Churn Analysis: Predicting Churners

Navid Forhad, Md. Shahriar Hossain, Rashedur M Rahman[1], M A Matin[2]
[1] North South UniversityPlot-15, Block-B, Bashundhara, Dhaka 1229, Bangladesh
[2] Institut Teknologi Brunei, Brunei Darussalam
{nforhad, matin.mnt@gmail.com}, shahriarshayan@yahoo.com, rashedur@northsouth.edu

ABSTRACT: *Churners have always been a big issue for any service providing company. Churning increases cost of the company as well as decreases the rate of profit. Generally, customer attrition can be identified when they initiate the process of service termination. At the same time, the individuals and the institutions that provide the data residing on the government databasesas well as the agencies who sponsor the collection of such information- are becoming increasingly aware that extend analytical capabilities also furnish tools that threaten the confidentiality of data records. However, using predictive analysis using customers past service usage, service performance, spending and other behavior patterns, the likelihood of whether a customer wants to terminate service can be determined. In this paper, the authors address the issue of churn analysis considering a scenario in which a company owning confidential databases wish to run a churn analysis technique on the union of their databases, without revealing any unnecessary information. The aim of the paper is to predict whether a customer will churn in the near future or not based on the predictive analysis using billing data of a telecom company.*

Keywords: Churn Analysis, Customer Attritions Analysis, Client Defection Analysis, Churn Prediction

Received: 20 May 2014, Revised 24 June 2014, Accepted 30 June 2014

1.Introduction

1.1 Definition

Data mining[6] is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patters discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities.

Data mining is a field, connecting the three worlds of Databases, Artificial Intelligence and Statistics. The information age has enabled many organizations to gather large volumes of data. However, the usefulness of this data is negligible if "*meaningful information*" or "*knowledge*" cannot be extracted from it. Data mining, otherwise known as knowledge discovery, attempts to answer this need. In contrast to standard statistical methods, data mining techniques search for interesting information without demanding a priori hypotheses. As a field, it has introduced new concepts and algorithms such as association rule learning. It has also applied known machine-learning algorithms such as inductive-rule learning (e.g., by decision trees) to the setting where very large databases are involved. Data mining techniques are used in business and research and are becoming more and more popular with time.

Data mining is the process of extracting patterns from data. As more data are gathered, with the amount of data doubling

every three years, data mining is becoming an increasingly important tool to transform these data into information. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery.

Data mining is used for a variety of purposes in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. For example:

• The insurance and banking industries use data mining applications to detect fraud and assist in risk assessment (e.g., credit scoring).

• Using customer data collected over several years, companies can develop models that predict whether a customer is a good credit risk, or whether an accident claim may be fraudulent and should be investigated more closely.

• The medical community sometimes uses data mining to help predict the effectiveness of a procedure or medicine. Pharmaceutical firms use data mining of chemical compounds and genetic material to help guide research on new treatments for diseases.

• Retailers can use information collected through affinity programs (e.g., shoppers' club cards, frequent flyer points, contests) to assess the effectiveness of product selection and placement decisions, coupon offers, and which products are often purchased together.

• Companies such as telephone service providers and music clubs can use data mining to create a "*churn analysis,*" [1,4] to assess which customers are likely to remain as subscribers and which ones are likely to switch to a competitor.

One distinct part of data mining is Churn analysis. It is the calculation of the rate of attrition in the customer base of any company. It involves identifying those consumers who are most likely to discontinue using a service or product.

Churn analysis is extremely helpful in developing a sustainable and robust strategy for customer retention in a company.

When a company is aware of the percentage of customers who end their relationship with them in a given time period they can easily come up with a detailed analysis of the causes for the churn rate using churn analysis. This helps in developing effective customer retention programs for the company.

Churn rate typically applies to many industries chiefly among them are subscription services, such as longdistance phone service or magazines. Churn analysis [2] helps in understanding the behavior of customers that unsubscribe and move their business to a competitor and predicting the likelihood of this event to occur. Other uses vary from calculating employee attrition in any given company.

1.2 Motivation

Our work is motivated by the need to both protect privileged information and enable its use for commercial or other purposes.

2. Related Work

We have researched to find out how much work is done in the field of churn analysis. In our search we found some works that discuss churn analysis in details. We are describing a few of the works here.

