

Neural Network Model Using Back-Propagation Algorithm with Momentum Term for Credit Risk Evaluation System

Dionicio D. Gante ¹⁺, Bobby D. Gerardo ² and Bartolome T. Tanguilig III ¹

¹ Technological Institute of the Philippines, Quezon City, Philippines

² West Visayas State University, Iloilo City, Philippines

Abstract. This paper present the results of an experiment made with the aid of KangarooBPNN, a graphical user interface software in order to find the Mean Squared Error (MSE) of a supervised neural network model. In the previous experimentation or study, NN-1B model was considered to be a good neural network model with 20 input neurons, 10 hidden neurons and 1 output neuron using 0.3 and 0.4 learning rate and accuracy rate respectively at 10,000 epochs. The German credit dataset was used to train and test the said neural network model using the back-propagation algorithm with momentum term for credit risk evaluation system. The results were recorded in a tabular form, compared and analyzed carefully. Then it was compared with the result of the previous study wherein the traditional back-propagation algorithm was used. Moreover, based on the comparison made by the researchers, it shows that it is better to use the back-propagation algorithm with momentum term than the traditional back-propagation algorithm.

Keywords: neural network, neural network model, back-propagation algorithm, back-propagation algorithm with momentum, credit risk evaluation, credit scoring.

1. Background of the Study

The accurate assessment of consumer is of uttermost importance in the banking sectors and other leading financial organizations. Credit loans and finances have a risk of being defaulted. These loans involve large amounts of capital and their non-retrieval can lead to a major loss for the financial institution. Therefore, the accurate evaluation of the risk involved is a crucial matter for banks and other such organizations. It is not only important to minimize the risks in granting credit, but also, on reducing errors in declining valid clients. Through this correct assessment, it will save the banks from lawsuits. [17]

Loan applications can be categorized into good and bad applications. Good applications are the applications that are worthy of giving the loan. Bad applications are those that should be rejected due to the low probability of the applicants ever returning the loan. The institution usually employs loan officers to make credit decisions or recommendations for that institution. These officers are given some hard rules to guide them in evaluating the worthiness of loan applicants. After some period of time, the officers also gain their experiential knowledge or intuition other than those guidelines given from their institution in deciding whether an application is loan-worthy or not. Generally, there is a widespread recognition that the capability of humans to judge the worthiness of a loan is rather poor. [6] Some of the reasons are: (1) there is a large gray area where the decision is up to the officers, and there are cases which are not immediately obvious for decision-making; (2) humans are prone to bias, for instance the presence of a physical or emotional condition can affect the decision-making process. Also, personal acquaintances with the applicants might distort the judgmental capability; and (3) business data warehouse store historical data from the previous applications. It is likely that there is knowledge hidden in this data, which may be useful for assisting the decision making. Unfortunately, the task of discovering useful relationships or patterns from data is difficult for humans. [7]

⁺ Corresponding author. Tel.: +639177054202.
E-mail address: dionygante@yahoo.com.

The reasons for such difficulties are the large volume of the data to be examined, and the nature of the relationships themselves that are not obvious.

Credit scoring methods are statistical tools employed by various banks and other financial institutions, marketing and advertising companies to estimate the probability whether the loan applicant could be categorized as potential defaulter or not. [21] Basically, credit scoring aims to classify the dependent variable with respect to the response variables. Banks collect the information about the applicant from various sources such as historical data, questionnaires, and interviews. It aims to collect all demographic details of the applicant such as age, sex, type of loan, nationality, job and income. [19] Credit scoring is a vital activity conducted by financial institutes which involve a discrete decision-making process in which the loan is accessed to avoid financial risk. However, this process is often biased as it involves a degree of personal preference due to the unavailability of suitable decision-making models. [13]

The use of Artificial Neural Network has become prominent in different sectors and field of finance. The neural network has shown its superiority over other systems in predicting future stock, bankruptcy, exchange rate and detecting credit card fraudulence. It is because the highly distributed parallel structure of neural network making it possible to make crucial decision solved. Now with the rapid growth of economic and money flow, making loan decision has become a great concern of bank authority of any country, because fraud in providing loan is spreading rapidly. [15]

2. Neural Networks

Neural networks are an emerging artificial intelligence technology that imitates the human brain on the computer. These techniques are based on the parallel, distributed processing design. The parallel structure makes neural networks proficient at analyzing problems with many variables [20; 22]. Scientists have been inspired by the capabilities of the human brain for information processing and problem-solving. Therefore, neural networks designers try to put intelligence into these systems in the form of generalized ability to learn and recognize patterns to exhibit similar intelligent functionality like humans [18].

