Luchao Qi

Research Data Scientist

Baltimore, MD
+1 (443) 839 9129
lqi9@jhu.edu
luchaoqi.github.io
Linkedin: https://www.linkedin.com/in/luchaoqi/

Skills

Programming	Python, R (Shiny), SQL, Linux	Visualization	Tableau, Matplotlib, Seaborn, ggplot2, plotly
Packages	NumPy, Pandas, NLTK, scikit-learn, Keras, PyTorch, TensorFlow	Machine Learning	GLM, Random Forest, SVM, PCA, CNN
Data Science	A/B testing, Hadoop, Kaggle		

Education

2018–2020 **M.S.E.**, *Biomedical Data Science*, Johns Hopkins University, Baltimore, MD.
Associations between Body Mass Index (BMI) and Accelerometer Time Series Data: National Health and Nutritional Examination Survey (NHANES) 2005-2006
Advisor: Brian Caffo, PhD

2014–2018 **B.S.**, *Biomedical Engineering*, Northeastern University, Shenyang, Liaoning.
Calcium fluorescence response of human breast cancer cells by 50-MHz ultrasound microbeam stimulation
Advisors: Professor Kwok Ho Lam, Professor Ming Qian, Professor Kun Yu

Professional Experience

2019–2020 **Research Data Scientist**, *The Johns Hopkins Data Science Lab*, Baltimore, MD.
Association Analysis Between Lifestyle Patterns and Body Mass Index (BMI).
Cleaned data from National Health & Nutrition Examination Survey (NHANES) by using deplyr and tidyverse.
Performed principal component analysis (PCA) to reduce data dimensionality.
Trained a generalized linear model (GLM) to predict user BMI with 46.07 mean squared error (MSE).
Reduced prediction error by 13% using nested ANOVA (F-tests) on principal component groupings through 10-fold cross-validation.
Identified statistically significant (p-value < 0.5) associations between BMI, age, race, and physical activity level to encourage multiple healthy behaviors.

2019 **Data Analyst Intern**, *Johns Hopkins Bloomberg School of Public Health*, MD.
Survival Analysis of Accelerometer Time-Series Data.
Wrangled time-series data of 32971 subjects & built a pipeline to front-end dashboard using MySQL.
Explored user distribution on Hadoop with MapReduce to maximize data value.
Constructed a spectral-based convolutional neural network (CNN) on subjects using Keras to predict mortality with 71% accuracy.
Improved mortality prediction accuracy to 86.45% using regularized logistic regression.
Hosted R Shiny website comparing machine learning algorithms (PCA, k-means, UMAP, and t-SNE); visualized clustering results using ggplot2 and plotly. demo: `https://luchaoqi.github.io/Shiny_clustering/#1`

2016–2018 **Research Associate**, *Paul C. Lauterbur Lab*, Shenzhen, CHINA.
EMG Signal Pattern Recognition for Hand Gestures Using Spectral Analysis.
Designed, constructed and assembled EMG data acquisition system for recognition of arm activities.
Denoised time domain signals of 200 gestures using Fast Fourier Transform.
Classified different hand movements using support vector machines (SVM) with 82% accuracy.
Improved accuracy by 3% in training a neural network, providing insight for medical rehabilitation systems.

Projects

Mar–Apr, **REINFORCEMENT LEARNING: Super Mario Bros (NES)**.
2020 **AI that Learns to Play Super Mario Bros Using Deep Q-Network (DQN) in TensorFlow.**
Demo: https://github.com/LuchaoQi/Reinforcement_Learning
Built a reinforcement learning environment using OpenAI Gym and emulated Nintendo Entertainment System using Nes-Py.
Designed 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 various levels of Super Mario Bros successfully without "death," achieving twice the speed of averaged human players.

Sep–Dec, **NATURAL LANGUAGE PROCESSING: Amazon Rating Prediction**.
2019 **Use of Machine Learning to Detect Fake or Abusive Amazon Product Reviews.**
Demo: https://www.kaggle.com/luchaoqi/amazon-review-rating-prediction
Extracted Amazon Food Review data from Kaggle; cleaned data using pandas, NumPy and dfply in Python.
Tokenized unstructured text of user reviews using NLTK for feature construction.
Converted text to vector using bag-of-words model (unigram/bigram) with scikit-learn.
Predicted customer ratings using logistic regression with 0.94 AUC.
Improved bad review detection by 3% to find abusive entities (sellers & reviewers) via random forest.

Jan–Mar, **INVESTIGATION OF YELP user funnels, Key Performance Indicators (KPIs)**.
2019 **Yelp User & Restaurant Performance Analysis Through SQL.**
Demo: https://github.com/LuchaoQi/Yelp_Data_Set_SQL
Programmed a web crawler to scrape / parse unstructured data from Yelp using Xpaths & BeautifulSoup.
Created a database using MySQL Workbench and imported 10 GB data file into the database.
Visualized geographical distribution of restaurants with average ratings using Tableau.
Created metrics (bracket retention, DAU/MAU) to measure customer engagement; suggested methods to improve upon KPIs via A/B testing.

Software

R Packages

MRIPCA: Principal component analysis (PCA) on MRI data.
MRIcloudT1volumetrics: T1 volumetric analysis of MRIcloud output.

R Shiny Web Applications

Clustering analysis using K-means, PCA, T-sne, and Umap,
https://github.com/LuchaoQi/Shiny_clustering.
BMI Calculator,
https://luchao-qi.shinyapps.io/BMI_Calculator/.

Publications

[1] Luchao Qi, Qi Zhang, Kwok Ho Lam, Ruibiao Guo, Runkang Chen, Jiangbang Huang, Ruoyan Meng, Zhaoyang Wang, Hairong Zheng, and Ming Qian. Calcium fluorescence response of human breast cancer cells by 50-mhz ultrasound microbeam stimulation. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–3. IEEE, 2017.

[2] Luchao Qi, Qi Zhang, Yan Tan, Kwok-ho Lam, Hairong Zheng, and Ming Qian. Non-contact high-frequency ultrasound microbeam stimulation: A novel finding and potential causes of cell responses. *IEEE Transactions on Biomedical Engineering*, 2019.