Predicting Injury Severity Outcome Based On Neural Networks Architecture

Thin De Aye
University of Computer Studies, Mandalay
thindiaye@cumandalay.edu.mm,

Htar Htar Aung
University of Computer Studies, Mandalay
htarhtaraung@cumandalay.edu.mm

Abstract

One of the major causes of unnatural loses of human beings all over the world in road traffic accident. The aim of the study is to apply Radial Basic Function (RBF) architecture of neural network for predicting injury severity scale. RBF network type models are applied to predict injury severity level of alcohol presence data from US traffic accident. This study acquired as categorical variables from the National Automotive Sampling System which is reported to police in the United States. This study reveals that extracted factors of alcohol presence data have the highest predictive power to explaining injury severity scale of vehicle crash.

Keyworks: ***RBF, Neural Network, Categorical***

1. Introduction

Motor vehicle traffic safety is one of the main priorities of governments in every country [1]. There were 29 million licensed drivers age 65 and older in the U.S. which are reported by National Highway Traffic Safety Administration in 2005 [6]. The speedy improvement of motorization has enlarged in the traffic-related casualties. The application of ANNs has been a lot of success in number of different areas of specialization, including transportation engineering. Neural network is based on the intelligent computation which uses the computer network system to simulate the biological neural networks.

The research of NN algorithm [5] based on factor analysis is relatively rare. The aim of this study is to understand how to investigate to analyze risk perceptions in traffic accidents. In this paper, the development of neural network model was carried out for the examination and prediction of accident rate using the United State as case study.

2. Proposed Framework

This aim of the study is to estimate the most relevance factors for different type of injury scale in roadway traffic accidents. Ten different situation of factors related injury scale are extracted from the NASS-CDS from the National Highway Traffic Safety Administration (NHTSA) [6]. The extracted data are converted to the categorical variables (nominal, ordinal and interval) for analysing in Neural Network. Afterwards, the converted factors are used to predict the injury scale of traffic accident by using radial basic function (RBF) of Neural Network technology. RBF network type models are applied to predict injury severity level of alcohol presence data from US traffic accident. The statistical distribution of the data does not need to be known when developing an ANN model [3]. The following figure 1 shows the procedure of proposed framework.

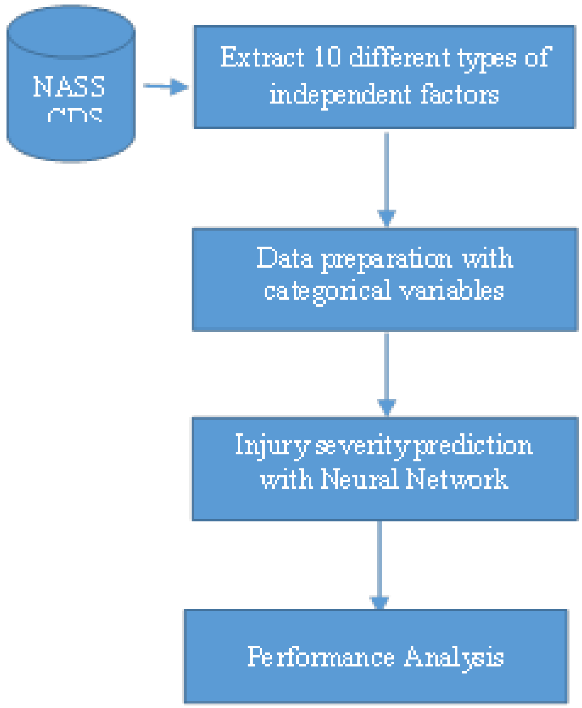

Figure 1. Proposed system architecture

The following section describes the related methodology and discusses the analytical results based on the proposed framework of the injury severity analysis for motor vehicle traffic accidents.

3. Related Methodology

Radial Basic Function (RBF) is a two layer feed-forward type network in which the input is transformed by the basics function at the hidden layer. At the output layer, linear combinations of the hidden layer node responses are added to form the output. The behaviour of any arbitrary function f(x) [2] is described in a compact area of the input space, by a combination of simpler functions

$$O_i\ (x) = y(!x-!x_i) \qquad (1)$$

Where y (.) is normally a Gaussian function. The appropriate to the function f(x) is given as $f(x,w)=?\ W_i\ G(!x-!x_i)$, where w_i are real value entries of the coefficient vector

$$W= [w_1, w_2, w_3,\ldots,w_n] \qquad (2)$$

F(x) being a real valued vector

$$X=[x_1, x_2, x_3,\ldots, x_n] \qquad (3)$$

Implements the input-output map of the NN RBF. Any arbitrary continuous function can be approximated with an NN RBFF if localized Gaussian are placed to cover the space, the width of each Gaussian is controlled, and the amplitude of each Gaussian is set.

