SUT Journal of Mathematics
Vol. 50, No. 2 (2014), 311–324

An approach for constructing survival tree based on combination of covariates

Asanao Shimokawa, Yohei Kawasaki and Etsuo Miyaoka

(Received May 28, 2014; Revised July 25, 2014)

Abstract. We propose a new approach for constructing a survival tree which is used to analysis of time to event data. The construction of survival tree is performed by the classification and regression tree (CART) algorithm, which constructs a tree model by recursively dividing the data. In the traditional approach, internal-node data are divided by a splitting rule which consider only one covariate for each node. In the proposed approach, two or more combinations of covariates are considered for dividing the internal-node data in the algorithm. We show that the proposed method is more suitable than the traditional approach in some situations through the comparative research by simulations. We also present the result of an actual analysis based on proposed approach.

AMS 2010 *Mathematics Subject Classification.* 62N01, 62P10.

Key words and phrases. Survival tree, CART, splitting method.

§1. Introduction

Prediction of time to event like death or recurrence based on patient information is one of an important subject in medical research. To achieve this purpose, a regression model is constructed based on learning samples. Then, we can predict time to event of a new patient based on covariates of the patient. However in survival analysis, the construction of the predict model cannot be considered as a simple regression problem owing to the presence of censored data. Censored data has no exact information about time to event but has only information about time to a point that is confirmed the event has not occurred. Therefore, we cannot treat this information as a response value directly.

In this situation, several modeling methods which are dealt with the problem are proposed in parametric, semi-parametric, and non-parametric frameworks. One of a most famous and widely used model is the Cox proportional

hazard model (Cox [4]). This semiparametric model has advantages in that it is easy to estimate the coefficients in model and easy to understand the covariate effects. However, this model has a disadvantage that is difficult to detect interaction terms when there are many covariates. As the other choice of the regression model in survival analysis, there is the survival tree which is constructed in non-parametric framework. This model has an advantage that the relationship between covariates and hazards can be easily determined, and moreover, it can detect interaction terms from the hierarchical structure of the model.

Tree-based methods were introduced by Morgan and Sonquist [13]. In Breiman et al. [2], the classification and regression tree (CART) algorithm, which is used extensively for constructing tree models, is confirmed. The CART is composed by splitting, pruning, and selection steps. In the splitting step, all data for learning are recursively dichotomized and a large tree is constructed. The splitting rule of a node in tree is determined as $Z_j \leq c$, where Z_j is a numerical covariate and c is a cutting point of it ($j = 1, \cdots, p$). The data which satisfy the criterion are assigned to the left child node, and the others are assigned to the right child node. To determine the splitting rule from several possibility of the choice of covariates and splitting points in each node, we need to determine an evaluation criterion for each split preliminarily. Criteria for this purpose have been proposed by various authors (Gordon [7], Davis and Anderson [5], Segal [16], Therneau et al. [18], Zhang [19], and Keles and Segal [9]). Comparative research of these criteria were performed by Radespiel-Tröger et al. [14] and Shimokawa et al. [17]. Since there is a possibility of the overfitting (or over learning), we can not use the large tree model obtained by splitting step as the predictive model directly. Therefore, an optimal tree size is searched and a prediction model is constructed through the pruning and selection steps.

As described in above, the splitting rule of a node in traditional CART algorithm is restricted by only one covariate, that is $Z_j \leq c$ or not. Although constructing the tree model under this restriction has several merits like short learning time and inhibition of over learning, there is some difficult problem to learn like exclusive OR (XOR) problem. That is, if true model is linearly-inseparable then the tree becomes prohibitively large and sometimes it becomes cause to lower prediction accuracy of the obtained model. As other example, consider the case that the true model has high probability of event in a small partial space in p-dimensional covariates space. Then, if we construct the tree model based on traditional CART algorithm $2 \times p$ splits are needed in the model. Since the data in each node are divided repeatedly, the number of sample which are used to evaluate a split in a node becomes small by splitting. Therefore, the estimation of an optimal splitting point becomes bad by splitting in generally.

