# A Hybrid Churn Prediction Model in Mobile Telecommunication Industry

Georges D. Olle Olle and Shuqin Cai

***Abstract*—In most industries where switching costs are prevalent, the landscape of activities is painted of Customers attrition or *churn* (Clients who want to switch or change their suppliers for various reasons). This phenomenon is ubiquitous in the telecommunication industry and every aspect related to it, leads to believe that it's steeply growing. As the market is fiercely competitive, and the number of prepaid customers is increasing, it is vital that companies proactively tackle the defection of their customers by determining behaviors that might ultimately create *churn*. In this paper we proposed a hybrid learning model to predict churn in mobile telecommunication networks. Experiments were carried out using WEKA a Machine Learning tool; along with a real dataset from an Asian mobile operator to evaluate the performance of the model. The results show that the new hybrid model is more accurate than single methods.**

***Index Terms*—Churn prediction, logistic regression, telecommunications, voted perceptron.**

## I. Introduction

Whatever IT industry you belong to, whether it's Mobile Service, ISPs, IT products or Social Networks, make no mistake; telecommunication customers continue to fully use their rights of switching from one service to another: That phenomenon is known as *Churn.* Among industries that are mostly hit, Mobile Telecommunication comes on top with more than 30% of defection in Europe; reaching 60% in African sub-Saharan countries and in Asia [1], [2]. The causes are among others: the technological development of telecommunications, the liberalization, the globalization and the fierce competition that have followed. An increasing churn rate is considered as the plague of all carriers because losing a high valued client means a loss of future incomes. In China for instance, if a customers' category spends CNY 200 /month and with 10 million customers in that category, then 0.5% churn rate is CNY 1million/month of lost income from then on. Also, the cost of winning a new customer is 5 to 10 times higher than the cost of satisfying and keeping an existing customer [3]. Therefore, customer retention has become the leitmotiv of marketing campaigns. It's more profitable for mobile telecom operators to invest in those customers that already have an experience with the service by renewing their trust, rather than constantly trying to attract new customers characterized by a higher churn rate. Mature Mobile Operators (MO) despite possessing a churn prediction system that can detect their churners, still cannot accurately explain in time the moment when their deactivation shall occur, neither can they produce with certitude the reason why a particular customer wants to give up. Tremendous volume of data is being generated from mobile telecom networks; however, customer's data is complex and under privacy regulation, whereby very limited information about the subscriber or his experience with the service is available. Mindful the conjuncture, most of the studies on churn prediction mainly focus on using customer base and Data Mining (DM) techniques to detect customer's behavior related to the churn event. Basically, they can detect the customers with high propensity to churn, but not necessarily providing the reason of churn [4], [5]. It's necessary for companies to know the reason of churn before applying the retention strategy [6]. Also, a survivability analysis is usually required when it comes to know the possible deactivation time. But, most prepaid customers would have churned already long ago before the company will realize [7], [8]; and most companies are aware of their postpaid-client's churn only at the moment the latter refuses to keep on with the contract. To the best of our knowledge, an efficient churn prediction model that raises the trilogy: Who wants churn; why does he or she want to churn and when would that happen, is very less found in the research of churn prediction. Accordingly, the goal of this study is to show that hybrid models built on DM techniques can explain the churn behavior with more accuracy than single methods; and that in some extend the reason of churn can be revealed, as well as explaining the gap between the decision to churn and the deactivation time. Our hybrid model uses Logistic Regression in parallel with Voted Perceptron for classification, and combined with clustering. The model is built using WEKA, a well-known tool of Machine Learning. 2000-instances of a real world dataset with 23 variables from an Asian mobile operator are used for evaluation. In particular, the moment of deactivation is subsequent to an overseen grace period, favorable to the measures adopted to proactively and efficiently interfere. The remainder of this paper is organized as follows: In Section II we provide the background on churn prediction and present the selected Data Mining techniques. Section III describes the dataset and process of the proposed method. In Section IV the model evaluation is made with discussions, and Section V concludes the study.

## II. Background on Churn Prediction and DM Methods

### A. Customer's Data and Churn Prediction

The new market trend has been reshuffled from product-

Manuscript received June 6, 2013; revised August 12, 2013.

G. D. O. Olle and S. Q. Cai are with the Department of Management Sciences and Information Management, Huazhong University of Science and Technology, 430074-Wuhan, PR–China (e-mail: georgesdelortolleolle@gmail.com, caishuqin@sina.com.cn).

