

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.
Courses: Statistical Learning, Probabilistic Reasoning and Machine Learning, Database Principles, Recommender Systems & Web Mining, Biomed NLP, Graph Mining & Network Analysis, Parallel Computing.

B.S. in Biotechnology, Huazhong University of Science and Technology 09/2011 - 06/2015
Courses: C++ Programming, Applied Statistics, Genetics and Genomics, Linear Algebra, Advanced Calculus

SKILLS

Programming 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, MS Office

SELECTED RESEARCH EXPERIENCE

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

- Constructed a functional gene regulatory network model based on ODE Systems for quantitative description of gene expression using python with multiple optimizations combined.
- Implemented clustering methods and graph models joint with Gene Ontology analysis API and NetworkX for detecting gene regulation mechanisms.
- Performed data analysis based on the RNA Sequencing data and PCR data and completed bioinformatics pathways analysis by comparing model with data and predicted hidden pathways using methods of Link Prediction of Network. (This work is under submission.)

SELECTED PROJECTS

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

- Processed a clinical trial placebo dataset aiming at diabetes with more than 8k observations and 15 features and a test dataset with 48 observations.
- 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.

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

- 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 the models' performances by comparing several statistical indicators such as F1 score, precision and ROC curve to determine the best model that 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 08/2017 - 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.

Astra-Zeneca Drug Combination Prediction in Drug Synergies from Cancer La Jolla, CA
UC San Diego 01/2016 - 03/2016

- Implemented a comprehensive model based on Bayes Model and Random Forest Regression Model, and we performed a combination and optimal adjustments.
- Used the model to predict synergy scores based merely on mono-therapy data and measured model accuracy using molecular data provided by AstraZeneca.