Optimizing Detection Time and Specificity: Early Classification of Time Series with Sensitivity Constraint

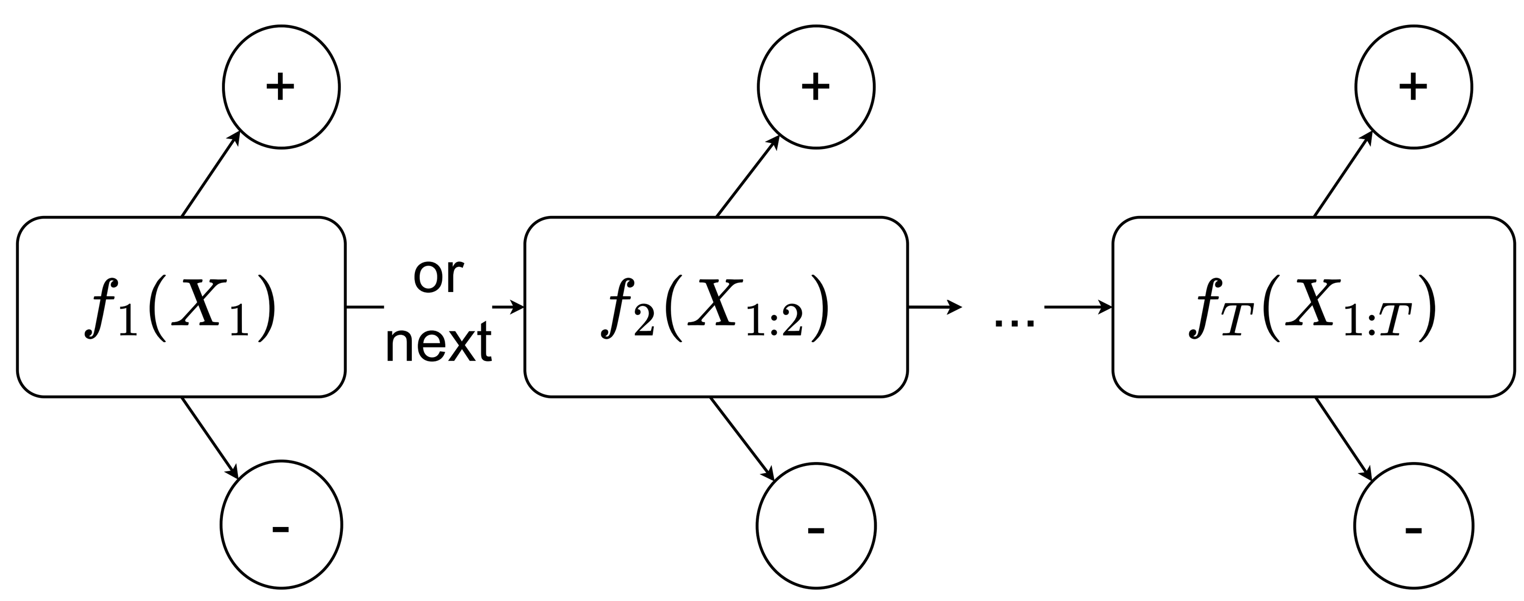
Jiaming Qiu, Ying-Qi Zhao, Yingye Zheng; Public Health Sciences, Fred Hutchinson Cancer Center

Earliest Decision & Highest Accuracy

- Early classification of time series (ECTS) as a multi-objective optimization (MOO) problem.
- Balancing timeliness, specificity, and sensitivity, the optimal trade-off among which quantified by Pareto front.
- Lagrange duality provides optimal solution to the MOO, tractable via neural networks.

Poblem Formulation

Data: finite time series $X_{1:T}$; binary label Y. Sequentially:



- Under some mild conditions, the scores (1) are Pareto optimal that maximize TPR, minimize FPR and cost/detection time simultaneously.
- Collection of all Pareto optima forms the Pareto front, which characterizes the optimal trade-off among the objectives.

- Scores: $f_t = f_t(X_{1:t})$; classification: $\varphi = \sum_{t=1}^{T} \mathbb{1}(f_{0:t-1} = 0, f_t > 0)$; detection time: $\tau = \inf\{1 \le t \le T : f_t \neq 0\}$;
- Objectives: false positive rate (FPR) $\mathbb{E}[\varphi \mid Y = 0]$; true positive rate (TPR) $\mathbb{E}[\varphi \mid Y = 1]$; cost/timeliness $\mathbb{E}[C_\tau]$ given $0 < C_1 < \cdots < C_T$.
- Goal: maximize TPR, minimize FPR and cost (multi-objective).

