IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 5 (May. 2019), ||S (XI) || PP 56-60

Application of Machine Learning For Survival Analysis- A Review

Mrs. R.A.Fadnavis,
Assistant professor, Yeshwantrao Chavan College of Engineering , Nagpur, Maharashtra

Abstract: *There is exponential growth in data over the last decade. Due to the advancements in various data acquisition and storage technologies, different disciplines have attained the ability to not only accumulate a wide variety of data but also to monitor observations over longer time periods. In many real-world applications, the primary objective of monitoring these observations is to estimate when a particular event of interest will occur in the future. Survival analysis is a model for time until a certain "event.".Time-to-event data encounters several research challenges such as censoring , instance/feature correlations, high-dimensionality, temporal dependencies, and difficulty in acquiring sufficient event data in a reasonable amount of time. The machine learning techniques are being used recently to develop more sophisticated and effective algorithms that either complement or compete with the traditional statistical methods in survival analysis.This paper proposes overview of fundamental and state of art techniques used for survival analysis.*

Keywords: *Survival data; censoring; survival analysis; machine learning, statistical methods*

I. Introduction

Survival analysis is a branch of statisticsused for analyzing the expected duration of time until one or more events of interest happen. Survival analysis (time-to-event analysis) is widely used in economics and finance, engineering, medicine and many other areas. It is used to model many different events like time from birth to death in biological organisms,time after cancer treatment until death,Time from first heart attack to the second,Time until tumor recurrence, failure in mechanical systems,and time from manufacture of a component to component failure. In this context, death or failure is considered an "event" in the survival analysis literature. Generally only a single event occurs for each subject, after which the organism or mechanism is dead or broken. The response is often referred toas a failure time, survival time, or event time.Survival time can be measured in years, months, days, or even fractions of a second.

As well as estimating the time taken to reach a certain event, Survival analysis can also be used to compare time-to-event for multiple groups. For example, two production lines for products could be compared to see if there is a different in lifetimes. In medicine, two groups with different attributes (like normal/overweight, diabetic/non-diabetic, high/low blood pressure) could be compared to see how those factors contribute to survival time in patients with different diseases.

One of the main challenges for survival data is that it is not completely observed, Instead some of the data is censored.i. e there exist censored instances. So the eventof interests is not observed for these instances due to either the time limitation of thestudy period or losing track during the observation period. Censoring refers to missing data in a study such as subjects dropping out of trials or data that is otherwise lost. Censoring is present when we have some information abouta subject's event time, but we don't know the exact event time. The reasons for occurrence of censoring can be :

A subject does not experience the event before the studyends
A person is lost to follow-up during the study period
A person withdraws from the study

These are called right censored subjects. They are usually counted as active for the duration of the study for purposes of data analysis.

Another type of censoring happens when a subject simply doesn't experience the event in question during the study. In other words, the time to event is incomplete. Therefore, it is not suitable to directly apply predictive algorithmsusing the standard statistical and machine learning approaches to analyze the survival data.

1.1 Terminology and Notation

1. The survival function S(t)is the probability that a subject survives longer than time t.
T denotes the response variable, $T >= 0$.

The survival function is: S(t) = Pr(T > t) = 1 − F(t).

As t ranges from 0 to 1, the survival function has thefollowing properties:
It is non-increasing
At time t = 0, S(t) = 1. In other words, the probabilityof surviving past time 0 is 1.
At time t = 1, S(t) = S(1) = 0. As time goes toinfinity, the survival curve goes to 0.
In theory, the survival function is smooth. In practice, events can be observed on a discrete time scale (days, weeks, etc.).

2. **The hazard function, h(t),** is the instantaneous rate at whichevents occur, given no previous events

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

3. The cumulative hazard describes the accumulated risk up to time t, $H(t) = \int_0^t h(u)du.$

If any one of the functions S(t), H(t), or h(t), is known then the other two functionscan be derived :

$$h(t) = -\frac{\partial \log(S(t))}{\partial t}$$

$$H(t) = -\log(S(t))$$

$$S(t) = \exp(-H(t))$$

Recording and Representing survival data with censoring:

- Ti denotes the response for the ith subject.
- Let Ci denote the censoring time for the ith subject
- Let δ_i denote the event indicator

$$\delta_i = \begin{cases} 1 \text{ if the event was observed } (T_i \leq C_i) \\ 0 \text{ if the response was censored } (T_i > C_i). \end{cases}$$

- The observed response is Yi = min(Ti,Ci).