DOI: 10.7763/IJEEEE.2014.V4.302

based to now customer-first based strategy, also known as *customer-centricity*. Despite the fact that this concept accordingly with the true customer knowledge is still elusive [9], the causality is well explained by the evolving customer needs and behavior, which when not fed and satisfied lead to churn There are several reasons why people churn in the mobile telecommunication sector and the churn taxonomy considers two main groups of churners: Involuntary churn (unpaid bills, fraudulent behavior) which is initiated by the service provider party; and the Voluntary Churn that means on one hand deliberately switching to competitor for many reasons; and one the other hand incidentally switching when moving to another geographic area, or due to Change in the financial situation. Detecting behaviors associated with deliberate churn is the great concern of MO [6], [10]-[13]. Rather than to show all contexts that drive deliberate churn, in this paper we just analyzed the variables available in the dataset and have represented the tributary effects of churn in its wide taxonomy. Thus, Fig 1 is a compiled taxonomy of churn, in which the context mentioned are the levers that convince subscribers to stay or deliberately to leave the MO that we've investigated. It's also important to mention the necessity to measure the influence of an unsatisfied customer in his/her call graph using Call Details Records (CDR). [14] Argues that churn is a social phenomenon and that the churn value of a subscriber or customer, is a combination of influence on his own usage patterns and his neighbors' usage activities. A customer base that possesses many attributes with less mutual information for purpose of churn management is a bonus.

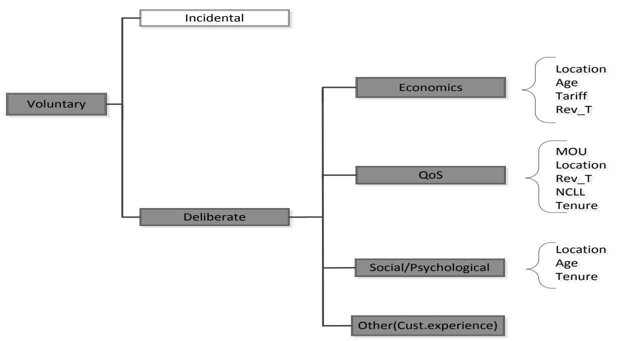


Fig. 1. Compiled churn taxonomy.

A wide variety of techniques have been adopted to detect churn behaviors in different contexts including DM, Social Networks Analysis, Genetic Programing (GP), and Generalized Additive Models. With DM techniques, it basically consists of building a prediction model based supervised or unsupervised learning using customer data. Data Preparation step is usually the most time consuming phase in data analysis and DM processes in particular. Raw data are not suitable for DM and most Telecom Data are time-series data features; a summary of features is necessary and critical before the DM task in order to obtain a useful description of the customers [15]. Data should be integrated and cleaned from redundancy because there could be a burden training the model with all the features. Also, features can be transformed to create new variables in order to get more information out of the customer data [16]. Variable extraction for model training is very important because some variable might have negative impact on the prediction ability of the model. The selection can be made assigning weights for collinearity to the predictive feature [17] or by establishing correlation among attributes using statistics method for information gain like Chi-squared test [18]. A most convenient feature set for model training though not easily obtained, can improve the prediction rate of True Positive (TP) while logically reducing the False Positive (FP).

### B. Related Works

There are many researches on churn predictive model in the telecommunication. [19] Investigates the causes of telecom churn using Fuzzy Logic. [18] used Ordinal Regression to model customer's satisfaction and predict churn time and compared the results with those of survival analysis methods. [11] proposed an Artificial Neural Network (ANN) integrated prediction model for prepaid customers that could explain the reason of churn using the data set of complaints records from subscribers and thus the appropriate measure to be taken for retention strategy. Many other studies opt for a hybrid model and in general, hybrid methods in Predictive Data Mining techniques follow two strategies of serial combination. In the first strategy, the input from the preprocessing step passes through a classification technique and subsequently to a clustering phase to get the final output [20]. In those models, the classification technique like Decision Trees (DT) can be used for dimensionality reduction before the clustering, just as a supervised filter to select the attributes. Alternatively, they can be directly used for prediction; the clustering is latterly used to detect outliers or even to derive new patterns in labeled groups [9], [11]. The second strategy uses clustering technique as a preprocessing step prior to classification. In Some methodologies, the two classification techniques are being cascaded respectively for attribute selection and prediction. This is done isolating the first method to accomplish specific preprocessing task, like finding the correlation among the attributes to optimize the information gain; or finding their collinearity to build the causality towards the predicted variable [13]. This practice is proved to be very useful to identify causalities especially when applying Bayesian Belief Networks (BBN). Although high credit can be attached to such models when it comes to build the attributes that most likely influence the event to be predicted, yet the maximum likelihood of that event is disregarded. In some scenarios, like in telecom data this aspect need to be investigated because placing attributes as linearly dependent variables does not forcibly embrace all the aspects raised by the churn taxonomy. Neural Network (NN) along with DT and Support Vector Machine (SVM) were used in [21] for the construction of a hybrid model for churn prediction in telecommunication. [22] Proposed a hybrid learning method as a combination of DT and GP to derive the rules of classification based on customer behavior.

All These models though robust, mostly were restricting the churn phenomenon. Good churn prediction system should not only pinpoint potential churners successfully, but further provide the reason and the moment of churn in its predictions [23]. In fact, when the company deactivates a prepaid user, in most cases he has already churned long ago.

