Cheng CHEN

Knoxville, TN 37916, United States
cchen67@vols.utk.edu | 4709856704

INTERNSHIP

Ph.D. Software Engineering Internship
Google LLC 05/2023-08/2023

- Smart display anomaly detection and analysis by data mining

Data Scientist Internship
Amazon.com Services Inc, Sunnyvale, CA, USA 05/2022-08/2022

- Building an ensemble system of multiple Machine Learning techniques to predict project head counts(**AWS, SageMaker, Redshift, Python, Pandas**)
- Performing ad hoc data retrieval and large-scale analysis to get actionable outcomes(**Redshift, SQL**)

Data Science Engineer Internship
ShunFeng Express Co.LTF, Shenzhen, China 05/2017-07/2017

- Built a data processing pipeline to clean and update the database (**Hive, Hadoop**)
- Built a customer behavior prediction pipeline, including feature engineering, machine learning modeling, and error analysis. (**Python**, **Scikit-learn, TensorFlow, statistical analysis**)

EDUCATION

University of Tennessee, Knoxville, Tennessee
Ph.D. in Computer Science 01/2020-present

- Advisor: Dr. Michael A. Langston

Georgia Institute of Technology, Atlanta, Georgia
M.S. in Bioinformatics 08/2015-05/2017
Shanghai Jiao Tong University, Shanghai, China
B.S. in Bioinformatics 09/2011-06/2015

- Scholarship of Shanghai Jiao Tong University

PHD RESEARCH & COURSE PROJECTS

Graph theoretic analysis of End-Stage renal disease and fungi relationship 01/2023-present

- Analysis data on a demographically diverse patient population
- Found and studied the overrepresented fungus in the patients

Unsupervised disease subtyping by paraclique with comprehensive feature selection 08/2022-present

- Built a sklearn-estimator that could apply the unsupervised cluster paraclique algorithm
- Built a comprehensive feature selection workflow that could utilize the advantages of filter, wrapper, and embedding feature selection methods.
- Applying the algorithm and developed tools to different disease data

Efficient MapReduce Implementation on a Graph Algorithm: Para-clique Use Case 03/2022-05/2022

- Speed up para-clique algorithm implementation by MapReduce to process a large-scale data in parallel
- Languages and platform: **Python, Spark, Jetstream**

Gene product-Disease-Drug Link Prediction Using Tripartite Graph 08/2020-12/2022

- Built gene product-disease-drug tripartite graph and analyzed its density subgraphs.
- Predicted novel interactions among the three groups by novel graph theory algorithm.
- Evaluated the gene product-disease predations by GO and BLAST analysis, and cross-referenced a comprehensive database to evaluate the novelty.
- Evaluated the gene product-drug predations by molecular docking
- Evaluated the drug-disease prediction by fingerprint similarity analysis

PUBLICATIONS

- **Cheng Chen**, Stephen Grady, Sally Ellingson, and Michael A. Langston. “Gene-Disease-Drug Link Prediction Using Tripartite Graphs” (preparing to submit)
- **Cheng Chen**. "Advantages and Disadvantages of SVM and NRWRH in Drug-gene Interaction Prediction." International Conference on Industrial Technology and Management Science 2015