Customer Churn Prediction in Irancell Company Using Data Mining Methods

Homa Meghyasi and Abas Rad

Customer Churn Prediction in Irancell Company by Using Data Mining

Homa Meghyasi
Master Student of Information Technology, Islamic Azad University, Tehran, Iran
Homa.Meghyasi@gmail.com

Abas Rad
Faculty of Management and Accounting, Shahid Beheshti University, Tehran, Iran
A-raad@sbu.ac.ir

Abstract—At present, in competitive space between companies and organizations, customers churn is their most important challenge. When a customer becomes churn, organizations lose one of their most important assets, which can lead to financial losses and even bankruptcy. Customer churn prediction using data mining techniques can alleviate these problems to some extent. The aim of the present study is to provide a hybrid method based on Genetic Algorithm and Modular Neural Network to customer churn prediction using Irancell Company data. Based on the results of similar methods in this field, the accuracy of the proposed method is about 95.5% compared to other methods. Using modular neural network with two modules of feedforward neural network and also using genetic algorithm to obtain optimal structure for modules of the neural network are the most important indicators of this method.

Keywords—Customer Churn, Neural Network, Genetic Algorithm, Data Mining

I. INTRODUCTION

Since the customer relationship management (CRM) concept is an organizational strategy aimed at improving attraction, retention, loyalty, and profitability of customers by properly understanding their personal information through meaningful communication, this concept has emerged since 90's. Defining customer relationship management from another viewpoint is a comprehensive strategy involving sales, providing customer service, marketing activities integration and supply chain activities to enhance system efficiency and optimization [1]. At present, mostly, telecommunications companies are encounter the challenge of holding down operating costs. In general, the aim of customer churn prediction in CRM is to increase correct predictions and reduce false predictions in customer churn notification warnings at an acceptable level of cost. Simply put, the customer is an organization or person who receives products or services, one who chooses his needs, consumes goods and services produced, he is willing to pay for them. In present's language, customer refers to someone who benefits from services and products of organizations. Telecommunications companies work in fields of customer loyalty and identify potential customers who are leaving the company. Customers churn in this company, despite revenue reduction, entails the cost of attracting new customers. Moreover, it is also effective in losing potential earnings. Recent studies explain, predicting customers who want to leave telecommunications companies, providing solutions to satisfy them & promoting their loyalty are the most important concerns of this company. Losing or Customer Churn is a business term used in a variety of topics such as erosion, divestment or customer defection to lose a customer. Churn is classified into two categories: optional and nonoptional. Optional churn occurs when a customer first leaves the company. Services disruption due to repeated nonpayment of bills on time or for such reasons by the service provider is a nonoptional churn operation [4]. Customer churn reduction as a complex process requires more knowledge in relevant scientific fields. Preventive actions may be taken whenever a customer churn occurs at a telecommunications company or institution. Telecom customer churn prediction is one of the well-known problems of this classification in which a dataset containing various information (attributes) about telecom is used [5]. The aim of the customer churn management operation is to minimize the obtained losses from customer churn and maximize profits from retaining valuable customers [3]. On the other hand, different ways of analyzing these transactions are used to identify customer churn in telecommunications. Although in the past, these transactions were manually analyzed by experts, at present, due to computer development and increasing volume of transactions, automated methods are used. A collection of statistical techniques, machine learning, pattern recognition, artificial intelligence, and various algorithms are used as data mining methods in a variety of fields, including customer churn detection in telecommunications. [5]

Since, customer churn prediction methods in telecommunications include neural networks, decision trees, Bayesian networks, genetic algorithms, etc., however, recent studies in this field explain, the combination of genetic algorithms & neural networks has not used so far in customer churn prediction in the telecommunications industry.

The aim of the present study is to provide a hybrid model based on genetic algorithms and feedforward neural networks for predicting the customer churn in Irancell Company.

Here's a brief overview of what has been done in this field in the past.

