2009 International Conference on Future Computer and Communication

Comparison of Artificial Neural Networks with Logistic Regression in Prediction of Kidney Transplant Outcomes

Fariba Shadabi
Faculty of Information Sciences and Engineering
University of Canberra, ACT, 2601, Australia
Fariba.Shadabi@canberra.edu.au

Dharmendra Sharma
Faculty of Information Sciences and Engineering
University of Canberra, ACT, 2601, Australia
Dharmendra.Sharma@canberra.edu.au

Abstract— Predicting the outcome of a graft transplant with high level of accuracy is a challenging task. To answer the challenge, data mining can play a significant role. The goal of this study is to compare the performances and features of an Artificially Intelligent (AI)-based data mining technique namely Artificial Neural Network with Logistic Regression as a standard statistical data mining method to predict the outcome of kidney transplants over a 2-year horizon. The methodology employed utilizes a dataset made available to us from a kidney transplant database. The dataset embodies a number of important properties, which make it a good starting point for the purpose of this research. Results reveal that in most cases, the neural network technique outperforms logistic regression. This study highlights that in some situations, different techniques can potentially be integrated to improve the accuracy of predictions.

Keywords-Logistice Regression, Neural Network

I. INTRODUCTION

Data mining techniques can be employed to support clinical data analysis. The logistic regression model is a statistical data mining method, which has been used widely by researchers in medicine for many years [1]. Logistic regression is popular mainly because it enables the researcher to avoid the need for many precise assumptions required by other regression methods. Recently AI-based data mining techniques such as Artificial Neural Networks (ANNs) have drawn the attention of computer scientists and clinicians for intelligent information retrieval from clinical data sources [2], [3].

Both ANN and logistic regression techniques have the ability to model non-linear relationships between dependent and independent variables. However research shows that with the growing power of ANN tools, ANN can often be a good analytical alternative to logistic regression techniques [2], [4].

In this paper we discuss our experience of applying data mining techniques for the purpose of predicting the outcome of medical procedures or events. The case study described in this paper is from the kidney transplantation domain. In earlier works, we addressed some of the practical issues associated with the use of ANN in the prediction of kidney transplant outcomes [5], [6]. In this paper we compare the performances and features of an ANN approach to logistic regression in predicting renal transplantation outcomes.

II. RENAL TRANSPLANT CHALLENGES

The first successful human organ transplantation was carried out on December 1954 by Dr. Joseph Murray, in Brigham hospital, Boston. In this instance Richard Herrick from Northboro, Massachusetts was given a kidney from his healthy identical twin brother, Ronald. He survived another eight years before the original medical condition destroyed his new organ.

With the advent of more sophisticated anti-rejection drugs and antibiotics, organ transplants from related or unrelated donors have become much more common. Over the last five decades, thousands of kidney and other vital organ grafts such as heart and liver transplants have been performed successfully by surgeons around the world.

Although anti-rejection drugs have helped to boost the success rate of transplants, the biggest challenge that continues to face transplant patients and surgeons is the risk of rejection of transplanted organs. Over the years, there has been substantial research into methods to predict graft outcomes and identify key parameters influencing the success or failure of transplanted organs [7]. Successful kidney transplantation will not only extend the longevity and quality of life for the recipient but also reduce medical expenses and increase the access to donor kidneys by reducing the need for multiple kidney transplants in the one patient.

Until now most clinical prediction methods have largely been focused upon the use of standard statistical models. However, statistical techniques often do not provide

978-0-7695-3591-3/09 $25.00 © 2009 IEEE
DOI 10.1109/ICFCC.2009.139

adequate knowledge for solving highly complex clinical prediction problems. ANN have the ability to provide good solutions in situations where large number of variables contribute to an outcome but their individual influence is weak and not well understood. Clinical data gathered from patients who have undergone graft transplant surgery have this characteristic and are known to be complex [8], [9].