The issue is also discussed in our study when attempting to build the gap between the churn and the deactivation times.

### C. Data Mining Methods

The two supervised classifiers selected are Logistic Regression a well-known cookie and widely applied for prediction in supervised models, along with Voted Perceptron which is a single node ANN algorithm also applied in prediction.

#### 1) Logistic regression (LR)

For a given instance, the hypothesis $h_\theta(x)$ of prediction is given by the probability estimated for each class. To understand how LR classifies an instance to be of class 1 or class 0 according to the estimated probability when the dataset has many features, let's predict that the instance will belong to class 0 for $h_\theta(x) < \beta$ , with β a value of interval [0-1]. If the condition is not satisfied, then the instance definitely must belong to class 1. Firstly, the hypothesis $h_\theta(x)$ will be estimated by equation (1):

$$\hat{y} = h_\theta(x) = \frac{1}{1+e^{-\theta^T X}} \tag{1}$$

$$for, y \in \{0,1\}$$

$$P(y/x;\theta) = [h_\theta(x)]^y [1-h_\theta(x)]^{1-y} \tag{2}$$

where $\theta^T x$ the product matrix of parameters and vector associated to all features is generated at that instance. The probability $P(y/x; \theta)$ to know whether the prediction y is class 0 or 1 is then obtained from equation (2).

#### 2) Voted perceptron (VP)

The Perceptron model is a single layer and simplest feed forward of ANN. The Inputs node is used to compute the attributes and output node is used to represent the model output. The computation in the input node respects the function

$$\hat{y} = sign(\theta x) \tag{3}$$

Voted Perceptron is a type of linear classifier which means that its makes its predictions based on a linear function, combining a set of weights with future vector describing a given input. It processes elements in the training set one at the time. Let's consider the true output classification of an $i^{th}$ instance of a training set to be $y_i$ and the predicted output be $\hat{y}_i$ . The perceptron algorithm will meet the consistency of the predicted value $\hat{y}_i$ with the true value $y_i$ by using the update formula:

$$\theta_j^{(k+1)} = \theta_j^{(k)} + \alpha(y_i - \hat{y}_i^{(k)})x_{ij} \tag{4}$$

where: $\theta_j^{(k)}$ is the weight associated with the instance for the attribute j after the $k^{th}$ iteration. $\theta_j^{(k+1)}$ Is the new (updated) weight associated with the instance; $(y_i - \hat{y}_i^{(k)})$ is the predicted error after the $k^{th}$ iteration; $x_{ij}$ is the input attributes at the instance *i*. $\alpha$ is the learning rate. The model learns throughout the whole training instances by adjusting for each instance the weight vector in order to be consistent with the true output of the example.

## III. Data and Hybrid Model Description

### A. Data Description

The Dataset used in this study was provided by an Asian mobile telecommunication operator. It recapitulates 6-month activity of 2000 subscribers, over 23 different data variables well described in the next lines. As shown in Fig. 2, 70% where used to train the model and the rest for testing. The Class distribution registers 1466 non-churners and 534 churners. The Data selection is made according to the empirical research on what features are more related to churn event in mobile telecommunication as explained in the introduction (Section I); also with some experts guidance and of course the availability of the required attributes.

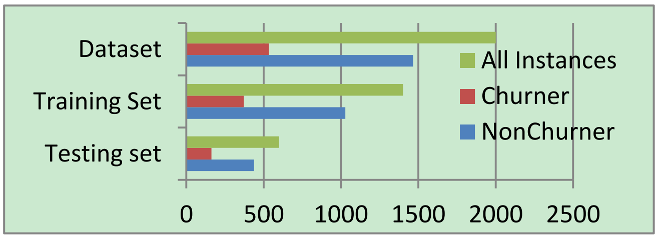


Fig. 2. Data distribution.

Following are the details on different attributes and class distribution of the raw data available.

**Rev_1-6:** Represent the revenue (billing amount) of the $n^{th}$ month. **Rev_T:** The total revenue (billing amount) of the last year. **NCLL_1-6:** Represent the number of outgoing calls during the $n^{th}$ month.

**MOU_T1-6:** Represent the total number of minutes of usage during the $n^{th}$ month. **Location**: Represents the place where most of the subscriber's activity in the network is localized. Customer experience with the operator and the usage of their rights of switching is highly related to their location. [24]. **Age**: As another demographic variable affecting churn. The mobile operator investigated has mention the fact that high propensity to churn is young people concern. The dataset counts 271 churners with age lesser than 45.

**Tenure:** This variable measures the age of the subscriber in the network. **Tariff**: It represents the type of billing available and adopted by individual subscribers. The first tariff (A) is profitable for customers with high number of outgoing calls on a minute pulse basis, as the unit cost per minute is cheaper than the second tariff (B) which has a lower constant fee. The elaboration of the final attribute set to train the model is explained in the next part of this section.

