软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(8):2177−2179 [doi: 10.13328/j.cnki.jos.005534]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

数据驱动的软件智能化开发方法与技术专题前言*

谢 冰[1], 魏 峻[2], 彭 鑫[3,4], 孙海龙[5,6]

[1](北京大学 信息科学技术学院,北京 100871)

[2](计算机科学国家重点实验室(中国科学院软件研究所),北京 100190)

[3](复旦大学 计算机科学技术学院,上海 200433)

[4](上海市数据科学重点实验室(复旦大学),上海 200433)

[5](软件开发环境国家重点实验室(北京航空航天大学),北京 100191)

[6](大数据科学与脑机智能高精尖创新中心(北京航空航天大学),北京 100191)

通讯作者: 谢冰, E-mail: xiebing@pku.edu.cn

中文引用格式: 谢冰,魏峻,彭鑫,孙海龙.数据驱动的软件智能化开发方法与技术专题前言.软件学报,2018,29(8):2177−2179. http://www.jos.org.cn/1000-9825/5534.htm

当今社会软件无所不在,追求高质量和高效率的软件开发是软件工程研究的核心目标.软件开发经历了从结构化方法、面向对象方法到网络服务化,逐步向基于互联网和开源模式构造的方法发展.软件开发工具与环境也是伴随着开发方法不断更替变化,从命令行开发工具到集成化开发环境,再到扩展开发环境、协同开发环境,正向智能化开发环境方向演进.

软件开发智能化一直是软件工程追求的目标之一.以开源软件为代表的互联网软件开发呈现出边界开放、群体分散、交付频繁、知识复杂等特征,同时贯穿全生命周期的软件开发活动中积累了快速增长、规模巨大的软件大数据.这为软件智能化开发建立了数据基础,但需要解决基础性的数据采集分析、知识抽取利用问题,并以智能搜索、推荐、问答等方式提升软件开发工具智能化程度,提高软件开发的效率和质量.智能化的软件工具可以基于数据和知识向开发人员提供推荐和智能检索,由此形成“人-工具-数据”融合的新一代软件智能化开发技术体系和环境.

本专题选题为“数据驱动的软件智能化开发方法与技术”,突出软件工程数据收集和汇聚、面向软件工程大数据的知识提炼,以及面向软件构造、测试验证、运行维护等不同阶段智能释放的新方法、技术与工具.

专题通过公开征文征得投稿 27 篇.这 27 篇论文通过特约编辑的形式审查,有 26 篇论文进入到评审阶段.上述稿件研究内容涉及数据驱动的软件智能化开发的多个方面.特约编辑先后邀请了 25 位软件工程、程序设计语言和大数据等相关领域的专家参与审稿工作,每篇投稿邀请 2 位专家进行评审.其中有 18 篇投稿通过专家评审受邀参加了 NASAC 2017 会议宣读,并在修改后参加了终审.整个稿件评审历经 5 个月,经初审、复审、NASAC 2017 会议宣读和终审共 4 个阶段,最终有 13 篇论文入选本专题,覆盖了如下 4 个方面的内容.

一、智能化程序搜索与构造

智能化程序搜索与构造关注于通过可复用代码搜索、理解、推荐以及缺陷检测等技术支持程序的智能化构造.本专题共收录了 5 篇论文.

《智能化的程序搜索与构造方法综述》从代码搜索、程序合成、代码推荐与补全、缺陷检测、代码风格改善、程序自动修复等方面对当前的国内外研究工作进行了综述,在对相关的理论和技术途径进行梳理的基础上,对该领域研究过程中面临的挑战进行了总结并给出了建议的研究方向.

《安卓应用用户界面交互模式抽取与检索》提出了一种安卓应用界面交互模式抽取与检索方法,通过安卓应用分析获得各个 Activity 的界面交互描述,允许用户基于预定义的模板表达交互模式检索需求,从而支持开

* 收稿时间: 2018-03-20

发者在选择、试用、学习安卓应用过程中高效地获得感兴趣的安卓界面交互模式.

《基于 StackOverflow 数据的软件功能特征挖掘组织方法》提出了一种基于 StackOverflow 问答数据的软件功能特征挖掘组织方法,获取以动宾短语形式描述的软件功能特征并自动生成一种以层次化方式展示的软件项目功能特征文档,从而形成对软件官方文档中功能描述的有益补充.

《融合结构与语义特征的代码注释决策支持方法》从大量的代码注释实例中学习通用的注释决策规范,提出了一种基于代码结构和代码语义信息的代码注释决策支持方法,根据代码与其上下文之间的逻辑耦合关系强弱程度确定合适的注释点,从而提高代码的可理解性和可复用性.

《一种基于关联分析与 *N*-Gram 的错误参数检测方法》针对代码中的函数调用参数错误,提出了一种基于关联分析和 *N*-Gram 语言模型的静态检测方法.该方法基于大量开源代码构建关联分析模型,并对参数间存在强关联规则的函数调用构建 *N*-Gram 语言模型以计算其正确性概率,以此作为错误检测依据.

二、面向开源生态的软件数据挖掘

面向开源生态的软件数据挖掘关注于通过对开源社区的数据挖掘理解开源生态并促进基于开源社区的大众化软件生产活动的发展.本专题共收录了 3 篇论文.

