This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

doi:10.3233/FAIA200688

Customer Churn Prediction Based on HMM in Telecommunication Industry

Huisheng ZHU a,1, Bin YU b

^a School of Mathematics and Information Technology, Jiangsu Second Normal University, China

^b Marketing Department, Taizhou Branch of China Telecom, China

Abstract.

The rapid development of technology and increasing numbers of customers have saturated the communication market. Communication operators must give focused attention to the problem of customer churn. Analyzing the customer's communication behavior and building a prediction model of customer churn can provide the advance evidence for communication operators to minimize churn. This paper describes how to design a HMM to predict customer churn based on communication data. First, we oversample churners to increase the number of positive samples and establish the relative balance of positive and negative samples. Second, the continuous numerical attributes that affect communication customer churn are relatively discretized and their monthly values are converted into monthly change tendencies. Next, we select the communication features by calculating the information gains and information gain rates of all communication attributes. We then construct and optimize a prediction model of customer churn based on HMM. Finally, we test and evaluate the model by using a Spark cluster and the communication data set of Taizhou Branch of China Telecom. Experimental evaluation provides proof that our prediction model is exceptionally reliable.

Keywords. Customer churn, Prediction, HMM

1. Introduction

The rapid development of technology and increasing numbers of customers have saturated the communication market. With the arrival of 5G era, the communication industry is facing a new round of reshuffle, and the competition among communication operators on customer resources has entered a white-hot stage. A customer will change to a different communication operator when his or her needs are not met, which is called customer churn. The communication operators should pay more and more attention to the problem of customer churn because customer churn is closely related to the profits of communication operators. For communication operators, the cost of acquiring a new customer is approximately six times higher than the cost of retaining an existing customer[1], and reducing 5% customer churn rate will lead to 25%-85% increment of profit [2].

¹Corresponding Author: Huisheng ZHU, Jiangsu Second Normal University, 6 Xinhe West Road, Shiqiu street, Nanjing, China; E-mail: zhs@jssnu.edu.cn.

Communication operators possess massive communication data. How to find the behavior patterns of communication customers hidden in these data by data mining and build a prediction model of communication customer churn has become a research hotspot in industry and academia. Communication operators can identify potential churners in advance and take corresponding measures to retain them by adopting a prediction model of customer churn. Therefore, the research of communication customer churn is of great practical significance for communication operators to adjust their marketing strategies, change their ways of service, expand their market share and enhance their competitiveness.

In this paper, we propose a prediction model of communication customer churn based on HMM (Hidden Markov Model) and evaluate its performances by using a Spark cluster and a real communication data set. The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 introduces our prediction model of communication customer churn. The experiment results for evaluating our prediction model are presented in Section 4, and finally, Section 5 concludes this work.

2. Related work

Since the 1990s, many scholars have focused on the research of communication customer churn prediction and proposed a series of prediction methods. Mozer et al. [3] put forward a prediction model of mobile customer churn based on linear regression, decision tree, neural network and ensemble learning. In [4-6], a prediction model of customer churn based on neural network was proposed. Song et al. [7] designed an algorithm for predicting customer churn by employing a hybrid process neural network based on the Fourier orthogonal basis function. Kisiglu and Topcu [8] proposed an algorithm of customer churn prediction based on Bayesian belief network. In [9][10], the activation diffusion algorithm of psychology was used to calculate the impact of a churner on other customers in social networks and determine whether a customer will be a churner or not. Xie et al. [11] applied sampling and cost sensitive learning in a standard random forest and proposed a balanced random forest algorithm for customer churn prediction. In [12][13], a prediction model of customer churn based on decision tree is proposed. In order to settle the problem of imbalance data in customer churn prediction, Li et al. [14] proposed an algorithm of customer churn prediction based on cluster sampling and support vector machine training. Hossam et al. [15] proposed a prediction model of customer churn based on genetic algorithm. In [16-18], three prediction models of customer churn based on the rough set theory were proposed. Huang et al. [19] used linear regression, decision tree, naive Bayes, neural network and support vector machine to predict customer churn and compared the advantages and disadvantages of these methods.

