A Case Study for the Churn Prediction in Turksat Internet Service Subscription

Mehmet Gök
Turksat Corporation
TOBB University
Ankara, Turkey
mgok@etu.edu.tr

Tansel Özyer
TOBB University
Ankara, Turkey
ozyer@etu.edu.tr

Jamal Jida
Lebanese University,
Lebanon
jamal.jida@ul.edu.lb

***Abstract*— Churn prediction is a customer relationship process that predicts for customers who are at the brink of transferring all the business to competitor. It is predicted by modeling customer behaviors in order to extract patterns. An acquaintance of a customer is more costly than retainment of an existing customer. Churn predictions shed light on members about to leave the service and support promotion activities. These attempts are utilized to avoid subscription cancellation of existing customers. Nowadays, telecommunication companies take churn prediction very serious. They strive for monitoring customers in the business by using various applications in systematic approach. Our study is based on leading internet service providing company, Turksat Satellite Communications and Cable TV Operations Company's customer behavior analysis. It is the leading internet service provider of Turkey operating in telecommunications sector. We have created a two-phase solution utilizing data mining techniques. These are time series clustering and classification techniques.**

Keywords — Customer relationship management, churn prediction, data mining, time series clustering, k-means clustering, hierarchical clustering, classification, support vector machines, recursive partitioning

I. INTRODUCTION

The interest in the customer relationship management (CRM) has begun in 1990s[3]. However, although several definitions made for CRM, there is no widely accepted definition. Berry and Linoff [4] divides customers into 5 group (Figure 1).

Promotional activities for the existing customers are gaining importance. The rationale behind this is that existing customers are valuable and it is harder to regain a customer working with a competitor than a new customer acquisition in all business areas. This is also necessary because decline in volume growth of different agile sectors at business markets. Telecommunication markets especially concentrate on promotions to keep attraction of costumers sufficient. Hence, it is the keystone for preserving and even more improving their market shares.

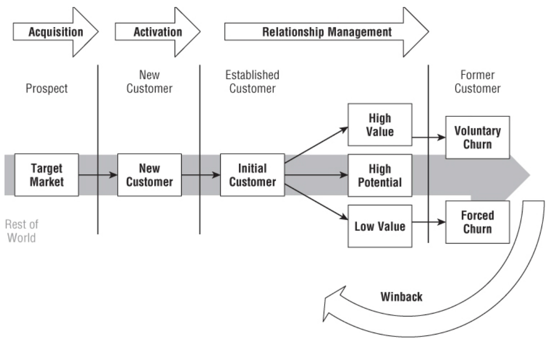

Figure 1. Customer Life Cycle Process (Source: [4])

Saturation of the market and competition between rivals makes a serious struggle for prestige and difficulty for gaining new customers. Even if cost criteria is considered, it can easily be figured out that less effort in promotional activities are less costly than that of gained customers in terms of sales, establishment and etc. There is an ongoing fierce competition environment telecommunications sector in addition to taking care of surviving. Churn prediction, which is one particular decision support system, is yet to be implemented for the betterment of customer satisfaction.

There are many CRM applications since retention of customers is a valuable act. It is more preferable to focus on demands of existing customers rather than potentially risky new customers. Satisfaction of customers will keep them as long term customers and they will tend to buy more; long term customers will be insensitive to competitors' offers in the long term. Their positive opinion will affect new customers to join. Especially negative opinions are disclosed in public. It must be prevented to have new customers[10].

Telecommunication market is highly dynamic and service providers offer appealing deals. The reason of why churn prediction is so substantial in the telecommunication sector can be explained as:

Monthly customer cancellation is 2.2% and that makes approximately 25% in a year. Cancellations cost around one quarter of the endorsement of a telecommunication

ASONAM '15, August 25-28, 2015, Paris, France

DOI: http://dx.doi.org/10.1145/2808797.2808821

company. On the other hand, promotional activities are five times cheaper than the cost of gaining new customers [2].

The customers, who should be subject to churn prediction, are former customers, new customers and win back customers given the situations of the customers in the telecommunications sector. Determining the new and win back customers, who have the same pattern with former ones by analyzing the behavior of former customers at once and revealing the reasons of the churn are considerable and eligible information for companies regardless of the sectors.

Customer churn detection by utilizing data mining techniques in telecommunications has been well studied in literature[5,11-23]. Subscription agreement and call details are used for churn prediction due to the lack of demographic data of the customers. When it is compared through experiments the decision tree model is more practicable and the time elapsed on training data is less than artificial neural network model.

