YU YANG

Email: yuyang.stat@gmail.com Webpage: https://yuyangyy.com LinkedIn: yuyangstat

EDUCATION

University of Minnesota, Twin Cities *August 2018 - June 2023*
Advisor: Prof. Xiaotong Shen
Ph.D. in Statistics GPA: 3.980/4 Minneapolis, MN

Shanghai University of Finance and Economics *September 2014 - June 2018*
Bachelor of Science in Statistics GPA: 3.89/4 Shanghai, China

EXPERIENCE

JPMorgan Chase & Co. 06/2022–09/2022
AI & Data Science Summer Associate *New York, NY*

· Developed modules for time series causal discovery

Seagate Technology 09/2019–04/2022
Research Assistant (Advisor: Prof. Xiaotong Shen; Manager: Sthitie Bom) *Minneapolis, MN*

· Explored large-scale and multi-sourced datasets and proposed systematic preprocessing pipelines
· Constructed interpretable predictive models for defected wafer products
· Proposed causal structure learning methods to unveil the causal relations among abnormal events
· Developed a Python package and an R package for causal structure learning

PUBLICATIONS

A hierarchical ensemble causal structure learning approach for wafer manufacturing
Authors: Yu Yang, Sthitie Bom, Xiaotong Shen *Published: 15 August 2023*

· Proposed a hierarchical ensemble approach to learn the causal structure in wafer manufacturing
· Validated the effectiveness through simulation experiments and a practical application involving data obtained from Seagate Technology

Boosting Summarization with Normalizing Flows and Aggressive Training
Authors: Yu Yang, Xiaotong Shen *Accepted by EMNLP 2023*

· Proposed FlowSUM, a normalizing flows-based variational encoder-decoder framework for Transformer-based summarization.
· Proposed a controlled alternate aggressive training (CAAT) strategy and an improved gate mechanism to improve training efficacy.
· Demonstrated that FlowSUM could significantly enhance the summary quality and unleash the potential for knowledge distillation.

PROJECTS

R Package: glmtlp 08/2021–01/2022
Package Developers: Chunlin Li, Yu Yang

· Developed an R package to fit generalized linear models with truncated lasso penalty
· Performed experiments to compare *glmtlp* with other competitors in terms of accuracy and time cost

Causal Discovery for Mixed Data with Temporal and Group Constraints 01/2021–04/2021

· Proposed three causal discovery methods for high-dimensional mixed data with special constraints

· Performed experiments on the simulated data to examine the performance of the three methods

Topic-Aware Abstractive Text Summarization 01/2021–04/2021

· Proposed a new model by marrying Pointer-Generator Networks with Replicated Softmax RBM
· Experimented the model on the CNN/Daily Mail data

Retro-BiDAF: A Retrospective Reader Over BiDAF 10/2020–12/2020

· Proposed a question answering model for the SQuAD 2.0 Challenge
· Examined the idea of retrospective reading in the non-PCE scenario

Kaggle: Lyft Motion Prediction for Autonomous Vehicles 09/2020–11/2020
*Team Members: Xuesong Hou, Chunlin Li, Yu Yang (Ranked **top 6%**)*

· Explored the large-scale image data and visualized the paths of vehicles
· Built an ensemble model upon EfficientNet and DenseNet to predict the motion of on-road objects

Wells Fargo Campus Analytics Challenge 2020 07/2020–08/2020
*Team Members: Xuesong Hou, Chunlin Li, Yu Yang (Won the **Grand Prize** of the year)*

· Identified proper encoding schemes from model fitting details and proposed a top-performing classifier
· Proposed a novel method called Sparse Grouping Pursuit which efficiently reduced feature dimensions

R Package: ImbCalib – Probability Calibration for Imbalanced Data 04/2020–05/2020

· Wrote an R package to calibrate probabilities for imbalanced data
· Compared probability calibrations visually and quantitatively

MinneMUDAC 2019 Student Data Science Challenge 09/2019–11/2019
*Team: Women in Math and Stats (Won the **Analytical Acumen Award**)*

· Collected data from a wide range of sources and applied creative feature engineering
· Built an ensemble model upon XGBoost, LSTM, and VAR to predict the soybean futures closing prices

Learning Rate Decaying Scheme Investigation 11/2019–12/2019
Team Members: Liwei Huang, Yu Yang

· Proposed several learning rate decaying schemes and applied them to MNIST and CIFAR-10
· Analyzed the decaying schemes in terms of convergence time and model performances

Kaggle: Travelers Claim Fraud Detection 11/2018–12/2018
*Team Members: Somyi Baek, Sam Piehl, King Yiu Suen, Xun Xian, Yu Yang (Won the **2nd** place)*

· Proposed a new feature which greatly improved the predictive capability
· Constructed an ensemble model for prediction and applied LIME for interpretation

TECHNICAL STRENGTHS

Languages	Python, R, Shell Scripting, C/C++, SQL, HTML
Tools	Git, VS Code, Google Cloud, AWS, LaTeX