XianZhu Shao

(+86)15655933280 | charon0328@163.com

EDUCATION

Tianjin University of Commerce, *College of Science*(GPA84.8/100) **Tianjin, China**
Bachelor of Science in Mathematics and Applied Mathematics *Sep. 2019– Jun. 2023*

- **Key Courses:** Data Mining, Optimization Methods, MATLAB, Regression Analysis.
- **Skills:** Python, Matlab, MySQL, COMSOL, Verilog HDL, FPGA.

RESEARCH INTERESTS

- Risk control and management, Data imbalance
- Financial time-series modeling and analysis, Financial deep learning
- Optimization theory, Machine learning, Deep learning architectures

RESEARCH EXPERIENCE

Machine Learning-Based Inclusive Credit Assessment Model for Farmers **Tianjin, China**
Team Leader, National Innovation Programme for University Students *Mar. 2021 – Mar. 2023*

- Aimed to develop a high-precision credit prediction model tailored for rural farmers. The goal was to break free from the current credit assessment system primarily reliant on collateral, credit history, and social guarantee relationships. The inherent limitations faced by rural farmers within this system constitute a key impediment to rural economic development.
- Reviewed over 100 articles on the application of machine learning models in the field of credit prediction. Programmed and implemented methods from the literature, including techniques for handling data imbalance issues and cost-sensitive learning methods.
- Conducted empirical research using the AdaBoost algorithm on rural farmer data provided by a commercial bank in China. Achieved an accuracy rate of over 95%. Ultimately, authored and published the paper titled "Based on Adaboost Algorithm for Loan Default Prediction Model" as the first author.
- With the deepening of the research, I improved the variant Adacost algorithm of the Adaboost algorithm and mathematically demonstrated that the improved algorithm's error upper bound is nearly zero. Eventually, I authored "A Dynamic Cost-Adjusted Adacost Model for Credit Prediction of Smallholder Farmers" as a co-first author and corresponding author. It has been submitted to the *International Journal of Forecasting* and is currently undergoing revisions based on reviewer feedback.

Machine Learning-Based New Energy Vehicle Sales Forecasting **Beijing, China**
Team Member, Chinese Academy of Sciences (CAS) Science and Technology Innovation Programme Jun. 2021 – Sep. 2022

- I developed a machine learning model aimed at improving existing research deficiencies in predicting sales of new energy vehicles and providing policy implementation recommendations for achieving global carbon neutrality goals. The model employed a fusion of Empirical Mode Decomposition (EEMD) and Deep Neural Networks (DNN) to obtain a high-precision price time series forecasting model, and a corresponding research report was authored.

Identification of Solid-State Diffusion Coefficients in Lithium-Ion Batteries **Shanghai, China**
Optimisation Algorithm Engineer, HaiFang Technology Co. *Jun. 2021 – Jun. 2022*

- I developed a parallelized improved simulated annealing algorithm that achieved an optimization efficiency improvement of over 20 times compared to the traditional simulated annealing algorithm while obtaining solutions of the same order of magnitude. Furthermore, under the same number of iterations, the algorithm I developed exhibited an average improvement in solution accuracy of seven orders of magnitude.
- I Improved the efficiency of the enhanced simulated annealing optimization for solid-phase coefficient identification through the use of random solution generation using distributions such as Levy flights and exponential distributions.
- I designed and developed an app using MATLAB, which comprises 6 functional pages. I integrated the improved Simulated annealing algorithm into the app. Additionally, I designed and implemented the software's database using MySQL to store and retrieve relevant data.

Papers & Publications

A Dynamic Cost-Adjusted Adacost Model for Credit Prediction of Smallholder Farmers **Under Review, Revise**
International Journal of Forecasting, Co-First authors&Corresponding author

- I proposed a dynamically adjusted cost-sensitive matrix, resulting in an exponential growth of misclassification costs for defaulting samples during the training process. As the model advances in its training, it acquires a more comprehensive understanding of the characteristics of defaulting samples, leading to a substantial reduction in the misclassification rate of defaulting samples.
- Through rigorous mathematical theory, I conducted a formal derivation and proof to establish the effectiveness of the exponentially growing cost-sensitive matrix during the training process. The model's training process is

characterized by an exponential rate of error reduction. Error analysis unveiled that the misclassification rate of the proposed model is confined to a range asymptotically approaching zero.

- I modified the AdaBoost algorithm code in scikit-learn to develop the Adjust-Adacost algorithm for this study. Using a real dataset from a commercial bank in China involving loans to famer, I compared the Adjust-Adacost model with the Adacost and AdaBoost models. Additionally, I compared it with other cost-sensitive algorithms like Cost-Sensitive Random Forest (CS-RF), Cost-Sensitive XGBoost (CS-XGBoost), Cost-Sensitive Support Vector Machine (CS-SVM), and Cost-Sensitive Decision Trees (CS-DT). The results showed that the Adjust-Adacost model significantly outperformed the others, with an AUC improvement of over 20% compared to Adacost and more than 10% compared to the other cost-sensitive algorithms.
- I conducted extensive experiments on international public datasets such as UCI and KEEL, and the results consistently demonstrated that the proposed Cost-Adjusted Adacost model outperformed other cost-sensitive machine learning models.