The optimal scores are

$$\tilde{f}_t(X_{1:t}; a, b) = \begin{cases} 1, & \text{if } \eta_t - aC_t > \nu_t \vee -aC_t, , \\ 0, & \text{if } \nu_t \ge \zeta_t, \\ -1, & \text{if } -aC_t > (\eta_t - aC_t) \vee \nu_t, \end{cases} \tag{1}$$

where a, b are Lagrangian multipliers; for $t = 1, \ldots, T-1$, $\zeta_t := \eta_t^+ - aC_t$; and

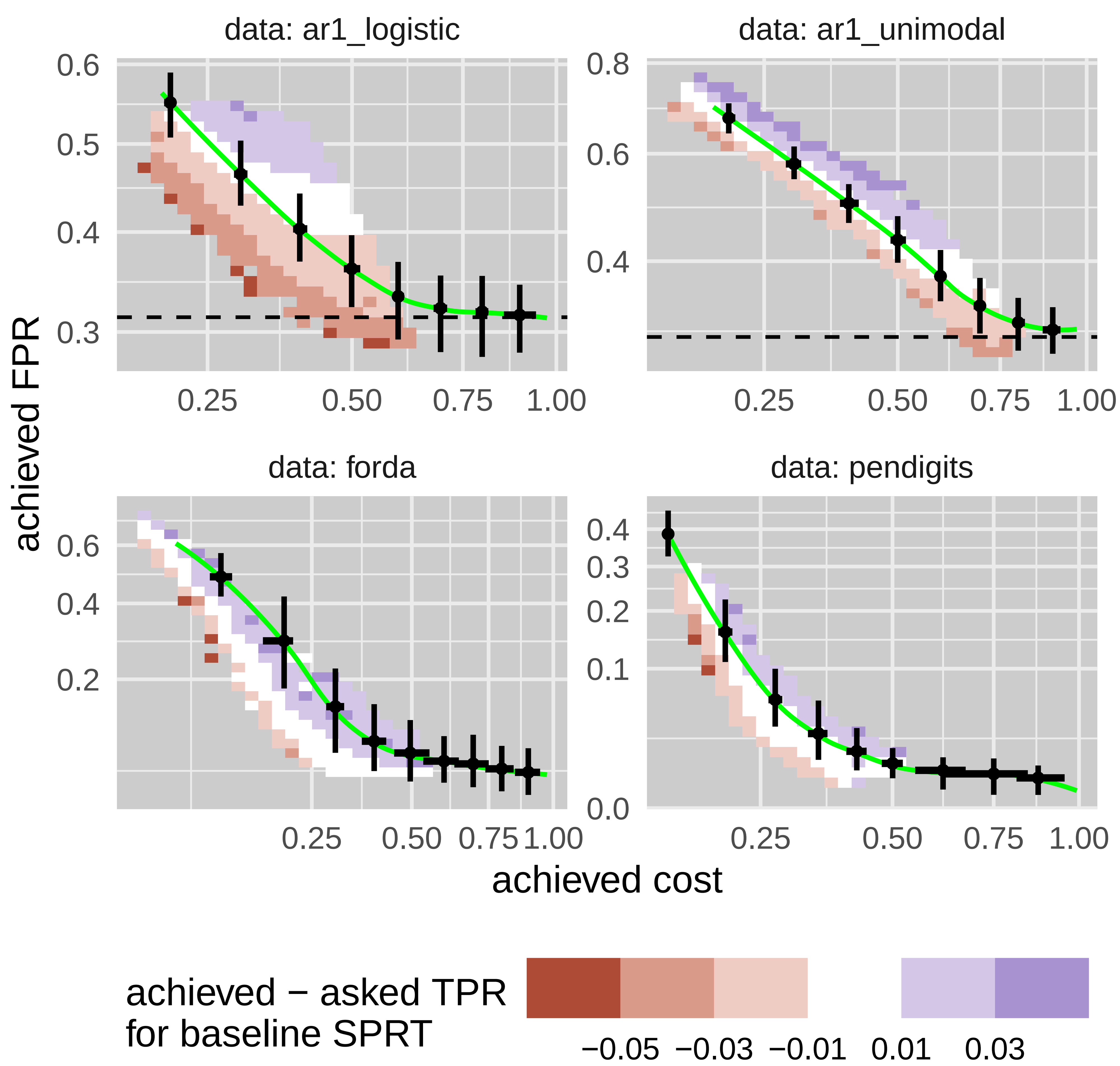
- $p_1 = \mathbb{P}(Y = 1) = 1 - p_0$, $\mu_t := \mathbb{E}[Y \mid X_{1:t}]$, $\eta_t := \left(bp_1^{-1} + p_0^{-1}\right)\mu_t - p_0^{-1}$;
- $S_T := \eta_T^+ - aC_T$, $S_t := \max\left(\eta_t^+ - aC_t, \nu_t\right)$, $\nu_t := \mathbb{E}[S_{t+1} \mid X_{1:t}]$.

Experiments

1. AR(1): X is a stationary AR(1) with $T = 10$ time points; while $\mathbb{P}(Y = 1 \mid X_{1:T}) = \text{sigmoid}\,(c\sum_t X_t)$ (i.e., logistic) or $\exp(-c(\sum_t X_t)^2)$ (i.e., unimodal).
2. Ford-A: sensor data with 500 time points and a binary response, 3601 series. Downsampled to $T = 50$ by blocking every 10 consecutive time points into a 10-dim vector.
3. Pendigits: X is a sequence of 2-dimensional coordinates of pen strokes when writing digits ($T = 8$); Y is whether the digit written is 1, 2, 3, 5, 7. Total 10992 series.

Elbow shape of the **Pareto front** $\Longrightarrow$

- descreases of FPR eventually deminished as cost grows;
- do not need the entire sequence to reach full-length performance.



Right figures: with TPR = 0.9, the estimated Pareto front (green lines); the 90% interval for achieved FPR (vertical crosses) and cost (horizontal ones) over 100 repeats; the Wald's SPRT performance (red-white-blue pixels); and the best-possible FPR when the true class probability is available (horizontal dashed lines).