EDUCATION

Boston University, Boston, MA September 2021-December 2022

Master's Degree in Computer Science -- GPA:3.7/4

• Major Courses: Advanced Database Management, Machine Learning, Data Mining, Data Analysis and Visualization with R, Big Data Analysis

Shanghai DianJi University, Shanghai, China September 2016 - May 2020 Bachelor's Degree in Computer Science -- GPA:3.4/4

Major Courses: Database Foundation, C++Programming, Data structure and algorithm, Cloud Computing, Internet of Things Engineering

RESEARCH EXPERIENCE

Lichtman Laboratory, Harvard University, Boston Position: Research Assistant, Neuroscience Period: January 2023-Pro

Overview:Our research focuses on studying the development of the cerebellum in mice across four key developmental stages: P0, P3, P7, P10 and P14. My responsibility in this project centers on the application of machine learning techniques to analyze Purkinje cells within the Electron Microscopy (EM) images of cerebellar tissue. I trace granule cell axons and their synapses, which involves significant pre-imaging preprocessing and subsequent data extraction and analysis from EM images.

Publication: In prepare

1. Image Segmentation: I performed image segmentation on the EM images using U-net implemented in Python.

2. Error Detection and Repair: I developed an efficient system for detecting and rectifying errors in the segmented images, manu ally correcting anomalies, such as displacement or distortion, that were unidentifiable by the model.

3. **Image Labelling and Data Extraction**: I labelled cells, axons, and synapses in the image, and from these labels, I was able to extract specific data related to each entity and layer.

4. **2D Image Error Correction**: I implemented error detection and repair mechanisms on the extracted 2D image data to ensure the highest accuracy in our analyses.

5. **3D** Image Creation and Analysis: By overlaying 2D images, I successfully created a 3D image and employed markers in the 2D images to analyze the data in the 3D representation.

6. **Data Analysis and Predictive Modeling**: Analyzed axon clusters in different spatial locations. Utilizing these data, I trained a model to predict axon and cell propensity and distribution, and identified axon and cell clusters in the dataset.

7. **Model Training Methodology**: Gained extensive practical experience in model training methods, utilizing various techniques such as K-means clustering, DBSCAN, and graph theory, among others.

Overview:Reconstructing neural circuits is a crucial technique in connectomics research, mainly accomplished through near-nanometer resolution electron microscopy (EM) imaging of ultrathin sections of brain tissue. In this project, my responsibility is to stitch together the image tiles of interest collected from EM into a complete 2D image and then align the 2D images to achieve a 3D reconstruction.

1. **Feature Extraction**: First, I use the ORB feature extraction algorithm on each image tile. Despite the lower accuracy of ORB, it computes quickly and can produce satisfactory results if enough good correspondences are found.

2. Alternative Feature Extraction: If ORB cannot find enough correspondences in some image pairs, I switch to the SIFT or SURF

algorithms. Although these two algorithms are slower, they perform better in terms of accuracy. To accelerate matching, I apply a FLANN-based matcher to SIFT and SURF.

3. **Global-Perception Outlier Rejection**: To maintain a reasonable quantity of output correspondences, I introduce a learning-based global-perception outlier rejection method. This method can handle varying amounts of correspondences.

4. **RANSAC Method**: Next, I use the classical RANSAC method to further detect and exclude outliers.

5. **Image Stitching**: Finally, based on the obtained correspondences, I can stitch these image tiles together to form an overall 2D image.

PROJECT EXPERIENCE

Boston University, Boston

A Predictive Analysis of Heart Disease base on R

Overview:This project is a heart disease data analysis based on R language. In this project, I performed an analysis on a dataset containing more than 4000 cardiac patient records using the R programming la nguage. The main goal is to explore the association between heart disease and other related factors.

- 1. Leaded the research and creation of a for detection to better predict and prioritize heart disease problems among more than 40000 patients, and ensure the success rate of data prediction in the model
- 2. Used python to perform data ELT and data processing on patient data

3. Various classification models are trained on the training set to determine which model has the highest accuracy. We compared the accuracy of logistic regression, KNN, SVM (support vector machine), naive Bayesian classifier, decision tree and random forest

4. The experimental results show that the random forest algorithm can produce the highest accuracy. Now we can correctly diagnose patients and provide them with the help they need for rehabilitation

League of Legends Streaming Database Management System

Period:September 2022-November 2022

Overview: The Database Management System project focuses on creating an efficient system for managing and analyzing real-time game data. By integrating multiple technologies and database management principles, this system enables seamless data streaming, storage, and analysis for League of Legends gameplay. The project encompasses various components, including data ingestion, data processing, real-time analytics, and interactive visualization.

- 1. Python integration with the League of Legends API for seamless data extraction and import.
- 2. Thorough data cleaning and transformation processes to convert semi-structured data into a structured format.
- 3. Real-time data transmission to a relational database on a cloud platform, enabling efficient data storage and retrieval.

4. Implementation of advanced functionalities within the database such as stored procedures and views, facilitating faster and more streamlined data querving.

