

Semi-Supervised Churn Clustering for Fault and Constraints Prediction in Telecom Industry

Narasimhan Kamalraj* and Arunachalam Malathi

Department of Information Technology, Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore - 641049,
Tamil Nadu, India; tpkamal@gmail.com, malathi.arunachalam@yahoo.com

Abstract

Objective: Churn prediction process on telecom industry is performed using background knowledge such as user information and application oriented constraints with the aim of clustering churn items of telecom communication users.

Methods: This paper proposes Semi-supervised Constraint based Churn Clustering (SCCC) method. Semi-supervised learning method identifies different types of churns for labelled data items in telecom industry. PCK Mean based Clustering with Expectation Maximization (EM) algorithm finds cluster labels and distance metric for easy grouping of faults set. Similar types of issues are mapped with fast traversal procedure. Constraint based Cluster Membership achieves effective churns clustering by considering relative pairing of users. **Findings:** The proposed Semi-supervised Constraint based Churn Clustering (SCCC) method is implemented by using JAVA language. This JAVA language uses the code to effectively identify the churns in telecom industry. The SCCC method uses a churn dataset for conducting the experimental work. Experiment is conducted on the factors such as execution time, accuracy, clustering efficiency and support rate on predicting the churns in telecom industry. Experimental results show that the proposed Semi-supervised Constraint based Churn Clustering (SCCC) method out performs than the existing methods. **Improvement:** Proposed method is able to minimize the execution time, improve accuracy and clustering efficiency and also increase the support rate on predicting churns in telecom industry.

Keywords: Cluster Membership, Churn Prediction, Expectation Maximization, Semi-Supervised Clustering,

1. Introduction

One of the foremost industries to apply data mining technique is the telecom industry. This is due to the fact that most of the telecom industry cache higher amount of data, with an ever increasing stock of customers in a highly competing scenario. In the telecom industry, the customers have their own rights of switching between service providers and they are able to select from among numerous service providers. As a result, the average churn rate in telecom industry is experiencing, the newly evolved problem of retaining the customers which has become one of the major issues to be handled by them.

Association Rule Mining task within a Corporate Privacy-Preserving (ARM-CPP)¹ framework analyzed the association rule mining task to protect privacy and ensure scalability. However the churn in this system is unable to cluster resulting in computational complexity.

Multivariate Reconstructed Phase Space (MRPS)² system identified multiple temporal patterns to provide feasibility and effectiveness of the approach using Gaussian Mixture Model Classification (GMMC) though the relationship between multiple temporal patterns was analyzed but an effective method for churn analysis was not performed. A churn management model was designed in ³ using Partial Least Square (PLS) optimization with the motive of providing a prediction and classification model.

Another method was presented in ⁴ called the partial prediction of churn using logistic regression pattern for effective classification of churning and non-churning. However, sequential information was obtained using Markovian transition matrices with little focus on sequence alignment. A comparative analysis was made in ⁵ using Decision Tree (DT), Artificial Neural Network (ANN), Support Vector Machine (SVM) and Pruning to Produce Error Reduction (PPER) to provide the best

* Author for correspondence

churn prediction model. Experiments conducted using the four methods gave a track record that ensemble methods were better than base learners.

Numerous service providers in the telecom industry are heavily concerned with the churn prediction and the probability of its occurrences. The activities related to Customer Relationship Management (CRM) work with the main aim of increasing the lifetime of a customer with a specific service provider. In⁶, Self Organizing Maps (SOM) and Genetic Programming (GP) were applied to improve the retention of the customer using enhanced classification tree. However, the most influencing factors for churn were not concentrated. With the motive of identifying the influential factors for churn, a Multi Layer Perceptron (MLP) was applied in⁷ for constructing efficient churn classification models. One another method was introduced in⁸ to analyze the influential factors using fuzzy correlation analysis.

The fuzzy correlation analysis was proved to be efficient with respect to optimal retention timing. Yet another method for post-paid customer segment was designed in⁹ and measures were taken to analyze the churn rate. A hybrid churn prediction model using Neighborhood Cleaning Rules (NCR) and Particle Swarm Optimization (PSO) was introduced in¹⁰ to improve the accuracy and actual churn rate. Though accuracy was improved, an efficient differentiation between churners and non-churners were not carried out. A new framework¹¹ to examine and predict the churn rate in social network was analyzed by constructing a customer network model to differentiate between churners and non-churners. A new prediction model was presented in¹² with the objective of improving the customer retention strategy using data mining techniques.