A neural network model is composed of a number of neurons that are organized in several layers: an input layer, a hidden layer(s) and an output layer [11; 2]. The input layer of neurons feeds the input variables into the network. The hidden layer is a bridge between the input layer and the output layer. The neurons in this layer are fundamentally hidden from view, and their number and arrangement can typically be treated as a black box to those who are carrying out the system. [4]

The function of the hidden layer is to process the input variables. It is achieved by summing up all weighted inputs, checking whether the sum meets the threshold value and applying the transformation function. The weights between the input neuron and hidden neurons determine when each unit in the hidden layer may fire or not and by modifying these weights, the hidden layer may fire or not [25; 20]. In other words, the hidden layers learn the relationship between inputs and outputs in a way similar to that of the human brain by adjusting the weights during the training process [14]. On the other hand, the function of the output layer is similar to that of the hidden layer. Each input for this layer is possessed as in the hidden layer [12]. A specific neural network model is determined by its topology, learning paradigm and learning algorithm [8].

In the previous study, [5] in order to propose a good neural network model for credit risk evaluation system, the researchers developed twelve neural network models with different network parameters which are shown in Table 1. The researchers used 20 input neurons in the input layer and 1 output neuron in the output layer of every neural network model with three learning rates such as 0.3, 0.5 and 0.7. The output neuron will determine whether the loan applicant is risky or not. In the experimentation and simulation, the researchers grouped the neural network models into four groups and used four different hidden neurons in the hidden layer of every neural network model such as 5, 10, 15 and 20 hidden neurons. With regards to the training end parameters of every neural network model, four accuracy rates were used which are 0.2, 0.4, 0.6 and 0.8 while 10,000, 20,000 and 30,000 epochs were utilized as well. The said twelve neural network models were tested using the KangarooBPNN. Searching for the Mean Squared Error (MSE) and program cost time were the deciding factors in choosing which among the twelve neural network models is good for the development of credit risk evaluation system. Based on the result of the previous study, NN-1B model

with 20 input neurons, 10 hidden neurons, and 1 output neuron at 10,000 epochs and 0.3 learning rate were considered to be a good neural network model for credit risk evaluation system.

Table 1. Neural Network Models with Learning Rates

Model	Input Layer	Hidden Layer	Output Layer	Learning Rate
NN-1A	20	5	1	0.3
NN-1B	20	10	1	0.3
NN-1C	20	15	1	0.3
NN-1D	20	20	1	0.3
NN-2A	20	5	1	0.5
NN-2B	20	10	1	0.5
NN-2C	20	15	1	0.5
NN-2D	20	20	1	0.5
NN-3A	20	5	1	0.7
NN-3B	20	10	1	0.7
NN-3C	20	15	1	0.7
NN-3D	20	20	1	0.7

3. Back-Propagation Algorithm with Momentum Term

The back-propagation neural network as proposed by a group of scientist led by Rumelhart and McClland in 1986, is a kind of multilayer feed-forward network trained by error reverse propagation algorithm. It is one of the most widely used neural network models today. The BPNN can learn and store a large amount of input-output schema mapping relations, without having to build the mathematical equations that describe the mapping relation. The learning rule is the steepest descent method, which constantly adjust the network weights and threshold through the reverse propagation to minimize the sum of squared error of the network.[23]

Rojas claimed that BP algorithm could be broken down into four main steps. After choosing the weights of the neural network randomly, the back-propagation algorithm is used to compute the necessary corrections. The algorithm can be decomposed in the following four steps: (1) feedforward computation; (2) back-propagation to the output layer; (3) back-propagation to the hidden layer; and (4) weight updates. [16]

