A Survey of Survival Analysis Techniques and Statistical Methods

Achom Surjeet Singh,
Research Scholar, Department of Mathematics, School of Science, MUIT, Lucknow.
Dr. Parul Saxena,
Professor, Department of Mathematics, School of Science, MUIT, Lucknow.

Abstract--- The field of statistics known as "survival analysis" studies how long it will take for a certain event to occur based on historical data. Since a suitable data structure is available, it may be used to almost any issue area, not only biological ones. This method of risk classification has recently been proved to be effective in recent research. Many research has attempted to improve performance by combining survival analysis with machine learning algorithms, which is a well-established statistical approach. Additionally, in the area of machine learning, the use of diverse data modalities has been proved to improve prediction models' accuracy. Many scientific findings in survival analysis have been based on the modelling of survival data and constructing solid prediction models for the assessment of the time to event. For example, clinical decision making relies heavily on the use of risk-profile clustering that has been largely overlooked in the world of machine learning. Survivability-based clustering is a great way to find previously unknown subpopulations in the entire data set. The goal of these methods is to find clusters with dramatically different lifespan distributions, which can't be accomplished using ordinary clustering algorithms. Research works in the aforementioned field are presented in this survey with a focus on strategies for grouping censored data and detecting distinct risk levels. In many clinical studies, patients are monitored for an extended period of time. In these trials, mortality, recurrence, adverse medication response, or the development of a new illness are the key outcomes of interest. The research might have a follow-up period of a few weeks or several years. The data is analyzed using a separate set of statistical processes, which increases the amount of time it takes to complete a study. Useful in clinical research, it gives crucial data on an intervention. The various survival analysis techniques are explained in this article.

Keywords---Survival Analysis, Clustering, Machine Learning, Risk Stratification, Hazard Ratio, Survival Analysis.

I. Introduction

When Graunt and Healey published the first lifetable and weekly mortality bill in London in the seventeenth century, they laid the foundations for survival analysis. Since then, actuaries, statisticians, and biomedical researchers have commonly employed the lifetable technique. The dependability of military equipment became a major concern during World War II. As a result, researchers began looking into industrial equipment's "lifetime" and how long it can withstand use. Reliability analysis techniques established during World War II and used to the study of cancer patients' survival times were further expanded and refined. In the field of cancer research, the word "lifetime analysis" was renamed "survival analysis" by industrial reliability engineers. Among the most widely used approaches for evaluating data in fields as diverse as medicine, epidemiology, and environmental health, as well as criminal justice, marketing, and astronomy during the last four decades is survival analysis.

The fundamental variable in survival analysis is survival time, and it is no longer restricted to the mean time to death that was previously used. It's common to refer to the length of time between a certain start time and an event as "survival time." Examples of survival time include the length of a person's first marriage, the time it takes for diabetic retinopathy to develop after a diagnosis of diabetes, the length of a prisoner's sentence, and the lifespan of electronic equipment and computer components. New treatment modalities are put through their paces in clinical trials to see whether they are effective. Among the most common outcomes in clinical trials include mortality, adverse reactions, regression after remission, and the appearance of a new disease entity. Cox's proportional hazards regression model is often used in medical literature dealing with survival analyses. In these statistical models, we take into account the period until an event of interest happens and compare the cumulative probability of occurrences across time for two or more cohorts, while correcting for other relevant factors. An overview of survival analysis is provided here. The reader is introduced to commonly used words in survival analysis.

Almost all human undertakings are motivated by a desire for correctness. Practitioners, particularly in the healthcare area, may anticipate a high degree of accuracy in their forecasts due to the incredibly narrow margin of error. In healthcare, survival analysis is a critical decision-making tool. There are a variety of reasons why it can be used such as a deeper understanding of the effect of some genetic or proteomic bio markers on the prognosis of cancer patients to understand the impact of risk factors such as diabetes, hypertension, and other cardiovascular

diseases (CVD) on chronic kidney diseases (CKD) or even know the outcomes of physical exercises, diets, or family health history on understanding cardiac heart problems in patients. In general, it's used to estimate how long it will be before a certain event occurs. In spite of the fact that it was originally developed for medical research, it may be used in a variety of different contexts.