Many companies have established a good rapport with their valued customers and through different types of transactions and gather numerous amounts of data. In order to increase the retention rate, the companies even obtain demographic attributes, by identifying prospective customers and their requirements. In this manner, a method to accurately predict the data was introduced in¹³ using Technology Acceptance Model (TAM).

Another framework using Bayesian network was designed in¹⁴ with the objective of effective identification of churn through Six Sigma Methodology. However, the operating cost for prediction was not considered. To address this issue, an effective Genetic Programming (GP) model was introduced in¹⁵ to enhance the robustness of

the design using customer based features. But, the cost related to retain the customer was not addressed.

As a result in¹⁶, a hybrid model including K-Means Clustering and Genetic Programming was introduced for the effective classification of customers into churners and non-churners. However, customer-oriented strategies were less focused in the hybrid model. Random forest and boosted tree techniques¹⁷ were introduced to address customer-oriented strategies rather than product-oriented and to improve the rate of sensitivity and specificity. But, churn factors with respect to cluster and retention rate remained unaddressed.

Since its origin, the area of Data Mining and Knowledge Discovery has come with many solutions to solve numerous practical issues. In¹⁸, decision support system was constructed for efficient identification of churn prediction in Telecommunication Company. However, with respect to scalability, the churn prediction remained unsolved. A Neural Network model¹⁹ to predict the customer churn with respect to scalability was included for non-steady customer state. However, issues related to dimensionality reduction remained unsolved. To address dimensionality reduction, a new rule algorithm on the basis of rough set theory was introduced in²⁰ to increase the positive predictive value and frequency.

Technological Changes in Telecom Sector²¹ provides Customer Relationship Management (CRM) such that changes in the approach to various customers are taken place to achieve fast and better service to end users. However, predatory pricing leads to harmful for the sectors. Instead of using voice and messaging (SMS), Telecom Service Providers offer Over the Top (OTT) Services²² to solve the issues related to mobile data traffic and data revenue. However, the investment required for such an approach is high.

By using K-Means Clustering Algorithms²³, machine learning and pattern recognition are performed for discovering the unknown knowledge. Though the distance among clusters is increased and distance within cluster is decreased, there is no detailed information about accident, scene and damaged personal information.

Based on the aforementioned techniques and methods, in this work, with the objective of identifying the fault and constraint prediction for churns in telecom industry, an efficient Semi-supervised Constraint based Churn Clustering (SCCC) method is presented. The contributions of Semi-supervised Constraint based Churn Clustering (SCCC) method includes the following:

- Technically, an efficient method called the Semi-supervised Constraint based Churn Clustering (SCCC) has been proposed to cluster the churn items (i.e. issues or faults) in the telecom communication industry.
- To improve the support rate on predicting the churn in SCCC method using Semi-supervised learning method that efficiently account with labeled data items and helps in easy identification of churns, both involuntary and voluntary churn in the telecom industry were used.
- To improve the clustering efficiency using PCK Mean based Clustering with the aid of Expectation Maximization (EM) algorithm that identifies the cluster labels and measure the distance for easy grouping of faults set.
- To significantly reduce the execution time for performing an efficient guiding operation by satisfying the pair wise constraints using PCK Mean based Clustering that directly maps similar type of issues in the specific cluster with fast traversal procedure.
- To increase the rate of accuracy using Constraint based Cluster Membership that effectively performs high churns clustering based on the relative pairing of users.

An efficient Semi-supervised Constraint based Churn Clustering in telecom industry is provided in the forthcoming section.

2. Semi-Supervised Constraint based Churn Clustering in Telecom Industry

In this section, a brief explanation is given on Semi-supervised Constraint based Churn Clustering (SCCC) method for easy prediction of faults and constraints in telecom industry. Telecommunication specifically depends on the electromagnetic waves and the main objective is to identify the customer churn (i.e. faults or issues). The goal of SCCC method is to cluster similar types of issues and optimize the system effectively. The telecommunication service provider identifies proficient and effective tools to predict customer churn. The objective of the Semi-supervised Constraint based Churn Clustering is to obtain the customer telecommunication function. Then the function is used to gain knowledge

about the issues using the semi-supervised learning.

