Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2020;6(1):62-71

Review Article

Review of Random Survival Forest Method

Majid Rezaei[1]*, Leili Tapak[2], Masoomeh Alimohammadian[3,4,5], Alireza Sadjadi[4,5], Mehdi Yaseri[1]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.
[2]Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.
[3]Department of Human Ecology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.
[4]Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran.
[5]Digestive Oncology Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran.

ARTICLE INFO

Received 22.01.2020
Revised 29.02.2020
Accepted 01.03.2020
Published 10.03.2020

Key words:
Machine learning;
Ensemble methods;
Random survival forest;
Survival

ABSTRACT

Background: Over the past years, there has been a great deal of interest in applying statistical machine learning methods to survival analysis. Ensemble-based methods, especially random survival forest, have been developed in various fields, especially medical sciences, due to their high accuracy and non-parametric nature and applicability in high-dimensional data sets. This paper aims to provide a methodological review and how to use random survival forests in the analysis of right-censored survival data.

Method: We present a review article based on the latest research in the PubMed database on random survival forest model methodology.

Results: This article begins with an introduction to tree-based methods, ensemble algorithms, and random forest (RF) method, followed by random survival forest framework, bootstrapped data and out-of-bag (OOB) ensemble estimators, review of performance evaluation indicators, how to select important variables, and other advanced topics of random survival forests for time-to-event data.

Conclusion: When analyzing right-censored survival data with high-dimensional data, while the relationships between variables are complex and their interactions are taken into account, the nonparametric random survival forest (RSF) method determines important variables affecting survival times with high accuracy and speed and also does not need to test the restrictive assumptions.

Introduction

Although Cox models (1) are very famous for survival data analysis, but these models have restrictive assumptions. Also, these models become inadequate in high dimensional settings and model assumptions are often violated (1, 2). Moreover, identifying high-order interactions requires the presence of these effects in the model, which in cox models makes it more difficult to interpret the results (3). To alleviate these problems, nonparametric and flexible methods such as tree ensembles have been developed (4-7).

In this paper we describe random survival forests, a well-known ensemble tree method for right-censored time-to-event data analysis. Findings of application of this method show that making ensembles by the average base learners, like trees, can

* . Corresponding Author Email: majidrezaei2019@gmail.com

Please cite this article in press as Rezaei M, Tapak L, Alimohammadian M, Sadjadi A, Yaseri M. Review of Random Survival Forest Method. *2020; 6(1):62-71*

significantly improve the performance of prediction. In addition, ensemble learning can be improved further by adding randomization to the basic learning process. The Random Forest Survival (RSF) method is a generalized Breiman's Random Forest (RF) method for the analysis of survival data. In RF, randomization is done in two ways. First, for the growth of each tree, a random sampling with replacement (bootstrap method) is taken from the data set. Second, in each tree node, a number of explanatory variables are randomly selected.

The purpose of this two-stage randomization is to make trees independent of each other, which due to the bagging property, reduces the variance for the ensemble (8). The use of deep trees (to reduce bias), when combined with reduced variance due to bagging (randomization and averaging), enables the random forest method to fit valuable models with low generalizability error. Therefore, the implementation of RSF is similar to the RF method and follows its general principles, the steps of which are: (i) Survival trees grow using bootstrapped data, (ii) When splitting tree nodes, several explanatory variables are randomly selected from the set of variables, (iii) Trees usually grow deep until the cessation condition is met, and (iii) the survival forest ensemble is calculated by the average predicted survival measures of the trees.

After fitting a model, it is very important to check the performance of the model. In the case of RSF survival analysis, C-index and the Brier statistics are usually calculated and presented. C-index is an extension of the area under ROC curve (9). The large values of the C index indicate the suitability of the model in fitting the data.

Tree-Based Methods and Ensemble algorithms

Recently, Machine Learning (ML) techniques have been rapidly applied in various fields to automatically analyze large volumes of data. The main concept of ML is to learn the algorithm from repeated input data, and to recognize hidden patterns and relationships from huge, noisy and complex data. An important feature of the ML approach is the construction of a prediction model with the presence of nonlinear effects and complex interactions, among several variables. Therefore, ML is widely used in medicine to identify important risk factors and predict diseases. In addition, ML techniques have been adapted to statistical learning concepts and principles (10).

Tree-Based methods are from ML techniques. A tree is composed of numerous nodes. Tree estimation is in general based on recursively performing binary dividing on the variable space using some pre-defined splitting rule. The result is a collection of candidate's nodes, which continues from the top of the tree (root) to a number of terminal nodes (leafs) (11).