To handle these situations, we propose a new approach that two or more combinations of covariates are considered for dividing the data in each node. That is, the splitting rule of a node is defined as $Z_j \leq c \cap Z_k > d \cap, \cdots$ or not $(j \neq k \neq \cdots)$. Although this approach has disadvantage that it requires a long learning time, there is possibility to construct a more suitable model in some situations like XOR problem because the size of model is constricted. We show that through the comparative research by simulations. As the simulation-based evaluation method, we use the integrated Brier score (Graf et al. [8]) , which is widely used for evaluation of the predict model in survival analysis. Moreover, we present the result of an actual analysis based on proposed approach in the end of this paper.

The remainder of this paper is organized as follows. In Section 2, we define the notations and constructing method of survival tree. Further, the method of proposed approach are described. The simulation methods and results are given in Section 3. The result of an actual analysis is shown in Section 4. Finally, the conclusion of this paper is given in Section 5.

§2. Methods

2.1. Notation and Survival model

Let Y and C be the true survival and censoring time, respectively. Then, we can observe the time $X = \min(Y, C)$. Let $\delta = I(Y \leq C)$ be the censoring indicator, which is 1 if the observation experiences an event and 0 if the observation is censored. Let $\boldsymbol{Z} = (Z_1, \cdots, Z_p)$ denote a p-dimentional covariate vector. An observed learning sample set is represented by $\mathscr{L} = \{(x_i, \delta_i, \boldsymbol{z}_i); i = 1, \cdots, N\}$.

We define the tree model as T, and a node in the tree is defined as t. If the node is exist in the bottom layer of the tree, we call it terminal node. The set of terminal nodes in the tree T is represented as $\tilde{T}$. The node which is not a terminal node is called internal node.

If we consider the survival function at each node of T as non-parametric framework, then the estimation of survival function at each node is given by Kaplan-Meier method. Let $\mathscr{L}_t = \{(x_i, \delta_i, \boldsymbol{z}_i); i = 1, \cdots, N_t\}$ be the set of learning samples in the node t, then the Kaplan-Meier estimation of t is given by

$$\hat{S}_t(x) = \begin{cases} 1 & (x < y_{(1)}) \\ \displaystyle\prod_{i(x_i \leq x)} \left(1 - \frac{d_i}{n_i}\right) & (y_{(1)} \leq x) \end{cases}, \tag{2.1}$$

where $y_{(1)}$ represents the earliest event occurrence time in $\mathscr{L}_t$. d_i and n_i represent the number of events and risk at time $x_i (i = 1, 2, \cdots, N_t)$, respectively.

As one of a merit of tree structure, once a tree model is composed and survival function for each terminal node in the model is estimated by using $\mathscr{L}$, it is easy to apply a new patient to the model. That is, the survival function for the new patient is estimated as the Kaplan-Meier survival function of the terminal node t which the patient is assigned.

2.2. Construction method of survival tree

The CART algorithm, which is widely used to construct a tree model, is composed three steps. In the first step, which is called as splitting step, all learning samples $\mathscr{L}$ are recursively dichotomized and a large tree is constructed. For the set of samples $\mathscr{L}_t$ in an arbitrary node t, the data which satisfy the following rule are assigned to the left child node and the others are assigned to the right child node:

$$Z_j \leq c, \tag{2.2}$$

where Z_j is a numerical variable in $\boldsymbol{Z}$ and c is a cutting point of it ($j = 1, 2, \cdots, p$). If Z_j is the categorical variable and its possible values are $s_1, \cdots, s_l$, then the data which satisfy the following rule are assigned to the left child node:

$$Z_j \in \{s_1, \cdots, s_m\}, \tag{2.3}$$

and the others are assigned to the right child node ($m < l$). For the set of $\mathscr{L}_t$, let $\mathscr{L}_{tL}$ and $\mathscr{L}_{tR}$ be the set of samples in left and right child nodes, respectively. Then, all possible splits are evaluated using $\mathscr{L}_{tL}$ and $\mathscr{L}_{tR}$ to select an optimal splitting rule of t.

