SafePredict: A Machine Learning Meta-Algorithm That Uses Refusals to Guarantee Correctness

David Ramirez, Mustafa A Kocak, Elza Erkip, Dennis Shasha

SafePredict: A Machine Learning Meta-Algorithm That Uses Refusals to Guarantee Correctness

David Ramirez (dard@princeton.edu), Mustafa A. Kocak, Elza Erkip, and Dennis E. Shasha

PRINCETON UNIVERSITY BROAD INSTITUTE NEW YORK UNIVERSITY

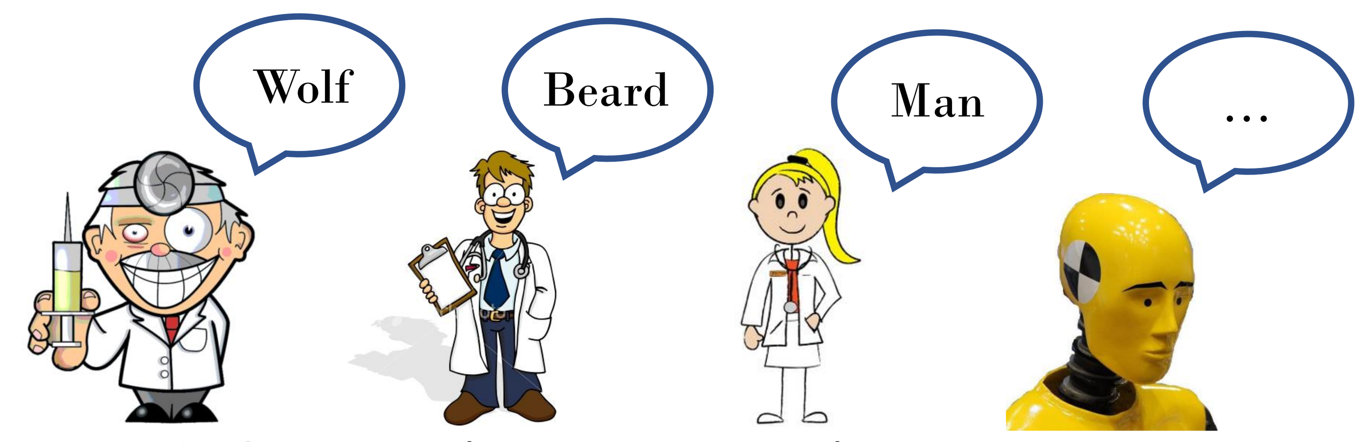
Introduction

Machine learning and prediction algorithms are the building blocks of automation and forecasting.

Algorithms benefit from a lower error rate.

SafePredict, a meta-algorithm, takes predictions from underlying algorithms and decides whether or not to predict with them.

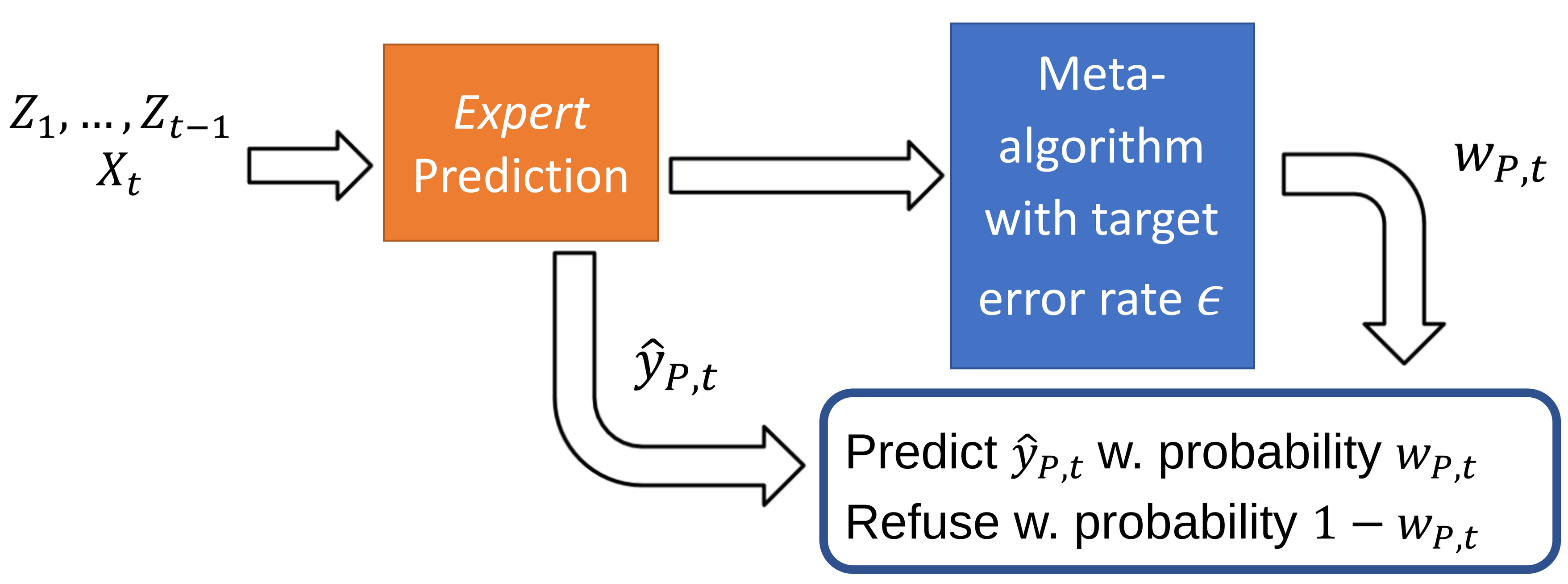
Observation



Crowd of experts (i.e., algorithms) are asked to predict. Dummy expert always *refuses* to predict.