II. Literature Review

George Marinos & Dimosthenis Kyriazis (2021): The field of statistics known as "survival analysis" studies how long it will take for a certain event to occur based on historical data. Since a suitable data structure is available, it may be used to almost any issue area, not only biological ones. This method of risk classification has recently been proved to be effective in recent research. Many research has attempted to improve performance by combining survival analysis with machine learning algorithms, which is a well-established statistical approach. Additionally, in the area of machine learning, the use of diverse data modalities has been proved to improve prediction models' accuracy. Many scientific findings in survival analysis have been based on the modelling of survival data and constructing solid prediction models for the assessment of the time to event. For example, clinical decision making relies heavily on the use of risk-profile clustering that has been largely overlooked in the world of machine learning. Survivability-based clustering is a great way to find previously unknown subpopulations in the entire data set. The goal of these methods is to find clusters with dramatically different lifespan distributions, which can't be accomplished using ordinary clustering algorithms. Research works in the aforementioned field are presented in this survey with a focus on strategies for grouping censored data and detecting distinct risk levels.

Silvia Liverani et. al, (2021): For censored survival data with variables, we present a Dirichlet process mixing model. This concept may be used in two different situations. You may discover clusters based on both the censored survival data and the predictors by using this approach first. For a second, this strategy may be used with strongly correlated predictors, which are difficult to describe using traditional survival models because of multicollinearity. This study extends the Dirichlet process mixture model for mixtures of Weibull distributions, which can be used to estimate survival times and also allow for censoring, to include a response vector and covariate data via cluster membership. It is possible to have a global shape parameter for the Weibull distributions (referred to as a global parameter) and one particular to each cluster. There are two ways to estimate the survival curve. The first is more rigid and requires that the proportionate hazard assumption be met. The latter, however, is more flexible and may be used whether or not this assumption is met. We offer a simulated study and a practical application to sleep surveys in older women from the Australian Longitudinal Study on Women's Health to show the method's applicability in reality. PrMiuM is a R package that contains the approach described in the study.

Eleonora Giunchiglia et. al, (2018): Clinical guidelines for the "average" patient guide current medical practice. Because of advances in deep learning, healthcare may now be tailored to the unique needs of each individual patient. A novel recurrent neural network model termed guarantee for individualized survival analysis is presented in this study. Using censored data, our model is able to calculate both the patient's risk score and their survival function. The network constructs an embedding and outputs the value of the survival function for each time step, taking as input the patient's characteristics and the time step identifier. The unique risk score is calculated by linearly combining the values of the survival function. Our model has a higher concordance index (C-index) than existing techniques because of the structure of the model and the training that takes use of two loss functions.

Jared L. Katzman et. al, (2018): Patients' variables (e.g., clinical and genetic traits) and treatment outcomes may be explored and understood using survival models by medical practitioners. Cox proportional hazards models need considerable feature engineering or previous medical knowledge to represent therapy interactions on an individual level. Despite the fact that nonlinear survival techniques, such as neural networks and survival forests, may intrinsically describe these high-level interaction terms, they have yet to be shown as viable treatment recommendation systems A state-of-the-art survival technique for modelling interactions between patient variables and treatment success, we offer DeepSurv, a Cox proportional hazards deep neural network. DeepSurv is trained on both simulated and actual survival data in a series of studies. It has been shown that DeepSurv is able to accurately predict a patient's risk of death as good as or better than current state-of-the-art survival models. As an example, we'll demonstrate how DeepSurv models the association between a patient's characteristics and the success of various treatment alternatives. Last but not least, we put Deep Survey through its paces on real-world clinical trials to show how its individualized treatment suggestions might boost a patient population's overall survival time. Patients' risk of failure may be explored, understood, and predicted with the use of DeepSurv's predictive and modelling skills for medical researchers. DeepSurv is a powerful instrument.