In [8] the authors explain predictive modeling for churners based on data mining methods. The paper also discusses how to use decision tree analysis model in detail. The paper mainly discussed customer churning from a business perspective. However at the end of the paper they also discussed case studies along with process flows and modeling techniques.

In [9] Teemu Mutanen described a case study on customer attrition. The paper described in details the methods used for the prediction, data used and also the result that was achieved. The author described two methods for churn analysis. The first one is logistic regression. Logistic regression is used to predict a discrete outcome on the basis of continuous and/or categorical variables. In this method only one dependent variable can exist. This method applies maximum likelihood estimation after transforming the dependent variable into a logistic variable.

The second method analyzes the estimation results of the logistic regression. It is known as the lift curve. This curve is

related to ROC curve of signal detection theory and precision-recall curve. The lift is a measure of predictive model calculated as the ratio between the results obtained with and without the predictive model. In [10] Shyam V. Nath describes a case study in which an Oracle based database of fifty thousand customers of wireless telecommunication industry was analyzed to predict churners. The study used JDevelpoer tools and the analysis was done using Naïve Bayes algorithm with supervised learning.

Marco Richeldi and Alessandro Perucci [11] wrote a paper on case study of churn analysis. This paper discusses the use of Mining Mart, an churn analysis tool. It mainly discusses the preprocessing of data to analyze with Mining Mart.

3. Methodology

3.1 Input Data

Churn analysis is basically done with a lot of historical data. This data is accessible from the data warehouse of respective company. For a complete analysis of customer attrition of a telecom company, the data that are needed are

- Customer Demographics, i.e., age, gender, marital status, location, etc.
- Call statistics: length of calls at different times of the day, number of long distance and local calls.
- Billing information of each customer – what the customer is paying for local and long distance.
- Extra service information, that is, what extra plan the customer is registered on, e.g. special long distances rates.
- Voice and data product and services purchased by the customer, e.g., broad-band services, private virtual network, dedicated data transport links, etc.
- Complaining information: how many customer service calls are made for disputed billing, dropped calls, slow service provisioning, un working special services, and so on.
- Credit history.

From all these data, relevant data are gathered and four tables are created.

The first table, we call it Customer table contains customer information, a key that identifies customers in the source systems and other basic information, e.g., age and gender.

The second table contains call data records (CDRs) in transitional form. Each record provides the duration of a specific type of calls, i.e., peek, weekend, international calls aggregated on a monthly basis.

The third table, known as service table contains basic information concerning service subscripted by customers, such as: handset model or class, length of service, number of dropped calls, tariff plan and tariff type used for billing, and target attribute, that is, the churn state in last month.

The last table contains basic billing data in transactional form. Each record provides the revenue related to each user aggregated on a monthly basis.

3.2 Our Data

The data we gathered are of a renowned telecom company. However, the data only contains billing record. To do some preprocessing we loaded the data in MySql. We used sql queries to find out more information about the data. The attribute set that we could use from our data set are phone number, bill payment date, payment location and payment amount. Let us name them ph, date, loc and amount accordingly.

By executing queries on the data set, we found out many aspects of the data. We named the table bill. Below is the list of aspects that we found along with the queries which were used to find the respective aspect.

- The data set has 6938 records. **Query:** select count(1) from bill
- The data set contains record of 26 months. **Query:** select count(distinct substring(date,1,7)) from bill
- It has 880 phone numbers. **Query:** select count(distinct phone) from bill

• Bills were paid from 25 locations. **Query:** select count(distinct loc) from bill

3.3 Methods of Churn Analysis

There are many ways to do churn analysis. However, these methods can be primarily categorized in two sections. They are

• Supervised methods

• Unsupervised methods

3.3.1 Supervised Methods

In supervised methods [7], the method basically learns to classify the data based on what it learns from given training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object and a desired output value. In our case the input object would be the rows of our data and the output value would be a binomial value depicting whether the customer has churned or not. However, our dataset does not have any attribute that will be saying who will churn. Therefore we cannot use supervised approach with our dataset.

3.3.2 Unsupervised Methods

On the other hand, unsupervised methods refer to the problem by trying to find hidden structure in unlabeled data. We can use unsupervised methods to cluster our data set. This way the churners could become part of a separate cluster. Again, we can use methods which self-learn the data.

3.4 Rule-based Classifier

One method of unsupervised learning is the rule-based classifier. The method learns the data based on preset rules. This means that the method self-learns. The rules can be generated by extracting features from our dataset. Some of facts that could be used to create rules are

• If a customer doesn't pay his bills on time he could become a churner.

