UNIVERSITEIT GENT

FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Modeling Partial Customer Churn: On the Value of First Product-Category Purchase Sequences

Vera L. Miguéis [1]

Dirk Van den Poel [2]

A.S. Camanho [3]

Joao Falcao e Cunha [4]

May 2012

2012/790

[1] PhD candidate, Faculty of Engineering, University of Porto

[2] Professor of Marketing Modeling, Ghent University, Tweekerkenstraat 2, B-9000 Gent, dirk.vandenpoel@UGent.be
http://www.crm.UGent.be

[3] Professor, Faculty of Engineering, University of Porto

[4] Professor, Faculty of Engineering, University of Porto

D/2012/7012/23

Modeling partial customer churn: On the value of first product-category purchase sequences

V.L.Miguéis [a,*], Dirk Van den Poel [b], A.S. Camanho [a], João Falcão e Cunha [a]

[a] *Faculty of Engineering, University of Porto*

[b] *Faculty of Economics and Business Administration, Ghent University*

Abstract

Retaining customers has been considered one of the most critical challenges among those included in Customer Relationship Management (CRM), particularly in the grocery retail sector. In this context, an accurate prediction whether or not a customer will leave the company, i.e. churn prediction, is crucial for companies to conduct effective retention campaigns. This paper proposes to include in partial churn detection models the succession of first products' categories purchased as a proxy of the state of trust and demand maturity of a customer towards a company in grocery retailing. Motivated by the importance of the first impressions and risks experienced recently on the current state of the relationship, we model the first purchase succession in chronological order as well as in reverse order, respectively. Due to the variable relevance of the first customer-company interactions and of the most recent interactions, these two variables are modeled by considering a variable length of the sequence. In this study we use logistic regression as the classification technique. A real sample of approximately 75,000 new customers taken from the data warehouse of a European retail company is used to test the proposed models. The area under the receiver operating characteristic curve and 1%, 5% and 10% percentiles lift are used to assess the performance of the partial-churn prediction models. The empirical results reveal that both proposed models outperform the standard RFM model.

Key words: Marketing, Customer relationship management, Churn analysis, Predictive analytics, Sequence analysis, Retailing, Classification, Logistic regression

1 Introduction

In the last decades, the emerging computing technologies led to a deep evolution in the ability of companies to collect, store and analyze large datasets. For each customer, thousands, or even millions of data objects are stored, enabling the analysis of the complete purchasing history. Moreover, the changes in the relationship between companies and customers, due to the recent economic and social changes, has made companies change from transaction marketing to relationship marketing. This change is commonly acknowledged as the paradigm shift in marketing (Brodie et al., 1997; Grönroos, 1994). In this context, Customer Relationship Management (CRM) is becoming an essential component of business management. Indeed, both practitioners

* Corresponding author. Tel: +351225082132. E-mail address: vera.migueis@fe.up.pt

and academics have concluded that important business knowledge can be extracted from the analysis of the data stored in their warehouses.

Ling and Yen (2001) define CRM as a set of processes and systems that support companies' business strategy to build long term and profitable relationships with specific customers. Ngai et al. (2009) summarizes customer relationship management as the combination of four dimensions: customer identification, customer attraction, customer development and customer retention. In today's competitive environment, customer retention is gaining particular attention from companies. Customers' life cycles are becoming more transitory than in the past, mainly due to the impact of competitors' actions on the existing relationships. In retailing, some customers are classified as cherry pickers, i.e. they visit multiple retailers when shopping for a basket of goods (Mogelonsky, 1995) while others are store switchers (Popkowski Leszczyc and Timmermans, 1997). This behavior is due to the fact that customers experience very little switching cost and do not have to inform companies about their churn intention. According to EFMI and CBL (2005), a total of 87% of grocery customers use two or more different supermarkets for their grocery shopping and, on average, those grocery shoppers visit 2.8 different supermarkets each month.