In this paper we compare a widely used ANN approach known as Multi-layer Perceptron (MLP) networks with logistic regression, with the challenge being to select the right kidney from the available pool of organs for a particular patient, thereby maximizing the chances for the successful transplantation.

III. MATERIAL AND METHODS

A. Dataset

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this template and download the file for A4 paper format called "CPS_A4_format".

B. Artificial Neural Networks and logistic regression

The inspiration of ANN came from the desire to simulate features of the brain, namely biological neural networks and learning systems, which show high power in pattern recognition tasks and adaptability. ANN architectures are generally divided into two categories namely feed-forward network (networks without any loops in their path) and feedback networks (networks with recursive loops). Different network architectures and training algorithms can affect networks capabilities and usually a lot of trial and error experimentation is necessary to determine the optimal network topologies and training parameters. Detailed information about the foundations of ANNs can be found in [12], [13]. For the purpose of this study, multilayer, feed-forward networks were used to differentiate between successful and unsuccessful transplants. All neural network classifiers described in this paper were implemented using the Delphi programming language.

Binomial (or binary) logistic regression is an effective supervised learning method that can be used to estimate the probability of a certain event occurring.

C. Bagging

An ensemble consists of several individually trained classifiers that are jointly used to solve a problem. The most popular methods for generating training sets for classifiers are Bagging and Boosting.

Bagging (bootstrapping aggregates) was originally proposed by Breiman [15]. Bagging is a popular method for training component classifiers. This technique generates several training sets using random drawing (with replacement) from the original training set. Consequently, in every new training set there are data points, which appear more than once while others, do not appear at all. Each individual classifier is trained with each of the training sets. As a result each classifier may produce a different prediction [16]. Bagging could offer a significant improvement in prediction accuracy. It is especially useful for a classifier with poor performance, or where a classifier has been presented with a small training sample set, or where small changes in the data can result in large changes in the classifier predictions (low stability issue). The pseudo-code of a classifier ensemble with bagging is shown in Table 1.

Table 1. The bagging approach

```
Given: Training set S (with n cases and their classes), learning algorithm L, number of bootstrap samples T

Process: 1.Model Generation

    for i = 1 to T:

        Generate a new training set(a bootstrap
        sample) with n cases, using random
        drawing (with replacement) from S

Apply the learning algorithm L to the bootstrap sample

keep the resulting model M(i) for future use

2.Prediction for a given test case

    for i = 1 to T:

        Predict class of case using M(i)

    Return class that appears most frequently
```

IV. METHODOLOGY

Before The following methodology was employed:

1. Pre-process the data set. This includes: extracting the data from different tables, cleaning the data, transforming the nominal attributes into numeric attributes, and choosing the appropriate parameters to be included in the dataset with the help of domain expert.
2. Split the dataset for training and testing (with balanced distribution of success and failure cases).
3. Perform classification using the MLP network and MLP network coupled with bagging.
4. Perform classification using the logistic regression and logistic regression coupled with bagging.
5. Assess and compare the predictive accuracy of the classifiers.

We conducted two main experiments:

Experiment 1. Neural Networks

- A single MLP classifier: Here we constructed a single MLP classifier. This experiment is reported in more details in our previous study [5].
- Neural network ensemble: In this experiment, we employed the bagging strategy to generate different

training sets from the original training set, using random drawing with replacement technique. Here we constructed 500 MLP classifiers. Each ANN classifier (network 1 through network 500) was trained independently of the others with each of the training sets, to differentiate between successful and unsuccessful transplants.

Experiment 2. Logistic Regression

- A simple logistic regression: Here a simple logistic regression classifier was used to model the relationship between a binary dependent variable (with values success/failure) and independent variables.
- Logistic regression with bagging: This is the same as previous experiment, however in this experiment, we implemented 500 classifiers, using bagging technique

V. Experimental Results

A. Experiment 1- Neural Networks

In this experiment the ANN classifier was used to predict the outcome of kidney transplants. All networks consisted of a set of 16 input neurons and 2 output neurons. Sigmoid transfer functions were used for both hidden layer and output layer. The training algorithm uses a trial and error approach to determine the appropriate neural network topology (i.e. the best training constant, number of epochs and hidden nodes) automatically.