II. Methods and Applications of Survival Analysis

Survival analysis methods can be classified into two main categories: statistical methods and machine learning based methods. In addition to the difficulty in handling the censored data, there are also several unique challenges to perform the predictive modeling with survival data and hence machine learning researchers have developed more sophisticatedand effective algorithms which complement or compete with the traditionalstatistical methods.

Statistical methods and machine learning methods both are used to make predictions of thesurvival time and estimate the survival probability at the estimated survival time. However, Statistical methods focus more on characterizing both the distributions of the event times and the statistical properties of the parameter estimation by estimating the survival curves, while machine learning methods focus more on the prediction of event occur renceat a given time point by incorporating the traditional survival analysis methods with various machine learning techniques. Also, Machine learning methods are usually applied to the high-dimensional problems, while statistical methods are generally developed for the low-dimensional data. In addition, machine learning methods for survival analysis offer more effective algorithms by incorporating survival problems with bothstatistical methods and machine learning methods and taking advantages of the recent developments in machine learning and optimization to learn the dependencies between covariates and survival times in different ways.

The traditional statistical methods can be further subdivided into three categories: (i) parametric methods (ii) non parametric methods, and (iii). semi-parametric methods whereas Machine learning algorithms, such as survival trees, Bayesian methods, neural networks , support vector machines, and advanced machine learning methods like ensemble learning, active learning, transfer learning and multi-task learningmethods, are used for survival analysis.

Parametric methods are more efficient and accurate for estimation when the time to event of interest follows a particular distribution specified in terms of certain parameters. It is relatively easy to estimate the times to the event of interest with parametric models. Linear regression method is one of the main parametric survival methods. The parametric survival models tend to obtain the survival estimates that are consistent with a theoretical survival distribution. The commonly used distributions in parametric censored regression models are: normal, exponential, logistic, log-logistic and log-normal. If the survival times of all instances in the data follow these distributions, the model is referred as linear regression model.

Non-parametric methods are more efficient when there is no underlying distribution for the event time or the proportional hazard assumption does not hold. In nonparametric methods, an empirical estimate of the survival function is obtained using Kaplan-Meier (KM) method, Nelson-Aalen estimator (NA) or Life-Table (LT) method More generally, any KM estimator for the survival probability at the specified survival time is a product of the same estimate up to the previous time and the observed survival rate for that given time.

In semi-parametric methods , Cox model is the most commonly used regression analysis approach for survival data and it differs significantly from other methods since it is built on the proportional hazards assumption and employs partial likelihood for parameter estimation. Cox regression method is described as semi-parametric method since the distribution of the outcome remains unknown even if it is based on a parametric regression model. In addition, several useful variants of the basic Cox model, such as penalized Cox models, Cox Boost algorithm and Time-Dependent Cox model (TD-Cox), are also proposed in the literature.

Machine learning techniques, have an ability to model the non-linear relationships and the due to quality of their overall predictions made, they have achieved significant success in various fields. In survival analysis, the main challenge of machine learning methods is the difficulty to appropriately deal with censored information and the time estimation of the model. Machine learning is effective when there are a large number of instances in a reasonable dimensional feature space.

Survival trees are one form of classification and regression trees which are tailored to handle censored data. The basic intuition behind the tree models is to recursively partition the data based on a particular splitting criterion, and the objects that are similar to each other based on the event of interest will be placed in the same node.

Naive Bayes, a well-known probabilistic method in machine learning, is one of the simplest but effective prediction algorithms. One drawback of Naive Bayes method is that it makes the independence assumption between all the features, which may not be true for many problems in survival analysis.

A Bayesian network, in which the features can be related to each other at various levels, can graphically represent a theoretical distribution over a set of variables. Bayesian networks can visually represent all the relationships between the variables which make it interpretable for the end user. It can acquire knowledge information by using procedures of estimating the network structures and parameters from a given dataset. In Artificial neural network, the simple artificial nodes denoted by “neurons” are connected based on a weighted link to form a network which simulates a biological neural network. A neuron in this context is a computing element which consists of sets of adaptive weights and generates the output based on a certain kind of activation function. Artificial neural network (ANN) has been widely used in survival analysis.

