

# 王子源

联系方式: (+86)188-1787-6918

电子邮箱: [wangziyuan@163.sufe.edu.cn](mailto: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/机器学习/量化分析

个人网站: [HomePage](#); [in LinkedIn](#); [Google Scholar\(300+引用\)](#); [Github\(1.2K+Star\)](#); [CSDN](#)



## 教育背景

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 2018.9-2024.6

研究方向: 自然语言处理、深度学习、量化分析

指导教师: [黄海量教授](#)

博士论文: 《文本表示的改进方法及其在金融领域的应用》

GPA: 3.58/4.0 学院排名: top10%

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、语义相似度计算、设计科学

指导教师: [樊卫国教授](#)

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%

指导教师: [马勇教授](#)

在校期间所获荣誉: 多次获得国家奖学金、研究生一等奖学金、二等奖学金、校优秀学生、湖南省优秀毕业生、花旗银行未来精英奖学金、校一等奖学金、校三好学生标兵、校优秀班干部、校百优党员等荣誉。

在校期间所担任职务: 担任年级长、研究生学生会副会长、班长、党支部组织委员等。

## 实习经历

浙商基金, 智能权益投资部量化组, 量化分析研究员

2024.2 至今

- 基于文本信息的图神经网络选股。根据文本内容探究 A 股股票间相关关系, 通过文本分析揭示潜在的市场动态。研究了 FinBERT、ERNIR3.0、DMETA 等多个 SOTA 的文本表示模型, 结合 Whitening 等技术改善语言模型的各向异性表示问题。基于 GCN、GAT 等图神经网络方法建立与股价之间的联系的投资策略, 基于量价特征在中证全指 2021-2023 区间盘样本外测试 IC 达到 14.44%、RankIC13.94%, ICIR、RankICIR 超 1.2 的效果, 效果优于基线模型及图信息在相同模型表现水平。

Millennium Management(千禧年对冲基金), CRTC, Quantitative Analyst

2023.7-2023.12

- 基于 NLP 在 Earnings Conference Call 情感因子的量化投资应用。通过对美股上市公司盈利电话会议进行文本分析, 基于 FinBERT, FLANG-BERT, RoBERTa 及 FinGPT 等大语言模型, 探索语言模型进行推断、微调以及上游预训练+下游微调等多种深度学习训练方式, 并对 LLMs 的不同 Prompts 情感分析推断进行对比分析。利用以上多种模型预测集成 Ensemble 因子; 基于学术界和业界研究 SOTA 成果, 建立基于行为金融学的 NLP 特征因子; 结合上述多种因子信号搭建市场投资情绪挖掘的 Pipeline 框架。
- 基于组内回测框架实测效果达到夏普率 3.03, 年化收益率 2.98%, 最大回撤 1.29%, 日均换手率 6.26%。相同策略交易条件下收益表现超过 Amenity, Alexandria 等多个数据供应商提供的因子表现水平。并对 ChatGPT 等大语言模型在量化投资中的应用有着较为广泛的研究了解, 实践应用及前沿的视角思路。

## 科研及项目经历

How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection 2022.12-2023.7

- 共同第一作者。我们作为全球最早发起 ChatGPT 对比与检测项目的团队, 通过开放域问答任务下收集人类-ChatGPT 对比数据并发布第一个 ChatGPT 语料库 HC3, 对语料库文本进行图灵测试、文本统计、

情感分析、语言学分析等。结合实例总结 ChatGPT 文本范式及与人类回答的不同点，利用深度学习、机器学习等方法开发了基于单文本、问答对等多个 ChatGPT 内容分类检测器，取得了显著的检测效果。

- 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 20 万次，开源模型月均下载量超 300000+，数据集月均下载量 30000+，Github stars 1200+，论文引用 300+。目前已被国际顶级计算机会议 [LLM@IJCAI](#)(CCF-A, CORE-A\*)录用，欢迎访问我们的 [Demo](#) 和 [Paper](#)。

#### **Investigating Effectiveness of Whitening Post-processing on Modifying LLMs 2022.9-2023.6**

- **第一作者。**探究了 PCA, ZCA 等向量矩阵白化方法对大语言模型的文本表示的去相关化作用，从而规范化语言模型的文本表示学习能力，以求解决传统语言模型忽略了余弦相似度基底正交的假设，及在不同 NLP 任务中近 20 个数据集的表现，在 Bert, DistilBert, GPT2 等多个模型上均取得了显著效果。
- 被 35th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2023, CCF-C, CORE-B, , 接收率=21%)接收为 Full Paper，国家自然科学基金项目（编号：72271151），上海财经大学研究生创新基金项目（编号：CXJJ-2021-052）。欢迎访问我们的 [Paper](#)。

#### **IDEA: Interactive Double Attentions from Label Embedding for Text Classification 2022.6-2022.9**

- **第一作者。**提出了一种利用有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，提出了一种新颖的融合文本和标签信息的 Siamese BERT 模型结构，并在 AGNews, DBPedia, Yahoo Answer, Yelp 等多个公开数据集上取得了显著的提升效果。
- 发表于 34th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2022, CCF-C, CORE-B, 接收率=15.7%)，国家自然科学基金项目（编号：72271151）。欢迎访问我们的 [Paper](#)。

#### **New Hybrid Model of Crowdfunding Project: A Perspective of Prospect Backers 2022.8-2024.3**

- **通讯作者。**本工作通过引入前景理论（PT）来捕捉众筹中投资者支持行为的异质性，基于多种机器学习、深度学习模型根据众筹绩效对出资人行为相关决策进行预测，根据评估的前景效用进行支持决策。
- 已被 ESWA(Expert Systems With Applications, JCR Q1, 中科院 SCI 期刊 I 区 Top 期刊) 接收。

#### **网页开发工具: SUFE-CS-CONF-DDL 2022.1-2022.3**

- **项目发起人/负责人。**基于 Vue-CLI 搭建上财 Tenure Track/CCF 双重检索计算机会议 DDL 系统工具。
- 得到本校师生一致好评，欢迎访问可视化网站 [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 2021.2-2022.1**

- **算法工程师。**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。
- 上海市统计局重点研究项目。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的政策业务分类(上海帆讯信息&平安科技) 2020.8-2021.1**

- **文本分类组负责人、算法工程师。**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。

#### **技能及其他**

- **学术审稿人:** Applied Intelligence, International Journal of Computational Science and Engineering
- **IT:** 擅长使用 Python 软件及 Pytorch、Keras、TensorFlow 等主流深度学习框架，具备 R、MATLAB、SPSS、Vue-CLI 编程能力，熟练使用 Git, Linux、MySQL 以及 LaTeX 等工具；计算机二级证书（C）
- **金融财会:** 证券从业资格；课余时间修读 ACCA F1-3，熟悉管理会计知识
- **语言:** 普通话证书（一级乙等），雅思 7.0(R/L/S/W:7.5/7.5/6.5/6.0)
- **其他:** 国家发明专利，C1 驾驶证，萨克斯七级，篮球队队长及女篮教练