Yufan Zheng

Tel: (+86) 188-2293-5122 | E-mail: zhjpre@gmail.com | Personal Website: https://yufanzheng.github.io

EDUCATION

NanFang College of Sun Yat-Sen University, Electrical and Computer Engineering   Guangdong, China
Bachelor of Engineering   Sep 2018 - Jun 2022
- GPA 86.36/100 (3.53/5.0), Major in Computer Science and Technology.
- Core Courses: Advanced Mathematics, Linear Algebra, Data Structure and Algorithm, Discrete Mathematics, Data Principles and Applications.
- Scholarship: National Scholarship (2020), Second price Scholarship (2020), Third price Scholarship (2021).

AWARDS AND FUNDINGS

- Provincial third prize in 2021 China Undergraduate Mathematical Contest in Modeling (Oct 2021).
- Best volunteer in International Conference on Neural Computing for Advanced Applications 2021 (Aug 2021).
- Software copyright, "Data analysis and prediction system of epidemic infectious disease change based on machine learning" (2020).
- Software copyright, "Chinese film box office data analysis and prediction system based on machine learning" (2020).
- Participated in and completed Provincial College Students' Innovative Entrepreneurial Training Plan Program: "E-business strategy mining based on big data and machine learning" (2020-2021).
- Responsible for and accomplished Provincial College Students' Innovative Entrepreneurial Training Plan Program: "Box office forecasting based on big data and deep learning" (2019-2020).
- Merit award conference paper in 2019 IEEE International Symposium on Product Compliance Engineering-Asia (Oct 2019).

PUBLICATIONS

Peer-reviewed Journal Articles
- Zhan C, **Zheng Y**, Zhang H, Wen Q. Random-Forest-Bagging broad learning system with applications for COVID-19 pandemic[J]. IEEE Internet of Things Journal, 2021 (**SCI** Q1, 2020 IF: 9.936).
- Zhan C, **Zheng Y**, Lai Z, Hao T, Li B. Identifying epidemic spreading dynamics of COVID-19 by pseudocoevolutionary simulated annealing optimizers[J]. Neural Computing and Applications, 2020: 1-14 (**SCI** Q1, 2022 IF:5.606).

Conference Papers
- **Zheng Y**, Zhen Q, Tan M, Hu H, Zhan C. COVID-19's impact on the box office: machine learning and difference-in-difference[C]//2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, 2021 (**EI**).
- Li J, **Zheng Y**, Hu H, Lu J, Zhan C. Predicting video game sales based on machine learning and hybrid based feature selection[C]//2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, 2021 (**EI**).
- Lin J, Tan M, **Zheng Y**, Wu K, Zhan C. Detection capability prediction based on broad learning system during the COVID-19 pandemic[C]//2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, 2021 (**EI**).
- Wu S, Hu H, **Zheng Y**, Zhen Q, Zhang S, Zhan C. The impact of COVID-19 on online games: Machine learning and Difference-In-Difference. CCF Conference on Computer Supported Cooperative Work and Social Computing. Springer, Singapore, 2021 (**EI, accepted**).
- Wu S, **Zheng Y**, Lai Z, Wu F, Zhan C. Movie box office prediction based on ensemble learning. IEEE Symposium on Product Compliance Engineering-Asia (ISPCE-CN). IEEE, 2019 (**EI**).

CONFERENCE ACTIVITIES

- Volunteer in International Conference on Neural Computing for Advanced Applications 2021 (NCAA 2021) held in Guangzhou, China, assisted in the organization and conduct of academic conferences, led the group in the preparation and commissioning of conference site equipment, and the coordination of conference site services (Aug 2021).
- Participated in IEEE 22th International Conference on E-health Networking, Application & Services (HEALTHCOM 2020) held in Shenzhen, China (Dec 2020).
- Participated in 2019 IEEE International Symposium on Product Compliance Engineering-Asia held in Hong Kong, China, and delivered an oral presentation (Oct 2019).

NanFang College of Sun Yat-Sen University, Research Institute of Big Data and Artificial Intelligence (RIBDAI)
Research Assistant to **Prof. Choujun Zhan** (Artificial Intelligence)   Guangdong, China   Mar 2019- Mar 2022

1. **Evaluation of epidemiological interventions considering multi-directional mutation and vaccination**

     Simulation of multi-directional variation in the transmission of COVID-19 and vaccination was performed by combining real data, and Susceptible-Exposed-Unreported-Asymptomatic-Confirmed-Recovered-Deceases model with consideration of vaccination population (SEAUCRD-VP) model was proposed considering the transmission characteristics of COVID-19, evaluation of individual interventions based on the proposed epidemiological model framework, and recommendation of combined interventions.
- Participated the experimental design, and independently programming modeling, and wrote the first draft of the paper.
- First draft of peer-reviewed journal paper in preparation: "Evaluation of epidemiological interventions considering multi-directional mutation and vaccination".

2. **COVID-19 pandemic impact quantification and prediction**

     The impact of COVID-19 pandemic on global movie box office and global popular online game players was quantified using a double difference model, and a prediction model for global movie box office and global popular online game player numbers was developed by considering epidemic factors combined with machine learning. The data-driven prediction modeling by considering pandemic factors during epidemic spread can improve the accuracy of the models, and the prediction errors are reduced by 1.3% and 48%, respectively.
- Led a team on a research project, participated in experimental design, part of data collection and cleaning, part of the programming modeling and revised the first draft of the paper.
- First author conference paper published in 2021 16th International Conference on Intelligent Systems and Knowledge Engineering: "COVID-19's impact on the box office: machine learning and difference-in-difference".
- Second author conference paper published in CCF Conference on Computer Supported Cooperative Work and Social Computing, 2021: "The impact of COVID-19 on online games: machine learning and difference-in-difference".

