Chun-Hao **Chang**

Research Scientist at Meta

Seattle, WA, 98121

(206) 459-9283 | chkchang21@gmail.com | http://www.cs.toronto.edu/ kingsley/ | chkchang | Chun-Hao Chang

Summary

My PhD research focuses on building machine learning models in critical settings such as Healthcare, which require models to be explainable to human and adapting to distribution shifts. Towards this goal, I work on topics in Deep Learning, particularly the intersections of Explainability (**XAI**), Distribution Shifts (**DS**), Reinforcement Learning (**RL**), Healthcare (**HC**) and Anomaly Detection (**AD**). I have interned in Google, Microsoft, and Meta. Before my PhD, I was a web and Android engineer for 2 years.

Education

University of Toronto — *Toronto, Canada*

Ph.D. and M.S. in Computer Science — *Sep. 2016 - Mar. 2022*

National Taiwan University — *Taipei, Taiwan*

B.S. in Electrical Engineering and Life Science — *Sep. 2010 - Jun. 2015*

- GPA: 4.07 / 4.30. Double major with Presidential Awards 5 times that only given to top 5% students. Graduated with Departmental Honor.

Experience

University of Toronto — *Toronto, ON, Canada*

Graduate Research Assistant. Advisor: Anna Goldenberg — *Sep. 2016 - Mar. 2022*

- [**XAI**] Developed NodeGAM, a self-explainable deep learning model on tabular data by training a Generalized Additive Model (GAM) composed of layers of differentiable trees. The GAM formulation makes it easily visualized and understood by human, and differentiability allows it to learn representations and its accuracy outperforms other GAMs or tree models. (*ICLR 2022* ***Spotlights***) [**ICLR'22**]
- [**XAI, RL, HC**] Developed CAIRL, an explainable inverse reinforcement learing (IRL) method on a clinicial ICU dataset that recovers the clinicians' treatments goals. IRL aims to recover the reward from expert demonstrations to help reward design in RL. CAIRL successfully recovers clinicians' specified reward and match clinical guidelines and outperforms previous methods. [**Arxiv'21**]
- [**DS**] Proposed two data augmentations to remove spurious correlations in image classifiers given the annotations of the causal features (i.e. bounding boxes). Improved both generalization accuracy and out-of-class detection in multiple challenging datasets whose distributions differ from the original distributions. [**CVPR'21**]
- [**DS,HC**] Introduced the feedback loop problem that the future data are corrupted by human interventions caused by model's predictions; if continually updated from new data, models degrade almost exponentially. Analyzed how clinician trusts to the model affect the feedback loop. Showed that being skeptical of the model's predictions can reduce the effects of the feedback loop by a large degree. Proposed various sample removal techniques to minimize the effects of the feedback loop. [**MLHC'20**]
- [**XAI**] Proposed an efficient way to purify interaction effects towards main effects to make models interpretable. Showcased large disparity, including contradictions, between the main effects before and after purification. [**AISTATS'20**]
- [**RL,HC**] Proposed a reinforcement learning approach to prioritize healthcare resouces for sequential measurements of patients in Intensive Care Units (ICU) by a Deep Q network (DQN). To handle the large combinatorial action spaces (2^{40}) in the multi-action setting, we developed a sequential action approach to learn it more efficiently; we show it outperforms other baselines. Applied to a real-world clinical ICU dataset (MIMIC3) and increase 3 times informative gain or reduce 31% measurement costs compared to physicians. [**ICML'19**]
- [**XAI**] Proposed a local explanation method that highlights important features for any deep neural networks. First, proposed using a generative model to interpret a black-box model to make the alternate image natural. Improved the search efficiency by a factorized Bernoulli distribution that speeds up 10x over prior arts. [**ICLR'19**]

Google Cloud AI — *Toronto, ON, Canada*

Student Researcher. Host: Jinsung Yoon — *Aug. 2021 - Jan. 2022*

- [**XAI, AD**] Developed DIAD, an explainable anomaly detection method for tabular data by optimizing NodeGAM on an Anomaly Detection (AD) objective (PID Objective). This has the explainability of GAM, allows fine-tuning on labeled anomalies, and achieves best accuracy in both unlabeled and partially labeled AD setting in 20 tabular datasets. [**Arxiv'22**]

Microsoft Research — *Redmond, WA, USA*

Research Intern. Host: Rich Caruana — *Jun. 2019 - Sep. 2019*

- [**XAI**] Compared total 9 kinds of inherently explainable models (Generalized Additive Models, GAMs) in terms of human explainability, inductive biases and performances across multiple real-world datasets. Found GAMs with high feature sparsity (only using a few variables to make predictions) tend to hide patterns in the data and be unfair to rare subpopulations. Found inductive bias plays a crucial role in model explanations and tree-based GAMs are recommended for their low feature sparsity and high data fidelity which make them more trustworthy. [**KDD'21**]

Ads Ranking Team, Facebook — *London, UK*

PhD Machine Learning Intern. Host: Damien Lefortier — *Jun. 2018 - Sep. 2018*

- Modeled counterfactual inference of the ads data to measure lift effects of ads exposure on users.
- Visualized and analyzed the data and removed the outliers.
- Investigated various causal approach such as nearest neighbor, causal trees, and multi-task learning.

