OXFORD UNIVERSITY
DATA SCIENCE AND DECISIONS
RESEARCH GROUP

Boosted Trees for Risk Prognosis

Alexis Bellot and Mihaela van der Schaar

Machine Learning for Health Care 2018

August 22, 2019

Objective: Survival Analysis

Predictions should be individualized

OXFORD UNIVERSITY
DATA SCIENCE AND DECISIONS
RESEARCH GROUP

We study the survival distribution which describes the time to event probability as a function of a patient's covariates,

$$S(t|\boldsymbol{x}_i) = \mathbb{P}(T_i > t|\boldsymbol{x}_i) \tag{1}$$

Problem: Heterogeneous patients.

- The average predictions of many current methods are not enough to accurately assess a patient's health state and progression.
- Complex diseases often result in nonlinear relationships between $\boldsymbol{x}_i$ and T_i.
- Simple approximations of S lead to misdiagnoses for large portions of patients with atypical disease presentation.

Examples: **chronic** or **multimorbid** patients whose risk factors are poorly understood (e.g. cardiovascular diseases and elderly patients).

Our Approach

Flexible Boosting Algorithm for Survival Analysis

OXFORD UNIVERSITY
DATA SCIENCE AND DECISIONS
RESEARCH GROUP

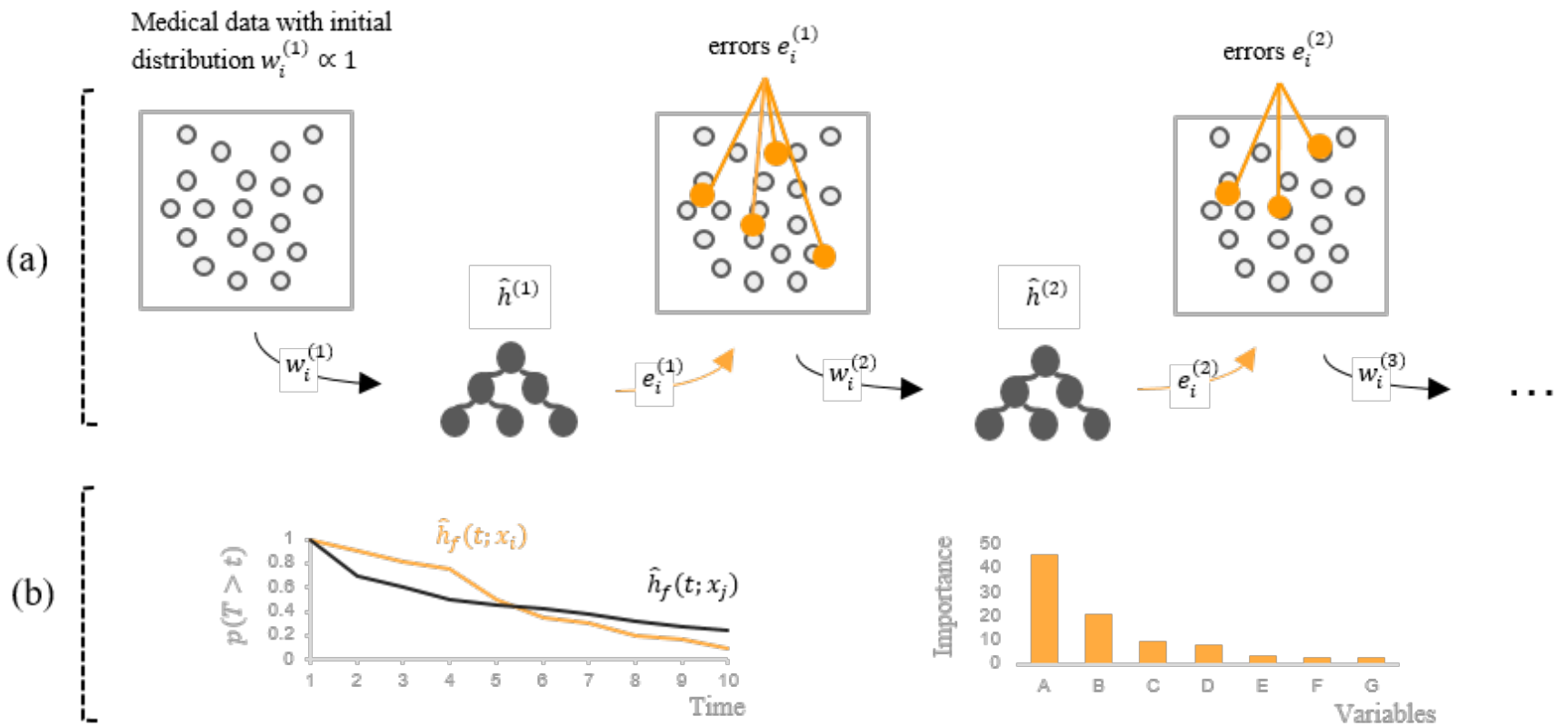
Focus on complex patterns and subgroups of patients that are consistently being mis-estimated

• We estimate a collection of nonparametric survival estimators sequentially where *each one of them* is designed to improve predictions on those patients that have been previously mis-estimated.

Advantages

- ▷ Efficient scheme for learning in high-dimensional settings.
- ▷ No a-priori assumptions on patient behaviour.
- ▷ Very flexible and thus able to provide *individualized* predictions.

Outline of the Algorithm



Experiments and Results

Performance improvements for patients at risk of Cardiovascular diseases

OXFORD UNIVERSITY
DATA SCIENCE AND DECISIONS
RESEARCH GROUP

- **Preventive care** : UK Biobank and MAGGIC.
- **End stage cardiac patients**: UNOS.
- **Co-morbid patients**: SEER.

Models	UNOS	MAGGIC	UK Bio.	SEER-I	SEER-II
Cox	0.603 ± 0.04	0.645 ± 0.01	0.679 ± 0.02	0.772 ± 0.03	0.740 ± 0.03
CBL	0.605 ± 0.04	0.644 ± 0.01	0.679 ± 0.02	0.774 ± 0.03	0.738 ± 0.04
CBM	0.635 ± 0.03	0.625 ± 0.01	0.673 ± 0.02	0.768 ± 0.03	0.740 ± 0.04
CindexBoost	0.564 ± 0.06	0.592 ± 0.01	0.655 ± 0.03	0.764 ± 0.03	0.742 ± 0.04
SRF	0.634 ± 0.04	0.642 ± 0.01	0.627 ± 0.01	0.686 ± 0.03	0.680 ± 0.01
CSRF	0.635 ± 0.05	0.652 ± 0.02	0.638 ± 0.02	0.755 ± 0.03	0.717 ± 0.04
SurvivalBoost.R	0.636 ± 0.03	0.676 ± 0.02	0.702 ± 0.02	$\mathbf{0.780 \pm 0.03}$	$\mathbf{0.752 \pm 0.03}$
SurvivalBoost.T	$\mathbf{0.647 \pm 0.04}$	$\mathbf{0.675 \pm 0.04}$	$\mathbf{0.725 \pm 0.03}$	0.775 ± 0.04	0.740 ± 0.04

Table 1: C-index figures (higher better) and standard deviations on all data sets.

COME SEE ME AT MY POSTER TO LEARN MORE

Try our survival prediction tool at:

mlhcprojects.shinyapps.io/survival_boosting_app