

LONG WANG

Addr: 5030 Centre Ave, apt 459, Pittsburgh, PA 15213

Email: lwang010@gmail.com | Tel: 412-880-9374

Github: [longw010](https://github.com/longw010) | Blog: <https://longw010.github.io/>

EDUCATION

Carnegie Mellon University (CMU) Pittsburgh, PA May 2018

Master of Science in Computational Biology, GPA: 3.63

Huazhong Agricultural University (HZAU) Wuhan, China Jul 2015

Bachelor of Science in Bioinformatics, Valedictorian, GPA: 3.68

WORK EXPERIENCE

SemanticMD, Inc. Houston, TX. May 2017–Aug 2017

Deep Learning R&D Intern

- Preprocessed on images and videos and extracted informative features using computer vision algorithms
- Developed three deep learning pipelines to 1. detect sperm cells from cell cluster images and recognize their classes; 2. identify diabetic retinopathy patients; 3. classify two-chambers vs. four-chambers ultrasound videos.

Jianbing Yan's Lab *National Key Laboratory of Crop Genetic Improvement.* Wuhan, China. Aug 2014–Jul 2015

Research Assistant (published one paper)

- Developed a tree-based machine learning method to detect SNP associations with rare variants in maize

Lingling Chen's Lab *College of Informatics, HZAU.* Wuhan, China. Jul 2013–Jun 2016

Research Assistant (published three papers)

- Developed an ensemble model to infer transcriptional regulation using RNA-seq in Prokaryotes.
- Developed a statistical method to predict protein-protein interaction network and performed network analysis.

SELECTED COURSE PROJECTS

Diabetic Retinopathy Classification with Imbalanced, Low-resolution retina images (Python | Independent)

- Designed and implemented a pipeline to do the classification (applied contrast limited adaptive histogram equalization (CLAHE) to enhance the contrast of the images, constructed a GAN to run data augmentation and applied transfer learning to avoid overfitting; embedding resolution-aware CNN with VGG16)
- Improved the accuracy from 92.9% (baseline) to 96.8%.

Machine learning in Large datasets (Python | Independent)

- Designed an efficient way to run a scalable regularized logistic regression on stochastic gradient descent (faster 10x than the classical LR model within a small memory)
- Implemented an automatic reverse-mode differentiation platform to train the CNN models (such as LSTM)
- Implemented a memory-optimized label propagation algorithm in Spark

Twitter Analytic Web Service (Java, MapReduce, HBase, MySQL | Team)

- Exerted ETL on five datasets (>1T each) by MapReduce and Spark, with data stored in MySQL and HBase
- Designed schema to load, developed Undertow and Vert.x web servlets and high-performance backend servers (by scaling, profiling, and finding perfect hyper-parameters sets in MySQL and HBase) to handle high loads of queries (29k RPS effective throughput) and respond to a live load

Automating and auto-scaling distributed service (Independent)

- Configured and deployed an Elastic Load Balancer with an auto-scaling group on AWS
- Built RESTful APIs with a machine learning task, dockerized the servers, and deployed to the Kubernetes cluster on GCP
- Configured FaaS with different tasks and events (including a video searching pipeline) on Azure and AWS

Scrabble with Stuff (Java, JUnit Test | Independent)

- Designed a multiplayer Scrabble game from scratch using UML diagrams and design patterns
- Implemented core mechanics and the GUI of the game with Java Swing using MVC pattern
- Completed comprehensively unit tests using JUnit and Apache Ant

SKILLS

Programming: Java, Python, Scala, Bash, R, Matlab, Golang, Perl, C++.

Tools: Hadoop, Spark, Tensorflow, Pytorch, OpenCV, Hbase, MySQL, HDFS, MapReduce, Kafka, Scikit-learn.

ACADEMIC HONORS AND AWARDS

MSCB Merit Fellowship, *Computational Biology Department. CMU, USA.* (twice) Aug 2017

Honorable Mention, Interdisciplinary Contest in Modeling, *Comap Inc, USA.* Apr 2014

First Prize, Chinese Undergraduate Mathematical Contest in Modeling, *Ministry of Education, CN.* Nov 2013

National Scholarship, *Ministry of Education, China. (for top 0.5% students)* May 2012