Expert Systems With Applications 115 (2019) 16–26

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A Deep Active Survival Analysis approach for precision treatment recommendations: Application of prostate cancer

Milad Zafar Nezhad [a,*], Najibesadat Sadati [a], Kai Yang [a], Dongxiao Zhu [b]

[a] Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth Street, Detroit, MI 48202, USA
[b] Department of Computer Science, Wayne State University, 5057 Woodward Ave, Detroit, MI 48202, USA

ARTICLE INFO

Article history:
Received 10 May 2018
Revised 12 July 2018
Accepted 30 July 2018
Available online 31 July 2018

Keywords:
Survival analysis
Deep learning
Active learning
Treatment recommendation
Electronic health records
Prostate cancer

ABSTRACT

Survival analysis has been developed and applied in the number of areas including manufacturing, finance, economics and healthcare. In healthcare domain, usually clinical data are high-dimensional, sparse and complex and sometimes there exists few amount of time-to-event (labeled) instances. Therefore building an accurate survival model from electronic health records is challenging. With this motivation, we address this issue and provide a new survival analysis framework using deep learning and active learning with a novel sampling strategy. First, our approach provides better representation with lower dimensions from clinical features using labeled (time-to-event) and unlabeled (censored) instances and then actively trains the survival model by labeling the censored data using an oracle. As a clinical assistive tool, we introduce a simple effective treatment recommendation approach based on our survival model. In the experimental study, we apply our approach on SEER-Medicare data related to prostate cancer among African–Americans and white patients. The results indicate that our approach outperforms significantly than baseline models.

1. Introduction

Survival analysis has been applied in several real-world applications such as healthcare, manufacturing and engineering in order to model time until the occurrence of a future event of interest (e.g., biological death or mechanical failure) (Hosmer, Lemeshow, & May, 2011). Censoring attribute of survival data makes survival analysis different from the other prediction approaches. One popular survival model is the Cox Proportional Hazards model (CPH) (Cox, 1992) which models the risk of an event happening based on linear combination of the covariates (risk factors). The major problem of Cox-based models is linear relationship assumption between covariates and the time of event occurrence. Hence, there have been developed several models to handle non-linear relationship in survival analysis like as survival neural network and survival random forest models (Ishwaran et al., 2014).

In the healthcare area, medical researchers applied survival analysis on Electronic Health Records (EHRs) to evaluate the significance of many risk factors in outcomes such as survival rates or cancer recurrence and subsequently recommend treatment schemes. There exist two specific challenges in survival analysis from EHRs: (1) Clinical data is usually high dimensional, sparse and time-dependent which in this case applying traditional survival approaches do not perform well enough to estimate the risk of a medical event, (2) In many health survival applications, labeled data (time-to-event instances) are small, time-consuming and expensive to collect. In this situation, it is hard to learn a survival model based on traditional approaches which able to predict the relative risk of patients precisely.

To address the first challenge, recently, semi-supervised learning using deep feature representation has been applied in number of areas and could improve the performance of different machine learning tasks as well as survival analysis. In other words, applying unsupervised learning using deep learning can reduce the complexity of raw data and provide robust features with lower dimensions (LeCun, Bengio, & Hinton, 2015). Using these represented features in the supervised learning algorithms (e.g., survival models) establishes a semi-supervised learning framework which achieves higher performance.

To overcome the second challenge, active learning is well suited to get high accuracy when the labeled instances are small or labeling is expensive and time-consuming (Settles, 2010). Active learning approach from censored data has been rarely addressed in the literature. However it has been widely used in the other aspects of health informatics where the labeled data are scarce.

* Corresponding author.
E-mail addresses: m.zafarnezhad@wayne.edu (M.Z. Nezhad), n.sadati@wayne.edu (N. Sadati), kai.yang@wayne.edu (K. Yang), dzhu@wayne.edu (D. Zhu).

https://doi.org/10.1016/j.eswa.2018.07.070
0957-4174/

According to the current works in the literature, no research has been conducted to overcome these two challenges by developing an integrated method for survival analysis. Although, there exist few studies in the literature that focus to one of these challenges, our goal is to address both of them simultaneously. In addition, several applications (especially in the healthcare domain) demand to develop such that integrated approach when they deal with a few amount of labeled instances that are high-dimensional and training a precise survival analysis model is difficult based on the current baselines. To address this research gap, first, we propose a novel survival analysis approach using deep learning and active learning termed DASA. Our method is capable to learn more accurate survival model using high dimensional and small size EHRs in comparison with some baseline survival approaches. Second, we introduce a personalized treatment recommendation approach based on our survival analysis model which can compare the relative risks (or survival times) associate with different treatment plans and assign the better one. We evaluate our approach using SEER-Medicare dataset related to prostate cancer. We consider two racial subgroup of patients (African-American and whites) in our analysis and apply our model on each dataset separately.

