Indian Journal of Science and Technology, Vol 9(33), DOI: 10.17485/ijst/2016/v9i33/99103, September 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Effective Customer Churn Prediction on Large Scale Data using Metaheuristic Approach

K. Sivasankar*

Department of Computer Applications, National Institute of Technology, Trichy - 620015, Tamil Nadu, India; sivasankar_kk@yahoo.com

Abstract

Objectives: Customer retention is one of the major requirements of any organization to gain competitive advantage. Accurately predicting the customer's status can help organizations reduce and prevent churns. **Methods/Analysis:** This paper presents an analysis of churn data and issues related to churn data in terms of data size, attribute density, data sparsity and abstraction contained in the data. It discusses the advantages of utilizing metaheuristic techniques for churn prediction and in specific analyses ACO for churn prediction and performs a comparison with other metaheuristic algorithms and emphasizes the importance of using ACO. **Findings:** Experiments were conducted by implementing ACO and applying it on Orange Dataset. It was observed from the ROC curve that the points plotted falls to the top left of the graph, hence indicating good efficiency and a fluctuation from low to moderate false positive rates were observed. It could be observed from the PR curve that the ACO algorithm exhibits high recall rates and moderate precision rates. ROC and the PR plots indicate that there is still scope for enhancement in terms of reduction in false positive rates and increase in precision levels. It was identified that though ACO exhibits effective performance, the size of the dataset acted as a huge downside increasing the time taken. Due to the huge size of the data, memory requirements are very high, but due to the skewed nature of the data most of them contain null values. **Applications/Improvement:** Findings exhibited scope for improvement, hence research directions namely data structure identification to reduce memory requirements, graph based churn prediction and fuzziness incorporation in the prediction process were proposed.

Keywords: ACO, Churn Prediction, Churn Prevention, Classification, Graph Models

1. Introduction

Customer churn prediction is one of the key factors in maintaining customer value for an enterprise. Increase in the amount of data generated and a better storage capability has ensured that all the generated data is stored and available for mining[1,2]. This data corresponds to the customer's usage patterns, will provide valuable insights on their behavior. Churn is defined as the tendency of a customer to cease their operations with an organization. This most probably occur either due to dissatisfaction with the current organization or an alluring offer from a competitor. It is often quoted that obtaining a new customer is 5 to 6 times costlier than maintaining an existing customer[3]. Hence, predicting churn has become one of the inevitable requirements of today's competitive business scenario[4]. Customer churn is a common occurrence in both service and product related areas. The probability of occurrence of churn is inversely proportional to the cost involved and the complexity of the migration process. Customer dissatisfaction remains to be the major contributor for the occurrence of churn[5]. Identifying customers with dissatisfactions and identifying the reasons for dissatisfaction can reduce customer churn to a large extent[6].

The most commonly used techniques for churn prediction include artificial neural networks, support vector machines, logistic regression and decision trees[7–10]. A feature selection based dynamic transfer ensemble model for predicting churn was presented[11]. This method also incorporates the theory of transfer learning in the domain of churn prediction. Feature selection is performed in two phases; the first phase selects the initial feature subset and the next phase identifies the

*Author for correspondence

mandatory properties and filters them. A tree based model that combines the advantages of ADT and logistic regression was presented[12]. This method uses feature selection to identify the best attributes and builds the tree. Tree construction is performed entirely on the basis of customer characteristics. A B2B based churn prediction in logistics industry was presented[13]. This method uses an ensemble of classifiers (C4.5, MLP, SVM and LGR) to predict churn. A similar ensemble based churn prediction method was presented[14]. This method uses Random Forest, Rotation Forest, RotBoost and DECORATE in combination with minimum redundancy and maximum relevance (mRMR). A Markov model based churn prediction model was presented[15]. Several application based churn prediction models have also been proposed in recent years. Bank based churn model[16,17], insurance based churn prediction[18,19], customer churn prediction in telecommunications[20, 21] etc.

2. Churn Data: Issues

Churn is the tendency of a customer to cease doing business with an organization. Churn data is customer based, hence, contains all the properties that are related to a customer, both existing and past. Customer churning exists in all product and service related organizations. The following discussion presents the issues related to churn data and problems faced due to these intrinsic properties.

2.1 Data Size

Churn data deals with the customer's properties corresponding to an organization. Hence, it has representations of all the customers operating with and operated with the organization. Hence, churn data is very huge in terms of the number of records especially in service related areas such as telecom and e-commerce, where migration is less complex from all aspects.

