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

Predicting Partial Customer Churn Using Markov for Discrimination for Modeling First Purchase Sequences

Vera L. Miguéis[1]

Dirk Van den Poel[2]

A.S. Camanho[3]

Joao Falcao e Cunha[4]

August 2012

2012/806

[1] Faculty of Engineering, University of Porto
[2] Faculty of Economics and Business Administration, Department of Marketing, Ghent University
http://www.crm.UGent.be
[3] Faculty of Engineering, University of Porto
[4] Faculty of Engineering, University of Porto

D/2012/7012/39

Predicting partial customer churn using Markov for Discrimination for modeling first 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

Currently, in order to remain competitive companies are adopting customer centered strategies and consequently customer relationship management (CRM) is gaining increasing importance. In this context, customer retention deserves particular attention. This paper proposes a model for partial churn detection in the retail grocery sector that includes as a predictor the similarity of the products' first purchase sequence with churner and non-churner sequences. The sequence of first purchase events is modeled using Markov for discrimination. Two classification techniques are used in the empirical study: logistic regression and random forests. A real sample of approximately 95.000 new customers is analyzed taken from the data warehouse of a European retailing company. The empirical results reveal the relevance of the inclusion of a products' sequence likelihood in partial churn prediction models, as well as the supremacy of logistic regression when compared with random forests.

Key words: Customer relationship management, Churn analysis, Retailing, Classification, Logistic regression, Random forests

1 Introduction

Nowadays, due to the intense competition between companies and the changes in lifestyle, customer relationships with companies are changing and, in some cases, are becoming vulnerable. This fact, combined with the boom of data observed in recent decades, has made companies focus on customer relationship management. This is supported by the information extracted from the data that companies collect through each interaction with their customers. Indeed, it has been concluded that valuable knowledge can be obtained from the analysis of POS data and RFID data (e.g. Nakahara and Yada, 2012).

Ngai et al. (2009) summarizes customer relationship management as the combination of four dimensions: customer identification, customer attraction, customer development and customer retention. Customer retention is receiving particular attention from companies since customer life cycles are becoming shorter than in the past. Some customers present switching behavior in their purchases (Peterson, 1995) and others split their purchases between several competitors

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

(Dwyer, 1989). Particularly in non-contractual settings, such as the retail sector, this tendency is of utmost relevance, since customers do not have to inform companies about their churn intention and experience very little switching cost. According to EFMI and CBL (2005), a total of 87% of grocery shoppers use two or more different supermarkets for their grocery shopping and, on average, those grocery shoppers visit 2.8 different supermarkets each month.

Companies' concern about customer attrition is based on the benefits associated with retaining customers. Previous research suggests that retaining customers costs less than attracting new customers (Dick and Basu, 1994; Saren and Tzokas, 1998). Moreover, when the relationship between companies and customers is strong, these customers tend to be less sensitive to competitors' actions (Strandvik and Liljander, 1994) and prices (Kotler, 1999), and are less costly to serve (Bejou et al., 1998). These customers present an important role in the word-of-mouth (WOM) process (Martin et al., 1995) and show high purchase levels (Kamakura et al., 1991). In this context, a small improvement in customer retention can mean a significant increase in profit (Reichheld and Sasser Jr., 1990; Larivire and Van den Poel, 2005). Despite all the advantages associated with retaining customers, churn analysis in retailing can still be considered incipient (Buckinx and Van den Poel, 2005).

The framework presented by Ngai et al. (2009) shows the data mining tools that usually assist each CRM dimension. Classification tools are those which are more frequently used to deal with customer churn prediction. Several classification techniques have been applied in this setting in order to distinguish between churners and non-churners. Examples of these techniques include neural networks, decision trees and logistic regression. Besides the technique used, churn problems presented in the literature also differ in terms of the explanatory variables considered. Buckinx and Van den Poel (2005) classifies defection explanatory variables, or predictors, using three categories: behavioral antecedents, demographics and perceptions. This paper seeks to introduce a new churn predictor which measures the similarity of the product sequence first purchased with churner and non-churner sequences. These sequences of first purchase events are modeled as a Markov process. Moreover, unlike most of the previous churn models proposed in the literature, this paper proposes a churn prediction model which identifies which customers are going to partially leave the company. Furthermore, in the empirical study we compare the predictive performance of both logistic regression and random forests.

