Zejun Xie

153-2171-9365 | zjx1113@hotmail.com
WeChat: sco_1751
Third year undergraduate | Objective: Data mining, Machine learning

EDUCATION

Renmin university of China *Sep 2017 - Jun 2021*
undergraduate Major:Data science and big data technology School of statistic *Beijing,China*

- Honors/Awards: National project approval of Undergraduate Innovation and Entrepreneurship Programs(Rank 1 in school),President assistant in RSS Statistical Survey Association,Second Price of Contemporary Undergraduate Mathematical Contest in Modeling,2019
- Course Highlights:
  - Math & Statistic: Mathematical Analysis, Higher Algebra, Probability, Mathematical Statistics,Real Analysis,Complex Analysis,Regression Analysis,Stochastic Process;
  - CS & DS : Machine Learning,Deep learning,Statistical Software,Programming, Data structure, Algorithm Design,Database System,Spoken Language Processing;
  - Next Semester proposal: Parallel Computing, Software Design, Combinatorics, Statistical Calculation and Optimization

PSU-PKU Summer school *Jul 2019 - Aug 2019*
Deep learning *Beijing,China*

- Achieved best performance in course project and attained Prof.Jinchao Xu's recommendation.

PROFESSIONAL EXPERIENCE

DiDi global *Aug 2019 - Nov 2019*
Business intelligence intern , International Business Technology *Beijing,China*

- **Supply & Demand Forecasting :** Predict hourly measures of supply & demand for three Mexican cities and six Brazilian city; Produced reports with prediction results on 3 holidays for the operations team to rake necessary precautions; Self-learned time series related knowledge, compared with Seasonal ARIMA models and other methods, and finally selected the most suitable *Prophet* framework; Implemented *Prophet* in Python; Modified daily seasonality for special events and incorporated holiday multipliers to improve performance;5% performance improvement over previous methods
- **News Data Visualization :** Crawl thousands of news about Didi on the large local news site in Brazil; Classify news by LDA model and visualize it on the map; Use *Requests, NLTK* and *Folium* in Python
- **Information & Competition Analytics :** Queried internal data using SQL to evaluate validity and significance of drivers' complaints as well as competitors' new functions found on social media groups every two weeks; Create travel information table for maintaining airport and hotel scenarios, and regularly check for abnormal data

RESEARCH EXPERIENCE

Multi-KPI Causal Graph Discovery *Nov 2019 - Present*
Research intern , Mentor:Dr.Xidao Wen,new postdoc in THU Dan Pei's NetMan LAB *Beijing,China*

- **Project purpose** : There are a number of KPIs in Autonomizing It Opeations(AIOps).Under Dr.Xidao Wen's advisor, I am committed to improving the existing methods to discover the causality of temporal data, expecting to obtain the causal graph faster and more accurately.
- **Project timeline** : Finish the current work in March, submit to conference (e.g. CIKM); The next phase will be combined with AIOps Knowledge Graph for new work

Constrained Linear Data-feature Mapping for Classification *Aug 2019 - Oct 2020*

Research Assistant , Advisor : Jinchao Xu, Penn State University *Beijing,China*

- **Background** : Xu's group unified **MgNet**, that simultaneously recovers some CNN for image classification and multigrid (MG) methods for solving discretized PDEs.(link)
- **Project purpose** : Constrained linear data-feature mapping model as an interpretable mathematical model for image classification using CNN such as the ResNet. (preprint : arXiv:1911.10428)
- **My contribution** : Assist to implement Model and experiment in PyTorch with Python. In the experiment, I obtained the model with the minimum number of parameters with accuracy greater than 99%.

Relevant competition experience(Teamwork) *Sep 2019 - Oct 2019*

- **Airport taxi driver's decision model and feasible arrangement and scheduling of ride system :**
  - Aiming at the question of whether taxi drivers arriving at the airport should choose to queue for the return trip, we use the update process and queuing theory to model and ARIMA model to predict;
  - MCMC method was used for 100000 simulations; Designed a two-lane scheme that prioritizes short-distance orders according to the priority scheme, which can encourage drivers to take short orders;
- **ECG abnormal event prediction :**
  - Multi-label classification of patient symptoms using waveform electrocardiogram data (ECG). According to the characteristics of ECG data, Resnet34 model with 1d-conv was selected;BCEloss and upsampling methods are used to solve the severe category imbalance in the sample data;
  - Entered the finals in the Tianchi platform Hefei ECG artificial intelligence contest, and finally got a top 2% (58/2353) score.

Relevant course project(Teamwork) *Sep 2017 - Present*

Here are some of the course projects I completed in class :

- **Use reinforcement learning to train SpaceShooter game AI : Tensorflow & Pygame**
- **TCN-based vocal and musical instrument separation : Pytorch**
- **Train ticketing systems : Django & mysql**

MISCELLANEOUS

- **Skills:** Python,Pytorch,Tensorflow, R, SQL, C, C++
- **Languages:** English (Fluent), Chinese (Native)
- **Activities:** Sociology, Comic, Movie, Chinese chess
- **Research Interests:** Data Mining, Machine Learning, Natural Language Processing