The advantages related to customers' loyalty justify the increasing concern of companies about customer attrition. Previous research states that retaining existing customers is less costly than acquiring new customers (Rosenberg and Czepiel, 1984; Reichheld and Sasser, 1990). Loyal customers tend to increase their spending over time (Reichheld and Teal, 1996) and spread positive word-of-mouth (Gremler and Brown, 1999). Moreover, these customers are less costly to serve (Knox, 1998) and exhibit reduced sensitivity to competitors' actions (Stum and Thiry, 1991) and prices (Zeithaml et al., 1996). As a result, reductions in the rates of customer defection result in profit increases (Reichheld and Sasser, 1990). Despite all these related benefits, churn analysis in retailing can still be considered incipient (Buckinx and Van den Poel, 2005).

Ngai et al. (2009) show the data mining tools that usually assist each CRM dimension. Classification tools are the most frequently used to model customer churn prediction. Examples of classification techniques used for this purpose include Neural networks (e.g. Tsai and Lu (2009)), Decision trees (e.g. Nie et al. (2011)), Random Forests (e.g. Xie et al. (2009)), Support vector machines (e.g. Coussement and Van den Poel (2008)) and Logistic regression (e.g. Nie et al. (2011)). The predictors included also distinguish the churn prediction models present in the literature. Buckinx and Van den Poel (2005) classify defection predictors in three categories: behavioral antecedents, demographics and perceptions. This paper proposes a new churn prediction model for retailing that introduces the first category purchase succession as a predictor which reflects behavioral antecedents. Unlike most of the previous churn models proposed in the literature, this paper aims to identify those customers who are going to defect even if it is only partially.

The structure of the remainder of the paper is as follows. Section 2 includes a brief revision of churn prediction modeling in the literature. Section 3 introduces the motivation for the inclusion of the first purchase succession in the proposed churn model. Section 4 introduces the methodology followed in this paper, namely the predictors used, the classification technique used and the performance evaluation criteria used. Section 5 presents the application, i.e. the company used as case study and the results. The paper finishes with the conclusion and some issues for future research.

2 Modeling customer churn

The identification of customers who exhibit large potential to abandon the existing relationship has deserved particular attention of companies in several domains. In the telecommunications setting, churning is usually referred to as changing phone operator. For example, Marcin (2010) predicts the churn probability of prepaid clients of a cellular telecommunication company. Also in this domain, Hung et al. (2006) compare the performance of several data mining techniques in the definition of a churn propensity score. In financial services (banking and insurance), churn is usually seen as closing accounts. Larivire and Van den Poel (2005) investigate the inclusion of several explanatory variables and compare some modeling methods for customer churn prediction in a financial services company. Hur and Lim (2005) predict the switching probability of an insured to other auto insurance company and compare the performance of some classification techniques. In subscription-based businesses, churn refers to the failure to resubscribe to the service under consideration. For example, Coussement and Van den Poel (2008) predict the probability of customers to cancel their newspaper subscription. Burez and Van den Poel (2007) develop a churn prediction model and test it by using data from a pay-TV company. In grocery retailing, churn has been considered as the partial and progressive defection of customers. Buckinx and Van den Poel (2005) use several classification techniques to build partial defection models in grocery retail. Burez and Van den Poel (2009) propose a method to handle class imbalance in churn prediction, namely in retail context. A review of the literature on the churn prediction modeling in the several domains can be found in Verbeke et al. (2011).

In contrast to other churn prediction domains, in grocery retail, it is difficult to identify the exact moment when clients discontinue their relationship with companies. Indeed, there is no particular point in time at which customers have to reveal their intention to suspend the relationship. Moreover, customers typically do not switch their grocery supplier suddenly. Most customers exhibit partial defection (Buckinx and Van den Poel, 2005), which may subsequently lead to a complete switch. Buckinx and Van den Poel (2005); Burez and Van den Poel (2009) use the concept of partial churn to identify customers that the company should focus on when concerned about customer retention.

Prior studies in the grocery-retail context include a large set of predictor variables. Buckinx and Van den Poel (2005) include predictors that intend to express: interpurchase time, frequency of purchases, amount of money spent, shopping behavior across product categories, brand purchase behavior, length of the relationship, timing of shopping, mode of payment, promotional behavior and customer demographics. This study concludes that behavioral recency, frequency and monetary (RFM) variables are those that have more discrimination power. Burez and Van den Poel (2009) use a subset of the same variables considered in the previous study, resulting from a stepwise variable selection procedure.