As described in our previous paper on the validation phase, the prediction accuracy for our best feed-forward, back-propagation ANN with 16-input neuron and 2-output neuron reached only around 62% [5]. The positive predictive power of a single Multilayer Perceptron (MLP) network for prediction of outcomes of kidney transplants in our previous study was low, indicating a need for improvement.

In our next experiment, a series (T=500) of MLP networks were trained independently of the others, to differentiate between successful and unsuccessful transplants. In this experiment rather than reporting predictive accuracy alone to show best model choice, we modified the program in order to show the input patterns (examples) that were included across the ANN series in the final results. Patterns (examples) that consistently were in agreement across the classifiers can be considered as examples with positive impacts or higher predictive sources (for a detailed discussion, see [6]).

In this experiment, the balanced test set reached 70% accuracy rate with 87%-agreement among the networks (435 of 500 networks), based on 19% of data points. The results are shown in Fig. 1. This model was able to classify about 87% of successful transplants and 54% of unsuccessful cases. It should be mentioned that the accuracy rate also reached 76% with 89%-agreement among the networks (446 of 500 networks), using only 63 examples (14% of data points). The results revealed that in this case, 500 bagging reduced the variability of neural networks classifiers and improved their performance.

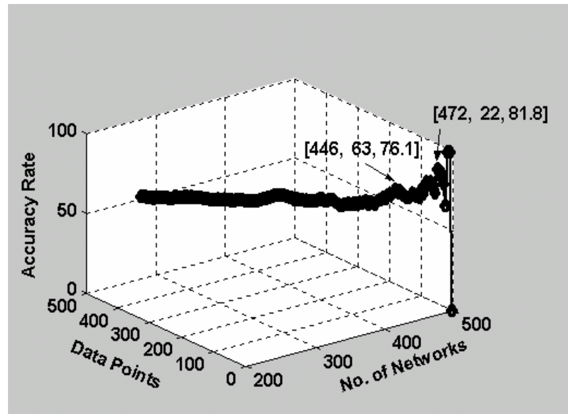

Figure 1. The results for neural network ensemble with 500 bagging

B. Experiment 2- Logistic Regression

In this experiment the logistic prediction model was used to predict the outcome of kidney transplants. We used the logistic function implemented in WEKA (version 3.4) [11]. This is a variation of an ordinary logistic regression model with ridge estimator that is known to be useful where a classifier has been presented with a small training sample set and can be used to improve the stability of classifiers [17]. Research shows that ridge estimators have been implemented successfully in logistic regression models by many researchers [17], [18]. The odds ratio, intercept and coefficients of the model are obtained by fitting the model to a set of training examples.

The dependent variable was the dichotomized transplant outcome (success vs. failure) obtained over a 2-year horizon. The independent variables that were used are listed in Table 1. An important point about the pre-processing step to note here is that missing values were replaced by a scheme that substitutes the mean for each missing value [11].

For the purpose of this study, we compared different data partitioning and models for their relative fit to the data. At one stage, we decided to test our model on a small balanced sample set using the same 84 examples (examples with positive impacts) that were consistently used across 435 ANN classifiers in our previous examples. In this case the model was able to predict only about 65% of successful transplants and 45% of unsuccessful cases, for an overall accuracy rate of 54%. As demonstrated in our previous experiment, using the same examples our ANN ensemble model achieved a 70% overall accuracy rate. However, by using a bigger balanced test set (439 examples), the best logistic regression model was able to classify about 54% of successful transplants and 73% of unsuccessful cases in the test set, for an overall accuracy rate of 64% (see also Fig. 2). It should be mentioned that logistic regression functions, especially when reasonably large data sets are provided for the training, are generally stable enough and usually do not benefit from bagging [19], [20]. However, to make a better comparison, we trained a logistic regression classifier using 500 bagging. Not surprisingly, the results revealed that bagging did not improve the classifier performance.