2.2 Attribute Density

Churn data should represent every property of a customer corresponding to the product or service that they utilize. Not all customers opt for the same type of service or product. Hence every property of every product/service dealt by the organization should be represented in the churn data. As the number of product/service categories increase, the attributes also tends to increase. This leads to the data structure representing churn containing a large number of attributes. Considering the telecom industry, a mobile service provider can offer several packages dealing with voice calls, data, SMS, social networking based packages, etc. All these properties must contain representations in the data set, even though a customer interested in one package might not be interested in others.

Increase in data size in terms of records and attributes will lead to the final data structure becoming very huge and eventually Big Data.

2.3 Sparse Data/Missing Values

As discussed previously, churn data contains a large number of attributes. Practically, not all customers will be related to all the products/services offered by the organization. Hence only the properties corresponding to the customer's interests will be filled in with appropriate values, while all the other properties contain null values. Hence churn data contains very sparse data with large number of missing values. But eliminating these properties is not possible, as the columns with missing values corresponding to one customer will contain valid entries for another customer. Hence eliminating the levels of data sparsity or missing values is not possible. The prominence of this section is observable in the telecom data sets. For example, a customer totally inclined towards voice call will have null entries in the data section. This process will automatically create an attribute irrelevancy scenario, where an attribute is totally irrelevant to a customer, but it is still present as a part of the customer's properties.

2.4 Privacy Vs Abstraction

Data privacy tends to be one of the major requirements due to the increased information associations in the form to data fusion. Maintaining abstraction in the customer data when utilizing it for analysis has become mandatory for an organization. However, this process tends to reduce the usability of the data. As the level of abstraction increases, user privacy increases, but usability of the data starts to reduce. Similarly reduced abstractions will provide an information rich data, but user privacy would be at risk. It is mandatory to maintain a balance between the level of abstraction applied on the data and the usability of the data.

2.5 Dataset Analysis

Orange Small and Orange Large datasets correspond to the French Telecom company's churn data[19]. This is a benchmark

data, also provided as a part of KDD 2009 challenge[22]. The properties of Orange data are presented in Table 1.

3. Metaheuristic Optimization for Churn Prediction in Big Data

Metaheuristic algorithms aimed at optimization have been in use for a long time, but their popularity has increased only recently. Though metaheuristics have been applied in several early studies, their results were not promising, hence their utilization was limited. This could be attributed to the size of the data being analyzed. In the earlier times, the data size was moderate, hence statistical techniques were able to produce effective results which were much better than the metaheuristic techniques. Since metaheuristic techniques were able to produce only optimal results and not the best results, their utilization scenarios were sparse.

Due to the increase in the amount of data generated, the efficiency of the statistical techniques is now being overshadowed by the huge amount of time taken by them to produce solutions[27]. Hence the need for metaheuristic algorithms has been realized. The small error that occurs as an intrinsic part of the solutions provided by metaheuristics has become acceptable in most applications whose major requirements is to provide faster results. The need for online processing has motivated the use of metaheuristics to a large extent[28]. Increased processing capacity from the hardware aspect has also enabled better and more effective processing, in turn increasing the accuracy of these optimization techniques. This paper uses Ant Colony Optimization (ACO) to identify probable customers for preventing churn.

4. ACO based Churn Prediction

4.1 Ant Colony Optimization: Working

Ant Colony Optimization (ACO)[23,29,30] is a metaheuristic technique that can be used to solve optimization problems.

Table 1. Dataset analysis

Property	Orange Small	Orange Large
Attribute density	230	15000
No of Records	50000	50000
Missing Values	60%	60%
No of Numerical Attributes	190	14740
No of Categorical Attributes	40	260

It has its inspiration from the food collecting behavior of the ants. It operates using a set of agents (ants) to identify the best solution in the search space. The agents contained in the ant system are primitive, however intelligence is brought about by communication and information sharing. Information is shared by the process of reinforcing the pheromone trail. Pheromone is a chemical deposited by the ant on the best path identified by it. This deposition, reinforcement and evaporation cycle of the pheromone deposits enables to identify the best solution in the search space.

