

一种基于特征选择与迁移学习的度量补偿软件缺陷预测方法

陈锦富^{1,2}, 王小丽^{1,2}, 蔡赛华^{1,2}, 徐家平¹, 陈静怡¹, 陈海波¹

¹江苏大学计算机科学与通信工程学院, 中国镇江市, 212013

²江苏大学工业网络空间安全技术江苏省重点实验室, 中国镇江市, 212013

摘要: 跨项目软件缺陷预测解决了传统缺陷预测中训练数据不足的问题, 克服了将多个不同源项目中学习的模型应用于目标项目的挑战。与此同时, 出现两个新问题: (1) 模型训练过程中过多无关和冗余特征影响训练效率, 降低了模型预测精度; (2) 由于开发环境等因素, 度量值的分布因项目而异, 当模型用于跨项目预测时, 预测精度较低。本文引入皮尔逊特征选择方法解决数据冗余问题, 采用基于迁移学习的度量补偿技术解决源项目和目标项目之间数据分布差异较大的问题。提出一种基于特征选择和迁移学习的度量补偿软件缺陷预测方法。实验结果表明, 用该方法构建的模型在AUC (接收器工作特性曲线下面积) 值和F1度量指标上取得较好结果。

关键词: 缺陷预测; 特征选择; 迁移学习; 度量补偿

<https://doi.org/10.1631/FITEE.2100468>