4. Results and Discussion

This study aim to determine whether a RBF neural network can help instructors to correctly predict the injury scale (minor, major and severe). In this analysis, different types of explanatory factors are used to predict and classify the response variables. In the preprocessing stage, the extracted data are changed to categorical variables such as (ordinal, interval and nominal) data. Ordinal means good information about the level of lane. Interval scales give the order of values like a different type of age group. Nominal defines the pertaining to nmaes such as gender, seatbelt use or not and airbag deploy situation factors.

Table 1 show the number of neuron in every layer and 10 independent factors with different categorical variables. Automatic architecture selection chose 10 nodes for the hidden layer, while the output layer 3 nodes to code the response varibles. For the hidden layer the activation function was the standardized, while for the output layer used the softmax function. Cross entropy was used as error function beacause of use of softmax function. The network diagram that SPSS used to predict injury severity from NASS-CDs database, in figure 2. This figure shown in Appendix of this paper. The diagram describes the 25 input nodes, 10 nodes of hidden layer and 3 nodes for output layer.

Table 1. Network Information

Input Layer	Factors	1	LIGHT CONDITIONS	
		2	ROADWAY ALIGNMENT	
		3	ROADWAY SURFACE CONDITION	
		4	ROADWAY SURFACE TYPE	
		5	NUMBER OF LANES	
	Covariates	1	gender group	
		2	Deltav group	
		3	age group type	
		4	seat belt used or not	
		5	Air bag deployed or not	
	Number of Units			25
	Rescaling Method for Covariates		Standardized	
Hidden Layer	Number of Units			10[a]
	Activation Function		Softmax	
Output Layer	Dependent Variables	1	AIS group level	
	Number of Units			3
	Activation Function		Identity	
	Error Function		Sum of Squares	

a. Determined by the testing data criterion: The "best" number of hidden units is the one that yields the smallest error in the testing data.

Table 2. Model summary for injury severity prediction

Model Summary

Training	Sum of Squares Error	6.888
	Percent Incorrect Predictions	17.0%
	Training Time	0:00:00.06
Testing	Sum of Squares Error	6.188[a]
	Percent Incorrect Predictions	22.9%

Dependent Variable: AIS group level

a. The number of hidden units is determined by the testing data criterion: The "best" number of hidden units is the one that yields the smallest error in the testing data.

The model summary shown in Table 2 provides information related to the results of training and testing sample. Cross entropy error indicates the power of model to predict injury severity outcome. According to the table, the percentage of incorrect predictions based on training and testing sample respectively are 17% and 22.9%.

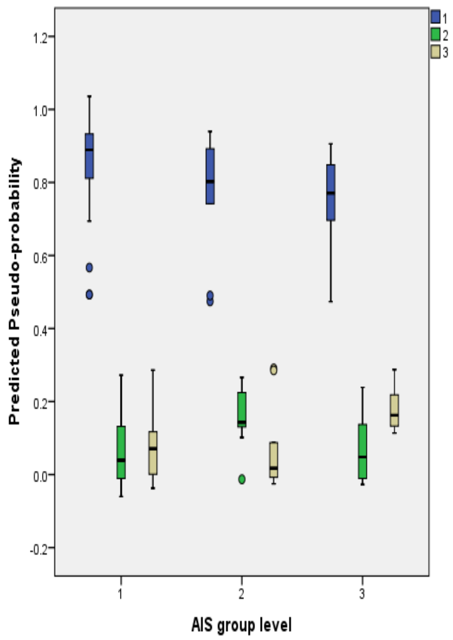

Figure 3. Predicted by observed chart

Box plots of predicted pseudo-probabilities are shown in figure 3. Pseudo probabilities are computed from the marginal survival function. Pseudo probabilities are conditionally independent and which do not need to consider the within subject correlation as in the neural network approach. For the response variables of injury severity scale, the chart describes box plots that classify the predicted probabilities based on the dataset. For each box plot, the value above 0.5 is correct predictions. From the left, boxplot shows the predicted probability of the observed AIS group 2 and 3 to be in group 1 category.

4.1. Performance Analysis

The ROC curve is a diagram of sensitivity versus specificity that shows the classification performance of all possible cutoffs. Based on the combined training and testing sample, figure 4 describes the sensitivity and specificity chart. The more the curves move away the 45 degree baseline, the more accurate is the prediction of traffic accident injury severity scale. According to the analysis, severe injury severity is around 81.6%, major injury is around 78.2% and minor injury is around 73.4% can be predicted based on the extracted 10 factors of vehicle crash.

