Hao Li

Phone: +1 (438) 927-8699 | Email: hao.li7@mail.mcgill.ca | GitHub: github.com/lhcaleo | Website: lhcaleo.github.io

Education

Bachelor of Science in Computer Science **September 2016 – May 2020**
McGill University, Montreal, Canada

Skills

Languages: Fluent in English, Intermediate French, Fluent in Chinese
Programming Languages: Java, Python, C++, C, C#, SQL, HTML, CSS, JavaScript, Bash, MATLAB, Assembly language (MIPS)
Tools: Linux, Git, Unit Test, Make file, PApplet, UML drawing, Scikit-learn, TensorFlow, Natural Language Toolkit
Core Courses: Algorithms and Data Structures, Computer System, Software System and Design, Artificial Intelligence, Concurrent Programming, Applied Machine Learning, Natural Language Processing, Probability, Database Systems, Numerical Method

Software Projects

Reddit Text Categorical Classification **October 2019 – November 2019**

- Developed supervised classification models such as **SVC** and **Bayes** Networks to predict which community a Reddit text is from
- The Reddit dataset is balanced and contains 70,000 randomly picked comments that belong to 20 communities.
- **Preprocessed** dataset by tokenization, removing stop words, converting text to lowercase, removing URLs and numbers, n-grams, TF-IDF encoding and normalization. Extracted titles of URLs and stored them as critical information for the text
- Applied parameter tuning with the help of cross-validation and effectively avoided the **over-fitting** issue
- Achieved a high accuracy which is 57.04% using the **Support Vector Machines** implemented from the SciKit learn package
- Ranked **top 5%** in the class Kaggle competition among more than 100 teams while leading a team of 3 people

Binary Classification with Linear Models in Machine Learning **August 2019 – September 2019**

- Implemented Logistic Regression using gradient descent and Linear Discriminant Analysis models from scratch **in Python**
- Performed binary class classification on two datasets while each data set contains more than 10 discrete features
- Applied k-fold **cross-validation** with randomly shuffled training and testing sets to unbalance the distribution of classes
- The accuracy of Linear Discriminant Analysis model is **94%** which is about 10% higher than Logistic Regression Model

Interactive Earthquake Map of the World **July 2019 – August 2019**

- Built an **Applet** of a 2D earthquake map of the world using Processing (**Java**) graphics methods and Unfolding Maps library functions to visualize earthquake data
- Collected recent earthquake and city data from live RSS feeds and plotted corresponding land quake, oceanic quake, and city markers on the map depending on parsed properties with a key besides the map
- Constructed markers with different shapes and colors according to their magnitude, category, depth and time
- Created abstract classes and interfaces to organize all marker classes for class hierarchy and polymorphism
- Built **event handlers** to interactively respond to mouse input to hide and show city markers affected by an earthquake or all the earthquakes that can threat the clicked city, to keyboard input to switch among three map providers

Enhanced Copy Button and Selection Search **June 2019 – September 2019**

- Built a Chrome extension to automatically create copy buttons for tags that represent code blocks in GitHub and Stack Overflow
- Combined **Bootstrap** tooltips with **CSS** code to display different button styles with the help of mouse event listeners and **jQuery**
- Enabled users to search selected text in multiple search engines such as Wikipedia and Bing in the form of new browser window to enhance fast search and user experience
- Created context menu items for user to choose search engines or read out the selected text with the support of text-to-speech API
- Designed a popup HTML page that shows a quick user guide of the extension when the icon on the toolbar is clicked

Image Convolution **December 2018 – January 2019**

- Created various image effects such as edge and line detections by performing convolutions on input image
- Implemented different kernel computations to compute every output pixel value without boundary problem and data races
- Improved program performance efficiently by 53% by implementing multithreads (8 threads) in **Java**

Activity

McGill Wave Soccer Club **September 2016 – May 2020**

- Managed team training sessions on campus and participated in three seasons of Soccer Intramurals of McGill