The above research works did not consider the temporal relation among communication data while building their prediction models. In fact, the churn tendency of communication customers are closely related to their historical communication behavior. HMM (Hidden Markov Model) is a probabilistic model of time series data, which describes the process of generating a state sequence from a hidden Markov chain, and then generating an observation sequence from the state sequence. HMM has been widely used in speech recognition [20], biostatistics [21], stock market prediction [22], fault

diagnosis [23], smart home [24] and other fields. However, the application of HMM in communication customer churn prediction is rare. The reason is that the traditional HMM has only one observation attribute, but there are multiple features hidden in the communication data which affect customer churn. How to recognize these features from the communication data and build a HMM based on the observation sequence of these features are the motivations of this paper.

3. Our Prediction Model of Communication Customer Churn

3.1. The framework of communication customer churn prediction

Communication operators usually stop service to the customer who owes for one month and recycle the SIM Card if the customer is in arrears for three months. In this paper, we aim to build a prediction model to identify the communication customer with churn tendency in advance, so a churner is defined as the customer who has not produced any communication data for more than one month.

The framework of communication customer churn prediction as shown in Figure 1 includes samples labelling, attributes discretization, features selection, model training, parameters optimization and model testing.

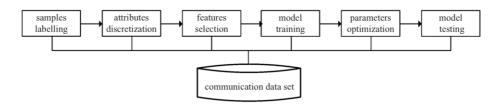


Figure 1. The framework of communication customer churn prediction

3.2. Samples labelling

The average churn rate of communication customers is about 2.2% [25], so the communication data set is an unbalanced data set in scenario of the two classification problem of predicting whether a customer is a churner or not. In the data set, churners are the minority and are called positive samples, while non-churners are majority and are called negative samples. The classification model trained directly from these samples will probably lead the prediction result belong to negative class.

Therefore, we oversample churners to increase the number of positive samples and establish the relative balance of positive and negative samples. As shown in Figure 2, we first label the monthly churners and non-churners of the training set composed of N months call data, then copy the data of churners in t+2 and t+3 in any month t ($1 \le t \le N-3$). Finally, the first N-3 months call data will be used as training samples, and the rest data will be used for model optimization.

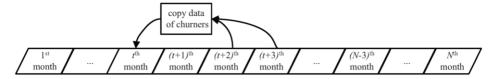


Figure 2. Churners oversampling

3.3. Attributes discretization

Each call record of communication customers has many attributes. Some of these attributes such as switch id, calling area code, calling cell id, etc. are not related to churn tendency of communication customers. The main communication attributes that affect customer churn are some continuous numerical attributes such as call frequency, call duration, call charge, etc. However, the monthly values of these continuous numerical attributes are not directly related to customer churn, instead their monthly change tendencies are most affinitive to customer churn. Therefore, these communication attributes need to be relatively discretized.

Let X_t be the value of communication attribute X of a customer in month t, the change tendency of communication attribute X of the customer in month t compared with month t-1 can be calculated by Eq. (1).

$$X'_{t} = \begin{cases} 0 & X_{t} = 0\\ 1 & X_{t} \neq 0 \land X_{t} - X_{t-1} \leq 0\\ 2 & X_{t} \neq 0 \land X_{t} - X_{t-1} = 0\\ 3 & X_{t} \neq 0 \land X_{t} - X_{t-1} \geq 0 \end{cases}$$
(1)

In Eq. (1), $X'_t = 0$ indicates that the customer did not produce any communication data in month t, $X'_t = 1$ indicates that the value of communication attribute X of the customer in month t is lower than that in month t-1, $X'_t = 2$ indicates that the value of communication attribute X of the customers in month t is equal to that in month t-1, and $X'_t = 3$ indicates that the value of communication attribute X of the customers in month t is higher than that in month t-1.

3.4. Features selection

Different communication attributes have different impacts on churn tendency. Let IF(X) be the impact factor of communication attribute X on churn tendency, its computational procedures are as follows:

1. Calculate the information entropy of customer class set $C=\{\text{churner}, \text{non-churner}\}\$ by Eq. (2).

$$H(C) = -\sum_{c \in C} p(c) log_2 p(c)$$
 (2)

2. Calculate the conditional entropy of customer class set *C* under the condition of a given communication attribute *X* by Eq. (3).

$$H(C|X) = \sum_{x \in X} p(x)H(C|X=x) = -\sum_{x \in X} p(x)\sum_{c \in C} p(c|x)log_2p(c|x)$$
(3)

3. Calculate the information gain of communication attribute X by Eq. (4).

$$IG(X) = H(C) - H(C|X) \tag{4}$$

4. Calculate the information gain rate of communication attribute *X*, namely the impact factor of attribute *X* on churn tendency by Eq. (5).

$$IF(X) = \frac{IG(X)}{H(X)} = \frac{IG(X)}{-\sum_{x \in X} p(x)log_2 p(x)}$$
 (5)

Let Avg JG be the average information gain of all communication attributes and Min JF be the threshold of information gain rate, attribute X is called a communication feature if $IG(X) \ge Avg JG$ and $IF(X) \ge Min JF$.

