



Heart Transplant Risk Learning

CS486 final project

Yue Huang(y4huang) 20329619

Yuxing Luo (y57luo) 20498935

Source Code Link: <https://github.com/yugiohatemu/MTP>

Abstract: Our intuition is to provide some suggestions for patients who tend to receive heart transplantation, by analyzing the effects of various related covariates. To be specific, for those who are not able to find a donor heart to process transplantation surgery, we assess the effects of age and prior surgery experience on the survival of patient. For those who found the donator, we predicate the possibility of heart rejection and provide estimation of the relative risk of transplantation by considering covariates on the success rate of heart transplant surgery.

Key words: Heart transplant; Survival data; Decision tree; Risk estimation;

§1. Introduction

Inspired by the Stanford Heart Transplant Data (Crowley & Hu, 1977; Kalbfleisch & Prentice, 2002), we propose an decision tree based model to predict the risk of heart transplant surgery. We investigate five different corresponding attributes that may have influence on the success rate of heart transplantation, as well as the survival rate.

Our intuition is to provide some suggestions for patients who tend to receive heart transplantation, by analyzing the effects of various related covariates. To be specific, for those who are not able to find a donor heart to process transplantation surgery, we assess the effects of age and prior surgery experience on the survival of patient. For those who found the donator, we predicate the possibility of heart rejection and provide estimation of the relative risk of transplantation by considering covariates on the success rate of heart transplant surgery.

§2. Database

§2.1 Sources

Two versions of the Stanford Heart Transplant data have been applied in our experiments. The first is from Kalbfleisch and Prentice (first edition) while the second is from Crowley and Hu. Each database contains 103 observations. The two versions are slightly different, but all encompass the attributes we plan to observe. The first edition is used as our test set, and the second is our training set.

Our implement mainly focused on the following attributes:

Age

Ages of patients who suffer from heart diseases.

Prior Surgery Status

Observe whether the patients have received surgery before, including all types of surgeries.

Mismatch of Alleles, Mismatch of Antigen and Mismatch Score

Allele is the alternative form of a gene for a character producing different effects. Antigen is a term in the field of immunology. An antigen is a substance that evokes the production of one or more antibodies. Both terms will be taken into consideration when we examine whether a donor heart matches the physiologic peculiarities of a certain patient. Mismatch score also measures the matching degree of the donor heart.

Reject and Death

Death denotes to the survival of the patient, while Reject denotes death by rejection.

§2.2 Pre-processing

Some of the attributes are merged for the convenience of classification. e.g., *Date of Birth* and *Date of Transplant* are merged into attribute *Age*.

Some of the attributes are given up due to their low contribution to our model. e.g., *Date of Acceptance* (by the hospital) and *Date Last Seen*.

We have to give up one useful attribute *Survival Time*, since this attribute can only be found in the first edition of Stanford Heart Transplant Data, the second database, i.e., our training set does not offer the records for this attribute.

Few noisy data are dropped during the pre-process of database. For example, the 39th record of the training set is listed below:

id	age	Date of Transplant	dead	Prior Surgery Status	Mismatch of Alleles	Mismatch of Antigen	Mismatch Scores	Reject
39	51	11/12/19	1	0	NA	NA	NA	-

Table 1

It is confusing when the date of transplant is settled while the mismatch attributes are unavailable. Therefore, we dropped such records on both training set and test set.

§3. Model

Our model contains two sub-models. The first model is designed for those who tend to receive heart transplantation but still on their way finding donor hearts. The second model is used for those who are lucky to found a donator but has not decided whether they should receive the transplantation surgery with the heart.

For the first case, our model will aim to assess the survival rate of patients by observing attribute *Age* and *Prior Surgery Status*. To be specific, each record will be classified into *dead* = true or *dead* = false by examine these two attributes. As showed

in Figure 2.

For the second case, attribute *Age*, *PriorSurgeryStatus*, *Mismatch of Alleles*, *Mismatch of Antigen*, *Mismatch Scores* are investigated, and all the records will be classified into *reject* = true or *reject* = false. As showed in Figure 3.

Model 1		
Attribute 1	Attribute 2	Classification
<i>Age</i>	<i>Prior Surgery Status</i>	<i>dead</i>

Table 2

Model 2					
Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5	Classification
<i>Age</i>	<i>Prior Surgery Status</i>	<i>Mismatch of Alleles</i>	<i>Mismatch of Antigen</i>	<i>Mismatch Scores</i>	<i>reject</i>

Table 3

§4. Experiment

§4.1 Result

The following is the accuracy of our first model, i.e. model for those who have not found donor hearts yet.