Table I Comparison of Customers Churn Methods

R	Source	Method	Positive points	Negative points
1	[6]	Perceptron and Regression	Proposed model can be used online.	Extreme attention in use of this data are essential.
2	[7]	Genetic Algorithm	Reduced costs and processing time	In order to accurately predict a customer churn record of being healthy, it needs to be tested in real terms and to be set limits again.
3	[4]	Multilayer Perceptron Neural Network	The proposed model is tested on three datasets.	The neural network performance is like a black box, analysis within algorithm is not understandable.
4	[9]	Decision Tree and Genetic Algorithm	The high accuracy and high cost of classification	Only predicts customer churn in personal cases.
5	[10]	Decision Tree of Cost-Driven	Its performance is 10 percent faster than other models. The model can detect an anonymous signal error despite identifying customer churn.	If a customer churn record is introduced, an expert is needed to make sure.
6	[9]	Neural network with multi-cost function	High accuracy and low cost of classification	At this step, identifying for an unsupervised state is difficult given momentary profiles to change.
7	[11]	Neural network with variable threshold	Customer churn prediction ability in unbalanced datasets	Limit determination leads some customers who are not in limit range may be mistaken for a customer churn rate.
8	[12]	Profit-Based Neural Network	This modeling with supervisor approach performed better than unsupervised approach.	Therefore, most of time is spent on feature extraction, while there is not enough time to develop model. This proposed model has only been investigated for available data, so future research will need to test new data in broad range.

II. PROPOSED RESEARCH METHOD

Artificial neural networks are used in classification because they have some characteristics such as variation in structure and training algorithms. Although customer churn prediction is one of the classification issues, it aims by training information & customer data to the neural network in form of churn or retention, to be able to predict when encountering new customers based on information learned. As well, a multilayer perceptron structure is the most well-known neural network structure due to its widespread use in classification issues. Should be mentioned, there are two major challenges in using this type of neural network, this tool can work very well if they are fixed. These two challenges are as follows:

- Selecting the proper structure for the neural network: This structure contains the number of hidden layers and the number of neurons in each layer. For example, if a designed neural network for a case has few hidden layers and few neurons, the neural networks may be ineffective. On the other hand, the opposite is also true, if the number of hidden layers and number of neurons per layer was high, the neural network structure will be complicated, which leads to wasting neural network generalization capability and required more time for data training to neural network. [13 and 14].

- Selecting the proper training algorithm: Training for Neural network means finding the right values for parameters of neural network in order to produce the expected output. In the present study, focus on the first challenge, genetic algorithm, one of the evolutionary algorithms, is used to fix this challenge.
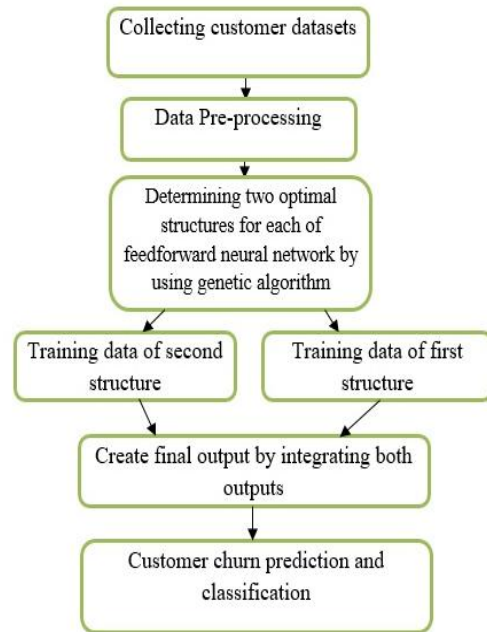

Fig. 1. Proposed Research Method Diagram

The Fig. 1 shows the proposed method diagram, it is explained as follows:

A. Collecting customer data

Customer data are the most important asset in different organizations. The proper analysis of data provide applicable

and useful information for senior executives in organizations. In Customer churn prediction by using the proposed combined method, first, due to Morgan's table, all customers data are sampled over a specific time period to the desired system predict current customer churn by analyzing them.

