

教育经历

软件工程硕士 | 上海交通大学

2018.9 -

大数据软件工程实验室 | 导师: 沈备军 (<http://base.sjtu.edu.cn/~bjshen/>)

中国上海

研究方向: 缺陷定位, 代码知识图谱

软件工程学士 | 上海交通大学

2014.9 - 2018.6

主要课程: 数据结构与算法 / 软件工程 / 数据挖掘 / 计算机系统

中国上海

研究经历

BugPecker: Locating faulty methods with deep learning on revision graphs

Junming Cao, Shouliang Yang, Wenhui Jiang, Hushuang Zeng, Beijun Shen, and Hao Zhong.

The 35th IEEE/ACM International Conference on Automated Software Engineering (ASE 2020, new idea)

CocoQa: Question Answering for Coding Conventions over Knowledge Graphs

Tianjiao Du, Junming Cao, Qinyue Wu, Wei Li, Beijun Shen*, Yuting Chen.

The 34th IEEE/ACM International Conference on Automated Software Engineering (ASE 2019, tool demo)

Constructing a Knowledge Base of Coding Conventions from Online Resources

Junming Cao, Tianjiao Du, Beijun Shen*, Wei Li, Qinyue Wu, Yuting Chen.

The 31th International Conference on Software Engineering and Knowledge Engineering (SEKE 2019)

项目经历

BugPecker | 基于缺陷报告的方法级缺陷定位系统

2019.9 - 2020.7

大数据软件工程实验室, 上海交通大学

<https://github.com/RAddRiceee/BugPecker>

挑战: 代码仓库中很多短方法缺乏足够的语义信息(例如Tomcat中46.49%的方法小于三条语句), 因此难以被匹配到

方法: 从过去的缺陷修复记录中构建修复图, 挖掘缺陷报告间及方法间相似关系

结果: 与最新的缺陷定位技术相比, BugPecker显著改进了MAP值

CocoQa | 面向编码规范的问答系统

2018.9 - 2019.5

大数据软件工程实验室, 上海交通大学

<https://github.com/14dtj/CocoQa>

将自然语言语句解析成SPARQL语句, 通过子图匹配从编码规范知识图谱中检索得到相关的节点与关系

将编码规范当做段落, 利用机器阅读理解技术(Match-LSTM模型)从中抽取相关的编码规范

使用logistic classifier来对上面两种的候选结果进行排序并返回

将CocoQA实现为一个IntelliJ IDEA的插件, 并也支持通过Web接口访问

DeepEye | 鉴黄与暴恐识别系统

2017.7 - 2017.9

后台开发实习生, 腾讯优图实验室

中国上海

通过Nginx模块开发的方式将机器学习模型部署为高性能的Web服务

基于Scrapy开发了分布式爬虫系统

每天可爬取上百万张图片, 用于图片识别系统的训练与测试

技能

Java / C++ (中级)

Git / Spark / PyTorch

Go/Python/Scala(初级)

比赛奖项

CUMCM (中国大学生数学建模竞赛 2018): 国家二等奖

MCM(美国大学生数学建模竞赛 2017): 一等奖