ABC Based Neural Network Approach for Churn Prediction in Telecommunication Sector

Priyanka Paliwal(✉) and Divya Kumar

Computer Science and Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, UP, India
priyanka07mnnit@gmail.com, divyak@mnnit.ac.in

Abstract. Customer churn prediction has always been an important aspect of every business. Most of the companies have dedicated churn management teams which work for both churn prevention and churn avoidance. In both of the scenarios it is highly required to identify customers who may change their service providers. In this paper we have tried to propose a neural network based model to predict customer churn in telecommunication industry. We have than used Artificial Bee Colony (ABC) algorithm for neural network training and observed a substantial improvement in accuracy. To prove the efficacy of our model we have compared it against Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Ant Colony Optimization algorithm (ACO). Simulation result shows that ABC trained neural network is more accurate than others in predicting customer churn in telecommunication sector.

Keywords: Customer churn · Neural network · Evolutionary algorithms · Swarm intelligence · Artificial bee colony

1 Introduction

In the present scenario, retaining old customers is as important as attracting the new ones. Total revenues of every industry is majorly dependent on their customer base. This situation is much realistic in telecommunication sector. In today's telecommunication market there is a good availability of service providers, each one trying to offer more and more services to its customers that too at a very reasonable price. Because of this competition, the customers of this sector are more susceptible to churn out. Thus an efficient churn prediction model is always a prime requirement of this sector [1]. To fulfill this purpose the customer relations team always maintain and use behavioral data base of customers for corporate decision making processes. Data-mining techniques and analytical algorithms are applied over this data-base to identify the customers who are at the risk of churning out [2,3]. Churn prediction is thus an interdisciplinary research area which encompasses a through data-base management, data-mining and model development for churn prediction. This problem is further magnified due to the high dimentionality of customer data-base, containing a lot of information about customer demographics, plans availed, complains

S.C. Satapathy and A. Joshi (eds.), *Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 2*, Smart Innovation, Systems and Technologies 84, DOI 10.1007/978-3-319-63645-0_38

raised, usage and billing etc. Although the problem is very complex but keeping in view the loss that may incur due to retention, a good number of statistical as well as meta-heuristic churn prediction models are proposed in the past.

In literature a lot of work is available on Logistic Regression [4,5], Decision Trees [6,7], Artificial Neural Network [3,8] and Support Vector Machine [9–11] models for churn prediction. These models are particularly successful and are most focused by the research scholars working in the domains of churn identification [12]. Artifical Neural Network (ANN) [13] is a simple and robust technique which is based on the structure and working of human brain. Its utility in machine learning, classification and pattern recognition is already well-proven [14,15]. A feed-forward neural network works with the help of neurons organized in a group of layers. These neurons are connected through weights which are to be optimized at the time of learning/training phase. Once the neural network is trained it can be used for classification. Tsai et al. in [3], Sharma et al. in [8], Parag in [16] and Song et al. in [17] and many others have used artificial neural networks in their research regarding churn prediction. In all the cases, authors have claimed that their ANN based models exhibits efficiency and substantial accuracy in churn case identifications.

Usually an ANN is trained through back-propagation algorithms [18]. However, latest trends of research shows that more sophisticated training can be provided to an ANN via Evolutionary and Swarm Intelligence algorithms [19–21]. The set of Evolutionary Algorithms (EA) majorly includes Genetic Algorithms (GA) [22], Evolutionary Strategies [23], Evolutionary Programming [24] and Genetic Programming (GP) [25] on the other hand Particle Swarm Optimization (PSO) [26], Ant Colony Optimization (ACO) [27] and Artificial Bee Colony (ABC) [28] are the most studied Swarm Intelligence (SI) based algorithms. ABC is a relatively new Swarm Intelligence optimization algorithm which based on the foraging behavior of honey bees swarms. It has been extensively used in the past as an efficient optimization technique [12]. Optimizing the neuron's weight in neural networks is one among various applications of ABC [29].

In this manuscript we have tried to show that ABC trained neural network can be efficiently used for customer churn prediction in telecommunication industry. To meet our objective the rest of the paper is organised as follows: In Sect. 2, we have described the churn prediction problem with its parameters and then have proposed a simple neural network based model to predict customer churn in telecommunication industry in Sect. 3. In the last sections we have proved the efficacy of our algorithm by comparing the results of ABC trained ANN with GA, PSO, ACO. It has been observed that the proposed ABC based ANN training algorithm considerably improves the performance of traditional back-propagation trained ANN.

2 Problem Statement

Customer churn prediction is the term used to determine the customers who can churn in near future, from a given service provider. Customer retention is one of the elementary characteristic of Customer Relationship Management (CRM) because it is always profitable to keep existing customers besides attracting the new one.

