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

Zexi Han

1094 Robbia Drive, Sunnyvale, CA 94087
zexihan@outlook.com | (617) 816-9210

Education

Northeastern University Boston, MA
*M.S. in **Data Science**, GPA: 3.8* Jan 2017-May 2019
Relevant Courses: Algorithms, Machine Learning (TA), Web Development, Computer Vision, Parallel Data Processing
Beijing University of Posts and Telecommunications, Joint Program with QMUL Beijing, China
*B.S. in **Telecommunications Engineering** with the First Class Honors, GPA: 3.5* Sept 2012-Jun 2016
Relevant Courses: Data Structures, Software Engineering, Calculus, Linear Algebra, Principles of Communications
Awards: Outstanding Final Project (Rank 12/680)

Technical Skills

Languages: Python, Java, Scala, JavaScript, Typescript, HTML, CSS, SQL, R, MATLAB
Web App Dev: React, Redux, Angular, Bootstrap, jQuery, Node.js, Express, Spring Boot, JPA, MongoDB, mongoose
Tools: MapReduce, Spark, AWS, Pandas, scikit-learn, Tensorflow, PyTorch, Docker, Tableau, D3.js, Git

Professional Experience

Data Scientist Co-op at Rue Gilt Groupe (Boutique Retailer) – Reseller Identification Jan 2018-Jun 2018
- Worked on feature engineering and XGBoost model training from an iterative perspective to identify resellers from over 2 million buyers, and put it into production to provide them with personalized boutique recommendations.
- Built docker apps for feature extraction, training and inference which were deployed to Amazon ECS and Airflow.
- Maintained daily ETL process for the recommendation system with robust SQL on Snowflake.

Research Assistant at National Laboratory of Pattern Recognition – Visual Search Aug 2015-May 2016
- Designed and 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 Taobao 5 million product image dataset with multiple CNN models using Caffe.
- Developed the backend of demo application and achieved real-time same-style product image retrieval.

Project Experience

TripElf – Interactive-Map with Neighborhood-Level Airbnb Review Summarization, NU Jan 2019-Apr 2019
- Proposed and developed an application to help travelers pick their favorite short-term rental neighborhoods before traveling by demonstrating the machine-generated overviews of the neighborhoods.
- Explored and applied various text models, such as KL-Sum, LDA-Sum and ELMo, to summarize Airbnb reviews and generate neighborhood overview from travelers' point of view.
- Implemented an interactive-map web app in React and Mapbox GL JS for data visualization, drawing travelers a vivid picture of NYC neighborhoods from the aspect of recreation, transit, noise, safety, expense, and Airbnb host.

TuneS – Social Music Website, NU Jan 2019-Apr 2019
- Developed a SPA using MERN stack and Spotify Web API that serves for music fans to engage with other music lovers and discover new songs and artists.
- Handled OAuth authorization, like/share/follow functions with a RESTful API built in Express and MongoDB back-end.
- Designed and wrote a responsive and interactive React front-end utilizing Bootstrap and AJAX techniques.

Parallel Matrix Multiplication in MapReduce, NU Oct 2018-Dec 2018
- Studied and implemented the different parallelization mechanisms for large matrix multiplication in MapReduce, including Horizontal-Vertical Partitioning and Vertical-Horizontal Partitioning for synthetic dense and sparse matrices.
- Measured and compared speedup and scalability performance for the two intelligent partitioning methods on Amazon EMR and S3 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.