KAIZHEN TAN

◇ Email: tkz@tongji.edu.cn ◇ Phone: +86 15214390380 ◇ Personal Website: tantansir.github.io

EDUCATION

Tongji University Sept. 2021-Present
Information Management and Information System, **GPA: 90.6/100, 3.94/4.0 (WES)** Shanghai
Relevant Courses: Python/C++/Java Programming, Data Structures, Computer Network, Operating System, Database Technology and Applications, Probability/Statistics, Discrete Mathematics, Data Mining/Data Analysis

RESEARCH EXPERIENCES

Modeling and Optimization of Air Traffic Management at Changi Airport Sept. 2024-Present
*Research Intern, Advisor: Dr. Yicheng Zhang, Dr. Sheng Zhang, A*STAR* *Singapore*

- To establish a behavioral model for air traffic controllers to predict communication tasks and optimize workload using radiotelephony communication and aircraft trajectory data.
- Gather historical and real-time air traffic data from various sources, including online flight tracking systems, flight logs, and weather condition platforms; perform data cleaning to address missing values, inconsistencies, and noise.
- Conduct exploratory data analysis to identify patterns, trends, and anomalies in air traffic; utilize statistical methods and visualization tools to uncover insights into traffic density, peak times, and common routes.
- Analyze factors affecting air traffic, such as weather conditions, airspace restrictions, and airport capacity; develop a CNN and Transformer-based model to test various scenarios and strategies for efficient air traffic management.

Analysis of Tourists' Focal Preferences and Recreational Experience in Historic Urban Quarters based on Deep Learning Jan. 2024-Present
Research Assistant, Advisor: Prof. Yujia Zhai, Tongji University *Shanghai*

- To assess tourists' perception evaluation of historic urban quarters in Shanghai using deep learning methods.
- Crawl travel reviews from social media platforms and use the SAM model for semi-automatic labeling, train a small model for automatic labeling to create the image dataset, and develop an improved semantic segmentation model for focal points extraction by fine-tuning various SOTA models and identifying the best-performing one.
- Use a fuzzy quantization method to categorize building facade colors, compare dominant colors in travel photos actual street views, and analyze tourists' preferences and expectations for urban street color schemes.
- Apply named entity recognition and sentiment word matching algorithms to calculate satisfaction scores for each dimension (Activities, Built environment, Facilities, Business types) for text data annotation, and employ BERT for multi-task learning through supervised fine-tuning, visualizing and presenting the analytical results.

Spatio-temporal Interaction Mechanism of Human Activities and Traffic Congestion Propagation — Project funded by the National Natural Science Foundation of China Apr. 2024-Sept. 2024
Research Assistant, Advisor: Prof. Wei Huang, Tongji University *Shanghai*

- Modeled traffic congestion propagation patterns (TCPP) using geospatial-temporal-semantic knowledge graphs.
- Used map matching algorithm to project Shanghai's taxi trajectory data onto real roads, calculated traffic state index based on speeds in each time slice to identify congested segments, and constructed feature-embedded graphs.
- Utilized an improved community detection algorithm to cluster congested segments, and observed significant shifts in congested subgraphs over adjacent time slices using ARI and NMI to detect TCPP.
- Identified the causality and correlation between TCPP and built environment by using causal inference and calculating weighted values of POIs within each buffer, and quantified the impacts through propagation probabilities.

CFGPT: Chinese Financial Generative Pre-trained Transformer Framework Jan. 2024-Apr. 2024
Research Assistant, Advisor: Prof. Dawei Cheng, Tongji University *Shanghai*

- Established the datasets of CFGPT, an open-sourced Chinese financial large language model (LLM), and contributed to model refinement as a prompt engineer in collaboration with Shanghai AI Lab.
- Collected online financial content using proxy-based distributed crawlers and filtered texts with regular expressions, banned word lists and locality-sensitive hashing algorithm, to assemble a pre-training dataset, including domain-specific financial data and general public data.
- Designed text prompts for further supervised instruction tuning, enhancing the model's generalization ability across 6 specific downstream financial tasks, such as sentiment analysis, topic decomposition and stock prediction.
- Crafted real-world financial application cases to provide textual support for retrieval-augmented generation (RAG).

