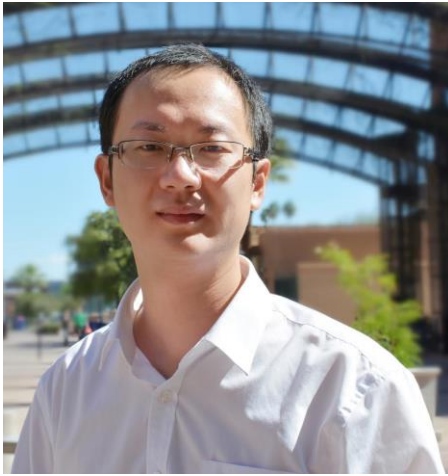# Interpretable Machine Learning: Concepts and Techniques

In many real-world applications, such as Healthcare and Cybersecurity, domain experts would appropriately trust and effectively manage prediction results only if they can understand the prediction model as well as results. For example, it is critical, for not only physicians but also patients, to understand why a patient is diagnosed with prediabetes and how the decision is made by a predictive model. This talk will cover fundamental concepts as well as the state-of-the-art algorithms with implementations in interpretable machine learning. First, we will introduce the background of this problem, and widely used concepts and problem definition of interpretable machine learning. Second, we will discuss the system architecture and main algorithms, as well as our current progress, to bridge the gap between powerful deep learning algorithms and interpretable shallow models through model and application perspectives. At the end, we will introduce XDeep, which is an open-source Python package developed to interpret deep models for both practitioners and researchers. XDeep takes a trained deep neural network (DNN) as the input, and generates relevant interpretations as the output with the post-hoc manner. From the functionality perspective, XDeep integrates a wide range of interpretation algorithms from the state-of-the-arts, covering different types of methodologies, and is capable of providing both local explanation and global explanation for DNN when interpreting model behaviors.

The first 1.5 hours of the tutorial is an introductory exposition of the topic. The second 1.5 hours is a practical session that helps the audience consolidate their knowledge through hands-on computational experience. Workshop participants should bring own laptop. Please visit TAMIDS website for this tutorial at https://tamids.tamu.edu/2020/02/09/workshop-interpretable-machine-learning/ and register for the tutorial at https://tinyurl.com/yx7b9u6e. Registration is free.

Background knowledge advisable: Participants must have at least a basic knowledge of machine learning concepts from supervised and unsupervised learning. Some understanding of deep learning and frameworks (e.g. TensorFlow or PyTorch) would be helpful. Basic experience with Python programming on machine learning problems is assumed.

Xia "Ben" Hu, Ph.D.
Assistant Professor and Lynn '84 & Bill Crane '83
Faculty Fellow
Dept. of Computer Science & Engineering
Texas A&M University

Date: Friday, February 21
Time: 2:00 – 5:00 p.m.
Location: ETB 1005

Biography

Dr. Xia "Ben" Hu is an Assistant Professor and Lynn '84 and Bill Crane '83 Faculty Fellow at Texas A&M University in the Department of Computer Science and Engineering. Hu directs the Data Analytics at Texas A&M (DATA) Lab. Dr. Hu has published over 100 papers in several major academic venues, including KDD, WWW, SIGIR, IJCAI, AAAI, etc. An open-source package developed by his group, namely AutoKeras, has become the most used automated deep learning system on Github (with over 6,000 stars and 1,000 forks). Also, his work on deep collaborative filtering, anomaly detection and knowledge graph have been included in the TensorFlow package, Apple production system and Bing production system, respectively. His papers have received several awards, including WWW 2019 Best Paper Shortlist, INFORMS 2019 Best Poster Award, INFORMS QSR 2019 Best Student Paper Finalist, IISE QCRE 2019 Best Student Paper Award, WSDM 2013 Best Paper Shortlist, IJCAI 2017 BOOM workshop Best Paper Award. He is the recipient of JP Morgan AI Faculty Award, Adobe Data Science Award, NSF CAREER Award, and ASU President Award for Innovation. His work has been featured in several news media, including the MIT Technology Review, ACM TechNews, New Scientist, Defense One, and others. Hu's work has been cited more than 6,000 times with an h-index of 36. He was the conference General Co-Chair for WSDM 2020. More information can be found at http://faculty.cs.tamu.edu/xiahu/.

For more information about TAMIDS tutorial series, please contact Ms. Jennifer South at jsouth@tamu.edu