

Nhat (Jonny) Tran

📍 Arlington, Texas, United States ✉ nhat.tran@uta.edu ☎ (469) 279-0297 🌐 in/nhatctran 📄 https://jonnytran.github.io/

SUMMARY

Experienced data science Ph.D. student from UT Arlington with background in software development and data engineering. Passion for building quality software for ambitious machine learning projects through innovative research, scalable technology, and team-work.

EXPERIENCE

Graduate Research Assistant / Lab Manager

BioMeCIS Lab, UT Arlington

August 2015 – Present, Arlington, TX

- Developed a regularized linear logistic regressor that resulted in a +10% accuracy improvement on early-stage lung cancer prediction by selecting robust groups of biomarkers through community detection analysis on scale-free interaction graphs.
- Performed statistical analyses of discovered disease biomarkers through t-tests and published findings to premier venues such as BIBM, PSB, and BMC Bioinformatics.
- Deployed and managed distributed frameworks on the lab computer infrastructures to enable parallel training of deep learning models across 6 machines with GPUs using Horovod on Kubernetes.

Business Intelligence Intern

USAA

June 2015 – August 2015, Plano, TX

- Designed data processing pipelines to compute order processing dates by developing ETL code on IBM DataStage and scheduling automated jobs on enterprise Linux servers.
- Developed a general-purpose data visualization tool for business reports using Bokeh and Python to interface with Hadoop and Hive using SQL queries.

Software Developer Intern

USAA

June 2014 – August 2014, San Antonio, TX

- Developed a web app for software teams to produce printouts of Agile wall cards that became widely adopted in 2+ branches of the company by constructing an interface to the project management tool.
- Collaborated in a mid-size team to gather customer requirements and participated in agile software development techniques that resulted in one of the most successful intern projects by rapid developments with the Java Wicket framework and JBoss webserver.

PROJECTS

rna2rna: Predicting Biological Interaction Network Using Deep Learning

BioMeCIS Lab • January 2020

- Built a novel graph embedding model for inductive prediction of functional interactions for newly-discovered molecular RNAs on a sparse network of lncRNA, MicroRNA, and mRNA heterogeneous sequence interactions.
- Employed a biased node sampling strategy on the scale-free graph that retains the degree distribution of sampled subgraphs, which resulted in the prediction result surpassing baselines with 85% precision on a large network of 32,000 nodes.

OpenOmics: Multi-Omics Data Integration Framework

BioMeCIS Lab • July 2018

- Developed an open-source framework for bioinformatics researchers to facilitate data preparation tasks by (1) integrating multi-omics expressions, genomics, images, and clinical data, (2) homogenizing with 20+ external annotation and interaction databases, and (3) providing an API to index and query for clinical subgroups or gene sets.
- Utilized Dask and Pandas to enable efficient processing of data-frames and designed the code modules to be extensible.

EDUCATION

Computer Science, Ph.D.

University of Texas at Arlington • Arlington, TX • Expected August 2021

Computer Science, B.S.

University of Texas at Arlington • Arlington, TX • 2015 • GPA 3.65

SKILLS

Languages: Python, Java, JavaScript, R, SQL, C

Algorithms: Graph embedding, neural networks, regularized linear models, correlation analysis, PCA

Frameworks: Pytorch, tensorflow, Keras, pandas, plotly, git, Linux, docker, Horovod, Kubernetes