 基于深度学习的增量学习
          - Encoder Based Lifelong Learning (ICCV 2017)
            - 基于低维特征映射的EBLL算法
          - Overcoming catastrophic forgetting in neural networks (PNAS 2017)
            - 基于贝叶斯框架的EWC算法
          - iCaRL: Incremental Classifier and Representation Learning (CVPR 2017)
            - 和LwF比较相似，它同样引入了蒸馏损失来更新模型参数，但又放松了完全不能使用旧数据的限制
          - Large Scale Incremental Learning (CVPR 2019)
            - 借鉴了知识蒸馏技术，从不同的角度来缓解灾难性遗忘问题
        - 与原有模型之间的关系
          - 目前基于机器学习只能做增量训练，而若进行增量学习的思想，必然要引入深度学习模块
- 特点
  - 学习新知识的同时能够保留以前学习到的大部分知识，也就是模型在旧任务和新任务上均能表现良好
  - 计算能力与内存应该随着类别数的增加固定或者缓慢增长，最理想的情况是一旦完成某一任务的学习，该任务的观测样本便被全部丢弃
  - 模型可以从新任务和新数据中持续学习新知识，当新任务在不同时间出现，它都是可训练的
- 实现方式
  - 正则化(regularization)
    - 主要思想
      - 通过给新任务的损失函数施加约束的方法来保护旧知识不被新知识覆盖
    - 参考文献
      - Learning without Forgetting (ECCV 2016)
        - 是基于深度学习的增量学习
      - Encoder Based Lifelong Learning (ICCV 2017)
        - 基于低维特征映射的EBLL算法
      - Overcoming catastrophic forgetting in neural networks (PNAS 2017)
        - 基于贝叶斯框架的EWC算法
      - Rotate your Networks: Better Weight Consolidation and Less Catastrophic Forgetting (ICPR 2018)
      - Learning without Memorizing (CVPR 2019)
      - Learning a Unified Classifier Incrementally via Rebalancing (CVPR 2019)
      - Class-incremental Learning via Deep Model Consolidation (WACV 2020)
      - 其他关于正则化的手段
    - 优点
      - 基于正则化的增量学习方法通过引入额外损失的方式来修正梯度，保护模型学习到的旧知识，提供了一种缓解特定条件下的灾难性遗忘的方法
    - 缺点
      - 目前的深度学习模型都是过参数化的，但模型容量终究是有限的，我们通常还是需要在旧任务和新任务的性能表现上作出权衡
  - 回放(replay)
    - 主要思想
      - "温故而知新"，在训练新任务时，一部分具有代表性的旧数据会被保留并用于模型复习曾经学到的旧知识，因此「要保留旧任务的哪部分数据，以及如何利用旧数据与新数据一起训练模型」成为问题
    - 参考文献
      - iCaRL: Incremental Classifier and Representation Learning (CVPR 2017)
        - 和LwF比较相似，它同样引入了蒸馏损失来更新模型参数，但又放松了完全不能使用旧数据的限制
        - 好处
          - iCaRL在训练新数据时为每个旧任务保留了一部分有代表性的旧数据(iCaRL假设越靠近类别特征均值的样本越有代表性)，因此iCaRL能够更好地记忆模型在旧任务上学习到的数据特征。
      - Experience Replay for Continual Learning (NIPS 2019)
        - 指出这类模型可以动态调整旧数据的保留数量，从而避免了LwF算法随着任务数量的增大，计算成本线性增长的缺点
      - End-to-End Incremental Learning (ECCV 2018)
        - 基于End-to-End算法
      - Large Scale Incremental Learning (CVPR 2019)
        - 借鉴了知识蒸馏技术，从不同的角度来缓解灾难性遗忘问题
      - Gradient Episodic Memory for Continual Learning (NIPS 2017)
        - 提出了梯度片段记忆算法(GEM)
      - Efficient Lifelong Learning with A-GEM (ICLR 2019)
      - Gradient based sample selection for online continual learning (NIPS 2019)
      - 其他
    - 优点
      - 能够保留原有的模型进行预测，尽可能保留之前训练样本的信息，并随时加入训练
    - 缺点
      - 需要额外的计算资源和存储空间用于回忆旧知识，当任务种类不断增多时，要么训练成本会变高，要么代表样本的代表性会减弱
      - 在实际生产环境中，这种方法还可能存在「数据隐私泄露」的问题
  - 参数隔离(parameter isolation)
    - 需要引入较多的参数和计算量，因此通常只能用于较简单的任务增量学习，不展开

对应关系

改进1

改进2

改进

存在重大区别，并改进