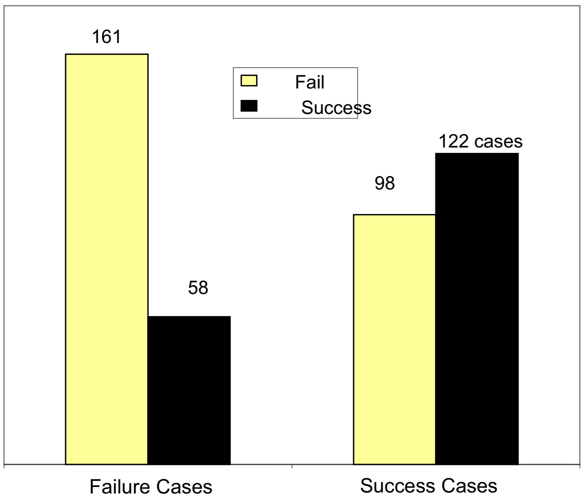

Figure 2. The histogram of confusion matrix for the best logistic regression model

VI. Discussion

This is not the first paper that investigates the performance of logistic regression and neural networks in clinical decision making. In fact, logistic regression and ANN approaches have been compared by many researchers, with mixed results [21], [22]. Though not conclusive, these studies seem to suggest that choosing the right technique is highly dependent on the underlying characteristics of the dataset of interest and the problem under investigation. Therefore, additional observational studies can identify certain medical domains that can benefit from the involvement of these techniques.

For the purpose of this study, the success of ANN and logistic regression models in graft outcomes prediction was assessed by looking at classification tables corresponding to each model. Classification tables show correct and incorrect classifications of the dichotomous dependent. The results revealed that the logistic regression approach provided better interpretability. However, overall, logistic regression models, regardless of training strategy, produced lower predictive accuracy than ANN.

Our observations suggest that where a model is presented with complex, incomplete and noisy data sets, an ANN ensemble approach provides a tighter model fit than logistic regression. Perhaps with modern modelling and transformation techniques, logistic regression models can also be used to adequately model complex set of relationships that are evident in many clinical data and uncover intricate structures. However, this requires developers with more advanced statistical knowledge.

It is important to keep in mind that although the developers of ANN models require less formal training, they still need to make some important design decisions in order to avoid over-fitting and thereby generate an accurate solution. Such decisions include choice of training data, training method, number of layers and number of processing elements in each layer [22], [23]. The good news is that the latest advances in ANN technology address some of these issues. Modern ANN packages[1] provide an easy-to-use interface that can be used to run an optimization process and find the appropriate ANN topology automatically.

VII. Conclusion and Further Research

In this study we compared features of logistic regression and ANN. We conducted an experiment on graft outcomes prediction using a kidney transplant dataset. Generally, for both techniques some careful pre-screening and calibrations are required. The results shown in this paper reveal that ANN coupled with bagging is an effective data mining method for predicting kidney graft outcomes. This confirms that different techniques can potentially be integrated to obtain a better prediction. Overall, the results reveal that in most cases, the ANN technique outperforms logistic regression.

A limitation of the ANN approach is that the way predictions are produced is not obvious. Unlike logistic regression models, the odds ratio and coefficients corresponding to each independent variable cannot be obtained. Due to these limitations, ANN models are not widely used by medical professionals. Our future research will involve studying the contribution of individual variables, by extracting and interpreting the weights corresponding to each independent variable, generated by the ANN model.

