Churn and Non-churn of Customers in Banking Sector Using Extreme Learning Machine

Ramakanta Mohanty and C. Naga Ratna Sree

Abstract With an invention been made in computational techniques, there has been increasing interest in the field of neural networks as major potential machine learning techniques have come to the front. In the current past, gradient descent-based algorithm is used in feed forward neural network and the parameter utilized, which devours extensively more opportunity for learning and tuned iteratively. The single hidden layer feed forward neural network have the input weights, the hidden layer and biases randomly assigned. The straightforward backward operation is utilized to discover output weight which thus relies on upon handling from hidden layer to output layer. In this paper, we propose to utilize Extreme Learning Machine (ELM) to foresee client churn. The goal of this paper is to propose a novel approach that enhances the precision of churn and non-churn of clients using banking data.

Keywords Extreme learning machine · Artificial neural networks
Customer churn · Single-layer feed forward network

1 Introduction

Customer churn implies loss of clients. Customer churn can be seen in numerous ventures like banking, broadcast communications and insurance organizations. Nowadays, Customer churn has turned into a vital issue in the banking industry, as well as in media transmission industry excessively [1]. To discover the early cautioning signs in lessened exchanges of clients, steps should be taken to retain the customer. The principal target of the customer retention is to cut through the

R. Mohanty (✉) · C. Naga Ratna Sree
Computer Science and Engineering, Keshav Memorial Institute of Technology, Narayanaguda, Hyderabad 500011, India
e-mail: ramakanta5a@gmail.com

C. Naga Ratna Sree
e-mail: cnratnasree@gmail.com

V. Bhateja et al. (eds.), *Proceedings of the Second International Conference on Computational Intelligence and Informatics*, Advances in Intelligent Systems and Computing 712, https://doi.org/10.1007/978-981-10-8228-3_6

customer's behaviour based on predictions. Because of the mechanical headways, information has been expanded in an awesome space, size and dimensionality step by step. Accordingly, it is essential to investigate some viable machine learning strategies that can be used to analyse the information and haul out the valuable information from the information.

Of late, Extreme Learning Machines (ELMs) has become one of the prominent machine learning methods for predictive analysis. ELMs are a feed forward neural network recognized by the introduction of their hidden layer weights, alongside the training algorithm [2–5]. These models can create great execution and are utilized to find a huge number of times quicker than other types of neural network like probabilistic neural network, decision trees and by use of ensemble method [6, 7].

Since client agitation has turned into a noteworthy issue, to keep the beat, we utilize a machine learning model where it is utilized to discover which sort of clients will probably churn.

This paper is organized as follows: Sect. 2 presents about related work on churn and non-churn different domain. Section 3 examines about the review of an extraordinary extreme learning machine. In Sects. 4 and 5 the outcomes are discussed and the paper is concluded.

2 Literature Survey

Building a functional model for customer churn has now become a decisive topic in recent days. The emergence of e-commerce has increased the information availability and offered for companies to respond efficiently to the clients, hence this gave rise to the notion of the customer churn by Lejeune et al. [8]. Kotler and Keller [9] analysed that customer relationship management is nothing but solving the customer's problem such as customer dissatisfaction, switching cost, service message and customer status and so on. Kincaid [10] defined CRM as the essential usage of information processing, use of technology, and people to manage the client relationship with the governing body for the industry to develop. Burez and Van den Poel [11, 12] proposed to manage the customer churn by two methods, viz. reactive and proactive. According to him, a customer requests the organization to offset the service supplied by the organization but on the other hand, in the proactive approach, organization solved the problem brought up by the customer and produced it as suits for the churn and find out a constructive way to solve the issues. Shaw et al. [13] proposed that genuine client relationship management can be found by accumulating the information revelation handle with the administration and by seeking after the promoting the market plans. Salman et al. [14] examined about logistic regression and time oriented, analytical technique to dissect client agitation. Zhang et al. [15] developed a model which consisted of logistic regression, decision tree and neural network (NN) learning algorithms by using Chinese mobile telecommunication dataset and also used SAS Enterprise Miner for training the models for prediction of customer churn. Coussement and Van sanctum Poel [16] utilized support vector machines and the result acquired by

SVM is that SVM performed superior to other methods. Mutanen [17] employed logistic regression in personal retail banking dataset to predict customer churn. Most recently, Mohanty and Jhansi Rani [18] employed FuzzyArtMap and counter propagation neural network on Indian Telecommunication data to predict churn of customers. From this, we can conclude that neural networks can be used to predict customer churn in different domains such as retails [19], banking [20] and finance [21]. This paper proposes a neural network-based approach, i.e., extreme learning machine to predict the customer churn with respect to the banking domain.