• If the bill amount decreases rapidly, the customer could churn.

4. Result & Analysis

4.1 Data Filtering& Rule Generation

A closure look at our dataset reveals some cases which could be considered as impurity or noise. To get a better analytic result, we have to remove this impure records from our data set.

One such case are the bill record for phone number that occurred only once in the total dataset. We could find them with the following query –

select count (1), phone from bill group by phone having count (1) = 1

The total number of records who only occurred once in the dataset can be found with the following query –

select count(1) from bill where phone in (select phone from bill group by phone having count (1) = 1)

This gives us total 303 records of phone number which only occurred once in the dataset. So among 880 phone numbers 303 can be considered as noise. Therefore the dataset contains 577 valid records.

We can create a new attribute from the dataset. That is the frequency of payment.

$$frequency = \frac{no.\ of\ months\ for\ which\ bill\ is\ paid}{total\ no.\ of\ months} \quad (1)$$

As we have data of 26 months and bill is paid once per month so frequency of payment would be one for the customer who paid bills in all 26 months. Lower frequency suggests that the customer is not paying his bills regularly. So the lower the fAs we have data of 26 months and bill is paid once per month so frequency of payment would be one for the customer who paid bills in all 26 months. Lower frequency suggests that the customer is not paying his bills regularly. So the lower the frequency the higher the chances for the given customer might churn. We created a new table named frequency with the frequency number for every phone number. Query is given below –

create table frequency select (count (1) / 26) as '*frequency*', phone from bill group by phone

We can find the maximum and minimum frequency for the customers using the following query –

select max (*frequency*), *min* (*frequency*) from *frequency*

From this query we get a maximum frequency value 1.1923 and minimum value 0.0385, a higher frequency rate means some customer has paid bills more than once in one month. This is plausible as a customer can pay his/her total dues in more than one installment.

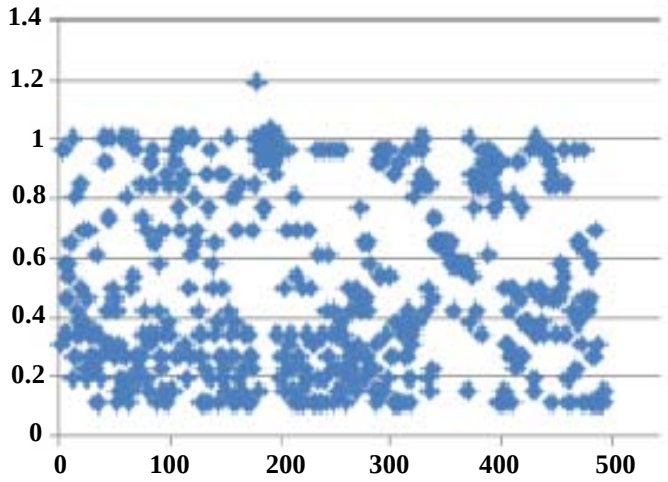

Figure 1. Scatter chart of frequency. [*x-axis*: frequency, *y-axis*: record no.]

Figure 1 shows scatter chart of the frequency of each customer whose frequency is higher or equal to .10. From figure 1 we can see that only one customer has frequency higher than 1. We can see a distinctive gap around the line 0.6. This is a good point to draw our border line. So any customer having frequency less than 0.6 is a churner. So we get the following rule –

$r1 : (\text{frequency} < 0.6) \rightarrow \text{Churner}$

The decision tree for this rule would be –

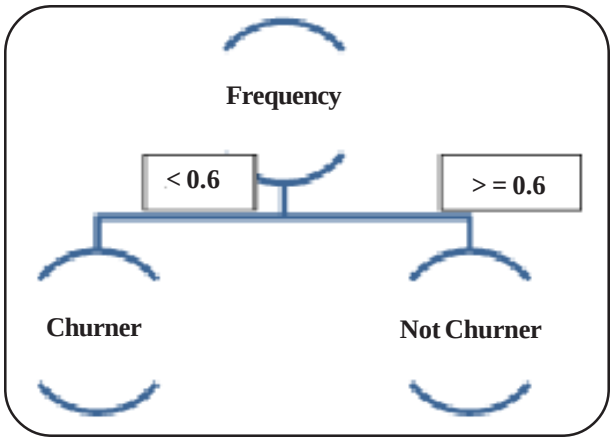

Figure 2. Decision tree for frequency rule

This rule can be checked by individually plotting amount against date for each customer in a line curve. If the line inclines, then the customer has less probability to hurn. Whereas declining line means that the customer might churn as it is a sign of customer slowly stopping using the service.

