An Artificial Neural Network based Credit Limit Prediction System

M A H D P weerasinghe[2] and H L Premarathne[1]

[1]Department of Communication and Intelligence Systems, University of Colombo School of Computing, 35, Reid Avenue, Colombo 03

[2]Technology Division, Head office, Bank of Ceylon, BOC Square, Bank of Ceylon Mawatha, Colombo 01

ABSTRACT

The decision to issue a credit card to a specific customer or offer the credit limit is more critical as wrong decisions affect huge financial losses to the banks. The effectiveness of artificial neural network based models in predicting credit limit by based on customer specific parameters, is presented. The implemented networks were based on feed-forward back-propagation technique. The neural network models were taking the customer specific parameters and predict a credit limit range for each input. A sample of 4000 customer details was used to train the networks. By providing combinations of several customer ground level parameters, effectiveness of the artificial neural network models was evaluated. Among those parameters age, designation level, loan amount and salary were chosen as the best inputs to the system. One neural network model was developed to predict the exact credit limit while the other model was developed by including the fuzzy techniques to predict the credit limit ranges. First model was able to predict the credit limits with an accuracy of 64±3% while other model was able to predict the limits with an accuracy of 78±3%. By introducing the Fuzzy logic the accuracy of the predictions has been improved. It was found that neural network models give better results when considering the lower credit limit ranges. When predicting the vague higher values the network prediction was considerably low.

1.0 INTRODUCTION

Credit cards are one of the most popular forms of payment for consumer goods and services in Sri Lanka nowadays. Most of the leading banks and financial institutes are offering credit cards to customers, with different credit limits. Parameters such as age, designation, salary, account credentials and past loan information are some of the inputs which are considered in evaluation process of a customer. Most of the banks use manual judgments by analysing the credit card application form of the each customer separately while other banks use credit scoring systems or combination of both. These manual techniques are more time consuming and it requires more domain specialists to get the correct decisions. Lack of domain knowledge or misunderstanding can produce wrong conclusions. [1-3]

In general stochastic models or numerical models are used in credit scoring systems. Neural network based models have not yet been developed sufficiently to forecast the credit limit. In recent past, a few research studies have attempted to predict credit scoring systems using neural techniques.

One such research has been done for investigate the most suitable neural network model by comparing the five models such as multilayer perception, mixture-of-experts, radial basis function, learning vector quantization, and fuzzy adaptive resonance. Results demonstrated that the multilayer perception may not be the most accurate neural network model, and both mixture-of-experts and radial basis function neural network models should be considered for credit scoring applications. [4]

Another research has compared the results using a combination of activation and error functions which applied differently on the hidden and output layers of the neural network. Experimental results of this paper demonstrate that the neural network computed satisfactory results when Sigmoid – Sigmoid - Sinh Combination of activation and error function is used for hidden and output layers. When properly and sufficiently trained by applying appropriate activation and error functions, the neural network performs remarkably better than any other statistical approach, such as logistic regression or discriminant analysis. [5]

Another useful literature was found that it has introduced the variables those are most frequently used in credit card scoring systems. And the secondary purpose is to initiate the comparison of the neural networks performance with other widely used statistical methods. [6]

The main purpose of this work was to study how the customers' ground level parameters have influenced in credit limit and to measure the accuracy of such model.

2.0 METHODS AND MATERIALS

2.1 Data classification

The credit card data used in this study was obtained from one of the leading bank in Sri Lanka. The data sample consists of recently granted credit limits with their ground level parameters. Considering the knowledge of the domain specialists and from the referred literature; age, occupation, salary and the loan amounts were considered as input parameters to the network.

The data set contained about 5800 data. In the data set there were some missing data values and vague values. Also there were some extreme values for granted credit limits for some VIP persons. Therefore to get a correct sample of data, above data set was rectified and choose 4000 data sample as input to the neural network.

When considering the credit limits, figure 1 shows the distribution of the whole data set. It illustrate that credit limit distribution is not a normal curve and it is rightly skewed. The histogram shows outliers in the distribution. Those extreme value frequencies were very less and those credit limits have been removed from the data set to provide a generalized sample to the network. Figure 2 demonstrates the descriptive statistics of the distribution.

