Predicting Hospital Length of Stay with Neural Networks

Steven Walczak
University of South Alabama
Computer & Information Sciences Dept.
FCW 20
Mobile, AL 36688-0002
walczak@cis.usouthal.edu

Ronald J. Scorpio
Division of Pediatric Surgery
Floating Hospital for Children
Suite 207, 750 Washington St.
Boston, MA 02111
ronald.scorpio@es.nemc.org

Walter E. Pofahl
University of South Alabama
Department of Surgery
3401 Medical Park Drive, Building 1, Suite 107
Mobile, AL 36693
wpofhal@usamail.usouthal.edu

Abstract

Critical care providers are faced with resource shortages and must find ways to effectively plan their resource utilization. Neural networks provide a new method for evaluating trauma patient (and other medical patient) level of illness and accurately predicting a patient's length of stay at the critical care facility. Backpropagation, radial-basis-function, and fuzzy ARTMAP neural networks are implemented to determine the applicability of neural networks for predicting either injury severity or length of stay (or both). Neural networks perform well on this medical domain problem. The backpropagation networks achieved the best performance for predicting a patient's length of stay, but the fuzzy ARTMAP produced superior performance in evaluating patient's level of injury (especially for the more severely injured patients). Thus a combination of backpropagation and fuzzy ARTMAP neural networks is recommended to produce the optimal combined (injury severity and length of stay) results.

Introduction

Hospital are faced with severely limited resources including beds to hold admitted patients. This resource constraint is particularly important in specialized areas of the hospital, such as intensive care units (ICU) or step down units. Evaluating length of stay (LOS) information is a challenging task (Weissman 1997), but is essential for the operational success of a hospital. Intensive care resources in particular are often limited and pose scheduling problems for hospital staff and administrators (Tu and Guerriere 1993). Predicting LOS is difficult and is often only done retrospectively. There are different levels in a patient's typical LOS. Lengths of stay may be evaluated for ICU, step down unit, and floor individually, and may also be evaluated as an aggregation of all areas to provide a total LOS. Existing research in this area looks at predicting or analyzing the LOS for patients in a specific hospital area (e.g., ICU). Our research will initially be concerned with ICU LOS, but will ultimately address predicting the LOS for a patient in all areas of the hospital and providing a total LOS.

A patient's LOS is highly correlated with that patient's injury or illness severity. Establishing a patient's injury severity may require extensive and invasive testing of the patient. Earlier identification of injury severity, and consequently LOS, would enable better resource planning and ultimately improve patient care by the hospital. The correlation between injury severity and LOS indicates that if characteristic variables of injury severity can be identified, they can be used to model LOS. Since the relationship between the characteristic variables and injury severity (and consequently LOS) is unknown, a nonparametric modeling technique is desired. Additionally, the modeling of injury severity or LOS appears to be a categorization problem. Neural networks are an ideal tool for creating nonparametric categorization models.

Hence, the research to be discussed in this article utilizes a supervised learning neural network approach to create a model that can accurately predict the LOS for pediatric trauma patients. Data for the pediatric trauma patients is taken from the National Pediatric Trauma Registry, which currently contains over 8000 records of patients that have been treated at a participating hospital's emergency room. The research will attempt to confirm the hypothesis that information which is available within the first ten minutes (without the use of invasive tests) of the patient's arrival at an emergency room can be used to accurately predict a trauma patient's LOS and the severity of the patient's injuries.

Background and Significance

Neural network applications in medicine have been primarily limited to laboratory applications (e.g., test evaluation) or medical imaging applications (Dybowski and Gant 1995, Montague and Morris 1994). Recently a few applications of neural networks have been made in the area of patient planning and medical resource allocation for patient care.

Previous studies that examine the use of neural networks for predicting patient outcomes have been limited to only predicting the occurrence of mortality (Dombi et al. 1995, Frye et al. 1996, Izenberg et al. 1997, Mobley et al. 1995). Most of these neural network predictors of mortality are specific to a particular type of injury, such as rib fracture

(Dombi et al. 1995), burns (Frye et al. 1996), or post-coronary care (Mobley et al. 1995). Our research will significantly extend this type of neural network research by categorizing those patients who will not die into four distinct groups: critical, severe, moderate, and low levels of illness or injury.