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 approach followed in this paper, identifying the explanatory variables, the classification techniques and the performance evaluation criteria used. Section 4 presents the application, i.e. the company used as a case study and the results. The paper closes with conclusions and some issues for future research.

2 Churn prediction modeling

Customer retention, and more specifically customer churn, has been widely discussed in the literature in the last decade. For an overview on churn prediction see Verbeke et al. (2011). Customer churn prediction models aim to detect customers which are easily persuaded to discontinue the relationship with the company. An accurate identification of these potential churners allows companies to target them in retention marketing campaigns. This topic has been studied in several domains and, in most cases, it is treated as total defection. In banking (e.g.

Kumar and Ravi, 2008; Larivire and Van den Poel, 2005) and insurance (e.g. Zeithaml et al., 1996; Morik and Kpcke, 2004)) churn is usually seen as account closure. In telecommunications (e.g. Hwang et al., 2004; Hung et al., 2006) it is usually seen as changing phone operator. In retailing, to the best of our knowledge, only Buckinx and Van den Poel (2005) and Burez and Van den Poel (2009) have analyzed churn. In both cases, churn was treated as partial churn since typically customers defect from companies progressively, rather than in an abrupt discontinuation. Buckinx and Van den Poel (2005) consider that in the long run partial churn may result in total defection.

Churn prediction problems may be decomposed primarily into the choice of the churn prediction techniques to be used and the definition of the churn prediction model. The model requires the identification of the explanatory variables which are relevant for the churn propensity. Furthermore, it includes the definition of the causality/link between these variables and the churn.

A wide diversity of data mining classification techniques have been used as churn prediction techniques (for an overview, see Verbeke et al., 2011). Neural networks proposed by Mcculloch and Pitts (1943) have frequently been used in this context (e.g. Hung et al., 2006; Hwang et al., 2004). Neural networks are analytical tools, inspired by the neural aspect of the human brain, which use simple processing units, linked to each other through weighted connections, to "learn" the relationships between variables. Despite usually presenting good performance, these tools are criticized for the fact they do not present the patterns underlying the data, being characterized as black boxes (Paruelo and Tomasel, 1997). Survival analysis (Kalbfleisch and Prentice, 1980) is a group of statistical techniques, also used in this context, that are concerned with the occurrence of a certain event, i.e. churning (as in this case). The main advantage of the use of survival analysis is its ability to estimate the occurrence time, despite requiring a long data series. Applications of these techniques include Larivire and Van den Poel (2004) and Mavri and Ioannou (2008). Decision trees, first introduced by Quinlan (1992), are frequently used in churn prediction by inducing a tree and subsequently extracting the rules that can be used to identify the defectors (e.g. Wei and Chiu, 2002; Hung et al., 2006). Rule inference is considered one of the advantages of this technique, while the lack of robustness and suboptimal performance are highlighted as disadvantages (Murthy, 1997). In order to deal with decision tree disadvantages, random forests (Breiman, 2001) have become popular (e.g. Buckinx and Van den Poel, 2005; Coussement and Van den Poel, 2008). Their classification is based on an ensemble of trees, avoiding misclassifications due to the weak robustness and sub-optimality of a single decision tree. The random forests technique allows a measure of the importance of each variable for the classification to be obtained. This technique is considered easy to use and provides robust results (Buckinx and Van den Poel, 2005). Despite having been extensively studied, no general consensus exists on the relative performance of churn prediction techniques. There are studies in which one technique outperforms the other and vice versa (Verbeke et al., 2011).

