Graft

http://gft.sagepub.com

Practical Remarks in Analyzing Kidney Transplant Data Using a Systems Approach with Neural Networks

George Diderrich and Ronald P. Pelletier

Graft 2002; 5; 399

DOI: 10.1177/152216202237627

The online version of this article can be found at:

http://gft.sagepub.com/cgi/content/abstract/5/7/399

Published by:

SAGE Publications

http://www.sagepublications.com

Practical Remarks in Analyzing Kidney Transplant Data Using a Systems Approach with Neural Networks

George Diderrich and Ronald P. Pelletier

In this article, the authors briefly examine the practical issues of applying neural network techniques to survival theory, followed by an outline of a systems approach that enhances the ability of the analyst to better understand the data. It is based on our prior experience in analyzing kidney transplant patients and the first author's experience using the second author's experience in industrial mathematics. In addition, the authors provide an introduction to neural network technology and its connection to simple linear models, along with a brief outline of the use of soft computing, such as scoring functions to incorporate subjective information into models.

Introduction

In this article, we examine the practical issues in applying neural network techniques to survival theory followed by an outline of a systems approach that enhances the ability of the analyst to better understand the data. We describe how one would apply neural network techniques with commercially available tools and how to integrate these tools to perform a systems analysis of the data.

Our recent attempt to quantify risk of kidney transplant patients using survival theory demonstrates the need to go beyond the standard statistical approach using survival theory methods, using neural network and other advanced methods.[1-4] Previous work dealing with manufacturing data[5,6] led to a greater confidence in the results as determined by this approach.

By survival theory, we refer to Kaplan-Meier analysis, Cox proportional hazard models, parametric models using exponential or Weibull distributions, and logistic regression of dichotomous variables. The commercially available exploratory data analysis tools[7,8] contain specialized tool kits to handle the preceding models and allow the user to program and develop customized analyses with their own scripting language. For technical details on the statistics of survival theory, Klein and Moeschberger's *Survival Theory: Techniques for Censored and Truncated Data*,[9] along with Therneau and Grambsh's *Modeling Survival Data: Extending the Cox Model,*[10] are among the many fine texts in the field.

Some Practical Remarks Using Neural Network Technologies

First, we briefly discuss neural network theory to give the reader a basic idea of what it is. We then follow with some practical remarks in applying this technology to survival theory.

Neural Network Models as Nonlinear Regression

What is a neural network and why is it so popular? What are some of the advantages of the technique and what are some of the downsides?

The easiest way to understand a neural network is to think of it as a warped linear regression model. A simple linear regression has the following format:

$$y = 3x + 6,$$

where y is a desired output, the predicted outcome variable, and x is an input variable. The factor 3 is a multiplicatiosn term or weight that scales the x, and the constant term 6 is a shift change to finally

George Diderrich, PhD
9265 N. Sleepy Hollow Lane
Bayside, WI 53217
Tel: 414.228.6554
Fax: 414.228.6397
email: gdiderrich@wi.rr.com
DOI: 10.1177/152216202237627

SURVIVAL THEORY STATISTICS

A branch of statistics that specializes in the analysis of the mortality and survivability aspects of people or processes in terms of the probabilities of exceeding a given time (e.g., the probability of a person surviving more than 4 years).

SYSTEMS ANALYSIS

An orderly procedure to assemble information or data with the intent to incorporate a larger scope of the interactions affecting a given object under investigation.

make the output y's after processing the arithmetic in the equation. A neural network is slightly more complicated, having, for example, the following form:

$$y = 1/(1 + e^{-3x}),$$

where e^{-3x} refers to an exponential function. The above function is actually a sigmoid function, which causes the y output to be warped by using an exponential function as the input x. The shape of this function is shown in Figure 1.

The idea of a neural network elaborates on this concept by incorporating a large number of interacting neurons, as shown in Figure 2. Because each of the interior neurons (shown as circles in Fig. 2) contains a sigmoid function, each neuron contributes both linear and nonlinear effects simultaneously. Therefore, one can think of this entity as producing a smoothly warped, higher dimensional surface that explains the y in terms of the x's. In any case, a linear model is a special type of neural network not having a nonlinear sigmoid function and, conversely, a neural network is a warped form of a linear model.

