Yanwen Lin

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

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

Work Experience

LinkedIn, System and Infrastructure Engineer Intern [source]   Sunnyvale, Jun.2020-Aug.2020
Tech stack: Java, RESTful API, MySQL, Dropwizard
- Developed RESTful API services for ad-hoc anomaly detection provided by LinkedIn business-wide monitoring platform ThirdEye to help customers interactively detect anomaly points inside various business data (page views, login failure rate) across different dimensions (country, web browser version, etc.).
- Designed an end-to-end workflow for serving ad-hoc anomaly detection within a distributed cluster including user-friendly API interface, communication between services scheduling worker threads, queuing requests and storing ad-hoc data source.
- Implemented & improved the service flexibility via supporting plug-and-play detection algorithms.
- Collaborated with other team members to implement and deploy the service to users across the whole LinkedIn infra teams.

Horizon Robotics, Backend Software Engineer Intern   Nanjing, Jan.2019-Mar.2019
Tech stack: Java, Apache Druid, Apache Kylin
- Developed an API service to ingest raw video streaming data and provide query services to downstream applications based on a big data cloud platform.
- 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
Tech stack: Python, Jupyter Notebook, Pandas
- Identified potential factors that leads to high fire risk based on model result and collaborated with Bureau of Fire's to prioritize the inspection of properties.
- 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.

Project Experience

Distributed File System Integrated with LRU Cache Proxy   Jan.2020-Feb.2020
- Developed a distributed file system including RPC protocol format and an RPC server from scratch.
- 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 (~ 1.2 TB).
- Developed user intimacy ranking system and topic word extraction system and provided web APIs.

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

Professional Skills

- **Programming Languages**: Java, Python, C, Scala, Bash, MATLAB, R
- **Tools**: AWS, Docker, MySQL, Hadoop, Spark, CUDA, OpenMP/MPI, Spring Boot, Hibernate, PyTorch, Git