Our contributions in this research lie into three folds: (1) To the best of our knowledge, we propose the first Deep Active Survival Analysis approach with promising performance, (2) In our active learning framework we develop a new sampling strategy specifically for survival analysis and (3) Our model with proposed treatment recommendation approach has highly potential to apply for evaluation of new treatment effect on new patients where the labeled data is scarce.

2. Background

In this section, we review some basic concepts for modeling of survival analysis, active learning and deep learning.

2.1. Introduction to survival analysis

Survival analysis is a kind of statistical modeling where the main goal is to analyze and model time until the occurrence of an event of interest. The challenging characteristic of survival data is the fact that time-to-event of interest for many instances is unknown because the event might not have happened during the period of study or missing tracking occurred caused by other events. This concept is called censoring which makes the survival analysis different (Wang, Li, & Reddy, 2017). The special case of censoring is when the observed survival time is less than or equals to the true event time called right-censoring, the main focus of our study.

Since the censored data is present in survival analysis, the standard statistical and machine learning approaches are not appropriate to analyze and predict time-to-event outcome because those approaches miss the censored/right-censored instances. Survival modeling provides different statistical approaches to analyze such censored data in many real-world applications.

In survival analysis, a given instance i, represented by a triplet (X_i, δ_i, T_i) where X_i refers to the instance characteristics and Ti indicates time-to-event of the instance. If the event of interest is observed, Ti corresponds to the time between baseline time and the time of event happening, in this case $\delta_i = 1$. If the instance event is not observed and its time to event is greater than the observation time, Ti corresponds to the time between baseline time and end of the observation, and the event indicator is $\delta_i = 0$. The goal of survival analysis is to estimate the time to the event of interest (T) for a new instance X_j (Wang et al., 2017).

2.1.1. Survival and hazard functions

Survival and hazard functions are the two main functions in survival modeling. The survival function indicates the probability that the time to the event of interest is not less than a determined time (t) (Kleinbaum & Klein, 2010). This function (S) is denoted by following formula:

$$S(t) = Pr(T > t) \tag{1}$$

The initial value of survival function is 1 when $t = 0$ and it monotonically decreases with t. The second function, hazard function indicates the rate of occurrence of the event at time t given that no event occurred earlier. It describes the risk of failure (dying) changing over time. The hazard function (or hazard rate or failure rate) is defined as following (Kleinbaum & Klein, 2010):

$$h(t) = \lim_{\delta(t)\to 0} \frac{Pr(t \le T \le t + \delta(t) | T \ge t)}{\delta(t)} \tag{2}$$

Survival and hazard function are non-negative functions. While the survival function decreases over time, the shape of a hazard function can be in different forms: increasing, decreasing, constant, or U-shaped.

2.1.2. Cox Proportional Hazards (CPH) model

There exist several models for survival analysis in the literature. Among all, Cox Proportional Hazards (CPH) model (Cox, 1992) is the most popular model for survival analysis. CPH estimates the hazard function $h(x)$ as a regression formulation:

$$h(t, X_i) = h_0 \; exp(X_i \beta) \tag{3}$$

where h_0 is the baseline hazard function which can be an arbitrary nonnegative function of time and X_i refers to covariate vector for instance i, and β is the coefficient vector estimated after survival model training by maximizing the cox partial likelihood. Because the baseline hazard function $h_0(t)$ in CPH is not determined, we cannot use the standard likelihood function in training process. The partial likelihood is the product of the probability of each instance i at event time T_i that the event has happened for that instance, over the summation of instances (R_j) probability who are still at risk in this time (T_i) (Cox, 1992):

$$L(\beta) = \prod_{i=,\delta_i=1} \frac{exp(X_i \beta)}{\sum_{j \in R_j} exp(X_j \beta)} \tag{4}$$

2.1.3. Evaluation metric for survival analysis

Since the censored instances exist in survival data, the standard evaluation metrics such as mean squared error and R-squared are not appropriate for evaluating the performance of survival analysis (Heagerty & Zheng, 2005). In survival analysis, the most popular evaluation metric is based on the relative risk of an event for different instances called concordance index or c-index. This measure is defined as following formula:

$$\frac{1}{N} \sum_{i,\delta_i=1} \sum_{j, y_i < y_j} I[S(\hat{y}_i | X_i) < S(\hat{y}_j | X_j)] \tag{5}$$

Where N refers to the all comparable instance pairs and S is the survival function. The main motivation for using c-index in survival analysis is originated from the fact that the medical doctors and researchers are often more interested in measuring the relative risk of a disease among patients with different risk factors, than the survival times of patients.

In general, the survival analysis models can be divided into two main categories: (1) statistical methods including non-parametric, semi-parametric and parametric and (2) machine learning based methods such as survival trees, Bayesian methods, neural networks and random survival forests. Readers for more comprehensive review can refer to the recent review provided by Wang et al. (2017).