Shijie Li

(402)-314-9216 | http://jerry-shijieli.github.io | sli41@ncsu.edu
1701 Gorman ST 206, Raleigh, NC 27606

EDUCATION

North Carolina State University (NCSU), GPA: 3.9/4.0 Aug. 2015 - Dec. 2017
Master of Computer Science

University of Nebraska-Lincoln (UNL), GPA: 3.8/4.0 Aug. 2011 - May 2013
Master of Science, Physics

Shandong University, GPA: 87/100 Sep. 2006 - Jun. 2010
Bachelor of Science, Physics

SKILLS

- Algorithms in Machine Learning, Deep Learning, Reinforcement Learning with mainstream frameworks.
- Programming languages: *Python, Java, R, Ruby, SQL, C/C++, JS, PHP* with major IDEs and tools.

EXPERIENCE

Research | **LexisNexis, Raleigh, NC**
Customer Profile Verification By Text Information Mining On Emails Jan. 2016 - May 2016

- Created Python scripts to preprocess labeled datasets obtained by Amazon Mechanical Turk crowd-sourcing platform.
- Applied rule-based algorithm with Talon library as base-line to extract customer email signatures through regular expressions, named-entity recognition(NER) and knowledge-based rules.
- Vectorized email documents using word2vec model and train classifiers to improve the accuracy of email signature recognition. Apply SMOTE method to handle unbalanced minor class, which increases precision by above 10%.

PROJECTS

Car Rental Web App Development Sep.2017 - Oct.2017

- Designed relational database based on objectives of web app and visualized primary structure with E-R diagram.
- Performed full stack development on *Docker container* using *Ruby on Rails* with RESTful web services.
- Deployed this app on Heroku Cloud platform with *PostgresSQL* and multi-process concurrency configurations.
- Lead a three-member team using *DevOps* and *TDD* methodology for software development.

Feature Selection For Reinforcement Learning In Educational Policy Development Mar.2017 - Apr.2017

- Modified *Python* Markov Decision Process Toolbox to compute Expected Cumulative Reward(ECR) of a Reinforcement Learning system, which studies student learning performance based on data of tutor actions and learning contexts.
- Designed a Forward-backward greedy search algorithm to select optimal subset of features using Wrapper method.
- Analyzed relationships among optimal features and generate corresponding policy map to make instruction decisions.

GMM-based Change Detection On Very High-resolution Satellite Imagery Feb.2017 - Apr.2017

- Implemented symmetric Kullback-Leibler(KL) divergence to measure temporal changes between Gaussian distributions.
- Built framework of Gaussian Mixture Model(GMM) class for clustering detected changes.
- Collaboratively performed *Python* code testing and parameter optimization for image processing.
- Detected over 85% of significant changes labeled by human experts on temporal satellite images.

Music Recommender System Using Apache Spark Jan.2016 - Feb.2016

- Cleaned and indexed raw data from Last.fm using Resilient Distributed Dataset(RDD) and MapReduce model.
- Implemented the collaborative filtering recommender model using *Spark MLlib* in *Python*.
- Obtained accuracy of top-K recommendation overlapping rate above 5% on large-scaled test data.

Library Database Administration And Management System Development Aug.2015 - Nov.2015

- Designed E-R diagram to define entities and relationships among components of the library system.
- Created *Java* and embedded Oracle *SQL* codes to implement administrations of publication check out, resource assignments for various user priorities, and monitoring usage and penalty statistics in the library database.
- Collaborated with teammates and potential users on software debugging and refactoring through Agile development.

CERTIFICATE

Microsoft Professional Program in Data Science Dec. 2016 - Feb. 2017