CHENG, YUQI

Bldg. No.12 China Agricultural Univ. West Campus, No. 2 Yuanmingyuan West Rd, Haidian Dist, Beijing, China 100083
✉: chengyuq@cau.edu.cn | : 086-18903215357

EDUCATION & QUALIFICATION

China Agricultural University (CAU) | Beijing, China
- **Bachelor of Science Degree, Biological Science** 09/2016-07/2020
- **GPA: 3.90/4.00** *Ranking 2/80 Among the Whole Grade*
- **Bachelor of Engineering Degree, Data Science and Big Data Technology (Dual Major)** 09/2018-07/2020
- **GPA: 3.82/4.00**
- GRE: Total 324: V 154/170 (64%), Q 170/170 (96%), AW 3.0/6.0 (15%) 08/2019
- **Gold Medal in iGEM Competition** (International Genetically Engineered Machine Competition) 11/2019
- **Merit student of Beijing** 12/2019
- **Baosteel Scholarship** (Top 3 among all the senior students of CAU) 10/2019
- 1st Class Scholarship for Academic Excellence 11/2019, 12/2018, 12/2017
- **National Scholarship** 12/2018
- **First Prize of Beijing College Students Biology Competition** 05/2019, 06/2018
- Jin Long Yu Scholarship 12/2017

RESEARCH EXPERIENCE

Study on Co-expression Network of Hexaploidy Wheat Gene Based on Transcriptome Big Data *Supervisor: Weilong Guo*
Group Leader 12/2018-06/2020
- RNA-seq Simulation and Process evaluation:
  - Simulated the RNA-seq data using Flux-simulator, achieved error model and non-error model files
  - Used STAR to align them with IWGSC v1.1 (wheat genome) in different parameters and programmed with python to make statistical analysis, find the best STAR alignment parameters.
- Building co-expression network based on public data:
  - Collected transcriptome data (SRA file) of hexaploid wheat, and checked the data quality with FastQC, cleaned the data with trimmomatic, used STAR to align them with IWGSC v1.1
  - Used HTSeq to count the align result and calculate FPKM with Python
  - Averaged and combined gene expression values of homology groups to form an expression matrix of homologygroups with Python (pandas, numpy)
  - Used WGCNA (R package) to clean data (goodSamplesGenes, hclust), selected a soft threshold to establish a adjacency matrix and turn it to topological matrix, used dynamic clipping method clip trees and plotted it, calculated the eigenvectors within each module and drew correlation heatmap
- Built interactive website based on shiny to visualize co-expression network and make it easier to query the gene's module location.

Primary Study of the Membrane Receptor of Verticillium dahliae Asp-f2 Like Protein *Supervisor: Junsheng Qi*
Group Leader 11/2017-12/2018
- Knocked out the gene with Overlap PCR, designed DNA primers and constructed prokaryotic expression vector
- Detected the transformation of targeted DNA into E. coli using colony PCR and agarose gel electrophoresis, checked the protein expression level with Western blot and got purified protein using affinity chromatography method
- Infected the arabidopsis and cotton with bacteria, detected the wilting effect, collected and analyzed experimental data and proposed to use water loss rate to analyze the wilting phenotype

Statistical Research upon Phenotypic Diversity of Papillionaceae in Xiaolongmen *Supervisor: Zhaohui Liu*
Group Leader 07/2017-08/2017
- Planned the project including confirming the indexes that would study
- Employed SPSS to calculate the coefficient of variation, and ANOVA and PCA to analyze the indexes
- Calculated the Pearson correlation coefficient between different indexes with SPSS and analyzed the possible relationships between different characters
- Made clustering analysis upon papillionaceae germplasm resources of different positions using K-means method and analyzed their similarity and difference
- Calculated the Shannon-weaver and estimated the diversity of papillionaceae

PYTHON PROJECTS

Probability prediction of heart disease incidence based on machine learning method （Kaggle） 02/2020-05/2020
- Used *numpy* and *pandas* to do Data Preprocessing (One-Hot encoding etc.)
- Used *sklearn* to build random forest (CART based), used ROC and confusion matrix to do model evaluation
- Used *sklearn*，*eli5* to conduct Feature Importance analysis
- Used *keras* based on *tensorflow* to conduct artificial neural network (activate fuction: Sigmoid, ReLU, epochs=5)

