

EDUCATION

Northeastern University (NU)	Boston, MA
<i>M.S. in Data Science, GPA: 3.89</i>	Jan 2017-May 2019
<i>Relevant Courses:</i> Algorithms, Machine Learning (TA), Web Development, Computer Vision, Parallel Data Processing	
Beijing University of Posts and Telecommunications (BUPT), Joint Program with QMUL	Beijing, China
<i>B.S. in Telecommunications Engineering with the First Class Honors, GPA: 3.5</i>	Sept 2012-Jun 2016
<i>Relevant Courses:</i> Data Structures, Software Engineering, Calculus, Linear Algebra, Principles of Communications	
<i>Awards:</i> BUPT Outstanding Final Project (Rank 12/680)	

TECHNICAL SKILLS

Languages:	Python, Java, Scala, JavaScript, SQL, R, MATLAB
Web Development:	React, Redux, Node.js, Express, Spring Boot, MongoDB, mongoose
Tools:	MapReduce, Spark, AWS, Pandas, scikit-learn, Tensorflow, PyTorch, Docker, D3.js, Git

PROFESSIONAL EXPERIENCE

Data Scientist Co-op at Rue Gilt Groupe (Boutique Retailer) – Reseller Identification	Jan 2018-Jun 2018
<ul style="list-style-type: none">Worked on feature engineering and <u>XGBoost</u> 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 <u>recommendation system</u>.Built docker apps for feature extraction, training and inference which were deployed to <u>Amazon ECS</u> and <u>Airflow</u>.Maintained various database applications and recommendation system with <u>robust SQL</u>.	
Research Assistant at National Laboratory of Pattern Recognition – Visual Search	Aug 2015-May 2016
<ul style="list-style-type: none">Designed and built a Three-stage Hybrid <u>Visual Search</u> 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 <u>CNN</u> models using <u>Caffe</u>.Developed the backend of <u>Android</u> demo and achieved real-time same-style product image retrieval.	

PROJECT EXPERIENCE

TripElf - Interactive Map with Neighborhood-Level Airbnb Review Summarization, NU	Jan 2019-Present
<ul style="list-style-type: none">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.Applied various text summarization techniques, such as <u>KL-Sum</u>, <u>LDA-Sum</u> and <u>RNN</u> models, to summarize Airbnb reviews and generate neighborhood overview from travelers' viewpoint.Implemented an interactive map web app in <u>React</u> for data visualization, drawing travelers an entire picture of NYC neighborhoods, including descriptive statistics such as entertainment, expense, transit, noise and safety.	
Feedback Collection Fullstack Web App, NU	Nov 2018-Feb 2019
<ul style="list-style-type: none">Developed a web app with <u>MERN</u> stack that serves for startup owner's app/service by sending customer an email requesting feedback and tabulating survey results to the user.Handled authentication, billing/payment and automated emails using <u>Express</u> and <u>Node.js</u> back-end.Designed and programmed the <u>React</u> and <u>Redux</u> front-end that communicates data with back-end server and MongoDB.	
Parallel Matrix Multiplication in MapReduce, NU	Oct 2018-Dec 2018
<ul style="list-style-type: none">Studied and implemented the different parallelization mechanisms for large matrix multiplication in <u>MapReduce</u>, including Horizontal-Vertical Partitioning and Vertical-Horizontal Partitioning for synthetic dense and sparse matrices.Measured and compared speedup and scalability performance for the two <u>intelligent partitioning</u> methods on <u>Amazon EMR</u> with different settings of the cluster.	
Business-Neighborhood Interaction on Yelp and Census Data, NU	Sept 2017-Dec 2017
<ul style="list-style-type: none">Extracted representative neighborhood-level features of business dynamics from Yelp dataset.Employed <u>K-Means</u> and <u>GMM</u> 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.	