Paper 344-2008

Mobile Operator Customer Classification in Churn Analysis

Piotr Sulikowski, Technical University of Szczecin, Poland

ABSTRACT

Customer churn is a grave problem for all mobile operators. Early identification of customers from the risk group could help retain them in the operator's network. This paper introduces a set of potential churn factors on which data can be relatively easily extracted from the operator's databases and analyzed using the SAS® Enterprise Guide®. A multi-stage research procedure utilizing such real-world data is proposed. It allows the identification of significant churn factors, the segmentation of customers, and finally the establishing of a rule model of the phenomenon for each customer segment. The method outlined in the paper is based on rough set theory and takes into account both qualitative and quantitative data. The new approach proposed in the paper relies upon data from the current period and also knowledge from the past.

INTRODUCTION

Over the last years the mobile telecommunications market has been dynamically developing in Poland as well as most countries in the world. According to GSM World Association, an association of GSM (Global System for Mobile communication) operators and telecommunications regulators, as of March 30th, 2007 there were almost 2.4 billion users of GSM & 3GSM networks in almost 220 countries and territories, which accounted for 82% of all worldwide mobile digital communication. It is one of the most dynamically developing standards of wireless communication. The need to serve so many mobile customers causes a number of problems, churn in particular.

The purpose of this paper is to suggest a solution which would help discover rules governing churn-related behaviour of mobile telephony users and identify possible churners.

PURPOSEFULNESS OF THE RESEARCH

Due to the increasing competitiveness of this market - the growing number of suppliers, the introduction of more and more sophisticated products, services and special offers *etc.*, the sublimation of tastes as well as ever easier access to market information (Freeland 2002, p.5-6) - customer churn constitutes an increasingly significant problem that must be dealt with by operators. Churn consists in changing the supplier of products or services (to the benefit of a competitive supplier) or resigning from them definitely. It is one of the most serious problems for many businesses, in which customer loyalty has a key influence on the market success.

The reasons for churn can be numerous and having considered them churn can be divided into two main types: compulsory and voluntary. Resignation of the first kind is usually caused by unexpected circumstances on which the operator has no direct influence, *e.g.* if the customer moves outside of the serviced area; due to the customer's health or financial problems; or finally due to a better offer presented by a competitive operator. Voluntary churn is caused by factors, which can easily be influenced by the provider or to which the provider can effectively react, *e.g.* dissatisfaction with the services; bad past experiences with the provider; low quality customer service; rejected customer complaints; or ill-treatment of the customer in the case of temporary payment problems.

It seems obvious that operators should build individual relationships with their clients and continuously care for their loyalty from the very beginning by providing them with special tailor-made offers, regularly checking their satisfaction level, trying to meet their individual expectations, *etc.* The reason for such loyalty-oriented approach is that acquiring a new customer is usually much more expensive than retaining an old one.

Furthermore, customer churn could be reduced if the operator were able to successfully identify customers from the risk group before they inform the service provider of their intention of leaving. Were such a risk situation somehow foreseeable, the operator could, for instance, present the customer with a special individualised offer early enough, and thus, retain them in their network. It could be extremely profitable and could help gain market advantage, especially when the risk group includes key customers characterised by a high LTV (lifetime value) or customers with high potential for that.

CHURN FACTORS IDENTIFICATION

As a result of a survey conducted among a group of experts and thanks to the analysis of data stored in mobile operators' computer systems a number of possible churn factors have been identified.

Usually the operator does not possess enough marketing data about satisfaction and loyalty levels of each of their clients, which would make individual churn prediction possible. Moreover, a wide-ranging research permitting the acquisition of such individualised data appears far too expensive or even impossible. The operators, however, do usually possess a lot of other data which might be helpful while identifying situations of high risk with respect to customer churn. Such situations can be signalled by a customer contacting the provider's customer service helpline, where they enquire about the consequences of resigning from the contract or about switching from postpaid to prepaid service. Another alarming signal can be a change in user behaviour, *e.g.* with regard to their usage level, call duration, target phone numbers or distance between callers (Todman 2003, p. 80).

