

有效警告识别技术文献综述

论文主要搜索关键词（不限于以下关键词）：每一行内的关键词是“或”关系，行与行之间关键词是“并”关系

序号	目的	关键字
1	后处理 警告	1) elimination, 2) reduction, 3) simplification, 4) ranking, 5) classification, 6) reviewing, 7) inspection
2	静态分 析	1) static analysis, 2) automated code analysis, 3) source code analysis, 4) automated defects detection
3	警告	1) alarm, 2) warning, 3) alert, 4) violations

基于置信学习的警告数据集去噪技术

现有的自动化警告标记技术主要基于关闭警告的启发式方法进行警告标记，给定同一个项目的多个版本，使用相同的静态分析工具对每个版本进行扫描并得到其扫描结果，通过比对不同版本之间的警告是否相同，如果一个版本中的某个警告 A 存在在后续所有的版本，那么该警告 A 被认为是无效警告（即误报警告），如果一个版本中的某个警告 A 在后续的版本中消失，那么该警告 A 被认为是有效警告（即正报警告）。然后，由于验证延迟（即有些警告会在未来的版本中被修复）或自动化警告标记方法固有的局限性（如：不精确的警告匹配算法），会使得自动化警告标记方法产生较多的噪音（即正报标记为误报，误报被标记为正报）。为了解决上述问题，本项目拟基于置信学习技术对自动化警告标记技术进行去噪，进而改善自动化警告标记技术标记的精度。

（i）初始警告数据集收集

- 项目来源于 apache 下面的开源的 Java 项目
- 基于关闭警告的启发式警告标记方法：
<https://github.com/lxyeah/findbugs-violations.git>，该链接中给出了该方法自动收集警告数据集的源码以及相关软件配置要求。在给定的源码中静态分析工具是 SpotBugs，在实验中可以选用其他针对 java 源码的静态扫描工具（如：PMD、CheckStyle、Jlint）

（ii）置信学习技术

- 置信学习原理可参考论文“Confident Learning: Estimating Uncertainty in Dataset Labels”
- 置信学习技术实现：目前已经有开源的 python 包，置信学习例子可见：
https://blog.5lcto.com/u_15346174/3833521

（iii）项目交付结果

- （i）初始警告数据集收集中的初始标记
- （ii）置信学习技术的处理结果
- 由于本项目主要是验证置信学习是否能够对现有的自动化警告标记去噪，因

此，需要去标记人工去标记警告的真实标记进行准确率评估，因此，本项目的交付结果还需包括**人工标记结果**和**警告标记精度统计结果**。本项目要求对（i）中警告标记结果中的所有有效警告进行人工标记以及对（i）中的无效警告进行抽样人工标记。人工标记的要求如下，其中，“unknown”表示无法判断该警告是一个有效警告还是无效警告，“close”表示该警告是一个有效警告，“open”表示该警告是一个无效警告。

- If method/file containing a warning was removed, label the warning as "unknown".
- If method/file containing a warning was removed, label the warning as "unknown".
- If the warning was removed, but a large amount of code was changed (e.g. a refactoring of the entire class, a change in functionality), label the warning as "unknown".
- If the method/field containing the warning/bug was renamed, but the warning/bug still appears to be present, label the warning as "unknown".
- If the statements containing the warning/bug was moved to a different method, but the warning/bug still appears to be present in these statements, label the warning as "unknown".
- If the warning was removed, but the code is still similar to the original code, and the warning/bug still appears to be present, label the warning as "open".
- If the warning was removed, but the original code looked to be a false positive, and you believe that the code was changed just to silence the false alarm from findbugs/spotbugs, label the warning "open".
- If the code appears to be automatically generated, label the warning "unknown".

(iv) 其他要求

- 项目实施可以使用多种语言，如：警告数据集收集是 Java，警告后续处理是 Python。
- 项目验收既可以是实验统计数据结果（如：各类图表），也可以用前端展示。
- 所有的源码、数据集、统计结果生成数据包。

联系方式

葛修婷, 1683245057@qq.com