The learning procedure is followed continuously and similar type of churns is grouped together for effective functioning. The data mining tool used for the churn identification and clustering is shown in Figure 1. The customers in the telecom industry often change their subscription due to higher amount of churns (i.e. faults or issues). Churn is also said to occur whenever a customer changes the account. The SCCC method applies data mining techniques using effective prediction and clustering of different type of churn. The data mining process in SCCC method applies the learning procedure to discover the customer behavior in telecom industry. The overall architecture diagram of describing the identification and clustering of churn in the telecom industry is shown in Figure 2.

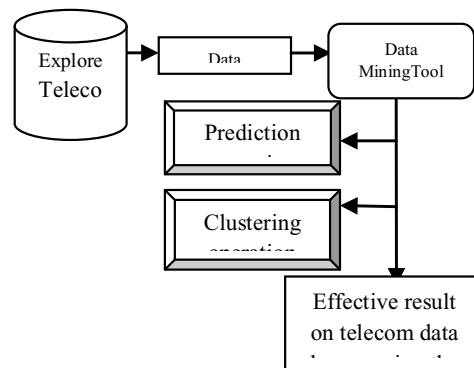


Figure 1. Data mining tool based representation of telecom industry.

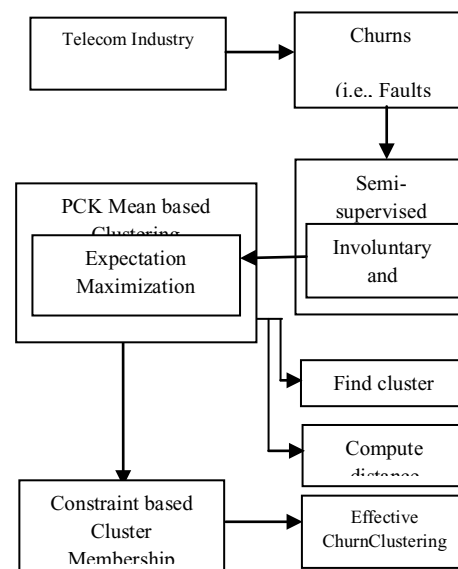


Figure 2. Architecture diagram of SCCC method.

As shown in Figure 2, each user (i.e. customer) uses different type of telecom industry to accomplish their communication. The data items of different telecom industry are taken for the experimental work. In the telecommunication industry, SCCC method provides broad definition for the churns on customer service. Service provider is validated to find whether churn arises in the communication path using data mining techniques. The customer initiated churn during the service is first identified using the semi-supervised learning. Then the semi-supervised learning tasks make use of the unlabeled items.

Semi-supervised learning easily identifies the involuntary and voluntary churns and then grouping is carried out on the proposed work using the PCK Mean based Clustering. The PCK Mean based Clustering in SCCC method uses the Expectation Maximization (EM) algorithm, where the parameter enables the system to easily group with minimal execution time. EM easily identifies the cluster label and the distance metrics. The cluster membership function is developed to predict easily the timing factor taken to analyze the churns in the telecom industry. For instance, in the telecom industry, the customer changes the service center, accounts and plan types and also the billing options such as online and offline. These services churn (i.e. faults and issues are identified) and occurrence of the churn in the communication line is identified in an efficient manner.

2.1 Semi-supervised Learning Method

To start with, a semi-supervised learning method is administered in SCCC to support and identify the churn. Both high and normal churn in the telecom industry is easily identified based on the services. Semi-supervised learning is related with supervised and unsupervised position. The information is associated for analyzing the customer previous services on the particular telecommunication path. Let us assume the unlabeled data item as $()$ for which the semi-supervised learning ω is provided.

The semi-supervised learning method in SCCC makes use of larger unlabeled data items and smaller amount of labeled data items. To perform both unlabeled and labeled data items, a Semi-supervised inductive learning is used in SCCC method and is formularized as:

$$SSLearning = mapping(L(I_n \rightarrow C)) \quad (1)$$

From (1), a semi-supervised learning ω consists of ω set. All the unlabeled data items ω are viewed with the learned (i.e. labeled) procedure to identify the churns in the telecommunication line. The voluntary and involuntary churns on the customer services are noted and then the quantitative data semi-supervised learning tool is used to understand customer category of learning, where most of the input is self-evidently unlabeled.

