International Journal of Engineering & Technology, 7 (2.15) (2018) 35-37

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Performance Comparison of Neural Network Training Algorithms for Modeling Customer Churn Prediction

Mohd Khalid Awang [1]*, Mohammad Ridwan Ismail [2], Mokhairi Makhtar[3] , M Nordin A Rahman[4], Abd Rasid Mamat[5]

[1,2,3,4,5] Faculty of Informatics and Computing,
Universiti Sultan Zainal Abidin, Besut Campus,
22200 Besut, Terengganu, Malaysia
**Email: khalid@unisza.edu.my*

Abstract

Predicting customer churn has become the priority of every telecommunication service provider as the market is becoming more saturated and competitive. This paper presents a comparison of neural network learning algorithms for customer churn prediction. The data set used to train and test the neural network algorithms was provided by one of the leading telecommunication company in Malaysia. The Multilayer Perceptron (MLP) networks are trained using nine (9) types of learning algorithms, which are Levenberg Marquardt backpropagation (trainlm), BFGS Quasi-Newton backpropagation (trainbfg), Conjugate Gradient backpropagation with Fletcher-Reeves Updates (traincgf), Conjugate Gradient backpropagation with Polak-Ribiere Updates (traincgp), Conjugate Gradient backpropagation with Powell-Beale Restarts (traincgb), Scaled Conjugate Gradient backpropagation (trainscg), One Step Secant backpropagation (trainoss), Bayesian Regularization backpropagation (trainbr), and Resilient backpropagation (trainrp). The performance of the Neural Network is measured based on the prediction accuracy of the learning and testing phases. LM learning algorithm is found to be the optimum model of a neural network model consisting of fourteen input units, one hidden node and one output node. The best result of the experiment indicated that this model is able to produce the performance accuracy of 94.82%.

Keywords: *Neural Network Learning Algorithm, Data Mining, Customer Churn Prediction, Multilayer Perceptron.*

1. Introduction

In the era where telecommunication technology changes rapidly, products keep changing to satisfy the needs of the customer. Getting new customers is rather challenging due to competitive pricing, advertisement, and attractiveness of the products. The service providers have incurred huge costs of marketing, channel, network, and manpower in obtaining new customers rather than retaining existing customers. In the telecommunication industry, the issue of customer churn has become an increasingly important contribution of revenue or profit to the company [1,2]. Customer retention program is one of the best strategies to keep customers' loyalty. The customer churn can occur at any time because service providers are unable to manage their customers on a one-by one basis.

To minimize churn rate, churn prediction has been proposed by previous studies to identify customers who are likely to churn [3-5]. This prediction has helped telecommunication companies in managing their customers with potential to move out through retention program. In order to manage the customers well, providers should investigate the reasons why their customers have developed the intention to leave. Furthermore, the companies should be efficient in terms of network, customer service, and product offering to meet the customers' needs.

2. Related Works

A variety of techniques have been proposed in data mining analysis to predict customer churn. No specific tool can be used for data mining because every data in the telecommunication companies present different types of information [6-8]. Previous studies have proposed many methods to predict customer churn using statistical and machine learning algorithm [9-10,12-16]. The scarcity of studies investigating customer churn using machine learning techniques is the impetus of this research to explore the potential of an artificial neural network to improve customer churn prediction. Predicting customer churn is not straightforward as various factors of predictor need to be analysed.

The neural network model is based on the existing human neuron. Understanding human neurons have led to the formation of an artificial neuron as the basic building block of neural network model [17-19]. Neurons, better known as the processing unit, is a very important element and are interconnected with one another to process such information. As with human neurons, each processing unit is connected to other neurons via input connection, such as a biological neuron dendrite. It also has a number of units that are connected to the output neuron axon. Each processing unit has a power potential or known as activation function, which has the same function as the internal potential of the neuron [20,21].