Pavement Disease Recognition Using Object Detection — Global Campus Artificial Intelligence Algorithm Elite Competition Sept. 2023-Nov. 2023
Member *Nanjing*

- Utilized YOLOv8 model for object detection of pavement diseases, added a detection module for small targets of 4x4 pixels and replaced the original loss function with Wise-IoU, enhancing the model's detection sensitivity, resulting in a 26.3% increase in F1 score and a 25.1% improvement in mAP50-95 on the competition dataset.
- Expanded the dataset through online enhancement of tuning parameters and offline random data augmentation techniques to mitigate the impact of sample imbalance and improve the generalization capabilities and robustness.
- Compared and analyzed different combinations of optimization methods, identified an optimized model encompassing small targets, Wise-IoU, and data augmentation, winning the Second Prize at national level (top 5%).

ACADEMIC PROJECTS

Design of a Medical Information Service Platform Website Jun. 2023-Aug. 2023
Leader *Shanghai*

- Led a team of five members to design and build a medical information service platform, providing authoritative medical advice, comprehensive healthcare information, and personalized treatment recommendations.
- Utilized the Bootstrap framework for frontend layout design and the Django framework (Python web) for backend development, and constructed a database containing information on hospitals, users, and diseases with MySQL.
- Developed an algorithm to provide patients with personalized recommendations regarding hospitals and departments based on their symptoms and the geographical locations of hospitals and patients using the Amap API.
- Completed cloud server deployment and successfully launched the project online.

Agent-based Modeling and Simulation System for Library Seat Selection Jun. 2023-Aug. 2023
Leader *Shanghai*

- Led a team of six members to collect data on the school library environment and reader behaviors through on-site investigation and questionnaire surveys, employing K-means clustering to classify reader agent types and fuzzy analytics hierarchy process (FAHP) to analyze demographic characteristics and decision-making patterns.
- Used Anylogic software to construct a virtual library environment and create agents representing readers and seats; simulate the decision-making process for seat selection based on reader attributes and environmental preferences, and incorporated random events to model changes in the library environment.
- Optimized and validated model inputs and outputs by adjusting reader attributes and seat booking preference parameters for hypothesis testing and model enhancement, achieving a grey correlation degree of 0.87.

INTERNSHIP

Shanghai Qiantan Emerging Industry Research Institute Jan. 2023-Feb. 2023
Data Analyst Intern *Shanghai*

- Used PyQuery and Requests to crawl news headlines from global think tanks, and employed Pandas for extensive data cleaning, transformation, and normalization, addressing missing values and outliers.
- Employed Jieba for word segmentation of news headlines and conducted statistical analysis on high-frequency terms; leveraged various machine learning techniques such as SVM, LSTM, XGBoost, Prophet and GAM to perform predictive modeling and trend analysis; visualized results using Seaborn and Matplotlib.
- Conducted an industrial study of Liaoyang with economic data including local businesses and recruitment information; applied clustering and association algorithms to obtain demographic profile and requirement diagram.

AWARDS

Global Sustainability Supply Chain Student Competition, Second Round Dec. 2023
Global Campus Artificial Intelligence Algorithm Elite Competition, Second Prize Nov. 2023

SKILLS

Languages: IELTS 7.5 (L:9, R:8.5, W:6.5, S:6), GRE 331 (V:161, Q:170, AW:3)
Programming & Tools: Python, C++, C#, Java, HTML, SQL, Git, Docker, Tableau, ArcGIS, Stata
Coursera Certificates: IBM Data Science Specialization, Stanford Machine Learning Specialization