Improving customer churn models as one of customer relationship management business solutions for the telecommunication industry

Slavescu, Ecaterina and Panait, Iulian

Academia de Studii Economice (Academy of Economic Studies), Hyperion University

2012

Improving Customer Churn Models as one of Customer Relationship Management Business Solutions for the Telecommunication Industry

Ecaterina Oana Slăvescu
PhD candidate, The Bucharest Academy of Economic Studies
ecaterinaslavescu@yahoo.com

Iulian Panait
PhD candidate, The Bucharest Academy of Economic Studies
iulian.panait@gmail.com

Abstract

Nowadays, when companies are dealing with severe global competition, they are making serious investments in Customer Relationship Management (CRM) strategies. One of the cornerstones in CRM is customer churn prediction, the practice of determining a mathematical relation between customer characteristics and the likelihood to end the business contract with the company. This paper focuses on how to better support marketing decision makers in identifying risky customers in telecom industry by using Predictive Models.

Based on historical data regarding the customer base for a telecom company, we proposed a Predictive Model using Logistic Regression technique and evaluate its efficiency as compared to the random selection. In the future, we will focus on extending our study by integrating more business considerations and mining models in order to adjust the churn models or redesign marketing activities for the telecom industry.

Key words: predictive models, data mining, churn, time series econometrics
J.E.L. classification: C25, C26, C32, C53, C81

1. Introduction

Competition in the telecommunications industry is very aggressive, since on each national market there are at least four carriers struggling to maintain their market share by offering very similar and very attractive products and services. This is the main reason why, in order to maintain their profitability, telecommunications operators are very concerned in controlling *churn* – generally defined as the loss of subscribers who switch from one carrier to another (Gopal, Meher, 2008) – and, based on predictive data mining techniques, developing new efficient CRM strategies to retain customers .

Keeping in mind the fierce competition on the market, the focus of telecommunication companies has therefore shifted from building a large customer base into maintaining and strengthening the previous developed base, in witch case it is valuable to identify in advance those customers who are likely to switch to a competitor in the near future, the so called churned customers.

Statistics have showed that is 5 to 6 times more profitable to keep existing customers than to attract new ones (LeBoeuf, 2000). This is why, churn prevention can be regarded as a popular way of reducing the company's costs, while increasing the profitability, which is the basic key for survival on the market, especially for listed companies.

This paper is organized as follows: in Section 2 we will present the concept of CRM and its close interference with churn, especially in telecom, while making a literature review on this subject, Section 3 will present the methodology insisting on logistic regression technique and its application in building a predictive churn model, Section 4 presents the predictive model build based on a data base from a Romanian telecommunication company, and, finally, Section 5 provides the conclusions.

2. Customer Relationship Management and telecom churn management

Since is an established fact that the building of long term relationships with customers provides multiple benefits to businesses (Dwyer et al., 1987) and since it has been acknowledged that old customers are more profitable in the long term, because new customers are attracted by promotional offers and tend to switch to another provider at the moment they receive a better discount (Gopal, Meher, 2008), and having to deal with this present necessity of better "Customer Relationship Management","(CRM) is a healthy and promising newcomer which has appeared on the business radar"(Greenberg, 2002).

Despite its large recognition as an important business strategy and the fact that the term gives the definition by itself: the management of relationship with customers, there is no widely accepted definition of CRM.

According to some authors (Lin, Xu 2009), CRM comprises a set of processes and enabling systems supporting a business strategy to build long term, profitable relationships with specific customers.
Kincaid (2003) defines CRM as a enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty and customer profitability.
Parvatiyar (2001) defines CRM as the strategic use of information, processes, technology, and people to manage the customer relationship with the company across the whole customer life cycle.
The same idea is considered by Swift 2001; Kim et al., (2003) – in their oppinion CRM consists of four dimensions: Customer Identification, Customer Attraction, Customer Retention and Customer Development.

The above given definitions have shown that in some way, CRM strategies are nowadays more focused on retention, and CRM is more and more becoming a process of retaining customers, with the help of business intelligence and data mining, to maximize the customer value to the organization.

So, in the business world (and we are referring especially to the telecommunication sector), CRM consists of all the activities that can transform a regular customer into a loyal one by exceeding its requirements.

