

个人简历

基本信息

姓名: 张峻伟 出生年月: 1995.05
电话: 13508307004 政治面貌: 中共党员
邮箱: jw.zhang@zju.edu.cn 在读院校: 浙江大学
个人主页: <http://www.jw-zhang.cn/> 学历: 博士研究生 (在读)



研究方向/兴趣

- 智能化软件工程**: 基于大模型的单元测试生成技术及优化、基于大模型的软件漏洞检测
- 代码仓库挖掘**: 开源软件社区中软件缺陷报告数据的挖掘与分析、软件开发人员社交信息挖掘
- 基于深度学习的推荐系统**: 基于超图的推荐系统研究、基于知识图谱的社交推荐系统研究

教育、工作经历

浙江大学 | 杭州, 浙江 2021.09 - 2025.06
- 博士在读 | 电子信息 - 导师: 李善平
华为杭州研究院 | 杭州, 浙江 2022.01 - 2025.01
- 实习生 | 软件工程应用技术实验室 - 导师: 夏鑫
重庆大学 | 重庆 2018.06 - 2021.06
- 硕士 | 软件工程 - 导师: 高旻
重庆大学 | 重庆 2014.09 - 2018.06
- 本科 | 软件工程 - 导师: 高旻

发表论文

近五年已录用/发表会议期刊论文 9 篇, 其中本人一作或导师一作本人二作论文 7 篇, CCF A 期刊/会议论文 3 篇, CCF B 期刊/会议论文 4 篇, SCI 一区论文 1 篇。

1. Less Is More: On the Importance of Data Quality for Unit Test Generation | 2025

- 发表渠道: *The ACM International Conference on the Foundations of Software Engineering (FSE)* (CCF-A 会议)
- 作者信息: **Junwei Zhang**, Xing Hu*, Shan Gao, Xin Xia, David Lo, and Shanping Li
- 论文简介: 该工作研究了数据噪声对基于大模型的测试生成模型性能影响, 首先提出了八种不同类型的噪声, 然后通过 17 位领域专家的访谈验证和评估噪声分类法的重要性、合理性和正确性, 最后提出了一个数据噪声清除框架 CleanTest 并在多个大模型上验证其有效性。

2. TRACETS4J: A Traceable Unit Test Generation Dataset | 2025

- 发表渠道: *2025 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)* (CCF-B 会议)
- 作者信息: Xuancheng Jin, Zhuang Liu, **Junwei Zhang**, Xing Hu*, and Xin Xia
- 论文简介: 该工作对现有单元测试生成数据集进行了实证研究, 分析可追溯性链接不准确的原因: 错误提取和遗漏提取, 并提出了一个联合跟踪启发式方法用于建立测试到被测试代码的可追溯链接。

3. Dual Prompt-Based Few-Shot Learning for Automated Vulnerability Patch Localization | 2024

- 发表渠道: 2024 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER) (CCF-B 会议)
- 作者信息: Junwei Zhang, Xing Hu*, Lingfeng Bao, Xin Xia, and Shanping Li
- 论文简介: 本文提出了一种基于大模型的安全补丁定位方法, 其利用提示微调策略捕捉漏洞描述和代码提交之间的语义相关性, 尤其是在数据稀疏的场景中。

4. Inside Bug Report Templates: An Empirical Study on Bug Report Templates in Open-Source Software | 2024

- 发表渠道: Proceedings of the 15th Asia-Pacific Symposium on Internetware (Internetware) (CCF-C 会议)
- 作者信息: Junwei Zhang, Zhongxin Liu, Lingfeng Bao*, Zhenchang Xing, Xing Hu, and Xin Xia
- 论文简介: 本文对 GitHub 中 3,194 个项目的 5,987 条缺陷报告模版进行了定量和定性分析, 发现采用缺陷报告模版可以减少评论评价数量, 增加缺陷报告被解决的可能性, 有助于开发人员识别重复的缺陷报告。

5. Pre-training by Predicting Program Dependencies for Vulnerability Analysis Tasks | 2024

- 发表渠道: ACM/IEEE 46nd International Conference on Software Engineering (ICSE) (CCF-A 会议)
- 作者信息: Zhongxin Liu, Zhijie Tang, Junwei Zhang, Xin Xia, and Xiaohu Yang
- 论文简介: 本文提出了两个新的预训练目标用于学习分析代码中细粒度依赖关系所需的知识, 即控制依赖预测 (CDP) 和数据依赖预测 (DDP), 旨在仅基于代码片段的源代码分别预测代码片段中的语句级控制依赖关系和标记级数据依赖关系。模型在三个漏洞分析任务 (即漏洞检测、漏洞分类和漏洞评估) 上对其进行了微调, 并在程序依赖性分析上对其进行了评估。

