

Shulin Cao

8332 Regents Road #1H, San Diego, CA 92122 | shc131@eng.ucsd.edu | (858) 666-5267

EDUCATION

Ph.D Candidate in Computational Biology and Bioinformatics, UC San Diego 09/2015 - Now
Thesis Topic: Gene Regulatory Network Modeling Using ODE Systems and Graph Mining Approaches.

B.S. in Biotechnology, Huazhong University of Science and Technology 09/2011 - 06/2015

COURSES & SKILLS

Courses Statistical Learning, Probabilistic Reasoning and Machine Learning, Database Principles, Biomed NLP, Recommender Systems & Web Mining, Graph Mining & Network Analysis, Parallel Computing, C++ Programming, Applied Statistics, Genetics and Genomics, Linear Algebra, Advanced Calculus
Skills Python, C/C++, R, HTML5, CSS, openMP, L^AT_EX, Matlab, Bash, Tensorflow, Java, MySQL
Tools MacOS, Linux, Github, Tensorflow, Keras, Pytorch, Eclipse, Vim, IntelliJ, Hive, Presto, Spark

RESEARCH

Gene Regulatory Network Analysis and Statistical Inference Modeling La Jolla, CA
Ph.D. Candidate, CMRG at UC San Diego, Advisor: Prof. Andrew McColluch 01/2017 - Now

- Constructed a functional gene regulatory network using ODE systems modeling, implemented clustering methods on genomic data and trained graph neural network to detect potential interactions of genes, which achieved 72% accurate prediction of gene expression.
- Performed data analysis based on RNA Sequencing data and PCR data and completed bioinformatics pathways analysis using network above. Also predicted by comparing model with existing networks using methods of link prediction (under submission).

WORK EXPERIENCE

Facebook Menlo Park, CA
Machine Learning Engineer Intern 06/2019 - 09/2019

- Trained the categorization and taxonomy of user interests using TF-IDF and K-Means which efficiently and accurately extract interest hierarchy. Then built the platform of user-group connection and deployed the ML algorithms in recommendations of users to potential groups based on these interests on different granularities. This is now under experiment testing with 1 million targeted users.
- Analyzed user and group level features to train models to determine group leaders among selected users for guiding new group creation(Model optimization included). It achieved 80% positive prediction and more than 60% true positive rate (5% and 10% increase compared with old model, respectively).

SELECTED PROJECTS

Biostatistics Case: A Drop Everything Situation Cambridge, MA
Novartis Institutes for Biomedical Research, Mentor: Dr. Brian Smith 08/2018

- Processed a clinical trial placebo dataset of diabetes with feature engineering from the 15 raw features and implemented several methods including logistic regression, multivariate regression, SVM and KNN in classifying key factors influencing body ALT values responding to the drug.
- Concluded dosing effect and pharmacodynamical effects of the drug are the main features affecting patients' physiological response and potential side effects and made corresponding suggestions (further experiments tested).

NLP Application in Predicting Psychological Health from Childhood Essays La Jolla, CA
UC San Diego 04/2018 - 06/2017

- Cleaned and processed clinical notes data using implemented language model and error model for spelling corrector and also collected features using bag-of-word and embedded word2vec algorithms.
- Implemented three main methods, KNN using counts, logistic regression using TF-IDF, single-layer neural network/adaboost/random forest using word2vec to train the data respectively.
- Measured models' performances by comparing several statistical indicators such as F1 score, precision and ROC curve and determined that single-layer neural network with TF-IDF can be utilized in predictions of psychological diseases from early stage resources.

E-commercial Recommendation Systems Based on Link Analysis La Jolla, CA
UC San Diego 04/2017 - 06/2017

- Implemented a machine learning algorithm which can calculate similarity based attributes of nodes and several other network features based on link prediction using graph mining methodology on the dataset of Amazon product co-purchasing network metadata for product recommendation.
- Compared the performance of product and customer similarity graphs and the accuracy of different algorithms and explore in detail the attributes and network properties of Amazon product co-purchasing network dataset.