If referring to the application of CRM strategies into the telecommunication sector, in order to upgrade regular loyalty and retention campaigns, we must specify that a current particularity of the telecommunications market is modelling the attrition risk (or the churn risk, where *churner* is a common term used both in academia and practice to denote the customers with propensity to leave for competing companies). According to some marketing studies (Berson et al., 2000), the average churn rate for mobile operators is approximately 2% per month, which means a loss of 25% of its customer base within a year.

For a customer, cancellation costs are relatively small, given the fierce competition in the market which consists into providing similar services (in terms of price and quality) and offering discounts on handsets and various other accessories.
Nowadays, price is no longer a differentiator, since many telecom providers tend to equal each other's prices. The main differentiator have become the additionally services.
In the present the customer is "serviced", not only "sold" (Cool, 1996).

Therefore, for almost all mobile phone companies, the biggest challenge is, at the present, switching from reactive to proactive retention, in other words retaining customers before they take the decision to end their contract and identifying those customers with a high churn risk (Slăvescu, 2011).

In order to maintain market share and profitability, telecom companies have used various approaches or management/marketing strategies to retain customers and prevent serious churn problems.

At present, most of the strategies are based on data warehousing - maintaining and managing historical and very detailed data obtained from many computing and point-of-contact devices – and data mining techniques that enriches and transforms data into meaningful information for business decision makers.

In this spirit, a common churn management process involves constructing a churn prediction model using past churn data, and determining key variables influencing churn. The churn model is then used to identify and classify a list of customers with potentially high risk of churn (potential churners) from existing customer data, to perform the appropriate retention activities (Berson et al., 2000; Au et al., 2003; Chu et al., 2007; Coussement et al., 2008; Ngai et al., 2009).

The efficiency of a company's combined CRM and churn management strategy is measured in terms of effectively decrease of the churn rate (or increase of the retention rate), and not only by whether it can locate a list of potential churners. In our study we will concentrate in building a model in order to identify customers with a high churn risk, the development of a model that can generate the increase of the retention rate, being on our list of future research interest.

3. The methodology: data mining techniques and logistic regression

Data mining is an integrated technique that involves analysis, filtering, extraction, and statistical analysis for large amounts of data; it has numerous applications in addressing business problems (Chu et al., 2007; Coussement et al., 2008; Lariviere et al., 2004; Lejeune, 2001; Luo et al., 2007; Ngai et al., 2009). Data mining can be used for analyzing the production process, investigating the marketing efficiency, studying the management effectiveness, or measuring the evolution of the company's stock exchange value (Panait, 2011).

Previous studies utilized various data mining technologies to assist telecom companies in resolving churn problems (Kim et al., 2004; Hung et al. 2006; Xia et al., 2008; Tsai et al., 2010).

Generally, churn is estimated using predictive models that are built based on a significant sample from the database, containing both churners and non-churners at a selected moment (i.e April 2010) and their historical data (at least last 3 months of information regarding their financial situation, their call detail records, their personal details- age, gender, address, etc).

Then, by using an appropriate software predictive tool (we have used Chordiant, but we can also use SPSS, UNIQA, etc.) the model is created by automatically applying a set of algorithms over the selected data set.

There are several categories of prediction models in statistics and data mining that can be applied to churn prediction problems. Past research shows that Decision Tree, Neural Network, Logistic Regression are well suited to this type of problem. For our model we have decided to use Logistic regression.

Logistic regression is a technique in which unknown values of a discrete variable are predicted based on known values of one or more continuous and/or discrete variables. A logistic regression model is distinguished from a linear regression model because the outcome variable in the logistic regression is binary or dichotomous.

If we consider a collection of n independent variables which will be denoted by the vector: $x_n = (x_1, x_2, x_3, \ldots, x_n)$, and the conditional probability that dependent variable is present be denoted be $P(Y=1|x) = y(x)$.

Then the multiple logistic regression model is given by the equation: $f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$, in which case the dependent variable is given by: $y(x)=e^{f(x)}/(1+ e^{f(x)})$.

In our case, the dependent variable is the churn prediction, and it can take the values: "churn" or "not churn" and the independent variables are the fields contained in our database that we have considered relevant for modeling churn. In this study, x_n =historical attribute variables (i.e. age, tenure, etc.), $y(x)$ = the probability of churn.

The model performance is measured by the hit and the lift rates. The hit rate is the ratio of the number of correctly classified targets to the number of classified targets. In other words, this is the ratio of true positives to the total number of hits identified.