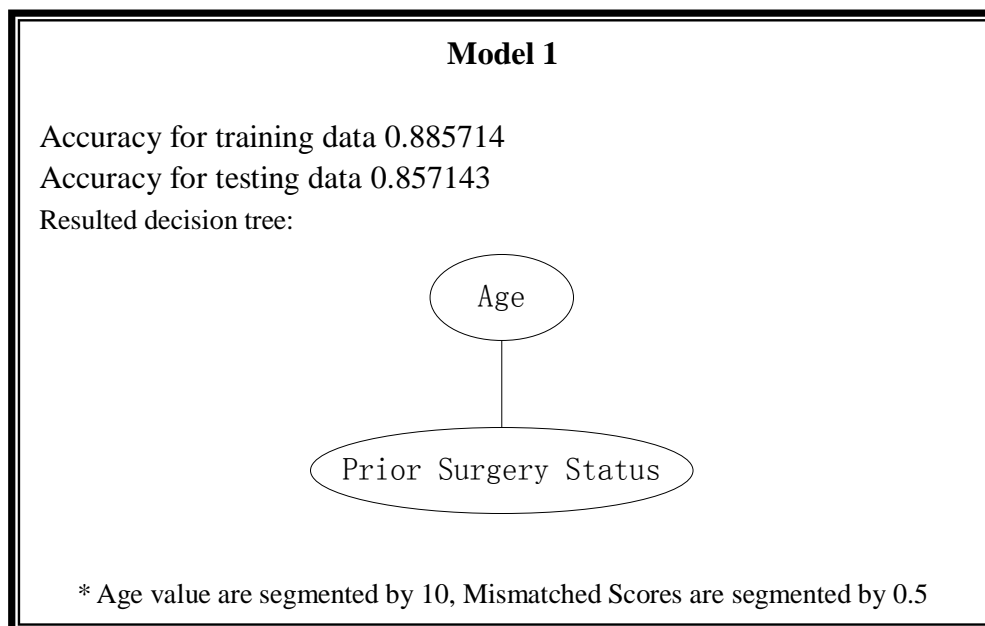


Table 4

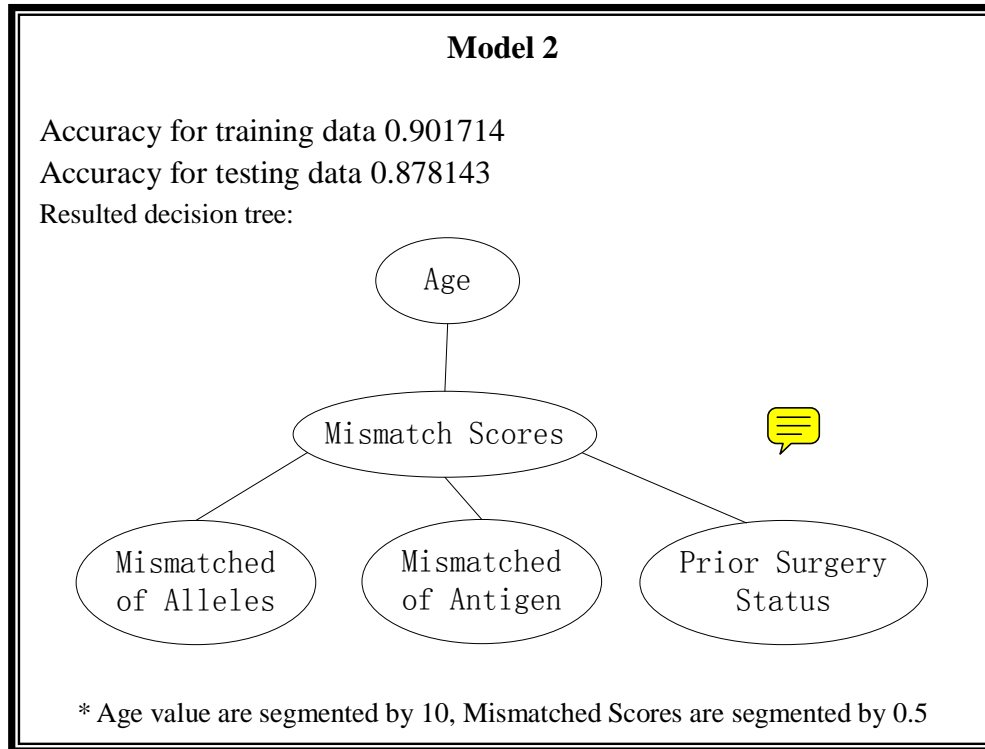


Table 5

According to the results, Model2 is more accurate. This is corresponding to our intuition since more attributes leads to a higher

§4.2 Optimization

During the experiment, we segment the values of continues attributes as the new values. By using our intuition, we guess that the smaller interval will lead to better performance.

To be specific, the following two attributes optimize the performance of our model.

Age Segmentation and Mismatch Score Segmentation

We change the segmentation interval of these two attributers, and get the following results:

Model 1 (For those who have not found donor heart yet)

Age Segmentation Interval	Training Set Accuracy	Test Set Accuracy
10	0.885714	0.857143
5	0.886014	0.858109

Table 6

Model 2 (For those who have already found donor hearts)

Age Segmentation Interval	Mismatch Scores Segmentation Interval	Training Set Accuracy	Test Set Accuracy
10	0.5	0.901713	0.878143
5	0.5	0.901914	0.878493
5	0.3	0.901914	0.879543

Table 7

According to Table 6 and Table 7, we can see that the accuracy improves if we set smaller intervals for the value of continues attributes.

§5. Conclusion

In this work, we provide two decision-tree-based models to predict the survival possibility of heart transplant surgery. We firstly parse the raw data into our desired format, and do some pre-processing to drop off some noisy data. Then we implement our models based on the database, and output the accuracies and tree models. Additionally, we consider the influence of interval size on continues data which need to be segmented during the parsing process. All in all, our models can provide reasonable predications for the survival rate of heart transplantation.

Acknowledgements

We would like to thank Prof. Poupart.

References

- [1] Kalbfleisch, Prentice. 2006. The Statistical Analysis of Failure Time Data. Stanford Heart Transplant Press, Appendix I, pages 230-232.
- [2] J Crowley and M Hu (1977), Covariance analysis of heart transplant survival data. Journal of the American Statistical Association, 72, 27-36.
- [3] Source of Our database: <http://lib.stat.cmu.edu/datasets/stanford>