Political Warfare Bureau, R.O.C. Armed Forces *Taipei, Taiwan*

Backend Engineer *Aug. 2015 - July 2016*

- Designed and programmed a search engine (ElasticSearch) with distributed computation to speed up.
- Led a team of 5 people including technical and non-technical personnel (designer/front-end engineers)

TripNotice Inc. *Taipei, Taiwan*

Android Engineer Intern *Jan. 2014 - Mar. 2015*

- Independently designed and implemented a trip-booking Android app (TripNotice) in Java from scratch.

Publications

Referred Publications

ICLR'22 **Chang, Chun-Hao**, Rich Caruana, and Anna Goldenberg. "NODE-GAM: Neural Generalized Additive Model for Interpretable Deep Learning." International Conference on Learning Representations. 2022.

CVPR'21 **Chang, Chun-Hao**, George Alexandru Adam, and Anna Goldenberg. "Towards robust classification model by counterfactual and invariant data generation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

KDD'21 **Chang, Chun-Hao**, et al. "How interpretable and trustworthy are gams?." Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021.

MLHC'20 Adam, G. A., **Chang, C. H. K.**, Haibe-Kains, B., & Goldenberg, A. "Hidden risks of machine learning applied to healthcare: unintended feedback loops between models and future data causing model degradation." In Machine Learning for Healthcare Conference. 2020.

AISTATS'20 Lengerich, B., Tan, S., **Chang, C. H.**, Hooker, G., & Caruana, R. Purifying interaction effects with the functional anova: An efficient algorithm for recovering identifiable additive models. In International Conference on Artificial Intelligence and Statistics (pp. 2402-2412). 2020.

ICML'19 **Chang, Chun-Hao**, Mingjie Mai, and Anna Goldenberg. "Dynamic measurement scheduling for event forecasting using deep RL." International Conference on Machine Learning. PMLR, 2019.

ICLR'19 **Chang, Chun-Hao**, et al. "Explaining Image Classifiers by Counterfactual Generation." International Conference on Learning Representations. 2019.

Preprints / Under Review / Technical Papers

Arxiv'22 **Chang, Chun-Hao**, et al. "Data-Efficient and Interpretable Tabular Anomaly Detection." arXiv preprint arXiv:2203.02034 (2022).

Arxiv'21 **Chang, Chun-Hao**, et al. "Extracting Expert's Goals by What-if Interpretable Modeling." arXiv preprint arXiv:2110.15165 (2021).

Arxiv'17 **Chang, Chun-Hao**, Ladislav Rampasek, and Anna Goldenberg. "Dropout feature ranking for deep learning models." arXiv preprint arXiv:1712.08645 (2017).

Honors & Awards

2019 **Student Travel Award**, ICML 2019 *Las Vegas, U.S.A*

2019 **Student Travel Award**, ICLR 2019 *New Orleans, U.S.A*

2010-2015 **Presidential Award 5 times**, Awarded to top 5% scores in the department, National Taiwan University. *Taipei, Taiwan*

2014 **Most Feasible Award**, NTU Creativity Competition, 4th prize out of 100 groups *Taipei, Taiwan*

2010 **1st Runner-Up**, International Biology Olympiad Selection Camp *Taipei, Taiwan*

Selected Talks

Apr. 2022 **Spotlight Talk**, ICLR'22, NODE-GAM: Neural Generalized Additive Model for Interpretable Deep Learning *Online*

Feb. 2022 **Job Talk**, Amazon Web Services. Neural Generalized Additive Models. *Online*

Feb. 2022 **Job Talk**, Borealis AI. Neural Generalized Additive Models. *Online*

Oct. 2021 **Seminar Talk**, Dr. Bo Wang's lab. Generalized Additive Models. *Online*

Aug. 2021 **Poster Talk**, KDD'21: How Interpretable and Trustworthy are GAMs? *Online*

Other Activities

Machine Learning Teaching Assistant, University of Toronto & Vector Institute *Toronto, Canada*

6 times in Machine Learning courses *Sep. 2016 - 2021*

Conference Reviewers, Machine Learning Conferences (ICLR, ICML, NeurIPS, CVPR, ...) *Online*

8 times. Total 39 papers. *2019-2022*

Computer Science Graduate Admissions *Toronto, Canada*

Appraiser *2019,2020,2021*

Sickkids GMCA miniMBA program *Toronto, Canada*

MiniMBA 12-week course completion *Jan. 2017*

References

Anna Goldemberg *PhD Advisor*

Professor, Computer Science, University of Toronto *anna.goldenberg@utoronto.ca*

Rich Caruana *Intern Host*

Principle Researcher, Microsoft Research *rcaruana@microsoft.com*

Jinsung Yoon *Intern Host*

Research Scientist, Google Research *jinsungyoon@google.com*

Marzyeh Ghassemi *Thesis Commitee Member*

Assistant Professor, MIT in EECS and IMES *mgassem@mit.edu*