3 Methodology

3.1 Extreme Learning Machine

The architecture of extreme learning machine is depicted in Fig. 1. Extreme Learning Machine (ELM) is a learning algorithm proposed by Huang et al. [2–5]. It consists of a single-layer feed forward neural network, which consists of an input layer, hidden layer and output nodes. The input weight is accessed to the outputs by series of weights.

The main advantages of ELM are that its parameters, hidden nodes, input weights and biases are randomly allocated and need not required to be tuned.

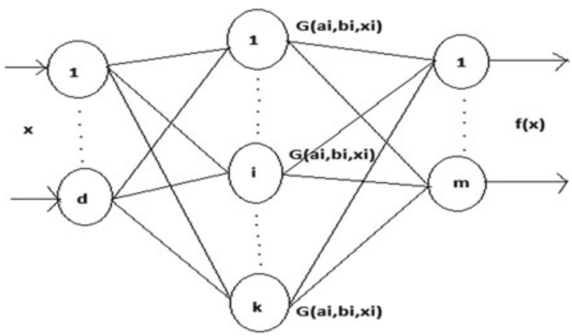
Let D be the training data of M samples, which can be represented as $(X_1 t_1)$, where $\{X_1)$ Є R^m, R^m is m times input vector. t_1 Є R^n, *where* R^n is n times Xl target vector. The ELM is a single-layer network having N nodes and let f(x) be the activation which can be defined as

$$\sum_{i=1}^{\bar{N}} \beta_i f_i(x_j) = \sum_{i=1}^{\bar{N}} \beta_i f(w_{i,} b_{i,} x_j) = o_j, 1 \leq \mathrm{j} \leq . \tag{1}$$

where w_i is weight vector of the input hidden nodes.

β_i be the weight vector of both the hidden nodes and the output nodes, whereas b_i is the threshold of the hidden node.

Fig. 1 Architecture of ELM



We need to derive the relation between X_i and t_i from β_i, w_i and b_i such that Eq. (2) can be written as

$$\sum_{i=1}^{\bar{N}} \beta_i f(w_{i,} b_{i,} x_j) = t_j, 1 \leq j \leq \mathrm{N}. \quad (2)$$

Equation (2) can be written alternatively as

$$\mathrm{H}\beta = T. \quad (3)$$

where

$$H = \begin{bmatrix} h(x_1) \\ h(x_N) \end{bmatrix} = \begin{bmatrix} f(w_{i,} b_{i,}, x_j) & \ldots & f(w_{N,} b_{N,}, x_1) \\ \vdots & \ddots & \vdots \\ f(w_{i,} b_{i,}, x_N) & \ldots & f(w_{N,} b_{N,}, x_N) \end{bmatrix}_{NXn} \quad (4)$$

H is the hidden layer output matrix of the network.

$\mathrm{h(x)} = f(w_{i,} b_{i,}, x) \ldots f(w_{N,} b_{N,}, x)$ is the hidden layer feature mapping. Equation (4) provides the hidden node output with respect to the input vector. If we know the activation function, we can assign randomly the value of input weight vector w_i and hidden layer biases b_i. Therefore, parameters need not required to be tuned without changing the hidden layer output matrix. Accordingly, ELM provides smallest training error and the smallest output weight as well to predict accurately.

$$Min = ||\mathrm{H}\beta = \mathrm{T}|| \quad (5)$$

If N (the no. of hidden neuron) = training sample, then, the output weight β can be found out by inverting the H. Thus, we can get less error in training samples by using ELM.

Further,

$$\boldsymbol{\beta} = H^{+}\mathrm{T}, \quad (6)$$

where H^{+} is the Moore–Penrose pseudo inverse of matrix H, and it can be computed by using the orthogonal projection method as

$$H^{+} = \left(\mathrm{HH}^T\right)^{-1}\mathrm{H}^T, \quad (7)$$

when HH^{T} is remarkable, or

$$H^+ = \mathrm{H}^T \left(\mathrm{HH}^T\right)^{-1}. \tag{8}$$

When HH^{T} is nonremarkable.
Where T means matrix transposition.

