Wayne Shukai Wang

332-260-8346 | sw3715@columbia.edu| New York, NY 10025 | www.linkedin.com/in/shukaiwang37

Education

Columbia University New York, NY
MS in Data Science, GPA 3.83 Expected 2023/12
Courses: Algorithms; Data Analysis using R; Machine Learning; Statistical Inference & Modeling; Database

Peking University Beijing, China
BS in Data Science, GPA 3.4 2018/9 – 2022/6
Honors: Research Innovation Award, 2nd Prize in 13th Chinese Mathematics Competitions
Courses: C/C++; Java; Python; JavaScript; Database; Algorithm; Probability & Statistics; Computer Systems; Graph Theory; ODE; Time Series Analysis; Parallel and Distributed Computing; Game Theory; NLP; Machine Learning; Numerical Analysis

Internship and Projects

Zhengren Quantitative Investment Management Co. Beijing, China
Algorithm Developer Intern 2021/11 – 2022/6

- Used time series analysis, data weighting, data pruning, deep learning and reinforcement learning to build the latest resampling model which analyzed the stock time series and solved the concept drift problem of stock data. Increased IC and IR by ~4%, from the initial model.
- Wrote, maintained and modified code of existing investing strategies and deployed execution platform based on Tensorflow frame. Conducted research on new systematic investing strategies and experimented with LightGBM, XGboost and Transformer on stock return and volatility forecasting.
- Used Apache Hive to manage and maintain investment data.
- Used C++ and OOD method to transform data and perform automatic data quality check; Used Python and Docker to build automatic investment strategy efficacy evaluation and visualization scripts.

Institute of Network Computing and Information Systems Beijing, China
Associated with Peking University Undergraduate Research Intern, advised by Prof. Ming Zhang 2019/3 – 2022/6

- Conducted research on GNN(Graph Neural Networks) and NLP models, collaborated with doctoral students in research projects.
- Publication: Liu, Zequn & Wang, Shukai & Gu, Yiyang & Zhang, Ruiyi & Zhang, Ming & Wang, Sheng. Graphine: A Dataset for Graph-aware Terminology Definition Generation. The 2021 Conference on Empirical Methods in Natural Language Processing. arXiv preprint arXiv:2109.04018.
- Used python to process data by applying algorithms incl. hash tables, quicksort to reduce time complexity.
- Created a large-scale graph dataset with over 2 million terminology definition pairs and 227 biomedical subdisciplines.
- Helped proposed a graph-aware definition generation model integrating transformer with GNN.

A Pre-trained Model for Cellular Responses to Drug Perturbations Remote
Associated with Peking University Research Intern Advised by Prof. Sheng Wang (U of Washington Seattle) 2021/7 – 2021/11

- Proposed pre-training model combining BERT with cvae to predict single-cell RNA sequencing (scRNA-seq) data of unseen drugs. Reduced RMSE by 36% by using muti-task learning method.
- Processed scRNA-seq data for training, pre-trained model to learn transcriptional drug responses from bulk RNA sequencing; fine-tuned model to adapt to scRNA-seq
- Paper Under Review: Wang, Shukai & Tan, Yuhao & Zhang, Ming (2021). A Pre-trained Model for Cellular Responses to Drug Perturbations.

Wangxuan Institute of Computer Technology (WICT), Peking University Beijing, China
Associated with Peking University Undergraduate Research Intern advised by Prof. Yang Liu 2021/4 – 2021/9

- Developed Computer Vision and Graph Neural Network models using Pytorch frame to accurately describe images taken by those visually incapacitated, which lack photo shooting techniques such as framing, proper lightning.
- Improved the performance of existing best-performing GNN model, adding background nodes and HOI transformer decoder for human object interaction capturing. Increased BLEU-4 by 4% amongother improved metrics.
- Attended the 2021 CVPR Workshop Vizwiz Grand Challenge and received 3rd Place in Image Captioning group.

Skills

- Python, C/C++, Java, SQL, MATLAB, JavaScript, R, Excel, Latex
  Data Analysis, Machine Learning, Deep Learning, Reinforcement Learning, NLP, Computer Vision, GNN