《面向开源生态的软件数据挖掘技术研究综述》首先界定了大众化软件生产活动的分布范围、基本过程和数据形态,然后从软件复用、协同开发、知识管理这 3 个核心环节对开源社区数据挖掘技术的研究工作进行了归类与分析,并总结了该领域研究工作存在的问题和未来发展趋势.

《基于贡献分配的开源软件核心开发者评估》通过设计开发文件的贡献度分配算法,以 9 个 Apache 项目为基础分析了开发者对项目的贡献度,以此区分核心开发者和外围开发者,然后在此基础上通过支持向量机建立分类模型,结合不同影响开发者地位的关键因素,提升了开发者分类的精确度.

《代码文件贡献组成模式的分析》基于代码所有权,从贡献组成的集中度、复杂度和稳定性这 3 个维度出发,提出了刻画代码文件贡献组成的 3 个量度,在此基础上基于 OpenStack 的核心项目 Nova 的版本控制数据建立贡献组成的量度,总结了 12 种通用文件类型以及 3 种贡献组成模式,并通过邮件交流和面对面访谈的方式验证了量度的有效性以及贡献组成模式的合理性.

三、面向开发任务分配的自动推荐技术

面向开发任务分配的自动推荐技术关注于基于软件开发历史数据分析,为新的开发或缺陷修复任务推荐开发者.本专题共收录了 3 篇论文.

《面向软件安全性缺陷的开发者推荐方法》针对安全性缺陷的修复提出了一种软件开发者推荐方法.该方法将开发者的历史开发内容(与安全性相关)以及开发者过往修复过的缺陷的复杂度和修复质量综合起来考虑,并且针对不同复杂度的安全性缺陷推荐不同经验级别的开发者进行修复.

《一种多特征融合的软件开发者推荐》设计了一种基于模糊综合评价的开发者能力模型,通过挖掘开发者与任务的动态交互行为、静态匹配度以及开发者能力这 3 个维度的特征并结合矩阵分解技术,提出了一种能力与行为感知的多特征融合协同过滤开发者推荐方法,同时结合开源社区以及企业内部开发数据总结出了适合大数据环境的多特征融合开发者推荐系统的实践.

《基于循环神经网络的缺陷报告分派方法》针对缺陷报告分派问题提出了一个基于循环神经网络的深度学习模型,利用双向循环网络与池化方法提取缺陷报告的文本特征,并使用单向循环网络提取特定时刻的开发者活跃度特征,在此基础上运用已修复的缺陷报告进行监督学习.

四、面向自然语言的智能化软件分析

面向自然语言的智能化软件分析关注于通过分析基于自然语言的需求描述和用户评价等文本信息获取软件设计方案或对软件进行质量评价.本专题共收录了 2 篇论文.

《自然语言数据驱动的智能化软件安全评估方法》在自然语言处理技术的基础上提出了一种基于待评估软件现有用户使用经验的软件安全性评估方法.该方法以自适应的方式爬取用户对相关软件的自然语言评价数据,然后利用深度学习方法与机器学习评估模型的双重训练来获得软件的安全性评估指标.

《基于限定自然语言需求模板的 AADL 模型生成方法》针对基于自然语言需求的安全性(safety)分析结果难以完整在软件设计中反映的问题,提出了一种面向构件化嵌入式软件的半结构化限定性自然语言需求模板,采用需求抽象语法图作为中间模型实现软件需求与 AADL 模型之间的转换并记录二者的追踪关系.

本专题主要面向软件工程、大数据分析与应用及相关领域的研究人员和专业工程师等,反映了我国学者在数据驱动的软件智能化开发方法、技术与实践应用中的最新研究进展.在此,我们要特别感谢《软件学报》编委会和软件工程专委会对专题工作的指导和帮助.感谢编辑部和软件工程专委会各位老师从征稿启示发布、审稿专家邀请至评审意见汇总、论文定稿、修改及出版所付出的辛勤工作和汗水,感谢专题评审专家及时、耐心、细致的评审工作.此外,我们还要感谢所有向本专题踊跃投稿的作者以及参加本专题 NASAC 2017 学术报告会的各位与会者.

最后,感谢专题的评审专家、编辑和读者们,希望本专题能够对数据驱动的软件智能化开发方法与技术的研究工作有所促进.

谢冰(1970－),男,湖南湘潭人,博士,教授,博士生导师,CCF 高级会员.现任中国计算机学会软件工程专业委员会和理论计算机科学专业委员会委员,《电子学报》英文版编委.主要研究领域为软件工程,分布式计算,计算机理论科学.

魏峻(1970－),男,博士,研究员,博士生导师,CCF 高级会员.现任中国计算机学会软件工程专业委员会委员和服务计算专委会副主任委员,《软件学报》编委.主要研究领域为智能化软件工程,分布式计算,服务计算.

彭鑫(1979－),男,博士,教授,博士生导师,CCF 高级会员.现任中国计算机学会软件工程专业委员会委员,《软件学报》编委.主要研究领域为智能化软件开发,移动云计算,软件维护与演化.

孙海龙(1979－),男,博士,副教授,博士生导师,CCF 高级会员.主要研究领域为智能化软件方法,群智计算,分布式系统.