



gozhen@gmail.com



(210) 630-3140



San Antonio, USA



linkedin.com/in/gozhen



github.com/gozhen

SKILLS

Python(7+ yrs)

Numpy

Scipy

scikit-learn

MATLAB(10+ yrs)

Ruby

Pearl

R

Java

C / C++

Clustering

Ensemble Learning

Network analysis

Hidden Markov Model

SVM

Random Forest

Naïve Bayes

NLP

Weka

Flask

Django

Linux

Git / Github

Parallel Computing

HPC

MongoDB

Zhen Gao

Research Scientist

Experienced Researcher with a demonstrated history in Machine Learning. Multiple first author publications on top ranked journals/conferences (most are Data Science related), including a groundbreaking work in data mining. Strong background in Data Science/Bioinformatics with a rich project experience and a wide range of engineering skills.

EXPERIENCE

Research Scientist / Postdoc

Dept. of Computer Science, UTSA

09/2016 – Present

Achievements/Tasks

- Proposed a novel multiple classifier generation and combination framework based on fuzzy clustering and individualized ensemble construction - a supv. learning framework for general data analysis and modeling.
- Built a sequencing - peak calling pipeline for histone marker modification data. Data analysis on genomic data.
- Web Scraping on sequence logo images using Python BeautifulSoup, Mechanize and Ruby Selenium.

Research Assistant/Teaching Assistant

Dept. of Computer Science, UTSA

09/2010 – 08/2016

Achievements/Tasks

- Built an application along with web server to recognize sequence logo images and re-generate position weight matrices for biologists. (Running server; 1000+ usage count; well maintained Github repository)
- Proposed an approach to model protein-DNA interactions by Multi-Instance Learning with significant improv.
- Proposed a network based approach for Cis-regulatory element discovery. Hidden Markov Model (HMM) used.

SELECTED PROJECTS/PUBLICATIONS

Z Gao, J Ruan*, A novel multiple classifier generation and combination framework based on fuzzy clustering and individualized ensemble construction, IEEE International Conference on Data Mining

- Highlights: Fundamental machine learning; novel topology based meta classifier; improves current classifiers, including deep learning in general; ability to reveal novel insights on data. Submitted to ICDM 2018, currently under review.

Z Gao, J Ruan*, Computational modeling of in vivo and in vitro protein-DNA interactions by Multi-Instance Learning, Bioinformatics, (2017) 33 (14): 2097-2105

- Highlights: First application of Multi-instance learning on target area; Impact Factor = 7.3; Well-maintained GitHub repository.

Z Gao*, Lu Liu and J Ruan, Logo2PWM: a web-based tool to convert sequence logo to position weight matrix, International Conference on Intelligent Biology and Medicine (ICIBM 2016)

- Highlights: *Corresponding author; Running web APP server with 1000+ usage count; Well-maintained GitHub repository.

D Lovelace, Z Gao, K Mutoji, Y Song, J Ruan, and B Hermann*, The regulatory repertoire of PLZF and SALL4 in undifferentiated spermatogonia, Development 143, no. 11 (2016): 893-1906.

- Highlights: Pipelining sequencing, peak calling and motif finding on HPC; Data analysis; Adjoint first author; IF = 6.06

Z Gao, R Zhao, and J Ruan*, A Genome-Wide Cis-Regulatory Element Discovery Method Based On Promoter Sequences And Gene Co-Expression Networks, BMC Genomics 14 (S1):S4, 2013.

- Highlights: Data engineering in combining heterogeneous types of data. HMM and graph based method have been used.

EDUCATION

Ph.D. in Computer Science

University of Texas at San Antonio

08/2010 – 08/2016

San Antonio, USA

B.S. in Computer Science

Shandong University of Science and Technology

09/2005 – 08/2009

Jinan, China