## IV. Description of the Hybrid Model (H M)

Our Research Framework adopts the second strategy of hybrid methods mentioned in Section II; with a slight difference that we combine in parallel two methods to reinforce the classification (Fig. 3). LR is used as major classifier and VP reinforce the prediction. The business objective determines the data mining task and the relevance of some variables. The Data preprocessing step transforms the selected input data in an appropriate format for analysis in WEKA knowledge flow. Attributes to be used for further steps are selected by first getting the predictive ability of each attribute described earlier from raw data. This step has contributed to obtain the information gain of each attribute. The results are generated by applying the Greedy stepwise forwards evaluation in WEKA. Afterward, the average of usage, revenue and number of call is elaborated and the correlation between the attributes and the class feature is separately established using the Ranker search methods to reduce information redundancy. Accordingly, 8 variables including by order: Tenure, AMOU_1-6, Rev_T, Age, Arev1-6, Tariff, ANCLL and Location were selected as their gains are higher with minimum redundancy. The MOU, NCLL and Rev over the 4th month were also added as the Rev on the 4th month was ranked the 3th attribute when applying the information gain on all the attributes on raw dataset. Selected continuous variables have large number of values and it's convenient to reduce the number of categories in those variables. Therefore, we've applied the discretization. Finally, 11 variables (Table IV and Table V) are used to train our model. K-means clustering is used to derive other patterns describing the reason of churn by building 4 groups over the sparsest attributes like month of usage and calling patterns.

**Confrontation stage:** The output predictions are afterward stored and balanced by a confrontation subpart that is elaborated to satisfy the condition, before choosing the final output of the prediction over the instances that create the conflicting predictions.

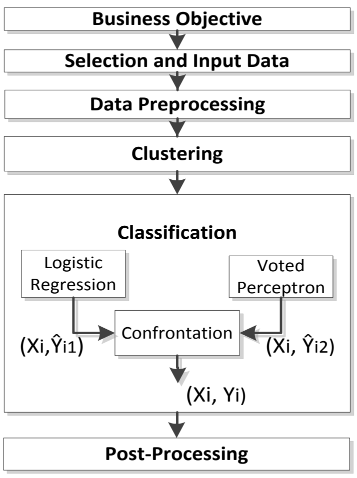


Fig. 3. Research framework.

This is done to reevaluate the false classification from the major classifier with a comparative prediction from the other classification.

$$P(y=0/\hat{y}=0)=TN/TN+FN \quad or \quad P(y=1/\hat{y}=1)=TP/TP+FP \tag{5}$$

Explicitly, since the two methods train their model differently as described in Section II and generate two predictions, the model utilizes the information of confusion matrices from each method where the predictions are not matching. The rates of false true positive or negative predictive value are compared over the class labeled by the major classifier. In case of contrast, the prediction method with the higher probability is selected. For instance, if the major classifier has predicted the instance class "zero" (or One) then the rates of True Negative (respectively True Positive) classification from each learner are compared. This represents the post probability of the class given the fact that the classifier has predicted the class. The technique measures how well the classifier rules churners and non-churners. We also impose the methods to process incrementally and follow the order of the instances to avoid instance order incompatibility. Simply to say, this requirement can ensure that we are not confronting $y_{i1}$ with $y_{j2}$ in place of $y_{i1}$ with $y_{i2}$. Since our two methods are independent but paralleled over the instances, the outputs of each single model are stored and some false predictions from the major classifier will be reexamined at the confrontation level before the ultimate classification.

TABLE I: CONFUSION MATRICES OF THE MODELS

| | | Churner | NChurner |
|---|---|---|---|
| VP | Churner | 5 | 157 |
| | NChurner | 2 | 436 |
| LR | Churner | 27 | 135 |
| | NChurner | 16 | 422 |
| **H.M** | **Churner** | **34** | **123** |
| | **NChurner** | **21** | **422** |

## V. RESULTS AND DISCUSSION

At the classification task, the results show that the hybrid mode (HM) built using LR and VP performs better than each of the individual methods. The precision of the HM is higher than the precision of the two other models and some false predictions are effectively corrected by confronting the predictions. The Receiver Operating Characteristic Curve (ROC) obtained for evaluation is shown in the Fig. 5 in comparison with Fig. 4. The area under curve is increased and more close to the upper left part and the correctly classified instances is also increased as shown in the confusion matrices (Table I). However the difference is not so high due to the number of contradiction predictions from the two algorithms. Also, the size of the dataset and the skewed distribution in the training and test datasets influences the results of the final prediction. In our perspective, we do not expect to ave all false classifications to be corrected, since it's utopic to get 100% of accuracy. Also, the churn taxonomy is broad and it's fair to say that noise might be due to some unavailable features that maybe would have contributed to the case that is being investigated. However, this result can be ameliorated by using a much bigger dataset or by employing other DM techniques at the lieu of LR and VP where the learning methods would be very different, in order to expect more cases of confrontation. Table II presents the detailed results of accuracy. In our clustering task, we regroup the samples in 4 groups according to the predominance on the eleven attributes that are most deterministic of the churn behavior (Table III). Table VI shows the different clusters obtained. The churners are concentrated in the first, the third and fourth groups. In the first group churners are in majority and are characterized by less tenure and high number in minutes of usage; consequently, we can observe that tariff A is the choice they've opted for billing system. Churners of that group might mostly raise concern about revenue of each month which is higher The retention strategy shall aim to propose incentives that alleviate the revenue while keeping high the usage, like free calls. The second group has few churners and is more financially stable. The third group shows that the bigger portion of subscriber have the revenue between 72- 318 and 75 %( 172/215) of group members are churners. It's also trivial that most subscribers are aged lesser than 34. They show social and economic instability as it's shown by their revenue. Retention shall propose incentives that satisfy their billing systems and help to grow their tenure to expect stable usage attitude in the future.