In tree-based estimation procedures, each tree is non-deterministic as the tree is grown on a subspace of individuals who were picked from bootstrapping the whole dataset (8). Growing a single tree is well known to exhibit high variances in predicted outcomes. By combining the trees, however, variance as well as bias in prediction can be substantially decreased (12).

Ensemble methods are based on collective votes or results average, that's mean in the classification mode, a new classifier by majority votes of classifiers, and in the predictive mode, it is based on the average of results (13). Ensemble algorithms in each tree use a random set of explanatory variables in each node to make the trees independent of each other and then aggregate the results. This is because averaging the results of several trees can improve the prediction for data outside the training samples (experimental sets) by reducing overfitting. The most important ensemble techniques are Bagging, Boosting and RF. In Bagging, trees grow with bootstrap samples are based on a large number of trees, each of which is a random sample with replacement and the same size of the original data set (8). Boosting trees are based on the idea of improving the fit, the algorithm is a linear combination of trees. In each step, the data that was not well fitted in the previous steps are re-fitted in the next steps. In this method we get strong from weak predictive trees.

Random Forest

Random forest is one of the most popular supervised ML methods and is one of the best classification and prediction algorithms ever designed (14). Random forests are known in the ML literature for their reliable performance that does not require excessive model tuning (15).

Random forests are an example of "model averaging". The prediction obtained with random forests is constructed by averaging over hundreds or thousands of distinct trees that differ from one another for several reasons. The name of the algorithm derives, in fact, from the characteristic of a random forest of being a multitude of trees that differ because of random selection of both the data and the variables. Random forests combine bagging with random selection of variables, an idea introduced first by (16).

Random forests overcome several problems with single decision trees. They reduce problems of overfitting by averaging several trees, also can handle automatically (i.e., without need or recoding, grouping, etc.) types of scales for explanatory variables and also missing data. They can capture non-linear effects and interaction terms. Another important feature of random forests is their ability to cope with a large number of explanatory variables, even if most of them are interrelated. In other words, collinearity is not a problem for random forests.

A disadvantage of random forests (like all ensemble methods) is that by averaging multiple trees, they do not provide a single tree for interpretation. However, several measures can be calculated to ease interpretation. For example, it is possible that each time a particular variable is used in a tree, a decrease in the fit indicators (eg Gini index) is calculated, and finally by averaging these values, the importance of variable in the prediction is obtained.

Random forests were initially used for cross-sectional data. However, recent methodological advances have also made it possible to use these methods for survival analysis (17-20).

Random Survival Forest

RSF is one of the ML methods that uses a set of decision trees and provides the most

effective explanatory variables related to survival time. In two stages of randomization, including random sampling with replacement (bootstrap) and random splitting of nodes, ensemble of decision trees in the RSF method is calculated (20).

Several independent bootstrap samples (usually at least 500 samples) are randomly taken from the dataset, and each bootstrap sample is used to calculate a separate decision tree (Figure 1). By using Bootstrap data, the problem of overfitting with other data is reduced (14). The next randomization stage is performed at the node splitting level. In each node of the decision tree, a certain (predetermined) number of variables are considered. To divide each node, from the considered variables, the variable that causes the most differences between the daughter nodes is selected.

In general, these two random processes create a balance between bias and variance and improve predictive performance relative to a single decision tree.

The ensemble is a cumulative hazard function (CHF) is based on averaging individual tree's Nelson-Aalen's CHF (21).

