LONG WANG

longw@andrew.cmu.edu | (412) 880-9374 | Linkedin

EDUCATION

Carnegie Mellon University (CMU) Pittsburgh, PA   May 2018(expected)
Master of Science in Computational Biology
Huazhong Agricultural University (HZAU) Wuhan, China   Jul 2015
Bachelor of Science in Bioinformatics, Valedictorian

WORK EXPERIENCE

SemanticMD, Inc. Houston, TX.   May 2017–Aug 2017
Deep Learning R&D Intern
- Prepossessed on images and videos to filter out essential features using computer vision algorithms
- Developed and implemented machine learning pipelines to 1. detect sperm cells from cell cluster images and recognize their classes; 2. identify diabetic retinopathy and glaucoma patients using optical coherence tomography images; 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
- Developed a tree-based ensemble method to detect SNP associations with rare variants in maize
- Participated in maize genomes assembly (hybrid PacBio and Illumina sequencing data)
Have published one paper as a participating author in **Nature Communication**.

Lingling Chen's Lab *College of Informatics, HZAU.* Wuhan, China.   Jul 2013–Jun 2016
Research Assistant
- Developed an ensemble model to infer transcriptional regulation using RNA-seq in Prokaryotes.
- Developed a computational method to predict protein-protein interaction network; constructed the protein-protein pipeline of *Bacillus licheniformis* WX-02; analyzed network features.
- Updated sequence and annotation of *Bacillus licheniformis* WX-02; analyzed asRNAs regulation under salt condition.
Have published two papers as a participating author in **FEBS Letters** and **Scientific Reports**, respectively and a paper as the first author has been accepted in **Current Bioinformatics**.

SELECTED COURSE PROJECTS

3D Structure Reconstruction of Human Chromosomes from Hi-C data (Python | Group)
- Extracted interaction frequency information from Hi-C data
- Estimated the 3D position for each nucleosome bead by posing the distance information to a constraint convex problem (solved two objective functions via gradient descent)
- Visualized 3D structure results in PYMOL

Machine learning in Large datasets (Independent)
- Applied sequential analysis on one-month Wikipedia data (set filters and MapReduce pipeline)
- Design an efficient way to run regularized logistic regression on stochastic gradient descent
- Implemented automatic reverse-mode differentiation for multilayer perceptron and long short term memory (LSTM) on large datasets

Retrieval Algorithms for Search Engine (Java | Independent)
- Implemented Boolean retrieval, BM25 and Indri ranking, and learning-to-rank capabilities.
- Implemented query expansion capabilities(pseudo-feedback) and diversification (PM-2 and xQuAD).
- Evaluate performance by calculating and comparing different metrics.

SKILLS

Programming: Java, Python, R, Perl, Bash, Matlab, Golang, C.
Tools: Git, Hadoop, Spark, Scikit-learn, OpenCV, Tensorflow, Weka, GATK.

ACADEMIC HONORS AND AWARDS

MSCB Merit Fellowship, *Computational Biology Department. CMU, USA.*   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