A few attempts have been made to use neural networks to predict a patient's hospital or ICU LOS. These attempts all have significant drawbacks. Most neural network LOS predictor models only predict a broad category for the LOS (Frye et al. 1996, Izenberg et al. 1997, Lowell and Davis 1994), such as less than seven days or greater than seven days, and all existing neural network attempts to predict patient LOS work only for a specific type of injury. Additionally, the only neural network attempt to predict the precise number of days for the LOS (Mobley et al. 1995) as well as others (Lowell and Davis 1994, Tu and Guerriere 1993) required extensive testing and pre-evaluation of the patient to acquire the input values used in the neural network model. Our methodology discussed below, uses values that are readily available within the first ten minutes of a patient's presentation to an emergency room. Furthermore, our system will predict the actual number of days (as opposed to a weekly or other LOS categorization value) for all patients who are treated in an emergency room. This approach is more generalized than previous approaches and supplies a more accurate determination of both injury severity and LOS with readily available data.

All of the referenced neural network applications have been developed using the neural network training technique of backpropagation. While backpropagation is a well known and widely used technique (Medsker and Liebowitz 1994, Alpsan et al. 1995) many other neural network training methodologies exist and may be more appropriate for medical domain applications. Our research will use various supervised-learning neural network training methods, including backpropagation, to determine the best possible modeling paradigm for this medical domain problem.

Neural Network Model for LOS Prediction

The National Pediatric Trauma registry provides 371 data values concerning the complete care of all pediatric patients that arrive at the emergency room of a participating hospital. The first step in the development of a neural network problem solution is to identify the independent variables (Tahai et al. 1998). Many of the data values are eliminated immediately because they are not available within ten minutes of the patient's arrival. The remaining values are screened using experts in the domain to determine all variables that could in any way contribute to a prediction of injury severity or LOS. Nineteen variables are identified as potential characteristic variables. Six of these variables are condensed into three because of partial overlap of information. Two of the variables (type of injury and mechanism of injury) are categorical and required three and sixteen neural network variables respectively to encode the desired information. Finally, two of the remaining variables are rejected due to statistical insignificance in the data distribution (not enough examples to train and test). Resulting in a 31 value input vector to the neural network, representing 14 independent variables.

One and two hidden layer architectures with varying quantities of hidden nodes per layer are implemented to minimize the problems of over-fitting and under-fitting the data (Barnard and Wessels 1992)(only those networks that produced the best generalization results for a particular training paradigm are reported in this article). Backpropagation, radial basis function, and fuzzy ARTMAP are the training methods used to optimize the neural networks. Radial basis function neural networks are selected as an alternative training paradigm due to their superior performance over backpropagation networks when extrapolation from the training population is required (Barnard and Wessels 1992). Fuzzy ARTMAP (Carpenter et al. 1995) neural networks are also implemented to evaluate the performance of adaptive resonance theory learning paradigms and because of their reliable performance when operating in noisy domains.

The test results of these three algorithms indicate that for similarly sized networks, the backpropagation function performs best for predicting LOS of pediatric trauma patients and the fuzzy ARTMAP paradigm works best for categorizing trauma injury severity. Why two different training algorithms each optimize on a different output set is a topic for future research.

Results

While all three neural network training methods categorize patient mortality with 80-88 percent accuracy and 100 percent sensitivity, the presence of these fatal patients decreases the accuracy of the LOS predictions. This is due to the fact that critically ill patients who die, normally die within the first 36 hours, while those that survive have extended hospital stays. The dichotomous nature of patient's who die versus those whom are near death, leads us to recommend that a fuzzy ARTMAP system that only predicts patient severity be used in combination with a second backpropagation neural network that predicts LOS. The backpropagation neural network will have patients that die in the hospital removed from the training set, to improve accuracy of the LOS predictions.

As mentioned earlier, the Registry contains data on over 8000 pediatric trauma cases. To simulate a prospective use of the neural network, the data is divided into training and test sets based upon date. All records recording injuries occurring prior to June 30, 1996, approximately 4200

Table 1: Backpropagation prediction of LOS.