In the churn prediction problem, for some n customers associated with a telecom company within last month, we have been provided with their *churn status* (whether they have left the company or not) and a data set containing the information regarding their *age*, *gender*, *marital status*, *dependents status*, *tenure*, *payment method*, *services availed*, *technical support*, *type of contract* and *monthly charges*. We have used the data-set provided by IBM Watson Analytics and for more information regarding this data-set we redirect the readers to [30]. From this data base we have to discover the patterns or associations for identifying the customers who may churn out. For this purpose a neural network based model is required to be proposed. *error ratio*, as defined in Eq. (1) is used to predict the accuracy of the model. It measures the number of falsely classified customers out of all n predictions. A customer is falsely classified if he/she is a churner classified as non-churner ($Case\,B$) or vice-versa ($Case\,C$). The values of $|A|$, $|B|$, $|C|$ and $|D|$ can be found using confusion matrix [3] as shown in Table 1.

$$error\ ratio = \frac{|B| + |C|}{|A| + |B| + |C| + |D|} \tag{1}$$

Table 1. Confusion matrix

		Actual	
		Churner	Non-churner
Predicted	Churner	*Case A*	*Case B*
	Non-churner	*Case C*	*Case D*

3 Proposed Methodology

Inspite of having various mathematical methods for classification like Support Vector Machines, Bayesian Networks, Adaboost and Decision Trees, we have used evolutionary approach. This is because of the motivation gained from the success of evolutionary approaches on other classification problems [31,32]. To tackle this problem we have used two Artificial Neural Network based approaches. The first approach uses the traditional ANN with back-propagated error based learning. The second approach uses an Artificial Bee Colony tuned ANN. The details of both the techniques are presented in the next sub-sections.

3.1 Back-Propagation Trained Neural Network

In this approach, Back-propogation algorithm [18] is used to train the neural network. The neural network which is trained, acts on 10 inputs for 1 output.

It contains 2 hidden layers, each with 5 neurons, thus there exists a total of 80 neuron weights in this network. For each sample input 20 epochs are used for training with log-sigmoid activation function.

3.2 ABC Trained Neural Network

For training a neural network with ABC an initial population P of size 200 is formed first. Each population member $P_i\, i \in \{1, \dots 200\}$ is a 80 dimensional vector of real numbers in the range $(0, 1)$ i.e. $P_i[p_{i1}, p_{i2}, \dots, p_{i80}]$, $j \in \{1, \dots, 80\}$, $p_{ij} \in \mathbb{R}$ *and* $0 < p_{ij} < 1$. In this 80 dimensional vector, each dimension represents a corresponding neural network weight between two neurons.

Once the initial population is formed ABC algorithm come into action. In our experiments we have used ABC algorithm as described in [28] to evolve the neural network weights. This ABC model consists of three types of bees: employed, onlookers and scouts. Employed and onlooker bees represents exploration and exploitation of the search space. The timbal bees turns into scouts which are randomly re-initialized. This model delineates two leading modes of behavior of bees which are sufficient for self-organization and coordinated decision making, *(1)* Recruitment of more foragers towards rich food sources searched by forager bees (positive feedback). *(2)* Probabilistic desolation of forager bees which have found poor food sources (negative feedback). Fitness assignment to artificial bee is an important issue in ABC algorithm. In the proposed approach fitness of a chromosome P_i is directly proportional to the number of cases correctly classified when neural network weights are according to the chromosome values. For n trials, i.e. while using a data-set of n different customers, over a chromosome P_i, Eq. (3) is used for fitness evaluation of P_i.

$$raw\_fitness(P_i) = \sum_{i=1}^{n} \begin{cases} 1, & for\, Case\, A\, and\, D, \\ -1, & for\, Case\, B\, and\, C, \end{cases} \tag{2}$$

$$fitness(P_i) = \frac{raw\_fitness(P_i)}{n} \tag{3}$$

4 Results

All the experiments are done on machines running $Matlab\, 2014a$ over $Windows\, 7$, 4 GB RAM and $corei5$ processor. *error ratio*, as defined in Eq. (1) is used to compare different algorithms. The available data-set [30] contains the information base of 7000 customers. Out of these 7000 entries, 5600 are used for training and 1400 are used for testing purposes. Each algorithm (Back-propagation [18], ABC [28], GA [22], PSO [26] or ACO [27]) is executed 5 times, each time with a fresh start. The best reading is chosen to be recorded. From our simulation results, we observed that ABC trained neural network is most accurate with an *error ratio* of 0.11 and Back-propagation is found to be least accurate with 0.19 *error ratio*. PSO came out to be second accurate algorithm with

0.13 *error ratio*. ACO and GA trained neural network exhibits an *error ratio* of 0.18 and 0.16 respectively (Fig. 1).