B. Pre-processing on data

After collecting customer data, pre-processing will be done on data. In this step some data that does not have any impact on result of prediction will be removed from the dataset such as name, surname, national number, etc.; on other hand data such as duration of service use, income from service use & etc. can have a decisive role in likely customer churn prediction. After this, data normalization will be done in this stage of data normalization. This enables data to be converted to new data with new change intervals and proper distribution, as well, normalization reduces data complexity.

C. Determining two optimal structures for two modules of feedforward neural network using genetic algorithm

This is the most important step of the proposed hybrid method. At this step, the genetic algorithm is used to finding the optimal structure for each module. All steps of the genetic algorithm to finding the optimal structure of the neural network are as follows:

Step 1: Determining genetic algorithm parameters and coding chromosomes: First, the values of various parameters in the genetic algorithm such as the number of population members (number of chromosomes), mutation rate, and intersection rate are determined, then chromosomes are coded. Permutation coding is the proposed method for coding chromosomes.

Step 2: Initialization of each population member and determining the merits rate in each chromosome: At first, each chromosome is initialized at random. The lower limit of 0 is used to create variations in the number of layers in random structures. After initializing, the merits rate of each chromosome is determined to obtain the best chromosomes (best neural network structure). So in the process of optimizing algorithm, we use them to find better answers. Therefore, a merit evaluation function is used for each chromosome.

$$(1) \qquad fittns = Mse(Target - Output)$$

In the above-equation (1), Mse is a function in order to compute the mean squared error of neural network after training. Target and Output represent expected output and produced output by the neural network, respectively. Any chromosome whose determined structure by it to be able to minimize mean squared error is the best determined structure for neural network, it can be to be selected as the best population member.

Step 3: beginning Genetic Algorithm Optimization Process: At this step, the genetic algorithm is repeated to the maximum number of replicates rates. In each iteration, the selection steps are performed in three ways: 1- Tournament (Roulette wheel, Random), 2- Intersection (Point-Single method) 3- Mutation

(Switching, Insert and Inverting). After performing these steps, the produced responses are evaluated using the evaluation function. Moreover, it is possible to replace appropriate answers with some of the previous answers in order to prepare a new population for the next iteration of the algorithm.

Step 4: Variation in number of layers modules: Although genetic algorithm by producing different responses creates diversity in number of neurons in each of hidden layers in neural network, it may be low diversity in number of hidden layers, on the other hand, number of layers in neural network has a great impact on producing appropriate responses. Therefore, in the proposed method, using the random method, the aim is to create a suitable variation in the number of hidden layers of the neural network.

Step 5: Selection of two Best chromosomes and network structure formation based on their data: After Optimization Process by Genetic Algorithm, Two of Best Chromosomes (The chromosomes that have most merit rates) are selected. Two feedforward neural networks are created due to them.

D. Training customer data to neural network modules

After applying genetic algorithm are obtained two optimal structures (i.e., the two chromosomes that have most merit). Data is trained to neural networks using the Levenberg–Marquardt training algorithm. In this implementation, a dataset is divided into three parts: training, testing, and validation. As mentioned earlier, the training dataset is used to train customer data to the neural network as long as the neural network predicts based on customer-derived patterns to encounter with new customers. The test dataset is used to evaluate proposed hybrid model performance in predicting customers whose information is unknown by the system. Moreover, the validation dataset is used to solve the overfitting issue. Simply put, the overfitting issue occurs when the neural network maintains a training dataset during a long training process, therefore, it will not have desired results in confronting with the test dataset. Therefore, the use of this dataset causes that with repeated validation of neural network using this data when confronted with this problem to be completed neural network training process. After data splitting, normalization is performed on all three databases.
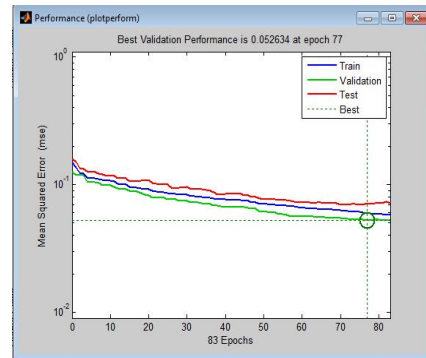