The feedforward computation or forward pass is a two-step process. The first part is getting the values of the hidden layer nodes and the second part is using those values from hidden layer to compute value or values of the output layer. Input values of input nodes are pushed up to the network towards nodes in the hidden layer. They are multiplied with weights of connecting nodes, and the values of hidden layer nodes are calculated. The sigmoid function is used for calculations. When hidden layer values are calculated, network propagates forward. It propagates values from hidden layer up to an output layer node. Below is the formulas in getting the weighted sum of the input to the j^{th} node in the hidden layer and the sigmoid activation function.

$$Net_j = \sum w_{ij} x_j \quad (1)$$

$$O_k = \frac{1}{1+e^{-Net_j}} \quad (2)$$

Once the forward pass is completed, the next step is to calculate the error of the output node. Once the error is known, it will be used for backward propagation and weights adjustment. It is a two-step process. The error is propagated from output layer to the hidden layer first. It is where learning rate and momentum are brought to the equation. So weights connecting the output layer nodes to the hidden layer nodes will be updated first. Before weights can be updated, the rate of change needs to be found. It is done by multiplication of the learning rate, error value, and output node value. When the rate of change is found, new weights can be calculated.

$$\mathbf{Oerror} = \mathbf{O}_k (\mathbf{t}_k - \mathbf{O}_k) * (\mathbf{t}_k - \mathbf{O}_k) \quad (3)$$

$$\Delta \mathbf{w}_{j,k} = \mathbf{l}_r \mathbf{Oerror} \mathbf{H}_k \quad (4)$$

$$\mathbf{w}_{OHnew} = \mathbf{w}_{old} + \Delta \mathbf{w}_{j,k} + (\alpha * \Delta(\mathbf{t} - \mathbf{1})) \quad (5)$$

Now errors have to be propagated from hidden layer down to the input layer. It is a bit more complicated than propagating error from output to the hidden layer. In the previous case, the output from output node was known beforehand. The output of hidden nodes was unknown. So, finding hidden nodes error will be calculated multiplying new weight value with an error for the output node and the same way error for other hidden nodes should be calculated. Once error for hidden layer nodes is known, weights between input and hidden layer can be updated. The rate of change first needs to be calculated for every weight. Then calculate new weights between input and hidden layer. [3]

$$\mathbf{Herror} = \mathbf{Oerror} * \mathbf{w}_{new} \quad (6)$$

$$\Delta \mathbf{w}_{j,k} = \mathbf{l}_r \mathbf{Herror} \mathbf{X}_k \quad (7)$$

$$\mathbf{w}_{HInew} = \mathbf{w}_{old} + \Delta \mathbf{w}_{j,k} + (\alpha * \Delta(\mathbf{t} - \mathbf{1})) \quad (8)$$

The introduction of momentum term is used to accelerate the learning process by encouraging the weight changes to continue in the same direction with larger steps. Furthermore, the momentum term prevents the learning process from settling in a local minimum by over-stepping the small hill. Typically, the momentum term has a value of between 0 and 1. According to Larose, he claimed that momentum term represents inertia. Large values of momentum term will influence the adjustment in the current weight to move in the same direction as the previous adjustment. [10]

4. German Credit Dataset

In the development of neural network model for credit risk evaluation system, the researchers used the German credit dataset. It is a real world dataset which has 1,000 instances that consist of 700 and 300 instances of creditworthy and noncreditworthy applicants respectively. [9] It has 20 attributes wherein 7 are numerical, and 13 are categorical. However, the dataset in the “German.data” file is not appropriate for back propagation neural network algorithm because it needs numerical values. Nevertheless, in the Applied Data Mining and Statistical Learning subject of the Eberly College of Science at the Pennsylvania State University, they have made some analysis on the German credit dataset. [1] Since the dataset have both categorical and continuous variable, they have performed Exploratory Data Analysis (EDA) and data cleaning. The proportions of applicants belonging to each classification of a categorical variable were grouped through tables. Those levels that have too few observations were merged for final analysis. And because most of the predictors are categorical with several levels, the full cross-classification of all variables leads to zero observations in many cells that cause the reduction of table size to make the dataset fit for an algorithm that needs numerical attributes. Hence, the edited German credit dataset produced by the Pennsylvania State University was utilized for the development of neural network model for credit risk evaluation system. Table 2 shows the descriptions of the German credit dataset including its class.

