

Hybrid Features for Churn Prediction in Mobile Telecom Networks with Data Constraints

Naimisha Kolli, N.Balakrishnan

Supercomputer Education and Research Centre (SERC)

Indian Institute of Science (IISc)

Bangalore, India

naimisha.k@gmail.com

balki@iisc.ac.in

Abstract— In a competitive Mobile telecommunications market, the customers want competitive pricing and high quality of service. A customer won't hesitate to change their telecom service provider if he/she does not find what they are looking for. This phenomenon is called churning. The telecom service providers often find that the cost of acquiring a new customer is much more than the cost of retaining one. Hence telecom operators are focusing their marketing strategies toward targeted customer retention campaigns and this is known as Churn management. One of the primary tasks of Churn management is to build an effective churn prediction models that can predict customers who are most likely to churn. The primary idea is to create profile of a customer using various data sources including call patterns, contractual information, billing, payment, customer service calls, demographic profiles and then predict the probability that he/she will churn based on his/her features. The apparent drawback of these approaches is that they require access to numerous other sources of information apart from Call Data Records (CDRs). More importantly, these models do not take into account any social influence. In our present work, we recognize the importance of the role played by social ties understanding the causal behavior of customers, and incorporate a novel feature of the social aspects of customers' social group along with the traditional individual customer profiles with potential practical implications. We propose hybrid feature sets that are based not only on the features extracted from CDRs but also on the changes in these feature sets combined with the changes in the social group patterns that would give improved performance over existing models with similar data constraints. Despite the data constraints, we demonstrate through our experiments that our model achieves improved prediction performance using these hybrid feature sets extracted from the CDRs as well as mobile social graphs even with our data constraints.

Keywords—churn prediction; communities; call data records

I. INTRODUCTION

In a competitive service provider markets, the customers want competitive pricing and a high quality of service. A customer won't hesitate to change their telecom service provider if he/she does not find what they are looking for. This phenomenon is called churning. Recent studies have shown that service providers like the mobile operators lose 20% to 40% of their customers yearly [1][2] and cost of acquiring a new customer is five times the cost of retaining one [3]. Due to these

huge discrepancies in cost, service providers are focusing their strategies toward targeted customer retention campaigns and this is known as Churn management. One of the primary tasks of Churn management is to build an effective churn prediction models that can predict customers who are most likely to churn. In this paper, we aim to build an efficient machine learning based churn predictor that can predict the customers who are most probable to churn for a mobile telecom operator.

Significant research on detecting churning of customers has used various data mining methods such as neural networks [4], clustering [5], regression [6,7], support vector machine [8,9], decision tree [10], and hybrid methods [11], deep learning [12] to provide churn prediction models. So far, there hasn't been any common criteria for the prediction models to identify churn. The primary idea is to create profile of a customer using various data sources including call patterns, contractual information, billing, payment, customer service calls, demographic profiles and then predict the probability that he/she will churn based on his/her features. The apparent drawback of these approaches is that they require access to numerous other sources of information apart from Call Data Records (CDRs). A key problem that one has to address while collecting data from numerous sources is that of privacy. The privacy constraints can be imposed by law or can be driven by competing business interests. Despite the potential mutual gain and utility of pooling multiple data sources for various stakeholders, like academicians or business analysts, this is often not possible due to privacy as well as cost to scaling issues. The businesses that allow the data collection are also aware of the potential threats that arise for the confidentiality of such the data made available. In our current work, the telecom operator has shown preference to models for churn prediction that are based on customer behaviors, for whom only Call Data Records (CDRs) were available without the need for other sources of information in an attempt to minimize privacy concerns. The Call Data Records (CDRs) are auto generated by the network operator whenever there is any customer activity. This data is extremely useful to build behavioral models as it focuses on customers' interactions rather than transactions and gives much more meaningful insights into individual customer behavior thus is used as the only source of input in this work to extract novel behavioral feature sets. Using predictive analysis, the aim of this work is to predict the probability that a customer will churn in the future based on billing data provided by the mobile telecom operator.

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/\$31.00 © 2020 IEEE

Moreover, the above prediction models do not take into account any social influence. Another line of research [13-15] deals with the mobile call graph where the decision to churn is solely based on social influence. The main premise here is that churning action of a customer is propagated through his social network and in turn influences the churn probability of other associated customers. They show an improvement in prediction performance through the use of link information alone. With the success of social network attributes for churn prediction, improved methods [16-18] for churn prediction have explored ways of integrating SNA along with traditional features.

In our present work, we recognize the importance of the role played by social ties for understanding the actions of customers, and incorporate novel features of the social aspects of customers' social group along with the traditional individual customer profiles from the CDRs with potential practical implications. However, there is still improvement that can be achieved in these combined models as well and this is addressed in this work. We investigate a hybrid behavioral model that not only combines the individual usage features extracted from the CDRs with the social network features extracted from the mobile social graphs but also introduces a novel hybrid feature based on rate of change in features. We demonstrate through experiments that the addition of this novel feature will be able to capture subtle changes in behavioral patterns and thus shows an improvement in the prediction performance of our churn prediction models.

The rest of the paper is organized as follows. Section II reviews the related work on churn prediction. Section III defines the problem of churn prediction as supervised classification task. The Section IV describe the dataset used, its characteristics and the data preprocessing involved while Sections V explain the various phases of our methodology consisting of construction of mobile call graphs, community detection in these graphs, feature extraction, and model generation and implementation phases. Section VI shows the results of experimental evaluations on the real-world CDR dataset. We make a conclusion and suggest future research in Section VII.