Eleonora Giunchiglia et. al, (2018): Clinical guidelines for the "typical" patient govern current medical practice. Because of advances in deep learning, healthcare may now be tailored to the unique needs of each individual patient. A novel recurrent neural network model termed guarantee for individualized survival analysis is presented in this study. Using censored data, our model is able to calculate both the patient's risk score and their

survival function. The network constructs an embedding and outputs the value of the survival function for each time step, taking as input the patient's characteristics and the time step identifier. The unique risk score is calculated by linearly combining the values of the survival function. Our model has a higher concordance index (C-index) than existing techniques because of the structure of the model and the training that takes use of two loss functions.

III. Survival Analysis Methods

Methods of Survival Analysis may be classified as either statistical or machine learning. There are three broad groups of statistical methods: parametric, non-parametric, and semi-parametric. Parametric approaches may be very fast and accurate for predicting the time until an event occurs if the data set is assumed to follow a specific distribution. Using a well-known theoretical distribution, such as the exponential, for example, makes it straightforward to estimate the time until an event occurs when using the studied data set. In real-world datasets, it is very difficult to collect data that follows a theoretical distribution exactly as it does in theory. Because there is no underlying distribution for the event time, non-parametric approaches may be utilized in this scenario. To date, the most widely used approach in this area is the Kaplan-Meier (Kaplan and Meier, 1958). Using both parametric and non-parametric methods, we arrive at a third strategy. Semi-parametric models, like non-parametric approaches, do not need an understanding of the distribution of time to event. Because it posits that the qualities have a multiplicative influence on the risks function over time, the Cox model (Cox, 1972) is a frequently used semi-parametric survival analysis technique in this area.

Machine Learning Techniques for Time to Event Predictions

Although statistical approaches try to describe the distribution of event timings and the statistical features of each (statistical) model's parameters, machine learning methods strive to make predictions about event phenomena at a particular moment in time.Adapted for survival analysis, a decision tree technique based on recurrent data splitting using a specific splitting criteria. Some study has been done on the development of a splitting criteria that can be efficiently employed for survival analysis. As a cornerstone of statistical thinking, Bayesian analysis ties together prior and post-hoc probabilities. According to a few studies, Bayesian models may be used to forecast the likelihood of an outcome of interest while taking use of some of its advantages, such as interpretability. Additionally, support vector machines are an important class of machine learning techniques that may be used for classification and regression and have been effectively applied to survival analysis applications.

A neural network-based machine learning approach based on the semi-parametric Cox Proportional Hazard model has also been developed to estimate a subject's lifespan. Deep learning approaches for life-time prediction, such as a recurrent neural network design, have also been developed. An embedding is created for each time step, and a survival function value is outputted for that time step as a result of the proposed neural network's input-output mechanism. This picture illustrates the taxonomy of all the survival approaches that have been presented in the literature, not only for addressing the job of time to event prediction, but also for strategies that have been offered for grouping based on risk profile.

IV. Survival Analysis Problem Formulation

A triplet represents an observation I in our dataset (Xi, yi, zi), where $Xi \in R^{1 \times P}zi$ is the binary event indicator, which is indicated as 0 if the subject has not experienced the event of interest, and is the feature vector. It's also important to note that Ti is the survival time for an uncensored observer, while Ci is censored observer's survival time.

$$y_i = \begin{cases} T_i, & \text{if } z_i = 1 \\ C_i, & \text{if } z_i = 0 \end{cases}$$

A new feature vector Xk is used in survival analysis to estimate the time until an event of interest for a new instance k.