Occupation was categories in to 6 classes and illustrate in the following tale.

Level	*Description*
1	Top Management/ *Executives*
2	Senior Management
3	Professionally Qualified & experienced specialists/mid-management
4	Skilled Technical & Academically Qualified/ Junior Management/ Supervisors/ Foremen/ Superintendents
5	Semi-Skilled & discretionary decision-making
6	Unskilled & defined decision-making

Table 1: Occupation levels

Age in years was provided to the network and also total outstanding bank loan amount (including pawning, leasing, trade finance and other term loans) of each user was calculated as an inputs.

2.2 Neural network model

Several neural network architectures were tested and it was found that the 4-10-1 (input layer – hidden layer –output layer) architecture provides the best performance. Therefore, it was chosen to implement as the final model. A graphical interpretation of the network structure is represented in Figure 3.

The Dataset of 4000 records were divided into 3 categories and 70% of data (2800) samples were taken for training and 15% of data (600) were taken for validation and remaining 15% of data (600) were taken for testing. Number of hidden neurons was taken as 10 (default values). To implement the neural network "MATLAB fitting" tool was used.

First model was developed to predict the credit limit directly as the output. The targets which have to train the network were the exact credit limit of a customer. Rather than predicting exact credit limit, another model was developed with fuzzy logic to predict a credit limit range. As a decision support system the later model would help for managers to get the final decision based on the output.

Limits were divided in to 11 fuzzy classes in the fuzzy classification model. For simplicity, the trapezoidal membership function was applied for the fuzzy classification. The expressions for the 11 membership functions are shown in below table 2.

Fuzzy class	Credit Limit(Rs)
Class1	limit <= 20000
Class2	limit > 20000 and limit <= 25000
Class3	limit > 25000 and limit <= 30000
Class4	limit > 30000 and limit <= 40000
Class5	limit > 40000 and limit <= 50000
Class6	limit > 50000 and limit <= 70000
Class7	limit >70000 and limit <= 100000
Class8	limit >80000 and limit <=150000
Class9	limit >150000 and limit <=200000
Class10	limit >180000 and limit <=250000
Class11	limit >=250000

Table 8: Fuzzy membership functions

When applying the trapezoidal membership function to the limits, they were overlapped at the boundaries of some classes, yielding improved success rate in overall neural network.

If the network is over-trained it will be trained to the noise in the data set and not for the actual patterns. Therefore, the correct number of epochs has to be used during the training. A predetermined number of epochs were considered and the performance of the network was checked with an early stopping. This process was carried out several times. During the training period, a limit of 150 epochs and a mean square error of 10-3 were set.

The remaining 784 data set was considered to test the neural network and calculated the performance by comparing the actual class against predicted input.

3.0 EVALUATION OF THE RESULTS

Mainly two neural network models were considered.

1. Model without fuzzy classes
2. Model with fuzzy classes

3.1 Methods used to evaluate the success rate

$$\text{Success rate} = \frac{Xc}{Xtot} \qquad (1)$$

Xc – Number of correct predictions
Xtot – Total number of predictions

Root mean square error (RMS)

$$\sqrt{\left(\frac{1}{n}\right)\Sigma[(Xp)i - (Xe)i]^2} \qquad (2)$$

Xp – Predicted output
Xe– Expected output

3.2 Model without fuzzy classes

For this model actual credit limit was taken as targets. The neural network was trained according to that target and predicts an exact credit limit to a customer. This method did not show much accuracy when predicting the direct credit limit.

3.3 Model with fuzzy classes

Since the whole dataset contains 8 years of data, some data were found being obsolete and therefore when deciding the boundaries, the outliers were not considered. For this model the credit limit has divided in to 11 fuzzy classes. After training the network the predicted output class was compared with the actual credit limits. This method shows much more accurate result than previous model.

Comparison between the two models is shown in the following table.

	Success rate %	*Root mean square error*
Model without fuzzy classes	*64±3*	*3.45*
Model with fuzzy classes	*78±3*	*1.07*

Table 3: Comparison of ANN models

By looking at the above figures it can be observed that the performance of the fuzzy model is high. Since the limit ranges consist of overlapping in the fuzzy classes, it predicts more accurate result within a range as the output, opposed to a direct limit prediction.