Backpropagation Architecture	Minimum Prediction (Actual)	Maximum Prediction (Actual)	ICU LOS Mean Absolute Difference (days)	Total LOS M.A.D. (days)
Single hidden layer				
(25 nodes)	0 (0)	55 (63)	**1.98**	2.81
(30 nodes)			2.09	**2.79**
Two hidden layer				
(30 nodes, & 10 nodes)	0 (0)	54 (63)	2.05	2.92
(30 nodes, & 20 nodes)			2.13	2.82

examples, are in the training set, while the remaining 3881 cases form the test set. Results for the backpropagation neural network predicting LOS values for ICU and total LOS are displayed in Table 1. The minimum and maximum stay predictions for the networks along with the mean absolute difference (MAD) between the predictions and the actual LOS is shown. Furthermore, 52 percent of the 3881 test cases have a neural network LOS prediction that is within one day of the actual LOS for that patient. The fuzzy ARTMAP network achieved a total LOS MAD value of 4.2, while the radial basis function network achieved a total LOS MAD value of 3.6.

However, the fuzzy ARTMAP did provide the best performance for predicting the patient's injury severity. The results for number of correctly classified patients for injury severity predictions are shown in Table 2. Unfortunately, there is only one example of a low severity patient in the test set, so this case is not shown. The distribution of injury severity in the test set population is 43% critical, 5% severe, 51% moderate, with the remainder being either low or mortality patients. Assuming that moderate patients, those discharged to a floor bed, have an innately higher recuperation rate than those classified as critical (discharged directly to an operating room or to an ICU bed) or severe, then the better performance of the fuzzy ARTMAP on these categories of patients provides the hospital with better patient care planning and ultimately with better patient recovery.

Table 2: Prediction of injury severity.

Training Algorithm	Critical (ICU)	Severe (Step-down)	Moderate (Floor)
Fuzzy ARTMAP	**59%**	**15%**	36%
Backpropagation	44%	0%	93%

Extending LOS Neural Network Research

Additional research has been performed using the same methodology as previously discussed, except for patients with acute pancreatitis. Acute pancreatitis patients are initially treated in a hospital's emergency room (ER), but are not considered as trauma patients. The neural networks for predicting the LOS for pancreatitis patients use patient information values that are available at the time of admission to the hospital or shortly thereafter (e.g., age, gender, heart rate, temperature, hemoglobin, hematocrit, and various other blood analysis results) as inputs to the network. The use of blood analysis delays the potential time of use for the neural network from the first ten minutes until one to two hours after ER admission.

Both a single hidden layer and two hidden layer backpropagation networks are implemented along with a radial basis function neural network and fuzzy ARTMAP neural networks. Two different input sets, one with 44 elements and one with 36 input elements are used for training the described neural networks. In addition to removing some input values that contained a lot of noise (i.e., missing values; which caused the elimination of training and test samples in the 44 input data sets), the 36 input value set also removed a RANSON score value. The RANSON score value is a heuristic method currently used to predict outcomes for pancreatitis patients (Ranson et al. 1974, Ranson 1982) including LOS. The neural networks utilizing the 36 input value vector consistently outperformed the 44 input value networks, indicating that the RANSON score may not be the best possible predictor of pancreatitis LOS. The reduction of required data elements also increased the number of data samples that contained all of the required data. The validation test set went from 56 cases for the 44 input variable model to 71 cases for the 36 input variable model, thus increasing the statistical significance of the research results.

The mean absolute difference between the neural network predictions and the actual patient LOS results for the various 36 input value neural networks are shown in Table 3 along with their LOS specificity scores. Specificity is defined as the accurate segmentation of the pancreatitis patient population into three categories of hospital stay: short (stays less than or equal to 1 week), medium (more than 1 week, but less than or equal to 2 weeks), and long (greater than 2 weeks). The LOS specificity for the pancreatitis domain is similar to the injury severity rating from the pediatric trauma domain, with the exception that a closeness criterion may be established for the LOS specificity values (whereas a patient is either admitted to ICU or is not admitted). For the single hidden layer back-propagation neural network, if a one-day overlap is per-mitted between the LOS categories, then the medium stay category accuracy increases to 50 percent. Furthermore, if a two-day overlap is used between categories, then the short stay accuracy increases to 74.5 percent and the medium stay accuracy increases to 75 percent, with no change in the long stay accuracy.