For further in-depth analysis of ABC algorithm, we have plotted the accuracy trends of ABC algorithm in Fig. 2. This figure shows the variations in the *error ratio* readings taken at different iterations. The *error ratio* apparently decreases from 0.75, as the ABC algorithms converges iteration by iteration and a minimum *error ratio* of 0.11 is obtained at iteration number 180. Thus *error ratio* has decreased nearly 7 times. An increase in the *error ratio* can also be observed after iteration number 180. This is due to the notification of the

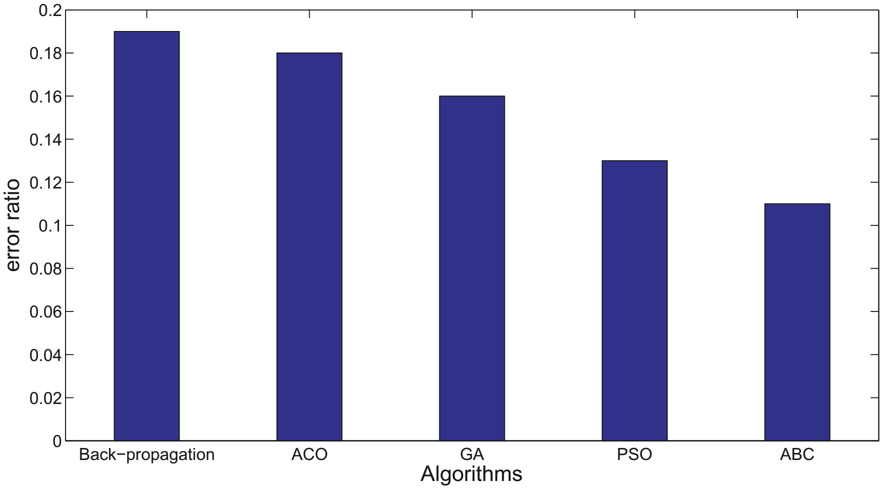

Fig. 1. *error ratio* obtained from different neural networks, trained through Back-propagation, ACO, GA, PSO and ABC.

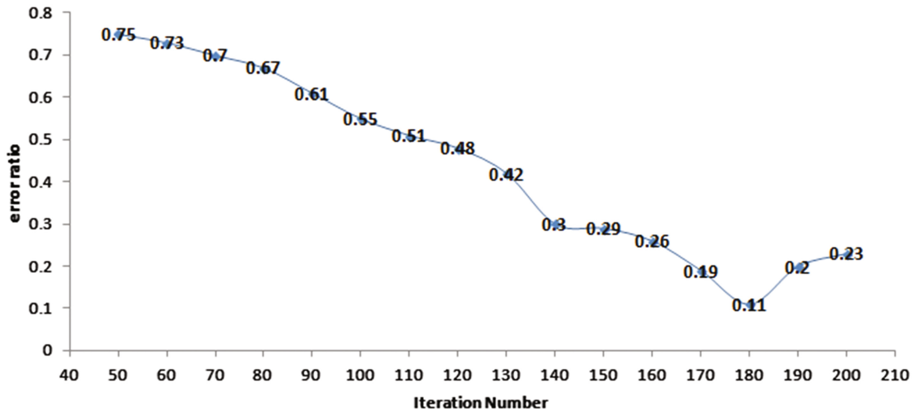

Fig. 2. *error ratio* obtained at different iterations, while running ABC algorithm. *error ratio* is minimum (0.11) at iteration number 180.

occurrence of scout bees, to be reinitialized, who had already converged in the previous iterations.

5 Conclusion

This manuscript has focused on the churn prediction aspect of telecommunication industry. We have first explained the importance of churn prediction for telecommunication and then described the role of Artificial Neural Networks (ANNs) on churn prediction. We have then proposed a Artificial Bee Colony (ABC) based model for training ANN. The ABC trained neural network is then used to predict the customers who are at a risk of churn in near future. To prove the validity of out model, we have compared the accuracy of ABC trained neural network with the neural networks trained through Back-propagation, ACO, GA and PSO. The experimental results clearly validate that ABC trained neural network out performs others.