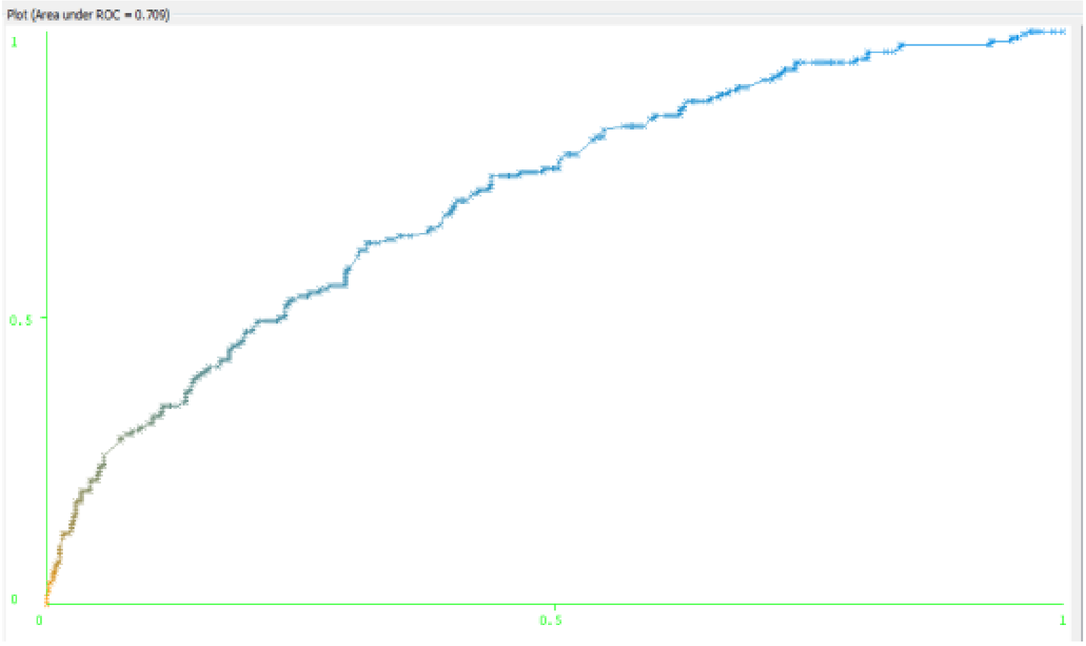


Fig. 4. Area under ROC–LR

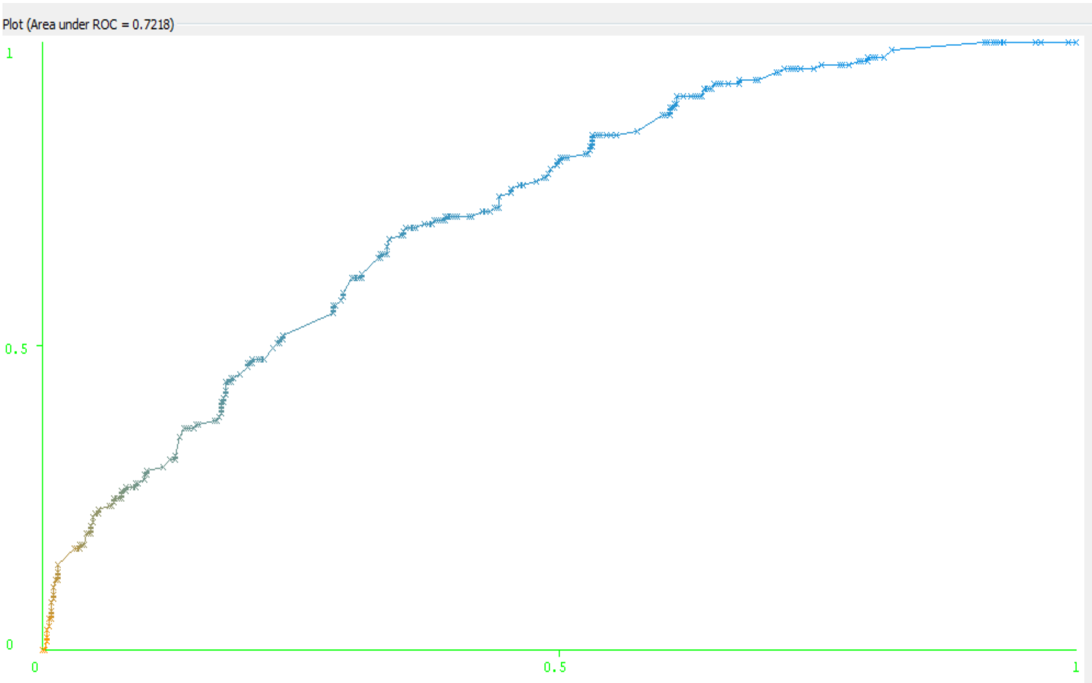


Fig. 5. Area under ROC hybrid model (HM).

During the experimental process, we observed that our hybrid model shows almost the same accuracy results when we use only the first 4 months of activity. The standard deviation over the attributes shows that the churn event is highly related to the Tenure, Age and Total revenue during the fourth month (Table IV). Thus, with subscriber behavior during the first 4 months and by adding the 5$^{th}$ and 6$^{th}$ month's activity, the accuracy results on churn classification does not change. Since we could not follow individual churner's activity from the fourth month, we regrouped churners with similar activity evolution during that period. Table V is the confusion matrix obtained for 4 months activity. Accordingly, the Clustering of churner's activity in Table VII reveals 3 groups with specific characteristics on their activity from 4th to 6th months. When Tenure is lower than 941 days, subscriber's activity is quite instable. For Group 2, after the 4th month the AMOU initially in the interval 122-245 degreases to lesser than 119 for Group 2 till the 6th month when deactivation is evident. The number of call after month 4 gradually decreases to 5, irrespectively with the group. However for Group 1 where tenure is higher than 941 days has very few churners. These results along with the table of confusion matrix indicate that most churners evaluate the service for 4 months considering the origin of the period where data are collected. Even though activities are registered after the first period they do influence much on the churn behavior and can be assumed as to be a grace period to convince to retain the subscriber from churning.