Concept of Censorship/ CensoredData

For a variety of reasons, obtaining entire data sets in real-world circumstances may be difficult. The data collecting phase, especially in clinical research, might run for many years, and participants in the study must remain consistent in order to maintain track of their data. For rare cases, the event of interest may not be visible. Due to missing data due to censorship, a typical regression model cannot be used to make predictions when dealing with this kind of data. We can see several different kinds of censorship, but the most common ones are right-censoring, which occurs when the observed survival time is less than or equal to the true survival time, left-censoring, which occurs when the observed survival time is greater than or equal to the true survival time, and third-interval censoring, in which we only know that the event occurs within a specific time interval. The time to the event of interest is the goal variable in survival analysis when using censored data. Only those who were there at the time of the incident are aware of this. Survival analysis may be used to a group of people shown in Figure 1. Four people

make up the cohort, three of whom fall within the purview of censorship and one of whom does not. Because we've tracked the subject's progress from the beginning to the end, we can't suppress it. Subjects 2 and 3 are deemed right censored since we tracked them from the beginning of our investigation, but they had not experienced the event of interest until the conclusion of the study, thus we do not have accurate information about the length of their lifespan. Due to the fact that it is uncertain when the subject 4 began the research, and since data collectors lost signals in the midst of the trial, the subject 5 is assumed to be interval censored.
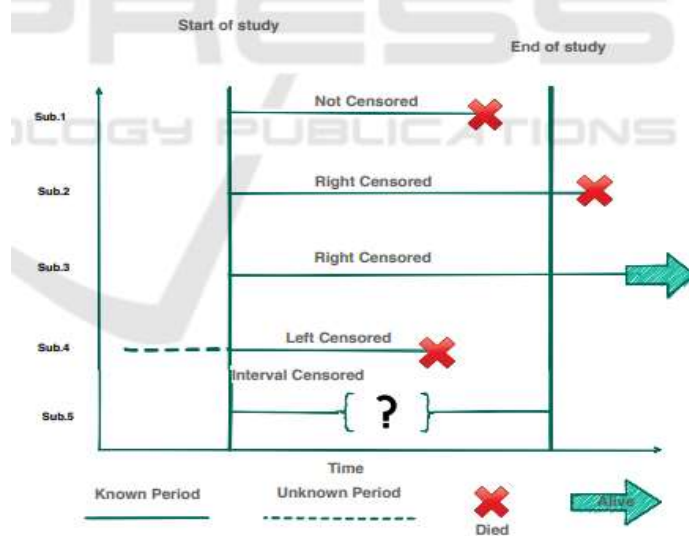

Figure 1: Concept of Censorship

Survival Analysis Formulas and Definitions

It is the chance that the time to the event of interest is not sooner than t. The survivor function or survivorship function in biological survival issues and the dependability function in mechanical survival problems are common names for the survival function. Survival function may be summarized in the following fashion.

$S(t) = P(T > t)$

A person who outlives t is represented by the above function. As t rises, so does the ability to survive. Its initial value is 1 for t = 0, which indicates that all subjects are alive at the beginning of the experiment. According to the formulation of the cumulative death distribution function, we may conclude F(t),

$S(t) = 1 - F(t)$

The likelihood that the event of interest happens before time t is represented by the cumulative death function. It follows that the probability density function associated with the survival rate is continuous.

$$S(t) = P(T > t) = \int_t P(x)dx$$

The cumulative death function depicts the probability that the relevant event will occur before time t. If the survival rate is continuous, then the probability density function associated with it must be too P(t) by

$$S(t) = P(T > t) = \sum_{T > t} P(x)$$

Survival Hazard Function

Also known as the force of mortality, instantaneous death rate, or conditional failure rate, the hazard function h(t) is often employed in survival analysis. As long as no previous event happened, the hazard function t (Lee and Wang, 2003) represents the rate at which an event will occur at time t. A hazard is a metric for danger in this context. In mathematical terms, the hazard function is:

$$h(t) = \frac{f(t)}{S(t)}$$

The probability density function to survival function ratio, to be precise. Since the probability density function is, by definition:

