ATML Fellow Class 2023

Dr. Zhiyong Yang

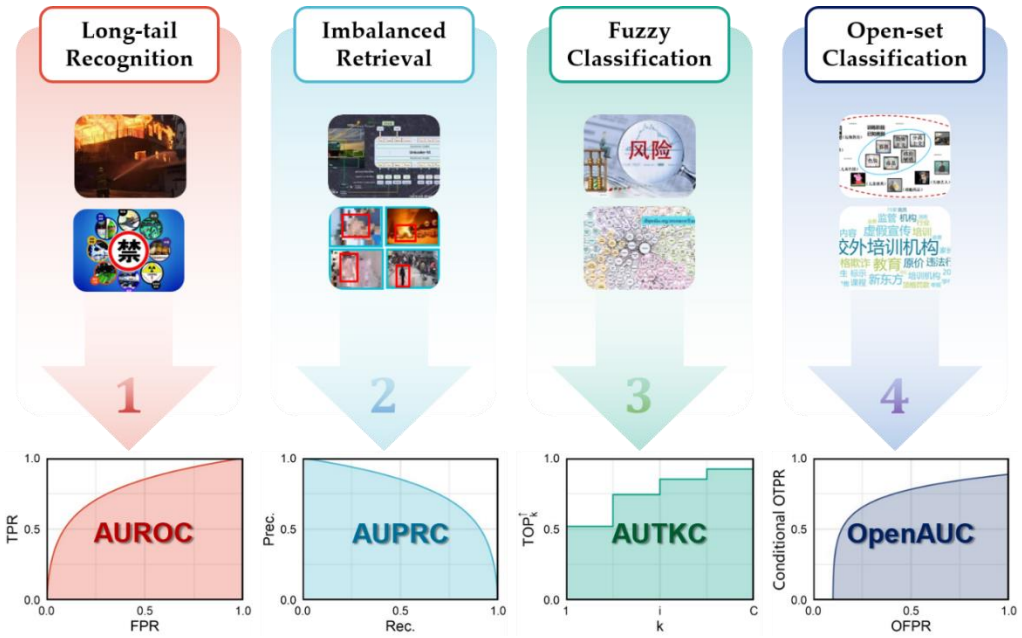
University of Chinese Academy of Sciences

https://joshuaas.github.io/

Bio: Dr. Zhiyong Yang is an Associate Professor (Tenure-track Assistant Professor) at the University of Chinese Academy of Sciences. His research focuses on AUC optimization, XCurve, long-tail learning and trustworthy machine learning. He has published 60+ papers in top-tier AI journals and conferences, including 30+ ICML/NeurIPS/T-PAMI. In terms of academic services, he served as an Area Chair of NeurIPS 2024, a Senior Program Committee Member (SPC) for IJCAI 2021, an Expert Reviewer (ER) for ICML, a PC Member for NeurIPS, ICML, and ICLR, as well as a reviewer for T-PAMI, T-IP and TMLR. He received several awards including CCF Excellent Doctoral Dissertation Incentive Plan (a.k.a the CCF Excellent Doctoral Dissertation Award, top 10 in China per year), 2021 Baidu AI Global Chinese Star Top 100 (Top 25 in machine learning area), Baidu Scholarship Global Top 20 nomination, Hundred Excellent Doctoral Dissertation Award in CAS, President's Special Award in CAS, and NeurIPS Top Reviewer.

Contribution:

Recently, machine learning and deep learning technologies have been successfully used in many complex high-stakes decision-making applications such as disease prediction, fraud detection, outlier detection, and criminal justice sentencing. All these applications share a common trait known as risk-aversion in economic and financial terms. In other words, decision-makers usually have very low risk tolerance. In this context, decision-making parameters, like classification thresholds, significantly affect model performance. In other applications like recommendation systems and information retrieval, decision parameters also play an important role since personalized decision needs must be met. In these scenarios, decision parameters change dynamically, causing biases that traditional ML framework can no longer address. To address these issues, we propose a decision-invariant machine learning framework called XCurve (https://xcurveopt.github.io/), inspired by the principle of AUROC optimization. Specifically, we optimize the Area Under X Curve (AUXC) on top of a specific performance curve X as a plot of two performance functions of the decision parameter. AUXC is decision-invariant because the integral of the performance curve implicitly considers all possible choices of decision parameters in some sense. Currently, the XCurve framework includes a series of learning algorithms for several performance curves including AUROC (for long-tail classification), AUPRC (for retrieval), AUTKC (for classification with semantic ambiguity) and OpenAUC (for open-set recognition).



