An Effective Architectural Model for Early Churn Prediction – NELCO

R. Manivannan, R. Saminathan, S. Saravanan

Abstract: Customer is an asset of any business organization, whose probable chances of churn is loss. Several challenges are to be considered towards controlling customer churn. Machine learning approach is needed to predict an early churn. Even though various soft computational approaches had been proposed, an optimized computational approach which identifies early churn prediction is necessary. The proposed approach NELCO predicts early customer churn using Negative Correlation Learning (NCL) which uses k-means neighbourhood discriminant similarity indices over network of ensemble values. NELCO proves to have an optimal accuracy towards early prediction of churn, as well as suggests that customer retention rate is higher over PSO, ACO approaches.

Keywords: Negative Correlation Learning, Early prediction, customer churn, customer retention.

I. INTRODUCTION

Customer is considered as king of any business organization[19], and loss factor or churn of a customer leads to imbalance of business institute [8][11], which is reported as loss. This paper works on Negative Correlation Learning (NCL) as a major approach towards early prediction chances of customer churn and suggesting on customer retention such that loss is avoided. Major business organizations consider customer churn as a chaotic event [4] [12], which needs to be predicted early in terms of retaining the customer such that the cost incurred towards bringing in new customers is minimized. Customer churn attributes to lot of aspects, which includes product unexpected cost, quality, other demographic features and competitive advantages compared to other business groups.

NCL can be defined as a neural network ensemble learning algorithm [17], which introduces a correlation penalty term to the cost function of each individual network so that each neural network minimizes its mean-square-error (MSE) together with the correlation factor ‘ α ’ , a scaling coefficient penalty factor ‘ δ ’ whose update interval is ‘ δ ’ as error function[20].This paper work discusses about NCL which is considered as an efficient ensemble based training method. NCL works on ensemble of learning machine as back propagation of feed forward neural networks along with bringing in error function () into each network. The primary aim of this research work is to predict on factors which contribute on early chances of customer churn which suggests on chances of customers who may potentially churn based on product quality related factors. This research work focuses on two major objectives: (a) To create and develop a strategy to improve customer product relationship model by understanding churn in a business organization. (b) To create a churn analysis model based on product buying patterns and determine risk tolerance of business organization.

Major research works pertaining to early prediction of customer churn suggests that its too challenging to understand the exact chances of churn[1][7], due to abstract features missing during analysis. Though research works are more surveyed in field of swarm intelligence [2] and machine learning[15], the demand for accurate prediction is always a misnomer[10]. This customer churn insight explains that NCL is prone to over fitting the noise in the training set. The paper analysis on the problem of over fitting customer churn prediction[3], due to irregularities in customer behavior and service. Hence this work proposes on the need for multi-objective of customer, product and service behavior which brings in optimized negative correlation learning approach. Fig-2 shows the functional architecture of NELCO on involvement of predicting churn based on feature analysis. NELCO works on customer risk analysis factors which are ranked primarily based on the type of risk faced by a customer during purchase process[9]. System identifies on various features of customer purchasing attitude including time of purchase, frequency of purchase, acceptable cost of purchase and time length of purchase of product. NELCO understands customer churn at an early rate such that any chance of error of a customer churn as miss may add to increasing the churn rate as each customer churn induces another new customer to be churned. Multiple transaction attributes [21] are suggested to verify on any churn parameter which may possess less probability to churn or suggest retention.

The model works on negative correlation based learning which involves a strong strength of a penalty value to be considered for iteration of neural networks whose average of ensemble values proposed in network always check on dissimilarity score[5].
NELCO depicts the feature of negative correlation learning factor, which impacts on the status of customer churn as an early prediction. Using impacts of customer’s after purchase feedback, the shopping evaluation is briefed in detail on various metrics of purchasing behavioural metrics, which include relative time of purchase, time spent towards purchase and other product time related metrics[12], which determine the correlation aspect of understanding on chances of customer churn.