References

[1] Bagley, S., White, H., Golomb, B.: Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. Journal of ClinEpidemiol, vol. 54, No. 10, (2001) 979-85

[2] Sheppard, D., McPhee, D., Darke, C., Shrethra, B., Moore, R., Jurewitz, A., Gray, A.: Predicting Cytomegalovirus disease after renal transplantation: an artificial neural network approach. International Journal of Medical Informatics, Vol. 54, (1999) 55-76

[1] http://www.ltn.net/T/Computers/Artificial_Intelligence/Neural_Networks/Companies/

[3] Michie, D., Spiegelhalter, D.J, Taylor, C.: Machine learning neural nets and statistical classification. Ellis-Horwood, Chichester, (1994)

[4] Astion, M.L. and Wilding, P.: The application of backpropagation neural networks to problems in pathology and laboratory medicine. Arch. Pathol. Lab. Med. Vol. 116 (1992) 995–1001.

[5] Shadabi, F., Cox, R., Sharma, D., Petrovsky, N.: Use of Artificial Neural Networks in the Prediction of Kidney Transplant Outcomes. KES'04, the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (2004)

[6] Shadabi, F., Cox, R., Sharma, D., Petrovsky, N.: Experiments with A Neural Network Ensemble to Predict Renal Transplantation Outcomes. AISAT 04 Proceeding, the 2nd International Conference On Artificial Intelligence Science and Technology, November (2004) 271-276

[7] Rapaport F.T.: The current status of the HLA controversy in clinical transplantation. Transplant. Proc. Vol.27, No. 1, (1995)

[8] Doyle H., et al.: Predicting outcomes after liver transplantation. A connectionist approach. Ann. Surg. Vol. 219, Issue 4, (1994) 408-415

[9] Matis, S., Doyle, H., Marino, I., Murad, R, Uberbacher, E.: Use of Neural Networks for Prediction of Graft Failure following Liver Transplantation. Proceeding of the Eight Annual IEEE Symposium on Computer-Based Medical Systems (1995) 0133

[10] Data Dictionary: ANZDATA Registry Database (2000)

[11] Witten, I.H. and Frank, E.: Data mining: Practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, San Francisco (1999)

[12] Simpson, P. K.: Foundation of Neural Networks. Artificial Neural Networks-Paradigms, Applications, and hardware implementation, IEEE Press (1992) ISBN 0-87942-289-0

[13] Lisboa, P.J.G.: A review of evidence of health benefit from artificial neural networks in medical intervention. Neural Networks, Vol.15 No.1, January (2002)11-39

[14] Hosmer, D.W. and Lemeshow, S.: Applied logistic regression. John Wiley, New York (1989)

[15] Brieman, L.: Bagging predictors. Machine Learning Vol. 24 Issue 2, (1996) 123-140

[16] Zhou, Z-H. , Wu, J., Tang, W. : Ensembling Neural Networks: Many Could Be Better Than All. Artificial Intelligence, Vol. 137 (2002) 239-263

[17] Le Cessie, S., Van Houwelingen, J.C.: Ridge Estimators in Logistic Regression. Applied Statistics, Vol. 41, No. 1, (1992)191-201

[18] Forman, G. and Cohen, I.: Learning from Little : Comparison of Classifiers Given Little Training. 15th ECML/PKDD, Pisa Italy, September (2004)

[19] Perlich, C., Provost, F., Simonoff, J.S.: Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. Journal of Machine Learning Research, Vol. 4 (2003) 211-255

[20] Bühlman, P. and Yu, B.: Analysing Bagging. Annals of Statistics, Vol. 30 (2002) 927-961

[21] Lette, J., Colletti, B.W., Cerino, M., et al.: Artificial intelligence versus logistic regression statistical modeling to predict cardiac complications after non-cardiac surgery. Clin. Cardiol, Vol. 17 (1994) 609–614

[22] Tu, J.V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of Clinical Epidemiology, Vol. 49, Issue 11, November (1996) 1225-1231

[23] Cox, R., Crowther, P., Lisle, I.: An Empirical Investigation into the Error Characteristic of Neural Networks. AISAT 04 Proceeding, the 2nd International Conference on Artificial Intelligence Science and Technology, November (2004) 92-97