3 Sequence mining

In the grocery retail setting, the available data set is generally timestamped transactional data and static data (e.g. demographics and address). Transactional data is a set of sequencial timestamped events which can be easily stored in relational database tables. Events represent interactions between customers and companies, and most represent purchases. Prinzie and Van den Poel

(2007) and Prinzie and Van den Poel (2006b) are examples of studies which analyze customer purchase events to support customer relationship management.

We believe that such event analysis may reveal the state of trust of a customer towards a company. Particularly the succession of first category purchases may represent the process of development of the relationship of trust of a customer in the company and consequent demand maturity. Li et al. (2005) considers that at different stages of customers' demand, customers present different requirements which are derived in a particular product-purchase sequence. Typically, new customers have little knowledge of the product categories they are trying to buy into. Most customers will try to reduce risk in this situation and consequently establish goal hierarchies. By doing so, customers break down the purchase process into portions which can take a share of the risk (Blythe, 2007). Usually, customers who did not enjoy the first experience risk not reaching their initial aspiration level are not motivated to take risks with late goals (Dhar and Novemsky, 2002).

Each customer perceives products' category risk differently and in addition each customer presents specific demand structures. However, we consider that risk and maturity stage may be connected with the promptness of a customer to churn or not. Therefore, we hypothesize that first-product category purchase sequences may support churners and non-churners discrimination.

Some authors claim that the beginning of the relationship between customers and companies is critical for the development of a long term relationship (Lawson-Body and Limayem, 2004), while others claim that the course of the relationship replaces the initial impressions (Redondo and Fierro, 2005). Thus, we analyze both the churn prediction power derived from the sequence of first product category categories purchases in chronological order and in reverse chronological order, placing the most recent categories at the bottom, respectively.

For each succession of first category purchases, according to the specific purchase process, the period of definition of a first impression can be different. Moreover, the relevant history of first category purchases can be distinct. Therefore, we consider in both forward and backward analysis a variable categories' succession length. Indeed, we use the idea underlying the variable memory concept, introduced by Rissanen (1983) in variable length markov chains (VLMC) context, to incorporate in the proposed model an adjusted category first purchase sequence. For each customer, in both analyses, we consider different succession lengths based on the discrimination power obtained for different succession lengths.

4 Methodology

This paper aims to predict partial churn by considering a variable length of the first-category purchase sequence for each customer, depending on whether a longer first sequence contributes to a more accurate model. We run two distinct models: one in which the first purchase succession is included in chronological order (forward model) and another in which the categories first purchase succession is included in reverse chronological order (backward model), placing the most recent categories at the bottom.

The relevance of the two proposed models is measured by comparing their performance with the standard RFM model performance. The classification technique used in this paper is Logistic regression.

In the next paragraphs we present the criterion used to infer partial attrition, the classification technique used, the evaluation criteria used to compare models' performance and the explanatory variables included in the models.

4.1 Partial churning

Since in non-contractual businesses the churn event is not explicit, in order to construct a binary classification model, we have to define an attrition criterion that identifies the switchers. Therefore, we derive the dependent variable of the model by grouping the purchases in periods of three months and by classifying as churners those customers who, from a certain period, did not buy anything else or those who in all subsequent periods spent less than 40% of the amount spent in the reference period. Figure 1 shows examples of the derivation of the independent variable for two hypothetical customers, according to their purchases distribution.

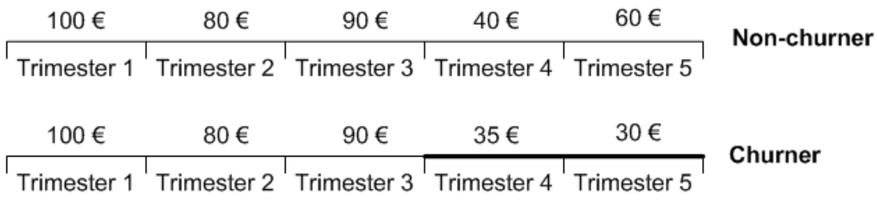

Fig. 1. Examples of derivation of partial churning variable.

