



# 计算机科学

COMPUTER SCIENCE

## 让未来世界透明可解释的智能软件技术与方法

刘璘, 邢颖, 孙昌爱, 李春芳, 石川

### 引用本文

刘璘, 邢颖, 孙昌爱, 李春芳, 石川. 让未来世界透明可解释的智能软件技术与方法[J]. 计算机科学, 2023, 50(5): 1-2.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于可解释性人工智能的软件工程技术方法综述](#)

Review of Software Engineering Techniques and Methods Based on Explainable Artificial Intelligence  
计算机科学, 2023, 50(5): 3-11. <https://doi.org/10.11896/jsjcx.221100159>

#### [非均衡数据分类经典方法综述与面向医疗领域的实验分析](#)

Imbalanced Data Classification: A Survey and Experiments in Medical Domain  
计算机科学, 2022, 49(1): 80-88. <https://doi.org/10.11896/jsjcx.210200124>

#### [一种基于变异分析的BPEL程序故障定位技术](#)

Mutation Based Fault Localization Technique for BPEL Programs  
计算机科学, 2021, 48(1): 301-307. <https://doi.org/10.11896/jsjcx.200900051>

#### [用户如何看待产品中的预测分析功能? ——面向非功能性需求的调研报告](#)

What Users Think about Predictive Analytics?—A Domestic Survey on NFRs  
计算机科学, 2020, 47(12): 18-24. <https://doi.org/10.11896/jsjcx.201200055>

#### [基于通道的螺旋型布图算法研究与实现](#)

计算机科学, 2002, 29(8): 126-127.

# 让未来世界透明可解释的智能软件技术与方法

刘 璘<sup>1</sup> 邢 颖<sup>2</sup> 孙昌爱<sup>3</sup> 李春芳<sup>4</sup> 石 川<sup>2</sup>

1 清华大学 北京 100084

2 北京邮电大学 北京 100876

3 北京科技大学 北京 100083

4 中国传媒大学 北京 100024

(linliu@tsinghua.edu.cn)

人工智能作为计算机科学最活跃的研究领域,在为人持续创造新的惊喜的同时,也带来了关于可信性与可解释性的问题与挑战。2022年11月27日,中国软件大会“面向可解释人工智能的软件工程方法与技术”论坛在线上召开。论坛联合主席石川教授应邀作了题为“异质图神经网络及其在软件工程领域的应用思考”的特邀报告,分享了他在异质图神经网络模型领域的最新研究进展,着重介绍了将现实复杂交互系统建模为图结构,并通过异质图神经网络进行预测分析的模型、系统和应用的新研究成果。论坛联合主席邢颖副教授应邀为本专题撰写《基于可解释性人工智能的软件工程技术方法综述》,分享了她对可解释性人工智能基本概念以及模型可解释性评估的理解,从软件工程研究者的角度分析了恶意软件检测、高风险组件检测、软件负载分配、二进制代码相似性分析等通过可解释 AI 提升智能软件系统可信度的方法,展望了软件工程与可解释人工智能相结合的重要研究方向。

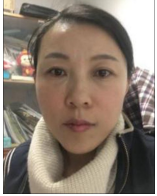
本专题收录了论坛中的4篇口头报告论文。中国联通研究院杨斌博士等的《基于注意力机制的可解释点击率预估模型研究》通过多头注意力机制对特征嵌入、线性特征交互和深度部分进行增强和解释。北京邮电大学李汇来等的《软件缺陷预测模型可解释性对比》对经典软件度量方法和部分深度学习模型的输入、输出及结构进行了对比分析,从数据差异程度和模型对代码的处理过程两个角度进行了差异性分析和解释。北京交通大学陈博、李红辉等合作完成的《基于BASFPA-BP的可靠性预测模型研究》针对基于传统BP神经网络的模型预测精度问题,提出了基于BASFPA-BP的软件可靠性预测模型,利用BASFPA算法优化网络权值、阈值,从而提高模型的预测精度。安徽理工大学李炳辉等合作完成的《基于BERT和弱行为轮廓的可解释性事件日志修复方法》针对现有方法在进行日志异常检测与修复工作中存在的可解释性较差的问题,提出将BERT4Log模型和弱行为轮廓理论与多层多头注意力机制相结合,进行低质量事件日志的可解释修复,该方式取得了更优的整体修复性能和弱行为轮廓学习能力,并给出了修复结果的详细解释。

与此同时,编辑部收到了多篇关于可解释性AI和面向领域人工智能应用的投稿。中国石油大学陈冲副教授等的《深度学习可解释性综述》从内在可解释模型、基于归因的解释和基于非归因的解释3个角度分析了深度学习解释方法,讨论了深度学习可解释性的应用以及未来发展方向。信息工程大学孙雪凯和蒋烈辉的《一种基于神经网络的代码嵌入方法》基于Code2vec算法,提出将一个代码函数分解为一系列抽象语法树路径,通过神经网络来学习如何表示每一条路径,最后将所有路径聚合成一个嵌入向量来表示当前的代码函数。解放军信息工程大学张文宁等的《基于灰狼算术混合优化算法的类集成测试序列生成方法》提出了灰狼优化算法和算术优化算法的一种混合优化算法,该算法保留了GWO的位置更新策略,选用群体领导层的中心个体替换AOA的引导个体,以平衡算法的全局探索和局部开发能力,实现用较短的时间生成测试桩代价较低的一类集成测试序列。清华大学董家祥等的《知识驱动的机械设备故障诊断》则设计了一个通用的机械设备故障诊断流程,包括故障问诊、故障定位、起因定位和故障维修指导等步骤,并在企业实践中验证了所提方法的可用性和面临的问与和挑战。华南师范大学黄迅迪和庞雄文的《基于深度学习的智能设备故障诊断研究综述》在调研设备故障诊断相关文献的基础上,分析了适用于设备故障诊断应用场景的深度模型框架,根据具体应用场景和设备监测数据类型对模型算法进行了分类介绍、列表对比及总结,最后针对存在的问题分析了未来的发展方向。