II. RELATED WORK

Churn prediction has seen a growing interest from both researchers and businesses. Using customer transactional and billing data, Ahn et. al [1] investigate Korean mobile service provider for factors affecting customer churn. In the absence of demographic profile, Wei and Chiu [10] predict churning using decision trees at contractual level from features such as service duration, mode of payment and type of contract along with pattern of calls extracted from call data records using a decision tree model. Hadden et.al [19] have investigated the suitability of customer complaints and repairs data for churn prediction and used neural networks, classification trees and regression models for prediction. A good survey of churn prediction models can be found in [20]. A recent work also proves the superiority of decision tree technique over logistic regression techniques [21].

The above models do not take into account any social influence. Recent research [13-15] deals with the mobile call graph where the decision to churn is solely based on social influence. Richter et.al [15] has proposed Group-First Churn Prediction, which detects closely-knit customer groups through

the use of second-order social features extracted with the information-theoretic framework. Based on these extracted group features, a decision-tree model has been applied to predict churning of groups. Dasgupta et.al [14] and Nanavati et.al [13] propose a spreading activation-based technique (SPA) for predicting customers who are most probable to churn. This method is constructed on that idea that a small number of identified influential churners spread the process of churning through their contacts, who in turn influence their own contacts and so forth. Potential churners are identified using a threshold on the weighted influence calculated for each customer. Numerous works have extended this model and propose combined models that incorporate various aspects of social influence with the traditional customer attributes. Phadke et.al [16] and Kusuma et.al [17] extended classical churn datasets with features extracted from social networks based on spreading activation model and proposed machine-learning methods to estimate the likelihood of churn for the customer. While all initial churners are considered homogenous in the SPA model, Kim et.al [18] applied community detection to the propagation and proposed various features of the initial churners for extracting the social features and combined them with the traditional customer features and then built a logistic regression and neural network models. A good survey of these combined methods can be found in [22]. In our current work, we propose the use of novel hybrid feature sets that are based on the features extracted from CDRs exclusively combined with the social group features to build an efficient churn prediction model with superior performance over existing models.

III. CHURN PREDICTION MODELLING

A. Problem Statement

Churn prediction is defined as a supervised classification problem. The focus is to build a churn predictor that learns from features extracted from CDRs exclusively to make an accurate forecast for the probability of no activity in the future predetermined churn period. A churn forecast is associated with a particular time boundary at which the churn period starts and the input features are extracted from the call data records for each customer based on the activity observed prior to this time boundary. The data of the customers after the time boundary is used to infer the predicted churn/non-churn labels for the classification task.

IV. DATASET DESCRIPTION

A. Data Description

In this paper, we analyze the Call Data Records (CDRs) provided by one of the popular mobile telecom operators in India. The data contains detail record of calls made by segment of customers active during a thirty-one-day (for the month of March 2014) period in a large metropolitan city in India. The original dataset is large (120GB) and consisted of 400+ million CDR entries. Each Call Data Record (CDR) contains fields such as the timestamp, call duration, call type, cell tower, originating number, destination number, etc. of the call as maintained by the mobile operator for billing purposes. All models were implemented using GraphlabCreate toolkit and SFrame package in order to scale to much large data than other available resources [23][24].

Further, we use a thirty-one days of CDR data to build and test our models due the data availability constraints set by the mobile telecom operator. Despite of this, after discussions with the experts in the area, the consensus is that most people leave the network at any random point within a month's time frame based on their billing cycles, their convenience periods or otherwise. Hence, we argue that one month's data can be segmented and used for churn prediction as has been carried out in this work.

B. Data Preprocessing

The data was too large and too widespread to be useful for our current purpose, so we have done the following preprocessing to the original dataset.

- We have removed CDRs of many customer numbers, which were simply utility, customer care or telecom operators themselves. These numbers have been deemed as outliers, which reduced about 8.5% of the size of the original dataset.
- We only considered those customers who have both made and received calls as active mobile customers. This further reduced the size of the dataset by another 1%.
- Further CDRs calls with a duration of five seconds are filtered.
- Next, we have only considered customers who have been active for more than one day in our dataset as we are interested in the change in their behavioral patterns.

V. METHODOLOGY

A. Constructing Mobile Call Graphs

A dataset of customers with call activity among them can be characterized by a network in which nodes are customers and links represent their call activity. Based on the information contained in the CDRs regarding who called whom, time at which the call was made, we have constructed a mobile call graph for each day in our dataset with customers as nodes and the calls between customers as edges. A mobile call graph G_t is denoted by $\langle V(G_t), E(G_t) \rangle$, and $V(G_t)$ is a non-empty set of customers active on day t and $E(G_t)$ are a finite set of reciprocal customer tuples from $V(G_t)$. We only consider reciprocal edges. The edge weights represent call frequency between nodes. A higher call frequency implies a stronger tie, while a lower call frequency implies a weak tie.

