Joshua Fan

Curriculum Vitae
205 Dryden Road (Apt 401) – Ithaca, NY 14850
(425)-628-7677 • jyf6@cornell.edu
Website: http://joshuafan.github.io

EDUCATION

Ph.D. in Computer Science, Cornell University, *Ithaca, NY, USA (in progress)*　　　**Aug. 2019 – present**
- ➢ Advisor: Prof. Carla Gomes
- ➢ Research focus: Machine Learning, Computational Sustainability

M.S. in Computer Science, University of Washington, *Seattle, WA, USA*　　　**Mar. 2017 – Jun. 2019**
- ➢ GPA: 3.84/4.0
- ➢ Advisor: Prof. Sreeram Kannan
- ➢ Courses: Deep Reinforcement Learning, Online and Adaptive Machine Learning, Machine Learning for Big Data, Probabilistic Graphical Models, Theory of Optimization and Continuous Algorithms, Databases

B.S. in Computer Science, University of Washington, *Seattle, WA, USA*　　　**Sep. 2013 – Mar. 2017**
- ➢ GPA: 3.97/4.0 *(summa cum laude)*
- ➢ Selected courses: Natural Language Processing, Algorithms, Computational Biology, Machine Learning, Artificial Intelligence, Operating Systems, Compilers, Computer Security, Accessibility Capstone, Data Management, Data Structures, Systems Programming, Discrete Math, Probability, Linear Algebra

PUBLICATION

Sumit Mukherjee, Yue Zhang, **Joshua Fan**, Georg Seelig, and Sreeram Kannan. "Scalable preprocessing for sparse scRNA-seq data exploiting prior knowledge." *Bioinformatics*, 34, 2018, i124–i132.

RESEARCH EXPERIENCE

Research Assistant, Cornell Institute for Computational Sustainability　　　**Aug. 2019 – present (part-time)**
(supervised by Prof. Carla Gomes)
- ➢ Exploring techniques for multi-entity modelling and demixing using deep learning, with applications to predicting agricultural crop yields and chlorophyll fluorescence

Research Assistant, UW Institute for Protein Design　　　**Jan. – Jun. 2019 (part-time)**
(supervised by Prof. David Baker)
- ➢ Implemented 3-D CNN, Wasserstein GAN, and Variational Autoencoder, and used them to assess and refine the quality of predicted protein structures

Research Assistant, UW Information Theory Lab　　　**Jun. – Sep. 2017 (full-time)**
(supervised by Prof. Sreeram Kannan)　　　　　*Mar. – Jun. 2017; Jan. – Jun. 2018 (part-time)*
- ➢ Developed scalable algorithms inspired by Latent Dirichlet Allocation and matrix factorization to discover cell types and find structure in large single-cell RNA-seq datasets (over 1 million cells) (Poster, Paper, Code)
- ➢ Used Siamese Seq2Seq neural networks and deep reinforcement learning to learn an edit embedding (which approximates Levenshtein distance between strings) (Poster, Report)

Research Assistant, UW Computing for Development Lab　　　**Mar. 2015 – Jun. 2016 (part-time)**
(supervised by Prof. Richard Anderson)
- ➢ Helped redesign a survey app which helps public health workers collect data according to medical protocol
- ➢ Collaborated with PATH (global health company) and field-testers to improve interactions and user experience

INDUSTRY EXPERIENCE

NLP Research Intern, Docugami   **Jun. – Sep. 2018, Jun. – Aug. 2019 (full-time)**
(AI Document Engineering startup)   ***Mar. – Jun. 2018, Jan. – Jun. 2019 (part-time)***
  ➢ Researched and implemented state-of-the-art NLP algorithms (including topic models, embeddings, clustering, and question-answering techniques)
  ➢ Applied techniques to address open problems in enterprise document analysis

Software Engineer Intern, Facebook *(Integrity Computer Vision Team)*   **Full-time: Sep. – Dec. 2018**
  ➢ Trained and evaluated a clip-based convolutional neural network to detect graphic and violent content in videos
  ➢ Created new datasets and achieved higher accuracy for violence detection than previous approaches

Software Engineer Intern, Facebook *(Search, Whole Page Ranking Team)*   **Full-time: Sep. – Dec. 2017**
  ➢ Trained a sequence classification neural network to predict which search result module the user will click on, based on recent query history; improved quality of search ranking and click rate

Software Engineer Intern, Facebook *(Search Indexing Team)*   **Full-time: Jun. – Aug. 2016**
  ➢ Built a web tool to help engineers debug and test changes to the search indexing pipeline
  ➢ Created a back-end C++ Thrift service to query data stores and generate expected indexing output

Software Design Engineer Intern, BitTitan   **Full-time: Jun. – Sep. 2015**
  ➢ Implemented infrastructure in C# to allow mailbox migrations to be simulated and tested in memory
  ➢ Improved speed of a key method by around 60% by optimizing SQL queries and calculations

TEACHING EXPERIENCE

Teaching Assistant, Cornell University
  ➢ Course: **Introduction to Artificial Intelligence** (CS 4700): Fall 2019

Teaching Assistant, University of Washington
  ➢ Courses:
      o **Probability & Statistics** (CSE 312): Fall 2015, Winter 2016, Spring 2017, Winter 2018, Winter 2019
      o **Foundations of Computing I/Discrete Math** (CSE 311): Fall 2016, Spring 2018
      o **Introduction to Machine Learning for Non-Majors** (CSE 416): Spring 2019
  ➢ In addition to teaching sections, holding office hours, and grading, I took initiative to create additional resources and host extra review sessions to clarify concepts

ADDITIONAL PROJECTS *(more info at http://joshuafan.github.io/Projects.html)*

Storage and Retrieval of Robotic Laser Range Data (Poster, Report)   *(Course: Graduate Databases)*
  ➢ Implemented a database for laser-range scans to allow for efficient content-based retrieval of images
  ➢ Used Flexible Image Database System and Locality Sensitive Hashing to speed up nearest-neighbor search

Contextual Bandits Notes (Notes)   *(Course: Online and Adaptive Machine Learning)*
  ➢ Surveyed recent research on contextual bandits and created a report synthesizing important results/algorithms

LANGUAGES AND TECHNOLOGIES
  ➢ **Significant experience:** Python, Java, C#, SQL, C++, PHP/Hack
  ➢ **Some familiarity:** R, Matlab, Julia, HTML/CSS, JavaScript, JQuery
  ➢ **Libraries/tools:** Tensorflow, Pytorch, Eclipse, Git, Visual Studio, Linux, Nuclide

HONORS & AWARDS
  ➢ **Bob Bandes Memorial Excellence in Teaching Award**, 2019 (one of 3 winners)
  ➢ Graduated summa cum laude (top 0.5%) from University of Washington
  ➢ Dean's List for 10 quarters (Winter 2014 – Winter 2017)