本专题将上述论文呈现给感兴趣的计算机科学研究者,希望为其带来一些启发。



**刘 璘** 清华大学博士,主要从事软件需求工程与知识工程领域的研究工作。主要研究方向为软件系统建模与分析、医疗大数据分析、新型网络服务软件开发方法等。现任《需求工程》国际期刊(REJ)编委,《信息系统建模与设计国际期刊》(IJISMD)副主编。曾编辑过服务计算通讯(TSC)期刊的“服务需求工程”特约专栏,《需求工程》国际期刊(REJ)的“云计算环境下的安全需求工程”专刊,《软件学报》的“大数据时代的软件工程”专刊,《计算机研究与发展》的“人工智能背景下的需求工程”专刊,《计算机科学》的“复杂系统的软件工程与需求工程”专刊等。曾参与与专刊主体相关领域的国家自然科学基金重点项目两项和主持863项目一项,致力于“面向大规模群体协同的软件建模方法与技术”以及“基于最优决策的需求工程方法”的研究与实践。



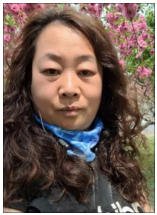
**邢颖** 北京邮电大学人工智能学院副教授、博士生导师、CCF 高级会员、软件工程专委会执行委员、九三学社社员、国家自然科学基金同行评议专家。主要研究方向为程序分析与理解、软件缺陷检测与预测、人工智能及其应用。主持国家自然科学基金、中国博士后科学基金及其他横、纵向项目多项,并承担了多个国家重点研发计划项目等国家及省部级项目。申请专利 10 余项,出版专著《源代码分析》《面向对象软件工程》等。在《软件学报》、《中国科学:信息科学》、EAAI、JCST、JSS、IET Software、KBS 及 ChinaSoft 等期刊和会议上发表论文数十篇。担任《IEEE Transactions on Reliability》《Chinese Journal of Electronics》

《软件学报》《电子学报》等期刊的审稿人。



**孙昌爱** 北京科技大学教授,长期从事软件工程、服务计算领域的研究与教学,主要研究方向为服务计算、软件测试、程序分析等。在《IEEE Transactions on Software Engineering》、《IEEE Transactions on Computers》、《IEEE Transactions on Services Computing》、《IEEE Transactions on Reliability》、《ACM Transactions on the Web》、《Journal of Systems and Software》、《中国科学》、《计算机学报》、《软件学报》等国内外重要学术刊物上发表论文 100 余篇,申请中国国家发明专利 10 项,登记计算机软件著作权 25 项,出版译著 3 部。主持完成国家自然科学基金、北京市自然科学基金、中国航空科学基金等 20 余项研究课题。

担任 TrustCom2008、UIC2015、CBPM2020 等国际/全国学术会议程序委员会主席,担任国际会议 IEEE SOCA2015、全国服务计算学术会议 NCSC2018 研讨会主席,担任 IEEE ICWS2020、COMPSAC2020 等软件工程、服务计算领域国际会议程序委员会委员 120 余次。IEEE 高级会员、中国计算机学会软件工程专委会委员、中国计算机学会服务计算专委会委员。曾在美国普渡大学、荷兰格罗宁根大学、澳大利亚斯文本大学、香港理工大学从事学术研究与交流。



**李春芳** 中国传媒大学计算机与网络空间安全学院副教授,北京大学艺术学院访问学者。主要研究方向为数据可视化、软件工程、计量电影学。主持国家社科基金艺术学、国家社科基金重大项目子课题、横向开发多项。主持开发了多个智能信息系统中国剧本库(ruyi.cool)、智能影视平台(yingshinet.com)、电影翻译库(film.yingshinet.com)、中国电影知识体系平台(movie.yingshinet.com),出版了《数据可视化原理与实例》《大数据技术导论》,申请专利 6 项、软著 6 项。



**石川** 北京邮电大学计算机学院教授、博士生导师、智能通信软件与多媒体北京市重点实验室副主任。主要研究方向为数据挖掘和机器学习,特别专注于图数据挖掘与学习。近 5 年以第一作者或通信作者在 CCF A 类期刊和会议上发表论文 50 余篇,中英文专著 5 部,授权发明专利 20 余项,相关研究成果应用于阿里巴巴、腾讯、华为、美团等公司。获得 ADMA2011/ AMDA2018 最佳论文奖和 WWW2019 最佳论文候选。研究成果获得省部级奖励 5 项,包括北京市/CCF 科学技术奖自然科学二等奖(第一)和吴文俊人工智能科技进步一等奖(第三),获得北京市高等学校青年英才和师德先锋等称号。