The evaluation criteria are proposed by various authors. In this paper, we use the criterion that using the exponential log-likelihood loss (Davis, and Anderson [5]). The assumption of exponential survival time is widely studied and used in many situations. As an example, this assumption is used in the decision of sample size (Lachin [11], Schoenfeld and Richter [15]). Because our proposed approach needs to evaluate a lot of splits for constructing a model, this criterion is suitable from simplicity of calculation. Moreover, this criterion has been shown a good performance in the comparative study of the splitting criteria by using simulations as previous research (Shimokawa et al. [17]).

Let the hazard of a node t be a constant:

$$\lambda(x|t) = \lambda_t,$$

where λ_t represents a constant parameter. Then the maximum log-likelihood estimator of λ_t is defined as

$$\hat{\lambda}_t = \frac{\sum_{i \in \mathscr{L}_t} \delta_i}{\sum_{i \in \mathscr{L}_t} x_i}.$$

Using $\hat{\lambda}_t$, the exponential log-likelihood loss of t is given by

$$R(t) = \sum_{i \in \mathscr{L}_t} \delta_i - \sum_{i \in \mathscr{L}_t} \delta_i \log(\hat{\lambda}_t). \tag{2.4}$$

For all possible splits, the sum of the exponential log-likelihood loss of the child nodes are evaluated by using (2.4). Then, the splitting point and the kind of covariate that has a minimum value of it are determined as the splitting rule. By recursively application of evaluation and splitting, the maximum size tree T_0 is constructed. As the stopping criterion of splitting, we can use the number of samples or events of the terminal node.

In second step, the set of subtree is constructed by recursively removing the splits in T_0 by using the cost-complexity measure (Breiman et al. [2]). This measure is defined by

$$R_\alpha(T) = \sum_{t \in \tilde{T}} R(t) + \alpha |\tilde{T}|, \tag{2.5}$$

where $R(t)$ is the exponential log-likelihood loss of t, which is obtained by (2.4) in splitting step, and $|\tilde{T}|$ is the number of terminal nodes of T, which is a complexity measure of the tree. α represents the penalty of $|\tilde{T}|$. This measure returns a small value when the values of the risk in each terminal node of T are small and the model is simple. For an arbitrary α, the subtree T_k which minimizes the (2.5) is obtained ($k = 1, \cdots, M$). If the value of α is 0, then the optimal subtree is the maximum size tree T_0. On the other hand, if α is ∞, then the tree T_M which has a root node only minimizes the measure. The optimal subtrees $T_0, T_1, \cdots, T_M$ are given by gradually increasing α from 0. The subtrees which are constructed by this method have a nested structure. That is, $T_k, \cdots, T_M$ are the subtree of T_{k-1}.

In the final step, the optimal subtree as prediction model is selected from $T_0, T_1, \cdots, T_M$. Because each subtree is specified for an arbitrary α in pruning step, this problem is equivalent to search the optimal penalty α in (2.5). There is some possible methods for this like test sample method, bootstrap, and we used V-fold cross validation method in this study.

2.3. Splitting method based on combination of covariates

In traditional CART algorithm for constructing a tree structured model, which is described in previous section, the splitting rule of the node is restricted by only one covariate as (2.2) or (2.3). As described in Section 1, this restriction has the possibility of overlooking a more suitable model in some situations. To handle these situations, we consider the splitting rule that two or more

combinations of covariates are considered for dividing the data in each node. For example, consider the case of combination of two covariates. If the covariate vector is defined as $\boldsymbol{Z} = (Z_1, Z_2, Z_3)$, where Z_j is the numerical variable $(j = 1, 2, 3)$, then the splitting rule in each node for traditional approach is one of the following structures:

