XIAOYANG CAO

Beijing, China | (+86)18164258180 | seancao0321@gmail.com | https://xiaoyangcao1113.github.io/

EDUCATION

Tsinghua University **Beijing, China**
B.Sc. in Mathematics and Physics + B.Eng. in Civil Engineering *Expected Graduation: June 2025*
GPA: 3.95 / 4.00
Rank: 3 / 60 (top 5%)
Relevant Coursework: Operation Research **(A+)** | Numerical Analysis **(A)** | Applied Statistics and Data Analysis **(A-)** | Data Structure and Algorithm **(A-)** | Transportation Planning **(A-)** | Principles of Economics **(A)** |Mathematical Physics Equations **(A)** | Quantum Mechanics **(A+)** | Fundamentals of Physics **(A+)**

University of California, Berkeley **Berkeley, CA**
Exchange Student *Jan 2024 – May 2024*
GPA: 3.93 / 4.0
Relevant Coursework : EECS 127 Optimization Models in Engineering **(A+)** | CS 188 Introduction to Artificial Intelligence **(A+)**

PUBLICATION

1. **Xiaoyang Cao**, Zhe Fu, Alexandre M. Bayen, "Pareto Control Barrier Function for Inner Safe Set Maximization Under Input Constraints." *Accepted for presentation at the 2025 American Control Conference (ACC).*
2. **Xiaoyang Cao**, Dingyi Zhuang, Shenhao Wang, Jinhua Zhao, "Virtual Nodes Improve Long-term Traffic Prediction." *Accepted for presentation at the 2025 Transportation Research Board (TRB).*

ACADEMIC RESEARCH

Virtual Nodes Improve Long-term Traffic Prediction
Advised by Professor Jinhua Zhao and Ph.D. Candidate Dingyi Zhuang at MIT *May 2024 – Present*
- Integrated virtual nodes into spatio-temporal graph neural networks (ST-GNNs) to improve long-term traffic prediction accuracy and address the *over-squashing* problem by constructing a semi-adaptive adjacency matrix using learnable node embeddings.
- Validated the virtual node-enhanced model on traffic data in San Diego, which includes 716 sensors and spans five years. The model achieved a 6.27% reduction in RMSE and a 5.04% reduction in MAPE for long-term horizons, with higher weights assigned to road intersections, enhancing model interpretability.
- The model improves interpretability by automatically assigning higher weights in the adjacency matrix to traffic-intensive nodes such as road intersections.
- Preliminary results have been accepted for presentation at ***Transportation Research Board (TRB) Annual Meeting 2025***, and the full manuscript is being prepared for submission to ***Transportation Research Part C: Emerging Technologies***. Preprint: https://arxiv.org/abs/2501.10048

Pareto Control Barrier Function for Inner Safe Set Maximization Under Input Constraints
Advised by Professor Alexandre Bayen and Ph.D. Candidate Zhe Fu at UC Berkeley *May 2024 – Oct 2024*
- Integrated the Pareto Multi-task Learning (PMTL) framework into Neural Control Barrier Functions (NCBFs) to develop the novel Pareto Control Barrier Function (PCBF) algorithm. The PCBF algorithm maximizes the inner safe set while ensuring system safety under input constraints.
- Validated the PCBF algorithm on a 2D inverted pendulum system, where the resulting inner safe set closely matched the viability kernel computed via Hamilton-Jacobi reachability; and succeeded on a 12-dimensional quadrotor obstacle avoidance task, where the inner safe set volume was 700% larger compared to the benchmark method.
- Results have been accepted for presentation at ***American Control Conference (ACC) 2025***. Preprint: https://arxiv.org/abs/2410.04260.

Travel Behavior Prediction by Fusing Socioeconomic Indicators with Satellite Imagery
Advised by Professor Jinhua Zhao and Ph.D. Qingyi Wang at MIT *Jun 2023 – Dec 2023*
- Leveraged image generation model VQ-GAN to extract latent features from satellite imagery. The latent

features were fused with socioeconomic indicators using LASSO to predict travel behavior.
- Applied k-means clustering algorithm to the VQ-GAN codebook to reduce the parameter size and address the *codebook collapse* problem.
- Predicted three types of travel behavior over a one-year period and achieved a 10% improvement in accuracy by using over 30,000 satellite images of Chicago. The codebook size was reduced from 1024 to 8, and its utilization rate was increased from 9.5% to 100%.

Predicting Urban Traffic Accident Duration: Integrating Deep Learning with Survival Analysis
Advised by Professor Ruimin Li at Tsinghua University *Sep 2022 – Jan 2024*
- Developed the Deep Competing-Risk Accelerated Failure Time (DeepCAFT) model by learning a generalized hazard function to capture non-linear relationships between accident duration and influencing factors.
- Introduced the competing-risk methodology to relax the i.i.d. assumption, which proved unrealistic in real-world accident data.
- Validated model on one year of accident data from Beijing's 4th Ring Road, comprising 5,386 records, and achieved a 9% improvement in Mean Absolute Percentage Error (MAPE) compared to traditional models.
- Under review by the journal ***IET Intelligent Transport Systems.***

AWARDS & HONORS

Outstanding Innovation in Science and Technology Scholarship, Tsinghua University *Oct 2024*
Top 15%, *awarded to student demonstrating exceptional performance and innovation in the fields of science and technology*

Comprehensive Excellence Scholarship, Tsinghua University *Oct 2023*
Top 10%, *award for academic excellence, leadership, and extracurricular involvement*

Outstanding Innovation in Science and Technology Scholarship, Tsinghua University *Oct 2023*
Top 10%, *awarded to student demonstrating exceptional performance and innovation in the fields of science and technology*

Outstanding Winner & INFORMS Award, Mathematical Contest In Modeling *May 2023*
Top 1%, *exemplifying superior analytical skills and creativity in a global mathematical modeling competition*

Outstanding Academic Scholarship, Tsinghua University *Oct 2022*
Awarded to students who achieve a GPA within the ***top 10%***

SKILLS & LANGUAGES

Technical Skills:
- **Python**: Proficient in data analysis and building machine learning models using scikit-learn, NumPy, Pandas, and PyTorch.
- **R**: Proficient in statistical tests (stats, car, lmtest), survival analysis (survival, flexsurv), and time series forecasting (forecast, tsibble). Experienced in data manipulation and visualization with dplyr, ggplot2, and tidyverse.
- **Optimization & Simulation:** Proficient in advanced optimization techniques using MATLAB, Gurobi, and Python (SciPy, CVXPY). Experienced in Monte Carlo simulations and dynamic systems modeling with Simulink and Python, applied to nonlinear control and real-time optimization.
- **Research & Report Writing**: Experienced in conducting academic research, authoring scholarly papers, and documenting results for publication.

Languages: English (fluent), Chinese (native)

ACTIVITIES

Latest Achievement in Transportation Engineering and Logistics Winter School **Beijing, China**
Online Course by Tsinghua University and Technical University of Munich *Nov 2022*
- Engaged in a comprehensive online course covering key aspects of transportation systems analysis and logistics, with a focus on e-commerce case studies
- Explored advanced topics including urban multi-level traffic state estimation, control methods, and the environmental impact of transport infrastructure