

Kai ZHANG

matthewzk@outlook.com · (+86) 18226139685

No. 82, Huoshan Rd., Student Apartment, Shushan Dist., Hefei, Anhui, P.R.China 233036

EDUCATION BACKGROUND

Anhui Agricultural University (AHAU), P.R.China

09/2016 – 06/2020

B.S. in Computer Science and Technology | GPA: 91.77 / 100

Core Coursework: Multivariable Calculus, Linear Algebra, Discrete Mathematics, Operating Systems, Algorithm Analysis, Software Engineering, Data Structure

RESEARCH EXPERIENCES

MIT Computer Science & Artificial Intelligence Laboratory (CSAIL)

07/2019 – 09/2019

Universal Village(UV) Group: Smart Home OS

- Designed and implemented the outline of Smart Home OS, including:
 - Established the feedback loops and the self-learning mechanism based on Reinforcement Learning
 - Conducted the literature review, and presented a qualitative description of the interactions among intelligent systems
 - Built the models to describe the relationships among subsystems with Neo4j
 - Optimized the task-oriented dialog systems, and completed the database design with MySQL.
 - Implemented the Natural Language Understanding (NLU) with Machine Learning model, handled the specific scenario with ruled-based model.
 - Conducted semantic analysis and implemented the Natural Language Generation (NLG) with CRF-based model.
 - Proposed the template-based model to deal with the abbreviations and incompetent statements.

Chinese Agricultural Knowledge Graph

Project 1: Chinese Agricultural Entity Relation Extraction (ERE) via Deep Learning

11/2018 – 03/2019

- Developed a ERE workflow for improving the speed of processing Chinese agricultural text information; key accomplishments included:
 - Built a Python-based crawler to retrieve text data of agricultural pests and diseases from online sources.
 - Performed data cleansing, segmentation, and vectorization using the Continuous Bag of Words (CBOW) algorithm.
 - Performed entity recognition using a customized Applied Conditional Random Field (CRF) algorithm, allowing the establishment of an entity-relationship diagram.
 - Built a Piecewise Convolutional Neural Network (PCNN) for encoding, followed by classifier and result validation based on selected performance indicators.
 - Built RNN, BiRNN and traditional CNN models as comparison to PCNN.
 - Optimized the classifier selection mechanism (consider using Attention mechanism) to improve the final result, leading to boosted classification accuracy and F1 score.
 - Completed a suite of tests using real-world data; a first-author conference paper *Chinese Agricultural Entity Relation Extraction via Deep Learning* accepted by International Conference on Intelligent Computing (ICIC2019) – Accept Rate:26%.

Project 2: Chinese Agricultural Named Entity Recognition

03/2019 - 06/2019

- Employed various libraries in Python to perform data preprocessing, covering data wrangling (NumPy, Scrapy, and Pandas), machine learning (Scikit-Learn and TensorFlow), data mining (Scrapy), data visualization (Matplotlib and Seaborn), and automated testing (Selenium).
- Proposed to build the Iterated Dilated CNN-CRF-Attention (IDCNN-CRF-A) Model, and carried out the comparative analysis with the Hybrid Bi-LSTM-CRF Model to verify the high accuracy.
- Wrote the academic paper- *Deep Neural model joint different features combination for better Pest and Disease Entity Recognition*, submitted to Knowledge-Based System(Under Review – Minor Revision now)

Recognition of False Online Product Reviews Based on Deep Learning

01/2019 – 06/2019

- Collected online product reviews from the Taobao website, followed by data pre-processing to remove outliers and reduce noise.
- Segmented the review texts using Hidden Markov model, forming a corpus, which was in turn vectorized via the CBOW algorithm.
- Built, trained and validated a CNN model for false review classification.
- Systematically compared the CNN-based model with the traditional SVM-based model, covering an array of performance indicators; composed a research article for documenting the findings.

EXTRACURRICULAR EXPERIENCES

Founder and President, Yiban Club, AHAU

03/2018 – Present

- Delivered 50+ lectures to junior students on the development of skills necessary for academic success such as time management, organization, critical thinking, study strategies, and problem-solving techniques.
- Helped a group of 100+ students in academic difficulty prepare for midterm and final exams, leading to a significant improvement of academic performance.
- Coordinated a series of activities to support study of quantitative subjects within AHAU student community, e.g., programming competitions, seminars, technical presentations, public services to local schools, among other fun activities.

Optimal Implementation of a Transportable Disaster Response System

Team Leader, MCM

02/2019

- Built and implemented a mathematical model to maximize the cost-effectiveness of drone-based disaster relief at Puerto Rico; key accomplishments included:
 - Scored the adaptive capability of drones using the fuzzy analytic hierarchy process; optimized the fuzzy fraction matrix

- by PSO algorithm to improve accuracy.
- Leveraged linear programming to optimize the composition of a drone fleet while honoring practical constraints.
- Optimized the cargo container locations using dynamic programming algorithm and Google map data.
- Established a container loading scheme based on genetic algorithm; Leveraged dynamic programming to reduce computational complexity.
- Composed a 20-page technical report summarizing methodology and key findings.
- Significantly strengthened mathematical modeling skills, covering model formulation, analytical/numerical solution generation, and solution interpretation carried out in an iterative manner.

