Nan Du

http://strideradu.github.io/•dunan@msu.edu•517-303-3341

Research Interest

Bioinformatics • Machine Learning • Artificial Intelligence

Summary

- Designing and implementing efficient, highly scalable algorithms to process the erroneous reads from large scale next generation sequencing data
- Good performance on graduate level class and solid understanding on algorithm design, machine learning, and software engineering

Education

Ph. D., Computer Science — **May, 2019 (Expected)**
Michigan State University, Michigan, USA
- Supervisor: Prof. Yanni Sun
- GPA: 4.0/4.0

M. Sc., Condensed Matter Physics — **May, 2015**
Michigan State University, Michigan, USA
- Supervisor: Prof. Chong-Yu Ruan
- GPA: 3.6/4.0

B. Sc., Physics — **June, 2011**
Fudan University, Shanghai, China

Experience

Research Assistant, Bioinformatics Lab — **Jan 2015 - present**
Michigan State University, Michigan, USA
- Working on developing of new algorithm for the challenge of the high error rate and long reads from Third Generation Sequencing like Pacific Biosciences and Oxford Nanopore. Designed the algorithm by modified one of popular error correction method. Implemented the algorithm and tested on simulated sequencing. dataset and real dataset

Research Assistant, Ultrafast Electron Microscopy Lab — **Jan 2013 – Dec 2014**
Michigan State University, Michigan, USA
- Design and setup the high-brightness ultrafast electron microscopy. Implement Mean Field Model to simulate the trajectories of electron bunch. Data analysis on 100 GB scale electron diffraction image data to track the crystal structure change in femtosecond level.

System Service Representative — **Jun 2010 – Aug 2010**
IBM Blue Pathway Summer Internship Program, IBM China
- Highly selective internship program (3% acceptance rate for Greater China Area)
- Maintenance IBM system with guide from mentor

Publication

- Du, Nan, and Sun, Yanni. "**Improve homology search sensitivity of PacBio data by correcting frameshifts**." *Bioinformatics* 32.17 (2016): i529-i537.

Conference

- Improve homology search sensitivity of Pacbio data by correcting frameshifts (Proceeding Talk), ***15th European Conference on Computational Biology (ECCB 2016)***, The Hague, Netherlands
- Poster, ***24th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2016)***, Orlando, Florida, USA

Projects

Silver Medal, Google Cloud & YouTube-8M Video Understanding Challenge — **March 2017 – June 2017**
Kaggle Competition
- Final ranking 35 of 650 with team RandomForest
- Data processing on large YouTube video dataset (1.7TB)
- Ensemble based on variations of LSTM and MoE models

Learning to be Poetic: Automatic Generation of Chinese Song Ci Using RNN — **Feb 2017 – May 2017**
CSE 847 Machine Learning class project
- Working with two class mates two train an LSTM model that can learned how to write Song Ci poems.
- Most popular project voted by classmates

Classifying mathematic teaching pins from Pinterest using deep learning — **Oct 2016 – Dec 2016**
CSE 881 Data Mining class project
- Implement Convolutional Neural Network model using TensorFlow
- Transfer learning using features pertained in Google Inception model on ImageNet

Skills

Efficient: Python, Machine Learning (Matlab, scikit-learn), Deep Learning (TensorFlow)
Familiar: C++, Java, MapReduce, MySQL

Rewards

Travel Fellowship, ECCB 2016 — **Sep 2016**
Summer Research Fellowship, College of Engineering — **Jun 2016**
Excellent Student, Fudan University — **May 2011**