
Teaching Models To Survive: Proper Scoring Rule and Stochastic Optimization with Competing Risks

Julie Alberge*¹

¹Inria Saclay – Equipe Soda – France

Résumé

In medical research, predicting the patient outcome is often complicated because of the presence of right-censored data, where certain outcomes remain unobserved. Survival analysis addresses this issue by estimating the "time to event," predicting the time when the event of interest may occur for the patient (e.g. the time of death).

However, when multiple possible events or outcomes can occur (e.g., the cause of death), a more complex scenario, known as competing risks, emerges. This framework aims to predict not only when the first event might occur but also which specific event is the most likely to occur first.

Competing risks remain relatively underexplored compared to survival analysis.

Here, our work focuses on the competing risks setting, with a focus on the estimation of event probabilities because those are essential for personalized medicine and decision-making.

To achieve this, we introduce a strictly proper, censoring-adjusted separable scoring rule designed for competing risks. This scoring rule can be optimized on a subset of the data because the evaluation is made independently of observation which makes it more efficient, contrary to the c-index that can only be assessed on the complete dataset. This approach facilitates stochastic gradient boosting trees, enhancing model flexibility and scalability.

We compared our proposed model, SurvivalBoost, to 12 alternative methods, and it achieves state-of-the-art performance in estimating the probability of outcomes under both survival analysis and competing risks frameworks.

SurvivalBoost not only provides accurate predictions across any time horizon but also significantly reduces computational time compared to existing alternatives, allowing it to be used on datasets with several million patients, which are now becoming more prevalent with the use of electronic health records or claims data. Its ability to model competing risks with high precision makes it a valuable tool for medical applications where multiple potential outcomes must be considered in patient care and prognosis.

A Python library called 'hazardous' is available with the model and several examples.

*Intervenant