![Functional Model of NELCO for customer churn analysis and Retention factors](Fig. 1)

The paper is organized based on primary discussions focusing on need for customer churn and its analysis as suggested in section-1, Section-2 elaborates on detailed survey and analysis of customer churn prediction approaches and challenges lying behind churn. Section-3 discusses on need for applying Negative Correlation Learning approaches and its influence of churn analysis. NCL is a primary learning approach which works based on correlation among churn metrics. NELCO supports on customer churn analysis based on effective scheduling of metric as discussed in Section-4. Section -5 suggests on the outcome of work based on experimental analysis and dataset preparation and Section-6 summarizes on outcome of work. Though detailed survey discusses [18] about various swarm models for prediction the need for an effective prediction model should be suggested.

### II. RELATED WORK

This section deals with an in-depth analysis of customer churn observed in research works related to customer churn and implementation of computational algorithms. Major research works [2][7] suggest on customer churn as a complex challenge which needs proper analysis and detailed survey to be carried out. Table-1 shows the feedback evaluation gathered from different customers on purchasing experience. The customer satisfaction level is obtained as Dissatisfaction, Neutral and Agree corresponding to the feedback queries. Detailed survey on customer purchasing behaviour [13] is dependent on multiple aspects of product quality and purchasing behaviour. This section discusses on the survey and analysis of evaluation models adopted in customer churn. Research works pertaining to churn analysis[8], describe the most frequently used approaches to Artificial Neural Networks (ANN), which adopts to multiple parametric analysis of several techniques being investigated with ANNs.

| Shopping evaluation | Dissagree (%) | Neutral (%) | Agree (%) |
|---------------------|-------------|-------------|-----------|
| Purchasing is desirable | 18.3 | 30.2 | 48.5 |
| The prices of products are very acceptable | 17.5 | 40.9 | 38.4 |
| I am confident that buying product is a good decision | 22.7 | 45.1 | 29.7 |
| I am saving significant amount of money | 20.8 | 48.7 | 28.6 |
| Purchasing product is definitely worth the money | 24.1 | 49.6 | 23.7 |
| Considering the price, products purchased are of excellent quality for the price | 22.6 | 51.0 | 22.6 |

Research community considers k-nearest neighbor (k-NN) [4] and random forests as another major two classification methods applied in literature for churn prediction. k-NN is considered as a major classification method in most of research works, which defines the instance classified by neighbourhood classification. Random forests research approaches are considered for research and survey on ensemble of decision trees. This approach finds major design and implementation in field of churn analysis [6] as discussed by Iridis et al[6]. The lifecycle of customer churn depends on customer acquisition, predicting profitable collection of customers, understanding their behaviour and maintaining or monitoring them effectively as discussed in [16]. Sakthikumar Subramanian and Arunkumar Thangavelu proposes on Bee Hive classifier approach applied over customer churn analysis. The analysis supports on identification of beneficial customers and chances of predicting the highly probability of customer to join organization [15]. Identifying and segmenting customers as sub-groups using classification algorithms are implemented for marketing techniques can be used to target different customer segments such targeted advertising and/or direct marketing. Langkvist et al [9] discussed on adoption of survey on two-layer Back-Propagation neural network (BPN) which showcased issues of customer behaviour and marketing challenges lying ahead. Different machine learning algorithms are suggested in literature analysis for churn prediction, but due to imbalanced class distributions whose primary task of earlier prediction is very challenging, the class of non-churning customers is found to out number the churners in analysis.

Few research works adopt Decision Trees to analyze on customer segmentation approaches suggest on analysis of customer retention model [8] which support on probable approaches to churn as well retain at an earlier state. Approaches to predict customer churn also adopts class imbalance behaviour which adopts PSO algorithm as well assigns on weight of customer's behaviour to optimize and suggest on structure of neural network model appropriately. This approach control and monitor issues of managing imbalanced data as well handles effective distribution using SVM in the training phase[5].
Based on detailed survey and analysis the research outcome which suggests on usage of soft computational algorithms to support on prediction of churn, it can be understood that multiple algorithms support on analysis criteria, which impact on customer behaviour, purchasing attitude and service monitored throughout the purchase activity. Though swarm intelligence is mainly adopted need for machine learning algorithms is felt to improve on factors such as minimize churn, increase retention, cost involved in adding more customers such that customer feedback is valuable for business organization.

