Zhi Chen

zhi.chen1@duke.edu | cs.duke.edu/~zhichen

EDUCATION

Duke University — Aug 2018 – Now
Ph.D. candidate in Computer Science (advisor: Cynthia Rudin)

Nanjing University — Sep 2014 – Jun 2018
B.Sc. in Computer Science (Honor Program)

WORK EXPERIENCE

Research Intern, Meta Responsible AI — May 2022 – Aug 2022
Mentor: Fulton Wang

- Doing research on globally interpretable deep learning models whose decisions are based on the influence functions from a sparse set of prototypes (training examples)

Research Intern, Microsoft Research — May 2021 – Aug 2021
Mentor: Rich Caruana

- Used Explainable Boosting Machines (EBM) to detect common flaws in data (e.g. poorly imputed missing values, confounders and treatment effects, bias, etc.)
- Developed an accurate and interpretable Bayesian GAM, based on BART, for better uncertainty quantification.

RESEARCH EXPERIENCE

Concept-based Interpretable Neural Networks. — Aug 2018 – Now

- Developed a module, concept whitening (CW), to decorrelate and align the axes of the latent space to predefined concepts. CW can provide a much clearer understanding for how the network gradually learns concepts over layers without hurting predictive performances.
- Developing methods that discover useful domain-specific concepts in an unsupervised way, and explicitly represent these discovered concepts in the latent space of neural networks.

Rashomon Set of Very Sparse Models — Jan 2022 – Now

- Constructing a set of almost-equally-good models (Rashomon set) and making it very accessible to users, giving them the flexibility to visualize, explore, select, and modify not only one but many well-performing models.
- Currently focusing on Rashomon set of very sparse models like optimal decision trees and additive models.

Interpretable Machine Learning for Metamaterial Designs — Aug 2019 – Now

- Developed interpretable machine learning methods to discover key local and global features related to important dynamic material properties such as mechanical band gaps. These physically interpretable features can also transfer information about material properties across scale.

TextGAN — Nov 2016 – Feb 2017

- Developed a generative adversarial network for text generation.

SELECTED AWARDS

- Finalist of Data Mining Best Paper Competition Award (student track), INFORMS — 2022
- Winner of the SPES+Q&P Student Paper Competition, American Statistical Association — 2022
- Outstanding Ph.D. Preliminary Exam Award, Duke Computer Science — 2021
- Ph.D. Fellowship, Duke University — 2018 & 2019
- Outstanding Graduate Award, Nanjing University — 2018

TALKS & PRESENTATIONS

- Invited talk, Interpretability in AI workshop, Banff International Research Station — May 2022
- Workshop oral talk, ECML-PKDD XKDD workshop — Sep 2021
- Invited talk, PAIR Team at Google Brain — Mar 2021
- Invited talk, HCAI Lab at Graz University of Technology, Austria — Jan 2021

TEACHING EXPERIENCE

- Teaching Assistant of Machine Learning (COMPSCI 671), Duke University — Spring 2019
- Teaching Assistant of Machine Learning (COMPSCI 671), Duke University — Fall 2019

PROFESSIONAL SERVICES

- Reviewer of CVPR, ICCV, ECCV, AISTATS

TECHNICAL SKILLS

- PyTorch, Tensorflow, Python, C/C++, Matlab, R

PREPRINTS & PUBLICATIONS

(* denotes equal contribution)

- Rui Xin*, Chudi Zhong*, **Zhi Chen***, Takuya Takagi, Margo Seltzer, Cynthia Rudin. Exploring the Whole Rashomon Set of Sparse Decision Trees. Accepted to *the Conference on Neural Information Processing Systems* (2022).
- **Zhi Chen**, Alexander Ogren, Chiara Daraio, Cate Brinson, Cynthia Rudin. How to See Hidden Patterns in Metamaterials with Interpretable Machine Learning, Accepted to *Extreme Mechanics Letters* (2022).
- Zijie Wang, Chudi Zhong, Rui Xin, Takuya Tagaki, **Zhi Chen**, Cynthia Rudin, Margo Seltzer. TimberTrek: Exploring and Curating Trustworthy Decision Trees with Interactive Visualization. *IEEE VIS* (2022).
- **Zhi Chen**, Sarah Tan, Harsha Nori, Kori Inkpen, Yin Lou, Rich Caruana. Using Explainable Boosting Machines (EBMs) to Detect Common Flaws in Data. *ECML-PKDD International Workshop and Tutorial on eXplainable Knowledge Discovery in Data Mining* (2021).
- Cynthia Rudin, Chaofan Chen, **Zhi Chen**, Haiyang Huang, Lesia Semenova, Chudi Zhong. Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *Statistics Surveys* (2021).
- **Zhi Chen**, Yijie Bei, Cynthia Rudin. Concept Whitening for Interpretable Image Recognition. *Nature Machine Intelligence* (2020).
- Yizhe Zhang, Zhe Gan, Kai Fan, **Zhi Chen**, Ricardo Henao, Dinghan Shen and Lawrence Carin. Adversarial Feature Matching for Text Generation. *Proceedings of the 34th International Conference on Machine Learning (2017)*.