Nhat (Jonny) Tran

Arlington, Texas, United States ✉ nhat.tran@mavs.uta.edu ☐ (469) 279-0297 ⊞ in/nhatctran ☁ https://jonnytran.github.io/

SUMMARY

Intellectually curious, innovative, and self-motivated data scientist with 4 years of experience in biomedical graph data analysis. Passion for building quality software in ambitious machine learning projects through team-work, innovative research, and scalable technology.

EXPERIENCE

RAAN Intern - Bioinformatics
Genentech   **August 2021 - February 2022, San Francisco, CA**
· Applied machine learning to identify genomics sequencing QC criteria for accurate HLA typing.
· Implemented, tuned, and interpreted Random Forest to identify predictive features in an imbalanced multi-task classification problem.
· Developed a scalable data generation framework to simulate sequencing quality of any user-provided dataset.
· Designed optimizations that would save 36% of computing jobs and terrabytes of storage.

Graduate Research Assistant II / Lab Manager
BioMeCIS Lab, UT Arlington   **August 2015 - Present, Arlington, TX**
· Inferred microRNA functional similarities to improve early-stage lung cancer prediction accuracy by +10% with a pipeline containing graph community detection analyses and a regularized logistic regressor with robust multivariate feature selection.
· Collaborated with cancer, immunotherapy, and neuropsychiatry experts to deliver statistical testing analyses of disease biomarkers.
· Deployed and shared knowledge to colleagues on distributed infrastructures, e.g., Horovod on a Kubernetes cluster of 6 GPU servers.

Business Intelligence Intern
USAA   **June 2015 - August 2015, Plano, TX**
· Built data pipelines to compute order processing dates with SQL-based ETL and deploying automated jobs on enterprise Linux servers.
· Developed multi-purpose data visualization plug-ins for business reports using Bokeh and Python to interface with Hadoop and Hive.

PROJECTS

RNA2RNA: Inferring Biological Interaction Networks with Deep Learning on RNA Sequences
BioMeCIS Lab · August 2019
· Built an efficient DNN model to recommend missing interactions for newly-discovered RNA sequences on a sparse knowledge-base network with link prediction result surpassing state-of-the-arts with 85% precision on a dataset of 32k sequences and 1.7m interactions.
· Utilized latent graph learning on unstructured annotation data to enable actionable, novel discoveries on the RNA knowledge-base.

LATTE: Layer-stacked Attention for Heterogeneous Graph Embedding
BioMeCIS Lab · September 2020
· Designed a new GNN model for heterogeneous information graphs with state-of-the-art inductive node classification performance.
· Optimized an attention mechanism to efficiently generate all possible higher-order relations and provide insights on salient meta paths.

OpenOmics: Open-Source Library for Multi-modal Data Integration and Visualization
BioMeCIS Lab · July 2018
· Contributed a package to facilitate reproducible bioinformatics multi-modal data integration with an API for data-frame manipulations.
· Utilized Dask to enable efficient parallel processing for data exploration and designed an interactive webserver for data visualization.

EDUCATION

Computer Science, Ph.D.
The University of Texas at Arlington · Arlington, TX · Expected May 2022

Computer Science, B.S.
The University of Texas at Arlington · Arlington, TX · 2015 · 3.65

SKILLS

Languages: Python, Java, R, JavaScript, SQL, NoSQL, C, C++, Scala, Unix.

Packages: Pytorch, TensorFlow, Keras, huggingface, Scikit Learn, Pandas, plotly, wandb.

Frameworks: Spark, Hadoop, Hive, Arrow, Docker, Kubernetes, snakemake, git.

Algorithms: Graph Neural Networks, Graph embedding, Attention-based networks, Sparse linear models.

Techniques: Deep learning, NLP, Computer vision, Time-series analysis, Cloud computing, Agile development, OOP.