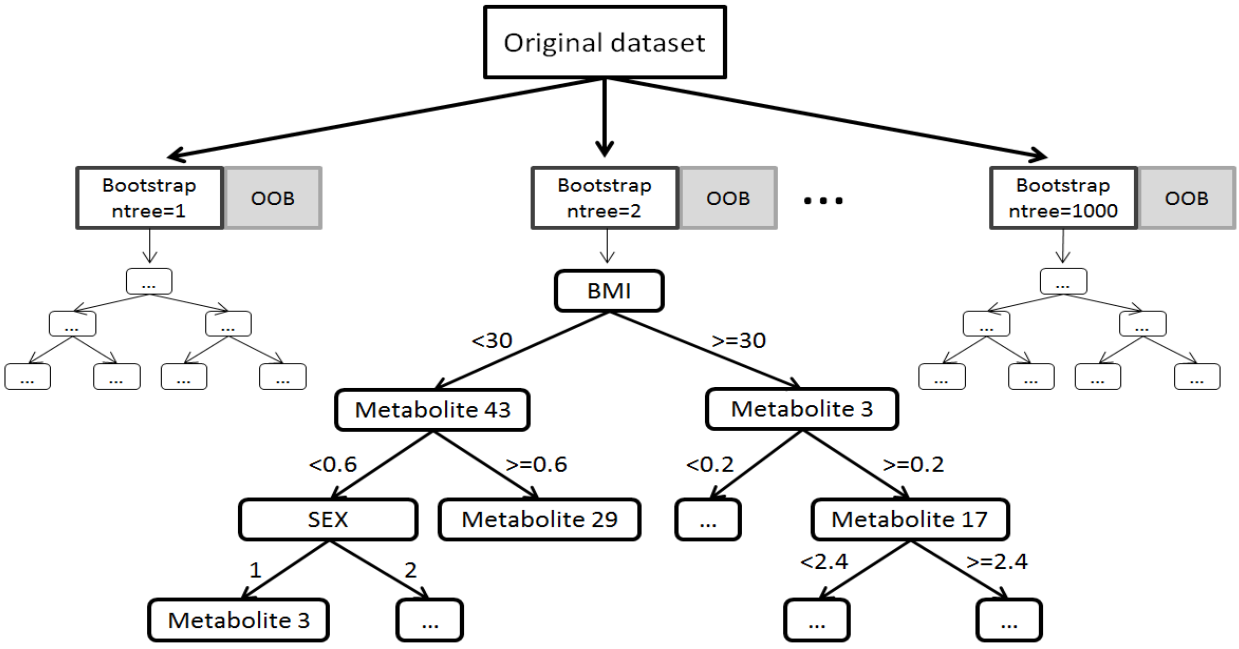

Figure 1: Illustration of the RSF method. Reference: (22)

The use of the RSF method is very usual and logical, because it is a completely data-driven method, has no restrictive assumption, and can work automatically with high-dimensional data and their interactions (20). For these reasons, in several studies, the application of RSF method in identifying disease risk factors has been successful (23-25).

Each Bootstrap sample contains an average of two-thirds of the data set. The remaining one third of the data is called out-of-bag data

(OOB). A single decision tree is created based on each bootstrap sample. To grow a decision tree, in each node a random set of explanatory variables are selected to split the random node. The number of variables selected is usually the square root of the total number of variables. The rule of splitting (Log-Rank statistics ...) is to maximize the difference in survival between nodes. The number of random split points for each selected variable can be predefined before calculating a RSF model when using the RandomForestSRC package in R statistical software.

Compared to regression methods, the RSF method has several advantages. RSF is completely data-driven and therefore independent of hypothesis testing (20). This method does not test the goodness of fit hypothesis, but runs a model that best describes the data. To use the RSF approach, no hypotheses such as examining the distribution of explanatory variables or proportional hazards need to be tested. Raw data can be used directly to run RSF models. The RSF process is largely automated and only a few key parameters such as the number of bootstrap samples or the number of node divisions need to be specified (20). Therefore, the RSF approach is also a suitable method for exploratory analysis of survival data in which prior knowledge is not yet complete.

For tree growth, RSF uses a random subset of variables at each node. As a result, to divide the nodes, the correlated variables are usually selected independently of each other, which leads to breaking the correlation structure of the variables. Therefore, there is less competition between highly correlated variables to participate in the model. Therefore, the selection of effective variables is possible even in the presence of multicollinearity (26). In addition, the overfitting problem is greatly reduced due to randomization through bootstrap sampling (27).

To evaluate the performance of the algorithm in predicting, calculate the "out of bag" error rate (OOB) and the concordance index (C-index) are calculated. The RSF does not require an independent data set for validation, because it is estimated during the implementation of the algorithm.

A disadvantage of the RSF method is that the relative risk and odds ratio cannot be calculated. Instead, the importance of each risk factor can be calculated through minimum depth and VIMP indicators. However, the variables selected by RSF can be analyzed in Cox regression models to estimate relative risks and odds ratios. Another disadvantage of tree-based methods is that in node splitting, they prefer to choose continuous variables (28), if the data set consists of a combination of continuous and qualitative variables, this can be done by selecting a little cut points numbers.

Random survival forests algorithm

The steps of the RSF algorithm are:

1- B bootstrap samples is extracted from data set. Each sample contains an average of two-third of the data, and removes the rest of the data (OOB sample).