Fig. 2. The convergence diagram of training, testing and validation data

This diagram explains that although appropriate structures for the neural network are obtained by genetic algorithm, Levenberg–Marquardt training algorithm can train the data properly to these neural networks. With this training, the overfitting issue does not occur, and the optimum output in the low number of repetitions will be obtained.

Final Output Production with Output aggregation Unit: Therefore, after completing the above-mentioned steps (i.e., training process and output production by neural networks) use the average method for aggregation, outputs of these two networks become into one output. [19]. Using this type of module will increase the efficiency of neural network output in problem-solving.

$$(1) \quad Output = w \times y1 + w \times y2$$

In above-equation (2), y^1 and y^2 are outputs of first and second modules of neural networks, respectively. The value of w or weight in this equation is 0.5. After that, each module product its output, this output is aggregated using the mean method and the final output is products by the proposed hybrid system and finally, the prediction is performed.

E. Classifying & Identifying Customers likely to Churn:

At this step of the proposed combined algorithm, customer churn prediction is performed due to neural network output. The Knowledge that is produced must be analyzed at the evaluation step to determine its value. In this regard, the Confusion Matrix is one of the methods to evaluate classification systems performance [20]. This matrix is used to show the percentage of correct and incorrect classification.
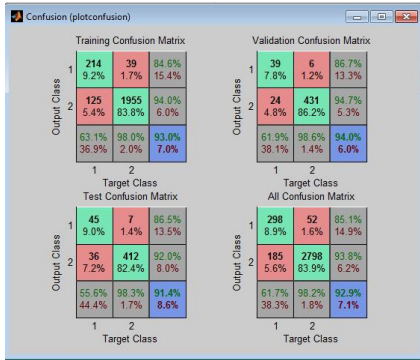

Fig. 3. The Result of Confusion matrix as neural network output

In Fig. 4, confusion matrix derived from the neural network output are shown for the dataset of training, testing, validation, and whole datasets. We compare the results after producing output and integrating proposed hybrid system performance with the results of some other methods used in recent years on similar datasets.

Table II: Comparison of the proposed hybrid method with other methods in customer churn prediction.

Source	Technique used	Recall	Precision	Accuracy
[15]	Neural Networks	81.75%	66.27%	93.20%
[16]	Decision Tree	76.47%	80.60%	77.90%
[16]	Neural Networks	83.90%	83.40%	83.70%
[16]	Support Vector Machines	83.37%	84.20%	83.70%
[17]	Support Vector Machines with RBF kernel- Based	52.12%	81.90%	85.60%
[17]	Support Vector Machines with LIN kernel- Based	39.15%	79.05%	77.90%
[12]	Naive Bayes	54.10%	52.54%	86.94%
[12]	Logistics Regression	70.55%	8.05%	87.94%
[5]	Decision Tree with C5 Algorithm	90.07%	77.60%	95.09%
[5]	Support Vector Machines with POL Kernel-Based	78.45%	91.72%	96.85%
[5]	Neural Network with training algorithm after publishing error & AdaBoost.M1 method	78.17%	84.13%	95.05%
-	First Neural Network with determined structure by Generic Algorithm	92.41%	93.77%	95.30%
-	Second Neural Network with determined structure Generic Algorithm	90.15%	92.86%	94.90%
-	Proposed Hybrid method	94.45%	94.70%	95.50%

III. CONCLUSION

Table (2) explains that the proposed hybrid method based on genetic algorithm and modular neural networks have better performance than other methods. The results show that this proposed hybrid method with two modules of neural network and a genetic algorithm, provide better results than the other methods. The reason for this superiority can be for the following factors:

The obtained optimization structure by genetic algorithm for modular neural network nodules leads to improving neural network performance

Levenberg–Marquardt training algorithm is one of the best neural network training algorithms, which can obtain neural network parameters for data training, in the best way.

