

基于主动学习的可追踪性自动化生成方法

杜天保¹, 沈国华^{1,2,3}, 黄志球^{1,2,3}, 喻垚慎¹, 吴德香¹

¹南京航空航天大学计算机学院, 中国南京市, 211106

²软件新技术与产业化协同创新中心, 中国南京市, 210093

³南京航空航天大学安全关键软件重点实验室, 中国南京市, 211106

摘要: 可追踪性生成 (traceability link recovery, TLR) 是一项重要且昂贵的软件任务, 需要开发人员在同一项目中建立源制品集合与目标制品集合之间的关系。之前研究提出通过机器学习创建可追踪性方法。但是, 当前机器学习方法无法很好地应用于没有追踪信息的项目, 因为训练有效的预测模型需要人工标记太多追踪链。为节省人力, 提出一种基于主动学习 (active learning, AL) 的TLR方法, 简称基于AL的方法。在7个常用可追踪性数据集上评估该方法, 并将其与基于信息检索的方法和最新机器学习方法比较。结果表明, 基于AL的方法在F-score方面优于其他两种方法。

关键词: 自动; 可追踪性生成; 人力; 主动学习

<https://doi.org/10.1631/FITEE.1900222>