3.5. Model training

HMM is a system embedded with two stochastic processes, one of which is a hidden state transition process (Markov chain), the other is a visible sequence of observation symbols which is related to the state transition process. If the state transition of a HMM depends on its previous n states, such a HMM is called a n-order HMM. The first order HMM is used in this paper.

There are two stochastic processes embedded in the scenario of communication customer churn prediction, one is the visible monthly call data, the other is the hidden states transition. The visible call data is closely related to the hidden state transition, so HMM is suitable for the scenario of communication customer churn prediction, and the prediction model can be trained by determining the following five elements.

- 1. State set $S = \{S_1 = \text{churn}, S_2 = \text{non-churn}\}$. A state indicates whether a communication customer has the tendency of churn or not. Let Q_t be the state S_1 or S_2 of the system in month t, the state sequence from the first to the T^{th} month is expressed as $Q = Q_1Q_2, \dots, Q_T$.
- 2. Observation symbol set $V = \{V_1, V_2, \dots, V_M\}$. For a traditional HMM, M is the number of all different values of only one observation attribute, and $V_j (1 \le j \le M)$ represents a value of the observation attribute. However, for the HMM applied in the scenario of communication customer churn prediction, the observed attributes are K communication features that affect customer churn. Therefore, M in the observation symbol set is equal to 3^K , here 3 indicates that there are three kinds of monthly change tendency of each communication feature (1 is descend, 2 is flat, 3 is ascend), and $V_j (1 \le j \le M)$ is a vector composed of the change tendencies of K communication features in the same month. Let O_t be the symbol V_j observed by the system at month t, the observation sequence from the first to the T^{th} month is expressed as $O = O_1 O_2, \dots, O_T$.
- 3. State transition distribution $A = \{A_{ij}\} = \{P(Q_{t+1} = S_j | Q_t = S_i)\}$. $P(Q_{t+1} = S_j | Q_t = S_i)$ represents the probability that the system is in state S_i at month t and in state S_j at month t+1, which is calculated by Eq. (6). *Num* in Eq. (6) indicates counting values that meets a given condition.

$$A_{ij} = \frac{\sum_{t=1}^{t=T-1} Num(Q_t = S_i, Q_{t+1} = S_j)}{\sum_{t=1}^{t=T-1} Num(Q_t = S_i)}$$
(6)

4. Observation symbol distribution $B = \{B(V_j|S_i)\} = \{P(O_t=V_j|Q_t=S_i)\}$, where $P(O_t=V_j|Q_t=S_i)$ represents the probability that the system is in state S_i at month t and the observed symbol is V_j , which is calculated by Eq. (7).

$$B(V_j|S_i) = \frac{\sum_{t=1}^{t=T-1} Num(Q_t = S_i, O_t = V_j)}{\sum_{t=1}^{t=T-1} Num(Q_t = S_i)}$$
(7)

5. Initial state distribution $\pi = {\pi(S_i)} = {P(Q_1 = S_i)}$. $P(Q_1 = S_i)$ indicates the probability that the system is in state S_i at the first month. If the initial state is a normal uniform distribution, $\pi(S_1) = \pi(S_2) = 1/2$.

Obviously, initial state distribution π and state transition distribution A can determine state set S, and observation symbol distribution B can determine observation symbol set V, so a HMM only requires the specifications of A,B and π . For convenience, the whole element set can be denoted by a triple, i.e., $\lambda = (A, B, \pi)$.

3.6. Parameters optimization

Given an initial model $\lambda = (A, B, \pi)$ and an observation sequence $O = O_1 O_2, \dots, O_T$, parameter optimization is defined as how to adjust A, B and π to maximize $P(O|\lambda')$, i.e., the probability of the occurrence of O in the new model λ' . We use the following Baum Welch algorithm [20] to optimize an initial model λ .

Step 1, estimate the parameters by model λ and observation sequence O, and a new model $\lambda' = (A', B', \pi')$ is obtained.