Horizontal data of customer behavior, usually cannot be assessed with static data. Especially horizontal behavior data converted to the static features to improve the performance of the prediction process. [6]. In this study, hierarchical multi-core support vector machines algorithm has been offered as a solution rather than traditional SVM solution. Both horizontal data of customer behavior and static data are taken as input. Algorithm complements the learning process by realizing also feature selection process in three phases. We have compared different classification algorithms with various parameters and data sets.

Our study contains introduction, related work, methodology, experiments followed by conclusion and future work.

II. RELATED WORK

A. Time Series Clustering Applications

Time series clustering can be applied on various domains including multimedia, finance, health-care, science, economy, and government[7,8,24,25,26]. a variety of data types expressed as time series can be seen in the researches.

Time series matching problem is defining similarity or distance between given and time series in database. Resulting proximity and similarity measures determines relationship between two distinct time series and helps for clustering. As shown in Figure 2, a time series specified as given a time series, the problem, boils down to finding distances between that one and existing time series.

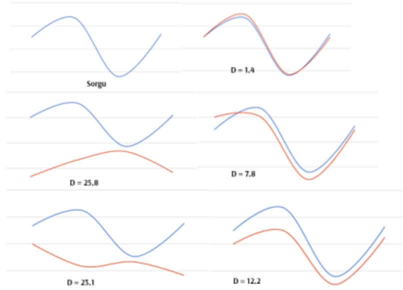

Figure 2. Time Series Matching Problem Example

III. METHODOLOGY

Our data structure and detailed explanation of two phased model and comparative results have been given in this section.

Data Structure

Our data structure regarding internet usage information has been given as below:

Unique Service Identifier: Unique number of customer service(Integer)

Gender : (Male, Female)
Age: Age of the owner of the service(Integer)
Internet Quota Group: Possible values are Quota, Fair Usage Quota, and Unlimited
Internet Quota Limit: Download quota for the customers who uses quota or fair usage quota (Unit: GB)
Speed of Internet Service: Internet service download speed (Unit: GB)
Undertaking: Contract information. Possible values are With Contract(True), Without Contract(False)
Total Delay of Fine: Total delay time on payments of invoices according their due dates.
Subscription Term Length:Duration of subscription status (Unit: Day)
Buying Trend Index: Numeric expenditure made by the subscriber in recent months. (6 months- Time Series)
Amount of Download: Numeric download quota used by the customer in recent days (last 2 months into 10-day periods- Time Series)
Failure Issue Count: Numeric Failure issues raised by the customer in recent days (last month into 10-day periods- Time Series)
Churn Class: Possible values are missed customer(True), Retained Customer(False).

Two Phased Solution Model

A solution model will be provided by making some improvements in addition to basic classification model within the scope of two phased model. The second phase of the stated two phase algorithm is classification.

Additional processes before the classification is referred to first phase.

It should be started by defining $behavior_{ijk}$. This variable holds the customers' behaviors in the form of time series. For instance, customer's total amount invoice is given a value in this variable with i^{th} customer j^{th} behavior type as buying trend index and k^{th} observation like k^{th} month in the data observation scope. On the other hand, a variable $feature_{ij}$ is set to define static features like age, sex, age of subscription and etc…

In this step, $tpBehavior_{ijk}$ is created by column binding on $behavior_{ijk}$ and $stBehavior_{ijk}$ to start first phase. Each customer's observation, which is realized in a certain behavior set, is hold by$behavior_{ijk}$. $stBehavior_{ijk}$ is represented by centering and scaling of the variable $behavior_{ijk}$. Furthermore these two variables are combined over behavior types and represented as $tpBehavior_{ijk}$.