5. Comprehensive data analysis of the stored information to gain valuable insights and identify trends within the game.

Spark-based Fitness Data Analysis on Google Cloud Platform

Period:September 2021-November 2021

Overview:The main focus was on analyzing the sports user data from Endomondo. The objective was to divide the users into different clusters based on various attributes and conduct analysis within each cluster. Additionally, a relationship analysis was performed to explore the impact of different factors on oxygen saturation and heart rate during exercise. The outcomes of the project provide valuable insights into fitness patterns, correlations, and user segmentation, contributing to a deeper understanding of the factors influencing exercise outcomes.

- 1. Utilized Spark and Google Cloud Platform to perform in-depth analysis on Endomondo's sports user data.
- 2. Conducted clustering analysis to group the data into different clusters based on various attributes.
- 3. Examined the heart rate differences between genders during exercise and compared other attributes such as height.
- 4. Calculated the correlation coefficients between heart rate and altitude, as well as heart rate and speed, using the Pearson distance formula.
- 5. Applied the K-Means clustering algorithm to categorize the data based on marking coordinates.
- 6. Employed data visualization techniques to present the findings and provide a clear and understandable classification of the data.

Shanghai DianJi University, Shanghai, China

Intelligent lighting system based on Zigbee - Graduation Thesis

Period: March 2019-January 2020

Overview:

I have successfully designed and integrated a Zigbee and WiFi-based IoT lighting system, incorporating various technologies such as Zigbee communication, WiFi connectivity, light sensors, and motion sensors, to achieve automated control and remote operation of the lighting system. Through a combination of hardware and software development, I have demonstrated strong skills in embedded system design and circuit development. My solution offers efficient energy management and a user-friendly interface.

1. **Integration of Hardware and Software**: Through embedded system development and circuit design, I effectively combined the CC3530 micro-controller with software, resulting in the successful design and development of a feature-rich and reliable smart lighting system.

System Functionality Design: I designed the system with features such as automatic control, energy optimization, remote control, and user interface, allowing the lighting system to intelligently adjust illumination based on ambient conditions and motion detection, providing flexible lighting solutions.

3. **Network Communication and Data Processing**: I established ZigBee and WiFi networks to enable wireless communication between internal and external devices. ZigBee is used for transmitting sensor signals, while the WiFi module connects to cloud platforms. Additionally, I performed data processing and analysis of sensor signals to ensure accurate control and feedback.

4. User Interface Design: I developed a user-friendly interface using C++, enabling users to remotely control the lighting system and adjust brightness, modes, and scheduling through a PC or smart devices.

5. **Innovative Solution Approach**: By creatively combining Zigbee and WiFi technologies, and utilizing the capabilities of light sensors and motion sensors, I achieved automation in the smart lighting system.

6. Implementation and Results: I successfully implemented the design and development of the smart lighting system, creating a feature-rich and reliable system. The system effectively improves energy efficiency and provides users with convenient lighting control experiences.

PROFESSIONAL EXPERIENCE

Premium Guard, Inc., Boston

Position: Data Scientist Intern

Period:May 2021 - August 2022

- Used Power BI/Report Builder to generate standard or customized business intelligence presentations and increased revenue by 15% in the first quarter
- Created and documented ETL processes in Python and SQL, reviewed and cleansed client data and reduced data processing time by 50%
- Leveraged R and Python to identify key trends and insights in customer behavior and presented actionable recommendations
- Developed a customer segmentation model using Python to improve customer retention by 20%
- Regularly communicate status and progress made on projects, developments towards clients and senior managers, ensured the accuracy and punctuality of the project

Dell EMC, Shanghai

Position:Business Intelligence Intern

Period:January 2020 - August 2020

- Led a team of 8 interns, effectively coordinating their work and ensuring the success of the project.
- Served as DBA in the completion of 20+ medium- to large-scale implementations, including Bank of America, Citibank, Bank of Japan, etc., managing projects from business requirements analysis to solutions delivery and support

- Used SQL, Dell Report Tool to compile monthly 40+page summarize of company operations, Identifying key areas for improvement to senior management
- Converted and loaded data for VMax and Power Max 2000 products into target database, allowing user corresponding solutions to happen on time and according to specifications and budget

Technical Skills

- SQL: Proficient in SQL Server and PostgreSQL, with experience in writing complex queries, creating views, and designing databases to support data analysis and reporting.
- Python/C++: Experienced in using Numpy, Pandas, Spark, Matplotlib, rpy2, and Hadoop to manipulate and analyze large datasets, as well as create data visualizations and predictive models.
- R: Proficient in using R to conduct statistical analysis and create visualizations.
- Matlab: Experienced in utilizing Matlab for data analysis and modeling.
- Business intelligence: Skilled in implementing business intelligence, including data cleansing, modeling, mining, and ETL processes.
- Software: Experienced in using a range of data analysis and visualization tools, including World, Power BI, Report Builder, Excel, Tableau, Microsoft Access, and SPSS.

Scholarship and Awards

Scholarship:

Erudite Scholarship - Shanghai DianJi University,2018 Erudite Scholarship - Shanghai DianJi University,2020

Sports competition awards:

Sixth place in Shanghai University Football League - Shanghai DianJi University,2019