$$f(t) = \lim_{dt \to 0} \frac{F(t+dt) - F(t)}{dt}$$

We may express the danger function using the following formula:

$$h(t) = \lim_{dt \to 0} \frac{Pr(t \leq T < t + dt | T \geq t)}{dt}$$

Assuming the person has made it to the beginning of this short period of time, it is characterised as a failure. The cumulative distribution function F(t) and the probability density function f(t) may also be used to create the hazard function as follows:

$$h(t) = \frac{f(t)}{1 - F(T)} = \frac{f(t)}{S(t)}$$

This is followed by an analysis of the cumulative risk, which shows how much risk has been accrued over time.

$$H(t) = \int_0^t h(x)dx$$

h(t) (or H(t)) and S(t) are also linked by another crucial relation besides the ones above.

$$S(t) = exp\left(-\int_0^t h(x)dx\right) = exp(-H(t))$$

V. Clustering based on Risk Profile

Parallel algorithms for building the latent components were provided by Li and Gui in their application of partial least squares regression (PLS) regression to censored survival data in Cox model (Li and Gui, 2004). Iterated least square fitting of residuals and Cox regression fitting are used to create predictive components in the proposed technique. The Cox model may then be used to develop a viable survival prediction model using these components. It is possible to utilize such a strategy for survival clustering despite the fact that the primary components have been generated. Using gene expression data and clinical data together, researchers discovered the value of discovering cancer subtypes. Subtypes that were shown to vary significantly in terms of patient survival were found to do so when the semi-supervised approach that was presented was applied. The authors tackled the challenge of identifying cancer subtypes without knowing whether or not there were any in the dataset. There are two aspects to this strategy. To begin with, just gene expression data is used to give each of them a "Cox" score and then pick only those genes with a high score. Only genes that were found to be statistically significant were then included in the dataset. The desired number of clusters may be obtained by using classic clustering approaches such as K means solely on data relating to gene expression. Second, in the suggested technique, the cluster assignment is tested using just clinical data. They employ classification methods to clinical data and use cluster assignment as the dependent variable.

To sum it all up, the classification method worked well, indicating that the cluster allocations were right. Using principal components for regression and survival models has the disadvantage of relying on a small number of principal components to summarise a large proportion of the variance in the data, with no guarantee that these principal components are linked to the desired outcome as noted by Bair (Bair et al., 2006) and Tibshirani (Bair and Tibshirani, 2004). To this end, they developed supervised principal components, a semi-supervised technique (SPC). Cox scores are calculated for each feature in this procedure, and the most important features are selected based on the best Cox ratings. It is similar to traditional principal component analysis in that it employs a subset of predictors chosen for their relationship to the result. The "pre-weighted sparse clustering" approach has also been improved upon. Sparse clustering and semi-supervised clustering, as previously discussed, both suffer from major drawbacks due to their reliance on the number of characteristics that are deemed "important." It is an attempt to overcome the restrictions of sparse clustering by implementing traditional sparse clustering in pre-weighted mode. Features with different mean values across the clusters are identified. Once this pre-weighted version of sparse clustering has been performed a second time, all characteristics that vary from each other in the first set of clusters are assigned a weight of 0 in this second run of sparse clustering.

As a result of this method, we may see secondary clusters that would otherwise be hidden by the bigger dissimilarity measures of the primary clusters. Another method that is presented in this paper is to use non-zero weights for the characteristics that are significantly connected with the outcome variable to determine how much initial weight to give each feature. Researchers have looked for "secondary" clusters that might be "covered" by larger main clusters with high variance properties. The data matrix X is subjected to classic hierarchical clustering as part of this procedure. With each branch of the hierarchical clustering tree being cut at a certain height, a new data matrix X is created, and this new data matrix X is specified to represent the expected value of residuals when regressed on group labels for each X row. All conceivable cuts are taken into consideration when determining the predicted value.

