Yanwen Lin

Phone: (+1)-412-708-5446 | Email: lyw1124278064@gmail.com
GitHub: https://github.com/jasonyanwenl

Education Background

Carnegie Mellon University
*M.S. in Intelligent Information Science (School of Computer Science), **GPA**: 3.91/4.0* Aug.2019-Dec.2020 (expected)
*M.S. in Civil and Environmental Engineering, **GPA**: 3.94/4.00* Aug.2017-Dec.2018
***Selected Courses**: Distributed Systems, Parallel Computer Architecture and Programming, Introduction to Deep Learning, Cloud Computing, Search Engines, Computer Networks, Introduction to Computer System, Data Structure and Algorithm*

Dalian University of Technology
*B.Eng. in Civil Engineering, **GPA**: 3.8/4.00 (top 10%)* Sep.2013 - Jun.2017

Professional Skills

- **Programming Languages**: Java, Python, C, Scala, Bash, HTML/CSS/Javascript, MATLAB, R
- **Software and Tools**: AWS, PyTorch, Numpy, Pandas, MySQL, HBase, MongoDB**,** Hadoop, Spark, Kafka, CUDA, OpenMP, MPI, Java Spring Boot/Hibernate, Java RMI, Terraform**,** Docker and Kubernetes, Pandas, Git, YACC/Flex

Work Experience

LinkedIn System and Infrastructure Engineer Intern [source] Sunnyvale, Jun.2020-Aug.2020
- Developed a backend server based on LinkedIn business-wide monitoring platform ThirdEye (open source).
- Implemented API services for various core functionalities supported by ThirdEye platform such as anomaly detection and root cause analysis.
- Refactored partial services to separate the logics of frontend and backend to use API service.

Horizon Robotics Backend Software Engineer Intern Nanjing, Jan.2019-Mar.2019
- Integrated Apache Druid with access control system using basic security and Kerberos extension.
- Coordinated pluggable Apache Kylin with Hadoop computing engine, HBase data storage and Hive data warehouse.

Metro 21 Institute Data Science Research Intern [source][paper] Pittsburgh, Jun.2018-Oct.2018
- Built an evaluating system which feeds ~600k lines of fire and property data within entire Pittsburgh into a XGBoost model to estimate its performance using Python Pandas and Jupyter Notebook.
- Identified potential factors that leads to high fire risk based on model result and informs the Bureau of Fire's prioritization of property fire inspections.

Paper

Jessica Lee, **Yanwen Lin**, Michael Madaio. A longitudinal evaluation of a deployed predictive model of fire risk. 32nd Conference on ***Neural Information Processing Systems*** AI for Social Good Workshop, Montréal, Canada.

Projects

AFS-style Distributed File System Based on Check-on-use Cache Policy Jan.2020-Feb.2020
- Developed a distributed file system including interposition shared library and RPC server from scratch.
- Designed a complicate RPC protocol message format for communication between RPC client and server.
- Implemented open-close session to resolve conflicts of sharing files between concurrent users.
- Integrated check-on-use cache proxy and LRU eviction policy to reduce file retrieval latency.

High Performance Web Service for Data Retrieval Oct.2018-Dec.2018
- Conducted Extract, Transform and Load on a large Tweets dataset (~ 1 TB).
- Developed user intimacy ranking system and topic word extraction system based on pre-processed Twitter data and provided APIs for client queries.
- Optimized various aspects of the system such as database schema, load-balancing, data sharding and replication.
- Achieved 6th in a 6-hour live server-performance competition out of 32 teams.

Learning Management Web Application Based on Spring Boot [source][demo] May.2020-Jun.2020
- Built a web application helping users manage their learning entities and add arbitrary tagging for each recorded entity.
- The Web application is based on Java Spring-Boot framework with MySQL as Database and Hibernate as ORM tool.

Multi-track Music Generation with Transformer Model [demo] Mar.2020-May.2020
- Developed a Transformer language model to generate multi-track music pieces including piano, guitar, drum, etc.
- Applied various data representation techniques such as composer and hybrid modes to model the interdependency between different music tracks.