CIDSE Invited Talk with Richard Caruana

ASU-Mayo Center for Innovative Imaging Invited Lecture Series

Friends Don't Let Friends Deploy Black-Box Models: The Importance of Intelligibility in Machine Learning

Tuesday, March 23, 2021

1:30–3 p.m.

Zoom https://asu.zoom.us/j/88650537033

Abstract

In machine learning often a tradeoff must be made between accuracy and intelligibility: the most accurate models usually are not very intelligible, and the most intelligible models usually are less accurate. This can limit the accuracy of models that can safely be deployed in mission-critical applications such as healthcare where being able to understand, validate, edit, and ultimately trust a model is important. We have developed a learning method based on generalized additive models (GAMs) that is as accurate as full complexity models, but even more intelligible than linear models. This makes it easy to understand what a model has learned and to edit the model when it learns inappropriate things. In the talk I'll present case studies where high-accuracy GAMs discover surprising patterns in data that would have made deploying a black-box model risky. I'll also show a case study that presents new findings about COVID-19 discovered by an intelligible model trained to predict COVID-19 mortality. And if there's time, I'll show how we're using these models to uncover bias in models where fairness and transparency are important. Every data set is flawed in surprising ways or hiding important secrets--you need intelligibility to uncover the hidden secrets and find the gold.

Bio

Rich Caruana is a Senior Principal Researcher at Microsoft. His research focus is on intelligible/transparent modeling and machine learning for medical decision making. Before joining Microsoft, Rich was on the faculty at Cornell, at UCLA's Medical School, and at CMU's Center for Learning and Discovery. Rich's Ph.D. is from CMU, and his thesis on Multitask Learning helped create interest in the new subfield of Transfer Learning. Rich received an NSF CAREER Award in 2004 for Meta Clustering.

https://www.microsoft.com/en-us/research/people/rcaruana/

Ira A. Fulton Schools of Engineering
Arizona State University

School of Computing, Informatics, and Decision Systems Engineering

Hosted by:
J. Caviedes, T. Wu, B. Li