Joshua Fan

Curriculum Vitae

5723 141st Place SE, Bellevue, WA
(425)-628-7677 • jyf2@uw.edu
Website: http://joshuafan.github.io

EDUCATION

M.S. in **Computer Science** *(in progress)* **2017-2019 (expected)**
University of Washington, *Seattle, WA*

- **3.93 GPA**
- **Relevant courses:** Deep Reinforcement Learning, Online and Adaptive Machine Learning, Databases, Machine Learning for Big Data, Probabilistic Graphical Models

B.S. in **Computer Science, *summa cum laude*** **2013-2017**
University of Washington, *Seattle, WA*

- **3.97 GPA**
- **Relevant courses:** Natural Language Processing, Algorithms, Computational Biology, Machine Learning, Compilers, Computer Security, Artificial Intelligence, Accessibility Capstone, Databases, Data Structures, Systems Programming, Software Design & Implementation, Hardware/Software Interface, Discrete Math, Probability, Statistics for Computer Scientists, Linear Algebra, Geographic Information Systems

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 at **UW Information Theory Lab** **June-Sept 2017**
(supervised by Prof. Sreeram Kannan)

- Researched and implemented scalable algorithms for parallel Latent Dirichlet Allocation, online optimization, and matrix factorization
- Applied algorithms to efficiently discover cell types and find structure in large single-cell RNA-seq datasets (over 1 million cells)
    - **Links:** Poster, Paper, Code (LightLDA warm start, LDA experiments)

Research Assistant at **UW Computing for Development Lab** **Mar 2015-June 2016**
(supervised by Prof. Richard Anderson)

- Helped redesign an app which helps public health workers collect data according to medical protocol
- Collaborated with team and PATH (global health company) to correct survey logic and optimize user experience
- Implemented a program to evaluate the accuracy of alignment algorithms (for scanning paper forms), and analyzed potential causes of misalignment (code)

INDUSTRY EXPERIENCE

Software Engineer Intern at **Facebook *(Integrity Computer Vision Team)*** **Sept-Dec 2018**

- Trained and evaluated a clip-based convolutional neural network to detect graphic and violent content in videos
- Achieved higher accuracy for violence detection than previous approaches
- Created data pipelines in SQL and C++ to produce training datasets for binary and multi-class classification

Machine Learning/NLP Intern at **Classify & Process, Inc.** **Mar-Sept 2018**

- Researched and implemented state-of-the-art NLP algorithms, including topic models, embeddings, and sequence neural networks
- Applied techniques to address open problems in enterprise document analysis

Software Engineer Intern at **Facebook** ***(Search, Whole Page Ranking Team)*** **Sept-Dec 2017**

- Improved quality of search ranking (and click rate) by taking user's previous queries into account
- Trained a sequence neural network to predict which search result module the user will click on
- Created data pipelines in SQL to produce training datasets, and engineered features in C++ and PHP

Software Engineer Intern at **Facebook** ***(Search Indexing Team)*** **June-Sept 2016**

- Built a 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
- Allows engineers to validate local indexing code changes by viewing formatted results in a webpage (created with PHP/XHP)

Software Design Engineer Intern at **BitTitan** **June-Sept 2015**

- Implemented an in-memory data migration provider in C# for simulating a mailbox migration in memory
- Improved the performance of a key method by around 60% by optimizing SQL queries and consolidating redundant calculations

TEACHING EXPERIENCE

Teaching Assistant at **University of Washington**

- Courses:
  - **Probability & Statistics** (CSE 312): Fall 2015, Winter 2016, Spring 2017, Winter 2018
  - **Foundations of Computing I/Discrete Math** (CSE 311): Fall 2016, Spring 2018
- Communicated difficult concepts in classroom, office hours, and grading
- Actively participated in creating additional practice problems, additional handouts, and hosting review sessions to clarify concepts

SELECTED PROJECTS

Edit Embedding via Reinforcement Learning (Poster, Report) ***(Course: Deep Reinforcement Learning)***

- Used Seq2Seq neural network and reinforcement learning to learn an edit embedding (which approximates Levenshtein distance between strings)

Storage and Retrieval of Robotic Laser Range Data in Database Systems (Poster, Report) ***(Course: Graduate Databases)***

- Implemented a database for laser-range scans to allow for efficient content-based retrieval of images
- Experimented with Flexible Image Database System and Locality Sensitive Hashing to speed up nearest-neighbor search

Political Speech Clustering (Python) (Report) ***(Course: Machine Learning)***

- Implemented unsupervised clustering algorithms (k-means, bisecting k-means, spectral clustering) on tf-idf features to analyze presidential campaign speeches (by candidates or issues)

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**: Java, Python, C#, SQL, C++, PHP/Hack
- **Some familiarity**: Matlab, Julia, HTML/CSS, JavaScript, JQuery
- **Tools/environments (past experience)**: Tensorflow, Pytorch, Eclipse, Git, Visual Studio, Linux, Nuclide

HONORS & AWARDS

- Graduated summa cum laude (top 0.5%) from University of Washington
- Dean's List for 10 quarters (Winter 2014 – Winter 2017)
- Robinson Center Paradise Scholarship (2014)