Also the prediction success rate was found to be less in the 'model with fuzzy classes' for the limits above Rs. 200,000 compared to lower limits. There were few inputs for such values and the parameters were vague and not consistence. Classes in between Rs. 50000 to Rs. 200,000 got higher prediction success rates when compared to others.

Following table illustrate the computed accuracy rates of the fuzzy classes.

Fuzzy class	Credit Limit(Rs) (l – Limit)	Prediction success rate (±3)%
Class1	l < 20000	72
Class2	l > 20000 and l <= 25000	75
Class3	l > 25000 and l <= 30000	73
Class4	l > 30000 and l <= 40000	77
Class5	l > 40000 and l <= 50000	79
Class6	l > 50000 and l <= 70000	81
Class7	l >70000 and l <= 100000	90
Class8	l >80000 and l <=150000	88
Class9	l >150000 and l<=200000	81
Class10	l >180000 and l<=250000	73
Class11	l >=250000	69

Table 4: Prediction success rates of fuzzy classes

4.0 CONCLUSION AND FUTURE WORK

Credit limit itself is a highly variable event in its nature and when it needs to be predicted by only being based on the customer ground level information, it becomes an extremely difficult task. However, in this work it has been shown that neural network models with fuzzy logic can be applied with reasonable success rate in predicting the credit limit for a particular customer in credit card issuing.

The network predicted credit limits within Rs. 50000 to Rs. 200000 showed much more accurate results than other class ranges. In Sri Lankan economy frequency of granting credit limits that falls within this range is generally high. Therefore better prediction success rate has been achieved in that particular range.

The dataset had missing values as well as outliers. However, an unbiased data set was taken as input and the credit limit was predicted for 11 fuzzy classes.

Normally in a bank the ultimate decision for a credit limit is taken by a senior management position. This system is a supportive system for managers to get some reasonable idea on the specific credit limit that should be granted to a particular customer.

According to the studied literature, found no other neural network model predicts the credit limit based on customer ground level data. Most of the credits limit prediction systems were based on credit scoring or statistical analysis. The main objective of this work was to test the effectiveness of neural networks in predicting the credit limit. The results show that model without fuzzy logic was able to predict the limits with an accuracy of 64±3% while other model with fuzzy logic was able to predict the limits with an accuracy of 78±3%. Therefore it can be concluded that there is much more efficiency when using fuzziness in the input, similar to the manner in which the human mind analyzes the data to predict such results.

The system could be further enhanced in, overall performance improvement, usefulness and user friendliness features as being illustrated below.

Individual customer's total loan data are stored in CRIB (Credit Information Bureau) system. Hence the neural network model accuracy can be further improved, if it is possible to consider the total loans of each customer via the crib system.

More parameters can be considered than 'age, salary, occupation level and current loan amount' to improve the accuracy of the network. It's interesting to investigate the granting of credit limit range in precise when parameters like, 'Past loan and settlement history, current social background, family background and other risk factors such as gambling, political influences' are been considered.

System can be enhanced to handle not only the credit card centre but also the whole bank when granting a loan or any other advance to a customer. This can be considered as the next stage of this project.

5.0 ACKNOWLEDGMENT

Assistance by the University Of Colombo School Of Computing is acknowledged.

6.0 REFERENCES

[1] Credit Scores - MyFICO [Online]. Available

www.myfico.com/crediteducation/creditscores.aspx

[2] Benefits and Pitfalls of Statistical Credit Scoring for Microfinance [Online]. Available

http://www.microfinance.com/English/Papers/Scoring_Benefits_Pitfalls.pdf

[3] Statistical Applications of Credit Scoring [Online]. Available

http://www.statsoft.com/Textbook/Credit-Scoring

[4] D.West, "Neural network credit scoring models," Computers & Operations Research,Volume 27, Issues 11–12, September 2000,PP. 1131–1152

[5] M. A. Doori and B. Beyrouti, "Credit Scoring Model Based on Back Propagation Neural Network Using Various Activation and Error Function," IJCSNS International Journal of Computer Science and Network Security, VOL.14 No.3, March 2014