III. NEED FOR NLC AND MODELLING APPROACH

Negative Correlation based Learning approach (NELCO) suggests on design and development of an effective architectural model to suggest accuracy in prediction of customer who are willing to churn. The demand for early churn prediction is primary which discusses on application of time series analysis of data adapted for analysis. This approach follows NCL learning approach which is considered as an improved learning of Fuzzy or neural networks which permit towards learning of multiple aspects of training all neural networks together based on ensemble of whole data set together. All the individual networks are trained simultaneously using correlation penalty terms or error terms as their error function. The approach of providing similar learning to all networks supports on the goal of optimizing individual training and deciding the best result as outcome. NCL approach brings in usage of \( \cdot \) as a parameter to suggest on convenient way of learning in order to balance covariance of trade off which exists among each approach models.

NELCO assumes a set of neural networks defined as \( X_i \), where \( i \) being \{1, ..., k\}, whose output pattern \( r \) defines the similarity index of each variable of an average output of layer nodes as

\[
A_i = \frac{1}{2} (Xi(n) - r(n))^2
\]

\( Xi(n) \) –Set of inputs for neural network function of network \( i \)

\( r(n) \) –output response pattern observed for input

Negative Correlation Learning introduces the penalty value \( \cdot \) whose error function is defined, where \( Pi \) is the correlated value of function.

\[
Y_i = A_i + \lambda. Pi(n)
\]

Here, \( Pi(n) = (Xi(n) - Y_i(n)) \sum_{k=1}^{n} (Xi(n) - Yi(n)) \)

This work uses discriminant analysis as the segmentation approach to classify or group customers as positive members who remain as customers within business institute and customers who refrain from buying with assigned weight \( b \) between them. To define proper discrimination among customers or variables, k-means neighbourhood classification approach, Equation (2) is discussed.

For any defined training set, \( R \) defines as discriminant score achieved, with \( b \) as the discriminant coefficient member variable, while \( X_1, X_2, ..., X_k \) are independent variables, shown in Eqn (3).

\[
R = \{ b \ (X_1, X_2, ..., X_k) \}
\]

Where \( X \in R \) which is scalar,

\[
O_i(n) = \frac{1}{K} \sum_{j=1}^{R} X_i^n
\]

NCL considers adding an extra term of error function of ensembling the neural networks as each independent network is expected to minimize the error (\( Zi(n) \)) shown in Equation (5).

Here, \( Y_i(n) \) indicates the output of an independent network and \( X(n) \) defines the average output of all networks. \( \cdot \) indicates the variable change in error analysis based on discriminant values.

Fig. 2. NCL approach applied over NELCO for customer churn classification

The error function of each network receives minimal response related to target response which is the average of all network node ensembles. This penalty strength parameter defines \( b = 0.0 \) suggests for independent training. The process of evaluation iteration continues until \( b > 0.0 \), where the NELCO system tries to understand the diversity of ensemble adopted and strengthen the penalty value. This journal uses double-blind review process, which means that both the reviewer(s) and author(s) identities concealed from the reviewers, and vice versa, throughout the review process. All submitted manuscripts are reviewed by three reviewer one from India and rest two from overseas. There should be proper comments of the reviewers for the purpose of acceptance/rejection. There should be minimum 01 to 02 week time window for it.

IV. NELCO APPROACH AND ALGORITHM

NELCO supports on analysis of churn management of customer and its related effect with product whose primary interest lies in buying the product with optimal factors of purchasing.
An Effective Architectural Model for Early Churn Prediction – NELCO

Fig. 3. Factors involving churn towards customer behaviour and product purchase

Fig-3 shows the factors influencing customer –product relationship and their essential bonding links. The chances of churn of a customer relates to multiple factors such as Product Interest over colour, cost of product, properties, quality, utilization factors as discussed in Fig-3. The feedback obtained from customer after analysis is as shown in Table-2.