It is important to realize that the nonlinear aspects of the information are not modeled by "intuitive" nonlinear functions. By this we mean that a statistical or mathematically oriented analyst would directly incorporate into the model a polynomial function or some other easily understood nonlinear function. In some form, the neurons already contain generic, nonlinear characteristics that allow the information content of the data to be effectively modeled. This is perhaps the greatest power of the neural network as well as its greatest downfall. We cannot intuitively "see" how the neurons bend and sculpt the hyper surface that represents the features of the data. Therefore, this is sometimes referred to as the "black box."

Training the network involves determining the weights, that is, the multiplication factors and other processing parameters in the network. This amounts to a generalized linear regression computation, an adequate discussion of which can be found in Bishop's *Neural Networks for Pattern Recognition.*[11]

This leads to the following practical paradigm:

$$\text{Linear model} \leftrightarrow \text{neural network model.}$$

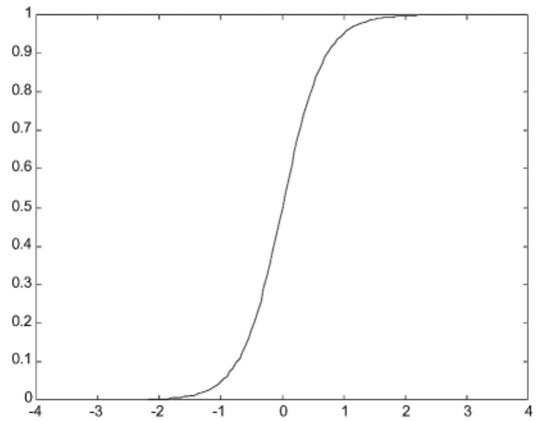

Figure 1. The sigmoid function (y = 1/(1 + exp - 3x)) is linear in the middle and warped at the "tails" near ±1.

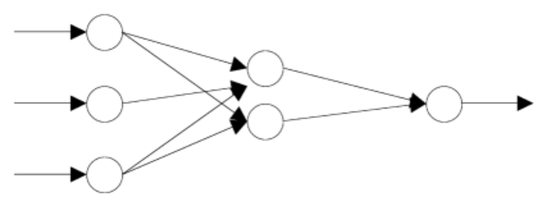

Figure 2. An example of a 3:2:1 feedforward network: 3 inputs, 2 hidden nodes, 1 output. The input signals x are processed by the weights by multiplying the input times the weight and then adding the combination. The resulting combinations are then passed through a sigmoid function at each neuron, producing a general output y.

Therefore, wherever a linear model is encountered such as

$$y = w_1x_1 + \ldots + w_nx_n,$$

the resulting model could be replaced by a neural network model,

$$y = NN(w_1, \ldots, w_n: x_1, \ldots, x_n).$$

Yet, as we have shown, this is essentially the same as the nonlinear regression model. This insight has far-ranging implications, which are still being developed, such as relationships to standard statistical methods, including nonlinear regression and Bayesian modeling.[5,12,13]

Applications of Neural Network Technology to Survival Theory

Liestol et al.[14] are pioneers in displaying how neural network technology can be applied to survival

Table 1 | **OUTLINE OF TECHNICAL STEPS IN APPLYING NEURAL NETWORKS TO SURVIVAL THEORY MODELS**

1. Formulate the problem probabilistically and identify covariates *x* and outputs *y*.
2. Set up the likelihood equations and write down the likelihood function.
3. Replace the linear form with an equivalent neural network form.
4. Calculate E = -log of the likelihood function.
5. Identify the individual terms as cost forms.
6. Minimize E with respect to the weights w embedded in the cost form. Note that one will need to calculate the partial derivatives of the terms in E with respect to the weights w.
7. Calculate the 2nd-order derivatives to obtain information on the correlation and standard deviations of the parameters and to provide "curvature" information to speed up the minimization.
8. Perform model evaluation and model validation procedures.

theory, as well as the most general probabilistic situation in any application. For a more recent and thorough discussion, we refer the reader to Neal.[12] Table 1 provides a brief outline of the steps for applying neural networks to survival theory models.