Concerning churn prediction variables, churn prediction models presented in the literature also differ considerably (see Buckinx and Van den Poel, 2005, for an overview). Perception variables are used in some studies and try to measure the way a customer appreciates the service/product of the company. They can be measured through customer surveys and include dimensions such as overall satisfaction, quality of service, locational convenience and reputation of the company. Most studies focus on demographic predictors, such as age, gender, education, social status and geographical data. A considerable number of prior studies also include behavioral antecedent variables. The number of purchases (frequency) and the amount of money spent (monetary value) are the most popular behavioral variables. Buckinx and Van den Poel (2005) concludes that, in addition to these two variables, recency is also part of the best-predictor group of variables. Despite having been disregarded in the literature, one behavioral dimension that seems to have huge potential concerning customer attrition detection is the purchase sequence of products. As stated by Grover and Vriens (2006), customers seem to follow purchasing patterns similar to other customers, by observing the purchasing behavior of other customers or due to the word-of-mouth effects (Bikhchandani et al., 1992, 1998) resulting from communication with other customers. As a result, one customer can follow a similar sequence to past customers, allowing companies to model their behavior (e.g. Prinzie and Van den Poel, 2006b,a). In particular, the first category purchase sequence expresses the development of the relationship of trust between a customer and the company (see Migueis et al., 2012). New customers have, by definition, 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 (Novemsky and Dhar, 2005). By doing so, customers break down the purchase process into portions which can take a share of the risk. Usually, people only take risks with later goals if the earlier goals in the sequence were accomplished successfully (Dhar and Novemsky, 2002). Therefore, we consider that the level of similarity between the first product category purchase sequence, chosen by a new customer, and the sequence chosen by churners and non-churners may be an indicator of the readiness of customers to churn or not. Thus, in this paper we hypothesize that first product category purchase sequences may support churner and non-churner discrimination.

3 Predicting partial churn

The methodology followed in this paper seeks to introduce and explore the predictive power of the likelihood of the first product category purchase sequence made by a new customer for the purchasing sequence made by customers who churned and customers who did not churn. The similarity measures are obtained by using Markov for discrimination. The value of the proposed predictor is assessed by comparing the accuracy obtained by the model including the similarity variables with the accuracy obtained by the model excluding these variables. Two data mining classification techniques are used and compared in the empirical study: logistic regression and random forests.

In the next paragraphs we present the criterion used to infer partial attrition, the explanatory variables used in the model, the classification techniques used and the evaluation criteria used to compare their performance.

3.1 Partial churning

Since the model proposed in this paper is intended to identify customers who may partially switch their purchases to another company and since in non-contractual businesses the defection is not explicit, we have to derive the dependent variable of the models. For this purpose, we first grouped the purchases in periods of three months. Then, we classified as churners those customers who, from a certain period, made no further transactions or those customers who in all subsequent periods spent less than 40% of the amount spent in the reference period. The granularity of the analysis and the amount spent threshold used was the result of a sensitivity analysis. We verified the impact of a higher/lower temporal aggregation and the impact of the variation of the threshold on the proportion of partial churners identified. The proportion which was considered as realistic by the domain expert from the company used for the case study dictated these parameters.

Figure 1 represents the amount spent by two customers over five quarters.

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

The first case represents a customer who partially churned. Notably, when the 2nd quarter is considered as the reference quarter, we observe that in all subsequent quarters this customer spent less than 40% of the amount spent in the reference quarter (100€×40%=40€). Therefore, we assume that this customer partially churned in the beginning of the 3rd quarter. Concerning the second case, there is no quarter in which the amount spent in all subsequent quarters is less than 40% of the amount spent in the reference quarter. Thus, this represents a customer who did not churn.

3.2 Explanatory variables

In this study we explore the potential of past behavior variables to distinguish churners from non-churners. Seven variables are included in the churn model proposed: recency, frequency and monetary – collectively known as RFM – and increased tendency, decreased tendency, churner product sequence likelihood and non-churner product sequence likelihood.

3.2.1 Recency

In this study, recency represents the number of days between the end of the period of analysis and the date of the last transaction. For model training purposes, the recency of churners is the number of days between the date in which those customers were classified as partial churners and the date of the previous transaction.

3.2.2 Frequency

Frequency is included in the model proposed as the average number of transactions per quarter. Regarding customers who churned partially, for model training purposes, frequency is naturally calculated by considering only the transactions observed up to the date in which those customers were classified as partial churners.

3.2.3 Monetary

This RFM dimension is covered in this study by the average value spent by customers in each quarter. As regards the partial churners used for training the model, this variable only takes into account the amount spent up to the date they were classified as partial churners.

3.2.4 Increased or decreased tendency

The model proposed includes two dummy variables that capture the recent changes in customers' behavior. Therefore, if a customer increased the amount of money spent in the last quarter of the period analyzed in relation to the amount spent in the previous quarter, the variable *increased tendency* takes the value 1, while the variable *decreased tendency* takes the value 0. The opposite happens if a decrease in the amount spent is observed. Concerning the partial churners used to train the model, these two variables were defined by analyzing the tendency in the two quarters before churning.