Table- II: Customer product influential metrics and their observed values

| Churn Factors | Description | Influential Value Observed during evaluation |
|---------------|-------------|---------------------------------------------|
| 1 Pc          | Customer – Product Price changes | 67.04                                      |
| 2 Pi          | Product – Purchase Influential factors | 74.99                                      |
| 3 Pm          | Product Market demand              | 65.14                                      |
| 4 Pd          | Product Demographic factors        | 58.04                                      |
| 5 Pf          | Product Purchase frequency         | 88.36                                      |
| 6 Pb          | Business Competition               | 85.69                                      |
| 7 Ci          | Customer Product change interest   | 79.05                                      |

The relative correlation existing between customer interest towards buying the product and product experiential features bring out the coherence among churn and non-churn. The following step wise algorithm explains the feature of NELCO with implementation discussed in section 5.0

Algorithm NELCO

Input: Customer purchase dataset (C1, C2, …Cn)

For ∀ churn predictors Oi to be verified

Do

(a) Training set $R = \{ (b_1, X_1, b_2, X_2, \ldots, (b_k, X_k) \}$

(b) Set $\lambda = (\lambda)^n$

For each R pattern from K=1 to n do

a) $O(n) = \frac{1}{K} \sum_{i=1}^{K} X_i(r_i) \cdot \lambda$

b) If $O(n) < R(i)$ then Assign Weight $O(i) \leftarrow W_k$ // Rank the features

Else i = i + 1

c) $Z(n) = \frac{1}{2} (Y_i(r_i) - r(r_i))^2 - \lambda (Y_i(r_i) - X(r_i))^2$

d) if $Y_i(n) - Z_i(n) < 0$ then Check next Pattern R // No similarity exists with trained pattern for churn

Else// If similarity exists

Assign Weight $O(i) \leftarrow W_n$

Set $\lambda = 0.0$ // Procedure ends

Stop.

NELCO algorithm explains on considering multiple inputs from customer behaviour, product purchase, and service outcome. Training pattern R defines values which are well observed and verified based on specific retention process. Weight is being assigned for churn probable predictors who suggest on maximum possibility to churn such that an early indication is measured.

V. EXPERIMENTAL TEST BED

The dataset consists of various customer details collected over period of two years 2015 to 2017, which considers dataset of 5,000 customer details including customer's product purchase information, service details, total number of visits per day, number of customer service calls, number of products purchased. This approach invokes all active customers who involve in retail shopping performance which are collected and finalized for 2 years of analysis. The customer dataset collected as real time store between 2017 to 2018 consists of 12730 records for 1117 products variable over differential parameters (Refer Table-3).

Table- III: Customer product influential metrics and their observed values

| Customer Details | Products | Purchase History | Service Questions |
|------------------|---------|-----------------|-------------------|
| Zip Code/Location| Type of Product | Frequency of Purchase | Number of Store Visits |
| Income Bracket   | Variety of Products | Date of Last Complaint | Complain Resolutions |
| Gender/ Age      | Coupon Usage | Time of Day/Season of Purchase | Complain Priority |
| Type of Occupation | Product Preferences or Combo interest | Value of Purchases | Type of complain? |
| Do they have kids in the household? | Customer Interactions | Payment Methods | Frequency of complaints |
| Interest in buying the product | Quality of product | Balances / Store Credit | Observable feedback |
| Activity/ Inactivity period/ Break | | | |

Three different datasets from various sources were considered; (1) Customer's past six month aggregate active period and churned customers data volumes, (2) Product data which is globally surveyed based on its usage and (3) Activities of service dataset which comprises of individual weekly data which includes quality, annoyance measure and churn scores which are analysed using Rapid Miner tool.
The NELCO results confirm with NCL based MLP ensemble achieves on an optimal generalization performance (high churn rate) compared with ensemble of MLP without NCL (flat ensemble) and other common data mining techniques used for churn analysis. The table-3 explains on inclusion of multiple metrics adopted for evaluation whose domains fall under Customer details, Products, Purchase History, and Service Questions. The experimental evaluation identifies any possible relation among these metrics leading to a churn.

The support provided by NELCO in prediction of churn process is shown in Fig-4, where performance of NELCO is on an average is 71.033 % in comparison to PSO which demonstrates on an average of 65.202 % and performance of ACO being 67.845%. Though performance of NELCO is relatively high in terms of accuracy, previous research works suggest that swarm intelligence approaches provide a major analytical model for churn analysis. NLC being a stochastic predictive machine learning based predictive approach, resolves the challenges in analysis of early churn. Fig-4 shows that the actual churn observed shows a similar pattern observed by NELCO approach. When churn rate is high, NELCO follows an adaptive predictive pattern which is not observed under PSO or ACO approaches.