Remarks Concerning Commercially Available Neural Network Tools for Survival Theory

Commercially available neural network software, such as StatSoft's *Statistical Neural Networks Manual*[7] and the Ward System Neural Network Classifier Package,[8] cannot easily incorporate the above steps for survival theory. One must construct his or her own neural network tools until these innovations become available. It is usually awkward to analyze survival theory models because the cost forms are not sum of squares forms. A cost form measures the probabilistic influence of the events in the likelihood function. A sum of squares form is derived from independent standard normal probability distributions. One therefore uses nonstandard forms. This is possible using the scripting language noted within the JMP Statistical Manuals,[15] but will again require extra effort.

However, because logistic regression is actually a special type of neural network (as noted above), classifier type neural network tools and logistic regression models are directly comparable. Hence, it is valid to compare the tools and results. The results obtained in Pelletier and Diderrich[2] were obtained using logistic regression[15] and classifier neural network[8] tools using a genetics training procedure. The tools in the neural network package[7] were also used.

Increasing the Confidence of the Analyst

We found that for the data set under investigation, there was no highly significant difference between the models; however, the benefit of using different tools was the ability to examine the data from fresh and different perspectives.[2] For example, the top 2 most significant variables should correlate with the ones identified by the statistical procedure. This helps to increase the user's confidence that the effects truly exist. On the other hand, incorporation of soft computing (i.e., scoring function methods) increased the number of viable explanatory variables that might lead to an advantage in understanding the data versus an approach that does not use these methods or models. We next present a brief discussion of these concepts.

Systems Approach to Analyzing Survival Theory Using Neural Networks and Soft Computing Methods

By systems approach, we mean the use of all available technologies to extract insights from the data through a systematic set of procedures. Ideally, we would automate this as much as possible. However, at present much of this is done by hand. These tools and methods allow one to provide the necessary level of insight to instill confidence in the results, which is worth the extra effort in the long run.

Brief Comments on Soft Computing Techniques

The idea of soft computing is to systematically incorporate intuitive information into a model using a scoring function or related methodology. Experiments have demonstrated that the incorporation of

Table 2 | **OUTLINE OF A SYSTEMS PROCEDURE**

A. Data cleansing
 1. Remove outliers.
 2. Remove inconsistencies.
 3. Fill in data if possible.
 4. During model assessment, remove or repair suspect data points.

B. Identify suitable response variables.

C. Using one-at-a-time models, identify the most likely variables that explain the responses. In survival theory applications:
 1. Apply Kaplan-Meier tool
 2. Logistic regression
 3. Parametric tools
 4. Cox semiparametric tools.

D. Repeat step 3 using neural network tools.
 1. Note that only binary variables using a classifier type neural network can be done easily with commercially available tools. These models are then compared with logistic regression models.
 2. The user must develop his or her own specialized neural network tools and or wait until the commercial neural network and statistical developers produce these innovations.

E. Based on intuitions developed from the above steps, experiment with adding scoring function variables. This is subjective and will require iterative steps.

F. Build multivariate models using the above tools.

G. Contemplate building consensus models that combine all of the above.

H. Experiment with iterative feedback of the models and further intuitive scoring function variables.

I. Evaluate the results and write an easily understood rationale of the data insights obtained using the above procedures.

J. Review and repeat the above with the intent of simplifying the models and discussion.

SCORING FUNCTION

A method, sometimes subjective, to code the importance of the impact of a series of input variables by assigning a value or score to the variable. The overall score and the procedure in obtaining the score is the "function" aspect of the definition.

additional variables expanded on from the primitive set of input columns improves models. The latter procedure was used in our recent study[2] and consists of assigning a score or weight to a variable, assigning for example a partial score of 1 for type of drug used to suppress immunological response and a score of 5 for no use of drug, a weight of 10 for differences in the race of the recipient, and so on. Users design their own scoring function based on the statistical, neural network experiments and clinical knowledge, and then test the results using the same tools. This procedure provides a general methodology to "amplify" the information content in the models while at the same time providing an immediate understanding of the relationships between the predicted outcome variable and the explanatory variables. We note in passing that this is also related to fuzzy logic methods.[16]

Multivariate Models Remove Cluttering Noise and Improve the Correlations

Another important technique in extracting intelligence from the data is to build multivariate models after the key variables have been identified. The addition of the variables tends to remove the cluttering noise in the data. In effect, in developing multivariate models, variables that had a weak or nonsignificant correlation when put in the context of other weak variables suddenly show much greater significance when placed in a model. This phenomenon could be likened to the effect that while 1 stage hand cannot lift a curtain very easily by himself, the help of 2 or 3 more can lift the curtain.

