DEEP LEARNING FOR PATIENT-SPECIFIC KIDNEY GRAFT SURVIVAL ANALYSIS

Group Members:

Nishat Zaman (011 151 264)
Nawshiba Tasnim Ahmed (011 161 145)
Ferdous Zaman (011 162 125)
Ebnul Mahmood Shovan (011 162 058)

Abstract

In this paper, they propose a deep learning method that directly models the survival function instead of estimating the hazard function to predict survival times for graft patients based on the principle of multi-task learning. By learning to jointly predict the time of the event, and its rank in the cox partial log likelihood framework, their deep learning approach outperforms, in terms of survival time prediction quality and concordance index.

INTRODUCTIONS

Survival data are characterized by the fact that we do not know many of the outcome values (e.g. death or disease recurrence in medical studies) because the event might not have occurred within the fixed period of the study or because subjects could move out of town or decide to drop out at any time. Instead, the date of the last visit (censoring time) provides a lower bound on the survival time, defined as the length of time between two events, called intake and endpoint. Such datasets are considered censored. There are various mode for survival time analysis. These models however suffer from several limitations when the objective is to get a precise estimation of the survival time of an individual patient. Two main issues of such models. The distributions chosen to model the data make it hard to provide accurate predictions in terms of survival time.

INTRODUCTIONS

In this paper they introduce a novel deep learning method for survival analysis, based on multi-task learning. They jointly predict the time of the event, and its rank in the cox partial log-likelihood framework. This allows better generalization, due in part to the fact that the model is able to account for the temporal aspect of the predictions. They compute the concordance index metric (C-index in what follows) in order to compare their results with the widely adopted Cox model. They demonstrate the performance of their method experimentally on real-life survival datasets where it yields better results in terms of C-index than the previous state-of-the-art methods.

A Deep Learning Survival Model

In this section they describe their main contribution. As opposed to most previous methods that attempt to estimate the survival or hazard function, they construct a deep neural network model to directly compute the time of the event (here, graft failure). Their model attempts to predict the probability of being alive at time t, for $t \in [0, T]$. The chosen modeling task is related to the problem of ordinal regression, with the exception that they have to take into account censored events.

A Deep Learning Survival Model

Model

The proposed model takes as input the different continuous and discrete features characterizing a patient and, in the case of the main dataset of their paper, a donor-recipient couple. The second-to-last layer consists of a single unit with linear activation. The value outputted, denoted s (1) in what follows is used to estimate the hazard function and can be considered a score indicative of the time of graft failure, and thus compare two patients. This allows them to compute the first loss, the Cox partial log-likelihood. The final layer has T units, where T is the number of time units (in our case, years