Consider m to be the number of ants and n to be the number of nodes in the network, then the probability that an ant mi will select a node is given by

$$p_{ij}(t) = \frac{[\tau_{ij}(t)]^{\alpha}.[\eta_{ij}]^{\beta}}{\sum_{j=1}^{n}[\tau_{ij}(t)]^{\alpha}.[\eta_{ij}]^{\beta}} \quad (1)$$

where τ_{ij} is the pheromone intensity in the edge ij and η_{ij} is the visibility range of the edge ij. α and β are the weights provided to the pheromone trial and the visibility respectively. Every time an ant needs to make a decision on selecting a node, the probability of all the nodes are identified using Equation (1) and Cumulative Distribution Function (CDF) is used to identify the destination node. Using CDF ensures that the ant also has scope for exploration, rather than using only the pre-identified paths (exploitation). After every movement, the amount of pheromone to be deposited on the edge and the trail intensity of the edge is given by Equation (2) and (3).

The amount of pheromone to be deposited and the point at which the pheromone is to be deposited differs on the basis of the variant that is being used. The Ant Density model deposits a total pheromone of quantity Q (determined by the user according to the application) on the path that has been selected.

$$\Delta\tau_{ij}^{k}(t, t+1) = \begin{cases} Q & \text{if } k-\text{th ant goes from } i \text{ to } j \text{ between } t \text{ and } t+1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Let $\tau_{ij}(t+1)$ be the intensity of trail on pathij at time $t+1$, given the formula

$$\tau_{ij}(t+1) = \rho.\tau_{ij}(t) + \Delta\tau_{ij}(t, t+1) \quad (3)$$

$\Delta\tau_{ij}$ is the pheromone level to be deposited on the path ij and ρ is the evaporation parameter.

This process is continued till the termination criterion is met and the final solution is obtained by identifying the best solution among all the solutions generated by the ants.

4.2 ACO Vs Other Metaheuristic Approaches

Several metaheuristic techniques such as Particle Swarm Optimization (PSO), Firefly, Bee Colony and Artificial Bee Colony Optimization (ABC) etc., exists in literature. On analysis it could be identified that most of the metaheuristic techniques are similar in working, in the sense one could be considered as a variant of another[25,26]. For instance, Firefly algorithm could be converted to standard PSO by setting the absorption co-efficient to zero and replacing the inner loop by the current global best.

ACO was the first concrete algorithm to be presented in the field of metaheuristics, while most of the metaheuristic techniques provided just an idea based on the analogy from nature. The analogy of ants solving an optimization problem was strong and quite simple to understand. Further, convergence of ACO has been analytically proved by Gutjhar[24], whereas most other algorithms are still being used based on experiments rather than mathematical proofs. Another major advantage is that ACO operates on a discrete search space and is dynamic in nature, which coincides with the churn prediction problem. While PSO operates in a continuous space, Firefly algorithm requires too many computations in the order of $O(n^2)$, where n is the number of records (cities) contained in the search space. Since churn data involves too many customers, the operating time becomes unacceptable. Hence, ACO has been our choice of algorithm to operate on churn data.

5. Experiments and Discussion

Experiments were conducted by implementing ACO using C#.Net on an Intel Core i7 (Quad Core) machine with 16GB RAM. Orange small dataset was used for the classification process.

Receiver Operating Characteristics (ROC) plot obtained from the test data is presented in Figure 1. It could be observed that the True Positive Rate (TPR) remains high representing effective classification of the positive entries. The False Positive Rate (FPR) however exhibits a maximum level of 0.45. Situation of points in the top left region of the ROC plot indicates good efficiency in classification algorithms. Though the false positive rate fluctuates from low to moderate, it could be observed that the points plotted falls to the top left of the graph, hence indicating good efficiency.

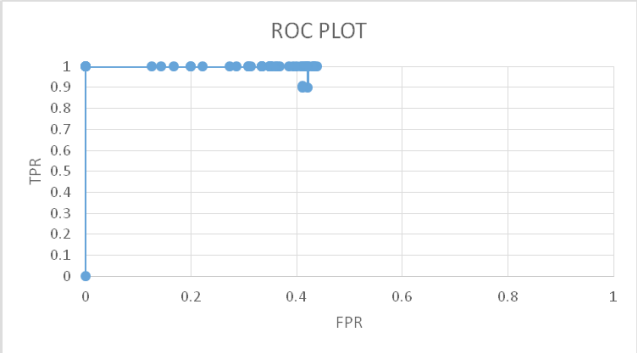

Figure 1. ROC plot.