Table 3: NN pancreatitis LOS

Network Type	MAD (days)	Accuracy 1-7 days	Accuracy 8-14 days	Accuracy > 14 days
Backprop 1 H layer	5.6	62.75%	**33.33%**	**62.5%**
Backprop 2 H layers	4.88	25.00%	100%	0.0%
RBF	4.76	68.63%	16.67%	50%
Fuzzy ARTMAP	**4.18**	100%	0.0%	0.0%

The mean difference is greater for the pancreatitis neural networks than it is for the pediatric trauma neural networks. This is a reasonable result given the fact that the pediatric trauma patient test population has an average LOS of 3.92 days with a standard deviation (σ) of 5.5 days, while the pancreatitis test patient population has an average length of stay of 7.75 days with a σ of 8.9 days.

The fuzzy ARTMAP neural network produces the minimum difference, but has difficulty in recognizing patients that have a LOS of greater than one week. Almost 30 percent of the acute pancreatitis patients have a LOS of greater than one week. Again, a combination of two different types of neural networks appears to be warranted for the pancreatitis LOS domain. The fuzzy ARTMAP neural network should serve as the primary LOS predictor due to its minimal absolute average difference in days from actual LOSs, but should be supplemented with the single hidden layer backpropagation neural network to identify those patients that are likely to require extended care. While the same two types of neural networks are recommended for the pancreatitis LOS prediction problem as for the pediatric trauma LOS prediction problem (Backpropagation and fuzzy ARTMAP), the reason for using each type of neural network has been reversed.

Finally, the fuzzy ARTMAP pancreatitis LOS prediction network, accurately predicts the precise LOS for over 15 percent of the 76 pancreatitis patients and predicts within one day the LOS for just over 35 percent of the pancreatitis patients. When used in combination with the single hidden layer backpropagation neural network, the accuracy of the perfect LOS prediction increases to 17 percent of all patients.

Summary

The research reported in this article has shown that neural network systems are capable of providing significant planning and patient care information with a relative paucity of diagnostic knowledge. All knowledge used as input for the neural networks implemented and tested in the presented research is available within the first ten minutes of a patient's arrival at the hospital for the pediatric trauma patients and within the first two hours after arrival for the acute pancreatitis patients. For pediatric trauma patients who had an average LOS of 4 days (ranging from 0 to 63 days), a backpropagation neural network predicted within one day the LOS for 52 percent of the patients. Extending this research to patients suffering from acute pancreatitis who had an average LOS of almost 8 days (ranging from 1 to 58 days), produced a fuzzy ARTMAP neural network that is able to predict within one day the LOS for over 35 percent of the patients.

An unusual finding of this research is that a two network system combination of a single hidden layer backpropagation neural network and a fuzzy ARTMAP neural network system performs better than just a single neural network system for predicting both a patient's injury severity and LOS. The same combination of a backpropagation neural network and a fuzzy ARTMAP neural network produce a combined performance better than either network type operating alone for both domains investigated in the presented research. However, the fact that the type of neural network which operates best on predicting injury severity or LOS is switched between the pediatric trauma and pancreatitis domains may reveal an underlying difference in the medical states of these two types of patients. As stated earlier, future research is needed to determine why specific (and different) types of neural networks operate better on different aspects of the LOS/injury severity problem.

The research results presented for the pediatric trauma domain are generated using in-sample training data from numerous hospitals subscribing to the National Pediatric Trauma Registry. It may be possible to further increase the accuracy and sensitivity of the implemented neural networks by limiting the training samples to a more localized region with similar patient demographics. Walczak and Scharf (1998) have shown that for the domain of surgical transfusions, the introduction of data that originates outside of a specific hospital's region introduces noise into the neural network model and subsequently reduces the accuracy of the network predictions. Future research will investigate the differences caused by local versus regional versus national data sources for modeling medical domains with neural networks.