Generally, the processing unit in artificial neural networks has three main functions: i) receiving input by the neurons; ii) ena-

Copyright © 2016 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

bling the calculation based on the input; and iii) ultimately deploying the calculated result or output to other units connected. ANN consists of three main elements, which are weight, bias, and activation function 21. Each input to the neural network is assigned with corresponding weight that is multiplied together and summed together with bias value, and then, computed by a mathematical function, which is known as activation function. By adjusting the weights of an ANN, the error between the predicted and the actual output can be minimized until the optimum network is achieved. Thus, previous studies have proposed various algorithms to adjust the weights of the neural network in order to obtain an optimum network and to minimize errors. This process is known as the training phase.

When the data are presented at the input of neural network, the networks perform calculation into successive layer. The neural network output is calculated and the error function ε(t) is determined based on the following equation. The error is then fed back to modify the connections weight known as backpropagation process. This process continues until the optimal network is achieved.

$\varepsilon(t)=y(t)-\hat{y}(t)$

where

$y(t)$ = actual output,

$\hat{y}(t)$ = neural network output.

The learning algorithm is a process to modify the weights and thresholds of a neural network with the objective to minimize the mismatch between the desired output and the obtained output from the neural network for any given input. In this project, the MLP network is trained using nine (9) different learning algorithms that are Levenberg Marquardt backpropagation (trainlm), BFGS Quasi-Newton backpropagation (trainbfg), Conjugate Gradient backpropagation with Fletcher-Reeves Updates (traincgf), Conjugate Gradient backpropagation with Polak-Ribiere Updates (traincgp), Conjugate Gradient backpropagation with Powell-Beale Restarts (traincgb), Scaled Conjugate Gradient backpropagation (trainscg), One Step Secant backpropagation (trainoss), Bayesian Regularization backpropagation (trainbr), and Resilient backpropagation (trainrp).

3. Methodology

MLP networks were trained to classify the input data based on the churn and non-churn. In this project, the dataset was obtained from a telecommunication company in Malaysia and it contained 312 data samples, which consisted of 177 churn customers and 135 non-churn customers. The distribution of data analysis is shown in Table 1.

Table 1: Summary of terms for measurements

	Predicted Churn	Predicted Non-churn
Actual Churn	TP	FP
Actual Non-churn	FN	TN

In engineering and statistics, the accuracy of the pattern classification system is to measure the level of closeness to the quantity of true value. Measurements of accuracy play an important role in the selection of prediction tools. In this research paper, accuracy is defined as:

Accuracy=(TP+TN)/(TP+TN+FN+FP)×100

The sensitivity of a test refers to how many cases of churn in a particular test that can be found. For example, 100% of sensitivity means the prediction tools are able to classify all actual positive cases. In mathematical form, sensitivity can be defined as:

Sensitivity=TP/(TP+FN)×100

The alternative side of sensitivity is known as specificity, which measures the capability of a prediction tool to identify the non-churn causes. A 100% of specificity means that the tools are able to identify all non-churn cases correctly. The specificity is defined as:

Specificity=TN/(TN+FP) ×100

The important part of churn prediction is the selection of the variables in the analysis. Table 2 shows the proposed features that were used as churn predictors in the telecommunication company. There were 14 out of 20 input features used as predictive factors for regression analysis and neural network based on p-value < 0.05. The result of p-value < 0.05 shows the significant of input variables to the output. The other features were excluded from the analysis. The features were manually extracted based on the proposed features from the data warehouse. The features were saved in text (.txt) format before they underwent regression and neural network analyses.

Table 2: The proposed input features for churn prediction system

Features	Input Features	
Customer Demographics Data	1.	Age
	2.	Gender
Customer Relationship Data	3.	Service Packages
	4.	Monthly Commitment
	5.	Call Plan
	6.	Service Rental
	7.	Method of Payment
Customer Billing Data	8.	Total usage bill
	9.	Account Receivable (AR) days
	10.	Total payment
	11.	Total bills for local calls
	12.	Total bills for national calls for same operator
	13.	Total bills for national calls for different operators
Customer Usage Data	14.	Total number of calls out to local calls
	15.	Total number of calls out to national calls for same operator
	16.	Total duration for national for same operator
	17.	Total number of free calls for same operator
	18.	Total number of national calls for different operator
	19.	Total duration for national for different operators
	20.	Total number of free calls for different operators