Precision Recall (PR) Curve obtained from the test data is presented in Figure 2. It could be observed that the ACO algorithm exhibits high recall rates and moderate precision rates. Efficiency of the algorithms are measured with the density of the point plotted in the PR Plot in the top right region. ACO exhibits an accuracy level of 66% and an F-Measure of 0.55.

ROC and the PR plots indicate that there is still scope for enhancement in terms of reduction in false positive rates and increase in precision levels. It was identified that though ACO exhibits effective performance, the size of the dataset acted as a huge downside increasing the time taken. Due to the huge size of the data, memory requirements are very high, but due to the skewed nature of the data most of them contain null values. The next section discusses research directions and algorithm enhancements to eliminate all the downsides of the current method to improve the accuracy levels and reduce the time and memory requirements.

6. Research Directions

The memory requirements of ACO indicate that the churn data requires a huge amount of memory, however the dataset analysis in Section 2.5, indicates that size of the usable data constitutes only 40% of the total data. This is attributed to the customer's irrelevancy towards most of the attributes in the dataset. Hence it could be concluded that unstructured storage of such data will lead to a huge reduction in the memory requirements. Hence an unstructured data storage scheme would be much suited compared to a structured storage for churn prediction.

Property graphs that depict data in the form of nodes and edges are the preferred structures for processing churn data. A property graph is a connected structure containing

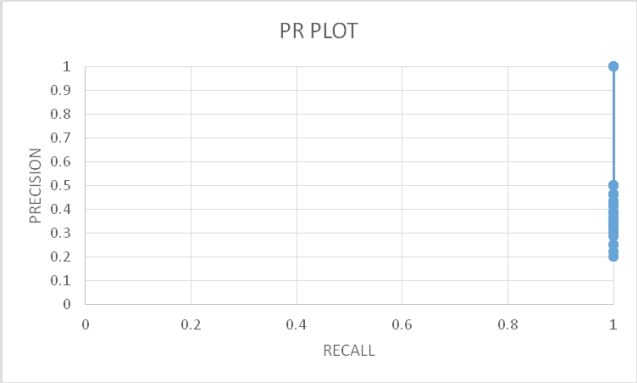

Figure 2. PR curve.

nodes and edges. Nodes contain several properties corresponding to the attributes of the node and the edges connecting two nodes can also have properties associated with them. Properties of nodes can vary in number and type, hence unstructured connected units can be created using property graphs. This tends to reduce the memory requirements to a large extent. Another major advantage is that ACO can operate well on graph data. Our research direction is to create a graph based structure and to provide optimal predictions using ACO.

Graph based representations of data has several advantages. In and out degree based analysis can reveal properties of importance. These properties can be used to incorporate specific property based weights, leading to better prediction. Extraction and aggregation of the plans corresponding to positive churn and plotting them in comparison with the competitor's plans can provide a deeper understanding of the reasons for a customer to move from one plan to another. Identifying the attractiveness of plans/features and the level of attractiveness could be identified by customer grouping. Levels of attractiveness required for each category of customer could be identified. The process of fuzzy classification can be incorporated into the model to identify the level of positive churn and in-turn use the properties and plans identified using the property graphs to prevent churn.

7. Conclusion

This paper presents a metaheuristic technique to predict customer churn. An analysis of churn data and its components is performed and the properties pertaining to churn data were identified. The reason for choosing ACO over other metaheuristics is discussed. The experimental results relating to accuracy of the classification process and the time taken to complete the process and the pros and cons related to this approach are discussed. The research directions are proposed based on the data analysis and the experiments conducted using ACO.