3.2 Algorithm

A dataset N with the number of hidden neurons L and Activation Function $f(x)$ is taken then

Step 1. Randomly assign input weights w_i and biases $b_{i.}$ i.e. hidden layer parameters $(w_{i,}\ b_{i,})$, i = 1———L randomly.
Step 2. Calculate the hidden layer output matrix H.
Step 3. Calculate the output weight β: $\beta = H^+ T$.

4 Results and Discussions

The churn and non-churn of customers is a Portuguese Banking Sector dataset, where it consists of both categorical and numerical values. The dataset consisting of 41,189 samples and having 20 attributes are presented in Table 1.

The dataset is divided into training and testing in the ratio of 80:20 and followed by tenfold cross-validation. We developed the Java code for ELM and the experiment is carried out in MATLAB environment. We use the input function for ELM as follows: training data, testing data, ELM type, number of hidden neurons, and activation function type. The outputs of ELM algorithm are average training time, average testing time, average training, and testing accuracy. We used the following formula for calculating the training and testing accuracies as follows:

$$Training\,Accuracy = 1 - MisclassificationRate\_Training/Number\ of\ Training\ data$$

and

$$Testing\,Accuracy = 1 - MisclassificationRate\_Testing/Number\ of\ Testing\ data.$$

In our experiment, we chose four types of activation function, viz. Sigmoid, Radbias, Hardlim and Tribias, respectively. From our experiment, we found that Tribias activation function took less time for testing data to be processed. Accordingly, it also provides best training accuracy of 0.0044 and testing accuracy of 0.0420 by using the Tribias activation function. We got training accuracy of

Table 1 Attributes of Portuguese telemarketing dataset

Sl. no	Attributes
1	Age
2	Job (type of job)
3	Marital (marital status)
4	Education
5	Default (has credit in default)
6	Housing (housing loan)
7	Loan (personal loan)
8	Contact (contact communication)
9	Month (last contact month year)
10	Day_of_week (last contact day of week)
11	Duration (last contact duration)
12	Campaign (no. of contact performed)
13	Pdays (no. of days that passed the client contacted)
14	Previous (no. of contact before campaign)
15	Poutcome (outcome of previous market campaign)
16	Emp.var.rate (employment variation date)
17	Cons.price.idx (consumer price index)
18	Cons.conf.idx (consumer confidence index)
19	Eribur3m (daily indicator)
20	nr.employed (no. of employees)
21	Output variable (yes or no)

0.0431 and testing accuracy of 0.0420 by using Radbias activation function compared to other activation functions, which are presented in Table 2.

Further, we simulate our experiment folds wise, we found that in case of fold 1 the best training accuracy is of 0.0045 and followed by fold 6 value of 0.0500, respectively. But, on the other hand, the best testing accuracy value is on fold 1 of value 0.0024 and followed by fold 5 of value 0.0344, respectively, which is shown in Table 3.

Table 2 The training and testing accuracies and time taken by applying different ELM activation functions

Type of activation function	No. of hidden neurons	Average training time (ms)	Average testing time (ms)	Training accuracy	Testing accuracy
Sigmoid	700	39.6875	0.3281	0.0566	0.0582
Radbias	600	45.1250	0.2031	0.0431	0.0420
Hardlim	500	16.5313	0.2500	0.0541	0.0585
Tribias	400	8.4375	0.1718	0.0044	0.0420

Table 3 The training and testing accuracies fold wise of dataset

Average	Training time	Testing time	Training accuracy	Testing accuracy
Fold 1	7.43	0.14	0.0045	0.0024
Fold 2	11.75	0.28	0.0509	0.0449
Fold 3	16.35	0.28	0.0520	0.0398
Fold 4	11.84	0.28	0.0540	0.0473
Fold 5	13.64	0.20	0.0536	0.0344
Fold 6	12.34	0.20	0.0500	0.0565
Fold 7	14.75	0.31	0.0509	0.0548
Fold 8	14.07	0.31	0.0515	0.0424
Fold 9	16.87	0.31	0.0517	0.0388
Fold 10	15.06	0.21	0.0532	0.0376
Total average	13.410	0.23	0.04714	0.03989

The novelty of this algorithm is that processing time for both training and testing data is in the order of milliseconds compared to other techniques what we observed earlier [22, 23]. In our experiment, for training data, it took an average time of 13.41 ms and testing data of 23 ms out of 41.189 samples.