2.2 PCK Mean based Clustering

Once the churns are identified using Semi-supervised learning, it has to be categorized in a significant manner and similar faults have to be grouped for easy prediction in telecom industry. The SCCC method uses PCK Mean based Clustering where K-Means clustering with pair wise constraints is performed. The pair wise constraints in SCCC method are checked for easy clustering of similar churns. Let us consider that ω is the set of the subscriptions with the ω links where the links are used as the centroid of cluster in telecom industry and ω partitions the churns for easy clustering of the churn (i.e. faults or issue) with cluster label. The objective function with the cluster label in SCCC method is described as:

$$PCK_{clusterlabel} = \sum_{I \in cluster} (|I_n - \mu_{In}|^2 M_{In} - \log(\det M_{In})) \quad (2)$$

In (2), the cluster label is measured on the Item ω count where ω is the centroid of the cluster with different set of customer service action churns in telecom industry and ω is the matrix metric of the cluster. The pair wise constraint is fixed and the logarithmic form is determined with matrix determinant factor. The cluster is labeled in the K-Means form to attain higher efficiency rate. The clustered label with pair wise objective function constraints helps to easily separate each item from the overall set of functions.

The clustered label is used for measuring the distance metric value based on pair wise limitation procedure in the proposed telecom industry. Pair wise constraints effectively represent the user's issues or faults view of similarity in the telecom industry domain. The distance metric is computed on the space where the clusters are sufficiently separated. SCCC method formularizes the distance metric as:

$$DistanceMetric = \sqrt{(C_i - \mu_{In})^2 + (C_j - \mu_{In})^2} \quad (3)$$

In (3), the distance metric is measured with the original data items for measuring the distance between the cluster objects ‘‘ and ‘‘. The cluster is discovered using the grouping of similar semi-learning of telecom data. The centroid of the cluster is measured where similar type of all clusters are measured and then summed up to measure the overall distance metric space in SCCC method.

2.2.1 Expectation Maximization (EM) Algorithm

The expectation and maximization step is carried out together in SCCC method to provide an improved clustering of the churns. Expectation Maximization (EM) algorithm enables the system to easily group the churn in telecom industry with minimal execution time. EM easily identifies the cluster label and the distance metrics. EM Algorithm is converged and it is described in step by step procedure.

//Expectation MaximizationAlgorithm

Begin

Step 1: Optimal bound of clustering identified churn in telecom industry.

Step 2: Clustering based on PC K-Means procedure.

Step 3: Perform expectation.

Step 3.1: $E[C(I)] = P(\text{Current Cluster} | \text{Expectation Count})$.

Step 4: Perform maximization.

Step 4.1: $M[C(I)] = \text{argmax } E(C(I)) + \log P \text{ value}$.

Step 5: Observed data items are discrete in structure and perform clustering to identify the churn set.

End

The SCCC method measures the execution time using the Expectation Maximization (EM) algorithm. The expectation ‘E’ of the cluster for ‘T’ user items is used to compute the current cluster based on the expectation count in the telecom industry for varying shifting service centers. Similarly, the next step, named maximization procedure ‘M’ of the cluster with ‘T’ items maximizes the clustering efficiency with minimal time rate. The same set of procedural links is used for identifying and measuring all types of the churn (i.e. voluntary and involuntary churn) clustering using the EM algorithm.

2.3 Constraint based Cluster Membership

Finally, a Constraint based Cluster Membership function is developed to predict easily the timing factor taken to analyze the churns in the telecom industry. The membership value helps to easily predict the set

of users who is actively taking part with the churns in telecommunication. The cluster membership count is formalized as:

$$\text{ClusterMembership} = C[I_1, I_2, \dots, I_n] \quad (4)$$

The clusters members are used for easy prediction of the time factor, where the user items services in the telecommunication are noted. The cluster churns with the membership rate is noted for various participants with the set rule ‘0’ or ‘1’. The churn system is represented in the form of ‘1’, if it is predicted to be true. Otherwise, all the other sets are zero. The overall system communication on various services set in SCCC method is shown in Figure 3.