The first distribution of purchases represents a non-churner customer, since there is no quarter in which the value spent in all subsequent quarters is less than 40% of the amount spent in such quarter. Concerning the second distribution, this represents a customer who partially churned. Considering as reference the 3nd quarter we conclude that in all subsequent periods this customer buys less than 36 ($90\times40\%$). Therefore, we assume that this customer churned in the beginning of the 4th quarter.

4.2 Logistic Regression

In the extant literature there are several regression techniques which differ in terms of research objectives and variable characteristics. Linear regression is the regression technique most frequently used. Linear regression considers continuous dependent variables and assumes that the relationship between the dependent variable and the independent variables is linear. When the objective is to examine the effect of a single independent variable on a dependent variable, we conduct a simple linear regression. When the objective is to analyze the influence of multiple factors on the dependent variable at the same time, we conduct a multiple linear regression.

Logistic regression is a particular case of general linear models. This regression technique considers binominal dependent variables, instead of continuous variables, and assumes a linear function between the dependent variables expressed in the logit scale and the independent variables, rather than in the original linear format. Ordinary linear regression techniques are insufficient to estimate the dependent variables, because they enable the dependent variables to fall outside the $0 - 1$ range. This is not admissible when dealing with binary dependent variables, since these are expressed as a probability which must fall between 0 and 1. Thus, the logistic regression is based on a mathematical transformation of the original linear regression to yield the logit or natural log of the odds of being in a dependent category (Y) versus in the other category ($1 - Y$).

The main reasons behind the application of logistic regression are the ease of use and the quick and robust results (Buckinx and Van den Poel, 2005). Moreover, logistic regression does not rely on assumptions of normality for the predictor variables or the errors and may handle non-linear effects (Fahrmeir, 1985).

4.3 Evaluation criteria

It is crucial to evaluate the models proposed in terms of performance. Rosset et al. (2001) present a brief review of the evaluation metrics in marketing context.

To measure the performance of the binary prediction models, we compute the receiver operating characteristic curve (ROC). ROC represents the trade off between the proportion of false non-churners and the proportion of churners for every possible cut off of the binary outcome predicted probability. Then, we use the area under the ROC curve (AUC) as a performance measure. An AUC close to 1.0 reveals that the model has perfect discrimination, while an AUC close to 0.5 suggests poor discrimination (Hanley and McNeil, 1982). We use the AUC to compare the predictive performance of the proposed models by conducting the non-parametric test introduced by DeLong et al. (1988).

Moreover, we use lift as evaluation metric. This measure focuses on the segment of customers with the highest risk to the company, i.e. customers with the highest probability to churn. The definition of lift depends on the percentage of customers the company intends to achieve on a retention campaign. Consider that a company is interested in the top p-th percentile of most likely churners, based on predicted churn probabilities. The top $p-th$ percentile lift then equals the ratio of the proportion of churners in the top $p - th$ percentile of ordered posterior churn probabilities to the churn rate in the total customer population. For example, a $p-th$ percentile lift of 3 means that the model under investigation identifies three times more churners in the top p% than a random assignment would do. Since the proportion of customers that a company is able and willing to target depends on the specific content, namely the available budget, this paper includes the lift performance for different percentiles (1%, 5%, 10%).

The performance measures are computed by splitting the dataset in 80% for training and 20% for test purposes. This subdivision is stratified, such that the percentage of churners in both training and test data is approximately the same as that in the initial dataset.

4.4 Explanatory variables

In this study we explore the potential of past behavior variables to distinguish churners from non-churners. We include in both churn models the recency, frequency and monetary variables. Moreover, we include in the forward and in the backward models a set of dummy variables which indicate whether or not a specific succession of categories was observed on the past transactional behavior of the customer.

4.4.1 Recency

Recency is a temporal metric that indicates how recently a customer has made a purchase. This is typically considered the most powerful predictor of customer future behavior among the RFM variables (Miglautsch, 2000). Most prior studies conclude that the lower the value of recency, the lower the switch probability (e.g. Buckinx and Van den Poel (2005)). In this study, recency measures the number of days between the last transaction date and the end of the period of analysis. For model training purposes, the recency of churners is the number of days between the date of the previous transaction and the date on which those customers were classified as partial churners.

