

# 许汤智

+86 13852554653 xutz@smail.nju.edu.cn 南京

## 教育经历

|             |                     |
|-------------|---------------------|
| 南京大学        | 2023年09月 - 2028年06月 |
| 计算机科学与技术 博士 | 南京                  |
| 南京大学        | 2019年09月 - 2023年06月 |
| 计算机科学与技术 本科 | 南京                  |

## 科研经历

### Comprehend, Imitate, and then Update: Unleashing the Power of LLMs in Test Suite Evolution (ASE25-一作在投)

- 在软件开发的过程中, 维护良好的测试套件对于确保软件质量有着重要作用。然而, 在强调快速功能迭代的开发周期中, 测试套件的协同进化经常滞后, 导致出现过时的测试套件。
- 提出基于LLM的框架来自动化更新过时的测试代码, 结合意图分析, 静态分析等来帮助LLM理解变更意图, 利用RAG技术来检索指导LLM更新的示例, 并最终结合错误定位和反馈技术来执行迭代更新。
- 建立了一个Java的数据集。结果表明本框架在96.4%的数据集上生成了可编译的测试更新, 在93.1%的数据集上生成完全覆盖产品代码的测试更新, 且在不同LLM基座上表现均比baseline高15%以上。

### Streamlining Repository Tasks with Effective Snippet Retrieval (TOSEM-一作在投)

实习工作 蚂蚁集团-智能平台工程 杭州

- 对于仓库级别的软工任务, 引入了一个新型检索器, 将语言模型和工具合理的组织成与人类搜索逻辑相似的代理工作流: 搜索文件-过滤文件-搜索片段-过滤片段, 并将检索结果直接用于下游任务。
- 结合本框架与不同规模LLM在SWE-bench-Lite进行检索, 基于GPT4o的片段Acc@10达到80.44%, 文件Acc@5达到87%, 基于Qwen2.5-3b片段Acc@10达到41%, 文件Acc@5达到58%。结合GPT4o的检索结果使用GPT4o进行直接修复的Pass@1达到33%, Pass@10达到45%。
- 使用本框架在RepoQA上检索指定函数, 基于GPT4o的检索准确率达到80%; 在RepoEval上结合本框架检索结果进行代码补全, EM比RepoCoder提升了7%以上。

### LoRA Decompose: Serving fine-tuned models into LoRA-like (ECAI25-二作)

- 提出了一种LLM的低秩分解方法, 在保持一系列Llama2-7b微调模型在下游任务上的准确性基本不变的情况下, 能够将GPU显存使用量减少至原来的1/32, 大幅减少了多模型服务的资源占用。
- 将多个微调模型分解为一个共享的类似预训练模型的基本组件和各自的类似LoRA的增量组件, 将问题构建为约束优化问题并设计了一个优化算法, 在几次迭代内在确保模型准确性的同时最小化增量组件的秩。

### Clause2Inv: A Generate-Combine-Check Framework for Loop Invariant Inference(ISSTA25-三作)

- 在循环不变式推导过程中, 在传统的 guess-and-check 框架基础上, 提出了一个新的 generate-combine-check 框架, 将不变式推导任务显式划分为子句生成和子句组合两个阶段, 充分发挥大语言模型在子句生成方面的优势, 规避了其在逻辑连接结构推导上的劣势。
- 基于该框架实现了一个具体方法实例, 由基于大语言模型的子句生成器和反例驱动的子句组合算法组成, 协同构造不变式。实验结果表明, 我们的方法在包含 316 个线性问题的基准测试集上成功解决了 312 个, 优于当时最优方法所解决的 219 个; 同时, 在非线性问题的基准测试集上的表现也超越了现有最优方法。

## 荣誉奖项

|                |            |
|----------------|------------|
| 南京大学研究生一等学业奖学金 | 2024, 2025 |
| 南京大学直博生校长特别奖学金 | 2023       |
| 南京大学人民奖学金二等奖   | 2022       |
| 南京大学人民奖学金三等奖   | 2021, 2020 |