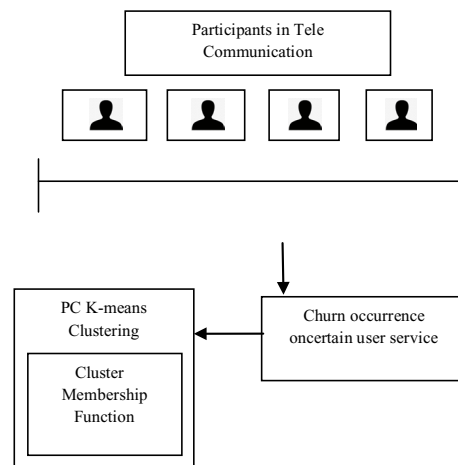


Figure 3. Design of constrain based cluster membership function.

Cluster membership function uses the discrete set of items (i.e. countable number of infinite steps) for easy way of predicting the time factor on churn clustering. Pair wise constraint propagation is used to easily identify clustering rate with better performance that is based on the cluster membership function for various service providers.

3. Experimental Evaluation

Semi-supervised Constraint based Churn Clustering (SCCC) method uses the JAVA platform. This JAVA platform uses the code to identify the churns in telecom industry. The SCCC method uses a churn dataset (URL: <https://www.sgi.com/tech/mlc/db/churn.data>) for the experimental work.

Semi-supervised Constraint based Churn Clustering (SCCC) method compares the work with the existing

system such as Association Rule Mining task within a Corporate Privacy-Preserving (ARM-CPP)¹ framework and Multivariate Reconstructed Phase Space (MRPS)² system. Experiment is conducted on factors such as support rate on predicting the churn, clustering efficiency, execution time and accuracy.

The support rate on predicting the churn measures the number of subscribers who move out over a period of time. The churn rate is the ratio of number of subscribers during a specific month to the average number of subscribers during the month (i.e. the subscribers during the start of the month and end of the month). It is measured in terms of percentage (%).

$$CR = \frac{\text{No. of subscribers}_{\text{mth}}}{(\text{No. of subscribers}_{\text{smth}} + \text{No. of subscribers}_{\text{emth}}) / 2} \quad (5)$$

Clustering efficiency in SCCC is measured on the basis of expectation and maximization process. It is measured in terms of percentage (%). Execution time using SCCC method measures the time taken to obtain the current cluster based on the expectation count and easily group the churn in the telecom industry. It is measured in terms of milliseconds (ms). The execution time to group the churn is formalized as:

$$ET = \text{Time} \{E[C(I)] + M[C(I)]\} \quad (6)$$

Accuracy using SCCC method measures the ratio of number of correct predictions to the total number of predictions.

$$A = \frac{\text{No. of correct predictions}}{\text{Total number of predictions}} \quad (7)$$

4. Results Analysis of SCCC

The Semi-supervised Constraint based Churn Clustering (SCCC) is compared against the existing Association Rule Mining task within a Corporate Privacy-Preserving (ARM-CPP)¹ framework and Multivariate Reconstructed Phase Space (MRPS)² system.

Table 1 evaluates the Support rate on predicting the churn in terms of percentage achieved with different number of subscribers ranging from 10 to 70 and comparison is made with the two existing methods namely, ARM-CPP¹ and MRPS² using the churn dataset.

Table 1. Tabulation for support rate on predicting the churn

No. of Subscribers	Support Rate on Predicting the Churn (%)		
	SCCC	ARM-CPP	MRPS
10	63.13	55.10	45.05
20	69.25	61.22	51.17
30	62.19	55.41	45.36
40	73.19	71.16	61.11
50	75.44	64.41	54.36
60	78.52	68.49	58.44
70	85.82	72.79	62.74

Figure 4 illustrates the support rate on predicting the churn based on subscribers using churn data. The proposed SCCC method performs relatively well when compared to two other methods Association Rule Mining task within a Corporate Privacy-Preserving (ARM-CPP)¹ framework and Multivariate Reconstructed Phase Space (MRPS)² system. The support rate on predicting the churn with respect to the number of subscribers is improved with the application of Semi-supervised learning method. By applying Semi-supervised learning method in SCCC method, the customer previous services are analyzed on the specific telecommunication path using both larger and smaller unlabeled data items resulting in the improvement of support rate on predicting the churn rate by 2-15 % compared to ARM-CPP. In addition using SCCC method, effective mapping of unlabeled data items to churns are performed with the aid of Semi-supervised learning method that efficiently understand the customer category of learning and therefore improves the support rate on predicting the churn by 16-28 % compared to MRPS.