Table 2. Description of the German Credit Dataset

Code	Description	Class	Code	Description	Class
A1	Account Balance	Categorical	A11	Duration in Current Address	Numerical
A2	Duration in Months	Numerical	A12	Most Valuable Asset	Categorical
A3	Payment of Previous Credits	Categorical	A13	Age	Numerical
A4	Purpose of Credits	Categorical	A14	Concurrent Credits	Categorical
A5	Amount of Credit in DM	Numerical	A15	Type of Apartment	Categorical
A6	Value of Savings/ Stocks	Categorical	A16	No. of Credits at this Bank	Numerical
A7	Current Employment Length	Categorical	A17	Occupation	Categorical

Code	Description	Class	Code	Description	Class
A8	Installment Percent	Numerical	A18	Number of Dependents	Numerical
A9	Marital Status/ Sex	Categorical	A19	Telephone	Categorical
A10	Further Debtors/ Guarantors	Categorical	A20	Foreign Work	Categorical

The German credit dataset is divided into two subsets, the training set, and the validation or testing set, which will be used in the development of neural network model for credit risk evaluation system. The training set is used to train the neural network model while the testing set is used for validating the neural network model that was trained. And, to avoid over-fitting in the training dataset, the cross validation method was used. The source dataset will be taken into 10 parts in each training process, wherein one part is used for the testing dataset and the remaining 9 parts are for the training dataset. Therefore, 9:1 is the training to validation ratio.

5. Testing and Experimental Result

In the development and determination of a good neural network model to be used for credit risk evaluation system, the researchers have used the result in the previous study where the NN-1B model was considered to be a good neural network model for credit risk evaluation system. It is based on the structure of a back-propagation neural network with one input layer, one hidden layer, and one output layer. It has 20 input neurons, 10 hidden neurons, and 1 output neuron. Figure 2 shows the structure of the NN-1B model.

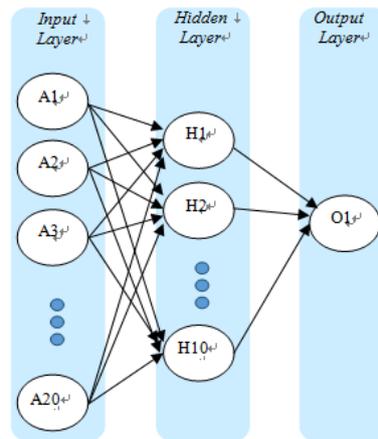


Fig. 1: The Structure of NN-1B model.

The KangarooBPNN was enhanced and utilized to help the researchers in developing a neural network model using back propagation algorithm with momentum term for credit risk evaluation system. It has four different groups of parameters, which are as follows: (1) data parameters; (2) network parameters; (3) train end parameters; and (4) file parameters. In the data parameters, it includes the dataset size both for the training and testing. The network parameters include learning rate, momentum term value and the number of nodes in the input layer, hidden layer, and output layer. The most important parameter is the learning rate and momentum term, which is the enhancement made in the KangarooBPNN. The epochs and accuracy are the two stopping criteria. The file parameters include the InputFilePath, which is the absolute path for the dataset to be used, the LogFilePath where the training process including the train and test results are saved while the well-trained model is saved on the WSavePath and VSavePath. In order to determine the best parameter values and come up with a good neural network model using back-propagation algorithm with momentum term for credit risk evaluation, NN-1B model was tested using the enhanced KangarooBPNN. The simulation using four different momentum term values such as 0.3, 0.5, 0.7 and 0.9 were obtained using a 1.8 GHz PC with 4 GB of RAM, Windows 8 OS and KangarooBPNN, which was developed using the C++ programming language.