Dr. Hwanjun Song

Korea Advanced Institute of Science & Technology

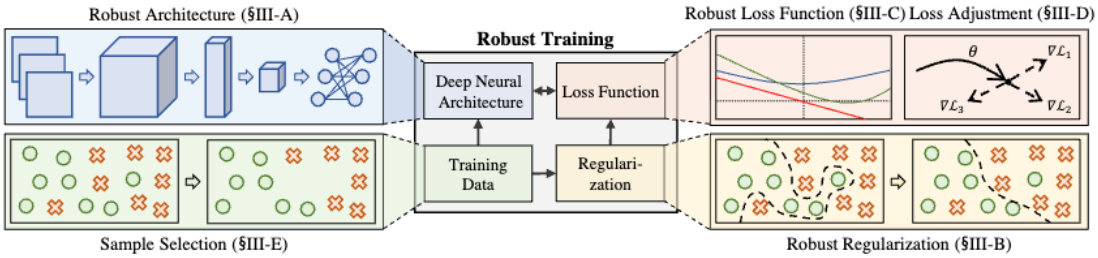
https://songhwanjun.github.io/

Bio: Dr. Hwanjun Song is an Assistant Professor in the Graduate School of Data Science and the Department of Industrial and Systems Engineering at Korea Advanced Institute of Science and Technology (KAIST), South Korea. Previously, he was a Research Scientist at the AI Labs of Amazon Web Services in 2023 and at the AI Lab of NAVER Clova in 2021-2022, and Research Intern at the Machine Perception of Google Research in 2020. He is broadly interested in in the fields of AI Safety and Robustness, with a particular focus on mitigating hallucination in generative AI, enhancing AI-Human value alignment, and learning with imperfect data, including noisy, imbalanced, and limited data. He has published over 40 papers in top-tier NLP/CV/ML conferences, such as ACL, NeurIPS, and CVPR. In terms of academic services, he served as a Program Committee member for ICLR, NeurIPS, and ICML, as well as a reviewer for TPAMI and IJCV. In addition, He sponsored by Microsoft through Azure for Research, received the Innovation Award from Qualcomm, N-Innovation Award from NAVER Corp., Paper Award from Samsung Human Tech.

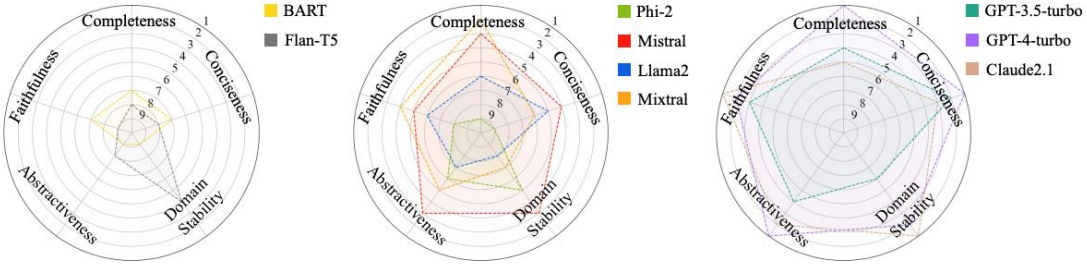
Contribution:

AI robustness and safety are crucial because they ensure that AI systems perform reliably under diverse and unpredictable conditions, minimizing the risk of errors or harmful outcomes. Specifically, AI models are widely known to be vulnerable to noisy labels, where inaccuracies or inconsistencies in the annotated data can significantly impact their training and performance, leading to compromised reliability and potentially misleading outputs. To address this issue,

Prof. Hwanjun has made significant contributions, including the release of a real-world noisy benchmark datasets, ANIMAL-10N (https://dm.kaist.ac.kr/datasets/animal-10n/), and the development of robust training approaches to cover diverse learning scenarios, including SELFIE (robust supervised learning), PuriDiver (robust continual learning), FedRN (robust federated learning), MetaQueryNet (robust active learning). Additionally, his survey paper titled "Learning from Noisy Labels with a Deep Neural Networks: A Survey" has become seminal in this field.



Beyond the AI robustness to noise labels, he has recently begun to create high-quality benchmark data for Large Language Models like GPT-4, which comprehensively evaluate them across various Safety Dimensions, including faithfulness and domain stability. This benchmark helps us gain a complete understanding of the current progress of generative AI. This benchmark dataset will bridge the gap between humans' preference and AI capabilities, inspiring numerous future studies on AI safety.



(a) Non-LLMs. (b) Open-source LLMs. (c) Proprietary LLMs.