Xie, Tianyang

tianyangxie@arizona.edu • LinkedIn

RESEARCH INTERESTS

FinTech, Network Science, Econ ML, Fairness & Disparity

EDUCATION

Ph.D. in Management Information System Sep 2021 - May 2026

University of Arizona, USA

M.S. in Statistics Sep 2017 - Dec 2020

University of Minnesota Twin Cities, USA

B.A. in Economics Sep 2013 - May 2017

Shanghai University of Finance & Economics, China

ACADEMIC WORKS

Publications

1. **Tianyang Xie and Yong Ge. Enhance knowledge graph embedding by mixup. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–12, 2023** In this research, we utilize the influence function and triplet mixup techniques to augment the performance of knowledge graph embeddings. The methodology we propose yields substantial improvements for TransE and DistMult embedding methods on FB15K237, WN18RR, and YAGO3-10 datasets.
2. **Tianyang Xie and Jie Ding. Forecasting with multiple seasonality. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 240–245, 2020** In this study, we develop an advanced parallel computing framework designed specifically for modeling multiple seasonal patterns in time series forecasting. This proposed approach demonstrates superior performance compared to established models such as LSTM, TBATS, and the Facebook Prophet model, with this superiority evidenced in both simulated and real-world datasets

Working Papers

1. **Tianyang Xie and Yong Ge. Identifying the power of cryptocurrency airdrop: A double machine learning approach. *In Preparation*** In this study, we use a double machine learning approach to identify the heterogeneous causal effect (CATE) of cryptocurrency airdrop (i.e., free tokens giveaway) on tokens' post-ICO performance. Our work contributes in three ways: 1) we focus on the aggregate causal effect, including the effect on announcement and execution. Unlike the previous studies, our work provides clear implications for the token project developers in suggesting when and how to initiate an airdrop campaign for the ideal post-ICO performance. 2) Unlike the existing ICO and airdrop studies that only consider human-crafted features, we include ML embedding as complex control variables, thus significantly lowering the model-misspecification risk. 3) We portray the problem in a high-dimensional setting and use a double machine-learning (DML) framework to identify the heterogeneous causal effect correctly. This work serves as one of the pioneering causal inference studies for high-dimensional problems in business topics.
2. **Tianyang Xie and Yong Ge. A scalable, end-to-end, and multi-label graph counterfactual explanation framework. *In Preparation*** In this research, we introduce a pioneering framework that effectively tackles the challenges of scalability, end-to-end training, and multi-label explanations within the domain of graph counterfactual explanations. Our methodology not only matches the performance of existing baseline models in the context of factual explanations but also exhibits superior efficacy in scenarios involving multi-label counterfactual explanations.
3. **Tianyang Xie and Yong Ge. Equal opportunity in survival analysis. *In Preparation*** In this research, we tackle the challenge of equalized odds fairness in survival analysis through the introduction of two innovative modules. The first is a trainable time-to-event imputation module, which incorporates imputed time for censored data as parameters within the loss optimization process. The second is a conditional mutual information fairness regularizer, theoretically ensuring equalized odds fairness. By combining the two modules in model training, the method not only achieves fair predictions but also maintains overall performance efficiency across both accelerated failure time and Cox models, applicable in both linear and deep learning frameworks.

4. **Tianyang Xie and Yong Ge. Accuracy, fairness, diversity all at once: An influence function guided data enhancement approach for recommender system.** ***In Preparation*** In this study, we have developed an innovative influence-guided data augmentation approach that simultaneously enhances the accuracy, fairness, and diversity of collaborative filtering recommendations. The comprehensive empirical evaluation has shown that our approach can improve accuracy, fairness, and diversity by up to 24.27%, 55.29%, and 1.85% and significantly outperforms the state-of-the-art baselines on multiple evaluation metrics.
5. **Tianyang Xie and Yong Ge. Mitigating the disparity of juvenile recidivism prediction: A novel data mixup approach with a human-in-the-loop design.** ***In Preparation*** We propose a novel data mixup approach to mitigate the disparity of juvenile recidivism prediction. Our approach consists of three connected components: mixup sample generation, influence function-based candidate evaluation, and virtual sample selection, and it iteratively generates virtual mixup samples to train a fair and accurate prediction model together with the original training set. Furthermore, on top of the iterative process, we design a novel human-in-the-loop (HITL) mechanism to integrate useful human knowledge into the algorithm further. After consolidating a unique U.S. JRP dataset from multiple sources, we evaluate our method against several state-of-the-art baselines, where we consider multiple classification models and evaluation metrics.

INDUSTRIAL EXPERIENCE

Data Scientist | *Metro Transit* Sep 2019 - May 2020

This project, guided by Prof. Qie He in the Industrial and Systems Engineering department, University of Minnesota Twin Cities, focused on assisting Metro Transit, the principal public transportation provider in the Twin Cities, in optimizing their workforce management. My primary responsibility was the development of day-ahead and season-ahead statistical models to predict absenteeism among bus operators. A significant challenge encountered was the discrepancy between individual-level responses and aggregate-level feature data. Traditional predictive statistical methods like logistic regression, SVM, etc., were not directly applicable due to this incompatibility.

To address this, we adapted the Poisson Binomial Distribution (PBD) model, employing a normal approximation technique and implementing the BFGS optimization method. Cross-validation using historical data demonstrated that our model outperformed traditional methods with feature engineering, including GLM, SVM, Decision Trees, Random Forest, and Spline models. The results were highly commended for their valuable insights, contributing substantially to the bus operation's scheduling and strategic planning.

SERVICES & HONORS

T.A. in Deep Learning | *University of Arizona* Jan 2023 - May 2023

T.A. in Applied Regression Techniques | *University of Minnesota Twin Cities* Jan 2019 - Jun 2019

T.A. in Experiment Design | *University of Minnesota Twin Cities* Sep 2018 - Dec 2018