References

1. Idris, A., Rizwan, M., Khan, A.: Churn prediction in telecom using random forest and PSO based data balancing in combination with various feature selection strategies. Comput. Elect. Eng. **38**(6), 1808–1819 (2012)
2. Au, W.-H., Chan, K.C., Yao, X.: A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Trans. Evol. Comput. **7**(6), 532–545 (2003)
3. Tsai, C.-F., Lu, Y.-H.: Customer churn prediction by hybrid neural networks. Expert Syst. Appl. **36**(10), 12547–12553 (2009)
4. Nie, G., Rowe, W., Zhang, L., Tian, Y., Shi, Y.: Credit card churn forecasting by logistic regression and decision tree. Expert Syst. Appl. **38**(12), 15273–15285 (2011)
5. Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression, vol. 398. Wiley, Hoboken (2013)
6. Bin, L., Peiji, S., Juan, L.: Customer churn prediction based on the decision tree in personal handyphone system service. In: 2007 International Conference on Service Systems and Service Management, pp. 1–5. IEEE (2007)
7. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
8. Sharma, A., Panigrahi, D., Kumar, P.: A neural network based approach for predicting customer churn in cellular network services. arXiv preprint arXiv:1309.3945 (2013)
9. Coussement, K., Van den Poel, D.: Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. Expert Syst. Appl. **34**(1), 313–327 (2008)
10. Zhao, Y., Li, B., Li, X., Liu, W., Ren, S.: Customer churn prediction using improved one-class support vector machine. In: International Conference on Advanced Data Mining and Applications, pp. 300–306. Springer, Heidelberg (2005)
11. Vladimir, V.N., Vapnik, V.: The Nature of Statistical Learning Theory (1995)
12. Kamalraj, N., Malathi, A.: A survey on churn prediction techniques in communication sector. Int. J. Comput. Appl. **64**(5) (2013)

13. Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O.: Neural Network Design, vol. 20. PWS publishing company, Boston (1996)
14. Basheer, I., Hajmeer, M.: Artificial neural networks: fundamentals, computing, design, and application. J. Microbiol. Meth. **43**(1), 3–31 (2000)
15. Zhang, G.P.: Neural networks for classification: a survey. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **30**(4), 451–462 (2000)
16. Pendharkar, P.C.: Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. Expert Syst. Appl. **36**(3), 6714–6720 (2009)
17. Song, G., Yang, D., Wu, L., Wang, T., Tang, S.: A mixed process neural network and its application to churn prediction in mobile communications. In: Sixth IEEE International Conference on Data Mining-Workshops, ICDMW 2006, pp. 798–802. IEEE (2006)
18. Hecht-Nielsen, R.: Theory of the backpropagation neural network. In: International Joint Conference on Neural Networks, IJCNN 1989, pp. 593–605. IEEE (1989)
19. Bullinaria, J.A., AlYahya, K.: Artificial bee colony training of neural networks. In: Nature Inspired Cooperative Strategies for Optimization, NICSO 2013, pp. 191–201. Springer, Heidelberg (2014)
20. Karaboga, D., Kaya, E.: An adaptive and hybrid artificial bee colony algorithm (aABC) for ANFIS training. Appl. Soft Comput. **49**, 423–436 (2016)
21. Branke, J.: Evolutionary algorithms for neural network design and training. In: Proceedings of the First Nordic Workshop on Genetic Algorithms and its Applications. Citeseer (1995)
22. Goldberg, D.E.: Genetic Algorithms. Pearson Education, India (2006)
23. Rechenberg, I.: Evolution strategy. In: Computational Intelligence: Imitating Life, vol. 1, pp. 147–159 (1994)
24. Fogel, L.J.: Intelligence Through Simulated Evolution: Forty Years of Evolutionary Programming. Wiley, Haboken (1999)
25. Koza, J.R.: Genetic Programming II: Automatic Discovery of Reusable Subprograms. MIT Press, Cambridge, MA, USA (1994)
26. Kennedy, J.: Particle swarm optimization. In: Encyclopedia of Machine Learning, pp. 760–766. Springer, Heidelberg (2011)
27. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. IEEE Comput. Intell. Mag. **1**(4), 28–39 (2006)
28. Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J. Global Optim. **39**(3), 459–471 (2007)
29. Karaboga, D., Gorkemli, B., Ozturk, C., Karaboga, N.: A comprehensive survey: artificial bee colony (ABC) algorithm and applications. Artif. Intell. Rev. **42**(1), 21–57 (2014)
30. Analytics, I.W.: Using customer behavior data to improve customer retention. https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/
31. Dasgupta, D., Michalewicz, Z.: Evolutionary Algorithms in Engineering Applications. Springer, Heidelberg (2013)
32. Freitas, A.A.: Data Mining and Knowledge Discovery With Evolutionary Algorithms. Springer, Heidelberg (2013)