After the experiment or simulation had been done, different figures of Mean Squared Error (MSE) both from the training stage and testing stage were recorded resulting in the table below. The said results were compared and examined carefully. The program cost time and the MSE during the training and testing stages are the deciding factor or basis in determining which among the four-momentum term values is fitted to be

used for credit risk evaluation system. Nevertheless, having the smallest MSE in the training or testing stages is not a guarantee that the value of the momentum term is a good one. Because the program cost time in generating the MSE both for training and testing stage is an important factor to be considered as well.

Table 3 shows the summary of the mean squared error both in the training and testing stage at 10,000 epochs with values of 0.3 and 0.4 for the learning rate and accuracy rate respectively. Using 0.5 momentum term value gained the smallest MSE in the testing stage and garnered the smallest program cost time while using 0.3 momentum term value have the smallest MSE in the training stage. Moreover, after a thorough analysis, it shows that using 0.5 momentum term value would be a good value to be used in order to have a good neural network model for credit risk evaluation system.

Table 3. Summary of the Mean Squared Error (MSE) and Time at 10,000 Epochs with 0.3 Learning Rate and 0.4 Accuracy Rate

Momentum	MSE Train	MSE Test	Time
0.3	0.003828	0.039700	5.539 s
0.5	0.004032	0.035630	0.672 s
0.7	0.003961	0.046203	1.25 s
0.9	0.004293	0.040886	7.781 s

Table 4 shows the comparison of the mean squared error and time gained in the training including the testing of the NN-1B model at 10,000 epochs with 0.4 accuracy rate. The use of the traditional back propagation algorithm gained the smallest MSE in the training stage while using the back propagation algorithm with momentum term garnered the smallest MSE in the testing stage and program cost time. Furthermore, Table 4 shows that using back propagation algorithm with momentum term is better compared to the traditional back-propagation algorithm.

Table 4. Comparison of the Mean Squared Error (MSE) and Time at 10,000 Epochs with 0.3 Learning Rate and 0.4 Accuracy Rate

Name of the Algorithm	MSE Train	MSE Test	Time
Back Propagation	0.003552	0.041434	6.829 s
Back Propagation with Momentum	0.004032	0.035630	0.672 s

6. Conclusion

The approval of loans by different banks is very risky and needs to be studied thoroughly in order to profit rather than to lose. Different banks around the world have recognized the importance of data mining and its techniques. The neural network as one of the data mining techniques has gained a lot of interest in the financial world. It is being used as a technique in developing a credit risk evaluation system that would help financial organizations or loan officers, in particular, to decide whether a loan should be granted or not. This paper presented the second part of the experimentation on the use of twelve developed supervised neural network models with different network parameters and training end parameters using the back propagation algorithm with momentum term which is embedded on the KangarooBPNN. Based on the results of the experiment done in the previous study, NN-1B model is considered to be a good neural network model for the development of credit risk evaluation system using back propagation algorithm compared to other eleven neural network models. And so, the said model was used in the simulation using the back propagation algorithm with momentum term and compared it with the MSE gained in the previous study. Furthermore, based on the comparison of the two algorithms, the researchers concluded that back propagation algorithm with momentum term is better than the traditional backpropagation algorithm. Hence, it should be used in the development of neural network model for credit risk evaluation system.