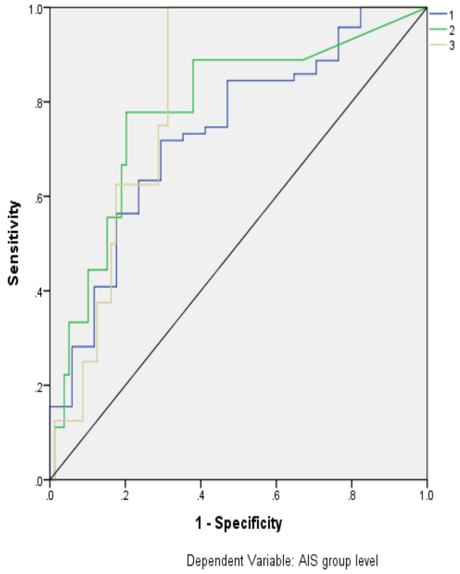

Figure 4. ROC Analysis

The chart in figure 5 shows the cumulative gain that is the presentence of correct classifications obtained by the RBF model of neural network. Above the baseline a curve lies which mean the greater the gain and indicate better performance. According to the analysis with environment factors, and demographic data, severe injury has better performance than other major and minor injuries.

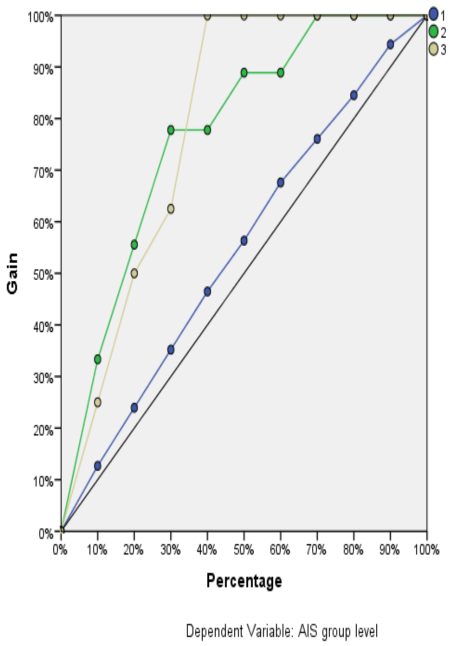

Figure 5. Cumulative Gain Chart

Table 3 and figure 6 explains the impact of each explanatory factor in neural network model in terms of relative and normalized importance which are calculated by using Gaussian funciton of neural network. This is used to find the relationship between the independent variable and dependent variables. DeltaV group is the most important factor effecting in traffic accident. In the environmental effect factors, number of lanes is the most experimental factor 43% for injury severity outcome. Roadway alignment factor is the second most important factor to expalin the injury outcome. Based on the factor analysis, roadway surface type is also a major determinant of model predictive power. In the US motor vehicle policies, there have a seatbelt usage law rather than other countries. Therefore, seatbelt usage is not emerge in the most relevance factor group for injury severity outcomes.

Table 3. Independent Varaible Analysis

Independent Variable Importance

	Importance	Normalized Importance
LIGHT CONDITIONS	.048	13.4%
ROADWAY ALIGNMENT	.105	29.7%
ROADWAY SURFACE CONDITION	.064	18.2%
ROADWAY SURFACE TYPE	.028	7.9%
NUMBER OF LANES	.154	43.3%
gender group	.057	16.1%
Deltav group	.355	100.0%
seat belt used or not	.042	11.9%
Air bag deployed or not	.058	16.2%
age group type	.089	25.0%

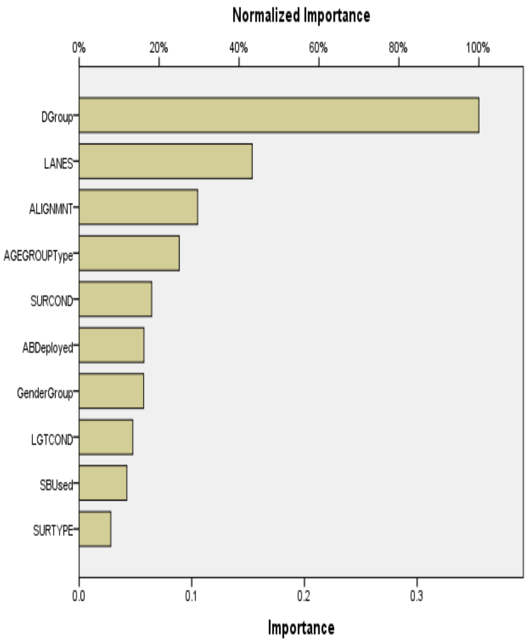

Figure 6. Bar chart analysis for important factors

5. Conclusion

The aim of the study was to evaluate the effectiveness of Radial Basic Function of neural networks in predicting the injury severity of traffic accident. A radial basic function neural network was trained by feed-forward algorithm, to predict and explain the injury outcome of motor vehicle crash of NASS-CDs database.

