UTHealth
The University of Texas Health Science Center at Houston
School of Biomedical Informatics

DSICCR Tuesday Seminar Series

Tuesday at Noon, Click Here to Join the Seminar

Towards Effective and Efficient Interpretation of Deep Neural Networks: Algorithms and Applications

Xia (Ben) Hu, Ph.D.
Director, Data to Knowledge Lab (D2K)
Associate Professor of Computer Science
Rice University

While Deep neural networks (DNN) have achieved superior performance in many downstream applications, they are often regarded as black-boxes and are criticized by their 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 among end-users. In this talk, I will present a systematic framework from modeling and application perspectives for generating DNN interpretability, aiming at dealing with main technical challenges in interpretable machine learning, i.e., faithfulness, understandability and the efficiency of interpretability. Specifically, to tackle the faithfulness challenge of post-hoc interpretation, I will introduce how to make use of feature inversion and additive decomposition techniques to explain predictions made by two classical DNN architectures, i.e., Convolutional Neural Networks and Recurrent Neural Networks. In addition, to develop DNNs that could generate more understandable interpretation to human beings, I will present a novel training method to regularize the interpretations of a DNN with domain knowledge. Finally, to accelerate the interpretation of DNNs, I will introduce a framework to significantly reduce the complexity of explaining DNNs without the degradation of interpretation performance. Fast and efficient DNNs explaining promotes the real-world application of XAI, especially for the online systems.

Tuesday, November 1st, 2022. 12p – 1p. Webcast

Contact: Xiaohong.Bi@uth.tmc.edu

#SBMIseminar