The Multilayer Perceptron (MLP) networks were trained using nine (9) types of learning algorithms, which were Levenberg Marquardt backpropagation (trainlm), BFGS Quasi-Newton backpropagation (trainbfg), Conjugate Gradient backpropagation with Fletcher-Reeves Updates (traincgf), Conjugate Gradient backpropagation with Polak-Ribiere Updates (traincgp), Conjugate Gradient backpropagation with Powell-Beale Restarts (traincgb), Scaled Conjugate Gradient backpropagation (trainscg), One Step Secant backpropagation (trainoss), Bayesian Regularization backpropagation (trainbr), and Resilient backpropagation (trainrp). The number of optimum hidden nodes was chosen based on the lowest hidden nodes with the highest accuracy at testing phase.

The next step was to compare the optimum structure and the performance of each neural network to identify the most suitable neural network that was used to develop the churn prediction system.

4. Experiment Results

The optimum structure and the performance of each neural network are shown in the following table.

Table 3: Performances of various types of neural network in predicting customer churn and non-churn

	Training Phase		Training Phase		
	Hidden Node	Acc (%)	Hidden Node	Acc (%)	Overal Acc (%)
trainlm	174	99.63	21	90.00	94.82
trainbfg	150	94.49	58	90.00	92.25
traincgf	100	88.24	5	87.50	87.87
traincgp	164	87.50	1	85.00	86.25
traincgb	184	88.97	4	87.50	88.24
trainscg	113	88.97	16	87.50	88.24
trainoss	114	92.65	62	90.00	91.33
trainbr	68	90.81	122	90.00	90.41
trainrp	28	77.57	13	87.50	82.54

As can be seen in Table 3, the Multilayer Perceptron (MLP) trained with the Levenberg Marquardt (LM) backpropagation gave the best performance among the tested neural networks. The overall accuracy of the MLP trained with LM algorithm was 94.82%. Besides, an analysis with LM algorithm yielded results of high accuracy with small number of hidden node. BFGS Quasi-Newton backpropagation was the second neural network that provided higher overall accuracy, which was 92.25%, followed by One Step Secant backpropagation (91.33%), Bayesian Regularization backpropagation (90.41%), Conjugate Gradient backpropagation with Powell Beale Restarts (88.24%), Scaled Conjugate Gradient backpropagation (88.24%), Conjugate Gradient backpropagation with Fletcher-Reeves Updates (87.87%), Conjugate Gradient backpropagation with Polak-Ribiere Updates (86.25%), and Resilient backpropagation (82.54%).
From the MLP neural network analyses, LM backpropagation was chosen as the best algorithm to predict customer churn and non-churn. The result of LM analysis is shown in Tables 4(a) and 4(b) for training and testing phases respectively.

Table 4(a): The performances of Levenberg Marquardt (LM) performance for training

Predicted	Actual		Totals
	Churn	Non- Churn	
Churn	20	3	23
Non- Churn	1	16	17
Totals	21	19	40

Table 4(b): The performances of Levenberg Marquardt (LM) performance for training

Predicted	Actual		Totals
	Churn	Non- Churn	
Churn	155	0	155
Non- Churn	1	116	117
Totals	156	116	272

Tables 4(a) and 4(b) show that the results of the accuracy, sensitivity, and specificity for the training stage were 99.63%, 99.36%, and 100.00% respectively, while for the testing set, the accuracy, sensitivity, and specificity were 90.00%, 95.24%, and 84.21% respectively.
Detailed submission guidelines can be found on the journal web pages. All authors are responsible for understanding these guidelines before submitting their manuscript.

5. Conclusion

This paper presents the comparison of nine neural network learning algorithms for customer churn prediction. The multilayer perceptron networks trained with LM algorithm contributed the best performance among the neural network learning algorithms. Although the LM algorithm yielded a result with large number of hidden nodes, its accuracy produced the highest priority in determining the best neural network model. The analysis shows that neural network trained with Levenberg Marquart (LM) algorithm is able to gain the prediction accuracy of 94.82%. The result shows that the proposed predictors are acceptable in churn prediction using neural network tools. The optimum model of neural network consists of fourteen input units, one hidden node and one output node with Levenberg Marquardt (LM) learning algorithm. As a whole, the findings suggest that a neural network in modelling techniques offers a viable alternative to traditional predictive approaches in customer churn prediction.