B. Community Detection in Mobile Call Graphs

The previous creation of mobile call graphs allows for the study of communities in these graphs. As mentioned previously, we recognize the importance of the role played by social ties for understanding the actions of customers, and we incorporate novel features of the social aspects of customers' social group along with the traditional individual customer profiles. From churn prediction point of view, the knowledge of social or community membership of a customer to well-connected structures in the network is of primary importance. To this end, we use the Louvain community detection algorithm [25], to efficiently partition large-scale graphs to get the communities in our customer mobile graphs for each day in the dataset. There have been numerous community detection algorithms proposed

in the literature [26]. Among those we have found Louvain method to be highly scalable and still able to produce consistent communities on very large graphs in various other domains such as Botnets [27] and other social graphs and hence our motivation for the choice of this community detection.

C. Feature Extraction

The feature extraction phase is divided two sets. The first feature set relates to the features to describe calling behaviors of a customer by aggregating his call usage while the second set of features correspond to the social group or community related features of the customers extracted from the previously constructed mobile call graphs. CDRs are auto generated by the network operator whenever there is any customer activity. Hence their timestamp is continuous. For our purposes we have aggregated these timestamps and used a time granularity of one day. The fundamental precursors for a user to churn is the change in his call usage pattern and the correlated reaction to churn motivated by the churn of his social group. In order to capture all of these features into our proposed model framework, we have extracted features in 1) call usage and its change and 2) the social groups and their change over the observed period.

1) Call Usage Features:

The following features are generated for each customer based on the call usage aggregated on a daily basis: i) Number of incoming calls ii) Number of outgoing calls iii) Number of incoming SMS iv) Number of outgoing SMS v) Total Duration of voice calls vi) total number of calls to churners vii) total number of SMS to churners

2) Social group Features:

The following features are generated for each customer based on their social affiliations or communities in the mobile call graphs aggregated on a daily basis: i) Number of Distinct Contacts within the network ii) Number of Distinct Contacts outside the network iii) Size of the community to which the customer belongs to iv) Fraction of Neighbors who are churners v) number of hops to nearest churner.

3) Rate of Change of Usage and Group Features:

Based on the premise that changes in call usage and social groups are likely indicators for churn, a novel hybrid feature called rate of change feature is introduced for each customer's features, and this is also given as feature to the model. We have calculated the rate of change in each of the above-mentioned features in the following manner: For a feature f and customer c , at time interval i , and a small constant $\epsilon > 0$, the rate of change in feature is calculated as

$$\text{Change}(f)_i^\epsilon = \frac{\text{Value}(f)_i^\epsilon - \text{Value}(f)_{i-1}^\epsilon + \epsilon}{\text{Value}(f)_{i-1}^\epsilon + \epsilon} \quad (1)$$

In our experiments, the constant ϵ is added to avoid numerical instabilities and the value is set to 0.001 and the time interval considered is one day. So, the rate of change is calculated daily for each of the features considered.

Features calculated in the above manner will be able to capture subtle changes in behavioral patterns and thus possibly improving the performance of churn prediction.

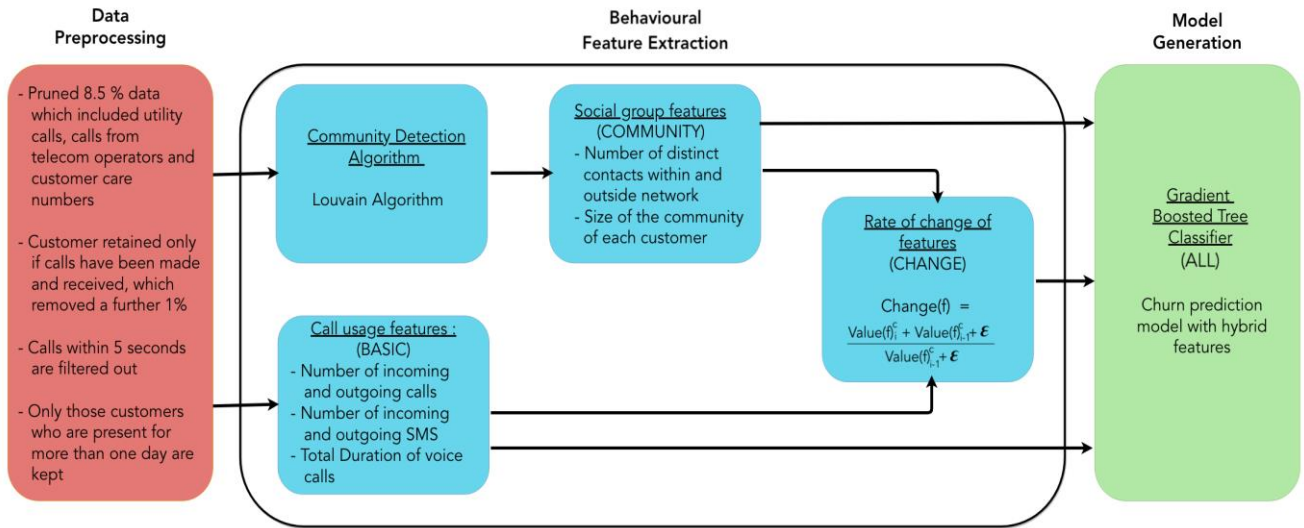


Fig. 1 Churn Prediction Modeling

4) Transformed Features:

Based on above described features, the toolkit also generates the following set of secondary features for input to the classifier to describe various usage and community change patterns based on a parameter T . We have set the parameter to generate the following features on a weekly and bi-weekly basis for each customer: i) Average number of events in the last T days, ii) Average value of a feature in the last T days, iii) Largest daily value of a feature in the last T days, iv) Largest time interval between two events in the last T days, v) days since the first activity in the last T days, vi) days since the most recent activity in the last T days, vii) days to the most recent activity in the last T days, viii) days with an activity in the last T days, ix) days with an activity in the last T days, x) Number of activities in the last T days, xi) Smallest daily value in the last T days, xii) Smallest time interval between two events in the last T days, xiii) Sum of a column in the last T days, xiv) T day trend in the number of events, xv) Time interval between the first two events in the last T days, xvi) Time interval between the two most recent events in the last T days.

