Yanwen LIN

Phone: +1-412-708-5446 | E-mail: lyw1124278064@gmail.com
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

Education Background

Carnegie Mellon University Aug.2019 - present
Master of Science in Intelligent and Information Systems (School of Computer Science)

Carnegie Mellon University Aug.2017 - Dec.2018
Master of Science in Civil and Environmental Engineering GPA: 3.92/4.00

Dalian University of Technology Sep.2013 - Jun.2017
Bachelor of Engineering in Civil Engineering GPA: 3.84/4.00

Research Experience

Longitudinal evaluation on Deployed Pittsburgh Fire Risk Model Metro21 Inst., CMU
Research Assistant, Partnered with the City of Pittsburgh's Bureau of Fire Jun.2018-Oct.2018

- Designed consistency assessment framework for deployed Pittsburgh fire risk model along timeline.
- Conducted longitudinal evaluation on deployed model via various metrics such as transition measurement between different risk-level groups, top-k empirical risk curve.
- Accepted by NIPS AI for Social Good as workshop paper.

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

High Performance Web Service for Data Retrieval Oct.2018-Dec.2018, CMU

- Implemented Extract, Transform and Load (ETL) on a large Tweets dataset (~ 1 TB) and loaded the data into MySQL and HBase systems (with customized MapReduce using Bulk-Loading API).
- Orchestrated frontend Undertow server and backend HBase/MySQL server on cloud infrastructure using Terraform.
- Optimized each piece of the system to improve throughput from ~1000+ to ~8000+ RPS.

Social Networking Timeline with Heterogeneous Backends Oct.2018, CMU

- Integrated RDBMS (MySQL), GraphDB (Neo4j) and NoSQL (MongoDB) in a social network web service context.
- Built a complex social networking web application with fan-out queries that span multiple databases.

Iterative Processing System on Social Relationship Graph Data via Spark Oct.2018-Nov.2018, CMU

- Developed Spark applications progressively using composite of RDDs, DataFrame and SparkSQL with Scala.
- Finished the execution model of graph processing by analyzing a social graph with the PageRank algorithm.

Cannes Film Prediction based on Imbalanced Twitter Data Modeling Apr.2018-May.2018, CMU

- Developed a pipeline for data collection, pre-processing and modeling to predict 2018 Cannes winner list.
- Tuned model Hyper-parameters based on Receiver Operating Characteristic (ROC) curve and Confusion Matrix.

Applied Machine Learning for Gene Expression Profile Classification Nov.2017-Dec.2017, CMU

- Performed dimension reduction on cell gene data using PCA, LDA to maximize variance inside sampled data and speed up data processing
- Constructed a multi-label classification model (ensembled by logistic regression, SVM, Gaussian Process) based on pre-processed training data and produced prediction for types of testing cells.

Professional Skills

- **Programming language**: Python, Java (with Maven), Scala, C, Bash, Terraform, MATLAB, R
- **Database System:** RDBMS (MySQL, SQLite), NoSQL (HBase, MongoDB), GraphDB (Neo4j)
- **Technical skills:** Docker and Kubernetes, Apache Hadoop MapReduce, Apache Spark, Machine learning, Statistical analysis, Git, Yaml, Latex, Mob Programming