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 - May 2019
PhD. Computer Science | Supervisor: Dr. Nagiza Samatova, Full professor

University of Nebraska-Lincoln (UNL), GPA: 3.8/4.0 Aug. 2011 - May 2015
M.S. Physics | Supervisor: Dr. Evgeny Tsymbal, Full professor

Shandong University, GPA: 87/100 Sep. 2006 - Jun. 2010
B.S. Physics | Supervisor: Dr. Mingwen Zhao, Full Professor

SKILLS AND INTERESTS

- Experience in Machine Learning, Artificial Intelligence, Data Mining and Web Development;
- Proficient in *Python, Java*, familiar with *R, SQL, C/C++, JavaScript*, experienced with major IDEs;
- Interested in machine learning, data analytics, quantitative modeling and web development positions.

SELECTED PROJECTS

Personalized Recommender System (Java, Python) Sep.2016 - Present

- Constructed personalized recommender systems with text analytics and social network mining.
- Developing feature extraction methods and content-based recommendation algorithms to improve the accuracy.

Graph Embedding for Recommender Systems (Python, Linux Shell) Apr.2016 - May 2016

- Created heterogeneous information network with nodes consisting of users, item-ratings, items, and related entities.
- Implemented DeepWalk algorithm in Python with networkx module to generate random walks over this graph.
- Based on random walks, embedded the graph in a low dimensional vector space using deep learning model word2vec.

Matching Algorithm Design for Adwords Problem (Python, MATLAB) Mar.2016 - Apr.2016

- Simulated Google AdWords online advertising marketing mechanism using random query sequence.
- Implemented Greedy, Balance and MSVV algorithms to maximize total revenue by advertisers on bid prices and budgets.
- Obtained best competitive ratio up to 99% in MSVV algorithm.

Market Segmentation using Attributed Graph Community Detection (Python, R) Feb.2016 - Mar.2016

- Divided social network into segments of common interests based on both vertex attribute similarity and edge intensity.
- Implemented using Python with the igraph module to facilitate the graph expression and manipulation.

Feature Extraction for Real-time Sentiment Analysis of Twitter Stream (Python) Jan.2016 - Feb.2016

- Set up data streaming pipeline using Kafka distributed streaming platform, Twitter API and Spark Streaming.
- Implemented Bag-Of-Words(BOW) sentiment analysis method using Spark RDDs and MapReduce API in Python.
- Implemented text feature extraction models using Python library module gensim and sklearn.
- Compared results by BOW and word2Vec using Naive Bayes and Logistic Regression with best accuracy above 80%.

Music Recommender System using Apache Spark (Python, HTML) Dec.2015 - Jan.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 90%.

Library Administration and Management System (Java, SQL) Aug.2015 - Dec.2015

- Designed E-R model and relational schema for library administration system and implemented in Oracle database.
- Organized library database using SQL queries embedded in Java code through JDBC APIs.
- Implemented menu-based UI, time reminder and late penalty messaging components in Java.

CERTIFICATE

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

TEACHING ASSISTANTSHIPS

CSC505: Design and Analysis of Algorithms, CSC230: C and Software Tools. (Fall 2015 - Spring 2017@NCSU)