The research finding, that a reduction in the quantity of diagnostic values used for modeling acute pancreatitis patient LOS gives direction for future research to further improve the accuracy and specificity of the pancreatitis LOS prediction neural networks. Currently, research is progressing to examine the predictability of 16 and 17 input variable models.

Neural networks have been shown to provide an accurate and reliable means for estimating the LOS for patients suffering different medical problems (trauma, pancreatitis, etc.). Usage of neural network systems by hospitals to predict patient LOS enables better resource allocation and resource planning.

References

Alpsan, D., Towsey, M., Ozdamar, O., Tsoi, A.C., and Ghista, D.N. 1995. Efficacy of Modified Backpropagation and Optimisation Methods on a Real-world Medical Problem. *Neural Networks* 8(6):945-962.

Barnard, E. and Wessels, L. 1992. Extrapolation and Interpolation in Neural Network Classifiers. *IEEE Control Systems* 12:50-53.

Carpenter, G.A., Grossberg, S., and Reynolds, J.H. 1995. A Fuzzy ARTMAP Nonparametric Probability Estimator for Nonstationary Pattern Recognition Problems. *IEEE Transactions on Neural Networks* 6:1330-1336.

Dombi, G.W., Nandi, P., Saxe, J.M., Ledgerwood, A.M., and Lucas, C.E. 1995. Prediction of Rib Fracture Injury Outcome by an Artificial Neural Network. *Journal of Trauma: Injury, Infection, and Critical Care* 39(5):915-921.

Dybowski, R., and Gant, V. 1995. Artificial neural networks in pathology and medical laboratories. *Lancet* 346(November 4):1203-1207.

Frye, K.E., Izenberg, S.D., Williams, M.D., and Luterman, A. 1996. Simulated Biologic Intelligence Used to Predict Length of Stay and Survival of Burns. *Journal of Burn Care and Rehabilitation* 17(6):540-546.

Izenberg, S.D., Williams, M.D., and Luterman, A. 1997. Prediction of Trauma Mortality Using a Neural Network. *American Surgeon* 63(3):275-281.

Lowell, W.E. and Davis, G.E. 1994. Predicting Length of Stay for Psychiatric Diagnosis-related Groups Using Neural Networks. *Journal of the American Medical Informatics Association* 1(6):459-466.

Medsker, L. and Liebowitz, J. 1994. *Design and Development of Expert Systems and Neural Networks*, New York: Macmillan.

Mobley, B.A., Leasure, R., and Davidson, L. 1995. Artificial neural network predictions of lengths of stay on a post-coronary care unit. *Heart and Lung* 24(3):251-256.

Montague, G. and Morris, J. 1994. Neural-network contributions in biotechnology. *Trends in Biotechnology* 12(8): 312-324.

Ranson, J.H.C., Rifkind, K.M., Roses, D.F., Fink, S.D., Eng, K., and Spencer, F.C. 1974. Prognostic signs and the role of operative management in acute pancreatitis. *Surgery, Gynecology and Obstetrics* 139:69-81.

Ranson, J.H.C. 1982. Etiological and prognostic factors in human acute pancreatitis: a review. *American Journal of Gastroenterology* 77:633-638.

Tahai, A., Walczak, S., and Rigsby, J.T. 1998. Improving Artificial Neural Network Performance Through Input Variable Selection. *Applications of Fuzzy Sets and The Theory of Evidence to Accounting*, Forthcoming.

Tu, J.V. and Guerriere, M. 1993. Use of a Neural Network as a Predictive Instrument for Length of Stay in the Intensive Care Unit Following Cardiac Surgery. *Computers and Biomedical Research* 26:220-229.

Walczak, S. and Scharf, J. 1998. Reducing Surgical Patient Costs Through Use of An Artificial Neural Network to Predict Transfusion Requirements. *Decision Support Systems*, Forthcoming.

Weissman, C. 1997. Analyzing intensive care unit length of stay data: Problems and possible solutions. *Critical Care Medicine* 25(9):1594-1599.