Intelligible Machine Learning Models for HealthCare

Rich Caruana, Ph.D.
Microsoft Research

Abstract:
In machine learning often a tradeoff must be made between accuracy and intelligibility: the most accurate models usually are not very intelligible (e.g., random forests, boosted trees, and neural nets), and the most intelligible models usually are less accurate (e.g., linear or logistic regression). This tradeoff sometimes limits the accuracy of models that can be applied in mission-critical applications such as healthcare where being able to understand, validate, edit, and trust a learned model is important. We have developed a learning method based on generalized additive models (GAMs) that is often as accurate as full complexity models, but remains as intelligible as linear/logistic regression models. In the talk I’ll present two case studies where these high-performance generalized additive models (GA2Ms) are applied to healthcare problems yielding intelligible models with state-of-the-art accuracy. In the pneumonia risk prediction case study, the intelligible model uncovers surprising patterns in the data that previously prevented other complex models from being deployed, but because it is intelligible and modular allows these patterns to easily be recognized and removed. In the 30-day hospital readmission case study, we show that the same methods scale to large datasets containing hundreds of thousands of patients and thousands of attributes while remaining intelligible and providing accuracy comparable to the best (unintelligible) machine learning methods.