D. Model Generation

Our framework aggregates the CDR data and generates the above described features including call usage patterns, social group features and also the rate of change in both usage and group features for each customer using the call data records and from the previously constructed mobile call graphs. The dataset has been randomly split into training dataset and testing dataset in 80:20 ratios. As mentioned in the Problem Statement section a churn forecast is always associated with a particular time boundary at which the churn period starts and the input features are extracted from the call data records for each customer based on the activity observed prior to this time boundary. The data of the customers after the time boundary is used to infer the predicted churn/non-churn labels for the classification task.

The next step in the model generation is to use the computed features including call usage patterns, social group features, the rate of changes in both usage and group features and the inferred

labels to train a classifier model. We build a gradient boosted tree classifier and train it using the training dataset to learn the model to classify potential customers who might churn. Gradient boosted tree [28] is an efficient machine learning algorithm for prediction applications. The Boosted Trees algorithm additively merges outputs from a system of weak classifier models. These class of models can be represented as:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \dots \quad (2)$$

where the classifier g is the formed by adding the simpler weak classifiers f_i . Here, each weak classifier is formed from a decision tree. Gradient Boosted trees uses gradient boosting [28]. Typically, the model makes fewer and fewer mistakes as more trees or weak learners are added. This classification approach works well for the current problem of churn prediction because the base classifiers or the weak learners are particularly suited to handle different feature types and also large amount of data in the training datasets.

We further split the training data into training and validation in the ratio of 90:10 in order to run a grid search to tune the hyperparameters. Once the classifier model is trained, we then test the model on the test dataset, previously split from the entire dataset. For a given time boundary, all the call activity in the dataset after this time boundary are not included in the training data or testing data for the model. The set of hybrid features including call usage patterns, social group features, the rate of changes in usage and group features are extracted for the customers in the test dataset and the trained boosted tree model is applied to predict the probability that a particular customer will churn. Once we have calculated this probability of churn for every customer, the model can be applied as a tool for creating prioritized ranked list for targeted marketing campaigns.

In our experiments, although we have experimented with other classification models such as Logistic Regression and Random Forest, the Gradient boosted trees have given superior results consistently. Hence for the purpose of illustration, we have presented our results using the Gradient boosted tree. Also,

it should be noted that the trends presented in our experiments for various models are consistent across other models as well.

VI. EXPERIMENTS AND RESULTS

A. Experimental Setup

Customer churn is defined to be no activity for a fixed period of time. Using this definition, a customer is said to have churned if he does not show any activity for the duration of time after t , the predetermined time boundary also known as the churn period.

For our experiments, to get a subset of the non-churners, we have considered the customers who have been active on day 1 up to day 31. We have identified 106,002 non-churners and this forms our non-churner subset. Moving forward, we have considered two cases for churner subsets: Case I are churners who have been inactive after the first 15 days in the dataset, that is those customers who have been active in the first fortnight, but inactive for the next two weeks. Case II are churners who have been inactive after the first three weeks in the dataset. So, the time boundary for Case I is day 16 and the time boundary for Case II is day 22. The number of churners in Case I are 89,745 customers and in Case II are 97,859 customers. Subsequently we build churn prediction models for both Case I and Case II and prediction performance on the test data is reported for both. We have reported results on both cases as it validates and further strengthens the applicability of our model. Here due to the data limitations, we have considered users who have been inactive for a fortnight in the network as churners. As this inactivity includes no phone calls made or received and no SMS sent or received, we believe that it is a reasonable assumption given the widespread usage of mobile phones as a mode of communication between users.

We build various models to evaluate the effectiveness of our hybrid feature subsets including call usage patterns, social group features and the rate of change in usage and group features in predicting churn from customer behavioral patterns. For each case considered, we build models based on call usage features only (CF), combined call usage and its dynamics (changes in call usage features) (CDF), hybrid model with call usage combined with community or social group features (CSF) and combined hybrid model with call usage, social group and their dynamics (CSDF).

In the following subsections we analyze the performance of four trained models namely call usage features only (CF), combined call usage and its dynamics (changes in call usage features) (CDF), hybrid model with call usage combined with community or social group features (CSF) and combined hybrid model with call usage, community and their dynamics (CSDF), on the testing dataset and report the results of each of these models for both Case I and Case II.

B. Evaluation Criteria

We evaluate the model performance with the Receiver Operating Characteristics (ROC) plot and the Precision-Recall (PR) plot. Telecom service providers prefer models which have high precision due to the disparity in the costs. The cost of incorrectly classifying a churner is higher than the cost of classifying a non-churner incorrectly. A tradeoff between high

precision and a rational recall is preferred by the mobile telecom service providers so that they can optimize their budgets to achieve improved customer retention.