TABLE II: DETAILED ACCURACY OF THE MODELS

| | Correctly Classified Instances | TP Rate | FP Rate | Precision | Recall | ROC Area | Class |
|---|---|---|---|---|---|---|---|
| **VP** | **441 73.5 %** | 0.031 | 0.005 | 0.714 | 0.031 | 0.513 | Churner |
| | | 0.995 | 0.969 | 0.735 | 0.995 | 0.513 | N-Churner |
| | | 0.735 | 0.709 | 0.730 | 0.633 | **0.513** | Weight-Avg. |
| **LR** | **449 74.83 %** | 0.167 | 0.037 | 0.628 | 0.167 | 0.571 | Churner |
| | | 0.963 | 0.833 | 0.758 | 0.963 | 0.572 | N-Churner |
| | | 0.748 | 0.618 | 0.723 | 0.748 | **0.71** | Weight-Avg. |
| **H.M** | **456 76 %** | **0.217** | **0.047** | **0.618** | **0.217** | **0.720** | Churn |
| | | **0.953** | **0.783** | **0.774** | **0.953** | **0.722** | N-Churner |
| | | **0.760** | **0.591** | **0.733** | **0.715** | **0.721** | Weight-Avg. |

TABLE III: RANKED ATTRIBUTES AND GAIN RATIO EVALUATION

| MOU_T4 | Rev_T | Tenure | AMOU1-6 | Rev _4 | Arev1-6 | Age | ANCLL | NCLL4 | Tariff | Location |
|---|---|---|---|---|---|---|---|---|---|---|
| **0.03187** | **0.03012** | **0.02949** | **0.02637** | **0.0254** | **0.01837** | **0.01751** | **0.01114** | **0.00852** | **0.00274** | **0.00106** |

TABLE IV: CORRELATION OF SELECTED ATTRIBUTES WITH THE CLASS

| Tenure | AMOU1-6 | Rev_T | Age | ARev1-6 | Rev_4 | MOU_T4 | Tariff | ANCLL | NCLL4 | Location |
|---|---|---|---|---|---|---|---|---|---|---|
| **0.1955** | **0.1116** | **0.109** | **0.106** | **0.1028** | **0.097** | **0.0964** | **0.0586** | **0.0396** | **0.0329** | **0.0329** |