2- For each bootstrap sample, a survival tree is grown. At each tree node, the p variable is randomly selected. Each node is split using one of the variables that maximizes the

survival difference between the daughter nodes.

3- Each tree grows under the constraint that a terminal node should have no less than d> 0 unique outcomes (events).

4- CHF is calculated for each tree and their average is used to obtain ensemble CHF.

5- Using OOB data, the CHF prediction error is calculated.

6- Estimates based on out-of-bag data (OOB) are used to calculate variable importance (VIMP).

Terminal node prediction

Finally, each survival tree reaches a final point due to the limit set for the number of outcomes in the terminal nodes, where no new daughter nodes are formed. The terminal nodes are denoted by £. Suppose a vector $(T_{1,h}, \delta_{1,h}), \ldots, (T_{n(h),h}, \delta_{n(h),h})$ indicates the survival times and δ (0=Censoring or 1=Occurrence of event) for persons in a terminal node h ∈ £. A person i is right-censored at time $T_{i,h}$ if $\delta_{i,h} = 0$; if not, $\delta_{i,h} = 1$, when the person have experienced outcome at $T_{i,h}$. Let $t_{1,h} < t_{2,h} < \ldots < t_{n(h),h}$ be the n(h) distinct outcome times. $d_{l,h}$ and $Y_{l,h}$ are the number of deaths and persons at risk at time $t_{l,h}$. The CHF estimate for h is the Nelson–Aalen estimator:

$$\widehat{H_h}(t) = \sum_{t_{l,h} \le t} \frac{d_{l,h}}{Y_{l,h}}$$

All persons within h have the same CHF.

Each person i has a d-dimensional covariate xi. Let H(t|xi) be the CHF for i. To compute this value, drop x_i down t_{he} tree. Due to the structure of the survival tree, x_i will fall into a unique terminal node h∈ £. The CHF for i is the Nelson–Aalen estimator for x_i's terminal node:

$$H(t|x_i) = \widehat{H_h}(t) \quad , \text{if } x_i \in h. \qquad (1)$$

Phrase (1) defines the CHF for all persons.

The bootstrap and OOB ensemble CHF

To calculate the ensemble CHF, the CHF values of the trees (B trees) are averaged. Here the CHF estimate in the out-of-bag (OOB) and bootstrap samples are computed. Recall that each tree in the forest is randomly generated by a bootstrap sample independent of the other trees.

$I_{i,b} = 1$ if i-th is an OOB case for b-th tree; if not, $I_{i,b} = 0$. $H_b^*(t|x)$ indicate the CHF (1) for a tree grown from the b-th bootstrap sample. The OOB ensemble CHF for i is

$$H_e^{**}(t \mid x_i) = \frac{\sum_{b=1}^{B} I_{i,b} H_b^*(t|xi)}{\sum_{b=1}^{B} I_{i,b}} \qquad (2)$$

See that (2) is an average over bootstrap data in which i is OOB.

Similarly, $H_e^{**}(t|x_i)$ can be computed as follows. Drop OOB samples down a survival tree grown from in-bag (bootstrap) data. Find i's terminal node and it's CHF. Compute the average of these CHFs. This gives (2).

Now, the bootstrap ensemble CHF for i is:

$$H_e^*(t|x_i) = \frac{1}{B}\sum_{b=1}^{B} H_b^*(t|xi) \quad (3)$$

See that (3) employs all survival trees.

C-index calculation

When an RSF model is fitted, model performance evaluation indicators such as C-index are required. How to calculate index C according to the following steps:

1. All possible pairs of cases are obtained.
2. Remove those pairs whose lower survival time is censored. Remove pairs i and j if $T_i = T_j$ unless at least one is a death (Occurrence of event). Permissible is the total number of permissible pairs.
3. For each permissible pair where $T_i \neq T_j$,count 1 if the lower survival time has worse predicted measure; count 0.5 if predicted measures are tied. For each permissible pair, where $T_i = T_j$ and both are deaths, count 1 if predicted measures are tied; if not, count 0.5. For each permissible pair where $T_i = T_j$,but not both are deaths, count 1 if the death has worse predicted measure; if not, count 0.5. Concordance is sum over all permissible pairs.
4. C-index = Concordance/Permissible.