3.2.5 Product sequence likelihood

We assume that the likelihood of the sequence of first products' category purchased compared with the same sequence recorded for churners and non-churners can reveal whether a customer is about to churn or not. We propose a measure of this similarity by using Markov for discrimination introduced by Durbin et al. (1998) and used, for example, by Prinzie and Van den Poel (2007) in a predictive purchase sequences context.

Consider a process whose states are defined by a discrete variable $X(t), (t = 0, 1, 2, ...)$ according to a stochastic process. The process can be considered a Markov process if:

$$P[X_t = a|X_{t-1} = b, X_{t-2} = c, ..., X_0 = d] = P[X_t = a|X_{t-1} = b] \tag{1}$$

In a Markov model, the probability of X_t taking a certain value depends only on the value of X_{t-1}. Each Markov process can be represented by means of a transition matrix. In the case of a process with N possible states, the transition matrix is a NxN matrix defined as:

$$M = [p_{ij}] \tag{2}$$

Each element of the matrix represents the probability of the system evolving from a state i, in period t, to another state j, in period $t + 1$. In this paper we consider the state variable X_t to be the product category that a customer purchases at the t-th store visit. Product category corresponds to the business unit level of the products' hierarchy defined by the company used as a case study (see Section 4 for details). Moreover, we assume that the sequences in each population, i.e. churners and non-churners, are generated by a specific Markov process for each population. Therefore, we build for each population a different transition matrix that reflects specific sequences of product category purchases. Following Durbin et al. (1998), we use these Markov transition matrices to calculate the log-odds ratio between the odds of observing sequence x given it originates from the non-churners' population and the odds of observing sequence x given it belongs to the churners' population:

$$S(x) = log \frac{P(x|\text{non-churners})}{P(x|\text{churners})} \tag{3}$$

This ratio $S(x)$ allows the affinity of a customer to be measured with respect to non-churners and churners, by means of their specific product category purchase sequence. A positive ratio indicates that the customer is not likely to churn while a negative ratio means the opposite.

For example, consider the purchase sequence of three products. Since we are interested in discriminating between churners and non-churners, we construct a transaction matrix for each population. Each matrix contains, for each population the probability of a customer buying for the first time from product category P_i, in period t, and then buying for the first time product category P_j, in period $t+1$. Consider the following transition matrices:

$$M_{non-churners} = \hookleftarrow \begin{array}{cc} & \begin{array}{ccc} & \text{t+1} & \\ P_1 & P_2 & P_3 \end{array} \\ \begin{array}{c} P_1 \\ P_2 \\ P_3 \end{array} & \begin{pmatrix} 0 & \mathbf{0.2} & 0.8 \\ 0.7 & 0 & 0.3 \\ \mathbf{0.4} & 0.6 & 0 \end{pmatrix} \end{array} \quad M_{churners} = \hookleftarrow \begin{array}{cc} & \begin{array}{ccc} & \text{t+1} & \\ P_1 & P_2 & P_3 \end{array} \\ \begin{array}{c} P_1 \\ P_2 \\ P_3 \end{array} & \begin{pmatrix} 0 & \mathbf{0.6} & 0.4 \\ 0.8 & 0 & 0.2 \\ \mathbf{0.7} & 0.3 & 0 \end{pmatrix} \end{array} \tag{4}$$

The log-odds ratio of a customer whose first purchase sequence is $P_3 \longrightarrow P_1 \longrightarrow P_2$ can be calculated as follows:

$$\begin{aligned} S(P_3 \longrightarrow P_1 \longrightarrow P_2) &= log\frac{0.4}{0.7} + log\frac{0.2}{0.6} \\ &= -0.7 \end{aligned} \tag{5}$$

The odds that the sequence stems from the non-churners' population is 0.7 times smaller than the odds that he sequence stems from the churners' population (see Equation (3)). Therefore, this hypothetical customer is likely to churn.

In this study, we transform $S(x)$ in two variables, i.e. non-churner likelihood and churner likelihood. As a result, positive log-odds ratios are assigned to the non-churner likelihood variable, while negative log-odds ratios are assigned to the churner likelihood variable.