To get a better visualization, we have chosen two customers and plotted their payment here. One customer has frequency of 0.9651 and the other one has 0.5769.

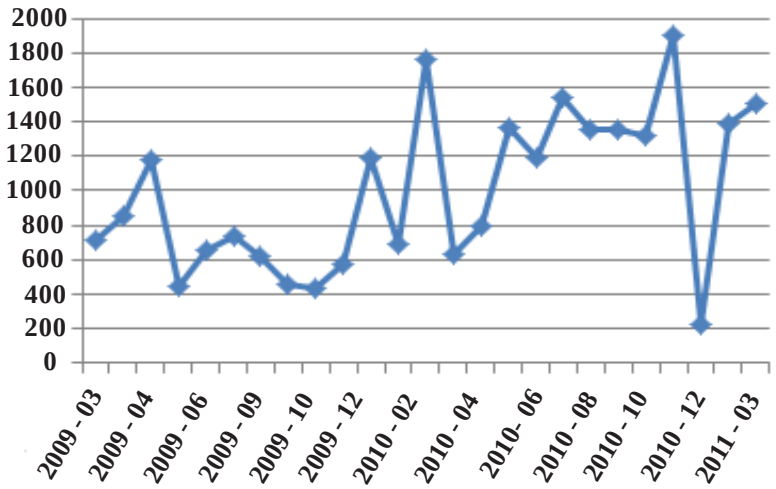

Figure 3. Line graph of a single customer with frequency 0.9651

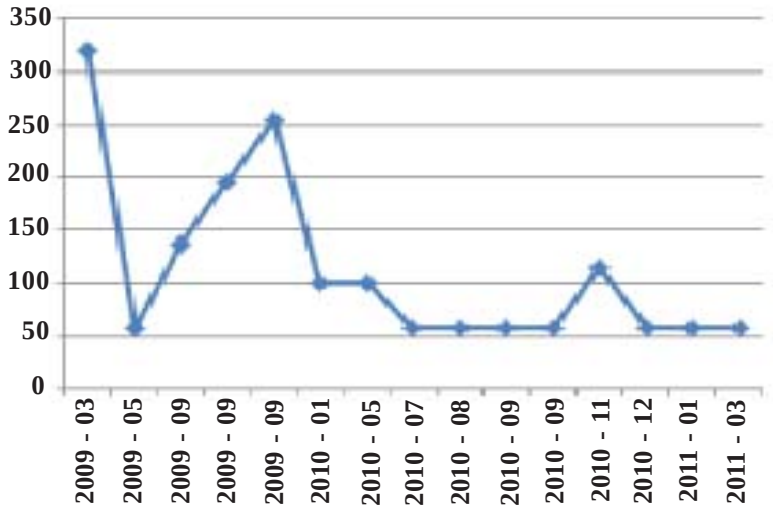

Figure 4. Line graph of a single customer with frequency 0.5769

By comparing the two graphs we can easily see that the line of the customer with higher frequency has inclined as time goes which means that the customer is using the service more over time. On the other hand, the line graph of the customer with lower frequency shows a declining line depicting lesser use of service over time. So our rule will correctly identify the first customer as not churner and the second one as churner.

4.2 Results

Based on the rule we can easily find out the churners. Below a table shows the probable number of churners based on this rule.

Churners (frequency < 0.6)	Not Churners (frequency > = 06)	Not Classified (frequency < 0.1)
315	187	387

Figure 5. Table showing result

We tried the RapidMiner data mining tool to cluster our dataset. We fed both the bill and frequency table to the tool. However the tool failed to cluster them properly. We were looking for two clusters, each representing one class. The classes are churners and not churners. However, in both cases the tool clustered the dataset to more than two segments.

5. Obstacles

5.1 Incomplete Data

The data that we worked with was incomplete. Therefore the result we achieved is not accurate. However, it gives us a picture of how churn analysis works.

5.2 Time

Churn analysis is a lengthy and complex process. Therefore the time frame was not enough to do a complete churn analysis and obtain a viable result.