General Outline of the Systems Procedure

In Table 2, we briefly outline the systems procedure for extracting the data.[13] In practice, due to project time pressure or other related issues, not all of the steps are rigorously carried out. Probably the most obvious is the data-cleansing step, which is sometimes given short shrift, and hence one is forced to repeat this step later. One is advised to allocate the necessary time in a project to this 1st step. A systems approach is the ideal method to systematically analyze data.

KIDNEY TRANSPLANT DATA

Patient database that includes information relating to the survival aspects of the ability of a kidney transplant graft to maintain proper functioning.

A Systems Procedure and Kidney Transplant Data

The above systems procedure was used to examine a large database of kidney transplant recipients transplanted at The Ohio State University Transplant Program between 1982 and 1996.[2] This database consisted of a number of recipient and donor clinical variables. The available variables for analysis included donor and recipient age, race, and gender, the degree of tissue matching between donor and recipient (i.e., the degree of human leukocyte antigen matching), the recipient panel-reactive antibody (PRA) level at the time of transplant, and the historical highest PRA level. Other recipient variables included the transplant number (1st, 2nd, etc.), the type of donor (living or cadaveric), the drugs used for immunosuppression (induction and maintenance), and the number and time posttransplant of all acute rejection episodes. Finally, the date of recipient death or return to dialysis was recorded when either event occurred. This database included 1952 kidney transplants in 1840 patients.

The initial challenge was to identify a suitable response variable for analysis. Initially, we chose kidney graft loss as the response variable. However, due to the relatively low frequency of this occurrence, we were unable to develop statistical models that were able to predict graft loss with an acceptably high degree of accuracy. Unfortunately, the initial models were better at predicting those patients who would not lose their grafts. From a clinically practical standpoint, we wanted to identify recipients at high risk of graft loss, not those at low risk of graft loss. Thus, we developed a different response variable, as discussed below.

After our initial efforts at developing a survival model following kidney transplantation, we decided on a different approach by focusing our survival model to identify recipients at risk for long-term graft loss (i.e., graft loss that occurred beyond 2 years after transplantation). Previous studies have shown that in our transplant program, the number of treated acute rejection episodes has a very significant negative impact on the length of graft survival. Thus, we reasoned that any recipient whose transplant survived beyond 2 years, but who lost their transplant before 4 years posttransplant, was a high-risk recipient. We also reasoned that any recipient who suffered 1 or more acute rejection episodes after transplantation was also at high risk for a shorter graft survival time. Thus, the new response variable was a high-risk variable defined as any recipient who lost his or her graft between 2 and 4 years after transplantation, or any recipient who experienced an acute rejection episode after transplantation. This new response variable divided the patients in the database into 2 roughly equally sized populations, namely, low-risk patients and high-risk patients.

Preliminary evaluation of the data was undertaken to identify the variables that appeared to influence our response variable. Univariate (or "one-at-a-time") models were employed to identify the variables most likely to explain the variation in graft survival or acute rejection and, by inference, identify other variables that influence our response variable. Analyses of Kaplan-Meier survival curves were very useful in identifying noncontinuous variables that determine populations of patients with a significantly poorer graft survival. Such variables include the recipient race, type of donor organ (living or cadaveric), type of immunosuppressant, and number of acute rejection episodes. Continuous variables, such as recipient age and the timing of the 1st acute rejection episode, were identified as variables that classify patients with poor graft survival, using parametric or Cox semiparametric tools. Both noncontinuous and continuous variables were also evaluated by logistic regression to verify significance in identifying recipients who lost their graft or experienced an acute rejection episode.

Our next approach was to experiment with intuitive scoring functions. Previously identified variables (listed above) that helped to explain our response variable were scored in an effort to enhance their ability to predict the response variable. Additionally, a sum of these scoring functions (termed the risk score) was evaluated to determine whether our ability to correctly predict the response variable was improved. To date, the use of scoring functions has provided an additional pool of explanatory variables that sometimes enhance the models but at other times do not appear to significantly improve the models. More research will be required to establish the statistical efficacy of these procedures.