3.3 Techniques

3.3.1 Logistic regression

Logistic regression is a well-known technique which can be used for classification purposes. This is a form of regression which is generally used when the dependent variable is a dichotomy. The independent variables can be of any type (Agresti, 1996). The relationship between the dependent variable and the independent variables is not assumed to be linear in a logistic regression. In fact, this technique assumes that the independent variable is linearly related to the logit of the dependent variables. The logistic regression technique does not require normally distributed variables. It is easy to use and provides quick and robust results (Buckinx and Van den Poel, 2005).

3.3.2 Random forests

As introduced in Section 2, decision trees are also very popular techniques for classification, mainly due to their ease of use and interpretability. This technique is variable scale independent and can deal with nominal variables. Nevertheless, conventional decision trees present several disadvantages, such as instability (Breiman, 1996) and suboptimal performance (Dudoit et al., 2002). To overcome these disadvantages, researchers have introduced new decision tree methods which optimize the decision trees. As mentioned in Section 2, the random forests technique is one of the most popular techniques that derive from decision trees. This is an ensemble of classification trees followed by a vote for the most popular class, i.e. labeled forests (Breiman, 2001). In this paper we ensemble the trees as proposed by Breiman (2001), such that, each tree is grown by using a subset of randomly restricted and selected independent variables. This approach implies the definition of two parameters, i.e. the number of trees to be used and the number of variables to be randomly selected from the set of independent variables. We follow the recommendations reported by Breiman (2001) and consequently we consider a large number of trees to be used (i.e. 1000 trees) and the square root of the number of variables as the number of independent variables selected (i.e. 3 variables).

3.4 Evaluation criteria

In order to measure the performance of the prediction model proposed, we compute the well known receiver operating characteristic curve (ROC) and we analyze the area under curve (AUC). The AUC measure is based on comparisons between the observed status and the predicted status. The status is predicted by considering all cut off levels for the predicted values. An AUC close to 1.0 means that the model has perfect discrimination, while an AUC close to 0.5 suggests poor discrimination (Hanley and McNeil, 1982). Moreover, we use the percentage of cases correctly classified (PCC), also known as accuracy, as an evaluation metric. All cases having a default probability above a certain threshold are classified as churners and all cases having a below threshold default probability are classified as non-defaulters. Then, PCC is defined as the ratio between the number of correctly classified cases and the total number of cases to be classified. According to Morrison (1969), considering α as the actual proportion of the class to be predicted, i.e. partial churners, PCC should exceed the proportional change criterion defined as:

$$\alpha^2 + (1-\alpha)^2 \tag{6}$$

In order to calculate the performance measures of the model proposed, the dataset is split with 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 Case study

In this section we present the company used as a case study and the prediction model results obtained for this case study.

4.1 Company

This study uses as a case study a European food-based retailing company selling through 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 city where they are located.

The establishment of loyalty relationships with customers became a main strategic goal for this company. The development of the company's information system and the implementation of a loyalty program have enabled the collection of data on each customer profile and on their transactions. This program is supported essentially by a loyalty card, and currently, approximately 80% of the total number of transactions are made by customers using the loyalty card.

At present, the company's customers are segmented in two ways. One division 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 internally referred to as the "frequency and monetary value" (FM) model. According to the values of these two variables, the company specifies 8 groups of customers. Each client is attributed to one of these groups, according to the average number of purchases made 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 those groups with relatively 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 together. Both methods of customer segmentation do not account for churn prediction. In fact, only the first method presented addresses this topic, but in a reactive way, i.e. after the churn event has occurred.

The analysis reported in this paper is based on transactional data of customers with a loyalty card. The database used includes the records of the 581.002 customers who have shopped in 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 new customers, since it is only for those customers that it is possible to trace the first category purchase sequence. New customers are considered to be those who made no purchase in the first half of 2009 but spent at least 100 euros up to the end of 2010. This corresponds to a total of 95.147 new customers. From these, according to the criterion presented in Section 3.1 to identify partial churners, 49% partially churned during the period of analysis, while the remaining stayed active.

