

新闻动态

热点新闻 (../rdxw2016/)	>
科研进展 (../)	>
科技动态 (../kjdt2016/)	>
通知公告 (../tzzg2016/)	>
内部公告 (http://work.iscas.ac.cn/index.php/Home/Service/NoticeList/t/1/o/0/p/1.html)	>

首页 (../) > 新闻动态 (../) > 科研进展 (../)

智能软件研究中心硕士生段旭荣获ISSRE 2021最佳实践论文奖

文章来源: | 发布时间: 2021-11-03 | 【打印】 【关闭】

近日, 中国科学院软件研究所智能软件研究中心武延军、吴敬征课题组在基于深度神经网络的静态代码分析研究中取得进展。该团队提出了基于多类型和多粒度的语义代码表示学习模型——MultiCode, 解决了工业场景中涉及多需求的开发任务时面临的开发开销大、模型集成困难、可扩展性受限等问题, 实现了在多需求工业场景下的高效开发和准确预测, 在漏洞检测、代码克隆检测等任务中得到了具体实践, 并在政府部门和互联网企业中实际应用。

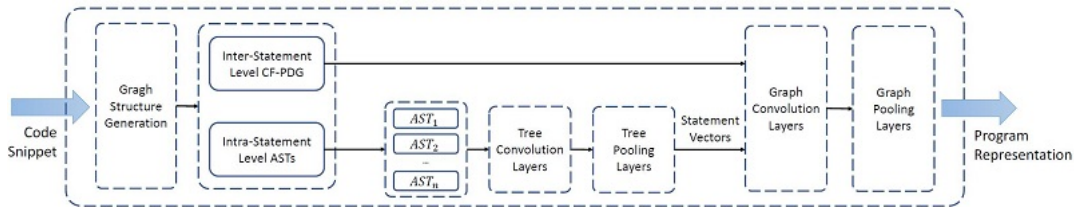
研究成果以MultiCode: A Unified Code Analysis Framework based on Multi-type and Multi-granularity Semantic Learning为题发表在软件可靠性工程领域旗舰国际会议ISSRE 2021的Industry Track上。在此次会议投稿的226篇论文中, 71篇论文被接受, 其中仅有本研究1篇论文被评为最佳实践论文。论文第一作者为课题组硕士研究生段旭。

基于深度神经网络的静态代码分析方法通常在不同代码分析任务中引入针对性设计, 导致模型呈现高度多样化的态势。在工业领域, 该现象会使开发者在开发涉及多需求的代码分析平台时, 面临开发开销大、模型集成困难、可扩展性受限等问题。

针对上述问题, MultiCode模型能够学习代码中多种类型和粒度的语义信息, 进而支撑多种代码分析任务。团队提出使用抽象语法树、控制流图、程序依赖图等结构, 对代码中不同类型和粒度的语义信息进行建模, 并利用树神经网络和图神经网络分别对不同的语义信息进行处理。在该过程中, MultiCode模型自底向上地先学习语句级别的表示, 再基于该表示学习代码段级别的表示。通过将模型作为编码器进行神经网络构建, 能够有效适配于不同的代码分析任务。在漏洞检测和代码克隆检测任务上的评估结果表明, 其能够在不需要重新构建编码器的情况下, 在不同任务中有效地识别并区分不同类别代码的语义, 进而支撑多种任务上的预测。

该研究得到国家重点研发计划、国家自然科学基金的支持。

论文下载 (<https://github.com/DuanXu-97/MultiCode-Paper>)



MultiCode基于多类型和多粒度的语义代码表示学习模型框架



ISSRE 2021最佳实践论文奖



Copyright © Institute of Software, CAS. All rights reserved. info(at)iscas.ac.cn
 版权所有 © 中国科学院软件研究所 京ICP备05046678号-1 (<https://beian.miit.gov.cn>) 文保网
 安备1101080077
 电话: 86-10-62661012 传真: 86-10-62562533 电子邮箱: info@iscas.ac.cn



method=show

