

WANG ZIYUAN

Address: 777 Guoding Road, Yangpu District, Shanghai 200433

Tel: (+86) 188-1787-6918; **E-Mail:** wangziyuan@163.sufe.edu.cn

Date of Birth: April, 1996; **Date of Graduation:** June, 2024; **Gender:** Male; **Nationality:** China

Research Area: Natural Language Processing, Deep learning, Quantitative Analysis, Design Science

My Website: [Home Page](#); [in LinkedIn](#); [Google Scholar \(300+Citation\)](#); [Github \(1.2K+Star\)](#); [CSDN](#)

Education

- **School of Information Management and Engineering, Shanghai University of Finance and Economics**
2018.9-2024.6, PhD Candidate, majoring in Management Science and Engineering.
PhD Thesis: An Improved Approach to Text Representation and its Application in Finance
GPA: 3.6/4.0 **Rank:** Top10%; **Supervisor:** [Professor Hailiang Huang](#)
- **Tippie College of Business, University of Iowa**
2022.4-2023.4, Visiting Scholar, majoring in Business Analytics; **Supervisor:** [Professor Weiguo Patrick Fan](#)
- **College of Finance and Statistics, Hunan University**
2014.9-2018.6, Bachelor of Science in Statistics, majoring in Statistics.
GPA: 3.8/4.5 **Rank:** Top5% **Supervisor:** [Professor Yong Ma](#)

Main Awards at school: Won National Scholarship, First Prize Graduates Scholarship, Outstanding Class leaders of School, Outstanding Students of School, Outstanding Graduate of Hunan Province, Citibank "Future Elite" Scholarship, School of Merit student, Outstanding Class Leader, First Prize Undergraduates Scholarship many times.

Positions held at school: Grade Leader, Vice President of Graduate Student Union, Monitor, etc.

Working Experience

ZheShang Fund, Quantitative Researcher

2024.3-Now

- **Stock selection by Graph neural network based on text information.** Explore the correlation between A-shares according to the text content, and reveal the potential market dynamics through text analysis. The graph information network structure is constructed on the all-index stock pool by using the text representation models of SOTA such as FinBERT, ERNIR, RoBERTa and DMETA, and the factors are mined based on basic volume and price factors combined with the graph neural network methods such as GCN and GAT for stock selection. In the total index of China Securities from 2021 to 2023, the out-of-sample IC reached 14.44%, RankIC 15.04%, the annualized RankICIR 5.3 effect, the annualized excess return of 39.18%, the information ratio of 3.29, the effect is better than the baseline model and other graph information under same model performance.

Millennium Management, CRTIC, Junior Quantitative Researcher

2023.7-2023.12

- **A quantitative investment application based on NLP in Earnings Conference Call sentiment factors.** Through text analysis of earnings conference calls of listed companies in the US stock market, based on FinBERT, FLANG-BERT, RoBERTa and FinGPT and other large language models, explore a variety of deep learning training methods such as language model inference, fine-tuning and upstream pre-training + downstream fine-tuning. Moreover, the sentiment analysis and inference of different Prompts in LLMs are compared and analyzed. The Ensemble factor is predicted by using the above models; Based on the research results of SOTA in academia and industry, the NLP feature factors based on behavioral finance are established. Combined with the above factor signals, we built Pipeline framework for market investment sentiment mining.
- Achieved an annualized return of 2.98%, a Sharpe ratio of 3.03, MaxDD 1.29% and daily turnover 6.26% in backtesting within the team's framework, outperforming the returns of factors provided by data vendors such as Amenity and Alexandria under the same strategy trading conditions. Possess extensive knowledge and practical application insights into the use of large language models like ChatGPT in quantitative investing.

Publications & Research Project

How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection 2022.12-2023.07
Accepted by Large Language Models @ International Joint Conference on Artificial Intelligence (LLM@ IJCAI 2023, CCF-C, CORE-B)

- **Co-First Author.** As part of one of the first teams globally to initiate a ChatGPT comparison and detection project, we collected human-ChatGPT comparative data in an open-domain Q&A task and published the first ChatGPT Corpus, HC3 (Human ChatGPT Comparison Corpus). We conducted Turing tests, textual statistics, and linguistic analysis on the corpus. We summarized the textual paradigms of ChatGPT and its differences

from human responses, and developed a series of ChatGPT classification detectors based on single texts and Q&A pairs using deep learning and machine learning methods, achieving significant detection results.