Step 2, calculate the probabilities $P(O|\lambda')$ and $P(O|\lambda)$ by λ' , λ and O.

Step 3, if $|P(O|\lambda') - P(O|\lambda)| > \varepsilon$, here ε is a very small positive number, let $\lambda = \lambda'$ and goto step 1. Otherwise, optimization is over.

3.7. Model testing

Communication customer churn prediction aims at calculating the churn probability of a customer in month T+1 based on the prediction model and the observed change tendency sequence of communication features in the T months of a customer, i.e., $O=O_1O_2,\cdots,O_T$ (O_t is an observation vector). The customer whose churn probability is greater than or equal to the churn probability threshold Min_CP is predicted to be a churner, otherwise the customer is predicted to be a non-churner. Obviously, the essence of communication customer churn prediction is seeking the maximum probability of hidden state at time T+1, i.e., the maximum $P(Q_{T+1}=S_j|O_1O_2,\cdots,O_T)$. $P(Q_{T+1}=S_j|O_1O_2,\cdots,O_T)$ can be calculated by Eq. (8).

$$P(Q_{T+1} = S_j | O_1 O_2, \dots, O_T) = \sum_i A_{ij} P(Q_T = S_i | O_1 O_2, \dots, O_T)$$
(8)

 $P(Q_T=S_i|O_1O_2,\cdots,O_T)$ in Eq. (8) can be calculated by Eq. (9).

$$P(Q_T = S_i | O_1 O_2, \cdots, O_T)$$

$$= \gamma P(O_T | Q_T = S_i) \sum_j A_{ji} P(Q_{T-1} = S_j | O_1 O_2, \cdots, O_{T-1})$$
(9)

In Eq. (9), γ is normalization constant which can ensure that the sum of the probabilities in the two hidden states of churn and non-churn is 1.

4. Experimental Evaluation

Spark is a fast and universal computing engine designed for large-scale data processing. It not only implements the map and reduce methods in MapReduce, but also provides rich components such as Spark Streaming for real-time streaming, MLlib for machine learning, GraphX for large-scale graph processing and Spark SQL for interactive query. Spark supports memory computing and greatly improves the efficiency of iterative algorithm. The experimental environment of this paper is a Spark cluster composed of one master node and four slave nodes. The hardware configuration of each node is 3.60GHz Intel (R) core (TM) i7-7820x CPU, 64GB memory and 2TB hard disk. The software environment is Ubuntu 16.04, Spark 2.3.0, Scala 2.11 and JDK1.8. All the programs are written in Java.

4.1. Data set

The experimental data set is from Taizhou Branch of China Telecom, which stores the call records of all communication customers from January 1 to December 31, 2019. Each call record in the data set includes 146 attributes as shown in Table 1. All call data are compressed by Snappy and stored in the form of plain text on the HDFS of Spark cluster.

Attribute	Description	
event_id	event id	
customer id	customer id	
calling_nbr	calling number	
calling_area_code	calling area code	
calling_cell_id	calling cell id	
calling_op_type	calling operator type	
called_nbr	called number	
called_area_code	called area code	
called_cell_id	called cell id	
called_op_type	called operator type	
switch_id	switch id	
call_frequency	call frequency	
call_duration	call duration	
basic_charge	basic charge	
prepaid_charge	prepaid charge	
plan_charge	charge of plan	
call_charge	charge out of plan	

Table 1. Attributes of each call record

The numbers of churners and non-churners in the data set are shown in Table 2. The ratio of the total number of churners to that of non-churners is about 2%.

Month	Churner	Non-churner
January 2019	15313	880016
February 2019	20345	895225
March 2019	18466	906232
April 2019	16328	915337
May 2019	11225	923441
June 2019	13532	932105
July 2019	18263	942081
August 2019	21611	956402
September 2019	32829	972033
October 2019	13147	981174
November 2019	16632	989005
December 2019	34206	1005233

Table 2. The numbers of churners and non-churners in the data set

4.2. Communication features

We extract a customer who became a churner in December 2019 from the data set and depict the change tendencies of main communication attributes of the customer in Figures 3-8.