$$\{Z_1 \leq c\},\ \{Z_2 \leq c\},\ \{Z_3 \leq c\}. \tag{2.6}$$

On the other hand in our approach, the splitting rule in each node is one of the (2.6) or following structures:

$$\{Z_1 \leq c \cap Z_2 \leq d\}, \{Z_1 \leq c \cap Z_2 > d\}, \{Z_1 > c \cap Z_2 \leq d\}, \{Z_1 > c \cap Z_2 > d\},$$
$$\{Z_2 \leq c \cap Z_3 \leq d\}, \{Z_2 \leq c \cap Z_3 > d\}, \{Z_2 > c \cap Z_3 \leq d\}, \{Z_2 > c \cap Z_3 > d\},$$
$$\{Z_1 \leq c \cap Z_3 \leq d\}, \{Z_1 \leq c \cap Z_3 > d\}, \{Z_1 > c \cap Z_3 \leq d\}, \{Z_1 > c \cap Z_3 > d\}.$$

In this example we considered the case that the variables included in covariate vector are numeric only, but if categorical variables are included in the vector we can consider as the same.

In traditional approach, as can be seen from this example, each splitting rule is practiced along a covariate axes. In proposed approach, it is practiced by whether data are included in a hyper-rectangle or not in some covariate spaces. The proposed approach has disadvantage that it requires a long learning time because it need to evaluate the large number of combinations of covariates. Specifically, the number of structures of splitting rule which are considered in this approach is $\binom{p}{k}2^k$ when the case of k combination of covariates is considered, while the number of it is p in traditional approach. However, PC technology in recent years may be able to resolves this disadvantage. As one merit of this approach, the pruning and selection steps in traditional CART algorithm can use as the same. Moreover, the processing time of these steps is not increased compared to the traditional approach.

§3. Simulation

3.1. Model and Setting

To compare the proposed and traditional approaches through simulations, we use the true tree model which are shown in Figure 1. The circles in the figures represent internal nodes. This model shows a typical XOR problem. The covariates used in this simulations are Z_1, Z_2 and Z_3. These are random values created from two patterns: one is a discrete uniform distribution with $\{1/50, \cdots, 1/50\}$, and the other is a Bernoulli distribution with parameter 0.5. This model assumes that the variables Z_1 and Z_2 are used in the true tree model and Z_3 is a nuisance.

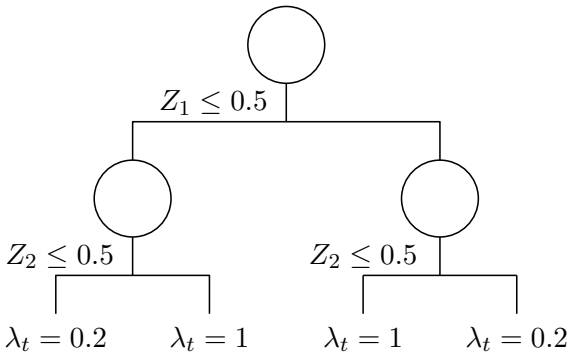

Figure 1: True tree structure used in simulation

We suppose exponential survival model, which has constant hazard to the change of time, for the simulations. This survival function is given by

$$S(y; \lambda_t) = P(Y > y; \lambda_t) = \exp(-\lambda_t x),$$

where the λ_t represents the constant hazard of node t. Based on the trees which are shown in Figure 1, we suppose λ_t as follows:

$$\lambda_t = \begin{cases} 0.2 & (Z_1 \leq 0.5 \cap Z_2 \leq 0.5) \\ 1 & (Z_1 \leq 0.5 \cap Z_2 > 0.5) \\ 1 & (Z_1 > 0.5 \cap Z_2 \leq 0.5) \\ 0.2 & (Z_1 > 0.5 \cap Z_2 > 0.5) \end{cases} \tag{3.1}$$

This model is difficult to detect in traditional approach, because the evaluations obtained by splitting in first node are nearly the same in each case that the covariate Z_1, Z_2, or Z_3 is used as splitting criterion.

