

Reducing Errors by Refusing to Guess (Occasionally)

Dennis Shasha

Department of Computer Science— New York University (NYU)
251 Mercer Street – 10012 – New York – NY – USA

shasha@cs.nyu.edu

Abstract. We propose a meta-algorithm to reduce the error rate of state-of-the-art machine learning algorithms by refusing to make predictions in certain cases even when the underlying algorithms suggest predictions. Intuitively, our SafePredict approach estimates the likelihood that a prediction will be in error and when that likelihood is high, the approach refuses to go along with that prediction. Unlike other approaches, we can probabilistically guarantee an error rate on predictions we do make (denoted the decisive predictions). Empirically on seven diverse data sets from genomics, ecology, image-recognition, and gaming, our method can probabilistically guarantee to reduce the error rate to 1/4 of what it is in the state-of-the-art machine learning algorithm at a cost of between 11% and 58% refusals. Competing state-of-the-art methods refuse at roughly twice the rate of ours (sometimes refusing all suggested predictions).

Dennis Shasha is a Julius Silver Professor of computer science at the Courant Institute of New York University (NYU) and an Associate Director of NYU Wireless. He works on meta-algorithms for machine learning to achieve guaranteed correctness rates, with biologists on pattern discovery for network inference; with computational chemists on algorithms for protein design; with physicists and financial people on algorithms for time series; on clocked computation for DNA computing; and on computational reproducibility. Other areas of interest include database tuning as well as tree and graph matching. Because he likes to type, he has written six books of puzzles about a mathematical detective named Dr. Ecco, a biography about great computer scientists, and a book about the future of computing. He has also written five technical books about database tuning, biological pattern recognition, time series, DNA computing, resampling statistics, and causal inference in molecular networks. He has co-authored over eighty journal papers, seventy conference papers, and twenty-five patents. He has written the puzzle column for various publications including Scientific American, Dr. Dobb's Journal, and the Communications of the ACM. He is a fellow of the ACM and an INRIA International Chair.