4.4.2 Frequency

Frequency is a measure of the strength of the customer relationship with the company. According to Calkins et al. (2005), frequency is a natural measure of behavioral loyalty. The more often a customer has purchased from a company, the more loyal a customer is. Therefore, frequency can be seen as a predictor of churn (e.g. Bolton et al. (2004)). Frequency is included in the model proposed as the average number of transactions by quarter. For model training purposes, frequency of partial churners is calculated by considering only the transactions observed until the churning date.

4.4.3 Monetary

Monetary is a measure of the amount spent by customers at a certain company. Typically, this RFM dimension is the least powerful concerning predicting ability, although still valuable when used in conjunction with the other two variables (Rud, 2000). This variable is covered in this study by means of the average amount of value spent by customers in each quarter. Regarding the partial churners used for training the model, this variable naturally only takes into account the amount spent until the date they were classified as churners.

4.4.4 Forward categories succession and Backward categories succession

Both forward categories succession and backward categories succession are included in the different models by means of a set of dummy variables. In the forward model, each forward category succession dummy variable represents a succession of first products' categories purchased in chronological order, i.e. the fist category bought is at the bottom and the most recent are on top. In the backward model, each backward category succession dummy variable represents a succession of first products' categories purchased in reverse chronological order, i.e. the category bought most recently is in the bottom and the most past ones on top.

The selection of the category succession dummy variables for the two models follows a similar procedure. In the first stage, we compute the dummy variables representing all sub-successions with length two, i.e. the first category bought and the second category bought, in the case of the forward model, and, the last category bought and the second last category bought, in the case of the backward model. Then, for those successions of length two, which are observed for more than 500 customers, we compute the corresponding successions of length three, and so on. We believe that this first selection process based on the successions frequency enables to exclude from the posterior analysis successions which would not contribute to the discrimination between churners and non-churners. In the second stage, we consider all dummy variables resulting from the first selection to conduct an in-depth logistic regression analysis. It means that we run a Logistic model including each dummy variable corresponding to the categories succession of length two and the RFM variables. For those successions which result in an increase of the AUC in relation to the AUC observed for the RFM model, we go deeper, i.e. we run a Logistic regression with each of dummy variable corresponding to the categories succession of length tree. This procedure is progressively executed and stops when the AUC does not increase. In the third stage, identified the dummy variables which individually increase the prediction performance in terms of AUC, we conduct a multivariate Logistic regression which includes all these variables.

5 Case study

This Section applies the proposed methodology to a real-world churn setting. First we present the company used as case study, then we present the results of the prediction models and finally some suggestions of retention actions.

5.1 Company

In this paper we use a European retailing company as a case study. This company has a chain of foodbased stores, i.e. hypermarkets, large supermarkets and small supermarkets. These formats differ essentially by the range and price of products offered, by the sales area and by the size of the town where they are located.

The company has adopted a customer centered strategy, in which customer relationship management is seen as a strategic issue. In this context, the development of the information system and the implementation of a loyalty program have enabled collecting data on each customer profile and on their transactions. This program is supported essentially by a loyalty card. Currently approximately 80% of the total number of transactions is done by customers using the loyalty card.

At present, the company identifies distinct subsets of customers in two ways. One of them consists of grouping customers based on their shopping habits. This segmentation model is a simplified version of the RFM model proposed by Bult and Wansbeek (1995), and is called internally: “frequency and monetary value” (FM) model. According to the values of these two variables, the company specifies 8 groups of customers. Each client ends up in one of these groups, according to the average number of purchases done in an 8-week period and the average amount of money spent per purchase. The changes in the percentage of customers belonging to each group are used to guide the marketing actions required to meet the company’s objectives. For example, if the number of customers in the clusters with more visits to the store decreases, the company is alerted to launch marketing campaigns in order to motivate customers to go to the stores more often (see Miguéis et al., 2011). The other method of segmentation is based on customer necessities and preferences. In this case, customers are grouped into 7 segments according to the mix of categories and products they purchase. Each segment is defined by using a clustering algorithm in which customers presenting similar percentages of products purchased belonging to predefined groups of products are grouped. Both methods of customer segmentation do not care about churn prediction. None of the methods of customer segmentation takes into account churn prediction. In fact, only the first method presented addresses churning, but in a reactive way, i.e. after the churn event has occurred.

