Li Zhe

139-8020-6966 zheli20@fudan.edu.cn
Fudan University, 220 Handan Road, Yangpu District, Shanghai

educational background

Fudan University Sep 2020 - Jun 2023

Statistics Master School of Data Science Shanghai

- Research Areas: Distributed Algorithms, Spatial Econometrics, Complex Network Analysis
- Honors/Awards: Postgraduate Outstanding Student Scholarship Second Prize (2020-2021; 2021-2022); Sixth National Statistics Doctoral Academic Forum Thesis First Prize (Top5); Fudan University Big Data School Second Doctoral Student Forum Third Prize

Fudan University Sep 2016 - Jun 2020

Physics (Data Science and Big Data Technology) Bachelor School of Data Science Shanghai

- Honors/Awards: Undergraduate Excellence Scholarship Third Prize (2016-2019); Undergraduate Professional Scholarship (2017-2019); Graduate Scholarship (2019-2020)
- Core courses: Advanced Mathematics, Linear Algebra, Probability Theory and Mathematical Statistics, Statistical Computing, Machine Learning, Deep Learning and Neural Networks

RESEARCH EXPERIENCE

- Wu, S., **Li, Z.**, & Zhu, X. (2020). Distributed Community Detection for Large Scale Networks Using Stochastic Block Model. "Computational Statistic & Data Analysis", accepted, [joint first author]
- Ren, Y., **Li, Z.**, Zhu, X., Gao, Y., & Wang, H. (2022). Distributed Estimation and Inference for Spatial Autoregression Model with Large Scale Networks. "Journal of Econometrics", under revision, [joint first author]
- **Li, Z.**, Zhu, X., & Zou, C. (2022). Consistent Selection of the Number of Groups in Panel Models via Sample-Splitting. submitted to "Journal of Econometrics", [first author]
- Han, S., Zhou, J., Zhu, X., **Li, Z.**, Liu, J., Wang, H., & Gong, Y. (2022). High-Resolution Image Classification with Rich Text Information Based on Graph Convolution Neural Network. submitted to "IJCNN 2023"
- **Li, Z.**, Zhu, X., & Zou, C. Consistent Selection of the Number of Factors in Factor Model via Sample-Splitting (working paper).
- **Li, Z.**, Zhu, X., & Zou, C. Statistical Inference for the Number of Groups in Panel Models (working paper).

PROJECT EXPERIENCE

Audi and other car companies shake sound live commercial cooperation projects Mar 2021 - Dec 2021

- Through the sound of the user to watch the live historical barrage and private letter for user portrait and use logical regression modeling, to find high-value potential users

.

- Combining simple paradigms with traditional machine learning models (random forests and naive Bayes) to construct a simple private letter automatic question answering system

.

- Participate in the design of the real-time vehicle recognition model, the CNN model and CRAFT text detection model fusion training, to obtain a high-accuracy classification model

  for vehicle recognition tasks.

Joint fund distributed statistical basic algorithm implementation Oct 2020 - Mar 2021

- Based on the Spark framework, the Pyspark is used to implement distributed linear regression and distributed logical regression algorithms

  based on DLSA algorithms.
- Using LARS algorithm and BIC algorithm for DLSA algorithm regression coefficient variable selection
- Based on the laboratory server cluster platform, the use of Spark cluster for 36GB size of about 0.1 billion pieces of simulated data using DLSA algorithm, the algorithm effect is close to the traditional full data estimation algorithm, excellent

Beijing Wanshang Lianxin Technology Company-Take the money: Jun 2019 - Aug 2019

- Responsible for nearly 10,000 merchants flow transaction data cleaning and according to the cleaned flow transaction data to establish a merchant credit evaluation model, for the bank to issue loans to provide reference

.

- Responsible for database interface testing and report writing

RESEARCH EXPERIENCE

Huawei Financial Elite Challenge May 2020

- Using the traditional time series model (ARIMA) and machine learning model (LSTM) to analyze and predict the 2021 world regional 5G market size, won the first phase of the competition third prize (campus top four) and individual winning award (Top3)

Alibaba Cloud Smart Capacity Planning Competition

- Using the window sliding LSTM model to predict the future period of server CPU utilization, won the first place (work link)

PROFESSIONAL EXPERIENCE

Byte Dance Jul 2022 - Sep 2022

Algorithm engineer Data-Comments Shenzhen

- According to the product manager's requirements, for the business line of products (trembles, watermelon video, etc.) sort of intervention, the use of power, power, filtering keywords and other strategy-based methods to improve the product review area quality rate, increase the review area users stay time

.

- Using the Kafka framework, the shaking audio and video comment data is captured in real time, and descriptive analysis is used to find useful features for comment intervention

.

- Using wind control data, identify the brush like, brush reply behavior, and further analysis of such comments whether there is a pornographic diversion, by analyzing the characteristics of the comments, the development of the corresponding suppression strategy and model, the Top 1000 comments of the somatosensory exposure index from 7.8 percent to 3.7 percent

  .

Microsoft Research Asia May 2022 - Jul 2022

Algorithm engineer Software Analytics Group Beijing

- Using Kusto to grab the company's server operation log, the Microsoft cloud server machine operation data descriptive statistics and analysis, focusing on the analysis of machine operation law and characteristics

  .
- uses Kusto to capture the company's server running logs, and by analyzing the server CPU, memory and other data, it is found that the data has obvious periodicity. Use ARIMA model to model and predict the subsequent utilization rate, and realize the "peak elimination and valley filling"

  of resources.
- Use Kusto to capture the company's server running logs, through the running logs, analysis of different nodes of the operating mode, and the use of nodes in a certain period of time the mean and standard deviation of the resources of the node classification, by identifying low utilization nodes to achieve resource migration, improve resource utilization

  .

Professional skills

- Programming languages and software used: Python, R, Matlab, Shell, SQL, Java, Scala, LaTeX
- Deep learning frameworks: Pytorch, TensorFlow, DGL
- Distributed frameworks: Hadoop, Spark
- Foreign language ability: CET6, IELTS 6.0