Algorithms for Interpretable Machine Learning

Cynthia Rudin
Massachusetts Institute of Technology
100 Main Street, E62, Room 576
Cambridge, MA 02142
1-617-715-4215
rudin@mit.edu

ABSTRACT

It is extremely important in many application domains to have transparency in predictive modeling. Domain experts do not tend to prefer "black box" predictive model models. They would like to understand how predictions are made, and possibly, prefer models that emulate the way a human expert might make a decision, with a few important variables, and a clear convincing reason to make a particular prediction.

I will discuss recent work on interpretable predictive modeling with decision lists and sparse integer linear models. I will describe several approaches, including an algorithm based on discrete optimization, and an algorithm based on Bayesian analysis. I will show examples of interpretable models for stroke prediction in medical patients and prediction of violent crime in young people raised in out-of-home care.

Collaborators are Ben Letham, Berk Ustun, Stefano Traca, Siong Thye Goh, Tyler McCormick, and David Madigan.

Categories and Subject Descriptors

I.2.6 [**Learning**]: *Knowledge Acquisition*

H.1.2 [**User/Machine Systems**]: *Human Factors.*

General Terms

Algorithms

KDD'14, Aug 24–27, 2014, New York, NY, USA
ACM 978-1-4503-2956-9/14/08.
http://dx.doi.org/10.1145/2623330.2630823

Keywords

Machine Learning; Interpretability; Comprehensibility; Understandability; Sparsity, Medical Calculators

Bio

Cynthia Rudin is an associate professor of statistics at the Massachusetts Institute of Technology associated with the Computer Science and Artificial Intelligence Laboratory and the Sloan School of Management, where she directs the Prediction Analysis Laboratory. Her expertise is in machine learning and knowledge discovery, and she aims to make predictive modeling closer to decision making. Previously, Prof. Rudin was an associate research scientist at the Center for Computational Learning Systems at Columbia University, and prior to that, an NSF postdoctoral research fellow at NYU. She holds an undergraduate degree from the University at Buffalo where she received the College of Arts and Sciences Outstanding Senior Award in Sciences and Mathematics in 1999, and she received a PhD in applied and computational mathematics from Princeton University in 2004. She is the recipient of the 2013 INFORMS Innovative Applications in Analytics Award, and a 2011 NSF CAREER award. Her work has been featured in IEEE Computer, Businessweek, The Wall Street Journal, the Boston Globe, the Times of London, Fox News (Fox & Friends), the Toronto Star, WIRED Science, Yahoo! Shine, U.S. News and World Report, Slashdot, CIO magazine, and on Boston Public Radio.