Appendix

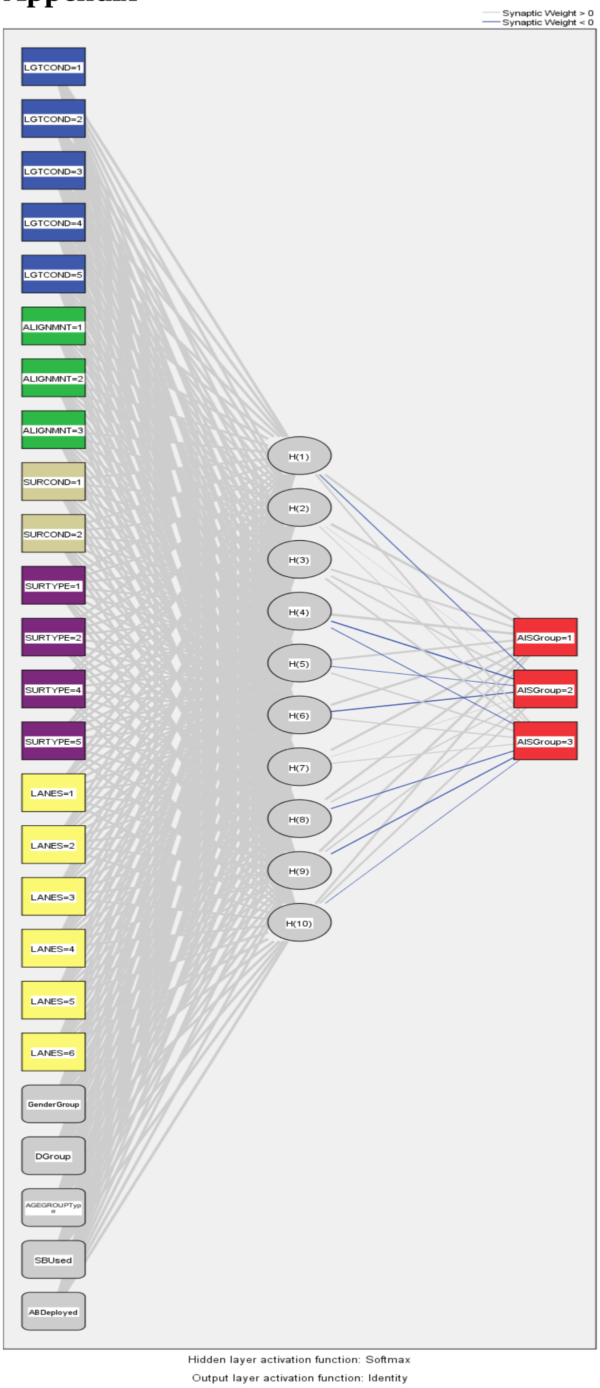

Figure 7. Network Diagram

References

[1] Abdelwahab, H.T. and Abdel-Aty, "M.A. Development of ANN Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections", Transport. Res.Rec., 1746, pp. 6-13, (2001)

[2] Adeolu J. O, "Road Accident Trends in Nigeria. Accident Analysis and Prevention", 25(4), pp.485-487, (1993)

[3] Delen, D., Sharda, R., and Bessonov, M., "Identifying Significant Predictors of Injury Severity in Traffic Accidents Using A Series of Artificial Neural Networks Accident Analysis & Prevention", 38 (3), pp. 434–444, (2006)

[4] Gopi, E.S., "Digital image forgery detection using artificial neural networks and independent component analysis", Applied Mathematics and Computation 2, 540–543 (2007)

[5] Mcclochw, S., "A logical calculus of the ideas immanent in nervous activity", Bulletin of Mathematical Biophysics 5, 115-133 (1943)

[6] National Highway Traffic Safety Administration, "National Automotive Sampling System (NASS) Crashworthiness Data System", Us Department of Transportation. Washington, DC, p.1.2014.