The use of modular neural networks with two modules of neural networks and average aggregation for output leads to increased accuracy of classification and precision and recall of the proposed method.

REFERENCES

[1] A. Jamalian, R. Fokurdi, "Application of Data Mining Techniques in Customer Churn Management", Second National Conference on Applied Research, pp. 1 to 4, 2014.

[2] N. Saadollahi Asmari, "Comparison of Data Mining of PSIRC Methodology and RMEES Methodology", Second International Conference on Research in Science and Technology, pp. 3-13, 2015.

[3] S. Ghasemi, M. Omidvarian, M., "Identification and Classification of Influential Variables in Predicting Customer Churn in Telecommunication Industry", International Conference on Modern Research in Industrial Management and Engineering, pp. 20 to 30,2015.

[4] S. Hung, D. Yen, H. Wang, "Applying data mining to telecom churn management", Expert Systems with Applications 31, p 515-524, 2006.

[5] T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K.C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction", Simul. Model. Pract. Theory, vol. 55, pp. 1–9, 2015.

[6] Caigny A. D., Coussement, K., Bock, K., A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees", European Journal of Operational Research, Volume 269, Issue 2, p 760-772, 2018.

[7] Jadhav, R., Usharani T., "Churn prediction in telecommunication using data mining technology", International Journal of Advanced Computer Science and Applications (IJACSA), Volume 2, 2011.

[8] Coussementa, k., Lessmann, S., Verstraeten, B., "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry", Decision Support Systems, Volume 95, p 27-36, 2017.

[9] Sahin, Y., Bulkan, S., Duman, E., "A cost-sensitive decision tree approach for fraud detection, Expert Systems with Applications,Volume 40, Issue 15, p 5916-5923, 2013.

[10] Qureshi, S., Rehman, A., Qamar, A., Kamal, A., Rehman, A., "Telecommunication subscribers' churn prediction model using machine learning", In Eighth International Conference on Digital Information Management (ICDIM 2013), pp. 131–136, 2013.

[11] Salchenberger, L. M., Cinar, M., Lash, N. A., "Neural networks: A new tool for predicting thrift failures, Decision Sciences", Vol. 23, p. 899-916, 2007.

[12] Zakaryazad, A,. Duman, E., "A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing", Neurocomputing, Volume 175, Part A, p 121-131, 2016.

[13] Zhang, X., , "Time series analysis and prediction by neural networks, Optimization Methods and Software", Vol. 4, pp. 151-170, 2015.

[14] Chester, D. L., "Why two hidden layers are better than one?, Proceedings of the International Joint Conference on Neural Networks", Vol. 1, pp. 1265-1268, San Diego, CA, USA, 1990.

[15] Sharma, A., Kumar, P., "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services", Int. J. Comput. Appl., vol. 27, no. 11, pp. 26–31, 2011.

[16] Shaaban, E., Helmy, Y., Khedr, A., Nasr, M., "A Proposed Churn Prediction Model",Int. J. Eng. Res. Appl., vol. 2, no. 4, pp. 693–697, 2012.

[17] Brandusoiu, I., Toderean, G., "Churn Prediction in the Telecommunications Sector using Support Vector Machines", Ann. ORADEA Univ. Fascicle Manag. Technol. Eng., no. 1, 2013.

[18] Hagan, M. T., & Menhaj, M. B., "Training feedforward networks with the Marquardt algorithm", IEEE transactions on Neural Networks, 5(6), 989-993, 1994.

[19] Melin, P. 2012, "Modular Neural Networks and Type-2 Fuzzy Systems for Pattern Recognition", Springer-Verlag Berlin Heidelberg, 2012.

[20] Piri, S., Delen, D., & Liu, T., "A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets", Decision Support Systems, 106, p 15-29, 2018.