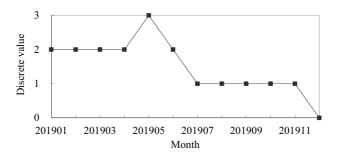


Figure 3. The change tendency of call_frequency

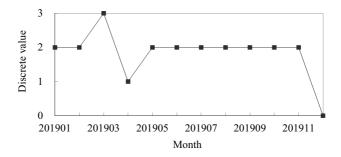


Figure 4. The change tendency of basic_charge

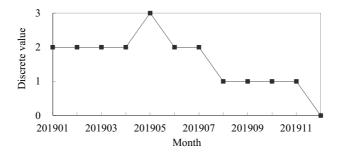


Figure 5. The change tendency of call_duration

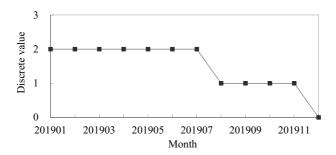


Figure 6. The change tendency of plan_charge

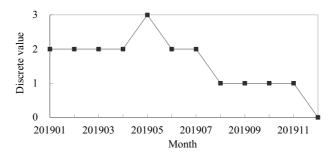


Figure 7. The change tendency of call_charge

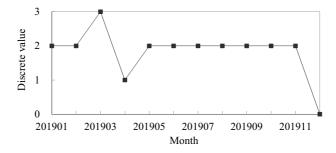


Figure 8. The change tendency of prepaid_charge

For the training set composed of the first 8 months call records in the data set, assume the threshold of information gain rate is 0.02, the communication features calculated from the positive and negative samples of the first 5 months (after samples labelling with the method described in Section 3.2) is shown in Table 3.

Communication feature	Impact factor
call_frequency	0.1443
basic_charge	0.1275
call_duration	0.1106
plan_charge	0.0983
call_charge	0.0619
prepaid_charge	0.0238

Table 3. The communication features in the data set

4.3. Performance evaluation criteria

Customer churn prediction is a problem of classification in two categories, whose confusion matrix is shown in Table 4. The table defines the churners as the positive class, and the non-churners as the negative class.

Table 4. The confusion matrix of customer churn prediction

	Prediction Positive	Prediction Negative
Positive Sample	True Positive (TP)	False Negative (FN)
Negative Sample	False Positive (FP)	True Negative (TN)

We use precision and recall to evaluate the performance of our prediction model. The higher precision and recall, the better the performance. Precision and recall are calculated by Eqs. (10) and (11).

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

We also use *AUC* (Area Under *ROC* Curve) [26][27] to evaluate the performance of our prediction model. The larger the *AUC*, the better the performance. *ROC* (Receiver Operating Characteristic) [26][27] refers to the connection of points drawn with FPR (False Positive Rate) as the horizontal axis and TPR (True Positive Rate) as the vertical axis. FPR and TPR are calculated by Eqs. (12) and (13).

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

$$TPR = \frac{TP}{TP + FN} \tag{13}$$

To calculate AUC, we first compute the churn probability of each communication customer by Eq. (9) and sort them in decending order, then we assign the customer with the highest churn probability rank n, the customer with the second highest churn probability rank n-1, and so on. In this way AUC can be calculated by Eq. (14).

$$AUC = \frac{\sum_{n \in TP} Rank_n - \frac{TP \times (TP+1)}{2}}{TP \times TN}$$
 (14)

4.4. Performance evaluation results

Experiment-1 (Performance vs. Size of training set). We choose the first 8 months call records of the data set as the maximum training set. Fixing *Min_CP* (churn probability threshold) at 0.57, Figures 9-11 depicts how size of training set (in month) affects precision, recall and *AUC*. From the figures we can notice that precision, recall and *AUC* increase with size of training set. The reason is that the larger training set, the more balanced the positive and negative samples. We also observe that it needs at least 7 months of training data to make the model have better prediction results.

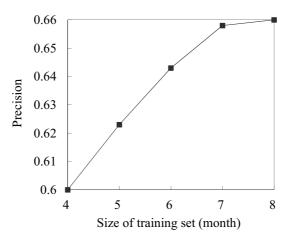


Figure 9. Precision vs. Size of training set

Experiment-2 (Performance vs. *Min_CP*). In this experiment, we present the performance comparison of our model and *LIBLINEAR* [28] on the real communication data set. LIBLINEAR has been a linear model widely used in massive data classification since it received the KDD CUP championship in 2010 [29].