By using uniform random numbers, the censoring rates are set as 0% and approximately 25% and 50%. The number of learning samples N are set to 200. The 5-fold cross validation is used in the selection step. We set 30 minimum number of events in nodes as the stop condition of splitting. Simulations are repeated 100 times in every setting. The number of covariates which are considered to construct a combination for splitting in proposed approach is restricted to two.

3.2. Evaluation methods

The integrated Brier score for censored data, which are proposed by Graf et al. [8], is used to evaluate the proposed approach. This score is calculated based on the test samples $\mathscr{L}_{\text{test}} = \{(x_i, \delta_i, \boldsymbol{z}_i); i = 1, 2, \cdots, N_{\text{test}}\}$. For each

simulations, this $\mathscr{L}_{\text{test}}$ is obtained from the simulated population of the same setting. Let $\hat{S}(x|\mathbf{z}, T)$ be the estimated survival function. Then the integrated Brier score for $\hat{S}(x|\mathbf{z}, T)$ is defined as

$$\text{IBS}_T = \frac{1}{\max(x_i)} \int_0^{\max(x_i)} \text{BS}_T(x)dx,$$

where $\text{BS}_T(x)$ is the Brier score of T, which are defined as

$$\text{BS}_T(x) = \frac{1}{N_{\text{test}}} \sum_{i \in \mathscr{L}_{\text{test}}} \{(0 - \hat{S}(x|\boldsymbol{z}_i, T))^2 I(x_i \leq x, \delta_i = 1)(1/\hat{G}(x_i)) + (1 - \hat{S}(x|\boldsymbol{z}_i, T))^2 I(x_i > x)(1/\hat{G}(x_i))\},$$

where $\hat{G}(x)$ is the Kaplan-Meier estimate of the censoring distribution of C, which are obtained by using (2.1) based on $\{x_i, 1 - \delta_i; i = 1, 2, \cdots, N_{\text{test}}\}$. The Brier score is construted as the mean square error between the $\hat{S}(x|\mathbf{z}, T)$ and the test data, which are weighted based on the loss of information due to censoring. We use the up to median time in learning sample to evaluate this score. The number of test samples N_{test} is set to 200 for every setting.

Moreover, we use the following criterion as the measure of the explained residual variation:

$$R^2 = 1 - \frac{\text{IBS}_T}{\text{IBS}_{T_M}},$$

where IBS_{T_M} is the integrated Brier score evaluated from the T_M, which has a root node only. Based on the IBS_T, R^2, number of terminal nodes, and the proportion of covariates which are used in the final tree model, we evaluate the each approaches.

3.3. Results

Table 1 shows the results of the simulations for traditional and proposed approach. The values in tables are average of each measure obtained through simulations. In each case of simulations, the proposed approach shows the good results than the traditional approach about the values of IBS_T, R^2 and the proportion of covariates which are used in the final tree model. Especially about the selected covariates in the final model, the proposed approach did not choose the nuisance covariate almost. When censoring rate is increased, the traditional approach make it difficult to detecting the Z_1 and Z_2 which is the variables used in true tree model. This will cause the results of explained residual variation of the traditional approach be negative when the censoring rate is 50%.

Table 1: The results obtained by simulations.

covariate	censor rate	approach	IBS_T	R^2	$\lvert\tilde{T}\rvert$	prop.	Z_1	Z_2	Z_3
quantitative	0%	traditional	17.3	3.1	4.3		1.4	1.3	0.6
		proposed	15.2	14.8	3.2		2.2	2.1	0.1
	25%	traditional	16.1	0.6	3.3		0.9	0.9	0.4
		proposed	14.3	11.8	3.0		2.0	2.0	0.0
	50%	traditional	13.8	-0.8	2.5		0.5	0.6	0.4
		proposed	12.9	6.0	2.6		1.6	1.6	0.0
binary	0%	traditional	17.1	9.6	3.7		1.2	1.2	0.3
		proposed	16.1	14.8	3.1		2.1	2.0	0.0
	25%	traditional	17.5	0.3	3.0		0.7	0.5	0.7
		proposed	15.3	12.7	3.0		2.0	2.0	0.0
	50%	traditional	14.8	-0.9	2.0		0.4	0.2	0.4
		proposed	14.1	3.6	2.1		1.1	1.1	0.0

