Luchao Qi

Research Data Scientist

(443)839-9129
lqi9@jhu.edu
3111 N Charles Street 4C
Baltimore, MD 21218

https://luchaoqi.github.io/
https://github.com/LuchaoQi
https://www.linkedin.com/in/LuchaoQi/

EDUCATION

Johns Hopkins University May 2020
M.Sc.Eng. Biomedical Engineering 3.7/4.0

Northeastern University Aug 2018
B.Eng. Biomedical Engineering 3.9/4.0

SKILLS

Programming: Python, R, SQL, Batch Scripting

Packages & Frameworks: NumPy, Pandas, Tidyverse, NLTK, Keras, PyTorch, TensorFlow

Machine Learning: GLM, Random Forest, SVM, PCA, CNN, LSTM

Data Visualization: Tableau, Matplotlib, Seaborn, ggplot2, plotly

Data Science: A/B testing, Hadoop, Kaggle

WORK EXPERIENCE

Research Assistant, The Johns Hopkins Data Science Lab Baltimore, MD | Nov 2019 – Jan 2020

Survival analysis of time-series data using Python, R

- Cleaned National Health and Nutrition Examination Survey (NHANES) data using **dplyr**, **tidyverse**
- Reduced dimensionality of data using **PCA** to capture essence of the data
- Selected features using **tree-based model**, **AIC/BIC** to achieve better predictive performance of model
- Constructed a spectral-based convolutional neural network (**CNN**) on 3000 patients using **Keras** to predict mortality with 71% accuracy
- Improved mortality prediction accuracy to 86.45% using **regularized logistic regression**
- Hosted **R shiny** website comparing **PCA**, **k-means**, **UMAP**, **t-SNE** and visualizing clustering results using **ggplot2**, **plotly** (demo: https://luchaoqi.github.io/Shiny_clustering/#1)

Data Analyst Intern, The Johns Hopkins Bloomberg School of Public Health Baltimore, MD | May 2019 – Aug 2019

Association analysis between lifestyle patterns and body mass index (BMI) via generalized linear model

- Wrangled time-series data of 32971 subjects and built pipeline to front-end dashboard using **MySQL**
- Explored user distribution on **Hadoop** using **MapReduce** to maximize the dataset's value
- Trained a generalized linear model (**GLM**) to predict user BMI with 46.07 mean squared error (MSE)
- Reduced prediction error by 13% using **ANOVA** and feature engineering method (**normalization, Random Forest**) through 10-fold **cross-validation**
- Identified statistically significant (p-value < 0.5) impact of lifestyle patterns on BMI to encourage the performance of multiple good health behaviors

SELECTED PROJECTS

Reinforcement Learning on Super Mario Bros (NES) Mar 2020 – Apr 2020

AI that learns to play Super Mario Bros using Deep Q-Network (DQN) in TensorFlow
Demo: https://github.com/LuchaoQi/Reinforcement_Learning

- Built **reinforcement learning** environment using **OpenAi Gym** and emulated NES using **nes-py** in Python
- Constructed a convolutional neural network (**CNN**) model with 5 hidden layers as an agent in **TensorFlow**
- Trained the agent using **deep Q-learning** and reduced training time by 20% using **Adam** optimizer
- Completed different levels of Super Mario Bros successfully without death which was twice as fast as averaged human players

Amazon Rating Prediction June 2019 – Aug 2019

Detection of suspicious or fake Amazon product reviews using machine learning in Python
Demo: https://www.kaggle.com/luchaoqi/making-predictions-over-amazon-recommendation-data

- Extracted Amazon Food Reviews data from Kaggle and cleaned data using **pandas**, **numpy** and **dfply**
- Tokenized unstructured text of user reviews using **NLTK** for feature construction
- Converted text to vector using **bag-of-words model (uni-gram/bi-gram)** with **scikit-learn**
- Predicted customer ratings using **logistic regression** with 0.94 AUC
- Reduced prediction error by 3% using **random forest** to improve detection of abusive reviews

Investigation of Yelp User Funnels, Key Performance Indicators (KPIs) Jan 2019 - Mar 2019

Performance analysis of Yelp users & restaurants using SQL

Demo: https://github.com/LuchaoQi/Yelp_Data_Set_SQL

- Wrote **web crawler** to scrape and parse unstructured data from Yelp using **Xpaths**, **BeautifulSoup** in Python
- Created a database using **MySQL workbench** and imported ~10 GB data file into the database
- Visualized geographic distribution of restaurants with average ratings using **Tableau**
- Performed metrics analysis (**bracket retention, DAU/MAU**) using SQL to measure customer engagement and making suggestions for ways to improve upon KPIs via **A/B testing**

PUBLICATIONS

1. **Qi L**, Zhang Q, Tan Y, et al. Non-contact High-frequency Ultrasound Microbeam Stimulation: A Novel Finding and Potential Causes of Cell Responses. *IEEE Trans Biomed Eng* 2019.
2. **Qi L**, Zhang Q, Lam KH, et al. Calcium fluorescence response of human breast cancer cells by 50-MHz ultrasound microbeam stimulation. Presented at 2017 IEEE International Ultrasonics Symposium (IUS), 6-9 Sept. 2017 2017.