Problem Setup

Online prediction setup with refusal option.



Prediction $\hat{y}_{P,t}$ or refusal $\hat{y}_D$ suffer a loss $l_{P,t},\ l_D \in [0,1]$.
Mistakes are costly, but we learn by observing.

Definitions

t = time index, T = total observations, η = learning rate

$T^* = \sum_{t=1}^{T} w_{P,t}$, expected predictions

$L_T^* = \sum_{t=1}^{T} l_{P,t} w_{P,t}$, expected cumulative loss

$V^* = \sum_{t=1}^{T} w_{P,t} w_{D,t}$, variance for number of predictions

$$w_{P,t+1} = \frac{w_{P,t} e^{-\eta l_{P,t}}}{w_{P,t} e^{-\eta l_{P,t}} + w_D e^{-\eta \epsilon}}$$ weight shift rule

Algorithm Properties

Def. A meta-algorithm is *valid* if, as $T^* \to \infty$, average expected loss $\leq$ target error rate.

Def. A meta-algorithm is *efficient* if, as $T^* \to \infty$, refusals occur only a finite number of times.

Main Results

Safe-Predict is valid and efficient!

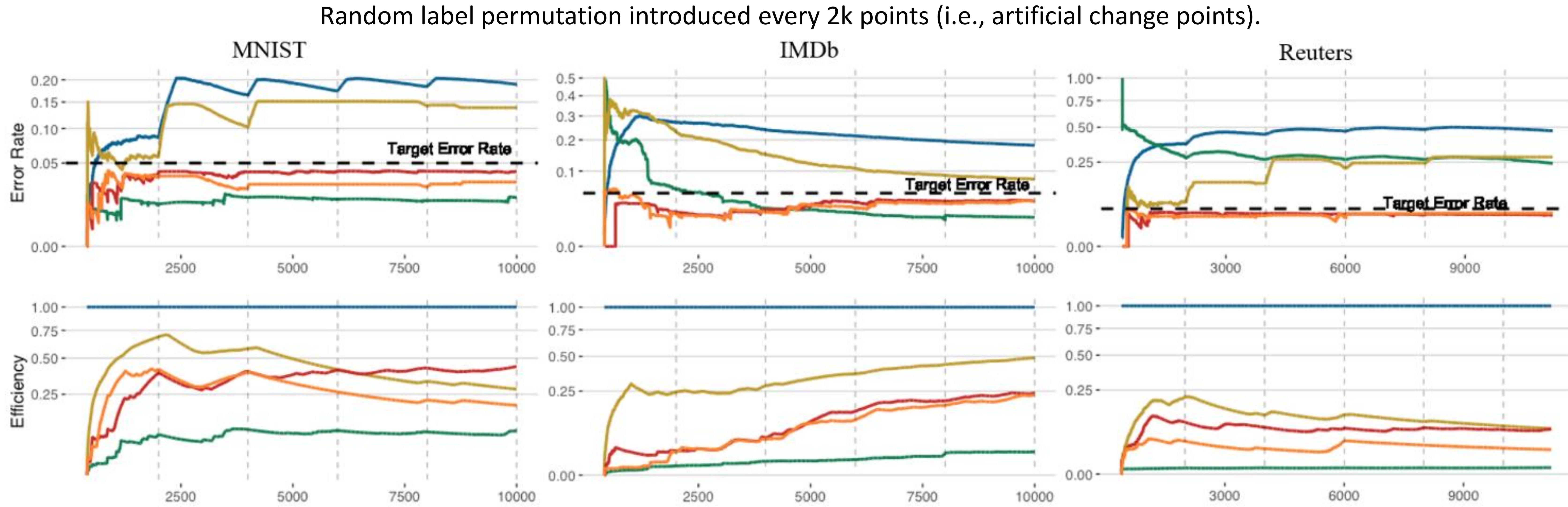
Guaranteed with no assumptions on data or underlying experts, but asymptotic in the number of non-refused predictions.

Theorem 1.- With learning rate $\eta = \Theta(\frac{1}{\sqrt{V^*}})$, *SafePredict* is guaranteed *valid* for any P. Particularly $\frac{L_T^*}{T^*} - \epsilon = O\left(\frac{\sqrt{V^*}}{T^*}\right) = \left(\frac{1}{\sqrt{T^*}}\right)$.

Theorem 2.- If $\limsup_{t\to\infty} \frac{L_{P,t}}{t} < \epsilon$ and $\eta T \to \infty$, then *SafePredict* is *efficient*.

Experimental Results

Randomly permute data, choose first 10k points for experiment. Target error rate $\epsilon = 0.05$.
Random label permutation introduced every 2k points (i.e., artificial change points).



Base Predictor | Confidence Based Refusal (CBR) | *SafePredict* | *SafePredict* + CBR | Amnesic *SafePredict*

References

Mustafa A. Kocak, D. Ramirez, et al., "SafePredict: A Meta-Algorithm for Machine Learning That Uses Refusals to Guarantee Correctness." *Available on arXiv.*
Nick Littlestone, and Manfred K. Warmuth, "The weighted majority algorithm." *Information and computation,* 1994.
Claudio De Stefano, et al., "To reject or not to reject: that is the question-an answer in case of neural classifiers." *IEEE Trans. on Systems, Man, and Cybernetics, Part ,* 2000.
Li, Lihong, et al., "Knows what it knows: a framework for self-aware learning." *Machine learning,* 2011.
Amin Sayedi, et al., "Trading off mistakes and don't-know predictions." *Advances in Neural Information Processing Systems,* 2010.