$feature_{ij} \leftarrow For\ each\ customer\ i.static\ feature\ j.$
$behavior_{ijk}$
$\leftarrow For\ each\ customer\ i.behavior\ type\ j.\ observation\ k.of\ the\ behavior\ type$
$stBehavior_{ijk} \leftarrow rowCenterRescale(davranış_{ijk})$
$tpBehavior_{ijk} \leftarrow columnBind(stBehavior_{ijk}, behavior_{ijk})$
$folds \leftarrow delaminate(feature_{ij}, k)$
$measurements \leftarrow \{TP, TN, Precision, Recall, Accuracy, F - Measure\}$
$behaviorTypes \leftarrow \{Purchase\ Trend\ Index, Download, Fault\ Count \ldots\}$
$\boldsymbol{for\ each}\ j \in behaviorTypes\ \boldsymbol{do}$
  $dk_{ij} \leftarrow cluster(method,\ tpBehavior_{ijk})\ \forall i \in 1 \ldots i\ \wedge j \leftarrow j$
  $dk_{ij} \leftarrow convertBinaryMatrix(dk_{ij})$
  $feature_{ij} \leftarrow columnBind(feature_{ij}, dk_{ij})$
$\boldsymbol{end\ for}$
$\boldsymbol{for}\ i \leftarrow 1 \ldots k\ \boldsymbol{do}$
  $trainSet \leftarrow divideFeatures(feature_{ij}, folds[i])\ \forall j \leftarrow i$
  $testSet \leftarrow divideFeatures(nitelik_{ij}, katlar[-i])\ \forall j \leftarrow i$
  $classification.model \leftarrow train(method, trainSet)$
  $classification.pred \leftarrow predict(classification.model, testSet)$
  $performance_{ij} \leftarrow performanceCalculate(classification.pred)\ \forall i \in measurements\ \wedge \forall j \leftarrow i$
$\boldsymbol{end\ for}$
$performanceMean_i \leftarrow \sum_{j=1}^{k} performance_{ij} \big/ k\ \ \forall i \in measurements$

Table 1 Two Phased Solution Model Algorithm

At the first step, we have clustered and obtained the behavior clusters from time series in the data set. For each separate behavior, we have produced meaningful features and then use these features which are composed of cluster summaries with $feature_{ij}$.

The first phase runs for each behavior set (time series). This phase's added value functionality is the *cluster* method. Each customer's observations in the system at each behavior set are assigned to a cluster. There may be intersection between two clusters as cluster membership has been represented in binary values. It is assumed that similar behavior customers have similar patterns in the customer life cycle. Therefore, this behavior data processing is meaningful.

As a result of clustering process, dk_{ij} variable consists of integer cluster number for each customer's specific behavior. As seen in performed experiments and making data meaningful process, when quantifying categorical data without a superiority on each other cause a problem if the data is represented as integers like 1,2,3.. etc. Instead of representing as mentioned, clusters should be converted to a binary matrix by the help of *convertBinaryMatrix* method. Instead of using a feature, there will be a feature set with a size of the number of the cluster size for each behavior in a behavior set.

Thus, behavioral clustering results can be divided into train set and test set for each customer separately. In other words, clustering results on the dk_{ij} variable is bound to the $feature_{ij}$ variable as a new feature by the *columnBind* method.

At the second phase is classification algorithms process on enriched and interpreted raw data. In brief, classification phase predicts each row whether the customer will churn or not. It uses below algorithms to compare according to the performance measurement criteria.

- Support Vector Machines (SVM)
- Recursive PARTitioning (RPART)

Furthermore, this process uses k-fold cross validation. Common usage of k-fold cross validation is defining k as 10. Sample data is divided into ten randomly selected groups. And then, 9 group train set and 1 group test set are evaluated in the SVM and the RPART algorithms. In order of groups the selected test set group is floating at each trial until k time is reached. Also, the second loop can be seen at the Table 1. It represents the second classifying phase of the model.

The performance of the whole two phased algorithm is measured at the end of the algorithm by taking means of the performance index at each trial of the second phase.

IV. EXPERIMENTS

A. Data Preprocessing

Each row represents a customer including his static and behavior data. Around 6000 customers are selected as sample data to give a demonstration of the survey. And around 70000 raw data was gathered from live environment to run on the real data where is chosen from a pilot district in Turkey. To improve results the data is queried from the active customers and the passive customers from one year before to the day of the survey.

The sample data have similar statistical attributes with the real data as the percentage of the churn class.

The raw data was obtained from the PostgreSQL database which is used by the corporation by using PL/pgSQL. Because such a large data seems not to be very possible with a single query through a live system, the logging data was provided by using JAVA with the paging technique in SQL statements. All of the data analysis is performed with R.

B. Parameters

One of the important parameter to decide is cluster size at first phase in the two phased solution model. It is decided by trial and error method by using various cluster sizes.

