

# Pei Fang

SENIOR UNDERGRADUATE, COMPUTER SCIENCE

---

**EDUCATION**      **Tongji University**, Shanghai, China  
*Bachelor of Engineering in Software Engineering*      *Sep' 2017 - Jul' 2021*  
**GPA: 4.57/5** (Ranked 18%)(Overall)  
*Mechanical Engineering*      *Sep' 2016 - Jul' 2017*  
**GPA: 4.14/5**(Overall)  
**Core courses:**  
**Math:** Advanced Mathematics (4), Linear Algebra (5), Probability and Statistics (5), Discrete Mathematics (5), Algorithm (5), Data Structure (5).  
**Computer Science:** Data Mining (5), Database Principle (5), Operating System (4), Computer Network (5), Computer System Structure(5), Cloud Computing (5).

---

**PUBLICATIONS**      **Manuscripts**  
**P. Zhou, P. Fang, LH. Lee, H. Pan** "Are You Left Out? An Efficient and Fair Federated Learning for Personalized Profiles on Wearable Devices of Inferior Networking Conditions," 2021 [UbiComp Submitted]

---

**RESEARCH EXPERIENCES**      **Federated Slimmable Neural Network**      *Jun '2021 - Present*

- On the basis of conventional Federated Learning paradigm, the clients download a sub-neural network models based on computation and communication overheads.
- The clients train the above models with local data. The server aggregates them with weights sharing. Use Multi-Loss to mitigate the gap of accuracy between different sub-neural networks.
- Use knowledge-distillation to distill the knowledge from large models to small models.

**Federated Sports Prediction on Wearables | HKUST SymLab**      *Sep '2020 - May '2021*  
*Supervisor : Prof. Dr. Pan Hui*

- Used Federated Learning techniques to preserve privacy. Applied a model-centric system design: A parameter server holds a global model and user mobile-phones train the model with their local data.
- Divided the training into many cycles. In each cycle, the phones send their updated model to the server, and the server aggregates them. The training data is collected by sensors in smart watches.
- Responsible for the performance improvement of the model, and the federated model implementation both in server and Android platform with [PySyft](#) and [KotlinSyft](#)
- The system predicts six different sports with three sensors at 95% accuracy. [[code](#)]

**Federated Learning Feature Engineering | Tongji University**      *Sep '2019 - May '2020*  
*Supervisor : Prof. Dr. Qingjiang Shi*

- Reproduced Federated Forest using MPI4py, in which participating agents exchange best split threshold and Gini purity without data leaks.
- Implemented Learning Feature Engineering (LFE) (F Nargesian et al. IJCAI, 2017), whereupon we collected datasets from OpenML, and generated quantile sketch arrays (QSA) to train MLP classifiers.
- Applied the transformation of which the confidence score returned by the classifier is greater than 0.9 on a real-world insurance dataset and compared the result with that from an auto-encoder. LFE reduced dimensionality from 330 to 130 and improved the f1-score from 85.6% to 87.5%.
- Studied the potential of LFE in multi-party feature engineering problem. Proposed FLFE, a framework that performs automated multi-party feature engineering.

See:[Federated Forest](#), [Automated Feature Engineering](#), and [Federated Learning Feature Engineering](#)

**COURSE PROJECTS**      **A simulative Red-Alert game**      *Mar '2018 - Apr '2018*  
*C++ Programming Language | Supervisor : Prof. Dr. Qingpei Zhao*

Used Cocos-2d as the game engine and was responsible for network server programming and basic manipulations. Completed the LAN multi-player game by using the Asio.boost net module. [code]

### **A simulative distributed file system**

*Cloud Computing | Supervisor : Prof. Dr. Jiangfeng Li*

*Dec '2019*

Implemented a simulative distributed system, which includes mechanisms such as heart-jump, redundancy, and distributed IO. Referred to the structure of UFS and Google File System. [code]

### **An OLAP movie analysis application**

*Data Warehouse | Supervisor : Prof. Dr. Hongming Zhu*

*Nov '2019 - Dec '2019*

Crawled 187,881 movies and more than 7,910,000 movie reviews. Analyzed the reviews' emotional inclinations. Tried MySQL (relational database), MongoDB (non-relational database) and Neo4j (graph database) to speed up queries. [code]

---

## WORK EXPERIENCE

### **Lightelligence, Inc**

*Position : Algorithm Engineer Intern*

*Dec '2020 - May '2021*

Reproduce the latest papers on Natural Language Processing, especially the computation acceleration of transformers, looking for the potential of these works on optic AI chips.

---

## LEADERSHIP & ACTIVITIES

### **Tongji Open-Source Student Club | Core Member**

*Oct '2017 - Apr '2020*

Gave lessons on: a. Python data analysis with NumPy and Pandas. b. React.js, basically the DOM and the transaction mechanism.

Awarded 'Best Lecturer' by **Tongji Google Camp Club**.

---

## SKILLS & OTHERS

**Languages:** C, C++, C#, Python, JavaScript, L<sup>A</sup>T<sub>E</sub>X

**Research Tools:** PyTorch, scikit-learn, PySyft, MPI, Tensorflow-federated.

**English:** TOEFL: 101 (Reading: 30, Listening: 30, Speaking: 22, Writing: 19).