The churn prediction analysis reported in this paper is based on transactional data of customers with a loyalty card. The database provided includes the records of two hypermarkets from January 2009 to December 2010. Each transaction includes: the client identity number, the date and time of the transaction, the product transacted and the price of the product. This study is only focused on the new customers, since only for those it is possible to trace the first category purchase sequence. We consider new customers those who did not buy in the first semestre of 2009 but spent at least 100 euros until the end of 2010. This corresponds to a total of 74.607 new customers. From these, according to the criterion presented in Section 4.1 to identify partial churners, 44% partially churned during the period of analysis, while the remaining stayed active.

The company classifies its products at several levels, such as: department, business unit, category and subcategory. Therefore, the company characterizes its products in 20 departments (e.g. Food, Textile), 65 business units (e.g. Grocery, Baby textile), in 247 categories (e.g. Milk/Soy drinks, Baby shoes) and in 1013 subcategories (e.g. Pasteurized milk, Baby MNO). For products succession analysis, we focus on the business unit level since this enables to incorporate discrimination between products and avoid large complexity. Moreover, we only select for the analysis the most representative business units concerning the purchases done by the new customers, i.e. those business units which were part of the market baskets of at least 10% of the new customers considered in the analysis. Therefore, we consider 20 business units, i.e. Grocery, Hygiene/cleanliness, Daily/frozen, Drinks, Fruits/vegetables, Bakery, Delicatessen, Culture, House, Butchery, Fishery, Takeaway, Bricolage/auto, Leisure, Stowage, Pets/plants, Women textile, Men textile, Baby textile, Child textile. In this paper these business units are referred to as categories.

5.2 Predictive Models Performance

By applying the forward churn prediction analysis proposed, we identified in the first stage 6968 forward categories successions which have relevance in terms of frequency. As shown in Table 1, 380 successions of length two are analyzed, i.e. combinations of two out of 20 products' categories, while 6588 successions of length tree are analyzed, i.e. combinations of tree out of 20 products' categories whose corresponding successions of length two are observed for more than 500 customers (366 successions of length two). The selection based on the frequency excludes from the analysis successions with a length higher than 3.

From the 6968 successions analyzed, only 820 allowed to increase the performance of the model in relation to the RFM model, whose AUC is 0.856. From these 820 successions, 96 have length two and the remaining have length tree. This means that not only the very beginning of the relationship is important to define whether a customer will stay active. It is interesting to note that most of the individually significant successions have as first category bought: stowage, takeaway, bakery and fishery. This type of analysis can be useful for the company to understand what is behind the churn process.

Table 1
Forward model - first stage successions selection.

Succession length	Successions considered
2	366
3	6588

By applying the backward churn prediction analysis proposed, we selected in the first stage 7220 backward categories successions. As shown in Table 2, we analyze all possible combinations of successions of length two and three. Once again, the frequency observed for the successions of length tree indicate that successions having higher length would not contribute to distinguish churners from non-churners.

Table 2
Backward model - first stage successions selection.

Succession length	Successions considered
2	380
3	6840

From these 7220 successions, 1631 enabled to increase the performance of the model in relation to the RFM model. From these 1631 relevant successions 184 have length 2 and 1447 have length 3. The analysis of these sequences reveals that the last categories bought by the first time that seem to have some discrimination power are: Delicatessen, Hygiene/cleanliness, Butchery, Daily/frozen, and Grocery.

The performance mesures of the RFM model, the forward and backward models in terms of AUC, top 1%, 5% and 10% percentiles lift are shown in Table 3.

Table 3
Performance results.

	Model		
	RFM	Forward	Backward
AUC	0.856	0.864	0.867
Lift 1%	1.790	1.898	1.959
Lift 5%	1.984	2.017	2.014
Lift 10%	1.936	2.026	2.010

From the analysis of Table 3, for this real example we can conclude that both forward and backwards models outperform RFM model in terms of AUC. The AUC increases about 1% by adding the dummy variables representing both the forward and backward sequences. In order to ensure that the differences in AUC are significant, we applied the DeLong et al. (1988) test which confirmed the differences. Moreover, the beneficial effect of the categories successions is also confirmed in terms of top 1%, 5% and 10% percentiles lift. Therefore, we can conclude that the state of trust and demand maturity of a customer towards a company reflected either by the initial categories successions either by the most recent categories successions can improve the partial churn prediction.