[6] Md. S Islam, L. Zhou, F. Li, "Application of Artificial Intelligence (Artificial Neural Network) to Assess Credit Risk: A Predictive Model for Credit Card Scoring", Thesis for the Degree of MSc in Business Administration, Spring 2009

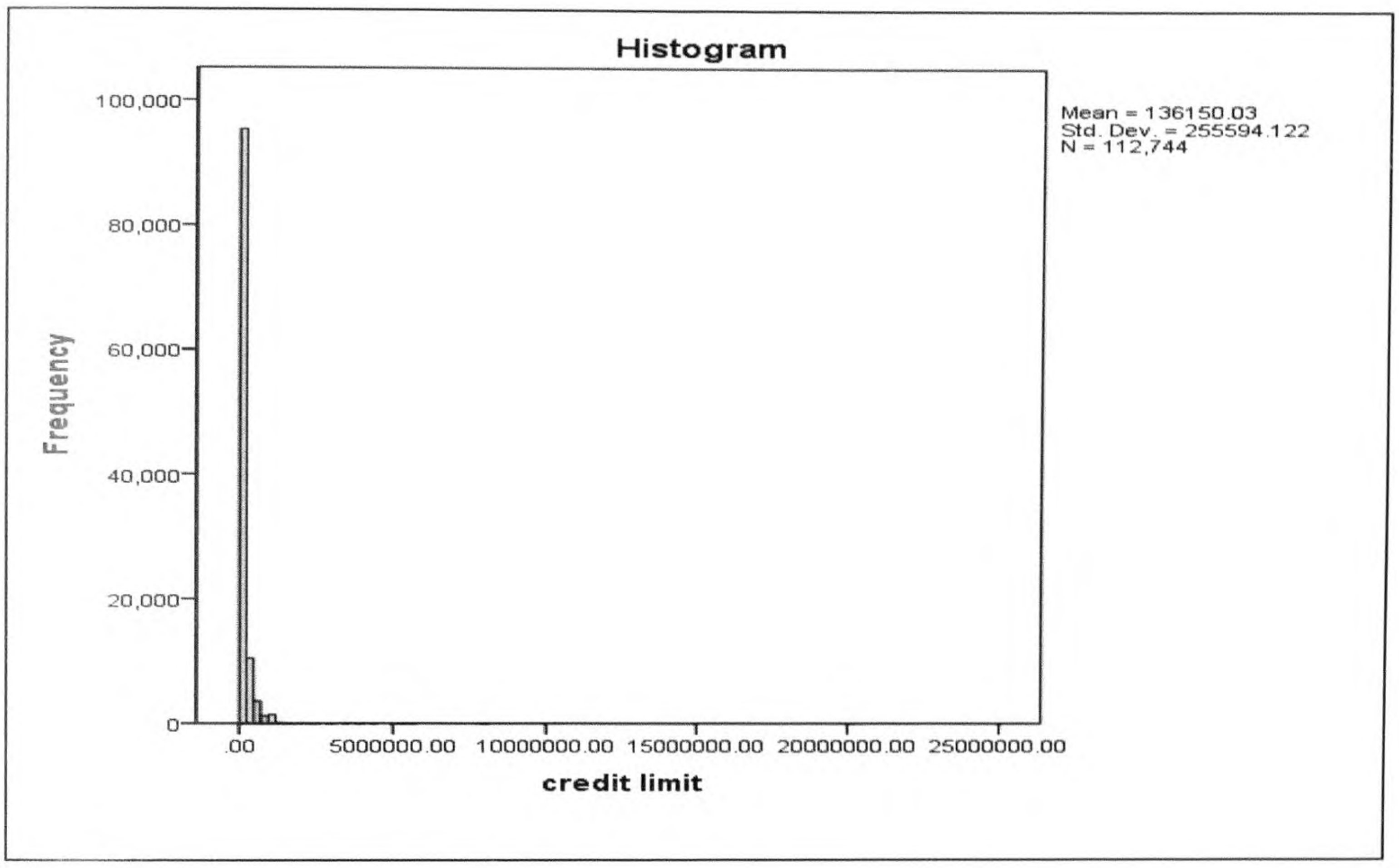

Figure 1: Distribution of credit limits

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
credit limit	112744	15000.00	20000000.00	136150.0272	255594.1223	18.472	.007	1050.869	.015
Valid N (listwise)	112744								