Based on Adaboost Algorithm for Loan Default Prediction Model **Published**
Journal of China Emergency Management Science, First authors *May. 2023*

- I constructed an Adaboost model and conducted empirical analysis on a loan dataset from a commercial bank in China. The empirical analysis confirmed the superior performance of the Adaboost model in default prediction compared to other ensemble learning methods.

Predicting China's Thermal Coal Price: Does Multivariate Decomposition-Integrated Forecasting Model with Window Rolling Work? **Under review**
Resources Policy, Fourth authors

- I propose rolling ICEEMDAN theory for time series prediction of the average coal price index in China's Bohai Rim region, which decomposes the time series data into multimodal modes and trains the prediction by neural network model for the sub-modal modes respectively.
- I constructed LSTM,TCN,CNN time-series combination prediction model, and the final error was reduced by more than 7% compared to the neural network model without time-series data decomposition.

WORK EXPERIENCE

AI Algorithm Engineer, *HaiFang Technology Co.* **Shanghai, China**
Responsible for algorithm development in AI-CFD and optimization algorithms *Apr. 2023 – Present*

- **LiFePO4 New Energy Vehicle Battery Life Prediction (SOH, RUL).**
  - I developed and trained an LSTM-based SOH time series prediction model capable of achieving seq2seq-style time series forecasting. This model demonstrated exceptional performance on two publicly available datasets (TRI and SNL), with maximum relative errors below 1% and average relative errors within 0.1%.
  - I addressed the challenge of significant SOH curve variations among identical batteries, which resulted in poor extrapolation performance in various operating conditions. After reviewing 50 related articles, I enhanced the existing model by incorporating IC curve features and applying transfer learning, resulting in a 50% reduction in maximum error.
  - Comparing the performance of SOH time series prediction achieved by the informer, GRU, and TCN models with the optimized LSTM model, the error of my optimized LSTM model is only 1/10 (or less than 1/10) of the errors produced by the three mentioned models.
- **Performance Optimization of Lithium Battery Electrode Infiltration Algorithms and Large-Scale Sparse Matrix Solving.**
  - About my company's in-house lithium battery electrode infiltration simulation algorithm, the algorithm's computational speed was improved by 60,000 times compared to ANSYS simulation software. Building upon this achievement, I further enhanced the computational speed by an additional threefold.
  - I employed GPU acceleration using tools like numba, cython, pypy, and cupy, as well as Just-In-Time (JIT) compilation techniques to optimize the computational aspects within the algorithm. Additionally, I refactored frequently used numpy functions to further enhance performance.
  - In the most performance-critical part, which involves solving large sparse matrices with dimensions approximately 1,000,000 x 1,000,000, I conducted extensive experimentation with direct solvers like Pardiso, Cupy, Scipy, Taichi, PETSc, AMGX, and ViennaCL, as well as iterative solvers (primarily the CG method). I compiled relevant CUDA computation packages and developed interface scripts to integrate them into Python code. As a result, I achieved an order-of-magnitude improvement in the efficiency of large matrix solving.
- **AI Optimization of Turbulence Models.**
  - I was responsible for further refining and optimizing the Bayes optimization algorithm under the Optuna framework. I also implemented the capability to resume and continue any incomplete optimization process from a prior session, even after terminating Python during any computation step in the Bayesian optimization. Additionally, I ensured the consistency of random sampling throughout the process.
  - Furthermore, I developed a multi-objective Bayesian optimization algorithm based on Optuna, as well as a simulated annealing algorithm utilizing Optuna. These enhancements led to a 32% reduction in error compared to the original approach.

EXTRACURRICULAR ACTIVITIES AND ACHIEVEMENTS

AI Second Class **Tianjin, China**
Founder and Leader *Sep. 2021 – Sep. 2022*

- Organized regular AI algorithm learning and sharing sessions for team members, including training for data competitions and mathematical modeling competitions held 1-2 times per week. Since its inception, I have guided team members, resulting in over 60 awards of third place or higher in various mathematical modeling and data competitions.

SCHOLARSHIPS & CERTIFICATES

SCHOLARSHIPS

- 2022-2023: Third Prize Scholarship.
- 2020-2021: National Encouragement Scholarship.
- 2020-2021: Second Prize Scholarship.
- 2019-2020: Second Prize Scholarship.

CERTIFICATES

- 2023: Mathematical Contest in Modelling **Honorable Mention**(Team Leader).
- 2023: **Outstanding Graduate Honour**.
- 2023: **Outstanding** National Innovation Project Completion(Team Leader).
- 2022: **Second Prize** in the 11th Asia Pacific University Mathematical Contest in Modeling(Team Leader).
- 2021: **Finalist Award** of the 2nd Tianjin "Ruisi Cup" Data Modeling Competition (Team Leader).
- 2021: **Third Prize** of the 9th "Teddy Cup" National Data Mining Challenge (Team Leader).
- 2021: **Third Prize** of the 9th "ShuWei Cup" National Mathematical Contest in Modeling(Team Leader).
- 2020: **University Merit Student**.