

# 基于主动学习的不确定性感知补标签查询

刘圣源<sup>1</sup>, 陈珂<sup>2</sup>, 胡天磊<sup>1</sup>, 毛云青<sup>3</sup>

<sup>1</sup>浙江大学浙江省大数据智能计算重点实验室, 中国杭州市, 310027

<sup>2</sup>浙江大学区块链与数据安全全国重点实验室, 中国杭州市, 310027

<sup>3</sup>城云科技(中国)有限公司, 中国杭州市, 310000

**摘要:** 许多主动学习方法假设学习者可便捷地向注释者询问训练数据的完整标注信息。这些方法主要试图通过最小化标注数量降低标注成本。然而, 对于许多现实中的分类任务来说, 精确标注实例仍然非常昂贵。为降低单次标注行为成本, 本文试图解决一种新的主动学习范式, 称为具有补标签的主动学习 (ALCL)。ALCL 学习器只针对样例特定类别提出是或否的问题。在收到标注者答案后, ALCL 学习器获得一些有监督实例和更多具有补标签的训练实例, 这些补标签仅表示对应标签与该实例无关。。ALCL 具有两个挑战性问题: 如何选择要查询的实例以及如何从这些补标签和普通标签中提取信息。针对第一个问题, 在主动学习范式下提出一种基于不确定性的抽样策略。针对第二个问题, 改进了一种已有的 ALCL 方法, 同时适配了我们的抽样策略。在各种数据集上的实验结果验证了本文方法的有效性。

**关键词:** 主动学习; 图片分类; 弱监督学习

<https://doi.org/10.1631/FITEE.2200589>