CONG CHENG

+1(706) 461-2259 ◇ Cong.Cheng@uga.edu
linkedin.com/in/cheng-cong ◇ oniontimes.github.io

EDUCATION

Ph.D. in Statistics, University of Georgia (UGA) 2021 - 2026
Areas of Expertise: Change Point Detection, High Dimension Statistics 3.9/4.0

B.S. in Mathematics, University of Science and Technology of China (USTC) 2016 – 2020
Honors: Graduation with Highest Distinction, Leadership Scholarship

EXPERIENCE

Sr. Data Scientist Intern June 2024 – Aug 2024
Walmart Inc. *Sunnyvale. CA*

- Constructed an AI ChatBot for interpreting large databases related to supply chain information, enhancing the efficiency of data query processes and reducing the learning curve for database understanding.
- Combined few-shot learning and prompt engineering to train LLM. Applied few-shot selection methods to mitigate concerns for bias as the number of few-shot examples and tables increased.
- Enhanced the ChatBot's capability to join across more than 17 tables, enabling the model to make a summary about key information such as package delay reasons and created a user-friendly UI.
- Implemented and compared several large language models such as Llama3, Llama3-SQL, Gemini and GPT-4 for text-to-SQL tasks on three self-created dataset levels, achieving significant improvements in query accuracy.

Quantitative Research Intern May 2023 – Aug 2023
Qilin Investment Management (AUM 30 billion) *Shanghai, China*

- Constructed value and momentum factors (three factors in the top 30 among the company) for China's A-share market using high-frequency data, including snapshot, order, and trade data. Collaborated with colleagues to create a handbook for DolphinDB database.
- Implemented ML models such as CNN and lightGBM models in Python to accurately forecast share returns within a 30-minute timeframe, resulting in a significant 25% increase in the information coefficient (IC).
- Developed a feature-selection framework utilizing integrated gradient analysis, successfully reducing the number of features from 152 to 92 while maintaining optimal performance.

PUBLICATION

1. Large precision matrix estimation with unknown group structure. *Submitted to the journal*, 2024
2. A homogeneity method for calibrating item response theory models with limited samples and sparse response data. *Submitted to the journal*, 2024

PROJECT

Character level CNN for text classification May 2022
Employed Python to categorize 65,000 Chinese articles into 14 distinct categories using the Character-level CNN algorithm. Abstained from utilizing the conventional Chinese-to-English conversion method and instead directly crafted the Chinese alphabet when converting the text into vectors. The ultimate accuracy rates consistently hovered around 90%.

SKILL

- **Programming Languages:** R, Python(Pandas, NumPy, Scikit-learn, Langchain), BigQuery, C/C++, Git
- **Frameworks:** Jupyter Notebook, Linux