Yu Pan

+86-13869099329 | p4nyu@foxmail.com | LoveCatc | Beijing, China | Homepage

EDUCATION

Beijing Institute of Technology **09 2018 ∼ 06 2022**
Bachelor (90.80/100) *Beijing, China*
Sensing and Control Engineering 2/31

Beijing Institute of Technology **09 2022 ∼ 06 2024**
Master (90.89/100) *Beijing, China*
Applied Statistics -

AWARDS

- **Full Semester Scholarships** for Undergraduate Students (8 Semesters) -
- **(National) 84781 Scholarship** (1%) for Undergraduate Students 09 2020
- Outstanding Cadres of Student Association 05 2021
- **(National) 84781 Scholarship** (1%) for Undergraduate Students 09 2021
- Outstanding Graduates 06 2022
- Outstanding Graduation Thesis 06 2022
- National Post-Graduate Mathematical Contest in Modeling 2nd Prize 12 2022

RESEARCH SUMMARY

My research interests lie at the intersection of machine learning and engineering applications. As a master's student in applied statistics, I have developed expertise in statistical and mathematical foundations, including linear algebra, calculus, information geometry, stochastic process, probability theory and so on. My master's thesis research focused on applying techniques from information geometry to improve representation learning in machine learning algorithms. Going forward, I aim to leverage my background in applied mathematics and statistics to advance machine learning methods as well as other related practical applications. Also, I am looking for opportunities where I can utilize my coding skills and statistical backgrounds to conduct some interdisciplinary researches, such as the challenges and solutions of Cloud Native era in the field of operational management, the application of deep learning models in downstream industries, etc.

PROJECT & RESEARCH EXPERIENCE

National Key R&D Program | 2019YFB1406303 **04 2022**
- *Application of Advanced Mathematical Techniques in University Academic Evaluation*
- Utilizing statistical manifold theories to extract the underlying relationships between universities. [1]
- Proposing to use TDA as feature generator to distinct and classify the academic growing potential. [2]
- Responsible for data crawling and cleaning, as well as algorithm implementation.

National Key R&D Program | 2020YFC2006201 **12 2021**
- *Research and Development of AI Identification Alarm System and Equipment for Public Stadium*
- *ECG classification based on geometrical and topological methods*
- Providing assistance on data processing and algorithm implementation.
- Co-inventor of patent *Classification Method of Heart Disease via Curvature of Statistical Manifold.*

General Program of NSFC | 61179031 **10 2021**
- *Information Geometry on Lie Groups and Its Applications*
- Providing assistance on the designing and implementing of simulation systems.

General Collaboration Program – CASIC 09 2022

- *Statistical Modeling of Regional Weather Conditions and Route Planning Based on Confidence*
- Proposing to utilize Gaussian Process Model for interval prediction. [3]
- Responsible for sensor data cleaning and algorithm implementation.

Other Practice

- Protein Identification Based on Persistent Homology
- Graph Convolutional Network Based Whole Heart Mesh Reconstruction

PUBLICATIONS

[1] Yu, Daohua, Xin Zhou, **Yu Pan**, Zhendong Niu, and Huafei Sun*. 2022. "Application of Statistical K-Means Algorithm for University Academic Evaluation" *Entropy 24*, no. 7: 1004.

[2] Yu, Daohua, Xin Zhou, **Yu Pan**, Zhendong Niu, Xu Yuan, and Huafei Sun*. 2023. "University Academic Performance Development Prediction Based on TDA" *Entropy 25*, no. 1: 24.

[3] **Yu Pan**, Xin Zhou, Fupeng Sun, Yin Ni, Xingming Gao and Huafei Sun*. 2023. "A Reliable Climate Model Based On Gaussian Process" accepted by *ASCMSE2023*.

[4] Yu, Daohua, **Yu Pan**, Xin Zhou, Zhendong Niu, Xu Yuan, and Huafei Sun*. 2023. "A Hybrid Approach of Timeseries Prediction Based on Gaussian Process and Its Application in University Academic Evaluation" under review of *Knowledge-Based Systems*

TECHNICAL SKILLS

- **Coding**: Skilled in **Python** for statistical models, machine learning and deep learning. Also familiar with data processing, software development and crawler techniques.
  Available for **MATLAB** and **C++**.
- **Language**: **CET4: 632 CET6: 603**
  TOEFL: 106 (Reading: 30, Listening: 25, Speaking: 23, Writing: 28)
- **Other**: Mastering usage of **Linux** and having knowledge of **Docker**, with basic operation and maintenance skills. Familiar with **git** and capable of collaborative development. Having basic knowledge about mechatronic systems design and embedded systems development.

SERVICES

- **Teaching Assistant**: Appointed as TA for course *"Calculus B"* (2022 Fall & 2023 Spring) and 2023 MCM/ICM Problem C *Advisor* for undergraduate students (**M Prize**).
- **Volunteering**: Participating in student association BITNP and providing volunteer services including computer repairing, IT consulting and coder's salons for nearly 300 volunteer hours.

INTERNSHIP

DeepLang Technology 03 2023 – Now

Data Engineer *Beijing, China*

- My work involves developing cleaning pipelines for TB-scale data, designing SFT templates and collecting data with web crawlers to serve for the creation of Large Language Model (LLM).
- During my internship, I handled the CommonCrawl data, building the entire pipeline of collecting, extracting and cleaning Chinese corpus from the raw data.
- I also participate in several experiments concerning the data proportion of training a Chinese LLM, where I learn a lot about the concepts and practices of LLM creation.