MINGYI HUANG

Beijing, China | +86-18210107390 | huangmy2024@gmail.com | https://github.com/hmy02

RESEARCH INTEREST

Inspired by the intricacies of the human brain as elucidated by biologists, my ultimate aspiration is to harness those insights to develop a machine-learning model emulating human cognition. Beyond that, I also hold a fervent passion for algorithmic innovation and its tangible applications.

EDUCATION

Beijing University of Technology, Beijing, China Expected June 2024

BS in Statistics, Minor in Computer Science and Technology, GPA:3.93/4.0 (Rank 2/26)

- *Selected Coursework:* Machine Learning, Data Structure and Algorithm, Principles of Database System, Data Mining, Time Series Analysis, Principles of Operating System, Computer Network, Principles of Computer Organization
- *Awards:* CUMCM 2021 – Provincial 2nd Price
  BJUT Scholarship, 2022, 2023
  BJUT, Faculty of Science, Scholarship, 2021

TECHNICAL SKILLS

- *Computer Languages:* Python, Java, C, C++, R, SPSS, MATLAB, SQL, HTML
- *Tools:* PyTorch, TensorFlow, Git, LaTeX

PUBLICATION

[1] Wang, L., Xie, J., Zhang, X., **Huang, M.**, Su, H., & Zhu, J. (2023). Hierarchical Decomposition of Prompt-Based Continual Learning: Rethinking Obscured Sub-optimality. *arXiv preprint arXiv:2310.07234.*

[2] **Huang, M.**, Wu, L., Wang, Y. (planned in Nov. 2023). Generalized OLS Calibration for Expensive Computer Models

WORK EXPERIENCE

JD, *Data Analyst Intern, Research Institute for Consumption and Industrial*, Beijing, China June 2022 – Aug. 2023

- Qixi (Chinese Valentine's Day) Customer data
  - Coded a web crawler to extract post titles and corresponding gifting recommendations posted on social media
  - Screened, compared, and analyzed different consumption characteristics of normal and long-distance relationships
  - Created a consumption trend chart to visualize the gift-purchasing dates among various age groups
- Treadmill, fitness equipment, and fitness industry market data
  - Extracted, organized, and provided reference market data to a treadmill brand client
  - Applied long short-term memory (LSTM) to predict fitness consumer market trends
- Home appliances replacement policy in the countryside
  - Collected home appliances selling data under this replacement policy in the past three years and 2008
  - Compared the data of 2008 with gray forecast, ARIMA, LSTM, and other methods to predict the main selling points and consumption trends of home appliances going to the countryside in 2022

GuoTai Asset Management, *Quant Intern, Active management department*, Beijing, China June 2023 – Aug. 2023

- Candlestick charts and convolutional neural network (CNN) for stock price prediction
  - Built a CNN model based on deep learning combining LSTM and convolutional filtering, to achieve time series analysis and price trend prediction
  - Improved the model through hyperparameter optimization and data normalization
  - Incorporated embedded coding, attention mechanism layer, and dummy variables into our Multi-task learning framework to enhance information coefficient (IC)
  - Based on the model results, generated alpha factors, conducted factor back testing, and calculated metrics such as annualized returns, which reached up to 28%
- Quantification of institutional research
  - Mined analytical research reports, constructed attribute sets, used FinBERT pre-trained model to encode the text and generated eigenvectors for cross-validation training of excess information in XGBoost
  - Studied the impact of different types of financial institution survey (type of company surveyed, form of survey, etc.), and predicted trends in the number of surveys based on AIRMA to further optimize the stock screening strategy

RESEARCH EXPERIENCE

BJUT Institute of Applied Probability and Statistics, *Undergraduate Researcher* Oct. 2022 – Now

- Generalized Ordinary Least Squares Calibration (GOLS Calibration)
  - Used Gaussian Processes as a surrogate model to mimic the outputs of complex computer experiments.
  - Derived the formula to prove that an additional variance term in the loss function can reduce fluctuation while enhancing accuracy
  - Reproduced other methods, such as KO and GALS calibrations, using data from the MATLAB FEM toolbox
  - Apply GOLS on stochastic computer simulation (ongoing project)

Tsinghua Statistical Artificial Intelligence & Learning Group, *Undergraduate Researcher* Mar.2023 – Aug. 2023

- Optimizing Task-identity inference and with-in task prediction for better Continuous Learning models
  - Compiled core research papers in Continual Learning, made an open-source paper list, and published it on GitHub
  - Benchmarked our HiDe-Prompt method against 4 other advanced prompt-based continual learning methods
  - Conducted parameter fine-tuning, recorded experiment results, visualized results, and improved code execution

PROJECT EXPERIENCE

Forecasting the Difficulty of Words in the Wordle Game, *Contest Team Leader* Feb. 2023

- Predicted player number of the Wordle game using selected ARIMA model improved by Markov Chain
- Fitted a Gaussian Regression model to the distribution of user attempt frequencies for a specific word
- Categorized words based on their difficulties using K-means and validated the clustering result using AdaBoost

Optimizing Raw Material Supply Chains for Manufacturers, *Contest Team Leader* Sept. 2022

- Applied time-series analysis to predict the 6th-year material supply accuracy using data from the past 5 years
- Designed a rating system for suppliers' supply volume and accuracy to find the key suppliers
- Utilized MINLP to optimize the supply chain by coming up with a new supply chain model with new transportation, storage, and ordering strategies

Water Sharing Strategy Solving Severe Drought in Five US States, *Contest Team Leader* Feb. 2022

- Obtained the water storage at minimum water level by fitting historical water level and water storage data
- Developed a model considering the water needs of 5 states from two lakes
- Maximized the economic benefit considering state GDP, industry water usage, historical stability of water usage

Chemical Composition Identification of Ancient Glass Relics, *Contest Team Leader* Sept. 2021

- Made descriptive statistics based on textures, category, and color to determine their information gain on relic weathering
- Conducted normality tests of chemical composition for relics to predict their composition before weathering
- Performed dimension reduction from 14 chemical composition attributes to 3 attributes using random forests
- Used K-means to classify relics based on their composition and determined the chemical relationships and differences of various types of relics with the correlation coefficient matrix method

Chatroom Using Python Socket, *Class Project* Mar. 2022 – June 2022

- Developed a chatroom app using Python, from account registration to database maintenance, using socket TCP/IP
- Created a socket binding with local IP and a thread to handle incoming messages whenever a new user logs in

LEADERSHIP EXPERIENCE

BJUT Faculty of Science Student Union Sports Department, *Member*, Beijing, China Sept. 2020 – July 2023

- Led a team of 4 to organize and coordinate college-level basketball tournaments
- Wrote articles with 700 views to promote college sports activities for students and faculty