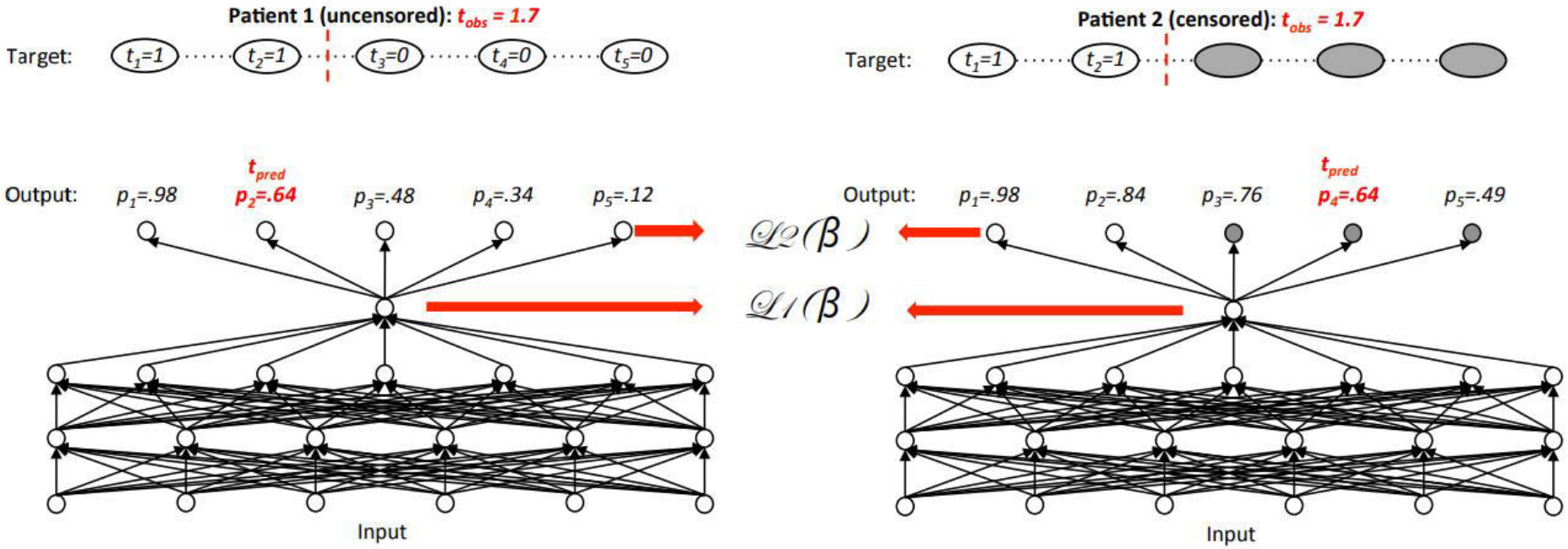
A Deep Learning Survival Model

Model

or months) considered in the study. The output is denoted s (2) in what follows. We use sigmoid activations for the units of the final layer. The output value at index t corresponds to the probability of not experiencing graft failure at time t (in our case at the t-th year). The second loss penalizes wrongly predicted times in a manner consistent with losses used in isotonic regression. The model is shown in a simplified form (removing some layers for clarity) in Figure 1. The two cases (censored and non-censored) are shown in the figure, which illustrates that when some right part of the history is missing, no loss and no gradient is computed for those output units.

A Deep Learning Survival Model

Model



A Deep Learning Survival Model

Figure 1: Our deep learning model. (Left) The case of an uncensored patient. (Right) The case of a censored patient. The greyed out output units in the censored case correspond to censored units for which no loss is computed and no error back-propagation is done. Two losses are computed: a ranking loss on the bottleneck scalar unit and a cumulative time prediction on the top output units.

A Deep Learning Survival Model

Handling ties and censored data

A Deep Learning Survival Model

Evaluation procedure

They first split the dataset into training (80%) and test (20%) sets in which the percentage of uncensored patients. We performed hyperparameter selection and early stopping on a subset of the training set (validation set corresponding to 20% of the total dataset and having the same proportion of uncensored patients and events occurring per time-step).

Survival Time Analysis

survival data analysis has three main characteristics: (1) it examines the relationships of survival distributions to features; (2) it models the time it takes for events to occur, and (3) the event we want to predict (such as time of death) is not always observed. Sometimes, a patient will drop out of the study (i.e., voluntarily or because he was still alive at the end of the study). We call such datasets right-censored. As medical events are granular by nature, a given unique time (in a given time unit, such as a month for their dataset) can correspond to multiple events (such as having 10 patients reject their graft on January). Such events are tied, making comparisons more complex, and requiring a modification to the loss function.

Survival Time Analysis

Linear models: Cox proportional hazards model

Cox proportional hazards model is not commonly used in the literature to perform prediction on new cases, but rather to characterize disease progression on existing cases, by highlighting the importance of the different features. As the baseline hazard function is never directly estimated, computing survival predictions is not directly possible without additional assumptions.

Survival Time Analysis

Non-linear models

Deep survival is an extension of the Cox model that uses a deep neural network to parametrize the hazard function. The part of their model that is trained on a Cox partial likelihood loss differs from other approach in the following regards:

- We perform the optimization per-batch instead of on the full dataset.
- We adapt the loss function to account for ties, using Efron's approximation.
- We did not limit ourselves to modeling the logarithm of the hazard function, and tried other forms. Modeling the hazard function directly yielded an improvement.

Ranganath uses deep exponential families (i.e., a class of latent variable models inspired by the hidden structures used in deep neural networks) to model event time.

Survival Time Analysis

Standard evaluation: the concordance index

The C-index is a standard measure in survival analysis that estimates how good the model is at ranking survival times by calculating the probability of correctly ranking the event time of cases taken two at a time. Let Ti denote the survival time for individual i and Ei be the associated event, censored or uncensored. The (Ti , Ei), ...,(Tn, EN) are all the events in the dataset. Considering all possible pairs (Ti , Ei),(Tj , Ej) for $i \leq j$, the C-index is calculated by considering the number of pairs correctly ordered by the model divided by the total number of admissible pairs. For our particular case of right censoring, a pair is considered admissible if it can be ordered in a meaningful way. A pair cannot be ordered if the events are both right-censored or if the earliest time in the pair is censored. A tied pair is counted as half correct in accordance with standard implementations of the C-index. Finally, a C-index equal to 1 indicates perfect prediction whereas a C-index equal to 0.5 indicates a random prediction. Survival models typically yield a C-index between 0.6 and 0.7.