- Pioneered in the academic and industrial communities by open-sourcing the comparative dataset and detector models; our detector demo has received over 200,000 global visits, with the open-source model averaging 300,000+ monthly downloads, the dataset averaging 30,000+ monthly downloads, and acquiring over 1200+ Github stars and 300+ paper citations. The work has been accepted by the international top-tier computer conference [LLM@IJCAI](#)(CCF-A, CORE-A*). Please feel free to visit our [Demo](#) and [Paper](#).

New Hybrid Model of Crowdfunding Project: A Perspective of Prospect Backers 2022.08-2023.02

- **Corresponding Author.** In this work, the prospect theory (PT) is introduced to capture the heterogeneity of investors' support behavior in crowdfunding. The empirical results show that there is a significant correlation between behavioral factors and crowdfunding performance, and the AUC of classification increases by 7.39% on average.
- **Accepted** by ESWA (Expert Systems With Applications), JCR Q1 and SCI I zone top journal. [Paper Url](#).

Investigating Effectiveness of Whitening Post-processing on Modifying LLMs 2022.09-2023.06

Published in 35th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2023, CCF-C, CORE-B), acceptance rate = 21%

- **First Author.** Investigated the decorrelation effects of vector matrix whitening methods such as PCA and ZCA on the textual representations of large language models. This research aimed to standardize the text representation learning ability of language models, addressing the traditional models' oversight of the assumption of orthogonal bases in cosine similarity. The study significantly improved performance across nearly 20 datasets in different NLP tasks, applying to models including Bert, DistilBert, and GPT-2.
- Our work has been accepted as a Full Paper at the 35th IEEE ICTAI 2023. This research was part of the National Natural Science Foundation of China project (Grant No.: 72271151) and SUFE Graduate Innovation Fund project (Project No.: CXJJ-2021-052). Welcome to visit [Paper](#).

IDEA: Interactive Double Attentions from Label Embedding for Text Classification 2022.06-2022.09

Published in IEEE ICTAI 2022, CCF-C, CORE-B, acceptance rate = 15.7%

- **First Author.** Proposed a succinct method to enhance BERT's performance in text classification by utilizing label embedding techniques in supervised learning. Developed a novel model structure that integrates text and label information on top of the existing BERT model, achieving significant improvements on public datasets.
- Published at the 34th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2022, CCF-C, CORE-B, acceptance rate = 15.7%). This research was part of the National Natural Science Foundation of China project (Grant No.: 72271151). Welcome to visit [Paper](#).

Web Tool: SUFE-CS-CONF-DDL 2022.01-2022.03

- **Project initiator/project leader.** We built a countdown system tool for the computer conference of Shanghai University of Finance and Economics based on Vue-Cli, providing dual retrieval of tenure track tier/CCF level.
- Please feel free to visit our visualization website [SUFE-CS-CONF-DDL](#), and the open-source code of the project is available on [Github](#).

Analysis of Shanghai's Biopharmaceutical Industry Chain Based on Big Data and Comparative Study of the Yangtze River Delta (with Suzhou Bank) 2021.02-2022.01

- **Algorithm Engineer.** Combining big data intelligent industrial research technology with traditional industrial economics, we constructed a knowledge graph of the biopharmaceutical industry chain and conducted an analysis of the biopharmaceutical industry chain in the Yangtze River Delta.
- In May 2021, the leaders of the Shanghai Municipal Party Committee visited an economic regulatory platform in a certain district and fully recognized the achievements of the platform construction, pointing out that further efforts should be made to build a "city brain" upgraded version in Shanghai.

Skills

- **Academic Reviewer:** Applied Intelligence, International Journal of Computational Science and Engineering,
- **IT:** Python (Pytorch, Tensorflow, Keras), R, SPSS, MATLAB, MySQL, Vue-Cli; LaTeX; Microsoft Office
- **ACCA:** Association of Chartered Certified Accountants, F1-3
- **Language:** IELTS: 7.0(R/L/S/W:7.5/7.5/6.5/6.0); Mandarin Certificate (Level 1 B)
- **Others:** National Invention Patent, C-1 Driver License, Saxophone Level 7