Joshua Fan

Email: *jyf2@uw.edu*
Phone: *(425)-628-7677*
5723 141st Pl SE – Bellevue WA, 98006
Website: *joshuafan.github.io*
GitHub: *github.com/joshuafan*

Education

University of Washington, Seattle, WA *(combined BS/MS program)* **2013-2019 (expected)**

- **M.S. Computer Science** *(expected June 2019)*; **B.S. Computer Science** *(completed March 2017)*
- **GPA: 3.97** (B.S.), **3.93** (M.S.)
- **Relevant courses** *(* indicates graduate course):* Deep Reinforcement Learning*, Online/Adaptive Machine Learning*, Databases*, Machine Learning for Big Data*, Probabilistic Graphical Models*, Algorithms, Computational Biology, Machine Learning, Compilers, Computer Security, Artificial Intelligence, Natural Language Processing, Accessibility Capstone, Data Management, Data Structures, Systems Programming, Software Design & Implementation, Hardware/Software Interface, Discrete Math, Probability, Statistics, Linear Algebra, Geographic Information Systems

Experience

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

- Trained a clip-based convolutional neural network model to detect violent and dangerous content in videos

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

- Researched and implemented state-of-the-art NLP algorithms, including for text segmentation and keyphrase generation
- Applied models to address open problems in enterprise document analysis

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

- Improved search click rate by taking user's previous queries into account for ranking search results
- Trained a sequence neural network to predict which search result module the user will click on, given sequence of queries

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

- Researched and modified algorithms for parallel Latent Dirichlet Allocation, online optimization, and matrix factorization to scale up cell type discovery on large single-cell RNA datasets (Poster, Paper, Code 1, Code 2)
  - **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.

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

- Built an indexing console tool which enables engineers to debug and test changes to the search indexing pipeline
- Created a C++ Thrift service to query data stores and simulate indexing process, and a PHP webpage to present data

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

- Implemented an in-memory data migration provider in C#, and used it to simulate a mailbox migration in memory
- Improved the performance of a frequently-called method by ~60% by optimizing redundant SQL computations

Teaching Assistant at **University of Washington Computer Science & Engineering** **Sept 2015-June 2017**

- Communicated concepts (in Probability and Discrete Math courses) in classroom, office hours, and grading
- Participated in creating practice problems, additional handouts, and hosting review sessions to clarify concepts

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

- Helped redesign an app for public health workers that collects data and makes diagnoses per medical protocol
- Collaborated with team and PATH (global health company) to correct survey logic and optimize user experience

Selected Projects

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

- Used Seq2Seq neural network and reinforcement learning to learn an embedding which approximates edit distance

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

- Implemented a database for laser-range scans to allow for efficient content-based querying and 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 project: Undergraduate 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)

Languages and Technologies

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