IBS_T: the integrated brier score evaluated from the selected tree $\times 100$, R^2: the explained residual variation $\times 100$, $|\tilde{T}|$: the number of terminal nodes about selected tree, prop.: the proportion of the covariates which are used in the final tree model.

As a natural outcome, the proposed approach tends to construct the small size tree compared to the traditional approach since more than one covariate is considered in each node at a time. In the present case, the optimal sizes of tree are 4 and 3 in traditional and proposed approach, respectively. If censoring is not occurred, both approach have almost correct number of terminal nodes. However, if censoring rate will be increased, the traditional approach rapidly decreases the number of nodes than the proposed approach. As a result, we conclude that the proposed approach give a more suitable model than the traditional approach when the true model includes the XOR problem.

§4. Example

We show the application of the proposed approach using leukemia patients bone marrow transplantation data. This data was collected from 1984 to 1989 at one of four hospitals. The data are composed of 137 patients. The observation period is defined from the data of transplant surgery to the date of relapse, death or last survival verified. The observation is considered as an event occurred if it was determined by death or relapse. The 54 patients are censored (the censoring rate is about 39%). We used five covariates: Z_1 is the indicator of Acute Myelocytic Leukemia (AML) low risk group or not, Z_2 is the indicator of AML high risk group or not, Z_3 is the indicator of French-American-British Classification (FAB) grade is 4 or 5 and the disease group is AML or not, Z_4 is the donor age, and Z_5 is the patient age. The patients

which are not included in AML group are included in Acute Lymphoblastic Leukemia (ALL) group. The details of this research are given in Copelan et al. [3]. The data used in this example are available from the Web site offered by Klein and Moeschberger [10].

We set 10 minimum number of events in nodes as the stop condition of splitting step. The number of cross validation is set to 10. The tree obtained by proposed approach is shown in Figure 2. The circle and square shapes in the figure represent the internal and terminal nodes, respectively. The values in the shapes represent the number of samples included in the node, and the values in parentheses represent the number of events. The tree has two splitting points and three terminal nodes ($t_1 - t_3$). As the result, the patients are divided to three groups:

$$
\begin{aligned}
\text{group } t_1 : \{ & \text{ 41 or younger donor} \cap \text{12 or older patient} \\
& \cap \text{ AML high risk or ALL} \cap \text{FAB grade is 4 or 5}\} \\
\text{group } t_2 : \{ & \text{ 41 or older donor} \cup \text{12 or younger patient } \} \\
\text{group } t_3 : \{ & \text{ 41 or younger donor} \cap \text{12 or older patient} \\
& \cap \{ \text{ AML low risk} \cup \text{FAB grade is 1, 2 or 3 }\}\}.
\end{aligned}
$$

The graphical representation of the covariate space which is separated by the obtained tree model is shown in Figure 3. Although the model obtained by the proposed approach seems difficult to understand than the classical approach, each terminal node is disjoint. The Kaplan-Meier survival curves for each group are shown in Figure 4. From the survival curves, we can understand that the group t_1 has the highest risk of death or recurrence and the group t_3 has the lowest risk of it. The group t_2 has the risk of death or recurrence between t_1 and t_3. The Kaplan-Meier survival curves are well separated from each other, and we conclude that the proposed approach give the reasonable result.

§5. Conclusion

In this paper, we proposed a new approach for constructing a survival tree based on CART algorithm. The proposed approach considers two or more combinations of covariates for dividing a node. As the motivation of proposing this approach, we have considered that the proposed approach is more suitable than the traditional approach to some problems like XOR.

