Pei FANG

Address: Room513, Pengyuan 20#, Tongji University, Shanghai China
Email: greilfang@gmail.com **Mob:** 0086 - 18916093316

EDUCATION

Tongji University, Shanghai, China Aug 2017 – Jul 2021
Bachelor of Engineering in Software Engineering (expected in July 2021)
Overall GPA: **89.3/100**; Major GPA: **91.8/100**
Core courses:
Math: Advanced Mathematics (4), Linear Algebra (5), Probability and Statistics (5), Discrete Mathematics (5), Algorithm (5), Data Structure (5).
Computer Science: Database Principle (5), Operating System (4), Computer Network (5),Computer Organization Principle (4), Computer System Structure (5), Human Computer Interaction (5), Compiler Principle (4), Data Warehouse (5), Java EE (5), In-memory Database (5), Cloud Calculation (5), Software Engineering (5).
Tongji University Excellent Student Scholarship (16/211) Sep 2018
National College Green Computing Competition Second Price (12/250+) Oct 2018
Undertaking a national innovation project: a mobile app promoting Chinese food culture Apr 2019 – Apr 2021

RESEARCH EXPERIENCES

Automated Feature Engineering on Bank Insurance | Tongji University | Research Assistant Oct 2019 – Present
Advisor: **QingJiang SHI** (Professor at the Network and Machine Intelligence Lab, School of Software Engineering)

- Read papers on different feature engineering with Q-Learning, Deep Feature Synthesis and DAG.
- Applied learning feature engineering (F Nargesian, H Samulowitz, U Khurana, EB Khalil, IJCAI, 2017) on the target dataset, whereupon we collected 40 datasets on OpenML and generated quantile sketch arrays (QSA), then defined MLP regressors for unary and binary transformations (multiplication, sigmoid, isotonic regression, etc.).
- Fit the MLPs with QSA, used the MLPs to predict every potential transformation for target datasets, applied the transformation of which the probability was greater than 0.9 and compared the result with auto-encoder.
- The Learning Feature Engineering reduced the dimensions from 330 to 130 and improved the f1-score from 82.6% to 88.9%. Moreover, it does not need a model evaluation. Due to the sample imbalance (1021 positive, 83 negative), I used Adasyn-Sampling to strengthen the robustness.
- Currently I am writing a paper on federated automated feature engineering (planning to submit it to ICDE 2020). It is a mechanism that transforms features from different agents without privacy leaks. We plan to optimize the communication mechanism design rather than use Homomorphic Encryption, which would save time.

See: https://github.com/Greilfang/Automating-feature-engineering

Research on federated machine learning | Tongji University | Research Assistant Sep 2019 – Oct 2019
Advisor: **QingJiang SHI** (Professor at the Network and Machine Intelligence Lab, School of Software Engineering)

- Generally viewed works in federated learning, especially the application in statistical machine learning and analyze the methods of Secret-Sharing and Homomorphic Encryption.
- Reproduced the work of Federated Forest using MPI4py. Based on Secret Sharing, each agent can exchange the information on best split threshold, gini purity without data leaks.

See: https://github.com/Greilfang/Federated-Forest

Co-forest on Multi-classification Model | Tongji University **|** Research Assistant Apr 2019 – Jun 2019
Advisor: **QingJiang SHI** (Professor at the Network and Machine Intelligence Lab, School of Software Engineering)

- Reshaped the data to 224x224 and applied VGG-16 net, reaching a precision of only 83.7%.
- Explained to the professor why I thought CNN seemed unsuitable, since, generally, the filters reveal relations between neighboring pixels while features in the datasets may not have such correlations.
- Based on Co-forest (Li M, Zhou Z H, 2007), I labeled 46,021 pieces of unlabeled data with existing 5,361 labeled ones. Then improved the confidence threshold θ to reduce the estimated error, ultimately adding Gaussian Noise, and achieve a precision of approximately 87%.
- Finally, we reached an 89.9% accuracy with XG-boost and an 87.8% accuracy in tri-training and Random-Forest, which, when compared with XG-boost, saved around 70% of the training time and made full use of unlabeled data.

COURSE PROJECTS

Development of Reversi-playing System on C | C Programming Language Oct 2017 – Jan 2018

- Applied a decision tree to discover the best strategy and used Alpha-beta pruning, bitmap and presort skills to speed up the searching procedure.
- Achieved a winning rate of 85% (ranking 8th out of 208 participants, in competition with other systems).

Development of a simulative Red-Alert game | Object-Oriented Programming Language Mar 2018 – Apr 2018

- Used Cocos-2d as the game engine and was responsible for path-finding algorithms, Network Server programming and basic manipulations.
- Completed the LAN multi-player game by using the Asio.boost net module.

See: https://github.com/Greilfang/Our-Red-Alert

A simulative distributed file system | Cloud Calculation Dec 2019

- Realized a simulative distributed system with Python, which includes mechanisms such as heart-jump, redundancy (3 replications) and distributed reading and writing.
- Referred to the structure of UFS and Google File System.

See: https://github.com/Greilfang/hive-hdfs-practise

An OLAP movie analysis module | Data Warehouse Dec 2019

- Crawled 187,881 movies and more than 7,910,000 reviews from Amazon, analyzing emotional inclinations.
- Tried the integrations of MySQL (relational database), MongoDB (non-relational database) and Neo4j (graph database) to increase the speed of searching.
- Used flask as backend.

See: https://github.com/Greilfang/Amzaon-movie-analysis

WORK EXPERIENCES

Internship at SAP Shanghai Lab, China | VT Internship Jul 2020 (expected starting date)

- Got familiarize myself with the SAP ERP system.
- Engaged in all ERP modules, including Sales and Distribution, Purchasing Order, Finance and Controlling.

LEADERSHIPS AND ACTIVITIES

Open-Source Student Club, Tongji university **|** Core Member Sep 2017 – Present

- Gave lessons on the following subjects:
  1. Python data analysis, using NumPy and Pandas to analyze changes in American names. About 40 people from school of CS and SE took part.
  2. React.js, basically on ECMA6 new features, the DOM and transaction mechanism of React.js. Awarded **'Best Lecturer'** by the **Tongji Google Camp Club**.
- Organized the ceremony for the release of Ubuntu Kylin at Tongji University in April 2018.

SKILLS AND OTHERS

TOEFL: Total: 101 (Reading: 30, Listening: 30, Speaking: 22, Writing: 19).
Computer: C/C++, Python, JavaScript, Java, experience in PyTorch, scikit-learn, knowledge of machine learning and deep learning. Experience backend development with flask, Spring Boot.
Optional Courses:

- Math: Game Theory**,** Combinatorics, Mathematical Modeling;
- Others: International Finance, Social Psychology.

Language: Chinese (Native), English (Fluent).
Interests: Football, Calligraphy.