Zhou Zhang

zhangzhouit@163.com

Educational Experience

South-Central University for Nationalities, Communication Engineering, *Bachelor Degree* 2011.09 - 2015.07

University of Wollongong, Computer Engineering, *Postgraduate Student* 2023.09 - now

- CET-4（555）CET-6（512）

Invention project

- Pedestrian Re-Identification Research Program Paper
- Main Work: The pedestrian re-identification system was successfully launched, and combined theoretical knowledge with practice, a paper was concluded.
- Core Technologies: Representation learning, metric learning, global and local feature extraction, video sequence action generation, GAN, CNN, LSTM

- Unsupervised Medical Image Segmentation
- Main Work: Unsupervised medical image segmentation based on nnUNet, PASS framework.
- Core Technologies: nnUNet, PASS

- Semi-supervised sentiment analysis of movie reviews
- Main Work: Sentiment analysis of semi-supervised joint models based on pytorch, transformer+cnn, transformer+lstm.
- Core Technologies: pytorch, transformer, cnn, lstm

- Read drug discovery papers and propose improvements
- Main work: Utilize integrated models and GAN networks to improve models.
- Core technologies: mlp, gnn, gan, integrated model

- Financial futures quantitative trading
- Main work: Use simple moving average, relative strength index, moving average convergence divergence, Bollinger Bands and other indicators to generate trading signals, introduce LSTM long-term signals, Transformer short-term signals, and then integrate these signals through the voting mechanism , finally forming a combined signal
- Core technologies: voting mechanism, moving average, relative strength index, moving average convergence divergence, Bollinger Bands, LSTM, Transformer

- Music recognition
- Main work: Use the PANNs algorithm to identify music in different scenarios and improve recognition accuracy
- Core technology: Short-time Fourier transform, PANNs algorithm

Work Experience

Zhangjiang Institute of Mathematics, Algorithm Intern 2024.07 - Present

- *Describe*: Name Title NER
- *Main Work*: Analyze requirements, write code, test interfaces, and ensure efficient and stable operation of the system.
- *Result Of The Work*: Build a complete labeling system
- *Core Technologies*: python ltp Fine-tuning large models bert word embedding

- *Describe*: Music recognition
- *Main Work*: Analyze requirements, write code, test interfaces, and ensure efficient and stable operation of the system.
- *Result Of The Work*: Accurately identify different scenes and different versions of music
- *Core Technologies*: Short-time Fourier transform, PANNs algorithm, docker, fastapi

Asia Innovation Group, Java Backend Engineer 2021.10 - 2022.7

- *Describe*: Supply chain system transformation, complete the transformation of supplier login, search, commodity order, category attributes and other functions.
- *Main Work*: Analyze requirements, write code, and test interfaces to ensure efficient and stable operation of the system.
- *Result Of The Work*: Use java spring framework, mybatis framework, mysql database to achieve all requirements.
- *Core Technologies*: springboot mybatis mysql rabbitmq

- *Describe*: Advertising system development, complete various expense report statistics, login, search functions.
- *Main Work*: Analyze requirements, write code, and test interfaces to ensure efficient and stable operation of the system.
- *Result Of The Work*: Use java spring framework, mybatis framework, mysql database to achieve all requirements.
- *Core Technologies*: springboot mybatis mysql rabbitmq