Through the simulation study, we have been shown the performance of the proposed approach. As the result, the proposed approach is considered that has a potential to construct a more suitable tree model than the traditional approach in some situations. The utility of the proposed method has been

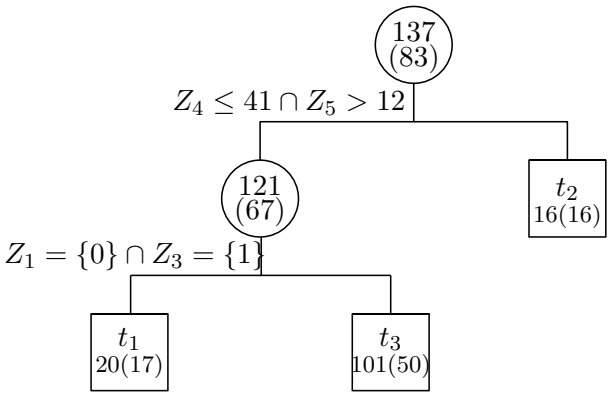

Figure 2: Tree constructed from bone marrow transplant patients data. The values in the shapes represent the number of samples included in the node, and the values in parentheses represent the number of events. The terminal nodes are represented as t_1 - t_3.

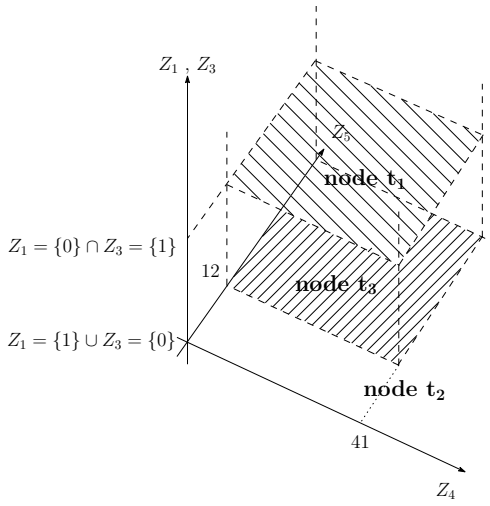

Figure 3: The covariate space represented by tree model in Figure 2.

shown by using an actual medical data. The tree constructed from the data divided the patients to three groups. From the Kaplan-Meier survival curves of each group, we conclude that the obtained results are considered to be reasonable.

Tree structure model has an advantage that the relationship between the covariates and hazards is easy to show. Moreover, there is another advantage that is easy to insert a new data to the model. The proposed approach has

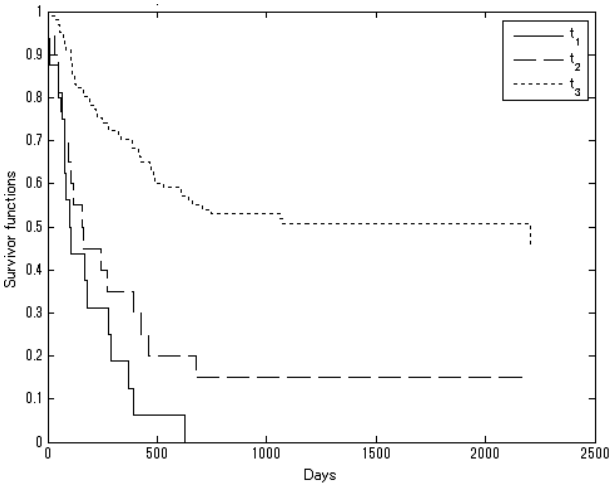

Figure 4: The Kaplan-Meier survival curves for each terminal node of Fig.2.

the possibility to construct an optimal model, while holding this advantages. However, from the aspect of "fragmentation" (Friedman and Fisher [6]), the proposed approach has a risk of building a low performance model than the model obtained by classical approach. That is, in a covering algorithm the low number of splitting step becomes a cause of a high value of bias and variance in the model. Further studies are needed in order to address this problem. As an another disadvantage, it requires a long learning time, but we consider that PC technology in recent years may be able to resolve this problem.