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. The DNN finds the correct mathematical manipulation to turn the input into the output, whether it be a linear relationship or a non-linear relationship. The network moves through the layers calculating the probability of each output. DNNs can model complex non-linear relationships.

Various researchers have used different methods of survival analysis for different applications. Some of them are discussed below:

Katerina Langova in her paper titled ” Survival analysis for clinical studies [15] have discussed about non parametric models like log rank test and generalized Wilcoxon test (also called Breslow test), The Cox proportional hazards model for survival analysis . here survivor function was estimated using the Kaplan-Meier method (K-M method). The Kaplan-Meier method provides very good estimations of survival probabilities also in such cases when the survival distribution is unknown. This method assumes that the pattern of censoring is independent of the survival times.

Yifei Chen,Zhenyu Jia ,Dan Mercola and Xiaohui Xie have used gradient boosting algorithm for survival analysis [12] They have proposed a nonparametric model for survival analysis that doesnot explicitly assume particular forms of hazard functions. Nonparametric model utilizes an ensemble of regression trees todetermine how the hazard function varies according to the associated covariates. The ensemble model is trained using a gradient boosting method to optimize a smoothed approximation of the concordance index, which is one of the most widely used metricsin survival model performance evaluation. They have applied this algorithm on large-scale breast cancer prognosis dataset. They implemented model using a software package called GBMCI (gradient boostingmachine for concordance index) in R and benchmarked the performance of model against other popular survival models and observed that GBMCI consistently outperforms other methods based on a number of covariate settings.

Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger have used A Cox Proportional Hazards Deep Neural Network For Personalized Treatment Recommender System[6] Medical practitioners use survival models to explore and understand the relationships between patients' covariates (e.g. clinical and genetic features) and the effectiveness of various treatment options. Standard survival models like the linear Cox proportional hazards model require extensive feature engineering or prior medical knowledge to model treatment interaction at an individual level. While nonlinear survival methods, such as neural networks and survival forests, can inherently model these high-level interaction terms, they have yet to be shown as effective treatment recommender systems. They introduced DeepSurv, a Cox proportional hazards deep neural network and state-of-the-art survival method for modeling interactions between a patient's covariates and treatment effectiveness in order to provide personalized treatment recommendations. The predictive and modeling capabilities of DeepSurv assist medical researchers to use deep neural networks as a tool in their exploration, understanding, and prediction of the effects of a patient's characteristics on their risk of failure.

Yan Li,Jie Wang,Jieping Ye,Chandan K. Reddy have proposed A Multi-Task Learning Formulation for Survival Analysis[10]. The popularly used survival analysis methods such as Cox proportional hazard model and parametric survival regression suffer from some strict assumptions and hypotheses that are not realistic in most of the real-world applications. To overcome the weaknesses of these two types of methods, they reformulated the survival analysis problem as a multi-task learning problem and propose a new multi-task learning based formulation to predict the survival time by estimating the survival status at each time interval during the study duration.

They proposed an indicator matrix to enable the multitask learning algorithm to handle censored instances and incorporate some of the important characteristics of survival problems such as non-negative non-increasing list structure into model through max-heap projection. They have proposed an optimization method which employs Alternating Direction Method of Multipliers (ADMM) algorithm to solve the proposed multi-task learning problem.

Changhee Lee,William R. Zame, Jinsung Yoon, Mihaela vander Schaar has proposed a DeepHit : A Deep Learning Approach to Survival Analysis with Competing Risks[2] DeepHit, that uses a deep neural network to learn the distribution of survival times directly. It makes no assumptions about the underlying stochastic process and allows for the possibility that the relationship between covariates and risk(s) changes over time. Most importantly, Deep Hit smoothly handles competing risks; i.e. settings in which there is more than one possible event of interest.

