LUCHAO QI

Research Data Scientist

Baltimore, MD 21218

lqi9@jhu.edu
443-839-9129

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

SKILLS

Programming: Python, R, SQL, Linux

Packages: NumPy, Pandas, NLTK, Keras, PyTorch, TensorFlow

Data Visualization: Tableau, Matplotlib, Seaborn, ggplot2

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

Data Science: A/B testing, Hadoop, Kaggle

EDUCATION

THE JOHNS HOPKINS UNIVERSITY, Baltimore, MD

M.S.E. Degree in Biomedical Data Science / Biomedical Engineering: May, 2020 **GPA: 3.7**

NORTHEASTERN UNIVERSITY, Liaoning, CHINA

B.Eng. Degree in Biomedical Engineering: August, 2018 **GPA: 3.9**

WORK EXPERIENCE

THE JOHNS HOPKINS UNIVERSITY, Data Science Lab, Baltimore, MD

Research Data Scientist May, 2019 – Apr, 2020

Association Analysis Between Lifestyle Patterns & Body Mass Index (BMI)

- Used DPLYR and TIDYVERSE to clean data from National Health & Nutrition Examination Survey.
- 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.

THE JOHNS HOPKINS UNIVERSITY, Bloomberg School of Public Health

Data Analyst Intern Jun – Aug, 2019

Survival Analysis of Accelerometer Time-Series Data

- Wrangled time-series data for 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 PCA, k-means, UMAP, t-SNE; visualized clustering results using ggplot2 and plotly. (demo: https://luchaoqi.github.io/Shiny_clustering/#1)

SIAT, PAUL C. LAUTERBUR LAB, Shenzhen, CHINA

Senior Researcher Nov, 2016 – May, 2017

EMG Signal Pattern Recognition for Hand Gestures Using Spectral Analysis

- Designed, constructed and assembled EMG data acquisition system for recognition of arm activities.
- Converted time domain data of 200 gestures into frequency domain using Fast Fourier Transform to help denoise signal.
- Classified different hand movements using support vector machines (SVMs) with 82% accuracy.
- Improved accuracy by 3% in training a neural network, providing insight for medical rehabilitation systems.

LUCHAO QI

PROJECTS

REINFORCEMENT LEARNING: Super Mario Bros (NES) Mar – Apr, 2020
Developed 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.

NATURAL LANGUAGE PROCESSING: Amazon Rating Prediction Sep – Dec, 2019
Use of Machine Learning to Detect Fake or Abusive Amazon Product Reviews
Demo: https://www.kaggle.com/luchaoqi/making-predictions-over-amazon-recommendation-data

- 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.

INVESTIGATION OF YELP user funnels, Key Performance Indicators (KPIs) Jan – Mar, 2019
Developed 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.

PUBLICATIONS

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.

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.