8. References

1. Delafrooz N, Farzanfar E. Determining the customer lifetime value based on the benefit clustering in the insurance industry. Indian Journal of Science and Technology. 2016 Jan; 9(1):1–8.
2. Dash P, Pattnaik S, Rath B. Knowledge Discovery in Databases (KDD) as tools for developing customer relationship management as external uncertain environment: A case study with reference to State Bank of India. Indian Journal of Science and Technology. 2016 Jan; 9(4):1–11.
3. Bhattacharya CB. When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science. 1998 Jan; 26(1):31–44.
4. Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH. Defection detection: Measuring and understanding the predictive accuracy of customer churn models. Journal of Marketing Research. 2006 May; 43(2):204–11.
5. Jana A, Chandra B. Mediating role of customer satisfaction in the mid-market hotels: An empirical analysis. Indian Journal of Science and Technology. 2016 Feb; 9(1):1–16.
6. Boroumandzadeh M, Mirsarraf MR, Movaghar A. Investigating the customer care based on enhanced telecom operation map standard in third generation of mobile networks. Indian Journal of Science and Technology. 2015 Sep; 8(22):1–6.
7. Au WH, Chan KC, Yao X. A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Transactions on Evolutionary Computation. 2003 Dec; 7(6):532–45.
8. Kisioglu P, Topcu YI. Applying bayesian belief network approach to customer churn analysis: A case study on the telecom industry of Turkey. Expert Systems with Applications. 2011 Jun; 38(6):7151–7.
9. Pendharkar PC. A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. Computers and Operations Research. 2005 Oct; 32(10):2561–82.
10. Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: A data mining approach. Expert systems with applications. 2002 Aug; 23(2):103–12.
11. Jin X, Yi X, Anqiang H, Dunhu L, Shouyang W. Feature-selection-based dynamic transfer ensemble model for customer churn prediction. Knowledge and Information Systems (Impact Factor: 1.78). 2014 Jan; 43(1):29–51.

12. Qi J, Zhang L, Liu Y, Li L, Zhou Y, Shen Y, Liang L, Li H. ADTreesLogit model for customer churn prediction. Annals of Operations Research. 2009 Apr; 168(1):247–65.
13. Chen K, Hu YH, Hsieh YC. Predicting customer churn from valuable B2B customers in the logistics industry: A case study. Information Systems and e-Business Management. 2015 Aug; 13(3):475–94.
14. Idris A, Khan A, Lee YS. Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost based ensemble classification. Applied Intelligence. 2013 Oct; 39(3):659–72.
15. Migueis VL, Van den Poel D, Camanho AS, e Cunha JF. Predicting partial customer churn using Markov for discrimination for modeling first purchase sequences. Advances in Data Analysis and Classification. 2012 Dec; 6(4):337–53.
16. Kumar DA, Ravi V. Predicting credit card customer churn in banks using data mining. International Journal of Data Analysis Techniques and Strategies. 2008 Jan; 1(1):4–28.
17. Lariviere B, Van den Poel D. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. Expert Systems with Applications. 2004 Aug; 27(2):277–85.
18. Zeithaml VA, Berry LL, Parasuraman A. The behavioral consequences of service quality. The Journal of Marketing. 1996 Apr; 60(2):31–46.
19. Morik K, Kopcke H. Analysing customer churn in insurance data– A case study. Knowledge Discovery in Databases: PKDD 2004. Springer Berlin Heidelberg; 2004 Sep. p. 325–36.
20. Hung SY, Yen DC, Wang HY. Applying data mining to telecom churn management. Expert Systems with Applications. 2006 Oct; 31(3):515–24.
21. Hwang H, Jung T, Suh E. An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. Expert Systems with Applications. 2004 Feb; 26(2):181–8.
22. KDD Dataset. Available from: http://www.kdd.org/kdd-cup/view/kdd-cup-2009/Data.
23. Dorigo M, Maniezzo V, Colorni A. Ant system: Optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 1996 Feb; 26(1):29–41.
24. Gutjahr WJ. A graph-based ant system and its convergence. Future Generation Computer Systems. 2000 Jun; 16(8):873–88.
25. Prakasam A, Savarimuthu N. Metaheuristic algorithms and probabilistic behaviour: A comprehensive analysis of Ant Colony Optimization and its variants. Artificial Intelligence Review. 2016 Jan; 45(1):97–130.
26. Prakasam A, Savarimuthu N. Metaheuristic algorithms and polynomial turing reductions: a case study based on ant colony optimization. Procedia Computer Science. 2015 Dec; 46:388–95.
27. Bottou L. Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT'2010 Physica-Verlag HD; 2010 Sep. p. 177–86.
28. Murata N. A statistical study of on-line learning. Online Learning and Neural Networks. Cambridge, UK: Cambridge University Press; 1999. p. 63–92.
29. Dorigo M, Gambardella LM. Ant colony system: A cooperative learning approach to the traveling salesman problem. IEEE Transactions on Evolutionary Computation. 1997 Apr; 1(1):53–66.
30. Dorigo M, Maniezzo V, Colorni A. The ant system: An autocatalytic optimizing process. Technical Report; 1991. p. 1–21.