Abstract:

We study the general problem of how to learn through experience to make intelligent decisions. In this setting, called the contextual bandits problem, the learner must repeatedly decide which action to take in response to an observed context, and is then permitted to observe the received reward, but only for the chosen action. The goal is to learn through experience to behave nearly as well as the best policy (or decision rule) in some possibly very large and rich space of candidate policies. Previous approaches to this problem were all highly inefficient and often extremely complicated. In this work, we present a new, fast, and simple algorithm that learns to behave as well as the best policy at a rate that is (almost) statistically optimal. Our approach assumes access to a kind of oracle for classification learning problems which can be used to select policies; in practice, most off-the-shelf classification algorithms could be used for this purpose. Our algorithm makes very modest use of the oracle, which it calls far less than once per round, on average, a huge improvement over previous methods. These properties suggest this may be the most practical contextual bandits algorithm among all existing approaches that are provably effective for general policy classes.

This is joint work with Alekh Agarwal, Daniel Hsu, Satyen Kale, John Langford and Lihong Li.

BIO:

Rob Schapire is a Principal Researcher at Microsoft Research in New York City. He received his PhD from MIT in 1991. After a short post-doc at Harvard, he joined the technical staff at AT&T Labs (formerly AT&T Bell Laboratories) in 1991. In 2002, he became a Professor of Computer Science at Princeton University. He joined Microsoft Research in 2014. His awards include the 1991 ACM Doctoral Dissertation Award, the 2003 Gödel Prize, and the 2004 Kanelakkis Theory and Practice Award (both of the last two with Yoav Freund). He is a fellow of the AAAI, and a member of the National Academy of Engineering. His main research interest is in theoretical and applied machine learning, with particular focus on boosting, online learning, game theory, and maximum entropy.