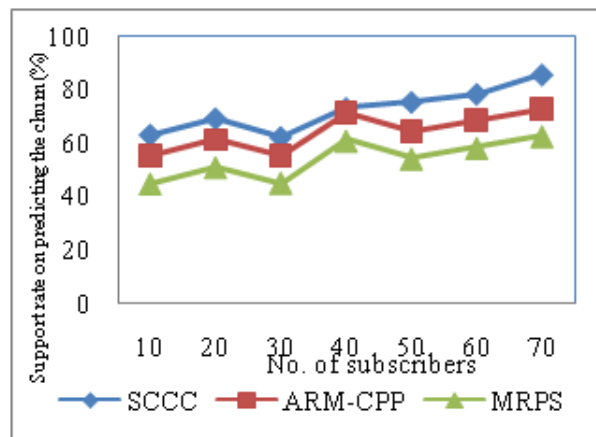


Figure 4. Impact of support rate on predicting the churn.

Table 2 represents the comparison results of clustering efficiency and performance with 70 subscribers considered for experimental purpose by applying the real world. The targeting results of clustering efficiency using SCCC method with two state-of-the-art methods ARM-CPP¹ and MRPS² is presented in Figure 5 for visual comparison. The proposed mechanism differs from the ARM-CPP¹ and MRPS² and the SCCC method is incorporated with PCK Mean based Clustering using Expectation Maximization (EM) algorithm to easily extend the determination factor and by tuning the churn observed grouped into clusters, rate of clustering efficiency increases the number of churns being identified in telecom industry.

Table 2. Tabulation for clustering efficiency

Method	Clustering Efficiency (%)
SCCC	83.88
ARM-CPP	74.12
MRPS	68.43



Figure 5. Impact of clustering efficiency.

The pair wise constraints are efficiently performed using PCK Mean based Clustering where centroid of cluster is used for easy clustering of the churn. This helps in improving the clustering efficiency. Furthermore, the objective function defined in (2), uses logarithmic form to determine matrix determinant factor aim to attain higher clustering efficiency rate in SCCC by 56% and 68% when compared to ARM-CPP and MRPS respectively.

In Table 3 the execution time for identifying the churn in telecom industry using expectation and maximization factor with the aid of churn data is examined. The experiments were conducted using the cluster with 70 subscribers taken into consideration for experimental purpose which is evaluated in terms of milliseconds (ms).

Table 3. Tabulation for execution time

No. of Subscribers	Execution Time (ms)		
	SCCC	ARM-CPP	MRPS
10	0.132	0.143	0.150
20	0.141	0.152	0.159
30	0.148	0.159	0.166
40	0.135	0.146	0.153
50	0.158	0.169	0.176
60	0.132	0.143	0.150
70	0.143	0.154	0.161

To explore the influence of execution time on SCCC method, the experiments were performed by varying the number of subscribers as depicted in Figure 6. It also shows that the SCCC method shows competitive results with the state-of-the-art methods, namely ARM-CPP¹ and MRPS² respectively. To reduce the execution time in Semi-supervised Constraint based Churn Clustering (SCCC) method by constructing an efficient EM algorithm. With the application of EM algorithm, the pair wise constraint map similar issues in the specific cluster, minimizing the execution time by 7-8% compared to ARM-CPP. In addition, the SCCC method uses Expectation for efficient identification of current cluster based on the expectation count for varying shifting service center and maximization procedure 'M' that improves the clustering efficiency with minimal time rate by 11-13 compared to MRPS.

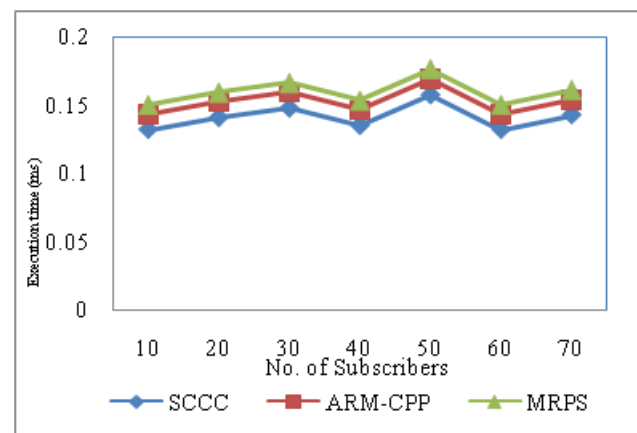


Figure 6. Impact of execution time.