Research on China Telecom User Offline Forecasting System based on Big Data Technology 01/2020
- Installed a Hadoop cluster based on Ubuntu 18.04, provided a distributed storage foundation through HDFS
- Used *pyspark.sql* to develop null data preprocessing and feature engineer (null value processing, etc), established the random forest model through the Spark machine learning algorithm library *Spark mllib (pyspark.ml)*
- Used AUC calculation of Spark to conduct model evaluation

Player unknown's Battlegrounds player ranking prediction 12/2019
- Used numpy and pandas to do Data Preprocessing (null value processing, etc)
- Used pandas to conduct feature engineering, matplotlib, seaborn to do visualization
- Used sklearn to build random forest model and conduct MAE (mean_absolute_error) evaluation

Online examination system-based on Python and SQL 05/2019-06/2019
- Used ER model (Entity-relationship model) to design mysql database
- Used *tkinter* to develop front-end of the examination system
- Used *SQL* construct MySQL database (containing exam questions)
- Used *pymysql* to connect database and record final score of tester
-

MATHEMATICAL MODELING EXPERIENCE

Team Leader of Model Analysis in iGEM 03/2019-10/2019
- Organized seminars to discuss and confirm the competition theme, and designed the gene circuit
- Decision model design: used AHP to provide quantitative decision basis, improved the RI value and independently designed a new algorithm for RI value
- Dynamic Model Design:

CHENG, YUQI

Bldg. No.12 China Agricultural Univ. West Campus, No. 2 Yuanmingyuan West Rd, Haidian Dist, Beijing, China 100083

✉: chengyuq@cau.edu.cn | : 086-18903215357

- Population Dynamics Model: regression-fitted the population density variation using Logistic Regression, independently designed the first order difference algorithm to promote the accuracy of parameters, and achieved the algorithm using MATLAB
- System Dynamics Model: gave a mathematical expression for the whole pathway with ODE, and used MATLAB to achieve and predict the changing of intermediate and final product and the final product yield

- Conducted marketing analysis using Linear Regression method, visualized the data with R, and analyzed the marketing potential and status-quo of cellulose, providing quantitative analytic basis for human practice
- Used HTML to complete the typography, layout and modification for web page

Team Leader in "Minsheng Cup" Mathematical Contest in Modeling−Shanghai and Shenzhen Index Volatility Analysis and Stock Valuation Modeling Based on Mathematical Model and Machine Learning 05/2019

- **Achieved 3rd Prize**
- Used python crawler to capture data of Shanghai and Shenzhen index fluctuations within half a year, used R to preprocess the Shanghai and Shenzhen index time series data, build VAR model and, made Multiple Regression Analysis
- Independently made factor analysis using SPSS for fundamental indicators, divided training sets and test sets, build BP neural network model with MATLAB and momentum gradient descent method, and used training sets and test sets to test their generalization performance, and K-means clustering (SPSS) to construct a valuation level measurement model

Team Leader in 2018 MCM/ICM−Does climate change really have influence on Regional Stability? 02/2018

- **Achieved Honorable Mention**
- Pre-processed and analyzed data using SPSS, and built up the graph
- Built the BP neural network model, implemented the model using MATLAB, verified the accuracy of algorithm using 10-fold Cross Validation method, and analyzed the data using trained BP neural network model
- Independently completed the English-written paper using LaTeX

MAIN ACTIVITIES

After iGEM academia & research committee 02/2020-10/2020

Committee member

- Connect iGEM alumni all over the world
- Conduct dry lab instruction to 2020 iGEM teams from different countries

Student Union of College of Biological Sciences, China Agricultural University 07/2017-08/2019

Director in Secretary Department (also known as Decision Supporting Center)

- Responsible for daily internal affairs of Student Union
- Statistically analyzed the number of participants and feedback information of various activities of the Student Union, and provides data analysis support for the decision-making of other departments.
- Innovatively built a staff selection system based on the AHP decision model, leading the staff selection mechanism of the Student Union to the era of data-driven decision making

Volunteering Teacher in Yihai Hope Primary School 07/2018-08/2018

- Taught in mathematics and sports, initiated their first flag class and organized the first flag raising ceremony

Freshmen Advisor for Grade 2019 08/2019-06/2020

- Organized orientation, and provided suggestions for their course selecting and university study planning and also psychological counseling service

Striker of Football Team of College of Biological Science 09/2016-present

- Partook in annual CAU Football Match and won the second place in team for 2019 CAU Football Match

ADDITIONAL

Language: Chinese (Native), English (Fluent), Japanese (Basic)
Computer: Python, R, MATLAB, C, Linux shell, SPSS, LaTeX, SQL, HTML
Other Skills: Go, Football