Prediction error rate

To specify the prediction accuracy of a RSF model, the prediction error rate of this model can be calculated based on the C-index. The prediction error rate is equivalent to the 1-C-index, whose values are between 0 and 1. The lower prediction error rate is equivalent to RSF models with higher prediction accuracy (20, 29).

Variable selection and VIMP

Variable selection is used for complex, high-dimensional data to identify variables related to the outcome under study. In general, the variable selection process is associated with a reduction in the size of the data set, which is desirable and practical for the following reasons (30-32).

1) Excessive predictive variables cause problems in statistical analysis that lead to bias, misleading estimates and reduced prediction accuracy.
2) Overfitting, multiple test problems and multicollinearity can be reduced with simple models.
3) Statistical interpretation is improved by reducing the number of variables.
4) In the field of clinical diagnoses and disease prediction, using several important variables to prevent misinterpretation and also reduce costs.

Important variables are selected based on their importance index (VIMP). To calculate the VIMP for a variable x, drop the OOB data down on their survival tree. While there is a split for x, assign the daughter node randomly. The CHF is calculated and averaged from each such tree. VIMP for x is the prediction error for the main ensemble that is subtracted from the prediction error for the new ensemble (14,33).

High values of VIMP indicate variables with high predictive power, while values of zero or negative indicate unpredictable variables.

Extensions of Random Survival Forests

A new generalization of RSF for use in competing risks is explored by (34). Two news dividing rules for growing competing risk trees, namely log-rank splitting and the

modified Gray's splitting, were defined to examine the equality of the cause-specific hazard and the equality of the cumulative incidence function (CIF), respectively. They also introduced several new ensemble estimators for competing risks such as ensemble CIF and event-specific estimates of mortality.

Cox boosting

Boosting is one of the ensemble methods that reaches a strong learner based on the combination of several weak learners and is one of the best methods, especially in classification issues (35). Therefore, due to its useful and accurate applications and performance, this method has been expanded in the field of statistics, including regression methods and survival analysis (36). Unlike the bagging method, this method does not use independent learners and its learning process is sequentially.

The main task of boosting is to improve the predictors sequentially, which in each iteration includes the weak predictors of the previous stage, and seeks to minimize the predefined loss function.

In survival analysis, most boosting methods are based on the Cox model, by using gradient boosting, with a loss function gained from the Cox partial likelihood function, as used in the R-packages mboost and CoxBoost (37). Mboost is based on model-based boosting, whereas CoxBoost is a likelihood-based boosting.

Discussion

Due to its high flexibility, ability to variable selection, and its nonlinear and nonparametric nature, the random survival forests method has become an active research topic and a promising approach for high-dimensional survival data in many bio-medical applications. This paper provides a partial survey of methodological developments of random survival forests in the past years.

In this article we have explained RSF, a new generalization of RF method (14), to right-censored survival data. A RSF includes of random survival trees. Using separately bootstrapped data, each tree is grown by random subset of variables at each node. Therefore, correlated variables will be chosen apart from each other, then splitting the node using a splitting rule such as Log-Rank. The tree is grown until each terminal node has no less than $d > 0$ unique experienced outcomes (deaths). The estimated CHF for a person is the Nelson–Aalen estimator for the person's terminal node. The ensemble is the average of these CHFs. OOB ensemble can be computed by dropping OOB cases down their in-bag survival trees and averaging.

Evidence from applied medical research papers shows that the RSF method performs better or at least better than its competitors in analyzing survival data and makes precise ensemble predictors. RSF is very useful for discovering very complex relationships between variables. While traditional methods are based on several restricted assumptions, and they are less automatic and do not work well in multicollinearity situation.

RSF is an attractive method when the target is to do prediction. Its advantage is more apparent when relationship between outcome and covariates are complex or when the proportional hazard assumption is at risk

(11). In addition, due to the process of random node dividing, highly correlated variables can also be in the model, and the selection of valid variables is possible even if there is multicollinearity (26). In addition, the overfitting problem is greatly reduced due to randomization done through the bootstrap sampling method (27). As expected, the results of the review of numerous articles show the applicability and accuracy of the ensemble methods as RSF, especially in the medical sciences.

