Zhaojun Ding

zhaojun.ding@uga.edu

Education

University of Georgia (UGA) *Ph.D. in Computer Science*   Aug 2024 - Present
Advisor: Assistant Professor Ninghao Liu (School of Computing, UGA)
- Research Interests: Large Language Model, Data Mining, Trustworthy AI

University of Wisconsin–Madison (UW-Madison) *M.A. in Mathematics   Cumulative GPA: 3.636/4* Sep 2020 - May 2022
Advisor: Associate Professor Hao Shen (Department of Mathematics, UW-Madison)
Supervisor: Professor Tonghai Yang (Director of VISP, Former Chair, Department of Mathematics, UW-Madison)
- Relevant Courses: Machine Learning, Math Methods in Data Science, Methods of Applied Math, Functional Analysis, Modern Algebra, Brownian Motion and Stochastic Calculus

University of Science and Technology of China (USTC) *B.S. in Mathematics & Applied Mathematics* Sep 2017 - Jun 2021
Advisor: Professor Congwen Liu (School of Mathematics, USTC)
- Relevant Courses: Computer Programming, Data Structures and Database, Computer Graphics, Operations Research, Application of Microcontroller, Time Series Analysis, Multivariate Analysis, Regression Analysis, Financial Derivatives

Work Experience and Internships

Institute of Data Space, Beijing, CN   *Natural Language Processing Engineer*   Jun 2023 - Jun 2024
Supervisor: Professor Fuli Feng (School of Information Science and Technology, School of Data Science, USTC) (Chief Scientist, Institute of Data Space)
- Applied cutting-edge large language models including GPT-3.5, LLaMA2, ChatGLM2, Qwen, and Baichuan to identify user attitudes towards concerned topics based on social media posts
- Employed advanced techniques such as k-NN Prompting, Self-Adaptive In-Context Learning, Automatic Prompt Optimization & Engineering, Prompt Ensemble, Chain of Thought to improve the performance of attitude recognition
- Improved persona-specific conversation ability of chatbot using Supervised Fine-Tuning, specifically LoRA Fine-Tuning
- Assisted text generations by referencing external knowledge using techniques such as Retrieval-Augmented Generation
- Combined regular expression-based filtering with large language model-based methods to increase both the coverage and accuracy of user feature extraction by 15% over LLM-only extractions
- Utilized vLLM, FlashAttention to accelerate model inference of the chatbot
- Packaged algorithms and their runtime environments, created APIs, and deployed them on company's servers

Business Big Data, Chengdu, CN   *Modelling Center Intern*   Jun 2020 - Sep 2020
- Constructed multivariable models between valuation and patent information of companies from Chinese "STAR Market"
- Extracted main patent features using Principal Component Analysis, and clustered companies into different types

Guolian Futures, Zhengzhou, CN   *Business Department Intern*   Nov 2019 - Feb 2020
- Analyzed futures trading transactions by collecting historical data and developing time series models
- Developed quantitative futures trading strategies of futures and continuously iterated for optimization

Research Experience

Titanic Survival Prediction Using Classification and Regression Machine Learning Models   Nov 2022 - Jan 2023
- Performed Exploratory Data Analysis by drawing histograms, box plots, correlation heat maps, etc.
- Applied random forest, SVM, k-NN, linear regression, logistic regression, etc. models and compared their performances

Multimodal Clustering of Amazon Review Data based on NLP Methods   Sep 2021 - Jan 2022
Advisor: Assistant Professor Frederic Sala (Department of Computer Sciences, UW-Madison)
- Embedded review text into sentence vectors using Word2Vec, S-BERT, and tuned (on Amazon reviews) S-BERT models
- Clustered reviews by their sentence vectors and 5-star ratings using k-medoid and k-means algorithms
- Evaluated the result of embedding and clustering by measuring the similarity to manual clustering using the Rand Index

Player Refund Model of Word Adventure Games and its Applications   Sep 2020 - Jun 2021
Chinese National Innovation and Entrepreneurship Program for University Students
Advisor: Associate Professor Yongjun Li (School of Management, USTC)
- Modeled single player's refund behavior by describing in-game demand and supply curves
- Developed methodology on how to determine parameters in refund model using sentiment analysis or players' game data
- Provided advice on how to increase profit separately for game developers and sellers by estimating players' refund time

Subway Passenger Flow Prediction based on Time Series Data   Feb 2020 - Jun 2020
Advisor: Associate Professor Yu Chen (Department of Statistics and Finance, School of Management, USTC)
- Applied GARCH model and ARIMA model to fit and predict passenger flow data of two different metro lines
- Compared and assessed the fitting and prediction by analyzing residual, variance, and other statistical variables

Leadership

Department of Mathematics, UW-Madison   *VISP-MA Program Student Coordinator*   Oct 2021 - May 2022
- Designed and hosted events for Math department. Events included seminars, VISP-MA program presentations, trips, etc.
- Built alumni database of the VISP-MA program and invited alumni to provide various support to our current students

Skills

- Programming Skills: Python (Machine Learning, PyTorch, scikit-learn, data analysis), C++ (Algorithms and Data Structures, LeetCode Contest Rating Global Top 2%), Docker, Linux Command Line, API, SQL, R (Statistical Analysis)

Awards

- Chinese National Innovation and Entrepreneurship Program for University Students   Sep 2020
- Scholarship for annual outstanding students, USTC   Dec 2020 & Dec 2018
- Scholarship for excellent fresh students, USTC   Dec 2017
- Third prize of the Chinese Mathematics Competitions (CMC), Anhui Province   Nov 2019 & Nov 2018
- Outstanding Volunteer Prize and Touching Achievements Prize   Oct 2018