

## 刘芳

pku.fangliu@gmail.com | +86 18801218303  
北京大学理科 1 号楼 1726 100871

### 教育经历

---

#### 北京大学

2017.9-现在

信息科学技术学院，博士学位在读，预计 2022 年 6 月获得博士学位  
导师：金芝 研究方向：深度学习，机器学习，程序分析

#### 重庆大学

2013.9 - 2017.7

专业：计算机科学与技术，2017 年 6 月获得学士学位  
GPA 排名：1/324

### 科研 & 实习经历

---

#### 基于全局上下文的函数名推荐

2020-2021

- 发表于 International Conference on Software Engineering
- 提出了基于全局上下文的方法名推荐模型，同时考虑了方法体的局部信息，项目相关的上下文信息，以及方法的注释信息。
- 构建了基于 Transformer 的序列到序列模型，并利用自注意力机制在生成方法名时关注不同级别的上下文信息。
- 该算法在 Java 数据集的方法名推荐任务中取得了比基准模型更高的准确率。

#### 基于预训练语言模型的代码补全

2019-2020

- 发表于 International Conference on Automated Software Engineering
- 构建基于 transformer 网络的预训练语言模型用于代码理解与生成，设计了三个预训练任务使得模型能够学到更加准确的程序特征向量表示。
- 利用程序代码中的变量的静态类型信息辅助模型更好地理解程序。
- 该算法在 Java 和 TypeScript 数据集上的代码补全任务中取得了比基准模型更高的准确率。

#### 基于自注意力机制和多任务学习的代码补全

2018-2019

- 发表于 International Conference on Program Comprehension (ACM Distinguished Paper Award)
- 构建了基于多任务学习的代码补全模型，在基于 AST 的代码补全任务中同时预测下一个节点的类型和值。
- 使用 Transformer-xl 作为基础模型学习代码中的长依赖关系，并考虑了根节点到待补全节点的路径信息增强对程序的结构信息的建模。
- 该模型在 Python, Java, JavaScript 三个数据集上的代码补全任务上取得了更高的准确率。

#### 层次化程序语言建模方法

2017-2018

- 发表于 Journal of Systems and Software
- 提出了可以对序列中的层次结构信息进行建模的语言模型，该模型基于 LSTM 神经网络，并引入了栈来记录序列中的层次结构。
- 该方法在代码补全、程序分类以及代码摘要任务中均取得提升。

## 算法研究实习生，微软亚洲研究院

2021.03-2021.08

- 数据知识智能组 (Data, Knowledge & Intelligence)
- spreadsheet 预训练模型设计与实现: 提出了基于二维 Transformer 的模型提取每个单元格周围的特征信息, 设计不同的 attention head 学习 cell 之间的依赖关系; 设计预训练任务使得模型能够学到 sheet 中数据的同秩性和语义特征。
- 大规模数据处理: 利用服务器集群处理千万级数据。
- 参与完成算法部署: 将表格边界检测模型部署到 client 端, 优化算法逻辑实现加速。

## 发表成果

---

- **Fang Liu**, Ge Li, Zhiyi Fu, Shuai Lu, Yiyang Hao, Zhi Jin, Learning to Recommend Method Names with Global Context, International Conference on Software Engineering (ICSE 2022)
- **Fang Liu**, Ge Li, Yunfei Zhao, Zhi Jin, Multi-task Learning based Pre-trained Language Model for Code Completion, International Conference on Automated Software Engineering (ASE 2020)
- **Fang Liu**, Ge Li, Bolin Wei, Xin Xia, Zhiyi Fu, Zhi Jin, A Self-Attentional Neural Architecture for Code Completion with Multi-Task Learning, International Conference on Program Comprehension (ICPC 2020, ACM Distinguished Paper Award)
- **Fang Liu**, Lu Zhang, Zhi Jin, Modeling programs hierarchically with stack-augmented LSTM, Journal of Systems and Software (JSS 2020)
- **刘芳**, 李戈, 胡星, 金芝, 基于深度学习的程序理解研究进展, 计算机研究与发展. 2019
- 胡星, 李戈, **刘芳**, 金芝, 基于深度学习的程序生成与补全技术研究进展, 软件学报. 2019
- 金芝, **刘芳**, 李戈, 程序理解: 现状与未来, 软件学报. 2019

## 奖励与荣誉

---

- The ACM SIGSOFT Distinguished Paper Award in International Conference on Program Comprehension (ICPC 2020)
- 北京大学优秀科研奖, 2018, 2020
- 重庆市优秀大学毕业生, 2017
- 国家励志奖学金, 2016
- 国家奖学金, 2014

## 技能

---

掌握 Python, Tensorflow, Pytorch, LaTeX

熟悉 C/C++, C#, Java, TypeScript, 基本数据结构及算法

英语流利