

高可信自适应机器学习专题前言*

王熙熙¹, 张敏灵², 曹付元³

¹(深圳大学 计算机与软件学院, 广东 深圳 518060)

²(东南大学 计算机科学与工程学院, 江苏 南京 211189)

³(山西大学 计算机与信息技术学院, 山西 太原 030006)

通信作者: 王熙熙, E-mail: xzwang@szu.edu.cn



中文引用格式: 王熙熙, 张敏灵, 曹付元. 高可信自适应机器学习专题前言. 软件学报, 2026, 37(4): 1447-1448. <http://www.jos.org.cn/1000-9825/7531.htm>

近年来,随着数据规模的不断扩大和计算能力的提升,机器学习模型的可信性和适应性需求日益凸显,高可信自适应机器学习应运而生.其发展的新趋势体现在3个方面:(1)机器学习算法的不断升级对模型的鲁棒性、安全性和可解释性提出了全新挑战;(2)各领域的复杂需求推动了机器学习系统的自动化演进,包括数据选择、模型调整和算法设计等;(3)算力资源的不断提升为高可信自适应模型的自动化调优和多样化数据处理提供了新的支持.专题聚焦高可信自适应机器学习在可信性与自适应性提升中的前沿探索,致力于呈现具有创新性和实用性的高水平研究成果.重点探讨基础理论、关键技术及系统实现,分享新兴学习模式和开放环境下的模型动态更新与自适应训练等前沿研究,进一步推动可信机器学习方向的学术交流与技术发展.

本专题公开征文,共收到投稿24篇.论文均通过了形式审查,内容涉及样本选择、异常检测、可解释性、鲁棒学习、联邦学习等.特约编辑先后邀请了30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、CCML 2025会议宣读和终审4个阶段,历时3个多月,最终有9篇论文入选本专题.这些论文的内容如下.

《[基于高质量样本选择的跨领域方面级情感分析](#)》以多领域情感分析为开放环境,结合高可信机器学习理论与领域适应方法,提出基于高质量样本选择的跨领域方面级情感分析方法,在跨领域迁移任务及方面级情感分析中验证了方法有效性.

《[基于自适应策略优化的鲁棒精度权衡学习](#)》利用演化优化提出一个自适应对抗训练框架,将遗传算法引入对抗训练过程以动态调整不同训练阶段的对抗攻击策略,实现对抗样本生成模式的渐进式复杂化.该机制还能通过策略优化记录实现训练早停以抑制过拟合.

《[基于强-弱互信息掩码学习的可解释动态不完整图异常检测](#)》提出一种可解释动态图异常检测方法,结合强-弱互信息优化的图掩码自编码器框架,通过掩码模拟动态不完整场景、强-弱互信息损失提升模型泛化能力,并借助掩码扰动增强可解释性.

《[面向欺诈检测的风险感知动态聚合图联邦学习](#)》提出面向欺诈检测的风险感知动态聚合图联邦学习方法,通过典型风险动态聚合策略增强模型对正常样本和典型欺诈样本的判别力,并使用多样化风险平均聚合策略进一步提升模型在复杂欺诈风险场景下的识别鲁棒性.

《[面向开放世界持续学习的任务敏感提示驱动混合专家模型](#)》在开放世界持续学习形式化定义基础上,提出任务敏感提示驱动的任务敏感提示驱动混合专家模型,通过即插即用的任务提示聚合机制与改进门限机制融合历史与当前任务知识,结合自适应开放边界阈值策略调整开放类别判别边界以提升检测与分类性能.

《[数值型标签噪声的渐进式区间校正方法](#)》提出回归噪声标签的渐进式区间校正算法,其先基于真实标签后

* 收稿时间: 2025-08-31; jos 在线出版时间: 2025-09-02

验分布给出标签校正有效性条件, 再对满足条件的标签进行最大后验校正, 最后通过逐步缩小可信区间渐进校正优化标签.

《全局与局部残差信息联合感知的可泛化图异常检测》提出全局和局部残差信息联合感知的可泛化图异常检测方法, 在图神经网络 (graph neural network, GNN) 建模局部节点关系基础上引入线性 Transformer 模块, 再从全局和局部视角构建残差, 最后将二者融合以构建数据集无关的通用节点表示.

《基于动量加速和任务均衡的目标检测对抗训练方法》提出基于 Nesterov 加速梯度的对抗训练方法, 设计自适应损失重加权策略以均衡对抗样本的数量占比, 促进模型聚焦定位以增强鲁棒性.

《可解释深度学习的概念建模方法综述》对比了基于信息流与基于概念的两类可解释性方法, 围绕可解释深度学习中的概念建模方法展开综述, 旨在为理解和构建语义可解释深度模型提供系统性参考与方法指引.

本专题主要面向人工智能、机器学习、数据挖掘等多领域的研究人员和工程人员, 反映了我国学者在高可信自适应机器学习领域最新的研究进展. 感谢《软件学报》编委会和中国人工智能学会机器学习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, 感谢踊跃投稿的所有作者. 希望本专题能够对高可信自适应机器学习相关领域的研究工作有所促进.

作者简介

王熙熙, 博士, 教授, 博士生导师, 主要研究领域为机器学习.

张敏灵, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为机器学习, 数据挖掘.

曹付元, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为数据挖掘, 机器学习.