This removes high-variance traits that may obscure secondary clusters, making it easier to identify them. Secondary clusters are produced by performing traditional hierarchical clustering on this updated matrix, X. This combined strategy, based on statistical and machine learning methodologies, was developed in accordance with Bair's work. Based on clinical and gene expression data, the suggested technique aimed to identify clusters. A semi-supervised method for discovering survival clusters has been presented by the authors. The suppressed lives were first estimated using solely the clinical data. Because they wanted to choose just the most relevant clinical characteristics, they used penalised logistic regression and penalised proportional hazard model using Expectation minimization technique. The survival time estimate for patients with censored survival time was performed using the K neighbors-based technique (with 10 neighbours) after a list of relevant clinical factors had been discovered. Following the use of the silhouette approach, they discovered the ideal number of clusters for the filtered data set. Gene expression data is then filtered using a fast correlation-based technique in order to identify the most important properties. As features are deleted from the list, those with lesser relevance are either dropped below a predetermined low limit or the list has run out of items to remove. Lastly, a classifier is employed to predict the label that was discovered from the clinical data set by analyzing the chosen genes. Clustering may be judged on the performance of the classifier, as well as whether or not groupings are found to have similar survival distributions.

Using a decision tree technique, Mouli sought to cluster survivors. In the end, the goal of this study was to identify two or more distinct groups with differing levels of risk. The paper's goal was to identify differences in survival distributions amongst clusters. First, the data set is divided into two groups, and then, using an attribute-values test, the survival distributions of the two groups are examined. The Kaplan-Meier estimations are used for this. After that, Kuiper statistics are employed to measure the importance of the difference between survival distributions. For each attribute-value pair, the process is repeated until the node with the best results is selected and utilised as a node in the decision tree. Kuiper statistic p-values are used to determine the optimal outcome in this stage of the method, where the significance level may be set by the user. The authors present a multiple hypothesis issue that can be fixed using the Bonferroni correction, as described in the study, as a result of doing several statistical tests at each node. As a consequence of the proposed technique, clusters may be found at the leaf nodes. There may not be a substantial difference in the degree of dissimilarity between detected clusters even when individuals with comparable survival distributions are put closer in the tree diagram.

In other words, the goal is to find groups with varying survival rates. A whole graph should be used, according to writers. In this graph, the nodes are referred to as leaf nodes, while the edges are referred to as p-values. The edges (normalized p-values) connecting each node indicate whether or not there is a meaningful link between them. The resultant graph is subjected to the Markov clustering process. Deep Life, an inductive neural network-based clustering model architecture, was presented in a recent research paper. The goal of this approach is to observe the empirical distribution of cluster lifetimes. This framework's ultimate goal is to perform studies with people that have various lifetime distributions, but who have the same lifetime distribution inside the same cluster. Proportional risks are not assumed in the suggested model. There are a number of advantages to this research study, including its potential to be used to data sets when termination signals have not been captured.

Kuiper two-sample test is the basis for this study's key contribution: a new clustering loss function. Since the test's infinite sum is not going to be included, the authors propose an upper limit on the Kuiper p-value that isn't dependent on computationally costly gradients. However, the suggested model isn't limited to a feedforward architecture, therefore this strategy is versatile and worth experimenting with alternative models to see how they vary in performance, as described in this article by the authors. An attempt is made to find clusters with the greatest disparity between their empirical lifespan distributions. Inspired by, Liverani developed a Dirichlet process mixing model for censored survival data with variables. It is recommended that the data set variables, which do not parametrically allocate data to clusters, be combined with Weibull distributions in the model.