Datacastle Big Data Competition

08/2018 – 09/2018

- Built a classifier in Python to predict whether an individual's annual income will exceed 50K based on a suite of factors, garnering the 1st prize (top 1%); key steps included:
 - Collected 50K+ data points, followed by data cleansing and establishment of training, test and validation datasets
 - Conducted feature selection and dimensionality reduction based on PCA and factor analysis; processed non-numeric features using one-hot encoding.
 - Systematically tested the performance of CART decision tree, random forest, boosted trees, GBDT, and Stacking classifier; employed GridSearchCV in scikit library for hyper-parameters tuning.
 - Validated the model using k-fold cross-validation.

COURSE PROJECTS

Campus Navigation App Based on Open Map Platform

07/2017 – 05/2019

- Developed a campus navigation system in Java, allowing users to add the location information unavailable on the open map platform so as to address the practical needs in the students and faculty communities.
- Followed the user-centered design philosophy by conducting on-campus survey, leading enhanced user experiences.
- Developed a workflow to retrieve geographical data from Baidu Map Platform API.
- Converted the map data into an undirected graph; employed Floyd algorithm for computing shortest paths.

Development of a Students Information Management System

Java Programming Course Project, AHAU

06/2019 – 07/2019

- Coordinated a team of four to develop an integrated software application for managing students' academic information.
- Employed various object-oriented programming concepts to facilitate development.
- Followed common software engineering practices and a suite of design patterns to enhance robustness, modularity, usability, and maintainability of the developed program.
- Completed a systematic software testing to verify the functionality of individual components and overall performance.

Multithreading Deadlock Avoidance and Resource Allocation using Banker's Algorithm

Operating System Course Project, AHAU

12/2018

- Implemented the banker's algorithm in Java to allow deadlock avoidance and resource allocation by 1) simulating the allocation of predetermined maximum possible amounts of all resources, 2) making an "s-state" check to test for possible deadlock conditions for all other pending activities, and 3) deciding whether allocation should be allowed to continue.
- Leveraged the Swing toolkit to develop a dashboard for visualizing the components and workflow of the banker's algorithm.

WORK EXPERIENCES

NLP Research Intern

Shanghai Soundwise Co., Ltd., China

07/2020 - Present

- **Project 1: Breast cancer ultrasound diagnosis report prediction task**
 - Processing medical text data: extract cancer description in the ultrasound diagnosis report by Regular Expression
 - Extracted description as feature and the diagnosis category as label to model a classifier (Supervised learning: [x1,x2..xn],label]), built several classifier like SVM, Decision Tree etc. to tackle the predict task.
 - Built LSTM+Attention model to perform the predict task as comparison
- **Project 2: From ultrasound image generate diagnostic text report (Incomplete)**
 - Relevant literature review – Image-Text embedding and alimnet topic
 - Label the image part and sentence description as a pair
 - Built Dual – Path Convolutional Embeddings to align image and text
 - Using trained model combine RNN to generate Text report from the new given image

SKILLS

Programming Languages: Proficient - Python, Java; **Familiar** - C, R, MATLAB; Latex; MySQL, Neo4j

Data Science Libraries: NumPy, SciPy, Pandas, Scikit-Learn, TensorFlow, Scrapy, Matplotlib, Seaborn, Selenium.

GitHub: <https://github.com/MatthewKKai>

AWARDS AND HONORS & PUBLICATION

- The 1st Prize, Datacastle Competition 10/2018
- The 3rd Prize, The 11th Chinese Collegiate Algorithm Design Competition (Anhui Division) 10/2018
- Outstanding Individual in Innovation and Entrepreneurship, AHAU 10/2017
- Zhang K.** et al. (2019) Chinese Agricultural Entity Relation Extraction via Deep Learning. In: Huang DS., Huang ZK., Hussain A. (eds) Intelligent Computing Methodologies. ICIC 2019. Lecture Notes in Computer Science, vol 11645. Springer, Cham.
- Q Yang, Y Wu, **K Zhang**. A Visual Analysis of the Research on Huizhou Intangible Cultural Heritage[J]. Journal of Huangshan University, 2019,21(06):5-10.
- G Wu, G Liang, **K Zhang**, L Tu. Comparative Research on SVM and RNN in Sentiment Analysis of Internet Comments[J]. Journal of Shanghai University of Engineering Science,2019,33(04):378-383.
- Y Yue, W Wang, **K Zhang**. et al. Remote Supervision on Relation Extraction via Multi-Layer Attention Mechanism[J]. Journal of Anhui Agricultural University, 2020,47(04):682-686.
- K Zhang**. et al. IDCNN: A Neural Architecture Model for Pest and Disease Entity Recognition. (Preprint - [Arxiv](#))