软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2022,33(7):2365−2366 [doi: 10.13328/j.cnki.jos.006591]

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

智能系统的分析和验证专题前言 *

明 仲[1], 张立军[2,3], 秦胜潮[4]

[1](深圳大学 计算机与软件学院, 广东 深圳 518061)

[2](中国科学院 软件研究所, 北京 100190)

[3](中国科学院大学, 北京 100049)

[4](华为香港研究所, 香港 999077)

通信作者: 明仲, E-mail: mingz@szu.edu.cn

中文引用格式: 明仲, 张立军, 秦胜潮. 智能系统的分析和验证专题前言. 软件学报, 2022, 33(7): 2365−2366. http://www.jos.org.cn/1000-9825/6591.htm

近年来, 基于深度学习算法的智能系统在一些长期未解决的任务中取得了与人类相当的能力. 然而另一方面, 智能系统同时面临着亟待解决的安全性和可靠性等可信性问题, 比如对于自动驾驶系统, 路标识别错误可能会导致灾难性的后果. 智能系统的可信性已经逐渐成为制约人工智能技术在实际生产和生活中应用的关键问题, 尤其是安全攸关领域. 专题围绕智能系统的可信性问题, 探讨可信智能系统在学术和产业界面临的难题、挑战和瓶颈. 该专题重点关注智能系统的安全内涵与可解释性、智能系统的形式化验证、智能系统的测试技术、智能系统的对抗攻击技术等相关技术方法, 并探讨可信智能系统的应用前景.

本专题公开征文, 共收到投稿 24 篇. 其中 19 篇论文通过了形式审查, 内容涉及智能系统的安全内涵与可解释性、形式化验证、测试、对抗攻击与应用研究. 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、ChinaSoft 2021 会议宣读和终审共 4 个阶段, 历时 6 个月, 最终有 8 篇论文入选本专题. 根据主题, 这些论文可以分为 4 组.

(1) 安全内涵与可解释性

《可信系统性质的分类和形式化研究综述》对可信系统的需求和形式化方法在不同系统中的应用进行不同维度的分类, 并介绍和总结了不同系统特征不同应用场景下的形式建模、性质描述以及验证方法与工具, 以更好地支撑基于形式化方法的可信软硬件系统的分析与验证.

《人脸识别反欺诈研究进展》综述人脸反欺诈技术(FAS)的最新研究进展, 总结了当前 FAS 所面临的主要科学问题以及主要的解决方法及其优缺点, 从理论和实践的角度说明了基于深度学习的 FAS 域泛化和可解释性问题, 给出了 FAS 算法的评估标准和实验对比结果, 并对未来研究方向进行展望.

《基于最小不满足核的随机森林局部解释性分析》提出了一种基于形式化和逻辑推理方法的机器学习可解释性方法, 用于解释随机森林的预测结果. 具体来说, 该论文将随机森林模型的决策过程编码为一阶逻辑公式, 并以最小不满足核为核心, 提供了关于特征重要性的局部解释以及反事实样本生成方法.

(2) 形式化验证技术

《基于多路径回溯的神经网络验证方法》提出了多路径回溯的概念, 并指出现有的基于线性抽象的符号传播方法仅使用单条回溯路径计算神经网络节点上下界, 是多路径回溯的一种特例.

《基于概率模型检查的树模型公平性验证方法》提出了一种基于概率模型检查的方法来形式化验证决策树和树集成模型的公平性. 具体来说, 该论文将公平性问题转换为概率验证问题, 为算法模型构建 PCPS#模型, 并使用 PAT 模型检查工具求解, 以不同定义的公平性度量衡量模型公平性.

* 收稿时间: 2022-01-30

(3) 测试与对抗攻击技术

《采用多目标优化的深度学习测试优化方法》提出了一种基于多目标优化的深度学习测试输入选择方法 DMOS. 具体来说, 该方法首先基于 HDBSCAN 聚类方法初步分析原始测试集的数据分布, 然后基于聚类结果的特征设计多个优化目标, 最后利用多目标优化求解出合适的选择方案.

《基于 Rectified Adam 和颜色不变性的对抗迁移攻击》提出了一种提高对抗样本可迁移性的方法. 具体来说, 该方法将 RAdam 优化算法与迭代快速符号下降法相结合, 并利用目标函数的二阶导信息来生成对抗样本; 同时, 通过优化对颜色变换图像集合的扰动, 针对防御模型生成更多可迁移的对被攻击的白盒模型不太敏感的对抗样本.

(4) 智能系统的应用研究

《安全强化学习算法及其在 CPS 智能控制中的应用》研究安全强化学习智能控制方法并运用于安全攸关 CPS 的控制器生成, 致力于保障系统的安全性质, 与此同时, 能够达到系统最优控制目标; 围绕一个工业油泵控制系统典型案例, 开展安全强化学习算法和智能控制应用研究.

本专题主要面向形式化方法、软件工程、机器学习、人工智能等多领域的研究人员和工程人员, 反映了我国学者在人工智能系统的分析与验证领域最新的研究进展. 感谢《软件学报》编委会和形式化方法专委会对专题工作的指导和帮助, 感谢专题全体评审专家及时、耐心、细致的评审工作, 特别感谢踊跃投稿的所有作者. 希望本专题能够对人工智能系统的分析与验证相关领域的研究工作有所促进.

明仲(1967－), 男, 博士, 教授, CCF 高级会员. 现任深圳大学研究生院执行院长, 光明实验室副主任, 深圳大学计算机科学与技术博士点带头人. 主要研究领域为人工智能, 软件工程, 推荐系统. 曾获国家教学成果奖 1 次和省教学成果一等奖 4 次, 曾获广东省科学技术奖一等奖(排名第一)、中国电子学会科技进步一等奖(排名第一)、吴文俊人工智能一等奖(排名第二)、广东省科技进步二等奖(排名第二).

张立军(1979－), 男, 博士, 研究员, 中国科学院大学特聘教授, 博士生导师, CCF 高级会员. 主要研究领域为形式化方法、程序语言/软件工程、人工智能. 曾在权威学术会议与期刊上发表论文 100 余篇. 曾主持科技部重点研发计划课题、国家自然基金委重点项目、国家自然基金委国际合作项目等多个科研项目. 曾任形式化方法国际会议 LICS 执行委员会委员, 并发系统会议 CONCUR 及形式验证会议 TACAS 等大会主席.

秦胜潮(1974－), 博士, 教授, IEEE 高级会员, ACM 高级会员. 现任华为高级技术专家. 主要研究领域为软件理论与形式化方法, 软件工程, 程序语言. 曾在国际知名期刊和会议发表高水平论文 130 多篇.