Fig-5 demonstrates on chances of early churn prediction rate, based on percentile of churn prediction relative to time as period of churn. NELCO approach shows that more than 80.36% of probability of churn is predicted earlier than compared to schemes such as PSO and ACO. Both these approaches are modular in their execution and follow regression policy to suggest on controlling churn.

NELCO approach suggests that early prediction is possible using training pattern and feedback mechanism which supports on early prediction. NELCO demonstrates 75.86% of early prediction well above the possible threshold demand. PSO shows 65.27% of early prediction while ACO demonstrates initially late prediction but performs similar to PSO scheme is linear regression model.

Fig. 4. Accuracy in observed customer churn

Fig. 5. % of early customer churn

The chances of an error to be predicted in any model is positive impression of demonstrating the behaviour of system. NELCO shows 12.29% of an average error being predicted while, PSO shows 22.19% of error and ACO performs with 14.77% percentage of error achieved as shown in Fig-7. The performance of NELCO is found to optimal in terms of achieved accuracy, % of error observed, chances of earlier probability of churn, and methods to retain the customer.

V. CONCLUSION AND FUTURE WORK

Understanding customer churn is a major research discussion among business analysts and research community. The support towards at retaining valuable customers is a demandable task which require regular monitoring and update as it has vital need towards improving the customer relationship management.
The need for an optimal early prediction of customer churn is felt as a major research motivation, which contributes to design and development of NELCO approach. Though this approach uses NCL approach along with K-means neighborhood approach to suggest of early prediction of customer churn, as well as supports on customer retention approach. An average error of 12.29% is an early predicted, ACO shows 14.77% percentage of error and PSO supports 22.19% of error achieved. The performance of NELCO is found to be optimal in terms of achieved accuracy, % of error observed, chances of earlier probability of churn, and methods to retain the customer.

Further research work can be extended in field of machine learning approaches which demand on construction of improved prediction performances, related to classifier ensembles (or multiple classifiers) with support over hybrid classifiers, outlier detection over negative churn analysis. These approaches support over improved memory consumption to detect and analyze on earlier churn prediction models and automated churn alarms with customer retaining strategies.

REFERENCES

1. Ali Rodan, Ayham Fayyoumi, Hossam Faris, Jamal Alisakran, and Omar Al-Kadi, Negative Correlation Learning for Customer Churn Prediction: A Comparison Study, Volume 2015, Article ID 473283, 7 pages, September 2014.

2. Chuh-Fong Tsai, Yu-Hsin Lu, Data Mining Techniques in Customer Churn Prediction, Recent Patents on Computer Science 2010, 3, 28-32.

3. H. Chen and X. Yao, “Regularized negative correlation learning for neural network ensembles,” IEEE Transactions on Neural Networks, vol. 20, no. 12, pp. 1962–1979, 2009.

4. H. Faris, “Neighborhood cleaning rules and particle swarm optimization for predicting customer churn behavior in telecom industry,” International Journal of Advanced Science and Technology, vol. 68, pp. 11–12, 2014.

5. Hossam Faris, A Hybrid Swarm Intelligent Neural Network Model for Customer Churn Prediction and Identifying the Influencing Factors, Information 2018, Pgs-9, Vol-288, November 2018.

6. A. Idris, M. Rizwan, and A. Khan, “Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies,” Computers & Electrical Engineering, vol. 38, no. 6, pp. 1808–1819, 2012.

7. Ismail, M.R.; Awang, M.K.; Rahman, M.N.A.; Makhtar, M. A multi-layer perceptron approach for customer churn prediction. Int. J. Multimed. Ubiquitous Eng. 2015, 10, 213–222.

8. M. R. Khan, J. Manoj, A. Singh, and J. Blumenstock, “Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty,” in 2015 IEEE International Congress on Big Data (BigData Congress), pp. 677–680, 2015.