From the confusion matrix the following can be computed. A True positive is defined as the number of correctly predicted churners. A False negative is defined as the number of churners wrongly predicted as non-churners. A False positive is defined as the number of non-churners wrongly predicted as churners. A True negative is defined as the number of non-churners correctly predicted. Accuracy is the percentage of correctly classified customers over the total number of customers. True positive rate is defined as the fraction of churners predicted correctly. False positive rate is defined as the fraction of non-churners wrongly predicted as churners. Precision is defined as the fraction of predicted churners that actually turn out to be churners in the group the classifier has declared as churners. Higher precision implies that the classifier has committed lower false positive errors. Recall is defined as fraction of actual churners correctly predicted by the classifier. This is equal to true positive rate. The higher recall values imply that there are fewer number of customers misclassified as non-churners.

The Receiver Operating Characteristic (ROC) curve plots the fraction of churners correctly classified as churners versus the fraction of non-churners wrongly classified as churners. It captures the tradeoff between benefits and expenses. The performance of the model is considered to be best when the ROC curve is closer to (0, 1). A model that passes exactly through (0,1) has no false positives and false negatives. The area under the ROC curve (AUC) can be used to compare different models. The value of the AUC is between 0.0 to 1.0. Models that have greater AUC have better performance. Besides, since the AUC for the diagonal is 0.5 and represents random models, models with AUC value greater than 0.5 have performance better than that of random.

In the Precision and Recall plot (PR plot), Recall is on the x-axis with Precision on the y-axis. Recall is the True Positive Rate (TPR), and Precision computes the fraction of customers correctly classified as churners. This precision-recall curve can be plotted at various thresholds, which illustrates the trade-off of precision and recall for the model, at those varying thresholds. This allows two models to be easily compared using their precision-recall curves. A superior model during comparisons is the one whose PR curve is more to the top-right hand side of the plot.

C. Average Churn Probabilities Plot

The output of each of the trained models on the testing dataset is the predicted likelihood or probability to churn for each customer in the test set. We look at the visual analysis of average predicted churn probabilities across various customers in the test set for all the tested models for both Case I and Case II. We plot average churn probabilities on the x-axis and percentage of users or customers predicted with the corresponding churn probability on the y-axis and call it the Average Churn Probability plot. It gives us the distribution of churn probabilities across the test dataset. In general, the percentage of users, associated with higher churn probabilities will be classified as churners while the percentage of users associated with lower churn probabilities will be classified as

non-churners. Hence in a way this plot gives the distribution of probabilities for churners versus non-churners in the testing dataset as predicted by that model.

Fig. 2 shows the distribution of churn probabilities in the testing set for CF, CSF and CDF in Case I. It is of interest to note the shift in the churn probabilities corresponding to highest percentage of users most likely to churn from 0.75 to 0.98 and the highest percentage of users least likely to churn from 0.25 to 0.15 between CF, CSF models and CDF model. With the addition of rate of change features, we see the model predicting the churners with increased confidence probabilities. Also, the distribution of churn probabilities is better separated in our proposed model, which seems to point at improved class separation in the learned model when the novel hybrid features of both call patterns and social group features are considered for classification. Fig. 3 shows the distribution of churn probabilities in the testing set for CF, CSF and CDF in Case II. There is a shift in the churn probabilities corresponding to highest percentage of users most likely to churn from 0.78 to 0.85 and the highest percentage of users least likely to churn from 0.3 to 0.05 between CF, CSF and CDF models. With the addition of rate of change features, we again observe our proposed model predicting the churners with increased confidence probabilities.

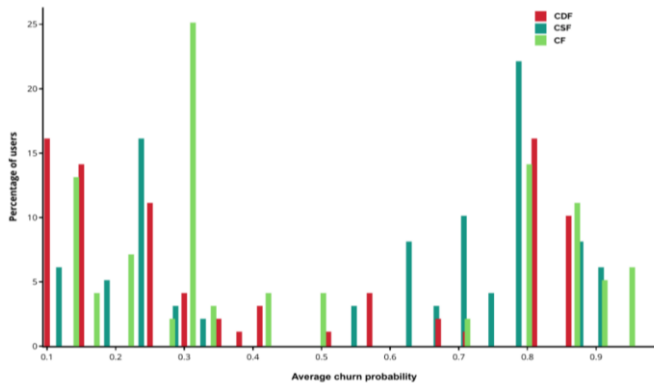


Fig. 2 Average Churn Probabilities Plot for Case I

CF: call usage features only, CDF: combined call usage and its dynamics, CSF: hybrid model with call usage combined with community or social group features

In both the cases considered it has been found that in the combined hybrid model the distribution of churners from non-churners is better observed than only call usage feature (CF) models. In the next section, we use the ROC curves to compare and analyze the four models.

D. ROC and PR Plots

As mentioned previously, The Receiver Operating Characteristic (ROC) curve plots the fraction of churners correctly classified as churners versus the fraction of non-churners wrongly classified as churners. The ROC captures the tradeoff between benefits and expenses. Fig. 4 and Fig. 5 compare the ROC curves and PR curves respectively of the four considered models for Case I. Fig. 6 and Fig. 7 compare the ROC curves and PR curves respectively of the four considered models for Case II.

