

王子源

联系方式: (+86)188-1787-6918

电子邮箱: wangziyuan@163.sufe.edu.cn

通讯地址: 上海市杨浦区国定路 777 号上海财经大学

出生日期: 1996 年 4 月 6 日

政治面貌: 中共党员

研究方向: 自然语言处理/深度学习

个人网站: [My Website](#); [CSDN\(77000+访问\)](#); [Github\(580+Star\)](#); [LinkedIn](#);



教育背景

国家留基委公派联合培养	美国爱荷华大学蒂皮商学院	商业分析	2022.4-2023.4
-------------	--------------	------	---------------

研究方向: 深度学习、自然语言处理、文本分类、语义相似度计算、设计科学

博士在读(硕博连读)	上海财经大学信息管理与工程学院	管理科学与工程	2018.9 至今
------------	-----------------	---------	-----------

GPA: 3.58/4.0 学院排名: top15%

核心课程: 机器学习、金融工程、高级计量经济学、高级统计学、高级运筹学、优化理论、随机模型

所获荣誉: 博士一等奖学金、二等奖学金, 硕士二等奖学金、校优秀班干部、校优秀学生

本科	湖南大学金融与统计学院	统计学	2014.9-2018.6
----	-------------	-----	---------------

GPA: 3.8/4.5 学院排名: top5%

核心课程: 统计学、多元统计分析、抽样技术、计量经济学、数理统计学、程序设计基础、统计软件应用

所获荣誉: 国家奖学金、湖南省优秀毕业生、花旗银行“未来精英”奖学金, 校三好学生、校优秀班干部

科研及项目经历

《How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection》2023.1-2023.2

- 我们发起的 ChatGPT 对比与检测项目通过收集人类-ChatGPT 对比数据, 进行丰富的统计、语言学等分析, 并基于深度学习、机器学习等技术, 开发了一系列 ChatGPT 内容检测器。
- 最早在学术界和工业界开源比较数据集和检测器模型; 检测器 demo 全球访问量超 20000 次, 开源模型月均下载量超 3000, 数据集月均下载量超 1000, Github stars 超过 580。访问我们的 [Demo](#) 和 [Paper](#)。

《IDEA: Interactive Double Attentions from Label Embedding for Text Classification》2022.6-2022.9

- 提出了一种利用监督学习中的标签嵌入技术来提高 BERT 在文本分类中的性能的简洁方法。在原有 BERT 模型的基础上, 提出了一种新的融合文本和标签信息的模型结构。
- 发表于 34th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2022, CCF-C), 国家自然科学基金项目 (72271151)。[Paper](#)。

《Investigating Effectiveness of Whitening Post-processing on Modifying LLMs》2022.10-2023.2

- 传统 Bert 语义表示忽略了余弦相似度基底正交的假设, 通过探究不同向量矩阵白化方法对 BERT 模型的向量化表示进行去相关的规范化, 及在不同 NLP 任务中的表现。
- 已投稿 SIGIR'23(CCF-A); 上海财经大学研究生创新基金项目 (编号: CXJJ-2021-052)。

《New Hybrid Model of Crowdfunding Project: A Perspective of Prospect Backers》2022.8 至今

- 本文引入前景理论 (PT) 来捕捉众筹中投资者支持行为的异质性, 对项目的前景效用进行评估, 根据评估的前景效用进行支持决策。

- 已投稿 Decision Support System(SCI I 区)。

Web Tool: SUFE-CS-CONF-DDL

2022.1-2022.3

- 基于 Vue-Cli 搭建上财常任轨计算机会议倒计时系统工具，提供常任轨 Tier 等级/CCF 等级双重检索
- 可视化网站可见 [SUFE-CS-CONF-DDL](#)，项目开源代码可见 [Github](#)。

基于大数据的上海市生物医药产业链分析及长三角比较研究

2021.8-2022.1

- 通过大数据智能化产业研究技术路线与传统产业经济学相结合，构建生物医药产业链知识图谱，从产业、企业、创新技术等以及地域分布等角度对长三角生物医药产业链展开分析。
- 2021 年 5 月上海市市委领导视察某区经济监管平台建设情况，对平台建设成果给予充分肯定，并指出要进一步打造上海升级版的“城市大脑”。

科技舆情推荐系统中基于 Albert 的政策业务分类

2020.8-2021.1

- 对政府发布的政策文本进行爬取与清洗，根据平安科技公司事业群业务类型进行多标签分类。使用 TF-IDF、Word2vec、CNN 等传统 ML/DL 文本分类方法，结合 BERT、RoBERTa、ZEN 等大型预训练模型进行微调。
- 目前在产品平台上取得了较好的效果并投入实际使用，项目部分开源代码可见 [Github](#)。

基于 BERT 的上市公司国民经济行业分类

2019.5-2019.9

- 根据 2017 年《国民经济行业分类》设定的 21 个门类作为分类类别，从公司简介业务介绍文本入手，采用 TF-IDF+ML、Word2vec+DL 等方法，最终在 Bert 模型上进行微调的文本分类取得了较好的效果。

《慈善之困——社交媒体下如何拯救众筹善意》

2017.2-2017.5

- 在 2017 年全国大学生市场调查与分析大赛担任项目负责人，全程参与问卷编写、汇总统计，数据建模与论文撰写，并获得全国市场调查与分析技能证书(CRA)；
- 获得国家二等奖、省级一等奖，并为湖南省赛区唯一晋级决赛的队伍。

《Optimization of The Direct Metro Lines》

2016.11-2016.12

- 在 2016 年“认证杯”数学建模国际赛中作为项目负责人承担数学建模、论文写作并进行基于 R 的数据多元统计分析；
- 获得全国 H 奖。

《农村垃圾现状及影响因素研究》

2016.8-2016.9

- 在 2016 年“农业发展银行杯”暑期调研大赛中作为团队数据分析师，主要担任问卷编写、数据整理分析与论文实证分析部分的任务；
- 获得国家一等奖。

技能及其他

- **英语:** 雅思 7.0(R/L/S/W:7.5/7.5/6.5/6.0)
- **IT:** 擅长使用 Python 软件及 Keras、Pytorch、TensorFlow 等深度学习框架，具备 R、MATLAB、SPSS 编程能力，熟练使用 Linux 环境、MySQL 数据库以及 LaTeX 编辑工具；计算机二级证书（C 语言）
- **金融财会:** 证券从业资格考试；课余时间修读 ACCA F1-3，熟悉管理会计知识
- **其他:** 国家发明专利，普通话证书（一级乙等），C1 驾驶证，萨克斯七级，篮球院队队长及女篮教练