

# 智能化软件新技术专刊前言\*

申富饶<sup>1,2</sup>, 李戈<sup>3,4</sup>



<sup>1</sup>(计算机软件新技术国家重点实验室(南京大学),江苏 南京 210023)

<sup>2</sup>(南京大学 计算机科学与技术系,江苏 南京 210023)

<sup>3</sup>(北京大学 信息科学技术学院,北京 100871)

<sup>4</sup>(高可信软件技术教育部重点实验室(北京大学),北京 100871)

通讯作者: 申富饶, E-mail: frshen@nju.edu.cn

中文引用格式: 申富饶,李戈.智能化软件新技术专刊前言.软件学报,2019,30(5):1203-1205. <http://www.jos.org.cn/1000-9825/5736.htm>

智能化软件新技术是使用人工智能与机器学习技术辅助软件工程的新兴技术.近年来以深度学习为代表的人工智能技术飞速发展,在计算机科学的各个领域中都开始发挥重要作用.随着软件工程领域对数据积累的重视,人工智能技术在软件开发过程中的应用也逐渐增加.利用人工智能技术辅助软件工程中的代码编写、纠错、测试等具体工作,可以大量节省人工劳动,提升软件开发效率.将人工智能技术更好地与工程各个环节相结合,进一步提升软件开发过程的自动化、智能化程度,是智能化软件新技术研究的主要关注点.

本专刊公开征文,共收到投稿 37 篇,其中 32 篇论文通过了形式审查,内容涉及智能化软件新技术的各个方面.特约编辑先后邀请了 20 多位专家参与审稿工作,每篇投稿至少邀请 2 位专家进行评审.稿件经初审、复审、全国软件与应用学术会议(NASAC 2018)宣读和终审 4 个阶段,共历时 5 个月.26 篇论文通过复审在 NASAC 2018 会议上进行报告并接受质询,最终有 24 篇论文通过终审,入选本专刊.本专刊主要包括以下 4 个主题.

## 一、自动化编程与纠错

本主题收录了 5 篇论文.《基于深度学习的程序生成与补全技术研究进展》针对基于深度学习的自动化编程的两种主要实现方式,即程序生成与代码补全,进行了广泛的调研与综述,揭示了该领域目前所面临的困难和挑战,并提出了几个未来的研究方向.本文对该领域内研究现状的分析总结有助于其他研究者在现有研究的基础上改进创新.《可微分抽象机混合编程系统》提出了一种可无缝结合高级编程语言与神经网络组件的混合编程模型.该模型可以同时使用过程化的高级编程语言元素和神经网络组件元素混合开发应用程序,发挥并整合两种编程方式各自的优势.《C/C++程序缺陷自动修复与确认方法》针对 C/C++程序的自动修复问题提出了一种基于机器学习和程序合成的解决方案,结合了深度学习技术进行修复选项的生成,并据此实现了原型工具 AutoGrader.《示例演化驱动的学生程序自动修复》针对学生程序设计应用背景,研究程序自动修复方法,利用模板示例程序指导补丁的演化,基于自动的错误分离及程序变异设计了细致的修复方案.《基于频繁模式挖掘的 GCC 编译时能耗演化优化算法》设计了一种基于频繁模式挖掘的遗传算法 GA-FP.该算法在演化过程中利用频繁模式挖掘得到出现频度高且能耗改进大的一组编译选项,并以此作为启发式信息设计了“增添”和“删减”两种变异算子,以帮助提高解质量和加快收敛速度.

## 二、缺陷预测与报告

本主题收录了 9 篇论文.《即时软件缺陷预测研究进展》从即时缺陷预测技术的数据标注、特征提取和模型评估等方面对即时缺陷预测研究进展进行综述,并展望了即时缺陷预测的未来发展,对读者掌握该领域的研究进展非常有帮助.《基于特征迁移和实例迁移的跨项目缺陷预测方法》在跨项目软件缺陷预测研究领域提出了一种两阶段跨项目缺陷预测方法 FeCTrA,同时考虑特征迁移和实例迁移两个角度,减少两个项目的数据分

\* 收稿时间: 2019-02-13

布差异性,提高缺陷预测精度.《C 程序内存泄漏智能化检测方法》提出了一种内存泄漏的智能化检测方法,通过使用机器学习算法学习程序特征与内存泄漏之间的相关性,构建机器学习分类器并应用机器学习分类器进一步提高内存泄漏静态分析的准确性.《基于深度学习的 API 误用缺陷检测》将深度学习中的神经网络模型应用于 API 使用规约的学习及 API 误用缺陷的检测,能够在一定程度上实现 API 误用缺陷的自动发现,为开发者节省开发和维护的时间开销.《一种基于深度学习的上帝类检测方法》提出了一种基于特征类属性和相关自然语言信息的上帝类检测方法,并使用神经网络作为主要检测的技术方法,其效果优于现有的上帝类检测方法,尤其是在查全率上有大幅度的提升.《基于维修日志的飞机设备故障原因判别方法》提出了一种基于卷积神经网络对非结构化文本提取文本特征建立字向量的方法,使用随机森林算法对长期积累下来的飞机故障日志数据建立故障原因分类器,可以帮助维修人员尽快确定故障原因.《基于代价极速学习机的软件缺陷报告分类方法》针对软件开发中的 Bug 分类问题提出了一种基于代价 ELM 的方法,将 Bug 分为严重 Bug 和不严重的 Bug 两类,将错误分类代价的概念引入 ELM 模型的损失函数,重点解决了报告分类中的数据不平衡和数据量不足的问题.《基于嵌入模型的混合式相关缺陷关联方法》针对自动化关联相关缺陷报告的问题,提出了一种基于嵌入模型的方法,将相关缺陷自动关联问题抽象为推荐问题,并且结合了传统的信息检索技术和深度学习中的嵌入模型,综合了词频相似度、单词相似度、文档相似度,有效地提高了传统方法的性能.《代码坏味对软件演化影响的实证研究》力求探究代码坏味产生的影响以及坏味与软件演化之间的关系,开展了关于代码坏味对软件演化的影响的实证研究,具体分析了代码坏味与新增文件、修改文件和移除文件这 3 类具体源文件操作之间的相关性.