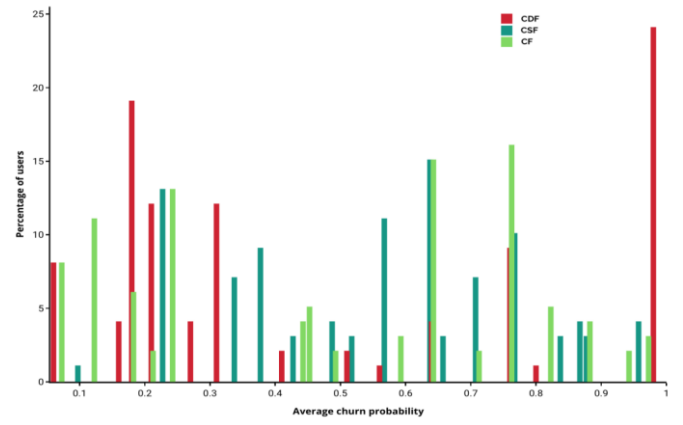


Fig. 3 Average Churn Probabilities Plot for Case II

CF: call usage features only, CDF: combined call usage and its dynamics, CSF: hybrid model with call usage combined with community or social group features

1) Case I:

As can be observed in Fig. 4 and Fig. 5, for Case I, where we train the model on two weeks of data, we get a significant improvement in the CSF model ROC (AUC=0.86) curve and PR curve (Precision=0.71, Recall=0.88 at threshold=0.5) over the CF model ROC (AUC=0.58) curve and PR curve (Precision=0.53, Recall=0.57 at threshold=0.5). AUC captures the class separation between churners and non-churners. The classes are perfectly separated if the AUC is 1 and randomly separated if AUC is 0.5. The higher values of AUC show that adding the social group-based features to the call usage patterns gives a significant improvement in prediction performance and that establishes the social influence as a key indicator for churn in customers. It can also be observed from Fig. 4 and Fig. 5 that there is a significant improvement in the CSDF model ROC (AUC=0.98) curve and PR curve (Precision=0.94, Recall=0.89 at threshold=0.5) over the CSF model ROC curve and PR curve. This shows the improved predicted power of our hybrid feature sets and significance of using the changes in community or social group sizes for predicting churn.

2) Case II:

As can be observed in Fig. 6 and Fig. 7, for Case II, where we train the model on three weeks of data, we get an improvement in the CSF model ROC (AUC=0.89) curve and PR curve (Precision=0.76, Recall=0.96 at threshold=0.5) over the CF model ROC (AUC=0.81) curve and PR curve (Precision=0.76, Recall=0.69 at threshold=0.5). Although the improvement is not as pronounced as in Case I, the AUC values are still high and it reinforces our earlier hypothesis of social influence as a key indicator for churn in customers. As in Case I, it can also be observed from Fig. 6 and Fig. 7 that there is an improvement in prediction performance in the CSDF model ROC (AUC=0.98) curve and PR curve (Precision=0.89, Recall=0.96 at threshold=0.5) over the CSF model ROC curve and PR curve.

Overall Fig. 4, Fig. 5, Fig. 6 and Fig. 7 demonstrate the improved performance of our combined hybrid model for Case I and Case II. This significant improvement is due to our additional features based on social groups and changes in both call usage patterns and social group-based features. It can be

seen that these features based on change patterns are able to capture subtle changes in behavioral patterns and thus possibly improving churn prediction performance. The call usage and social features alone may not give acceptable accuracies in practice, hence in our current work we propose the novel idea of also using their dynamics to improve the churn prediction accuracy. The results demonstrate the predictive power of using features based on changes in call usage patterns and social groups to augment the churn prediction models in the absence of other sources of information and more so when the duration of observed data used for feature extraction is short.

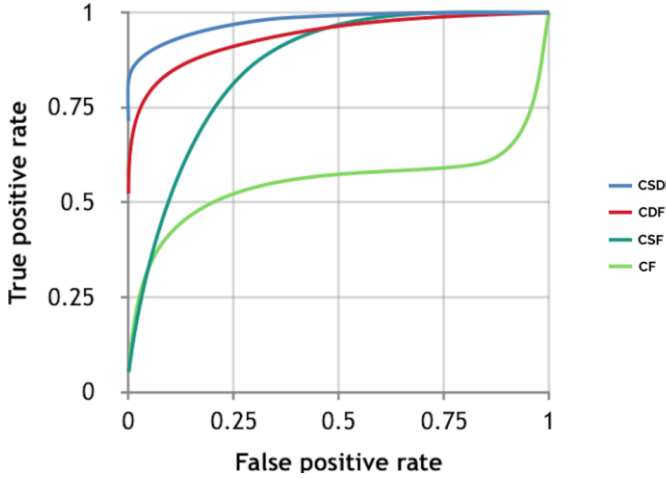


Fig. 4 Comparison of ROC plot for Case I

CF: call usage features only, CDF: combined call usage and its dynamics, CSF: hybrid model with call usage combined with community or social group features, CSD: combined hybrid model with call usage, community and their dynamics

