

RUI ZHANG

PHD CANDIDATE IN COMPUTER SCIENCE

✉ r.zhang@duke.edu | 🏠 ruizhang1996.github.io | 🐙 github.com/ruizhang1996 | 🔗 linkedin.com/in/rzhang9997

EDUCATION

Duke University

Ph.D in Computer Science

Durham, NC

Sep. 2021 - Present

- Interpretable Machine Learning Lab
- Advisor: Prof. **Cynthia Rudin**
- GPA: 3.96/4.0

University of British Columbia

Bachelor of Science in Computer Science (**Honors**)

Vancouver, BC

Sep. 2016 - May. 2020

- Graduated with Distinction
- GPA: 3.94/4.0

PUBLICATIONS & PREPRINTS

- Jiachang Liu*, **Rui Zhang***, and Cynthia Rudin, “FastSurvival: Hidden computational blessings in training cox proportional hazards models,” in NeurIPS 2024, Accepted.
- **Rui Zhang**, Margo Seltzer, and Cynthia Rudin, “Interpretable multi-label learning with concise models,” in AISTATS 2025, Submitted.
- **Rui Zhang**, Rui Xin, Margo Seltzer, and Cynthia Rudin, “Optimal sparse survival trees,” in International Conference on Artificial Intelligence and Statistics, PMLR, 2024.
- **Rui Zhang***, Rui Xin*, Margo Seltzer, and Cynthia Rudin, “Optimal sparse regression trees,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2023.
- Qijia Huang*, **Rui Zhang***, and Sicong Fan*, “Knowledge distillation to increase robustness under natural distribution shift and distance measurement of shifts,” Duke University, Tech. Rep., 2021.
- **Rui Zhang**, “An personalized experimenter platform for eye-tracking-based user adaptation,” Honor Undergraduate Thesis, University of British Columbia, 2020.

RESEARCH PROJECTS

Optimal Sparse Model Learning

Sep. 2021 - Present

- Developed practical and customized algorithms that efficiently generate SOTA interpretable models (e.g., decision trees and generalized additive models) to solve a **wide range** of machine learning problems on large-scale data, including supervised learning, multi-label learning, transfer learning and more.
- **Fast** algorithms (trained **in seconds**) identify provably-optimal models that are both **accurate**—guaranteeing predictive performance **at least as good as** SOTA black-box models, and up to **22% better**—and **sparse**, making them naturally human-understandable.
- **Open-sourced** packages, compatible with scikit-learn, that have been widely adopted by researchers and practitioners, with **over 200K** downloads on PyPI.

Interpretable Reliability / Risk Analysis

Sep. 2022 - Present

- Leveraged the dynamic programming with branch-and-bound method and the beam-search method, and discovered novel bounds and surrogate functions to accelerate optimization.
- Efficiently produced optimal sparse survival trees with **30% higher** accuracy compared to SOTA, and fast sparse survival GAMs with **33% better** accuracy than SOTA.

Rashmon Set

Sep. 2022 - Present

- Explored efficient methods for constructing, storing, visualizing, and deploying the Rashomon Set (a set of nearly equally good models) across various machine learning models.
- User accessible: users can select, modify, or even generate new models from common sub-modules within the Rashomon Set according to their personalized needs (e.g. fairness), **without retraining**.

Stable, Robust and Trustworthy ML

May. 2024 - Present

- Exploring methods to enhance the stability and robustness of explainable ML models including both inherently interpretable model and post-hoc explainable deep model.
- Developing benchmarks for various machine learning models by providing a general framework and reproducing models with unavailable public code.

Hierarchical Multidimensional Datawarehousing

Apr. 2019 - Sep. 2019

- Designed a fast dynamic-programming algorithm (**3.6x faster**) to construct the most concise tree-structured summary for hierarchical multidimensional database.
- Supervised by Prof. **Margo Seltzer** and Prof. **Laks V.S. Lakshmanan**.

TECHNICAL SKILLS

Programming C/C++, Python (PyTorch, Numpy, Pandas, scikit-learn, matplotlib. etc.)

DevOps Stack Docker, Github CI/CD, Git

Database MySQL, Oracle

PROFESSIONAL SERVICES

Journal Reviewer

- Journal of Machine Learning Research (JMLR)
- Theoretical Computer Science (TCS)
- INFORMS Journal on Computing

Program Committee for Conferences and Workshops

- INFORMS 2024 Workshop on Data Science
- NeurIPS 2024 Workshop on Interpretable AI

HONORS AND AWARDS

INFORMS Data Mining Society DMDA Workshop Best Paper Finalist (Winner TBD) 2024

AAAI Scholarship 2023

Ph.D Fellowship, Duke 2021 - 2022

Trek Excellence Scholarship (top 5%), UBC 2018 - 2019

Faculty of Science International Student Scholarship, UBC 2018 - 2019

Dean's Honour List, UBC 2018 - 2020

Science Scholar, UBC 2018 - 2020

Charles and Jane Banks Scholarship, UBC 2018

Outstanding International Student Award, UBC 2016

TEACHING

Duke COMPSCI 590: Data Science, Graduate Teaching Assistant 2023

Duke COMPSCI 671: Theory & Alg ML, Graduate Teaching Assistant 2022

UBC CPSC 404: Advanced Relational DB, Undergraduate Teaching Assistant 2019 - 2020

UBC CPSC 304: Intro to Relational DB, Undergraduate Teaching Assistant 2019

UBC CPSC 213: Intro to Computer Systems, Undergraduate Teaching Assistant 2018