After training our prediction model based on the first 8 months call records of the data set, we predict the average churn probability of all customers in December by their change tendencies of all communication features from September to November. We vary churn probability threshold *Min_CP*. The precisions and the recalls are shown in Figures 12 and 13. Clearly, the precisions of our model and *LIBLINEAR* increase with *Min_CP*, but the recalls of our model and *LIBLINEAR* decrease with *Min_CP*. This is because the

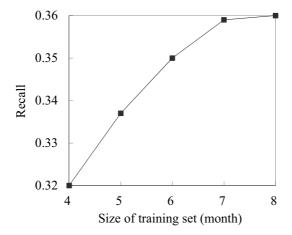


Figure 10. Recall vs. Size of training set

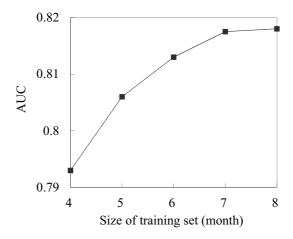


Figure 11. AUC vs. Size of training set

larger the *Min_CP*, the less false positive and the higher the precision by Eq. (10), the more the false negative and the lower the recall by Eq. (11).

We can also observe that the precision and the recall of our model are higher than those of *LIBLINEAR*. The reason is that our model based on HMM has one more time dimension than *LIBLINEAR*.

Figures 12 and 13 show that $Min_CP = 0.57$ is the intersection of precision curve and recall curve. Therefore, we generally choose $Min_CP = 0.57$ in the actual prediction in order to take into account of precision and recall.

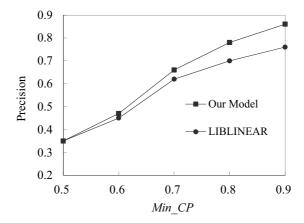


Figure 12. Precision vs. Min_CP

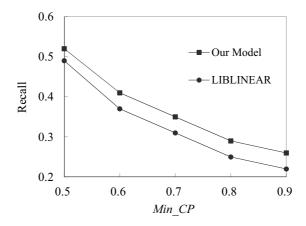


Figure 13. Recall vs. Min_CP

5. Conclusion

In this paper, we design a prediction model of communication customer churn based on HMM and evaluate its performances by using the Spark cluster and the real communication data set. The construction of the model includes samples labelling, attributes discretization, features selection, model training, parameters optimization and model testing. The experimental results shows that our model has better performances than *LIBLINEAR* in the scenario of communication customer churn prediction.

The prediction model designed in this paper only considers call records of communication customers. In the future, we will study on communication customer churn prediction based on call record, message record and network traffic.

Another future research direction is the examination of whether a higher-order HMM can be used to improve the performance of our predicted activities.

6. Acknowledgments

We very appreciate the anonymous reviewers for their valuable comments and suggestions. This research was supported by the National Natural Science Foundation of China grants 61672163, and the Chinese Ministry of Education Foundation for 'Integration of Cloud Computing and Big Data, Innovation of Science and Education' grants 2017B06109.