The results of 70 different subscribers and the impact of accuracy are listed in Table 4. As listed in Table 4, the SCCC method measures the accuracy while providing numerous predictions for telecom industry which is

measured in terms of percentage (%). The accuracy using SCCC offers comparable values than the state-of-the-art methods.

Table 4. Tabulation for accuracy

No. of subscribers	Accuracy (%)		
	SCCC	ARM-CPP	MRPS
10	71.33	64.30	59.25
20	74.29	67.26	62.21
30	65.13	56.10	51.05
40	68.47	61.44	56.39
50	78.23	71.20	65.15
60	81.41	73.38	68.33
70	85.33	78.30	73.25

Lastly, the accuracy to measure the correct number of predictions is evaluated via number of subscribers with a queue size of 10 to 70. From the Figure 7 it is illustrative that the proposed SCCC method potentially yields better results than existing ARM-CPP¹ and MRPS². The significant results achieved using SCCC method is because of the application of Constraint based Cluster Membership. The Constraint based Cluster Membership in Semi-supervised Constraint based Churn Clustering (SCCC) method analyses the churns using the membership count for various participants improving the accuracy rate by 8-10 % compared to ARM-CPP. In addition, the cluster membership function uses the discrete set of items for easy prediction of churn improving the accuracy using SCCC method by 14-21 % compared to MRPS respectively.

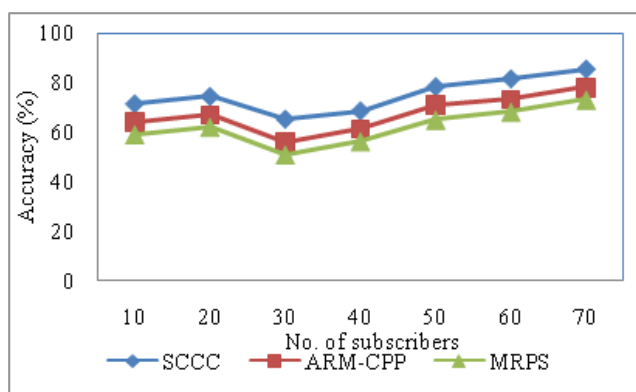


Figure 7. Impact of accuracy.

5. Conclusion

In this work, the problem of churn fault and constraint

in telecom industry is examined and an efficient Semi-supervised Constraint based Churn Clustering (SCCC) method is proposed to achieve support rate on predicting the churn and to minimize the execution time. It is shown how different subscribers can be used in a ubiquitous fashion to drastically reduce the time taken to execute the expectation and maximization procedure to identify the predictions in telecom industry. Also studied how this reduction in execution time translates into clustering efficiency and optimal number of churns being developed in the telecom industry. It is further shown, the attainable performance gains of the proposed strategy in terms of clustering efficiency that can achieve well above 10% compared with the state-of-the-art methods. Moreover, with the application of Expectation Maximization algorithms further increases the support rate on predicting the churn. It utilized Semi-supervised learning method and PCK Mean based Clustering using Expectation Maximization (EM) algorithm that results in marginal improvements in increasing the support rate on churns according to the size of subscribers. It has been considered how pair wise constraint map could prove beneficial in attaining clustering efficiency while maintaining the execution time taken for predicting the churns for different subscribers. Performance results reveal that the proposed SCCC method provides higher support rate on prediction and clustering efficiency and also strengthen the overall method by applying churn data set. Compared to the existing methods, the proposed SCCC method outperforms the state-of-art works in terms of execution time, accuracy, clustering efficiency and support rate on predicting the churns in telecom industry.