Table 2 Classification Results After Clustering with different number of clusters values

	Cluster Size							
Performance Measure	**3**	**5**	**8**	**10**	**20**	**30**	**40**	**50**
Accuracy	0,90	0,92	0,92	0,92	0,92	0,93	0,92	0,92
True Positive	0,64	0,73	0,69	0,72	0,73	0,74	0,73	0,72
True Negative	0,95	0,96	0,97	0,96	0,96	0,96	0,96	0,96
Precision	0,70	0,78	0,81	0,80	0,80	0,80	0,79	0,79
Recall	0,64	0,73	0,69	0,72	0,73	0,74	0,73	0,72
F-Measure	0,67	0,75	0,75	0,75	0,76	0,77	0,76	0,75

The performed tests with the sample data through the application with various cluster sizes shown in Table 2. The output belongs to the k-means for the first phase and the recursive partitioning for the second phase in the two phased solution model. Barely these results are valid for support vector machines and hierarchical clustering algorithms according to the performed parameter tests. Chosen cluster size affects the total result of the two phased model.

After utilizing hierarchical clustering at the first phase, the performance of the second phase has been given in Figure 3. When the measurement parameters of two separate clustering algorithms experimented in the frame of two phased solution model is compared that k-means algorithm is 2 percent more accurate then hierarchical clustering according to F-measurement values.

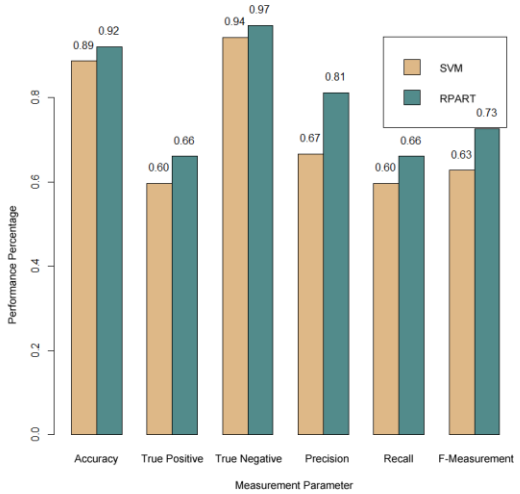

Figure 3. Two Phased Model Results with Hierarchical Clustering

C. Distance to The Cluster Centers Experiment

In this experiment, instead of individual clusters, there will be cluster center points for each behavior. And the binary matrix features will be converted to distance measures by measuring the distance between each cluster center to the customer behavior. This remains a relation degree as the distance measure for a customer to the each cluster center.

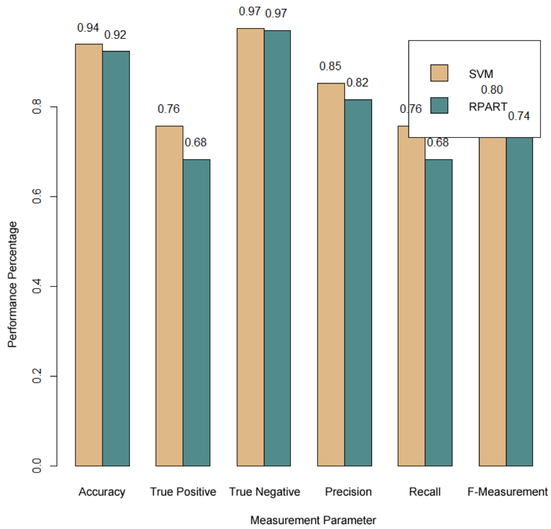

Figure 4. Distance to the Cluster Center Experiment Results

Experimental results point out that RPART works better than SVM. Representing time series features in terms of binary values would make difference for different classification algorithms as they work on different input types. In decision tree algorithms, categorical attributes can help improve our results whereas, support vector machines have ended up with worse results since the attributes contained features with mostly categorical type.

When we consider distance to cluster values instead of discrete boolean values, SVM works better than RPART as in Figure 4. Difference between F-measure values between SVM and RPART is 3% in favor of SVM.

V. CONCLUSIONS

It has become critical to monitor behavioral patterns of customers for the churn prediction for telecommunications companies for upholding their shares at the market. We have enriched the existing data by preprocessing time series data after time series data has been exposed to clustering for enrichment. Churn detection has been extended with unsupervised extra features to retain existing customers to give them a better quality of service in addition to having new customers. Our solution on real data has demonstrated its effectiveness in predicting their future acts.

The method we proposed extends dataset after further pre-processing time series data. At the end, we have two different types of datasets. The first part contains standard data and the second part contains cluster membership values for each time series. Each group of time series is the signature of the subscriber.