Kan Ren, Jiarui Qin, Lei Zheng, Zhengyu Yang,Weinan Zhang, Lin Qiu, Yong Yu in their paper titled "Deep Recurrent Survival Analysis" [1] proposed a Deep Recurrent Survival Analysis model which combines deep learning for conditional probability prediction at finegrained level of the data, and survival analysis for tackling the censorship. By capturing the time dependency through modeling the conditional probability of the event for each sample ,their method predicts the likelihood of the true event occurrence and estimates the survival rate over time, i.e., the probability of the non-occurrence of the event, for the censored data. Meanwhile, without assuming any specific form of the event probability distribution, model shows great advantages over the previous works on fitting various sophisticated data distributions.

III. Conclusion

The main objective of survival analysis is to predict the occurrence of specific events of interest at future time points. Due to the availability of data acquisition and storage methods, survival data collected from various real-world domains can be accumulated and then survival analysis can be applied on it with the recent developments in various machine learning methods, So there is an increasing demand for improving machine learning methods for effectively handling survival data. This paper provided a survey of the traditional survival analysis methods and various machine learning methods for survival analysis. All these methods can be integrated and then based on performance evaluation metrics , optimized result can be produced.

References

[1]. Kan Ren, Jiarui Qin, Lei Zheng, Zhengyu Yang,Weinan Zhang, Lin Qiu, Yong Yu "Deep Recurrent Survival Analysis" Copyright c 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org).

[2]. Changhee Lee, William R. Zame, Jinsung Yoon, Mihaela van der Schaar ,DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks", Copyright 2018, Association for the Advancement of Artificial Intelligence www.aaai.org

[3]. Stephane Fotso," Deep Neural Networks for Survival Analysis Based on a Multi-Task Framework", arXiv:1801.05512v1 [stat.ML] 17 Jan 2018

[4]. Alaa, A. M., and van der Schaar, M. 2017." Deep multi-task gaussian processes for survival analysis with competing risks" In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2017).

[5]. Ping Wang, Yan Li, Chandan K. Reddy "Machine Learning for Survival Analysis: A Survey, ACM Computing Surveys, Vol. 1, No. 1, Article 1, Publication date: March 2017.

[6]. Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger " DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network" arXiv:1606.00931v3 [stat.ML] 9 Aug 2017

[7]. Y. Li, V. Rakesh, and C. K. Reddy." Project success prediction in crowdfunding environments. " In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, WSDM '16, pages 247{256, 2016.

[8]. Y. Li, B. Vinzamuri, and C. K. Reddy "Regularized weighted linear regression for high-dimensional censored data". In Proceedings of SIAM International Conference on Data Mining, pages 45{53, 2016.

[9]. Y. Li, K. S. Xu, and C. K. Reddy." Regularized parametric regression for high-dimensional survival analysis". In Proceedings of SIAM International Conference on Data Mining, pages 765{773, 2016.

[10]. Yan Li,Jie Wang ,Jieping Ye ,Chandan K. Reddy,"A Multi-Task Learning Formulation for Survival Analysis**"**, KDD '16, August 13-17, 2016, San Francisco, CA, USA c 2016 ACM. ISBN 978-1-4503-4232-2/16/08. . . $15.00 DOI: http://dx.doi.org/10.1145/2939672.2939857

[11]. Sattar Ameri, Mahtab J Fard, Ratna B Chinnam, and Chandan K Reddy. 2016. "Survival analysis based framework for early prediction of student dropouts." In Proceedings of ACM International Conference on Conference on Information and Knowledge Management. ACM, 903–912.

[12]. Yifei Chen, Zhenyu Jia, DanMercolaand Xiaohui Xie"Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index", Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2013, Article ID 873595, 8 pages http://dx.doi.org/10.1155/2013

[13]. Imad Bou-Hamad, Denis Larocque, Hatem Ben-Ameur, and others. 2011. "A review of survival trees". Statistics Surveys 5 (2011), 44–71.

[14]. Paul D Allison. 2010. Survival analysis using SAS: a practical guide. Sas Institute.

[15]. Katerina Langova" Survival Analysis for Clinical Studies"*Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2008, 152(2):303–307.* 303,*November 25, 2008*

[16]. [16]Harald Binder and Martin Schumacher. 2008. "Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models". BMC bioinformatics 9, 1 (2008), 1–10.

[17]. [17] Bart Baesens, Tony Van Gestel, Maria Stepanova, Dirk Van den Poel, and Jan Vanthienen. 2005. Neural network survival analysis for personal loan data. Journal of the Operational Research Society 56, 9 (2005), 1089–1098.