Hit ratio (H%) = $A/(A+B)$, where A is the number of subscribers predicted to churn in the predictive time window who actually churned, B is the number of subscribers predicted to churn but did not. (Hung et al., 2006)

The lift rate is the ratio of hit rate (H%) to the overall monthly churn rate.

For instance, after building the model, we can rank order all customers by their churn score, and define hit ratio (H%) as the hit ratio of the "10% Customers with top churn score".

If the overall monthly churn rate = 2.5%and hit ratio (10% most risky churners) = 25%, then LIFT(10%)=25/2.5%=10

The lift curve is a measure of productivity for modeling: with random sampling of the entire customer base we would yield 2,5% churners; focusing instead on the 10% top-churn-risk-score customers we would yield 10% churners (Hung et al., 2006).

4. The churn model and the results

In our study we have used data from a mobile telecommunication company, taking a sample of 51691 cases – mobile postpaid customers, where a postpaid customer is a client having a contractual setting for a specific period of time (12 or 24 months) with the carrier – containing both customers that have churned in april 2010 (churners) and customers who have remained loyal in the same month (non-churners) and their historical data for a period of 6 months before the "churn moment" (October 2009 to march 2010). All chosen customers have had tenure of at least 9 months.

We have used 80% (41353 cases) of the entire data set for the training set, which is the data set used to build the model, 10% of the original data set for the test sample, which is the data set used to identify and correct wrong predictions of the model during the building phase, and 10% for the validation set, used to measure the performance of the model against the real outcome.

The training set contains 47.78% churners (19752) and 52.22% non-churners (21601).

There were 13 independent variables included in the model, that have been considered by our Chordiant application as statistically relevant in influencing the churn rate: average number of calls to the Call Centre in the past 6 months, number of visits in the representative stores of the carrier in the last 3 months, average invoice in Euro in the past 3 months, average invoice for voice services only in the past 3 months, average on net call duration (in minutes) for the last 4 months, number of months before expiring the contractual period, etc., each having a specific performance measured by the correlation coefficient (CoC). (See Figure 1)

Figure. 1 The list and performance of the independent variables

The model incorporates the following predictors:

Field	Performance
AVG_NO_ENQUIRIES_IN_PAST_6M	60.62
RETAIL_VISITS_3M	52.13
A3_TOTAL_NET_REV	55.87
NO_MONS_SINCE_LAST_SMS	56.26
SUM_MO_VOI_OTHER_SERV_MIN	56.94
SDV_MO_VOI_NAT_CALL_CENTER_MIN	54.86
RTS_MO_VOI_MIN	64.75
SDV_MO_VOI_NAT_CALL_CENTER_CNT	54.34
MONTHS_TO_OOC	54.98
A4_MO_VOI_MIN	56.73
A4_MO_VOI_NAT_OTHER_MIN	61.09
A3_AVERAGE_R_VOCE	52.86
A3_AVG_REVENUE_OFFNET_VOCE	52.12

Source: *Snapshot from the application*

The correlation coefficient (CoC) is a measure of how well trends in the predicted values follow trends in past actual values. It is a measure of how well the predicted values from a forecast model "fit" with the real-life data. In our case it measure the relative influence of each predictor to the outcome variable. We can observe in Figure 1, that the most "predictive" variable is the number of visits in the representative stores of the carrier in the past 3 months.

The performance of the model is given, in our case by the CoC, which is 76.001%, and the lift ratio, as defined in the previous section. The lift ratio for this model is 35,7% (see Figure 2), meaning that if targeting top 10% of the customers with the highest churn score, we can identify and retain 35,7% of them, compared with a 2,5% medium retention rate obtained if randomly selecting the customers.

