|  |  |  |
| --- | --- | --- |
| **LUCHAO QI**  Research Data Scientist | Baltimore, MD 21218  [lqi9@jhu.edu](mailto:lqi9@jhu.edu)  (443)-839-9129 | <https://www.linkedincom/in/LuchaoQi/>  <https://luchaoqi.github.io/>  <https://github.com/LuchaoQi> |

# Skills

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

# Education

|  |
| --- |
| JOHNS HOPKINS UNIVERSITY, Baltimore, MD |
| Master of Science in Engineering Degree -- Biomedical Data Science: GPA: 3.6/4.0 May, 2020 |
| NORTHEASTERN UNIVERSITY, Shenyang, Liaoning |
| Bachelor of Science Degree -- Biomedical Engineering: GPA: 3.8/4.0 May, 2018 |

# Professional Experience

|  |
| --- |
| THE JOHNS HOPKINS DATA SCIENCE LAB, Baltimore, MD  **Research Data Scientist** May, 2019 - April, 2020  **Association Analysis Between Lifestyle Patterns and Body Mass Index (BMI).**   * Processed 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; * Reduced prediction error by 13% using nested ANOVA (F-test) on principal component groupings; * Identified statistically significant (p-value < 0.5) associations between BMI, age, race, and physical activity level to encourage multiple healthy behaviors. |
| JOHNS HOPKINS UNIVERSITY, Bloomberg School of Public Health, Baltimore, MD  **Data Analyst Intern** Summer, 2019  **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. |
| PAUL C. LAUTERBUR LAB, Shenzhen, CHINA  **Research Associate** November, 2016 - May, 2017  **EMG Signal Pattern Recognition for Hand Gestures Using Spectral Analysis**   * Designed, constructed and assembled an EMG data acquisition system to recognize 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. |

**LUCHAO QI**

# Projects

|  |
| --- |
| **REINFORCEMENT LEARNING: Super Mario Bros (NES)** March - April, 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; 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 2X averaged speed human players. |
| **NATURAL LANGUAGE PROCESSING: Amazon Rating Prediction** September - December, 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. |
| **INVESTIGATINGF YELP user funnels, Key Performance Indicators (KPIs)** January - March, 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; * Developed a database using MySQL Workbench and imported 10 GB data file into the database; * Visualized geographical distribution of restaurants with average ratings using Tableau; * Designed 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: Volumetric analysis of MRIcloud output.  R Shiny Web Applications  Clustering analysis using K-means, PCA, T-sne, and Umap:  BMI Calculator: | <https://github.com/LuchaoQi/Shiny_clustering>  <https://luchao-qi.shinyapps.io/BMI_Calculator/> |

# Publications

|  |
| --- |
| [1] Qi, Luchao, et al. "Non-contact High-frequency Ultrasound Microbeam Stimulation: A Novel Finding and Potential Causes of Cell Responses." IEEE Transactions on Biomedical Engineering (2019).  [2] Qi, Luchao, et al. "Calcium fluorescence response of human breast cancer cells by 50-MHz ultrasound microbeam stimulation." 2017 IEEE International Ultrasonics Symposium (IUS). IEEE, 2017. |