5.3 Confidentiality Issues

A key problem that arises in any en masse collection of data is that of confidentiality. The need for privacy is sometimes due to law (e.g., for medical databases) or can be motivated by business interests. However, there are situations where the sharing of data can lead to mutual gain. A key utility of large databases today is research, whether it is scientific or economic and market oriented. Thus, for example, the medical field has much to gain by pooling data for research; as can even competing businesses with mutual interests. Despite the potential gain, this is often not possible due to the confidentiality issues which arise.

6. Future Work

The result that we got is not promising. This is due to the fact that we had a dataset that was incomplete. It is our failure that we could not manage a proper dataset to do churn analysis. In future there are many things that we want to do for this project. Some of them are listed below:

- Use a complete dataset to implement churn analysis method.
- Properly use data mining tool to predict the churning.
- Implement multiple methods to analyze churn.
- Compare different methods to find the optimal one.
- Use and compare multiple data mining tool.

What we will try to find out the effectiveness of a noise addition program which will mask the data to preserve privacy. The factors related to the choice of the privacy preserving algorithm are: characteristics of a good masking technique, disclosure risk, and, minimum disclosure risk. First we will finish the task of record generation. Following this we will do graph generation, pattern analysis, and, comparison of masked data and original data by turn.

7. Tools Used

7.1 MySQL Database: We used MySQL database to store the dataset. Basically, the bill and frequency tables were stored in the database.

7.2 HeidiSQL IDE: This is basically a MySQL front end. We used it to connect to our database and execute query. This tool also helped us to export data in csv format which is required by data mining tools.

8.Conclusion

In this paper we tried to present churn analysis in a telecom sector. The analysis focused on churn prediction is based on rule based classification. However the incompleteness of the dataset made it hard to create a predictive model. The findings indicate that we have to use a complete and big dataset if we want to achieve any kind of accuracy.

Churn analysis in data mining activities is a very important issue in many applications. Different techniques are likely to play an important role in this domain. However, this paper illustrates some of the challenges that these techniques face in churn analysis. It showed that under certain conditions it is relatively easy to breach the privacy protection offered by the different

techniques. It provided extensive experimental results with different types of data and showed that this is really a concern that we must address. In addition to raising this concern the paper offers a model churn analysis technique that may find wider application in developing a new perspective toward developing better churn analysis techniques. It is interesting for a company's perspective whether the churning customers are worth retaining or not. And also in marketing perspective what can be done to retain them. How long a time period for the data should be is also a matter of interest.

References

[1] Customer attrition. Retrieved from http://en.wikipedia.org/wiki/Customer_attrition, on, February 25, 2011.

[2] Predictive analytics, Retrieved from http://en.wikipedia.org/wiki/Predictive_analytics on April 14, 2011.

[3] Predictive modeling, Retrieved from http://en.wikipedia.org/wiki/Predictive_modelling, on March 17, 2011.

[4] Mattison, Rob. The Telco Churn Management Handbook, 2001.

[5] Churn Analysis. (n.d.). Retrieved from http://www.ambarasoft.com/researchservices/churnanalysis. html

[6] Witten, Ian H., Frank, Eibe. (2005). Data Mining Practical Learning Tools and Techniques, Morgan Kaufmann Publishers.

[7] Supervised learning Retrieved from http://en.wikipedia.org/wiki/Supervised_learning, retrieved on April 16, 2011.

[8] Khalida binti Oseman., Sunarti binti Mohd Shukor., Norazrina Abu Haris., Faizin bin Abu Bakar. (2010). Data Mining in Churn Analysis Model for Telecommunication Industry *Journal of Statistical Modeling and Analytics*, 1 (19-27).

[9] Teemu Mutanen., Customer churn analysis – a case study, Technical Report, Retrieved from, http://www.vtt.fi/inf/julkaisut/muut/2006/customer_churn_ case_study.pdf, April 12, 2014.

[10] Shyam V. Nath., Customer Churn Analysis in the Wireless Industry: (2014). A Data Mining Approach, Technical Report, retrieved from http://download.oracle.com/owsf_2003/40332.pdf, April s14.

[11] Marco Richeldi., Alessandro Perrucci. (2014). Churn Analysis Case Study, Technical Report, Telecom Italia Lab, Italy, retrieved from http://sfb876.tudortmund. de/PublicPublicationFiles/richeldi_perrucci_200 2b.pdf, April 12.