TABLE VI: CLUSTERING RESULTS

| | Tenure | Age | Tariff | AMOU1-6 | ARev1-6 | Rev-T | CuSTATUS |
|---|---|---|---|---|---|---|---|
| **Group1** (196 Sub)10% | 18 (<158)<br>**153(158-1908)**<br>25 (>1908) | 49 (<33.5)<br><br>**147(>=33.5)** | **142(A)**<br>52(B)<br>2(Ntype) | 50(<37.47)<br>**146(>=37.47)** | 66(<21.64)<br>**130(>=21.64)** | 14(<72.27)<br>34(72.27-318.94)<br>**148(>=318.94)** | **151(Churn)**<br><br>45(NChurn) |
| **Group2** (864Sub)43% | 8<br>220<br>**636** | 118<br><br>**746** | **610**<br>238<br>16 | 53<br>**811** | 74<br>**790** | 1<br>19<br>**844** | 65(Churn)<br>**799(NChurn)** |
| **Group3** (215 Sub)11% | 37<br>**166**<br>12 | **139**<br><br>76 | 35<br>**180**<br>0 | 78<br>**137** | **171**<br>44 | 16<br>**172**<br>27 | **164(Churn)**<br><br>51(NChurn) |
| **Group4** (725 Sub)36% | 17<br>115<br>**593** | 28<br><br>**697** | 28<br>**697**<br>0 | 336<br>**389** | **669**<br>56 | 1<br>**649**<br>75 | 154(Churn)<br>**571(NChurn)** |

TABLE VII: CLUSTERING RESULTS FOR 4 AND 6 MONTHS OF CHURNER'S ACTIVITY

| | Tenure | Age | Rev_4 | AMOU1-4 | AMOU1-6 | NCLL4 | NCLL5 | NCLL6 |
|---|---|---|---|---|---|---|---|---|
| **Group1** (45 Churn)8% | 17(>941d ) | 12 (33-40) | 25 (<38) | 21 (< 122) | 23(<119) | 36 (<=8) | 31 (<=6) | 30(<=5) |
| **Group2 (433Ch)81%** | 131(<=941) | 40-47 | 414(<38) | 414(122-245) | 410(<119) | 408 (<=8) | 388 (<=6) | 365(<=5) |
| **Group3** (56Chu)10% | 33(<=941) | 26-33 | 27(<38) | 42(122-245) | 31(119-238) | 33 (<=8) | 33 (<=6) | 30(<=5) |

TABLE V: CONFUSION MATRIX HM (4 MONTHS)

| | Churners | N-Churners |
|---|---|---|
| **Churner** | **29** | **128** |
| **N-Churner** | **16** | **427** |

## VI. CONCLUSION

This paper presented a new hybrid model for Churn prediction that predict customers with high propensity to churn, profiling the reason of churn and examining the gap between the churn decision and the deactivation time. In theory it contributes to the problem of increasing TP and decrease FP for more accuracy on the prediction by using a hybrid model. The evaluation of the model shows that its accuracy is higher than when using single model and that the results could be ameliorate when the data distribution are less skewed. The clustering results show that the churner's profile and the gap between the decision to churn and the deactivation time can be explained as a grace period for the company to provide better incentives. The case we investigated shows that 4 months period where enough for a churn to show his dissatisfaction before the deactivation; this depending on the Tenure and the data collection periodicity. This study also contributes to the expanding and newly adopted mixed model perspective in Data Mining applications and especially for the churn prediction in mobile telecom.


## ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China under Grant 71071066, and by the Social Science Foundation of Ministry of Education of China under Grant 11YJA630098.


## REFERENCES


[1] Winning the war against churn: How to use churn prediction techniques to improve customer retention. [Online]. Available: http://www.invest-aura.com/2011/02/winning-the-churn-reduction-war-how-to-use-churn-prediction-techniques-to-improve-customer-retention/

[2] Industry customer churn rate increases 15 percent. [Online]. Available: http://www.geoconnexion.com/news/industry-customer-churn-rate-increases-15-percent

[3] T. J. Gerpott, W. Rams, and A. Schindler, "Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunication market," *Telecommunications Policy*, vol. 25, no. 4, pp. 249-269, 2001.

[4] K. Clark, "What price on loyalty when a brand switch is just a click away?" *Qualitative Market Research: An international Journal*, vol. 4, no. 3, pp. 160-168, 2001.

[5] F. F. Reichheld, "Learning from customer defections," *Harvard Business Review*, vol. 74, pp. 56-70, 1996.

[6] A. Khalatyan, "Churn Management in Telecommunications," *From the Faculty Editor, 21, channel*, vol. 5, pp. 21-26, 2010.

[7] J. X. Lu, "Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS," in *Proc. SAS User Group International*, 2002, pp. 27-33.

[8] S. Rosset, E. Neumann, U. Eick, and N. Vatnik, "Customer lifetime value models for decision support," *Data Mining and Knowledge Discovery*, vol. 7, no. 3, pp. 321-339, 2003.

[9] R. B. Johnson and A. J. Onwuegbuzie, "Mixed methods research: A research paradigm whose time has come," *Educational Researcher*, vol. 33, no. 7, pp. 14-26, 2004.

[10] Y. Richter, E. Yom-Tov, and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," in *Proc. SIAM International Conference on Data Mining*, 2010.