References

1. Cox DR. Regression models and life tables (with discussion). J R Stat Soc, 1972, 34:187–220.
2. Cox DR, Oakes D. Analysis of survival data. Vol 21 New York, Chapman & Hall/CRC, 1984.
3. Radespiel-Troger M, Rabenstein T, Schneider H.T, and Lausen B. Comparison of tree-based methods for prognostic stratification of survival data. Artificial Intelligence in Medicine, 2003, vol. 28, no. 3, 323–341.
4. Gordon L, Olshen RA. Tree-structured survival analysis. Cancer Treat Rep 1985, 69:1065–1069.
5. LeBlanc M, Crowley J. A review of tree-based prognostic models. Springer, 1995, 113–124.
6. Hothorn T, Lausen B, Benner A, Radespiel-Troger M. Bagging survival trees. Stat Med, 2004, 23:77–91.
7. Hothorn T, Buhlmann P, Dudoit S, Molinaro A, Laan MJ. Survival ensembles. Biostatistics, 2006, 7:355–373.
8. Breiman L. Bagging predictors. Machine Learning, 1996, 24 (2), 123–140.
9. Heagerty P. J, Lumley T, Pepe M. S. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics, 2000, vol. 56, no. 2, 337–344.
10. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, 2001.
11. Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. The Wadsworth Statistics/Probability Series, Belmont, CA, 1984.
12. Breiman L. Heuristics of instability and stabilization in model selection. Ann Stat, 1996b, 24(6), 2350-2383.
13. Berk R. A. An introduction to ensemble methods for data analysis. Sociological methods & research, 2006, 34(3), 263-295.
14. Breiman L. Random forests. Machine Learning, 2001, 45: 5-32.
15. Athey S, Imbens GW. The state of applied Econometrics – Causality and Policy Evaluation. ArXiv 2016.
16. Ho, T. K. Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995, 278–282.
17. Hothorn T, Lausen B. Double-bagging: Combining classifiers by bootstrap aggregation. Pattern Recognition, 2003, 36: 6, 3 3–1309.
18. Hothorn T, Lausen B, Benner A, Radespiel-Troeger M. Bagging Survival Trees. Statistics in medicine, 2004, 3: 77–91.
19. Hothorn T, Buhlmann P, Dudoit S, Olinaro A. Laan J. Survival Ensembles. Biostatistics, 2006, 7: 3, 355–373.
20. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. The

Annals of Applied Statistics, 2008, 2(3), 841–860.
21. Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. Stat Anal Data Min 2011, 4:115-132.
22. Dietrich S, Floegel A, Boeing H, Schulze MB, Illig T, Pischon T, et al. Random Survival Forest in practice – a method for modelling high-dimensional metabolomics data in time to event analysis. In revision at International Journal of Epidemiology, 2016.
23. Datema FR, Moya A, Krause P, Back T, Willmes L, Langeveld T, et al. Novel Head and Neck Cancer Survival Analysis Approach: Random Survival Forests Versus Cox Proportional Hazards Regression. Head and Neck-Journal for the Sciences and Specialties of the Head and Neck, 2012, 34(1):50-8.
24. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying Important Risk Factors for Survival in Patient With Systolic Heart Failure Using Random Survival Forests. Circulation-Cardiovascular Quality and Outcomes, 2011, 4(1):39-45.
25. Omurlu IK, Ture M, Tokatli F. The comparisons of random survival forests and Cox regression analysis with simulation and an application related to breast cancer. Expert Systems with Applications, 2009, 36(4).
26. Siroky DS. Navigating Random Forests and related advances in algorithmic modeling. Statist Surv, 2009, Vol3, pp 147-163.
27. Van der Schaaf A, Xu CJ, van Luijk P, Van't Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: guarding against overfitting and effects of data set size. Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology, 2012, 105(1):115-21.
28. Loh W Y. Split Selection Methods for Classification Trees. Statistica Sinica, 1997, Vol 7, 815-840.
29. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. Journal of the American Medical Association, 1982, 247(18):2543-6.
30. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003, 3:1157-82.
31. Gentle J. E, Härdle W. K, Mori Y. Handbook of Computational Statistics - Concepts and Methods. Springer, 2012.
32. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning-with Applications in R. Springer ,2013.
33. Ishwaran, H. Variable importance in binary regression trees and forests. Electron. J. Statist, 2007, 1, 519–537.
34. Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. Biostatistics, 2014, 15:757–773.
35. Freund Y. Boosting a weak learning algorithm by majority. Inf Comput, 1995, 121:256-285.
36. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat, 2000, 28:337-407.
37. De Bin R. Boosting in Cox Regression: A Comparison between the Likelihood-Based and the Model-Based Approaches with Focus on the R-Packages CoxBoost and mboost. Technical Report No. 180, Munich: Department of Statistics, University of Munich, 2015.