Yujin Han

yujinhan@connect.hku.hk +86-188-1137-8895

EDUCATION BACKGROUND

The University of Hong Kong — **Hong Kong, China**
Ph.D. in Computer Science, Computer Science — *Sep. 2023-Jun. 2027*
Supervised by Difan Zou

Yale University — **New Haven, USA**
MSc in Biostatistics, Biostatistics (Data Science Track) — *Sep. 2021-Jun. 2023*
Classification: Honor (above 90/100), Supervised by Leying Guan

Renmin University of China (RUC) — **Beijing, China**
BS in Economics, Economic Statistics, School of Statistics — *Sep. 2016-Jun. 2020*
GPA: 3.74/4, Supervised by Maozai Tian

University of California, Berkeley (UCB) — **Berkeley, USA**
Summer School — *Jun. 2018-Jul. 2018*
GPA: 3.94/4

Short Bio

I am currently a first-year Ph.D. candidate in Computer Science at the University of Hong Kong (HKU), where I have the privilege of being supervised by Prof.Difan Zou.
I maintain a curiosity for any challenging research topics in machine learning and deep learning, exploring them with the aim of proposing principled methods and providing insightful understanding that can benefit the public.
My Research Topic. Guided by Prof.Zou, my research topic currently involves two aspects: (1) Developing new algorithms to address the issue of non-robustness in ERM caused by spurious correlations, and (2) Theoretically exploring the discrepancies between diffusion models' generative distributions and the oracle distribution to understand diffusion models' underlying mechanisms. Additionally, I am exploring how foundation models can help address specific data distribution biases.

PUBLICATIONS & PREPRINTS

* indicates equal contribution.

Publication in Statistics & Machine Learning

1. Conformalized semi-supervised random forest for classification and abnormality detection. [PDF]
   Yujin Han*, Mingwenchan Xu*, Leying Guan.
   International Conference on Artificial Intelligence and Statistics (AISTATS), 2024
2. Improving Group Robustness on Spurious Correlation Requires Preciser Group Inference.
   Yujin Han, Difan Zou.
   NeurIPS Workshop on Causal Representation Learning (CRL), 2023
   Submitted to ICML 2024 (Score: 7-6-5-4)

Publications in Other fields

1. Blockchain Technology for Electronic Health Records. [PDF]
   Yujin Han, Yawei Zhang, Sten H. Vermund
   International Journal of Environmental Research and Public Health, 2022. JCR Q1.

2. Suicidal ideation risk and socio-cultural factors in China: A longitudinal study on social media from 2010 to 2018. [PDF]
   Li He, **Yujin Han**, Yunyu Xiao, Xingyun Liu, Ang Li, Tingshao Zhu.
   International Journal of Environmental Research and Public Health, 2022. JCR Q1.

Preprints

1. FedGBF: An efficient vertical federated learning framework via gradient boosting and bagging. [ArXiv]
   Yujin Han, Pan Du, Kai Yang.

Patents

1. Clustering methods, devices, electronic devices and storage media. Application (Patent) No.:202110360295.
   Yujin Han, Yixin Li, Xiaolin Chen, et al.
2. A vertical federated learning modeling method, apparatus, device, and computer medium. Application (Patent) No. 202110417898.
   Yujin Han, Yixin Li, Xiaolin Chen, et al.

RESEARCH EXPERIENCE

HUAWEI Noah's Ark Lab **Beijing, China**
Research Intern, Decision Making & Reasoning Team *Jul. 2022- Sep. 2022*

- The cascade algorithm is designed to address the problems of difficult training and poor generalization of multimodal fiber data. The fiber data consisting of features and phases are trained with a cascaded neural network after out of distribution (OOD) detection by Auto-Encoder. Experiments shown the algorithm is efficient in training and alleviates the generalization problem. The algorithm has been adopted by the fiber optic communication department.
- Advisors: Bingshuai Li, and Yunfeng Shao

Chinese Academy of Sciences **Beijing, China**
Research Assistant, Computational Cyber-Psychology Lab *Mar. 2019- May 2021*

- Collected the text data related to suicide possibility of Weibo (Chinese Twitter) users by crawler and predicted suicide probability by DNN and SVM.
- Applied panel regression and Structural Equation Modelling to study the macro factors in Chinese society that can influence suicidal possibility of Weibo users. Two SCI papers were published based on the above reaserch.
- Advisor: Tingshao Zhu

INTERNSHIP

Jingdong Technology **Beijing, China**
Algorithm Engineer Intern, Federated Learning Team *Aug. 2020-Feb. 2021*

- Participated in federated modeling projects for Tencent and UnionPay; trained SecureBoost and FedLR (based on FATE), and evaluated the performance improvement.
- Proved the effectiveness and shortcomings of federated learning based on comparison between federated learning model of three participants and Xgboost; published the research report "How to play with multi-party federated learning?" on the official WeChat account of Jingdong.
- Co-authored the book "Big Data Risk Control Guide" and wrote two chapters (14.4 horizontal federated learning and 14.5 federated transfer learning); co-authored the book "Federal Learning and Technical Practice" and wrote the introductory part of 10+ pages; supported to revise the rest

chapters.

DiDi **Bijing, China**
Data Analyst Intern, Carpool Team *Sep. 2019 - Dec. 2019*

- Collected and consolidated data of 250,000+ App users in DiDi database through SQL.
- Analyzed and visualized the data to explain the behaviors of users and drivers; concluded that the DiDi App should pop up an interface page at fixed time showing the remain waiting time to reduce order cancellation; this advice was accepted by the Product department to optimize the consumer service.

AWARDS & ACHIEVEMENTS

- Jingdong Study Scholarship, RUC *2016/2018*
- First Prize, CUMCM *2018*
- Merit Student, RUC *2017*
- Outstanding Social Community Volunteer, RUC *2017/2018*
- Annual Speech Contest Champion, RUC *2018*
- Editor-in-Chief, Culture, Youth (Official Newspaper of RUC) *2016-2019*

ACADEMIC SERVICE & TEACHING

- Conference Reviews: AISTATS2023
- Teaching Assistant: CCST9047 The Age of Big Data (Spring 2023, HKU), BIS 620 Data Science Software Systems (Fall 2022, Yale)