[11] J. O. Daramola, O. O Oladipupo, and G. A. Musa, "A data mining process framework for churn management in mobile telecommunication industry," *IJCSS*, vol. 1. no. 1, 2009, pp. 34-40.

[12] C. Archaux, A. Martin, and A. Khenchaf, "An SVM based churn detector in prepaid mobile telephony," in *Proc. IEEE International Conference on Information and Communication Technologies: From Theory to Applications*, 2004.

[13] B. Q. Huang, T. M. Kechadi, B. Buckley, G. Kiernan, E. Keogh, and T. Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3657-3665, 2010.

[14] X. Long, W. Yin, L. An, H. Ni, L. Huang, Q. Luo, and Y. Chen, "Churn analysis of online social network users using data mining techniques," in *Proc. International Multi Conference of Engineers and Computer Scientists*, vol. 1, 2012.

[15] G. M. Weiss, *Data Mining in Telecommunications*, Data Mining and Knowledge Discovery Handbook, Springer US, 2005. 1189-1201

[16] X. Zhang, J. Zhu, S. Xu, and Y. Wan, "Predicting customer churn through interpersonal influence," *Knowledge-Based Systems*, vol. 28, pp. 97-104, April 2012.

[17] P. Kisioglu and Y. I. Topcu, "Applying Bayesian belief approach to customer churn analysis: A case study on the telecom industry of Turkey," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7151-7157, June 2011.

[18] R. Gopal and S. Meher, "Customer churn time prediction in mobile telecommunication industry using ordinal regression," *Advances in Knowledge Discovery and Data Mining*, 2008, pp. 884-889.

[19] K. H. Liao and H. E. Chueh, "Applying fuzzy data mining to telecom churn management," *Intelligent Computing and Information Science*, 2011, pp. 259-264.

[20] C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12547-12553, Dec 2009.

[21] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, "A proposed churn prediction model," *International Journal of Engineering Research and Applications*, vol. 2, no. 4, pp. 693-697, 2012.

[22] V. Yeshwanth, V. V. Raj, and M. Saravanan, "Evolutionary churn prediction in mobile networks using hybrid learning," presented at Twenty-Fourth International FLAIRS Conference, March 2011.

[23] H. Hohwald, E. Frias-Martinez, and N. Oliver, "User modeling for telecommunication applications: Experiences and practical implications," *User Modeling, Adaptation, and Personalization*, 2010, pp. 327-338.

[24] D Radosavljevik, P. van der Putten, and K. K. Larsen, "The impact of experimental setup in prepaid churn prediction for mobile telecommunications: What to predict, for whom and does the customer experience matter," *Transactions on Machine Learning and Data Mining*, vol. 3, no. 2, pp. 80-99, 2010.


**Georges D. Olle Olle** was born in Sangmelima, Republic of Cameroon, in 1980. He received the B.S. and M.S. degrees in Physics from University of Yaounde 1, Cameroon in 2007. In 2010 he received a Master of Engineering in Information and Communication Systems from Huazhong University of Science and Technology (HUST), Wuhan, China. His Master's thesis was related to RFID technology and the design of Reader Management Module for RFID Middleware. A project sponsored by the "863 high tech Projects" of China. He is currently Pursuing a PhD degree in Management Science and Engineering at Huazhong University of Science and Technology, Department of Management Science and Information Management. He is a student in Information System Research. He is also a member of Enterprise Business Intelligence and Engineering Institute (EBIE) at HUST. His research focuses on Churn prediction behavior and on inferring inter personal influence and its diffusion among users of same and different telecom operators. His interests also include multi relation social network analysis with Telecommunication data and its implications. His research participates in the design and optimization of information management system and other decision systems. He is also interested in Mobile Telecommunication, RFID middleware design, Data Mining, Machine Learning and Big Data analysis.

**Shuqin Cai** was born in 1955 in China. She is a professor in the department of Management Science and Information Management at school of Management, Huazhong Univertsity of Science and Technology (HUST). She received a PhD in Management Sciences and Engineering in 1996 from HUST.

She is the director of EBIE Institute. She teaches

Management Information System Analysis and Design, Logistic Information System, E- commerce and Internet Marketing. Her research interests include design of CRM, E-Commerce; IT based opportunity management, Networking, Enterprise Information Technology, and Government Information Technology, Opportunity/Risk identification methods, Business process reengineering, Social Networks, Web 2, Data Mining, Hyper graph, Cloud computing and others. She is the author and co-author of many Books and hundreds of articles in Logistic Information and Information system design, Marketing Intelligence and Planning, and Industrial Informatics.

Professor Cai is a committee member of the Wuhan International Conference of E-business and reviewer of many journals. She has visited many Business schools overseas such as University of Greenwich (UK) and Erasmus University (Netherland). She has designed many Information Systems for Governmental Institutions and Enterprises in China. She has completed several projects about “863 high tech projects” and has successfully completed 8 funding projects of National or Provincial Natural Science Foundation in China. She received the Wuhan Science and Technology progress award in 2007 and in 2012.