Figure. 2 The ten deciles of top-churns and the model lift for each

Interval	churned		loyal		
	Count	Percentage	Count	Percentage	Lift
Wide of scheme	0	0%	0	0%	0
[0.003245, 0.019647>	646	2.62%	33442	10.38%	27
[0.019647, 0.025205>	772	3.13%	33919	10.53%	31
[0.025205, 0.029215>	799	3.24%	33845	10.51%	32
[0.029215, 0.034333>	1012	4.1%	33979	10.55%	41
[0.034333, 0.042442>	1121	4.54%	33576	10.43%	45
[0.042442, 0.05266>	1651	6.69%	34068	10.58%	65
[0.05266, 0.067689>	1962	7.95%	30415	9.44%	85
[0.067689, 0.098453>	2761	11.18%	32413	10.06%	110
[0.098453, 0.167329>	5010	20.3%	30162	9.37%	200
[0.167329, 0.815802]	8951	36.26%	26241	8.15%	357
	24685	100%	322061	100%	

Source: *Snapshot from the application*

5. Conclusions

Predicting churn customers using CRM strategies and data mining techniques is an important issue for telecom companies in competitive markets of today.

Our study contains important insights in building predictive models and an empirical example based on real data from a telecommunication Romanian operator, which is quite an innovative research for the Romanian market, national scientific literature on this subject being rather scarce.

According to the performance measures generally used in the international studies, it seems our model is significantly better than random, with a lift ratio of 35.7%.

As most studies of the telecom churn problem, we have focused on constructing an effective churn prediction model to locate lists of potentially lost customers in advance. However, identifying potentially lost customers does not mean those potentially lost customers can be retained and that will be the main subject of our future research studies.

6. References

[1] Au W, Chen KCC, Yao X,. "A novel evolutionary data mining algorithm with applications to churn prediction". *IEEE Trans. Evol. Comput*., 2003,7: 532-545.

[2] Berson A, Simith S, Thearling K., *Building data mining applications for CRM*, McGraw-Hill, New York, 2000

[3] Chu BH, Tsai MS, Ho CS., "Toward a hybrid data mining model for customer retention.", *Knowl. Based Syst*., , 2007, 20: 703-718.

[4] Coussement K, Van den Poel D,. "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", *Expert. Syst. Appl*., 2008, 34: 313-327

[5] Dwyer, F.R., Schurr, P. and Oh, S., "Developing buyer-seller relationships"', *Journal of Marketing*, Vol. 51, 1987 pp. 11-27.

[6] Greenberg, L.S.. *Emotion-focused therapy: Coaching clients to work through their feelings.*, American Psychological Association, Washington, DC, 2002

[7] Hung SY, Yen DC, Wang HY,. "Applying data mining to telecom churn management", *Expert. Syst. Appl*.,2006, 31: 515-524.

[8] Kim MK, Park MC, Jeong DH,. "The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services.", T*elecomm. Policy*, 2004, 28: 145-159.

[9] Lariviere B, Van den Poel D (2004). Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. Expert. Syst. Appl.,27: 277-285.

[10] Lejeune M (2001). Measuring the impact of data mining on churn management, Internet Research. Electr. Network Appl. Policy, 11:375-387.

[11]Lin J, Xu X., "A Novel Model for Global Customer Retention Using Data Mining Technology", *Data Mining and Knowledge Discovery in Real Life Application* , Viena, Austria , 2009, pp. 438-452

[12]Luo B, Shao P, Liu D, . "Evaluation of three discrete methods on customer churn model based on neural network and decision tree in PHSS",. *Proceedings of the first international symposium on data, privacy, and e-commerce*, 2007, pp. 95-97.

[13]Ngai EWT, Xiu L, Chau DCK,. "Application of data mining techniques in customer relationship management: A literature review and classification",. *Expert. Syst. Appl.*, 2009, 36: 2592-2602

[14]Panait I., "Stock Market Diagnosis", *Hyperion International Journal of Econophysics & New Economy*, 2011, vol.4, issue 2, pag. 313-315, ISSN 2069-3508

[15]Parvatiyar, A. J. N,. "Customer relationship management: Emerging practice, process, and discipline", *Journal of Economic and Social Research*, 2001, Vol.3, pp.1-34

[16]Slăvescu *E O,* "The Implementation of Uplift Modelling to Telecommunications Marketing Campaigns. The Case of the Romanian Mobile Telecommunications Market.", *Proceedings of The 7th Int. Conf. Management of Technological Changes,* Alexandroupolis, Greece, 2011

[17]Tsai CF, Chen MY, "Variable selection by association rules for customer churn prediction of multimedia on demand.", *Expert. Syst. Appl.*, 2010, 37: 2006-2015.

[18]Xia GE, Jin WD, "Model of customer churn prediction on support vector machine.", *Syst. Eng. Theory Pract*.,2008, 28: 71-77.