7. References

- [1] Analysis of German Credit Data. Applied Data Mining and Statistical Learning. Eberly College of Science at Pennsylvania State University. Retrieved May 19, 2015, from the World Wide Web: <https://onlinecourses.science.psu.edu/stat857/node/215>
- [2] Cao, Q. & Parry, M. 2009. Neural Network Earning Per Share Forecasting Models: A Comparison of Backward Propagation and Genetic Algorithm, *Decision Support Systems* 47.
- [3] Cilimkovic, Mirza. Neural Networks and Back Propagation Algorithm. Institute of Technology Blanchardstown, Dublin, Ireland.
- [4] Eletter S.F & Yaseen, S.G. 2010. Applying Neural Networks for Loan Decisions in the Jordanian Commercial Banking System. *International Journal of Computer Science and Network Security*, Vol. 10, No. 1.
- [5] Gante, D.D., Gerardo, B.D. & Tanguilig III, B.T. 2015. Neural Network Model using Back Propagation Algorithm for Credit Risk Evaluation System. 3rd International Conference on Artificial Intelligence and Computer Science, Penang, Malaysia.
- [6] Glorfeld, L.W. & Hardgrave, B.C. 1996. An Improved Method for Developing Neural Networks: The Case of Evaluating Commercial Loan Creditworthiness. *Computer Operation Research*, 23 (10), 933-944.
- [7] Handzic, M. & Aurum, A. 2001. Knowledge Discovery: Some Empirical Evidence and Directions for Future Research. Proceedings of the 5th International Conference on Wirtschaft Informatics (WI'2001), 19-21, September, Augsburg, Germany.
- [8] Handzic, M.T., Rawibawa, F. & Yeo, J. 2003. How Neural Networks Can Help Loan Officers to Make a Better Informed Application Decisions, *Informing Science Institute*.
- [9] Hofmann, Hans. (1994). UCI Machine Learning Repository. Institut f"ur Statistik und "Okonometrie Universit"at Hamburg. Retrieved January 16, 2015 from the World Wide Web: <http://archive.ics.uci.edu/ml>
- [10] Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. Retrieved on December 16, 2015, from the World Wide Web: <http://romisatriawahono.net/lecture/dm/book/Larose%20-%20Discovering%20knowledge%20in%20data%20-%202005%20.pdf>
- [11] Malhotra R. & Malhotra D.K. 2003. Evaluating Consumer Loans Using Neural Networks, Elsevier Science Ltd.
- [12] Muller, G., Steyn-Bruwer & Hamman, W. 2009. Predicting Financial Distress of Companies Listed on the JSE-A Comparison Techniques, *S.Afr.J. Business Management* 40(1).
- [13] Palihakkara, V. & Peiris, M. 2011. An Investigation in to Application of Decision Making Models for Credit Scoring in State Sector Bank. *International Symposium on the Analytic Hierarchy Process 2011*.
- [14] Peel, M. & Wilson, N. 1996. Neural Network Simulation: A New Approach to Risk Assessment and Business Forecasting, *Management Research News*, Vol. 19. No. 6.
- [15] Rahman, Mahbubur, Ahmed, Samsudin and Shuvo, Hossain. 2014. Nearest Neighbor Classifier Method for Making Loan Decision in Commercial Bank. *I.J. Intelligent Systems and Application*, 08, 60-68.
- [16] Rojas, Raul. 2005. Neural Networks: A Systematic Introduction. Springer.
- [17] Sawant, A.A. & Chawan, P.M. 2013. Study of Data Mining Techniques used for Financial Data Analysis. *International Journal of Engineering Science and Innovative Technology*, Vol. 3, Issue 3, May 2013.
- [18] Shachmurove, Y. 2002. Applying Neural Networks to Business, Economics, and Finance.
- [19] Singh, R. & Aggrawal, R.R. 2011. Comparative Evaluation of Predictive Modelling Techniques on Credit Card Data. *International Journal of Computer Theory and Engineering*, Vol. 3. No. 5, October 2011, 598-603.
- [20] Tafti, M.H.A. & Nikbakht, E. 1993. Neural Networks and Expert Systems: New Horizons in Business Finance Applications, *Information Management and Computer Security*, Vol. 1, No. 1.
- [21] W. E. Henley "Statistical Aspects of Credit Scoring" Ph.D. Dissertation, The Open University, Milton Keynes, UK.1995.
- [22] Zhang, G.P. 2004 Neural Networks in Business Forecasting, Idea Group Inc.
- [23] Zhang, R. & Jiang, C. 2013. The Bank Risk Forewarning Model of Back Propagation Neural Network Based on the Cloud Computing. *International Journal of Intelligent Information Processing*, Vol. 4, No. 1, March 2013, 23-31.