Robust Learning with FeatureBoost

Joseph O'Sullivan, John Langford, Rich Caruana, Avrim Blum
School of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213
{*josullvn,jcl,caruana,blum*}*@cs.cmu.edu*

Abstract

PAC learning typically assumes that the training and test sets are drawn from the same distributions. This assumption often is violated in practice. How can machine learning algorithms be biased to output hypotheses that are robust to alterations in the test distribution? We create a framework for learning in environments where the test and training distributions differ because features in the test sets are likely to be missing or corrupted. Motivated by this framework, we present a new meta-learning algorithm, FeatureBoost. We demonstrate FeatureBoost on three learning problems using backprop nets, k-nearest neighbor, and decision trees.

Corresponding Author: Rich Caruana

Category: Algorithms & Architecture

Oral Presentation Preferred

This work has not been submitted or published elsewhere.