The performance of the two models suggested seems to be similar. The test proposed by DeLong et al. (1988) confirms that there is no evidence of significant differences between both AUC. This suggests that the impact of the first impression and the impact of the most recent risks taken by customers on the promptness to churn is approximately the same.

5.3 Retention actions

The contribution of the proposed models for the company lies in the prevention of wasting budget on mass marketing approaches. In fact, an accurate churn prediction model enables the company to target the real churners, by identifying those customers with the highest probability of churn.

Managing customer expectations to improve satisfaction is one of the best customer retention strategies that the company can develop. Customers have expectations concerning, for example product quality, range of products, service responsiveness, price stability, promotional activity and staff empathy. Therefore, it seems of utmost importance for the company to clearly know the expectations of the potential churners, by means of an individual contact. This can support the design of an effective customized service which may ensure the long term relationship of the customers with the company.

Moreover, other retention strategies, not directly related to customer expectations, can be undertaken. For instance, to be sure to stay in touch with the potential churners by placing phone calls or send emails, special offers, customized advertising, follow-ups, and cards or notes with a personal touch. The company can also consider to send good deals to the potential churners identified by the models. These might be, for instance, discounts on the purchases and an extra gift included with a purchase. Customers usually answer to these actions positively, since they will feel valued and important for the company.

6 Conclusion and Issues for further research

In this study, we have proposed two predictive models for partial customer churn in retail. Both models include the succession of first-category purchases as a proxy of the state of trust and demand maturity of customers towards a company. Considering the impact of the first impression on the current state of the relationship as well as the impact of the most recent risks undertaken, we model the first purchase succession in chronological order as well as in reverse order, respectively. Both successions of first categories purchase are modeled by considering a variable length, defined according to the models' accuracy.

In order to test the proposed model in a real context, we used as case study a European retail company. This fact allowed us to test the model with a large dataset. The results reported reveal that both proposed models outperform the standard RFM model, what highlights the relevance of the state of trust and demand maturity in the partial churn prediction. Models performance is measured in terms of area under the receiver operating characteristic curve (AUC) and $p-th$ percentile lift.

By using the prediction model proposed in this paper, companies can define the target of future retention marketing campaigns. We present in this paper some actions that can be conducted in these retention campaigns.

This study is limited to analyzing sequences in shopping baskets in the offline world (what people purchased in the physical store). Future studies might combine this information with customers' clickstreams online on the retailer's website (Van den Poel and Buckinx, 2005). Moreover, this study analyzed sequential information using Markovian transition matrices. Future studies may utilize other methodologies such as SAM (sequence alignment methods). See Prinzie and Van den Poel (2006a).

References

Blythe, J. (2007). *Consumer Behaviour*. Thomson Learning, London.

Bolton, R. N., Lemon, K. N., and Verhoef, P. C. (2004). The theoretical underpinnings of customer asset management: A framework and propositions for future research. *Journal of the Academy of Marketing Science*, 32:271–292.

Brodie, R. J., Coviello, N. E., Brookes, R. W., and Little, V. (1997). Towards a paradigm shift in marketing? an examination of current marketing practices. *Journal of Marketing Management*, 13:383–406.

Buckinx, W. and Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1):252–268.

Bult, J. R. and Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4):378–394.

Burez, J. and Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2):277–288.

Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36:4626–4636.

Calkins, T., Tybout, A., and Kotler, P. (2005). *Kellogg on Branding: The Marketing Faculty of The Kellogg School of Management*. Wiley, New York, 1 edition.

Coussement, K. and Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1):313–327.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3):837–845.

Dhar, R. and Novemsky, N. (2002). The effects of goal fulfillment on risk preferences in sequential choice. *Advances in Consumer Research*, 29:6–7.

EFMI and CBL (2005). Consumenten trends 2005. Technical report, Rotterdam/Leidschendam.

Fahrmeir, L. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, 13(1):342–368. Mathematical Reviews number (MathSciNet): MR773172; Zentralblatt MATH identifier: 0594.62058.

Gremler, D. D. and Brown, S. W. (1999). The loyalty ripple effect: Appreciating the full value of customers. *International Journal of Service Industry Management*, 10:271–293.