While signals of the first kind may not be noted anywhere in the operator's helpline or CRM system (Customer Relationship Management) in general, the signals connected with user behaviour are stored within the billing system. Some data which could also be utilised in modelling churn-related behaviour are also gathered in other IT systems of the operator.

In the initial research, the following churn-related factors were suggested (Roh et al. 2000, Sulikowski 2006):

- seniority (currently understood as the number of years the customer has been using the operator's network; before number portability, it also reflected the customer's attachment to the phone number),
- reaction to customer's complaints (the dominant reaction of the operator to the complaints forwarded by the customer – acceptance or rejection),
- charge dynamics (*e.g.* the percentage relation of customer charges in the last monthly clearing period and the average charges in the previous three periods),
- average charge (average monthly charge over the last 12 months),
- maximum charge (maximum monthly charge so far),
- customer's assessment of the provider's offer (in relation to other operators' offers),
- age,
- hometown and residential area,
- additional services (information on the use of extra services such as SMS, MMS *etc.*).

In a test survey conducted by the author, a group of 30 respondents were questioned on the first three factors (Sulikowski, Budzinski 2005). In the meantime, the leading Polish operators were consulted on the availability of customer data in their IT systems. Access to databases of one of the service providers was granted. Owing to that as well as a survey conducted among a group of experts, the following set of churn factors was identified basing on the available real-life data (*Table 1*).

Group of factors	Factor
Demographic data	age
	sex
	postal district of place of residence
	population of district ('powiat') of residence
Loyalty data	seniority
	type of the last promotional offer (and the loyalty contract)
	period of the last loyalty contract
	months left before the end of the current loyalty contract
	handing in the deactivation application
Financial data	monthly standing charges
	changes in the monthly standing charges
	monthly invoice amount
	changes in the monthly invoice amounts
Call data	number of monthly outgoing calls
	changes in the number of monthly outgoing calls
	duration of monthly outgoing calls
	changes in the duration of monthly outgoing calls
Additional services data	caller identification restriction on
	number of monthly outgoing short messages
	changes in the number of monthly outgoing short messages
	number of monthly outgoing multimedia messages
	changes in the number of monthly outgoing multimedia messages

Table 1. *Churn factors*

As can be noticed, the factors have been organised into 5 groups. Some of the initially suggested factors had to be ommitted in the final set since information on them was simply unavailable or impossible for the operator to easily access. Personal Data Act and the rules connected with granting access to internal operator data also had to be observed.

The necessary data have been successfully acquired and they concern a random set of postpaid individual customers over 12 consecutive billing periods. To enable classification as churners or non-churners, information on whether a customer is retained within the operator's network at the end of each billing period was also obtained.

CUSTOMER CLASSIFICATION METHOD

The problem of customer churn can be helped by using various statistical and data mining methods. Using data stored by the operator, a churn model can be created and the rules governing it can be discovered. However, the obtained data are not always certain and full, sometimes they are of a qualitative rather than quantitative nature. Moreover, the examined customers constitute a very heterogenous group. That implies what kind of methods can be used in mining the data and extracting knowledge.

In the proposed method, classical rough set theory is applied (Pawlak 1982). SAS® Enterprise Guide® and rough-sets-related software are used.

The procedure consists of 5 main steps:
1. Data formatting
2. Major factors identification
3. Initial clients segmentation
4. Rule model creation
5. Verification

In the first step, the source data are transformed into a format necessary for the consecutive calculations and interpretations. Then, a statistical analysis is performed to identify factors strongly correlated with churn. Hellwig's method in particular can be applied. Therefore, from the numerous above-mentioned factors, on which data is available, it is possible to select only those having major influence on loyalty and churn. In the next stage, an initial segmentation of clients into relatively homogenous groups is performed. Finally, a rule model of the churn phenomenon is created for each of the identified segments, using a modified rough set method. The modification is discussed further in the article. The rule model obtained is verified to check its reliability.

