Learning Optimal Predictive Checklists

Haoran Zhang[1], Quaid Morris[2], Berk Ustun[3,*], Marzyeh Ghassemi[1,4,*]

[1]MIT [2]MSKCC [3]UCSD [4]Vector Institute *Equal Supervision

Motivation

- Checklists are simple tools that are often used in clinical applications because they are easy to *use*, *deploy*, and *scrutinize*.
- The large majority of checklists are hand-crafted by panels of a experts (Huang et al., 2008, Blevins et al., 2015).
- However, creating checklists by hand is often a complex and time-consuming process.
- Learning checklists from data allows us to quickly create checklists that can be evaluated using measurable objectives.

We develop a method to learn predictive checklists from data.

Checklists as Simple Linear Models

Data	$(\boldsymbol{x}_i, y_i)_{i=1}^n \qquad \boldsymbol{x}_i \in \{0,1\}^d \quad y_i \in \{0,1\}$
	e.g. Age ≥ 30, Age ≥ 45, Age ≥ 65, Systolic BP ≥ 140, etc
Model	$\hat{y}_i = \mathbb{1}(\boldsymbol{\lambda}^T \mathbf{x}_i \geq M)$
Coefficients	$\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_d] \in \{0,1\}^d$
	where $\lambda_j = 1 \iff$ checklist uses item j

ERM Problem

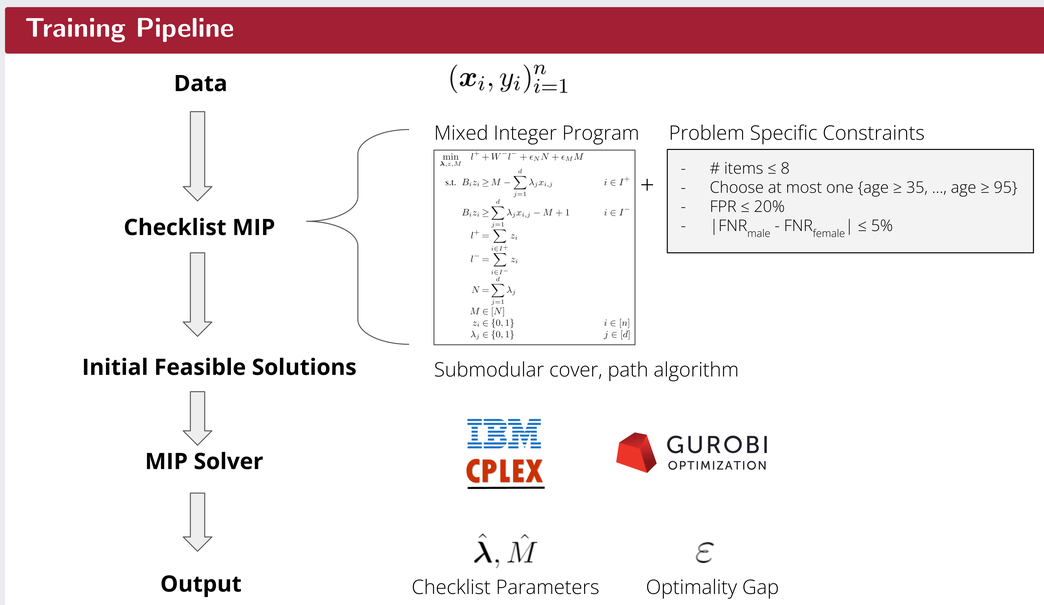
$$\min_{\boldsymbol{\lambda}, M} \; l(\boldsymbol{\lambda}, M, X, Y) + \varepsilon_N N + \varepsilon_M M \qquad \text{0-1 loss}$$

$$\text{s.t.} \quad \boldsymbol{\lambda} \in \{0,1\}^d \qquad \text{Whether to select item}$$

$$N = \sum_{i=1}^{d} \lambda_i \qquad \text{Total \# items in checklist}$$

$$M \in \{1, ..., N\} \qquad \text{Threshold for positive prediction}$$

Training Pipeline



Customizable Constraints

	Example	Constraint
Model Size	Use $\leq N_{max}$ items	$N \leq N_{max}$
Binarization	Choose at most one Age feature	$\lambda_{age \geq 45} + \lambda_{age \geq 60} + \ldots \leq 1$
Performance	FPR $\leq \beta$	$l^- \leq \lceil \beta \cdot n^- \rceil$
Group Fairness	Max FPR disparity of δ between males and females	$\left\lvert \frac{l_M^-}{n_M^-} - \frac{l_F^-}{n_F^-} \right\rvert \leq \delta$
Minimax Fairness	No group with FNR worse than δ	$l_g^+ \leq \lceil \delta \cdot n_g^+ \rceil \quad \forall g \in G$

Demo: Fair Mortality Prediction

Goal: Predict in-hospital mortality in ICU patients given Continuous Renal Replacement Therapy (CRRT) while maintaining fairness between intersectional groups.

(a) No Fairness Constraints

Predict Mortality Given CRRT if 3+ Items are Checked	
Age ≥ 66.0 years	☐
AST ≥ 162.6 IU/L	☐
Blood pH ≤ 7.29	☐
MCV ≥ 99.0 fl	☐
Norepinephrine ≥ 0.1 mcg/kg/min	☐
Platelets ≤ 65.0 $\times 10^3/\mu L$	☐
RDW ≥ 19.2%	☐
Time in ICU ≥ 14.1 hours	☐

Constraints
- FNR ≤ 20%
- At most one item per feature
- Use at most 8 items

Objective: Minimize FPR

	FNR	FPR	Worst FNR	Max FPR Gap
Training	20.0%	43.9%	33.3%	24.3%
Test	22.2%	52.6%	62.5%	54.5%

(b) With Fairness Constraints

Predict Mortality Given CRRT if 2+ Items are Checked	
ALT ≥ 16.0 IU/L	☐
Bicarbonate ≤ 17.0 mmol/L	☐
Blood pH ≤ 7.22	☐
Norepinephrine ≥ 0.1 mcg/kg/min	☐
RDW ≥ 19.2%	☐
Time in ICU ≥ 117.3 hours	☐

Constraints
- FNR ≤ 20%
- At most one item per feature
- Use at most 8 items
- FPR gap ≤ 15%
- Worst FNR ≤ 20%

Objective: Minimize FPR

	FNR	FPR	Worst FNR	Max FPR Gap
Training	17.5%	52.2%	18.1%	13.9%
Test	19.6%	55.1%	50.0%	38.3%