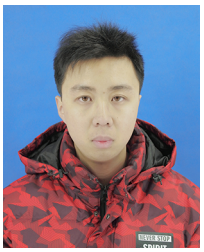
Zhong Junyan

M.U.S.T.,Taipa Weilong Road (Address)
999078 (postcode)

17665458796 (Tel.)
pzsx005@gmail.com (Mail)
github.com/W55699 (GitHub)



EDUCATION

- **M.U.S.T** ·*Applied Mathematics and Data Science* 2019.09 – 2021.08 M.S.
  - Numerical Linear Algebra, Applied Statistics, Machine Learning, Time Series Analysis ,Data Mining, Database Systems
- **ECJTU** ·*Vechile engnieering* 2015.09 – 2019.07 B.S.C
  - Mathematical modeling, Numerical calculation, Microcomputer principle, C++, Automatic control principle, Networke control system of the train

Skills and Research area

- **Computer skills**: Python, C/C++, MATLAB, LaTeX, Java
- **Database management tools**:MySQL,Hive
- **Reasearch area**:Graph neural network, recommendation system application and security, confrontation training

Project experience and internship experience

- **2017 MCM/ICM** *Captain* 2017.02
  - We use the analytic hierarchy process to carry out statistical analysis on the factors affecting the urban living environment, and at the same time analyze the factors affecting the urban living environment and the weight of the factors.
  - We use fuzzy comprehensive evaluation and gray prediction methods to predict the changes of various indicators.
  - We study urban sustainable development models, and after joint discussions within the group, we write the sustainable development papers to describe changes in various indicators.
- **The 6th TaiDi competition** *Captain* 2018.03 – 2018.04
  - Firstly,we cleaned the original data set and used regular expressions to extract the names of TV programs in the original data and categorized the programs. We also implicitly score TV programs based on the user's viewing time and frequency.
  - Secondly,we constructed a user label system table and a product label system table, and performed user portraits based on the time characteristics of user viewing information and we matched programs and program categories to categorize the TV programs.
  - To better implement the recommendation, we use user-base collaborative filtering, item-base collaborative filtering, SVD and hybrid Recommendation algorithm.The SVD algorithm has the best effect.
  - We use K-means method to package users and products, and recommend users without any historical behavior, effectively solving the problems of user cold start and product cold start.
  - After the competition, we optimized the original method and tried to use Text-cnn to recommend the original data , which improved the accuracy of the recommendation.

- **Douban web crawler data analysis** 个人项目 2019.09 –2019.12
  - https://github.com/W55699/doubanbook – web – crawler
  - 利用正则表达式，并通过创建线程池，多线程爬取豆瓣书籍信息。
  - 将信息生成 csv 文件，并将信息存入 mysql 数据库。
  - 利用 pandas 读取 csv，并做数据可视化分析以及统计分析。
  - 通过 pca 将数据进行降维，提取关键信息，然后通过 k-means 算法进行聚类分析。
  - 根据 pca 降维后的信息, 同时结合数据的标记，将数据分为训练集和测试集，并将数据进行二分类，比较各种分类方法如 SVM,LR，决策树，随机森林算法的优劣。
- **IMDB sentiment analysis** 个人项目 2020.10 –2020.12
  - 利用 stop-words 对数据集进行清洗，并通过 wordcloud 进行词云可视化。
  - 利用 python gensim word2vec 对文本进行向量化处理。
  - 训练并调整 bi-lstm 模型，使模型准确率在测试集中达到 85%。
  - 利用 docker,Tensorflow-serving,streamlit 对模型进行部署，实现可视化。
- **AIATSS(友邦资讯科技公司)** 测试组数据分析实习 2020.04-2020.6
  - 撰写 SQL 以及 python 脚本校核公司内部数据。
  - 利用 jira 实时监控工作流程进度，并通过 Excel pivot table 绘制组内测试进度报告。
  - 对测试流程以及 ETL 开发流程有了更深入的了解。
- **TCL 工业研究院** 数据挖掘实习 2020.07-2020.09
  - 通过组内讨论，参与制定推荐系统 CTR 的业务指标，并基于以上指标进行统计分析。
  - 利用 spark 负责数据清洗以及异常数据的核验。
  - 参与组内的论文讨论，并参与大规模特征数据的分类（Random Fourier features SVM)、聚类 (minitach kmeans) 工作，并参与特征筛选以及特征交叉工作。
  - 参与组内爬虫代码的日常维护，丰富自身挖掘经验。
- **品友互动** 策略算法工程师 2021.02-
  - 利用 Hive，以及 Sqoop 等工具等数据进行同步以及清洗（多张数据库表）。
  - 通过业务了解，划分并定义正负样本，并在实际项目中解决小样本训练问题。
  - 利用业务知识对缺失值进行补充，同时在特征工程中对特征交叉，特征构造做出尝试。
  - 利用 LR,xgboost,catboost 对所定义的问题进行分类，优化模型并对样本特征给出可行性解释。

获奖情况

- **2017 美国大学生数学建模竞赛** 二等奖 2017.02
- **第六届“泰迪杯”数据挖掘** 三等奖 2018.04

论文以及专利

- **一种可单轨行走的转向架** 实用新型专利 2017.11
- **一种联网联动小型化多功能晾衣机** 实用新型专利 2017.11
- **一种基于激光检测独立供源的共享单车区域护车系统** 实用新型专利 2017.11
- JUNYAN ZHONG,HUIBIN WANG,KIN TAK U. Matryoshka Attack: Research on an Attack Method of Recommender System Based on Adversarial Learning and Optimization Solution,ICWAPR, 2020.