The discovered rules governing customer behaviour can be generalised reliably to the whole population since the sample is random and large enough. Owing to that, one can infer about the other customers. The obtained decision algorithm allows the identification of customers from the risk group basing on past customer behaviour. It is also possible to couple the rules with LTV of each customer in the risk group, which would help the operator take appropriate measures.

Taking into consideration the fast-changing business climate, the analysis procedure described above is repeated regularly, for example every quarter, month *etc.* In this particular research procedure available customer data have been divided into 2 periods: one encompassing the first 6 billing periods, and the other encompassing the last 6 ones.

Repeating the analysis in the next period and discovering new rules basing only on data from that period usually eliminates the knowledge from the previous period. On the other hand, it does not seem reasonable to treat equally observations from both periods – due to the changes which separate them.

Therefore, it is proposed that data from both the previous and the current period be utilised differentiating their significance. This postulate can be realised by modifying the support level of particular rules assuming that *e.g.* observations from the previous period are combined with multiplied observations from the current period. It means that the number of the current observations is purposely increased by "cloning" them *n* times.

The multiplication coefficient n ($n \in N_0$) depends on the expert's assessment of the significance of the changes which separate the periods. In the suggested solution (see *Figure 1*) the coefficient $n = 2$, which causes that data from the last period are considered 2 times more significant than those from the previous one. Similarly, observations from more than two periods can also be analysed. *Figure 2* presents sample values of the n coefficient while considering the current period ($t = 0$) together with three past periods (t=-1; -2; -3).