Grönroos, C. (1994). From marketing mix to relationship marketing: Towards a paradigm shift in marketing. *Management Decision*, 32(2):4–20.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.

Hung, S., Yen, D. C., and Wang, H. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3):515–524.

Hur, Y. and Lim, S. (2005). Customer churning prediction using support vector machines in online auto insurance service. In Wang, J., Liao, X., and Yi, Z., editors, *Advances in Neural Networks ISNN 2005*, volume 3497, pages 928–933. Springer Berlin Heidelberg, Berlin, Heidelberg.

Knox, S. (1998). Loyalty-based segmentation and the customer development process. *European Management Journal*, 16(6):729–737.

Larivire, B. and Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2):472–484.

Lawson-Body, A. and Limayem, M. (2004). The impact of customer relationship management on customer loyalty: The moderating role of web site characteristics. *Journal of Computer-Mediated Communication*, 9(4).

Li, S., Sun, B., and Wilcox, R. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research*, 42(2):233–239.

Ling, R. and Yen, D. C. (2001). Customer relationship management: An analysis framework and implementation strategies. *Journal of Computer Information Systems*, 41(3):82.

Marcin, O. (2010). Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*, 37(6):4710–4712.

Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing*, 8:67–72.

Miguéis, V. L., Camanho, A. S., and Cunha, J. F. (2011). Mining customer loyalty card programs: The improvement of service levels enabled by innovative segmentation and promotions design. *Lecture Notes in Business Information Processing*, 82:83–97.

Mogelonsky, M. (1995). *Everybody Eats: Supermarket Consumers in the 1990's*. Amer Demographics Books, Ithaca, New York, 2nd edition.

Ngai, E., Xiu, L., and Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2, Part 2):2592–2602.

Nie, G., Rowe, W., Zhang, L., Tian, Y., and Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 38(12):15273–15285.

Popkowski Leszczyc, P. and Timmermans, H. (1997). Store-Switching behavior. *Marketing Letters*, 8(2):193–204.

Prinzie, A. and Van den Poel, D. (2006a). Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM. *Decision Support Systems*, 42(2):508–526.

Prinzie, A. and Van den Poel, D. (2006b). Investigating purchasing-sequence patterns for financial services using markov, MTD and MTDg models. *European Journal of Operational Research*, 170(3):710–734.

Prinzie, A. and Van den Poel, D. (2007). Predicting home-appliance acquisition sequences: Markov/Markov for discrimination and survival analysis for modeling sequential information in NPTB models. *Decision Support Systems*, 44(1):28–45.

Redondo, Y. P. and Fierro, J. J. C. (2005). Moderating effect of type of product exchanged in long-term orientation of firm-supplier relationships: an empirical study. *Journal of Product & Brand Management*, 14(7):424–437.

Reichheld, F. F. and Sasser, W. E. (1990). Zero defections: Quality comes to services. *Harvard Business Review*, 68(5):105–111.

Reichheld, F. F. and Teal, T. (1996). *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*. McGraw-Hill, Boston.

Rissanen, J. (1983). A universal data compression system. *IEEE Transactions on Information Theory*, 29(5):656– 664.

Rosenberg, L. J. and Czepiel, J. A. (1984). A marketing approach for customer retention. *Journal of Consumer Marketing*, 1(2):45–51.

Rosset, S., Neumann, E., Eick, U., Vatnik, N., and Idan, I. (2001). Evaluation of prediction models for marketing campaigns. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01, page 456461, New York, NY, USA. ACM.

Rud, O. P. (2000). *Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management*. Wiley, New York, 1 edition.

Stum, D. L. and Thiry, A. (1991). Building customer loyalty. *Training & Development Journal*, 45(4):34.

Tsai, C. and Lu, Y. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10):12547–12553.

Van den Poel, D. and Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research*, 166(2):557–575.

Verbeke, W., Martens, D., Mues, C., and Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3):2354–2364.

Xie, Y., Li, X., Ngai, E., and Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3, Part 1):5445–5449.

Zeithaml, V., Berry, L., and Parasuraman, A. (1996). The behavioral consequences of service quality. *The Journal of Marketing*, 60(2):31–46.