References

[1] S. Portelaa and R. Menezes, Modeling Customer Churn: An Application of Duration Models. Proceedings of the Australian and New Zealand Marketing Academy (ANZMAC) (2009)

[2] J. Hadden, A Customer Profiling Methodology for Churn Prediction. Cranfield University (2008)

[3] A. Sharma and P. K. Panigrahi, A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. International Journal of Computer Applications, (2011) 27(11), 26-31

[4] H. E. Chueh, Analysis of marketing data to extract key factors of telecom churn management. African Journal of Business Management, (2011) 5(20), 8242-8247

[5] A. Berson, S. J. Smith and K. Thearling, Building Data Mining Applications for CRM, McGraw-Hill, NewYork (2000)

[6] M. C. Mozer, R. Wolniewicz, D.B. Grimes, E. Johnson and H. Kaushansky, Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. IEEE Transactions On Neural Networks, (2000) 11(3), 690 - 696.

[7] B. H. Chu, M.S. Tsai and C.S. Ho, Toward a hybrid data mining model for customer retention. Knowledge-Based Systems, (2007) 20(8), 703-718.

[8] M.K. Awang, M.N.A. Rahman and M.R. Ismail, Data Mining for Churn Prediction: Multiple Regressions Approach, Computer Applications for Database, Education, and Ubiquitous Computing. Springer Berlin Heidelberg, (2012) 318-324

[9] D. Chiang, Y. Wang, S. Lee and C. Lin, Goal-oriented sequential pattern for network banking and churn analysis. Expert systems with applications, (2003) 25, 293-302.

[10] A. T. Kearney, European Mobile Industry Observatory 2011. In GSMA (Ed.), Rising to the Challenge of Intense Competition, (2011)

[11] A. D. Athanassopoulos, Customer Satisfaction Cues To Support Market Segmentation and Explain Switching Behavior. Journal of Business Research, , (2000) 47, 191-207

[12] S.K. Abi, M.R. Gholamian, M. R. and M. Namvar, Data Mining Applications in Customer Churn Management. Paper presented at the International Conference on Intelligent Systems, Modelling and Simulation, (2010)

[13] J. B. Ferreira, M. Vellasco,M. A. Pacheco and C. H. Barbosa, Data Mining Techniques on the Evaluation of Wireless Churn. Paper presented at the European Symposium on Artificial Neural Networks (2004)

[14] Y. He, Z. He and D. Zhang, A Study on Prediction of Customer Churn in Fixed Communication Network Based on Data Mining. Paper presented at the Sixth International Conference on Fuzzy Systems and Knowledge Discovery (2009)

[15] P. Datta, B. Masand, D. R. Mani and B. Li, Automated Cellular Modeling and Prediction on a Large Scale. Artificial Intelligence Review, (2000) 14(6), 485-502

[16] C. Kang and S. Pei-ji, Customer Churn Prediction Based on SVM-RFE. Paper presented at the International Seminar on Business and Information Management, (2008)

[17] S. Haykin, Neural Network : A Comparehsive Foundation (Vol. 2). New Jersey Prentice Hall International, (1999)

[18] A. Krogh, What are artificial neural networks? Nature Publishing Group, (2008) 26(2), 195-197

[19] D. M. Levine, P. P. Ramsey and R. K. Smidt, Applied statistics for engineers and scientists: using Microsoft Excel and Minitab. Upper Saddle River, NJ: Prentice Hall, (2001)

[20] S. A. Sweet and K. Grace-Martin, Data Analysis with SPSS: A First Course in Applied Statistics (Vol. 2), Pearson (2010)

[21] L. Fausett, Fundamentals of neural networks: architectures, algorithms, and applications. (1994)