VI. FUTURE WORK

We plan to focus on finding the number of clusters value leading to the best classification model for testing. Our results depend on empirical results but it would be a better solution to apply different number of clusters value for each time series group for the automation of churn prediction.

REFERENCES

[1] Chih-Ping, W. and I-Tang, C., Turning telecommunications call details to churn prediction: a data mining approach, Expert Systems with Applications, 23, 103-112, 2002.

[2]Kamalraj, N. and Malathi, A., A Survey on Churn Prediction Techniques in Communication Sector, International Journal of Computer Applications, 2013.

[3]Ling, R. and Yen, D., Customer Relationship Management: An Analysis Framework and Implementation Strategies, Journal Of Computer Information Systems, 2001.

[4]Linoff, G.S. and Berry, M.J.A., Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Wiley Publishing, Indianapolis, 2011

[5]Wei, C. P., and Chiu, I. T., Turning telecommunications call details to churn prediction: A data mining approach, Expert Systems with Applications, 23 (2), 103–112, 2002.

[6]Swift, R.S., Accelerating Customer Relationships: Using CRM and Relationship Technologies, Prentice Hall Professional, New Jersey, 2001.

[7]Jessica, L., Michail, V., Eamonn, K. and Dimitrios G., Multi-resolution time series clustering and application to images, Springer, Londra, 2007.

[8]Rakthanmanon, T., and E., Keogh, Fast shapelets: A scalable algorithm for discovering time series shapelets, Proceedings of the 13th SIAM International Conference on Data Mining, Texas, USA, 2013.

[9]Kamalraj, N., and Malathi, A,. A Survey on Churn Prediction Techniques in Communication Sector. International Journal of Computer Applications 64(5):39-42, February 2013.

[10] Poel, D. V. den and Larivi`ere, B. Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research, 157(1):196–217, 2004.

[11] Au, W., Chan, C., Yao, X., A novel evolutionary data mining algorithm with applications to churn prediction IEEE Transactions on Evolutionary Computation, 7 (2003), pp. 532–545

[12] Coussement, K., Poe, D.V. den, Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques, Expert Systems with Applications, 34 (2008), pp. 313–327

[13] Hung, S.-Y. , Yen, D.C., Wang, H.-Y., Applying data mining to telecom churn management, Expert Systems with Applications, 31, pp. 515–524, 2006.

[14] John, H., Ashutosh, T., Rajkumar, R., Dymitr, R., Computer assisted customer churn management: State-of-the-art and future trends. 2007.

[15] Luo, B., Shao, P., Liu, J. Customer churn prediction based on the decision tree in personal handyphone system service. In International conference on service systems and service management (pp. 1–5), 2007.

[16] Wei, C., and Chiu, I., Turning telecommunications call details to churn prediction: A data mining approach, Expert Systems with Applications, 23, pp. 103–112, 2002.

[17] Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H., (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. Neural Networks, IEEE Transactions on, 11(3), 690-696, 2000.

[18] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B., New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research, 218(1), 211-229, 2012.

[19] Amin, A., Shehzad, S., Khan, C., Ali, I., & Anwar, S.,Churn prediction in telecommunication industry using rough set approach. In New Trends in Computational Collective Intelligence (pp. 83-95). Springer International Publishing, 2015.

[20] Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U., Improved churn prediction in telecommunication industry using data mining techniques. Applied Soft Computing, 24, 994-1012, 2014.

[21] Lu, N., Lin, H., Lu, J., & Zhang, G., A customer churn prediction model in telecom industry using boosting. Industrial Informatics, IEEE Transactions on, 10(2), 1659-1665, 2014.

[22] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., and Chatzisavvas, K. C., A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1-9, 2015.

[23] Bandara, W. M. C., Perera, A. S., and Alahakoon, D., Churn prediction methodologies in the telecommunications sector: A survey, 2014.

[24] Doshi-Velez, F., Ge, Y., & Kohane, I., Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. Pediatrics, 133(1), e54-e63, 2014.

[25] Hebert, D., Anderson, B., Olinsky, A., & Hardin, J. M., Time Series Data Mining: A Retail Application. International Journal of Business Analytics (IJBAN), 1(4), 51-68, 2014.

[26] Choudhury, S., Ghosh, S., Bhattacharya, A., Fernandes, K. J., & Tiwari, M. K., A real time clustering and SVM based price-volatility prediction for optimal trading strategy, Neurocomputing, 131, 419-426., 2014.