There are no assumptions made about how these variables relate to one another in this technique, therefore it is possible to investigate the intricate relationships that exist between them. Non-linear correlations may be found between variables and response even if they are separately modelled. When compared to the Deep Life project, Chapfuwa's research is focused on characterizing the predictive distributions of time-to-event predictions using a clustered latent space conditioned on covariates. A complicated model developed by the authors of Survival Cluster Analysis offers not just risk profile-based clustering, but also functions as a deterministic encoder that translates variables into a latent representation, which in turn feeds a stochastic survival predictor. The clustering procedure in this study was carried out using a Bayesian non parametric technique. Distribution matching (in this work) followed a Dirichlet Process, whereas the Bayesian technique allowed for the latent representation to operate like a mixture of distributions. List all the algorithms created in the literature to solve the challenge of identifying people in a dataset according to their risk profile, taking into consideration the filtering that is often encountered in real-world datasets.

VI. Conclusion

Since a suitable data format is readily accessible, survival analysis may be used broadly across a wide range of fields. Scientific publications using machine learning approaches for time-to-event forecasts have been published several times, but the number of studies devoted to survival clustering algorithms has been much lower. Survivability-based clustering is very beneficial in identifying unknown subpopulations in a large dataset. To identify clusters whose lifespan distributions are considerably different, such strategies seek to use classic clustering algorithms, which are not capable of doing this task.

References

[1]   Marinos, George & Kyriazis, Dimosthenis. (2021). A Survey of Survival Analysis Techniques. 716-723. 10.5220/0010382307160723.

[2]   Liverani, Silvia & Leigh, Lucy & Hudson, Irene & Byles, Julie. (2021). Clustering method for censored and collinear survival data. Computational Statistics, 36, 10.1007/s00180-020-01000-3.

[3]   Giunchiglia, Eleonora &Nemchenko, Anton & Schaar, Mihaela. (2018). RNN-SURV: a Deep Recurrent Model for Survival Analysis.

[4]   Katzman, Jared & Shaham, Uri & Cloninger, Alexander & Bates, Jonathan & Jiang, Tingting & Kluger, Yuval. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology, 18. 10.1186/s12874-018-0482-1.

[5]   Giunchiglia, Eleonora &Nemchenko, Anton & Schaar, Mihaela. (2018). RNN-SURV: a Deep Recurrent Model for Survival Analysis.

[6]   Gaynor, S. and Bair, E. (2017). Identification of relevant subtypes via reweighted sparse clustering. Computational statistics & data analysis, 116:139–154.

[7]   Giunchiglia, E., Nemchenko, A., and Van Der Schaar, M. (2018). Rnn-surv: A deep recurrent model for survival analysis. In International Conference on Artificial Neural Networks, pages 23–32. Springer.

[8]   Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2014). Bayesian survival analysis. Wiley Stats Ref: Statistics Reference Online.

[9]   Liverani, S., Leigh, L., Hudson, I. L., and Byles, J. E. (2020). Clustering method for censored and collinear survival data. Computational Statistics, pages 1–26.

[10]   Molitor, J., Papa Thomas, M., Jerrett, M., and Richardson, S. (2010). Bayesian profile regression with an application to the national survey of children's health. Biostatistics, 11(3):484–498.

[11]   Molitor, J., Su, J. G., Molitor, N.-T., Rubio, V. G., Richardson, S., Hastie, D., Morello-Frosch, R., and Jerrett, M. (2011). Identifying vulnerable populations through an examination of the association between multipollutant profiles and poverty. Environmental Science & Technology, 45(18):7754–7760.

[12]   Mouli, S.C., Naik, A., Ribeiro, B., and Neville, J. (2017). Identifying user survival types via clustering of censored social network data. arXiv preprint arXiv:1703.03401.

[13]   Mouli, S.C., Teixeira, L., Ribeiro, B., and Neville, J. (2019). Deep lifetime clustering. arXiv preprint arXiv:1910.00547.

[14]   Nowak, G. and Tibshirani, R. (2008). Complementary hierarchical clustering. Biostatistics, 9(3):467–483.

[15]   Prinja, S., Gupta, N., and Verma, R. (2010). Censoring in clinical trials: review of survival analysis techniques.

[16]   Rupert Jr, G. et al. (2012). Simultaneous statistical inference. Springer Science & Business Media.