3. **COVID-19 pandemic medical resource needs prediction**

     Explored and analyzed the correlation between nucleic acid detection and COVID-19 pandemic, and built a prediction model for COVID-19 nucleic acid detection demand using traditional machine learning and broad learning, and experimentally found that broad learning could predict nucleic acid detection demand more accurately, and the prediction error decreased by 18.3% compared to the original best result.
- Led a team on a research project, participated in experimental design, data cleaning, part of the programming modeling and revised the first draft of the paper.
- Third author conference paper published in 2021 16th International Conference on Intelligent Systems and Knowledge Engineering: "Detection capability prediction based on broad learning system during the COVID-19 pandemic".

4. **COVID-19 pandemic evolution prediction**

     The proposed improved machine learning model of RF-Bagging-BlS using random forest for feature selection and fusion of broad learning and Bagging is applied to the prediction modeling of the number of infections in COVID-19. The proposed improved model has a smaller prediction error and better fit compared to the other 14 machine learning models, with a 12% reduction in prediction error compared to the original best result.
- Independently completed the experimental design, part of data collection, data cleaning, analysis, and visualization, and programming modeling, and wrote the first draft of the paper and responded to the review.
- Based on public health departments and multiple data sources from countries around the world, a COVID-19 data set containing 184 countries and 1241 regions from December 8, 2019 to October 15, 2021 was constructed.
- Second author paper published in IEEE Internet of Things Journal: "Random-Forest-Bagging broad learning system with applications for COVID-19 pandemic".
- Co-authored software copyright: "Data analysis and prediction system of epidemic infectious disease change based on machine learning".

5. **Identification of epidemic transmission dynamics model parameters**

     An improved epidemic dynamics model, SEIR-Migration, is proposed to describe the evolution of the spread of COVID-19 in different cities in China based on a realistic intercity migration network, and a pseudo-co-evolutionary simulated annealing algorithm is proposed to solve the problem of a large number of parameters in model optimization.
- Participated in data collection and cleaning, migration network construction, and improved design of epidemic dynamics model.
- Second author paper published in Neural Computing and Applications: "Identifying epidemic spreading dynamics of COVID-19 by pseudo-co-evolutionary simulated annealing optimizers".

6. **Video game sales analysis and prediction**

A novel hybrid feature selection method (PCC-RFFS) is proposed and compared with other feature selection methods, and machine learning is applied to build a video game sales prediction model, and PCC-RFFS leads to better prediction modeling results than Pearson correlation coefficient and random forest feature importance based feature selection, and the prediction errors are reduced by 13.6% and 8.4%, respectively.

- Led a team on a research project, participated in experimental design, data collection, cleaning and visualization, part of the programming modeling and revised the first draft of the paper.
- Based on a video game statistics platform, VGcharts, a historical video game sales data set containing 37,841 games and 17 gaming platforms in Japan, Europe and the United States between 1970 and 2018 was constructed.
- Second author conference paper published in 2021 16th International Conference on Intelligent Systems and Knowledge Engineering: "Predicting video game sales based on machine learning and hybrid based feature selection".

7. **Box office analysis and prediction**

A cumulative movie box office prediction model was constructed based on ensemble learning combined with movie box office time series data. The experimental results found that the integrated learning model has better prediction effect than the traditional machine learning model in building the movie box office model, and the prediction error decreased by 33.8% compared with the decision tree model. The prediction error decreased by 33.8% compared with the decision tree model.

- Independently data collection, cleaning, and visualization, and programming modeling, and wrote the first draft of the paper.
- Based on a movie statistics platform, Box Office Mojo, a US Box office data set containing 13,737 films from 1980 to 2017 was constructed, and a Chinese Box office data set containing 3,612 films from 2011 to 2019 was constructed based on a movie statistics platform, Manyan Movie.
- Second author conference paper published in IEEE Symposium on Product Compliance Engineering-Asia, 2019: "Movie box office prediction based on ensemble learning".
- Achievement of Provincial College Students' Innovative Entrepreneurial Training Plan Program: "Box office forecasting based on big data and deep learning".
- Co-authored software copyright: "Chinese film box office data analysis and prediction system based on machine learning".

PROFESSIONAL EXPERIENCE

Huangpu Institute of Materials, Industrial Software Development Division

Algorithm Intern   Guangdong, China   Mar 2022-Present

Responsible for improvement and predictive modeling of traditional industrial control using machine learning, deep learning and other techniques to assist in algorithm design in industrial software development.

- Explored machine learning applications in healthcare and transportation with sensor development engineer, using artificial intelligence algorithms to develop medical monitoring and road condition detection.

ADDITIONAL INFORMATION

Research Interests

My main methodological interest lies in machine learning, deep learning, data mining and time series theory with their applications in public health, healthcare, and industry, include:

- Decision optimization of epidemic prevention measures.
- Time series analysis and prediction modeling for public health, healthcare, and industry.
- Epidemiological dynamics modeling.

Language Skills

English (College English Test 6, CET6), Mandarin, Preparing for the IELTS test.

Computer Skills

Python (PyTorch, Scikit-learn, Numpy, Pandas, Matplotlib), C, MATLAB, Git, Linux, LaTeX.