IQiYi, Recommend System Engineer 2018.11 - 2020.04

- *Describe*: Distributed crawler project, associated robot account, randomly publishes some content every day, in order to enrich the content of the community, increase user stay time, ctr and other recommendation indicators, the server implements the background system based on the celery asynchronous framework, and adds an agent thread pool To avoid anti-climbing.
- *Main Work*: Analyze requirements, design and implement the entire crawler framework, daily maintenance and performance optimization of the server, and count the number of articles published every day.
- *Result Of The Work*: Use the celery framework to develop the entire business logic, greatly improve performance through distributed and asynchronous.
- *Core Technologies*: Celery asynchronous multi-threaded concurrent distributed
- *Describe*: Natural language processing project, evaluate the word segmentation tools of iqiyi nlp team through algorithms to see if the word segmentation tools can achieve better results in the field of animation.
- *Main Work*: By comparing a small part of manual word segmentation data and machine word segmentation data, the handwriting algorithm calculates the accuracy rate, recall rate, and F1 value.
- *Result Of The Work*: Evaluated the word segmentation tools popular in the industry and the company's internal word segmentation tools. After comparison, it was found that the company's word segmentation tools can meet the needs.
- *Core Technologies*: sklearn jieba machine learning algorithm
- *Describe*: Recommended index optimization project, through the algorithm to improve the recommended index of the vertical App Bata.
- *Main Work*: At first, I wanted to use the traditional tagging method to build a corpus. Later I found that this method was too time-consuming. I obtained the titles, article introductions, and recommended phrases of all the animation works of Petta through the corpus data center. The above text was segmented and the DBScan algorithm was used. Carry out word clustering, select words that can represent the animation works of the article, use the large-scale word vector data pre-trained by Tencent word2vec to find similar words of the previous clustered words, to achieve the purpose of expanding the vocabulary database representing the animation works, and at the same time combine these words Convert all into word vectors. Use these word vectors to represent animation works. According to the whitelist, the animation works viewed by the members of the group are obtained. The cosine similarity between the animation works viewed by the members of the group and other works is calculated, and works with a high degree of similarity are recommended. Members in the group, follow-up promotion to a small number of users, do AB test, improve the recommendation effect.
- *Result Of The Work*: Use sklearn to develop the entire algorithm logic and improve the ctr index.
- *Core Technologies*: sklearn machine learning algorithm
- *Describe*: Feature monitoring project, through the big data framework spark, redis middleware, monitor training model data, training model results, and play an automatic alarm effect。
- *Main Work*: The model training parameters and training results are sent out through the redis broadcast mechanism, blocking the waiting on the server side, and monitoring the server process to prevent the process from disappearing, sending emails, and the email results are rendered through html and displayed intuitively. Feature monitoring uses chi-square calculation, compares today' s and yesterday' s features, sorts, and divides 20 points into equal parts, checks the 95 confidence interval of the chi-square distribution table, and observes whether it exceeds the confidence interval threshold. If it exceeds, send an email to alert.
- *Result Of The Work*: Effectively monitor the characteristics, prevent the wrong characteristics caused by human factors, and clearly show the training results and the recent CTR trend.
- *Core Technologies*: spark hadoop redis offline task
- *Describe*: Data docking between my department and center department, the world-class production of front-row files and inverted indexes.
- *Main Work*: Communicate the data format, use scala spark to engineer all the logic to ensure the normal output of data, and use protobuf to compress the data, which greatly improves the memory utilization of the couchbase.
- *Result Of The Work*: Ensure the accurate output of data, and use protobuf to optimize performance.
- *Core Technologies*: spark hadoop couchbase protobuf offline task
- *Describe*: Play record items, spark streaming tasks, users will watch some videos in iQiyi App, these records need to be saved and displayed.
- *Main Work*: Designed and implemented the entire code logic, filtering long video, short video, small video, stream task mini batch packaged data, merging data, inserting data, inserting failed data processing, optimizing stream task, stream task monitoring alarm, and solving from upstream kafka Redis clients pile up, distributed write and play records to Couchbase cover the problem, and recursive thinking handles wrong data.
- *Result Of The Work*: The user's playback records for the past six months are recorded to ensure the stable operation of streaming tasks and accurate data calculations.
- *Core Technologies*: spark hadoop kafka distributed singleton mode reduceByKey operator
- *Describe*: Play record transfer projects, spark streaming tasks, transfer part of data to hikv, merge and display new and old data when the capacity of couchbase is limited.
- *Main Work*: Designed and implemented the entire code logic, using Cas optimistic lock to transfer old playback records in Couchbase to hikv, roughly segmenting, de-duplication and merging, maintaining and optimizing streaming tasks.
- *Result Of The Work*: It greatly relieves the pressure on the Couchbase cluster, realizes the dynamic balance of data, and enables the

entire system to operate efficiently.

- *Core Technologies*: spark hadoop cas optimistic lock
- *Describe*: Recall sorting breaks up the construction of back-end services, the efficient recommendation system engine ensures the normal operation of the entire recommendation system, and an AB test platform is also built for algorithm use.
- *Main Work*: Using springboot to design and implement the back-end system code, through the card stream recall, asynchronously read the inverted index in the couchbase, return different data through the connection, recall part of the data, and then sort the recalled data. Mainly realize the xgboost+fm sorting model code, discretize continuous features, and finally input the leaf nodes and discrete features into fm to get the sorting score, and finally realize the mmr algorithm, through the title of different channels, profile data, and edit distance Calculate the degree of difference and break up the data.
- *Result Of The Work*: Optimized the back-end interface, improved the interface speed, and realized the basic logic of a recommendation system.
- *Core Technologies*: springboot xgboost fm edit distance mmr