9. M. Langkivist, L. Karlsson, and A. Loutfi, “A review of unsupervised feature learning and deep learning for time-series modeling,” Pattern Recognition Letters, vol. 42, pp. 11–24, 2014.

10. Y. Liu and X. Yao, “Ensemble learning via negative correlation,” Neural Networks, vol. 12, no. 10, pp. 1399–1404, 1999.

11. Lu, N.; Lin, H.; Lu, J.; Zhang, G. A Customer Churn Prediction Model in Telecom Industry Using Boosting. IEEE Trans. Ind. Inf., 10, 1659–1665, 2014.

12. A. Saas, A. Guittart, and A'. Peria n’ez, “Discovering playing patterns: Time series clustering of free-to-play game data,” Computational Intelligence and Games (CIG), 2016 IEEE Conference on. IEEE, 2016, pp. 1–8, 2016.

13. A. Sharma and P. K. Panigrahi, “A neural network based approach for predicting customer churn in cellular network services,” International Journal of Computer Applications, vol. 27, no. 11, pp. 26–31, 2011.

14. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting.” The Journal of Machine Learning Research, vol.15, no. 1, pp.1929–1958, 2014.

15. Saktihumar Subramanian and Arunkumar Thangavelu, CCHUM - An adaptive Business Intelligence strategy to improve Customer Retention using bee-hive classification algorithm International Journal of Research and Reviews in Information Technology (IJRRIT) Vol. 1, No. 2, pp.35–40, June 2011.

16. Saktihumar Subramanian, Arunkumar Thangavelu, FACT—An adaptive customer churn rate prediction method using fuzzy mutecriteria classification approach for decision making, Asian Journal of Science and Technology, Volume-4, Issue-11, pp: 227-233, 2011.

17. Shuo Wang and Huahuax Chen and Xin Yao, Negative Correlation Learning for Classification Ensembles, WCCI 2010 IEEE World Congress on Computational Intelligence July, 18-23, 2010 - CCIB, Spain

18. Venkatesan, M., Arunkumar, T. and Prabhavathy, P. ‘A novel Cp-Tree-based Co-located classifier for big data analysis’, Int. J. Communication Networks and Distributed Systems, Vol. 15, pp.191–211, 2015.

19. T. Verbraken, W. Verbeke, and B. Baesens, “Profit optimizing customer churn prediction with Bayesian network classifiers,” Intelligent Data Analysis, vol. 18, no. , pp. –23, 2014.

20. G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” International journal of forecasting, vol. 14, no. 1, pp. 35–62, 1998.

21. X. Zhang, Z. Liu, X. Yang, W. Shi, and Q. Wang, “Predicting customer churn by integrating the effect of the customer contact network,” in 2010 IEEE International Conference on Service Operations and Logistics and Informatics (SOLI), 2010, pp. 392–397.

AUTHORS PROFILE

R. Manivannan received the B. E. degree in Computer Science and Engineering from Annamalai University in 2014. He received the M.E. degree in Computer Science and Engineering from Annamalai University, Annamalainagar in the year 2016. He is doing his Ph.D in Computer Science and Engineering at Annamalai University. He published several papers in International Conferences and Journals. His research interest includes Big Data, Computer Networks and Network Security.

Dr. R. Saminathan received the B. E degree in Computer Science and Engineering from Arunai Engineering College in 1997. He received the M.E. degree in Computer Science and Engineering from Annamalai University, Annamalainagar in the year 2005. He has been with Annamalai University, since 2000. He completed his Ph.D degree in Computer Science and Engineering at Annamalai University, in the year 2012. He published 55 papers in international conferences and journals. His research interest includes Computer Networks, Network Security, Mobile Ad hoc Networks, Network Simulator and Big Data.

Dr. S. Saravanam received the B. E. degree in Computer Science and Engineering from Annamalai University in 1998. He received the M.E. degree in Computer Science and Engineering from Annamalai University, Annamalainagar in the year 2005. He has been with Annamalai University, since 2005. He received his Ph.D in 2015 from Annamalai University. He published several papers in International Conferences and Journals. His research interest includes Computer Networks, Intrusion Detection System, Network Security, Mobile Ad hoc Networks, Mobile Computing, Software Engineering and Big Data.