5 Conclusions

This paper analyses the systematic way to predict customer churn by employing Extreme Learning Machine. The different activation functions used to predict the actual churners relatively well. Bank customers churning average Training time, Average Testing Time and Average Training Accuracy, Average Testing Accuracy is being computed from the given dataset. The ELM model gives more accurate results compared to other machine learning techniques, viz. SVM, Gradient descent, backpropagation neural networks, etc. Time to time, many data mining techniques have been implemented on the banking data to predict the customer churn and non-churn but our results outperformed all other machine learning techniques in terms of accuracies and time taken to process the dataset. This is the significant study in this paper.

References

1. Athanassopoulos, A. D.: Customer satisfaction cues to support market segmentation and explain switching behavior. Journal of Business Research, Warwick Business School, pp. 191–207, (2000).

2. Huang, G. B., Chen, L., Siew, C. K.: Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes. IEEE Transactions on Neural Networks, vol. 17, no. 4, pp. 879–892, (2006).
3. G. B. Huang, G. B., L. Chen., Convex Incremental Extreme Learning Machine: Neurocomputing, vol. 70, pp. 3056–3062, (2007).
4. Huang, G. B., Zhou, H., Ding, H., and Zhang, R.: Extreme Learning Machine for Regression and Multiclass Classification, IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 42, no. 2, pp. 513–529, (2012).
5. Huang, G. B.: An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels, Cognitive Computation, vol. 6, pp. 376–390, (2014).
6. Huang, G. B., Chen L.: Enhanced random search based incremental extreme learning machine, Neuro computing 71, 3460–3468, (2008).
7. Huang, G. B., and Ding, X., Zhou, H.: Optimization method based extreme learning machine for classification. Neuro computing 74, pp. 155–163, (2010).
8. Lejeune, M. A.: Measuring the impact of data mining on churn management. Internet research: Electronic Networking Applications and Policy, 11, pp. 375–387 (2001).
9. Kotler, P., Keller, L.: Maketing Management (12th ed.). New Jersey, Pearson prentice Hall (2006).
10. Kincaid, J.: Customer relationship management: Getting it Right NJ: Pretice-Hall PTR (2003).
11. Burez, J., Van den Poel, D.: Handling class imbalance in customer churn prediction, Expert system with Applications, 36, pp. 4626–4636 (2009).
12. Bue, J., Van den Poel, D.: Separating financial from commercial customer churn: Expert Systems with Applications, 35, pp. 497–514 (2008).
13. Shaw, M. Subramaniam, C., Tan, G., Weldge, M.: Knowledge management and data mining for marketing, Decision Support System, 31(1), pp. 127–137 (2001).
14. Salman, S.: Value based Time dimensioned Churn Prediction, Journal of Emerging trends in Computing and Information, vol. 4, pp. 180–183, (2013).
15. Zhang, X., Z. Liu, Z., Yang, X., Shi, W., and Wang, Q.: Predicting Customer Churn by Integrating the Effect of the Customer Contact Network: IEEE International conference on Service Operations, Logistics and Informatics (SOLI), Shandong, China, pp. 392–397, (2010).
16. Coussement, K., Van den Poel: Improving Customer attrition Prediction by integrating emotions from Client/Company interaction emails and evaluating multiple classifiers. Experts Systems with Applications, 36, pp. 6127–6134 (2009).
17. Mutanen, T. Customers churn analysis – a case study, Research Report No. VTT-R01184-06, Dated March 15, Available at: http://www.vtt.fi/inf/julkaisut/muut/2006/customer_churn_case_study.pdf, Retrieved on 19 August (2008).
18. Mohanty R, Jhansi Rani K.: Application of Computational Intelligence to predict churn and Non churn of customers in Indian Telecommunication. IEEE International Conference on Computational Intelligence and Communication Networks, pp. 598–603, (2015). India.
19. Buckinx, W., Van den Poel, D.: Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research, 164(1), pp. 252–268, (2005).
20. Van den Poel, D., Larivie're, B.: Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research, 157, pp. 196–217, (2004).
21. Chiang, D. A., Wang, Y. F., Lee, S. L., Lin, C. J.: Goal-oriented sequential pattern for network banking churn analysis. Expert Systems with Applications, 25, pp. 293–302, (2003).
22. Mohanty, R., Ravi, V., Patra, M. R.: Hybrid Intelligent Systems for Predicting Software Reliabilit, Elsevier, Applied Soft Computing 13, PP. 180–200, (2013).
23. Mohanty, R., Ravi, V., Patra, M. R.: Web Services Classification using intelligent Techniques, Elsevier, Expert Systems with Applications 37, PP. 5484–5490, (2010).