Chaoyang Zhu

Mobile: 979-985-7809 • **Email:** chaoyang@tamu.edu • **GitHub:** https://github.com/chaoyang-project

Objective: Seeking Summer/Fall 2021 Internship in Software Engineering

EDUCATION

Texas A&M University College Station, Texas

- Master of Science in Computer Engineering GPA: 3.88 08/2019 - 05/2022

Nanjing University of Information Science & Technology Nanjing, China

- Bachelor of Engineering in Internet of Things Engineering GPA: 3.86 09/2015 - 06/2019

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, C, C++, JavaScript, HTML
- **Packages:** sklearn, nltk, keras, numpy, pandas
- **Frameworks and Technologies:** CSS, Bootstrap, Node.js, MongoDB, MySQL, QT, Linux, AWS Cloud9
- **Software:** Jupyter Notebook, Pycharm, MATLAB

RELATED GRADUATE COURSE

- Data Mining and Analysis
- Machine Learning
- Analysis of Algorithms
- Information Storage and Retrieval
- Artificial Intelligence
- Software Engineering

CSCE 633 MACHINE LEARNING Fall 2020

Classify flower types from the UCI IRIS dataset

- Data exploration: compute each class number, plot the histogram of each feature and scatter plots of all pairs of features
- Implement a KNN classifier from scratch using the l2 norm as a distance measure to classify between the three classes
- Plot the accuracy of the model against with the different K values, choose the best K*

Machine learning with Pokemon GO

- Data exploration: compute the Pearson's correlation coefficient between the numerical attributes and the combat points
- Pre-processed the categorical attributes with the one hot encoding
- Implement a Linear Regression model to predict the combat points, use the 5-fold cross-validation to evaluate the model

Machine learning for facial recognition

- Visualize the grayscale images for each emotion
- Use the FNN and CNN to perform the emotion classification task respectively

Classify benign vs malignant tumors

- Data exploration: plot the histograms of the class and each feature, compute the number of each class
- Implement a function that computes the conditional entropy of each feature
- Use a Decision Tree and Random Forest classifier to classify between benign and malignant tumors based on their features

Perform COVID-19 diagnosis from chest Xray images

- Crop and resize the images, extract HoG features
- Train SVM and Random Forest classifier with that features
- Improve the model by VGG-19 neural-network under the guidance of the research papers

CSCE 670 INFORMATION STORAGE AND RETRIEVAL Spring 2020

Implement the Boolean Retrieval engine

- Tokenize entities and definitions using whitespaces and punctuations as delimiters
- Remove stop words and stemming with nltk package
- Build an inverted index to support Boolean Retrieval
- Rank the documents with sum of TF-IDF scores, vector space model with TF-IDF, and BM25 respectively

Find significant Twitter users

- Build a re-tweet graph by parsing the tweets in the dataset
- Implement PageRank algorithm to find the top ten "impactful" users with highest scores

Recommendation System

- Implement Matrix Factorization to predict ratings on MovieLens dataset, evaluate the model by computing the MAE and RMSE value on the testing dataset
- Use a BPR package to experiment with top-K item recommendation on a Spotify playlist recommendation dataset, evaluate the results with NDCG

Word Embeddings for Information Retrieval and Query Expansion

- Use the Word2Vec algorithm to generate word embeddings for tokens in the dataset
- Match the query and the document via the cosine similarity between the embeddings of them
- Expand the original query and redo the vector space model via word embeddings

ECEN 758 DATA MINING AND ANALYSIS

Fall 2019

Predict network protocol class from PORT and SIZE dataset

- Visualize the scatter of PORT vs. SIZE for each of the two classes: TCP and UDP
- Use sklearn package for computing prediction accuracy and generating graphical representations of decision trees
- Experience how different splits of dataset affect the final prediction accuracy

PRACTICAL PROJECT

Web Development 08/2020 - present

- Create my personal website using HTML and CSS, host it on GitHub Pages
- Design a beautiful startup website exhibiting an application using Bootstrap
- Develop a local blog system by Node.js and EJS, connect it with MongoDB database
- Develop a todoList web application by Node.js, deploy the server with Heroku and the database with MongoDB Atlas.

EXPERIENCE

Tarena 07/2018 - 08/2018; 01/2019-04/2019

- Received intensive and systematical training of Java and C++ programing
- Developed an E-reader application on Android Studio by Java. Provided user with different novels on the home page. Provided them with functions of changing the font size and background on the reading pages
- Developed a Pop Star game application on QT framework by C++. Designed a user-friendly gaming interface. Designed the algorithm to pop stars when the condition holds. Added the sound effects and scoring function when stars are popped.
- Implemented a client program speaking to a server by C++. Realized the functions of requesting data points, requesting binary files, and requesting new channel creation. Run the server process as a child of the client process.

PATENT

- Wei Sun, Chaoyang Zhu, A kind of bus, CHN, Patent No. 201610946647.2, Aug.10, 2018
- Wei Sun, Chaoyang Zhu, A vacant seat seeking system in library, CHN, Patent No. 201620415182.3, Oct.19, 2016
- Xiaorui Zhang, Chaoyang Zhu, A street lamp with auto-brightness setting, CHN, Patent No. 201620403151.6, Oct. 12, 2016

SELECTED HONORS & AWARDS

- The First Prize of Scholarship for Academic Excellence (Top 10%) 2017&2018&2019
- Dean Scholarship of NUIST (Top 0.5%) 2016
- The Second Prize of the 13th Advanced Mathematics Contest of NUIST 2016

WORK AUTHORIZATION

- Eligible to intern in the U.S. with CPT