References

- [1] Alsultanny YA. Database preprocessing and comparison between data mining methods. International Journal of New Computer Architectures and their Applications. 2011 Jan; 1(1):61-73.
- [2] Shin HM, Jeong DH. An application of data mining for marketing in telecommunication. Proceedings of Portland International Conference on Management of Engineering and Technology; 2001 July 29 -August 2; Portland, OR, USA. Wiley: IEEE Press; 2002. p. 247.
- [3] Mozer MC, Wolniewicz R, Grimes DB, Johnson E, Kaushansky H. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. IEEE Transactions on Neural Networks. 2000 Jan; 11(3):690-6.
- [4] Zhang Y, Liang R, Li Y, Zheng Y, Berry M. Behavior-based telecommunication churn prediction with neural network approach. Proceedings of International Symposium on Computer Science and Society; 2011 July 16-17; Kota Kinabalu, Malaysia. Wiley: IEEE Press; 2011. p. 307-310.
- [5] Yu R, An X, Jin B, et al. Particle classification optimization-based BP network for telecommunication customer churn prediction. Neural Computing and Applications. 2018 Feb; 29(3):707-720.
- [6] Hu J, Zhuang Y, Yang J, et al. pRNN: A recurrent neural network based approach for customer churn prediction in telecommunication sector. Proceedings of IEEE International Conference on Big Data; 2018 December 10-13; Seattle, WA, USA. Wiley: IEEE Press; 2018. p. 4081-5.
- [7] Song G, Yang D, Wu L, Wang T, Tang S. A mixed process neural network and its application to churn prediction in mobile communications. Proceedings of IEEE International Conference on Data Mining Workshops; 2006 December 18-22; Hong Kong, China. Wiley: IEEE Press; 2006. p. 798-802.
- [8] Kisiglu 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] Dasgupta K, Singh R, Viswanathan B, et al. Social ties and their relevance to churn in mobile telecom networks. Proceedings of the 11th International Conference on Extending Database Technology; 2008 March 25-29; Nantes, France. ACM Press; 2008. p. 668-677.
- [10] Yossi R, Elad YT, Noam S. Predicting customer churn in mobile networks through analysis of social groups. Proceedings of SIAM International Conference on Data Mining; 2010 April 29 - May 1; Columbus, Ohio, USA. SIAM Press. 2010. p. 732-741.
- [11] Xie T, Li X, Ngai EWT, Ying W. Customer churn prediction using improved balanced random forests. Expert Systems with Applications. 2009 Apr; 36(3):5445-9.
- [12] Wei CP, Chiu TT. Turning telecommunications call details to churn prediction: a data mining approach. Expert Systems with Applications. 2002 Aug; 23(2):103-112.
- [13] Deng W, Deng L, Liu J, Qi J. Sampling method based on improved C4.5 decision tree and its application in prediction of telecom customer churn. International Journal of Information Technology and Management. 2019 Jan; 18(1):93-109.
- [14] Li P, Yu X, Sun B, Huang J. Telecom customer churn prediction based on imbalanced data resampling method. Proceedings of the 2nd International Conference on Measurement, Information and Control; 2013 August 16-18; Harbin, China. Wiley: IEEE Press; 2014. p. 229-233.
- [15] Hossam F, Bashar AS, Nazeeh G. A genetic programming based framework for churn prediction in telecommunication industry. Proceedings of the 6th International Conference on Computational Collective Intelligence; 2014 September 24-26; Seoul, Korea. Springer Press; 2014. p. 353-362.
- [16] Amin A, hehzad S, Khan C, Ali I, Anwa S. Churn prediction in telecommunication industry using rough set approach. New Trends in Computational Collective Intelligence. 2015 Jan; 83-95.
- [17] Amin A, Anwar S, Adnan A, et al. Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing. 2017 May; 237:242-254.

- [18] Vijaya J, Sivasankar E. Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. Computing. 2018 Aug; 100(8):839-860.
- [19] Huang BQ, Kechadi MT, Buckley B. Customer churn prediction in telecommunications. Expert Systems with Applications. 2012 Jan; 39(1):1414-1425.
- [20] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989 Feb; 77(2):257-286.
- [21] Liu JS, Neuwald AF, Lawrence CE. Markovian structures in biological sequence alignments. Journal of the American Statistical Association. 1999 Mar; 94(445):1-15.
- [22] Hassan MR, Ramamohanarao K, Kamruzzaman J, Rahman M, Hossain MM. A HMM-based adaptive fuzzy inference system for stock market forecasting. Neurocomputing. 2013 Mar; 104:10-25.
- [23] Laxman S, Sastry PS, Unnikrishnan KP. Discovering frequent episodes and learning hidden Markov models: a formal connection. IEEE Transactions on Knowledge and Data Engineering. 2005 Nov; 17(11):1505-1517.
- [24] Cheng BC, Tsai YA, Liao GT, Byeon ES. HMM machine learning and inference for activities of daily living recognition. The Journal of Supercomputing. 2010 Oct; 54(1):29-42.
- [25] Berson A, Thearling K. Building data mining applications for CRM. New York: McGraw-Hill Press; 1999. p. 510.
- [26] T. Fawcett. An introduction to ROC analysis. Pattern Recognition Letters. 2006 Jun; 27(8): 861-874.
- [27] J. Davis, M. Goadrich. The relationship between Precision-Recall and ROC curves. Proceedings of the Twenty-Third International Conference on Machine Learning 2006 Jun 25-29; Pittsburgh, Pennsylvania, USA, ACM Press; 2006. p. 233-240.
- [28] Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research. 2008 Jul; 9: 1871-1874.
- [29] Yu HF, Lo HY, Hsieh HP, et al. Feature engineering and classifier ensemable for KDD cup 2010. Proceedings of the JMLR Workshop and Conference;2010 July 25-28; Sardinia, Italy. Cambridge: MIT Press; 2010. p. 153-160.