6. References

1. Giannotti F, Lakshmanan LVS, Monreale A, Pedreschi D, Wang HW. Privacy-preserving mining of association rules from outsourced transaction databases. *IEEE Systems Journal*. 2013 Sep; 7(3):385–95.
2. Zhang W, Feng X. Event characterization and prediction based on temporal patterns in dynamic data system. *IEEE Transactions on Knowledge and Data Engineering*. 2014 Jan; 26(1):144–56.
3. Kim YS, Lee H, Johnson JD. Churn management optimization with controllable marketing variables and associated management costs. *Expert Systems with Applications*. Elsevier. 2013 May; 40(6):2198–207.
4. Migueis VL, Poel DVD, Camanho AS, Cunha JFEC. Mod-

- eling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*. 2012 Sep; 39(12):11250–56.
5. Abbasimehr H, Setak M, Tarokh M. A comparative assessment of the performance of ensemble learning in customer churn prediction. *The International Arab Journal of Information Technology*. 2014 Nov; 11(6):599–606.
 6. Faris H, Shboul BA, Ghatasheh N. A Genetic Programming based framework for churn prediction in telecommunication industry. Springer. 2014 Jun; 8733(6):353–62.
 7. Adwan O, Faris H, Jaradat K, Harfoushi O, Ghatasheh N. Predicting customer churn in telecom industry using multilayer perceptron neural networks: Modeling and analysis. *Life Science Journal*. 2014; 11(3):75–81.
 8. Chueh HE. Analysis of marketing data to extract key factors of telecom churn management. *African Journal of Business Management*. 2011 Sep; 5(20):8242–7.
 9. Radosavljevik D, Putten PVD. Preventing churn in telecommunications: The forgotten network. Springer. 2013; 8207:357–68.
 10. Faris H. Neighborhood cleaning rules and particle swarm optimization for predicting customer churn behavior in telecom industry. *International Journal of Advanced Science and Technology*. 2014; 68:11–22.
 11. Pushpa, Shobha G. An efficient method of building the telecom social network for churn prediction. *International Journal of Data Mining and Knowledge Management Process (IJDKP)*. 2012 May; 2(3):31–9.
 12. Shaaban E, Helmy Y, Khedr A, Nasr M. A proposed churn prediction model. *IJERA*. 2012 Jun-Jul; 2(4):693–97. ISSN: 2248-9622. www.ijera.com
 13. Alawin A, Al-ma'aitah M. Proposed ranking for point of sales using data mining for telecom operators. *IJDMS*. 2014 Jun; 6(3):17–31.
 14. Trnka A. Position of retraining churn data mining model in six sigma methodology. *Proceedings of the World Congress on Engineering and Computer Science*. 2012; 2200(1):488–92.
 15. Khan I, Usman I, Usman T, Rehman GU, Rehman AU. Intelligent churn prediction for telecommunication industry. *International Journal of Innovation and Applied Studies*. 2013 Sep; 4(1):165–70.
 16. Obiedat R, Alkasassbeh M, Faris H, Harfoushi O. Customer churn prediction using a hybrid genetic programming approach. *Scientific Research and Essays*. 2013 Jul; 8(27):1289–95.
 17. Nabavi S, Jafari S. Providing a customer churn prediction model using random forest and boosted tree techniques. *Journal of Basic and Applied Scientific Research*. 2013 Jun; 3(6):1018–26.
 18. Jadhav RJ, Pawar UT. Churn prediction in telecommunication using data mining technology. *International Journal of Advanced Computer Science and Applications*. 2011 Feb; 2(2):17–9.
 19. Sharma A, Panigrahi PK. A neural network based approach for predicting customer churn in cellular network services. *International Journal of Computer Applications (0975-8887)*. 2011 Aug; 27(11):26–31.
 20. Dhandayudam P, Krishnamurthi I. Customer behavior analysis using rough set approach. *Journal of Theoretical and Applied Electronic Commerce Research*. 2013 Aug; 8(2):21–33.
 21. Yadav K, Tiwari S, Divekar R. Impact of technological changes in Telecom Sector in India. *Indian Journal of Science and Technology*. 2015 Feb; 8(S4):194–9.
 22. Sujata J, Sohag S, Tanu D, Chintan D, Shubham P, Sumit G. Impact of Over the Top (OTT) services on Telecom Service Providers. *Indian Journal of Science and Technology*. 2015 Feb; 8(S4):145–60.
 23. Karamizadeh F, Zolfagharifar SA. Using the clustering algorithms and rule-based of data mining to identify affecting factors in the profit and loss of third party insurance, insurance company auto. *Indian Journal of Science and Technology*. 2016 Feb; 9(7):1–9.