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

PUBLICATION

Huan, W., **Tan, K**., Liu, X., Huang, W. (2024). STALS: A spatiotemporal adaptive local search method for tracking congestion propagation in dynamic networks. Submitted to *Transportation Research Part C: Emerging Technologies*.

RESEARCH EXPERIENCES

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

- Established a behavioral model for air traffic controllers to predict communication tasks and optimize workload using radiotelephony communication and aircraft trajectory data.
- Collected and cleaned historical and real-time air traffic data from various sources, conducted exploratory analysis to identify patterns, trends, and anomalies, and utilized statistical methods and visualization tools to uncover insights into traffic density, peak times, and common routes.
- Analyzed factors affecting air traffic, such as weather, airspace restrictions, and airport capacity; developed a hybrid CNN-transformer 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-Dec. 2024
Research Assistant, Advisor: Prof. Yujia Zhai, Tongji University *Shanghai*

- Assessed tourists' perception of historic urban quarters in Shanghai using deep learning methods.
- Crawled travel reviews from social media, integrated the SAM model for automatic labeling to form a training dataset, and developed a semantic segmentation model for focal point extraction by fine-tuning SOTA models.
- Categorized building facade colors using fuzzy quantization method, compared dominant colors in travel photos with actual street views, and analyzed tourists' preferences and expectations for urban street color schemes.
- Applied named entity recognition and sentiment analysis to calculate satisfaction scores for various dimensions (Activities, Built environment, Facilities, Business types) and employed BERT for multi-task learning.

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.
- Mapped Shanghai's taxi trajectory data to real roads, calculated traffic state index to identify congestion, and built feature-embedded graphs by fusing node curvature, degree, spatial proximity, and semantic information.
- Proposed a novel spatiotemporal adaptive local search (STALS) method to track the propagation of traffic congestion by identifying multi-scale communities in the dynamic adjacency matrices.
- 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 with Shanghai AI Lab.
- Collected online financial content using distributed crawlers, filtered texts with regular expressions, banned word lists, and locality-sensitive hashing algorithm, to assemble a pre-training dataset.

- 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 to build a virtual library and simulate seat selection based on reader attributes, preferences and environmental factors, incorporating random events to model environmental changes for model testing.
- 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.
- Performed word segmentation with Jieba, analyzed high-frequency terms, and applied SVM, LSTM, XGBoost, Prophet, and GAM for predictive modeling and trend analysis; visualized results with Seaborn and Matplotlib.
- Conducted an industrial study of Liaoyang using economic data including local businesses and recruitment information; applied clustering and association algorithms to create demographic profiles and requirement diagrams.

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