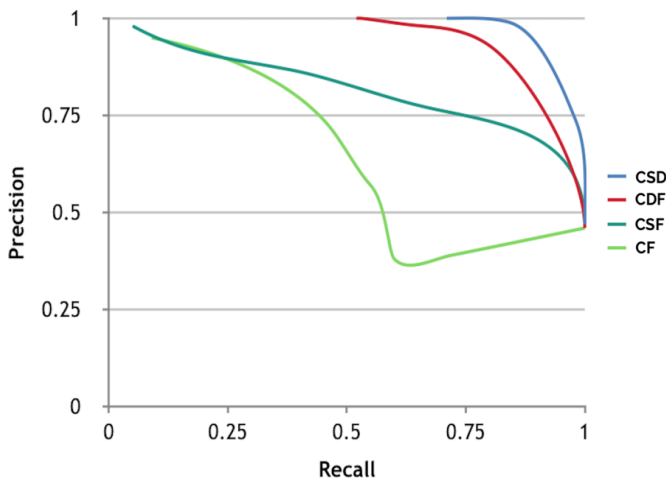


Fig. 5 Comparison of PR plot for Case I

CF: call usage features only, CDF: combined call usage and its dynamics, CSF: hybrid model with call usage combined with community or social group features, CSD: combined hybrid model with call usage, community and their dynamics

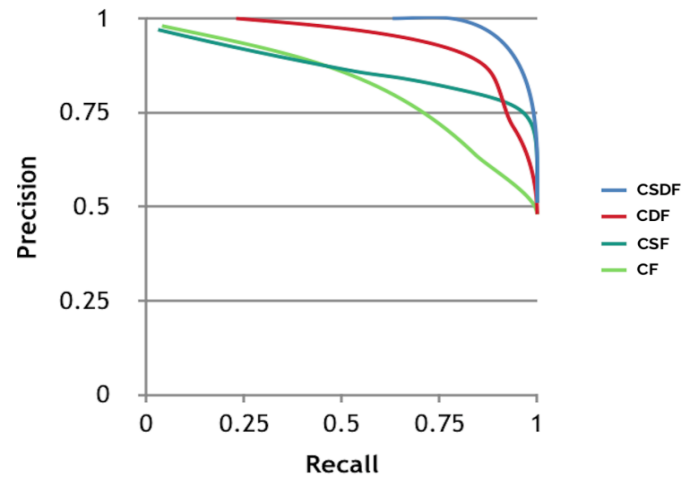


Fig. 6 Comparison of PR plot for Case II

CF: call usage features only, CDF: combined call usage and its dynamics, CSF: hybrid model with call usage combined with community or social group features, CSD: combined hybrid model with call usage, community and their dynamics

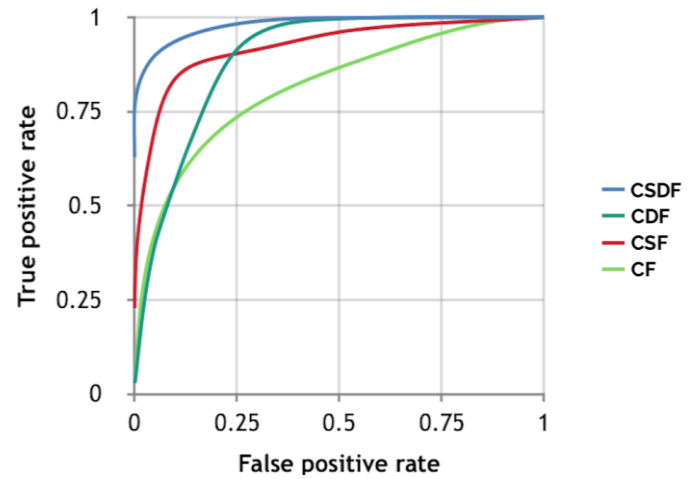


Fig. 7 Comparison of ROC plot for Case II

CF: call usage features only, CDF: combined call usage and its dynamics, CSF: hybrid model with call usage combined with community or social group features, CSD: combined hybrid model with call usage, community and their dynamics

VII. CONCLUSION AND FUTURE WORK

In this work some of the novel features based on communities and change in features have been extracted from call data records only and used to train our churn prediction model. Our experiments showed that our proposed features based on social influence and changes in both call usage patterns and social groups are good indicators of churn and have been shown to perform better than models without these features on the same dataset. In both cases that we have analyzed, the prediction performance has significantly improved compared to the baseline models. Taking into consideration the data constraints imposed by the mobile telecom operator, we proposed the use of novel hybrid feature sets that are based on the features extracted from CDRs exclusively combined with the social group features to build an efficient churn prediction model

with superior performance over existing models with similar data constraints. Experiments were performed using gradient boosted trees algorithm on the customers for identifying potential churners. Despite the practical limitations, our results demonstrate the improved prediction performance with the inclusion of these novel combined behavioral feature sets.

Once we have calculated the likelihood of churn for each customer, the model can be applied as a tool for creating prioritized ranked list for targeted marketing campaigns. The proposed technique is extremely advantageous to mobile telecom operators that only have inadequate customer profiles and hence provides wider applicability compared to existing models. Since the data for many months is available with telecom operators, they can use it over longer timelines and improve predictions.

It is of future interest to perform empirical comparisons with customers from diverse geographical locations. Future efforts will also be directed toward promoting the generalizability of the proposed technique not only in terms of extending the analysis duration of call data but also for extending to other industries where churning is a concern as well, like credit card which come with their own unique set of features. Another promising application area is for estimating the interest of the public in consuming information disseminated by the public visits to blogs, websites and other social media could be used as a dataset while retaining the same techniques adapted using CDR data. It is of special interest to extend our model to be able to handle detection of extreme churn events associated with competitive schemes and technological advantages mooted by competitors. Employing big data analytic techniques can reveal incident and occurrence sequences that lead to churn, which in turn can be utilized as features that can be fed back into our traditional churn prediction models.

