智能化软件可信构造与质量保障研究获进展

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/40813.html

智能化软件可信构造与质量保障研究获进展
。近日，中国科学院软件研究所团队在大规模开源生态治理、代码自动化审查、高安全领域代码生成理解等方面取得了系列进展。

在大规模软件生态治理方面，针对现有开源生态组织机制的扁平化、非规范化及覆盖率不足等问题，研究团队提出了基于多智能体协作的自主演化治理框架ATLAS。该框架通过设计智能体提取分类维度，并结合分类智能体在实际数据分布上的反馈，构建“设计—分类—修正”闭环，实现了开源仓库的自动化、多层级语义分类。实验结果表明，在GitHub 5.4万个真实仓库上，ATLAS自动构建高质量层级分类体系。同时，ATLAS支撑的软件替代品发现任务P@1达85.71%，超越人工维护列表，并可以揭示生态向AI/ML应用转型的趋势。

在自动化代码审查方面，针对现有自动化代码审查任务建模割裂、模型逻辑推理能力不足且审查过程缺乏解释性等问题，研究团队提出了端到端推理引导的对齐方法E4R-Reviewer。该方法将质量估计、缺陷定位、缺陷分类、缺陷描述、修复建议及代码优化等任务统一为单一推理链条，利用群组相对策略优化方案，将推理步骤转化为可优化的中间目标。实验显示，该方法在质量估计任务中达到74.61%的F1值，增强了审查过程的透明度与可解释性。

针对大语言模型在Bash脚本生成过程中缺乏透明推理且生成代码鲁棒性不足问题，研究团队提出了可鲁棒性感知的大模型生成框架BashCoder-R1。该框架引入长思维链监督微调，配套鲁棒性感知组相对策略优化，将语法正确性、静态分析规则和格式规范转化为强化学习奖惩信号，强制模型在生成代码前执行显式推理。在包含952项任务的BashBench基准测试中，BashCoder-R1多行脚本任务的FullRate达73.18%，较DeepSeek-V3.2提升效果突出。

针对大语言模型对Bash脚本中复杂命令语义细节的逻辑推理能力不足，导致生成注释准确性难以保证的问题，研究团队提出了语法感知偏好优化方法Bash-Commenter。该方法通过对脚本抽象语法树应用原子操作自动构造“最小语法对”，并以此作为偏好学习信号注入模型权重，使模型掌握细粒度的命令语义理解能力。实验评估表明，相较于主流方法，Bash-Commenter在单行及多行脚本生成的各项指标上均达到最佳性能。

在高安全智能合约生成方面，针对大语言模型对智能合约隐式安全规则逻辑推理能力不足，导致生成代码鲁棒性较弱的挑战，研究团队提出了高安全防御性生成架构SmartCoder-R1。该架构通过安全感知的组相对策略优化，将12类关键安全规则转化为强化学习硬约束，实现“逻辑—代码”端到端对齐。在涉及289个已部署合约的测试中，SmartCoder-R1将安全性指标FullRate提升至50.53%，较DeepSeek-R1相对提升45.79%。

上述成果为智能化软件开发与质量保障，提供了新的技术路径与理论支撑。相关方法在开源生态治理、代码审查自动化及高安全领域代码生成等场景中具有应用前景。

相关论文分别被软件工程领域国际会议ASE 2026、FSE 2026和ISSTA 2026录用。

研究团队单位：软件研究所