6. Vulnerability Detection by Learning From Syntax-Based Execution Paths of Code | 2023

- 发表渠道: IEEE Transactions on Software Engineering (CCF-A 期刊)
- 作者信息: Junwei Zhang, Zhongxin Liu*, Xing Hu, Xin Xia, and Shanping Li
- 论文简介: 本文设计基于贪心算法的路径选择策略从代码的控制流图中抽取多条可执行路径, 并利用代码大模型学习抽取路径的特征表示, 捕捉代码片段中与漏洞相关的重要信息, 保证原有代码片段语义信息的同时缩短代码长度, 过滤与漏洞无关的信息。

7. Double-Scale Self-Supervised Hypergraph Learning for Group Recommendation | 2021

- 发表渠道: The 30th ACM International Conference on Information and Knowledge Management (CIKM) (CCF-B 会议)
- 作者信息: Junwei Zhang, Min Gao*, Junliang Yu, Lei Guo, Jundong Li, and Hongzhi Yin
- 论文简介: 设计采用超图神经网络模型学习群组推荐中群组与群组、用户与用户之间的高阶交互关系, 并利用多粒度自监督学习策略缓解群组推荐中的数据稀疏问题。在 Douban、Weeplaces、GAMRa2011 等开源数据集中验证超图神经网络和自监督学习策略在群组推荐模型的可行性。

8. Recommender Systems Based on Generative Adversarial Networks: A Problem-Driven Perspective | 2021

- 发表渠道: Information Science (SCI 1 区, CCF-B 期刊)
- 作者信息: Min Gao*, Junwei Zhang, Junliang Yu, Jundong Li, Junhao Wen, and Qingyu Xiong
- 论文简介: 以问题驱动角度统计分析了三十多种基于生成对抗网络的推荐模型, 并将其按照解决推荐系统中数据稀疏问题和数据噪声问题进行分类。

9. Path-Based Reasoning over Heterogeneous Networks for Recommendation via Bidirectional Modeling | 2021

- 发表渠道: *Neurocomputing* (SCI 2 区, CCF-C 期刊)
- 作者信息: **Junwei Zhang**, Min Gao*, Junliang Yu, Linda Yang, Zonwei Wang, and Qingyu Xiong
- 论文简介: 该工作利用基于注意力机制的双向 LSTM 模型建模异构网络中的节点序列, 并利用注意力机制解释推荐原因。此外, 在节点向量中加入对抗性正则化项, 以提高模型鲁棒性, 减少数据噪声的影响。在 Nowplaying, Yahoo, Xiami 等开源音乐数据集中验证了所提推荐模型的可行性、准确性以及鲁棒性。

专 利

1. A Recommendation Algorithm Based on Adversarial Learning and Bidirectional Long Short-Term Memory Network | 2020.09

- 专利授权号: 202010903794.8
- 人员信息: 高旻, **张峻伟**, 余俊良, 王宗威, 熊庆宇, 赵泉午, 王旭

学 术 服 务

- 1. 报告:
 - 会议报告: SANER 2024, 罗瓦涅米, 芬兰, 报告题目: Dual Prompt-Based Few-Shot Learning for Automated Vulnerability Patch Localization (CCF-B 会议)
 - 会议报告: CIKM 2021, 线上, 报告题目: Double-Scale Self-Supervised Hypergraph Learning for Group Recommendation
- 2. 审稿服务:
 - IEEE Transactions on Software Engineering* (TSE, CCF-A 期刊) 审稿人
- 3. 会议服务:
 - Internetwork 2023 会议志愿者
 - 2020 年中国软件大会 ChinaSoft2020 会议志愿者

荣 誉 奖 项

优秀研究生	浙江大学	2023-2024 学年
研究生奖学金	浙江大学	2022-2024 学年
优秀共产党员	浙江大学计算机科学与技术学院	2023-2024 学年
优秀硕士学位论文	重庆市	2021 年
研究生国家奖学金	重庆大学	2020 年
航天奖学金	重庆大学	2019 年
学业奖学金	重庆大学	2018-2021 学年
优秀学生	重庆大学	2019 年
第二届互联网+创新创业大赛”全国银奖	教育部	2016 年