References

[1] I. Bou-Hamad, D. Larocque and H. Ben-Ameur, A Review of Survival trees, *Statistics Surveys* **5** (2011), 44-71.

[2] L. Breiman, J. H. Friedman, R. A. Olshen and C. Stone, *Classification and Regression Trees.* Wadsworth, California, 1984.

[3] E. A. Copelan, J. C. Biggs, J. M. Thompson, P. Crilley, J. Szer, J. P. Klein, N. Kapoor, B. R. Avalos, I. Cunningham, K. Atkinson, K. Downs, G. S. Harmon, M. B. Daly, I. Brodsky, S. I. Bulova and P. J. Tutschka, Treatment for Acute Myelocytic Leukemia with Allogeneic Bone Marrow Transplantation Following Preparation with BuCy2, *Blood* **78** (1991), 838-843.

[4] D. R. Cox, Regression Models and Life-Tables, *Journal of the Royal Statistical Society* **34**(2) (1972), 187-220.

[5] R. B. Davis and J. R. Anderson, Exponential Survival Trees, *Statistics in Medicine* **8** (1989), 947-961.

[6] J. H. Friedman and N. I. Fisher, Bump hunting in high-dimensional data, *Statistics and Computing* **9** (1999), 123-143.

[7] L. Gordon and R. A. Olshen, Tree-Structured Survival Analysis, *Cancer Treatment Reports* **69** (1985), 1065-1069.

[8] E. Graf, C. Schmoor, W. Sauerbrei and M. Schumacher, Assessment and Comparisons of Prognostic Classification Schemes for Survival Data, *Statistics in Medicine* **18** (1999), 2529-2545.

[9] S. Keles and M. R. Segal, Residual-based tree-structured survival analysis, *Statistics in Medicine* **21** (2002), 313-326.

[10] J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data 2nd ed.* Springer, New-York, 2003.

[11] J. M. Lachin, Introduction to sample size determination and power analysis for clinical trials, *Controlled Clinical Trials* **2** (1981), 93-113.

[12] M. Leblanc and J. Crowley, Survival Trees by Goodness of Split, *Journal of the American Statistical Association* **88** (1993), 457-467.

[13] J. N. Morgan and J. A. Sonquist, Problems in the analysis of survey data, and a proposal, *Journal of the American Statistical Association* **58** (1963), 415-434.

[14] M. Radespiel-Tröger, T. Rabenstein, H. T. Schneider and B. Lausen, Comparison of Tree-based Methods for Prognostic Stratification of Survival Data, *Artificial Intelligence in Medicine* **28** (2003), 323-341.

[15] D. A. Schoenfeld and J. R. Richter, Nomograms for Calculating the Number of Patients Needed for a Clinical Trial with Survival as an Endpoint, *Biometrics* **38** (1982), 163-170.

[16] M. R. Segal, Regression Trees for Censored Data, *Biometrics* **44** (1988), 35-47.

[17] A. Shimokawa, Y. Kawasaki and E. Miyaoka, Comparison of splitting methods on survival tree. *Proceedings of the 27th International Biometric Conference*, Florence, July 6-11, (2014), 37806.

[18] T. M. Therneau, P. M. Grambsch and T. R. Fleming, Martingale-Based Residual for Survival Models, *Biometrika* **77** (1990), 147-160.

[19] H. P. Zhang, Splitting Criteria in Survival Trees, *In Statistical Modelling : Proceedings of the 10th International Workshop on Statistical Modeling*: July 10-14; Innsbruck, Austria, (1995), 305-314.

Asanao Shimokawa
Department of Mathematical information, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
E-mail: a.shimokawa0226@gmail.com

Yohei Kawasaki
Department of Mathematics, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
E-mail: ykawasaki@ma.kagu.tus.ac.jp

Etsuo Miyaoka
Department of Mathematics, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
E-mail: miyaoka@rs.kagu.tus.ac.jp