zexihan.com/portfolio
linkedin.com/in/zexihan
github.com/zexihan

Zexi Han

20 Summer Street, Malden, MA 02148
zexihan@outlook.com | (617) 816-9210

EDUCATION

Northeastern University (NU) Boston, MA
*M.S. in **Data Science**, GPA: 3.8/4.0* Jan 2017-May 2019
Relevant Courses: Algorithms, Machine Learning (TA), Computer Vision, Parallel Data Processing

Beijing University of Posts and Telecommunications (BUPT), Joint Program with QMUL Beijing, China
*B.S. in **Telecommunications Engineering** with the First Class Honors, GPA: 3.5/4.0* Sept 2012-Jun 2016
Relevant Courses: Data Structures, Software Engineering, Calculus, Linear Algebra, Principles of Communications
Awards: BUPT Outstanding Final Project (Rank 12/680)

TECHNICAL SKILLS

Languages:	Python, Java, Scala, SQL, R, MATLAB
Machine Learning:	LR, Bayesian, SVM, Tree-based, Ensembles, NN, CNN, RNN, Clustering, EM, PCA, etc.
Database:	S3, Oracle Database, Snowflake
Tools:	MapReduce, Spark, AWS, Pandas, scikit-learn, Tensorflow, Caffe, OpenCV, Docker, Git, etc.

PROFESSIONAL EXPERIENCE

Data Scientist at Rue Gilt Groupe (Retail) Jan 2018-Jun 2018
- Worked on feature engineering and XGBoost model training from an iterative perspective to identify suspect resellers from over 2 million buyers, and put it into production as a block for the recommendation system.
- Built docker apps for feature extraction, training and inference which were deployed to Amazon ECS and Airflow.
- Maintained various database applications and recommendation system with robust SQL.

Software Engineer at Tsinghua University – Edge Sensing Interaction for Smartwatch May 2016-Jul 2016
- Discovered the accelerometer's data collected by tapping the edge of the smartwatch from 4/6/8 directions.
- Collected and preprocessed the accelerometer's data with Python in different scenes of daily life.
- Classified the tapping motion with logistic regression model and developed the Android Wear demo.

Research Assistant at Chinese Academy of Sciences – Visual Search with Deep Learning Aug 2015-May 2016
- Built a Three-stage Hybrid Visual Search Framework (Classification, Object Detection and Matching) to the task of same-style product image retrieval with convolutional neural networks.
- Experimented on the ALISC 5 million product image dataset with multiple CNN models using Caffe.
- Developed the backend of Android demo and achieved real-time same-style product image retrieval.

PROJECT EXPERIENCE

Neighborhood-Level Airbnb Review Generation with Spark+DL, NU Oct 2018-Dec 2018
- Proposed and implemented a neighborhood-level review generator that generates representative text review about the average living experience for the candidate neighborhoods of the traveler's destination city.
- Pre-processed the Airbnb reviews data by guests from different major cities in the world and built an LSTM nature language model with Keras and PySpark. Ran the distributed model training and inference on Amazon EMR.
- Explored different model parameters to address memory problems. Measured the speedup performances of the distributed model training with different settings of the cluster.

Business-Neighborhood Interaction on Yelp and Census Data, NU Sept 2017-Dec 2017
- Extracted representative neighborhood-level features of business dynamics from Yelp dataset.
- Employed K-Means and GMM clustering at both the Zillow Neighborhood and Census Tract level to identify clusters based on population characteristics and socioeconomic metrics.
- Investigated the relationship between local business dynamics and neighborhood characteristics.

Video Classification on YouTube-8M Dataset, NU Mar 2017-Apr 2017
- Developed a classifier with TensorFlow that could assign the class label based on given features of the video using a subset of the Google's large-scale YouTube-8M dataset.
- Compared the loss and accuracy performance of three machine learning algorithms (LR, SVM, ANN).