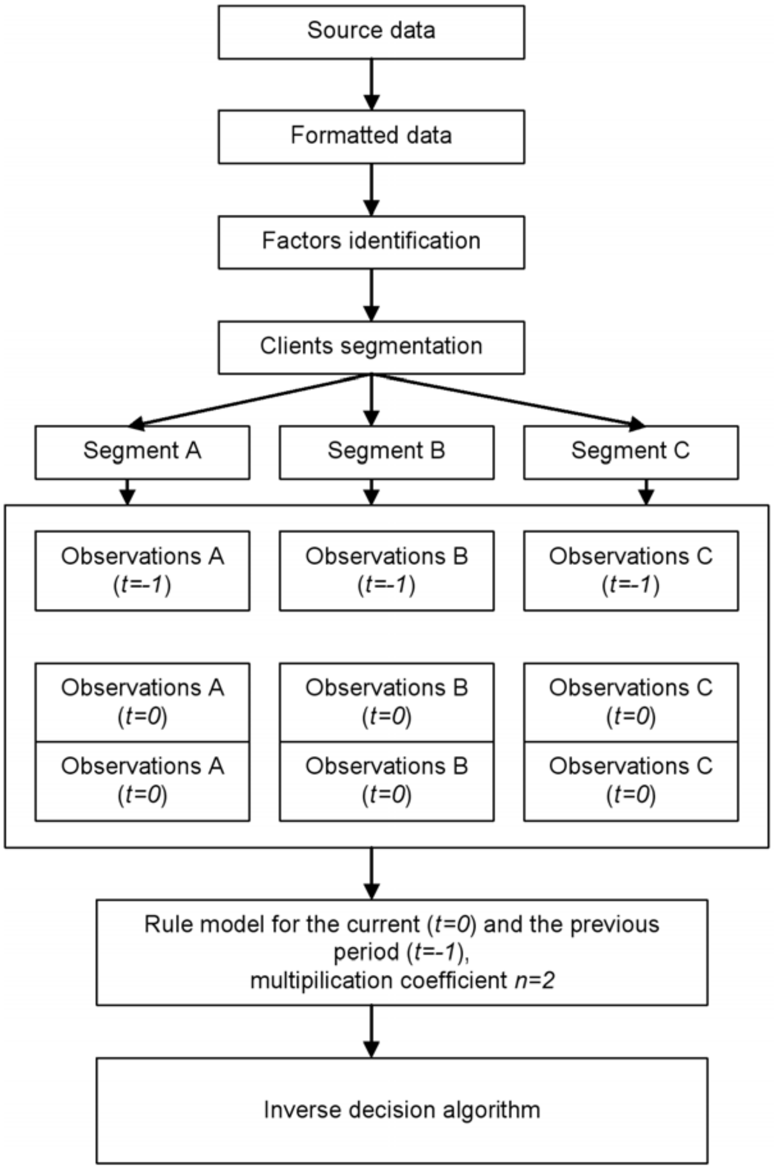

Figure 1. *Method outline for 2 periods*

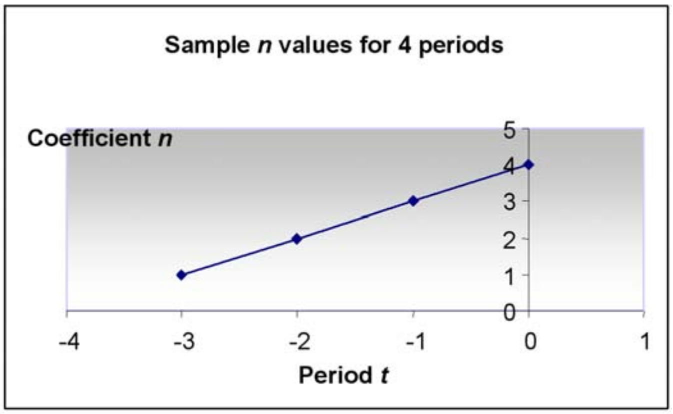

Figure 2. *Coefficient n for multiple periods*

Thanks to the described method, utilised for 2 or more periods, new rules governing customer behaviour can be discovered. They are based on a longer period and they differentiate the significance of information from different moments of time. Then, deducing an inverse decision algorithm allows for characterising loyal clients as well as the resigning ones. That can help identify potential churners.

CONCLUSION

The proposed method of customer classification allows fast and inexpensive churn analyses since it utilises data normally stored by mobile operators. A rule model of the churn phenomenon is obtained – a model taking into account both quantitative and qualitative data and, what is crucial, differentiating data from the current period from those from the past.

In this paper, the problem matter is related to the mobile telecommunications sector. Among many other companies which are particularly susceptible to churn the following types of enterprises should also be mentioned: Internet service providers, companies offering retail financial services, public utility enterprises and retail supermarket chains (Freeland 2002, p. 13). All of them are characterised by large numbers of customers. It has to be stated that the method outlined in the article can be applied in them per analogia and, thus, can help identify possible churners.

REFERENCES

GSM World Association. 2008. *Statistics.* http://www.gsmworld.com/news/statistics

Freeland, J. 2002. *The Ultimate CRM Handbook*: *Strategies and Concepts for Building Enduring Customer Loyalty and Profitability*. McGraw-Hill

Todman, Ch. 2003. *Projektowanie hurtowni danych. Zarządzanie kontaktami z klientami (CRM)*. WNT.

Roh, T.H., Han, I. and Jang, W.H. 2000. *The Churn Management for Telecom Market Using the Knowledge Discovery in Database*. Seul: KaistGSM

Sulikowski, P. 2006. „Metoda klasyfikacji klientów operatora telefonii komórkowej w analizie zjawiska rezygnacji." *Multimedia w organizacjach gospodarczych i edukacji.* CDiI Difin, 153-158

Sulikowski, P., Budziński, R. 2005. "Application of rough set theory to customer loyalty analysis and churn management in the mobile telecommunications sector". *Information Management*. Wydawnictwo Uniwersytetu Gdańskiego, 422-432

Pawlak, Z. 1982. "Rough Sets." *International Journal of Computer and Information Sciences 11*: 341-356

Mozer, M.C. et al. 2000. "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry." *IEEE Transactions on Neural Networks 11*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Piotr S. Sulikowski
Faculty of Computer Science and Information Technology
Technical University of Szczecin
Ul. Zolnierska 49
PL-71-210 Szczecin
Poland
E-mail: psulikowski@wi.ps.pl