REFERENCES

- [1] Ahn, J. H., Han, S. P., & Lee, Y. S. "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*", 30(10), 552-568, 2006. <http://dx.doi.org/10.1016/j.telpol.2006.09.006>
- [2] Jonathan, L., Janghyuk, L., & Lawrence, F. "The impact of switching costs on the customer satisfaction-loyalty link: Mobile phone service in France." *Journal of Services Marketing*, 15(1), 35-48, 2001. <http://dx.doi.org/10.1108/08876040110381463>
- [3] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *Eur. J. Oper. Res.*, vol. 218, no. 1, pp.211–229, April 2012.
- [4] P. Datta, B. Masand, D. R. Mani, and B. Li, "Automated cellular modeling and prediction on a large scale," *Artif. Intell. Rev.*, vol.14,no.6, pp. 485–502, December 2000.
- [5] D. Popović and B .D.Bašić, "Churn prediction model in retail banking using fuzzy C-means algorithm," *Informatica*, vol. 33, no. 2, pp.235–239, May 2009.
- [6] M. Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts," *Expert Syst. Appl.*,vol. 37, no. 6, pp. 4710–4712, June 2010.
- [7] J. Burez and D. V. Poel, "Handling class imbalance in customer churn prediction," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4626–4636, April 2009.
- [8] K. Coussement and D. V. Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert Syst. Appl.*, vol. 34, no. 1, pp.313–327, January 2008.
- [9] N. Kim, K.-H. Jung, Y. S. Kim, and J. Lee, "Uniformly subsampled ensemble (USE) for churn management: Theory and implementation," *Expert Syst. Appl.*, vol. 39, no. 15, pp. 11839–11845, November 2012.
- [10] C.-P. Wei and I. T. Chiu, "Turning telecommunications call details to churn prediction: A data mining approach," *Expert Syst. Appl.*, vol. 23,no. 2, pp. 103–112, August 2002.
- [11] D. A. Kumar and V. Ravi, "Predicting credit card customer churn in banks using data mining," *Int. J. Data Anal. Tech. Strategies* , vol. 1,no. 1, pp. 4–28, August 2008.
- [12] P. Spanoudes and T. Nguyen "Deep Learning in Customer Churn Prediction: Unsupervised Feature Learning on Abstract Company Independent Feature Vectors," *CoRR*, abs/1703.03869, March 2017
- [13] A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjee, A. Joshi. "On the structural properties of massive telecom call graphs: findings and implications." *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 435–444, New York, NY, USA, ACM, 2006.
- [14] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjee, A. A. Nanavati, A. Joshi. "Social ties and their relevance to churn in mobile telecom networks." *Proceedings of the 11th international conference on Extending database technology*, pages 668–677, New York, NY, USA, ACM, 2008
- [15] Y. Richter, E. Yom-Tov, and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," in *Proc. SDM 10, Soc. Ind. Appl. Math.*, Philadelphia, PA, pp. 732–741, 2010.
- [16] C. Phadke, H. Uzunalioglu, V. B. Mendiratta, D. Kushnir, and D. Doran, "Prediction of subscriber churn using social network analysis," *Bell Labs Technical Journal*, vol. 17, no. 4, pp. 63–75, 2013.
- [17] Kusuma, P.D., et al. "Combining customer attribute and social network mining for prepaid mobile churn prediction." in *Proc. the 23rd Annual Belgian Dutch Conference on Machine Learning (BENELEARN)*, 2013.
- [18] K. Kim, C.-H. Jun, and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6575–6584, 2014.
- [19] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends," *Comput.Oper. Res.*, vol. 34, no. 10, pp. 2902–2917, October 2007.
- [20] W. Verbeke , D. Martens , C. Mues , B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Systems with Applications An International Journal*, v.38 n.3, p.2354-2364, March 2011
- [21] Hassouna, M., Elyas, T., & AbouTrab, M. S. "Customer Churn in Mobile Markets: A Comparison of Techniques." *International Business Research*, 8(6), 224-237, 2015. <http://dx.doi.org/10.5539/ibr.v8n6p224>
- [22] Óskarsdóttir, M., et al., "Social network analytics for churn prediction in telco: Model building, evaluation and network architecture". *Expert Systems with Applications*, 2017. 85: p. 204-220.
- [23] "Ipython API Documentation." Online Available: <http://ipython.readthedocs.io/en/stable/overview.html>.
- [24] "GraphlabCreate resources" <https://pypi.python.org/pypi/GraphLab-Create>
- [25] V. D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. "Fast unfolding of communities in large networks." In *Journal of Statistical Mechanics: Theory and Experiment*, page 10008, 2008.
- [26] Fortunato, S., and Hric, D. "Community detection in networks: A user guide." *Arxiv*, abs/1608.00163, 2016
- [27] B Venkatesh, SH Choudhury, S Nagaraja, N Balakrishnan, "BotSpot: fast graph based identification of structured P2P bots", *Journal of Computer Virology and Hacking Techniques* 11 (4), 247-261, 2015
- [28] Friedman, Jerome H. "Greedy function approximation: A gradient boosting machine." *Ann. Statist.* 29, no. 5, 1189—1232, 2001. doi:10.1214/aos/1013203451.