NEURAL NETWORKS

A mathematical object consisting of a system of artificial neurons resembling a biological network of neurons that processes input patterns to produce output patterns of interest to an analyst.

Multivariate models were then employed and contrasted with neural network models to verify the significant variables that explain our response variable, as well as to examine variable interactions. Additionally, these models were used to evaluate the usefulness of our scoring functions. The nominal logistic regression and neural network models showed significant agreement in identifying the variables most likely to explain our response variable, which were the time to 1st acute rejection episode, type of induction agent, recipient race, and type of donor organ (living or cadaveric).

Future efforts to improve our models include addition of other clinical variables not available at the time we undertook these studies (including cause of recipient renal failure, level of recipient education and financial status—which might serve as surrogate markers for compliance with taking immunosuppresion medications—post operative creatinine levels, histology results in biopsy specimens [if obtained], as well as the presence or absence of detectable antibodies in recipient sera that are reactive to their donor organ), continued refinement of our intuitive scoring functions, and the assembling of consensus models using all of the above statistical tools.

Summary and Conclusions

In this article, we emphasized the importance of using a systematic approach (i.e., a systems approach) to analyze data by using advanced methods including neural network technologies and soft computing, such as scoring functions to incorporate intuitive information into models. We also emphasized the importance of building multivariate models that explain the data more precisely. Although we lament the practical fact that at present commercially available neural network tools cannot easily perform survival theory methodology (except that the neural network classifier type tools are directly comparable to logistic regression tools), this article is intended to report our experience in applying these methods to date and to promote these procedures in data analysis of biostatistical information.

Acknowledgment

The authors would like to thank Ron Landes of Landes Bioscience for encouraging us to publish this article.

References

1. Pelletier RP, Cosio FG, Henry ML, Bumgardner GL, Davies EA, Elkhammas EA, et al. Acute rejection following renal transplantation: evidence that severity is the best predictor of subsequent graft survival time. Clin Transpl 1998;12:543-52.
2. Pelletier RP, Diderrich G. Study of long-term survival in kidney transplant patients utilizing a multi-tool statistical, neural network and intuitive scoring function approach. Working Draft, April, 2002.
3. Pelletier RP, Hennessy PK, Adams PW, VanBuskirk AM, Ferguson RM, Orosz CG. Clinical significance of MHC-reactive alloantibodies that develop after kidney or kidney-pancreas transplantation. Am J Transpl 2002;2:134-41.
4. Pelletier RP, Orosz CG, Cosio FG, Ferguson RM. Risk factors in chronic rejection. Curr Opin Org Transpl 1999;4:28-34.
5. Diderrich G. Theoretical and practical techniques in applying neural network technology to data mining applications. 2002. Available: www.geocities.com/gtdiderrich/.
6. Klein C, Weyer B, Jones J, Diderrich G. Taking manufacturing plant financial simulation to the bottom line using neural networks. Paper presented at the 4th Annual Simulations Conference, Institute of Industrial Engineers, April 7, 2002.
7. Statistical neural networks manual. 1998. Available: www.stasoft.com.
8. Ward system neural network classifier package. 2000. Available: http://www.wardsystems.com.
9. Klein JP, Moeschberger ML. Survival theory: techniques for censored and truncated data, statistics for biology and health. New York, NY: Springer-Verlag; 1997.
10. Therneau TM, Grambsh PM. Modeling survival data: extending the cox model, statistics for biology and health. New York, NY: Springer-Verlag; 2000.
11. Bishop CM. Neural networks for pattern recognition. New York, NY: Oxford University Press; 1995.
12. Neal RM. Survival analysis using a bayesian neural network, joint statistical meeting of ASA. August 2001. Available: http://ftp.cs.utoronto.ca/pub/radford/nnsv-talk.pdf.
13. Hosmer DW, Lemeshow S. Applied logistic regression. New York, NY: John Wiley & Sons; 1989.
14. Liestol K, Andersen PK, Andersen U. Survival analysis and neural nets. Statistical research report no. 93/2, University of Copenhagen; 1993.
15. JMP statistical manuals. Cary, NC: SAS Institute; 2000.
16. Yen J, Langari R. Fuzzy logic: intelligence, control and information. Upper Saddle River, NJ: Prentice Hall; 1999.