### 三、软件测试

本主题收录了 3 篇论文.《面向持续集成测试优化的强化学习奖励机制》针对持续集成测试优化问题,提出了一种包含测试用例历史执行信息的奖励函数的强化学习方法,使用包含测试用例历史失效信息的奖励函数,使检错能力得到明显提高.《敏感变量和感知机结合的测试预言生成方法》提出了一种在已知部分测试用例集的情况下,基于敏感变量和线性感知机的新测试用例的测试预言自动生成方法,为测试预言缺乏的场景提供了解决思路.《基于 SOM 神经网络的二阶变异体约简方法》面向在软件测试中具有重要意义的二阶变异测试领域,详细分析了影响条件下二阶等价变异体的可能性因素,并基于这些因素提出了一种基于 SOM 神经网络的二阶变异体约简的方法.

### 四、代码、数据与用户管理

本主题收录了 7 篇论文.《基于图嵌入的软件项目源代码检索方法》将软件项目代码中的元素以图的方式进行组织,形成包含继承、实现、成员等关系的软件代码图,然后采用图嵌入的方式计算结点之间的相关性,以支持查询中子图的生成与推荐.《企业级海量代码的检索与管理技术》设计实现了一个对大规模代码库进行管理和搜索的工具平台,通过爬取大量代码、基于 Elasticsearch 工具设计了代码搜索引擎,并进一步研究了查询变换、搜索策略和结果排序、摘要生成等技术以提高代码搜索效果,在大型软件企业中具有较为重要的应用价值.《一种基于迭代的关系模型到本体模型的模式匹配方法》提出了一种考虑数据本地化特征的模式匹配算法,对单一模式匹配算法匹配不准的成因进行了较为深入的分析,并提出了一种迭代优化的模式匹配方案,能够更好地兼容数据源的本地化特征.《软件开发活动数据的数据质量问题》研究软件工程中的数据质量问题,通过分析现有文献,总结出了 10 种数据质量问题,涵盖了数据产生、数据收集和数据显示这 3 个阶段,也提出了相应的方法以帮助发现和解决数据问题,对其他研究人员有很大的参考价值.《基于本体推理的终端用户数据查询构造方法》针对企业数据在进行业务分析时,业务分析者无法直接使用数据库查询和使用数据的问题,提出了一种基于推理的终端用户本体查询构造方法,能够减少数据库设计细节的暴露并使得终端用户更易查询和操纵数据.《一种基于支持向量机和主题模型的评论分析方法》提出了一种结合分类和主题提取的用户评论分析方法 RASL,用于从移动应用的用户所提供的大量无结构化负面反馈评论中提取有效信息,为移动应用开发者提供更为易读且准确的结构化用户评论,并且也为用户评论的分组与记录带来好处.《基于多开发者社区的用户推荐算法》针对开发者社区中的用户标签自定义、不同社区的不对称活跃现象和关键词集封闭等问题,提出

了一种跨社区的用户推荐算法,旨在赋予用户更加准确的标签,进而拓展可推荐用户范围并提升用户推荐结果的准确度.

本专刊主要面向智能化软件新技术领域,突出反映我国计算机软件与人工智能学者在该领域的最新研究进展.在此,我们特别感谢《软件学报》编委会、中国计算机学会系统软件专委会和软件工程专委会对本专刊工作的指导、支持和帮助,感谢编辑部各位老师所付出的辛勤努力,感谢评审专家严谨、细致、及时的评审工作,感谢向本专刊积极投稿的所有作者,希望本专刊有助于促进人工智能、软件工程相关领域人员在智能化软件新技术方面的研究与实践工作.



申富饶(1973—),男,江苏泰兴人,博士,教授,博士生导师,CCF 系统软件专委会委员,历任日本科学振兴机构研究员、日本学术振兴会外国人特别研究员等.主要研究领域为神经网络,数据分析,机器人智能.在国内外包括IJCAI、AAAI、CVPR 等顶级国际会议和《IEEE TNNLS》、《Neural Networks》等领域内顶级 SCI 期刊上发表学术论文 80 余篇.



李戈(1977—),男,博士,副教授,CCF 专业会员.主要研究领域为深度学习,程序分析,知识工程.所在研究团队聚焦于基于机器学习概率模型的程序语言处理,是国际上最早从事相关研究的团队之一,在代码功能分析、代码自动补全、代码缺陷检测等方面取得了国际上领先的研究成果,累计发表论文 70 余篇.科研转化成果 aiXcoder 在业界得到了广泛应用,其技术指标长期保持国际领先水平.