Figure 2: Descriptive statistic of the credit limits

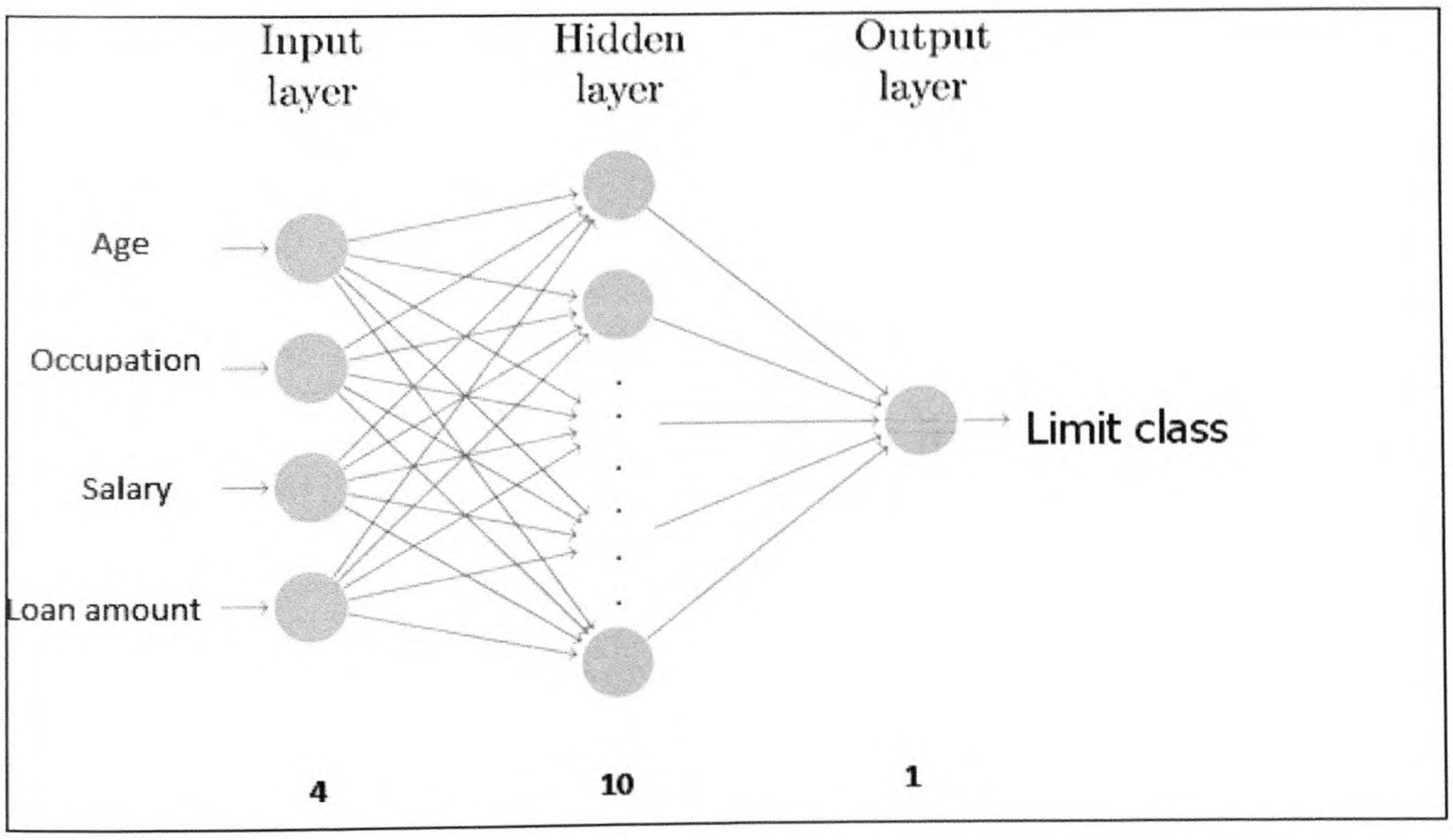

Figure 3: Neural Network model with fuzzy classes