February 26, 2023

Bishwamittra Ghosh
205 Bukit Batok Street 21, #02-46
Singapore, 650205
Singapore
https://bishwamittra.github.io
Email: bghosh@u.nus.edu
Phone: (+65)85990160

Subject: Application for attending IJCAI 2023 Doctoral Consortium Program.

Dear Organizing Committee,

I am writing to apply for the IJCAI 2023 Doctoral Consortium Program. Please find attached my application materials.

My research interests primarily lie in the areas of fairness and interpretability in machine learning. Specifically, in my dissertation research, I focus on formally verifying fairness in machine learning, interpreting fairness by identifying its sources, and designing interpretable rule-based classifiers. To achieve scalable and accurate solutions, I closely integrate automated reasoning, formal methods, and statistics with fairness and interpretability.

My research is motivated by prior works on fairness and interpretability. For example, Rudin [2019] emphasizes the design of interpretable classifiers, such as rule-based classifiers [Lakkaraju et al., 2016], for high-stake and safety-critical decision-making, instead of interpreting opaque black-box classifiers. We improve the scalability of interpretable classifiers based on maximum satisfiability (MaxSAT) and incremental learning by demonstrating interpretable classification on million-size datasets. In fairness, Albarghouthi et al. [2017] and Bastani et al. [2019] study formal fairness verification to verify multiple fairness metrics of classifiers given the probability distribution of features. In our research, we improve the scalability and accuracy of formal fairness verification by integrating stochastic satisfiability (SSAT) and Bayesian network capturing feature-correlations. Furthermore, motivated by the GDPR's "right to explanation", research on interpreting model predictions has surged [Lundberg and Lee, 2017]. However, interpreting prediction bias has received less attention [Lundberg, 2020]. In this context, we apply global sensitivity analysis (GSA) to interpret fairness metrics by decomposing model bias into individual features and the intersection of multiple features.

I have not previously attended a doctoral consortium. I expect to graduate in Summer 2023 and plan to pursue a research career in either academia or industry, with a research focus centered on solving novel and practical problems in fairness and interpretability in machine learning for the next five years. To this end, I would be delighted to have mentors working in fairness and interpretability, such as Xiting Wang, Arthur Choi, Alexey Ignatiev, Dominik Peters, and Mohamed Siala.

My research has thrived through multiple collaborations and internships in industry and academia. We have published our works at premier conferences and journals in artificial intelligence and machine learning (AAAI-2022, 2021, 2020, JAIR-2022, ECAI-2020, AIES-2019) and databases (VLDB-2018, TSAS-2022).

Thank you for considering my application. I look forward to hearing from you.

Sincerely,
Bishwamittra Ghosh

References

Aws Albarghouthi, Loris D'Antoni, Samuel Drews, and Aditya V Nori. FairSquare: probabilistic verification of program fairness. *Proceedings of the ACM on Programming Languages*, 1:1–30, 2017.

Osbert Bastani, Xin Zhang, and Armando Solar-Lezama. Probabilistic verification of fairness properties via concentration. *Proceedings of the ACM on Programming Languages*, 3:1–27, 2019.

Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1675–1684, 2016.

Scott M Lundberg. Explaining quantitative measures of fairness. In *Fair & Responsible AI Workshop@ CHI2020*, 2020.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777, 2017.

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. volume 1, pages 206–215. Nature Publishing Group, 2019.