

Department of Computer Science



Mr. Du Mengnan

PhD Candidate
Department of Computer Science and Engineering
Texas A & M University, USA



Date: 16 February 2022 (Wednesday)



Time: 11:00am – 12:00nn



Registration: <http://bit.ly/bucs-ereg>

(*Zoom details will only be provided to registrants)

Register Now



Deep Neural Networks Explainability: Algorithms and Applications



ABSTRACT

Deep neural networks (DNN) have achieved extremely high prediction accuracy in a wide range of fields such as computer vision, natural language processing, and recommender systems. Despite the superior performance, DNN models are often regarded as black-boxes and criticized for the lack of interpretability, since these models cannot provide meaningful explanations on how a certain prediction is made. Without the explanations to enhance the transparency of DNN models, it would become difficult to build up trust and credibility among end-users. In this talk, I will present our efforts to tackle the black-box problem and to make powerful DNN models more interpretable and trustworthy. First, I will introduce post-hoc interpretation approaches for predictions made by two standard DNN architectures, including Convolution Neural Network (CNN) and Recurrent Neural Network (RNN). Second, I will introduce the usage of explainability as a debugging tool to improve the generalization ability and fairness of DNN models.



BIOGRAPHY

Mengnan Du is currently a Ph.D. student in Computer Science at the CSE department of Texas A&M University, under the supervision of Dr. Xia Ben Hu. His research is on the broad area of trustworthy machine learning, with a particular interest in the areas of explainable, fair, and robust DNNs. He has had around 40 papers published in venues such as NeurIPS, AAI, KDD, WWW, NAACL, and ICLR. He received over 1,200 citations with an H-index of 11. Three of his papers were selected for the Best Paper Candidate at WWW'19, the Best Paper Candidate at ICDM'19, and the INFORMS 2019 Best Refereed Paper Finalists, respectively. Another of his paper was highlighted on the cover page of Communications of the ACM, January 2020 issue. He served as the Registration Chair of WSDM'22, and is the program committee member of conferences including NeurIPS, ICML, ICLR, AAI, ACL, EMNLP, NAACL, etc.

ENQUIRY

Tel: 3411-2385 Email: comp@comp.hkbu.edu.hk Website: <https://www.comp.hkbu.edu.hk/v1/?page=events>