The company classifies its products on several levels, such as: the department, the business unit, the category and the subcategory. This classification includes 20 departments (e.g. Food, Textile), 65 business units (e.g. Grocery, Baby-textile), 247 categories (e.g. Milk/Soy drinks, Baby-shoes) and in 1013 subcategories (e.g. Pasteurized milk, Baby-MNO). For product sequence analysis, we focus on the business unit level since this allows the incorporation of discrimination between products and avoids a large degree of complexity. Moreover, we only select for analysis the most representative business units, concerning the purchases made by new customers, i.e. those business units which were represented in the shopping baskets of at least 10% of the new customers considered in the analysis. As a result, 20 business units are considered, specifically grocery, hygiene/cleanliness, daily/frozen, drinks, fruits/vegetables, bakery, delicatessen, culture, house, meat, fish, takeaway, bricolage/auto, leisure, storage, pets/plants, women-textile, men-textile, baby-textile and child-textile. In order to respect the terminology used in marketing, these business units are referred to as categories in this paper.

4.2 Results

The transition matrices for both the non-churner and churner populations are shown in the Appendix. From the analysis of these matrices we can conclude that most of the non-churners and churners first purchase sequences include combinations between the following categories: grocery, hygiene/cleanliness, daily/frozen, drinks, fruits/vegetables, bakery, delicatessen and culture. The least common category transitions are those which include any category as an antecedent and men-textile, baby-textile and child-textile as subsequent categories.

Both matrices exhibit different probabilities of transition between categories. Note that the probability of buying hygiene/cleanliness products after having bought a child-textile product, and the probability of buying a daily/frozen product after having bought a men-textile product, are considerably higher for churners than for non-churners. Moreover, the probability of buying a product included in pet/plant category after having bought a storage product is higher in the case of the non-churners than in the case of churners.

Consider a real customer whose sequence of first purchase is: Grocery $\longrightarrow$ Women-textile $\longrightarrow$ Hygiene/cleanliness $\longrightarrow$ Fruits/vegetables $\longrightarrow$ Pets/plants *and* House. Since the first products from Pets/plants and House categories were bought on the same date, we computed two product sequence likelihood measures:

$$\begin{aligned} S(Grocery \longrightarrow Women-textile \longrightarrow & Hygiene/cleanliness \longrightarrow Fruits/vegetables \longrightarrow Pets/plants) = \\ &= log\frac{2.1\%}{2.3\%} + log\frac{6.0\%}{7.1\%} + log\frac{8.2\%}{8.2\%} + log\frac{4.1\%}{3.5\%} \\ &= -0.04 \end{aligned} \tag{7}$$

$$\begin{aligned} S(Grocery \longrightarrow Women-textile \longrightarrow & Hygiene/cleanliness \longrightarrow Fruits/vegetables \longrightarrow House) = \\ &= log\frac{2.1\%}{2.3\%} + log\frac{6.0\%}{7.1\%} + log\frac{8.2\%}{8.2\%} + log\frac{6.8\%}{7.1\%} \\ &= -0.13 \end{aligned} \tag{8}$$

Both sequences reveal that this customer is likely to churn. However, we only consider the maximum log-odds ratio in order to avoid an overestimation of the churner likelihood based on a non-unique product sequence.

To evaluate the impact of the inclusion of the sequence likelihood variables in the partial churn prediction models, we compare the prediction performance of a model including these variables with the performance of the model including only the recency, frequency, monetary value and increased or decreased tendency. However, first we consider the possible multicollinearity issues. One popular multicollinearity metric is the variance inflation factor (VIF) (see Gujarati, 2002). A value of VIF greater than 10 means that multicollinearity may be causing problems with the estimations (Neter et al., 1996). VIF for the independent variables considered in this study ranges from 1 to 2. This is within the acceptable range and thus multicollinearity is not an important issue for the analysis.

Table 1 presents the performance results of the two models using both logistic regression and random forests. It is important to note that in a retailing context, companies are interested in concentrating their efforts to keep customers on a small group of customers, due to the high costs related to a retention marketing campaign. Consequently, the threshold used to compute the PCC for both classification techniques is the value that allows a percentage of partial churners of approximately 5% to be obtained in the case of the model including the sequence likelihood variables.

Table 1
Performance results.

	AUC	PCC
	Logistic regression	
Without sequence likelihood variables	87.22%	54.94%
With sequence likelihood variables	87.42%	56.41%
	Random forests	
Without sequence likelihood variables	80.78%	54.69%
With sequence likelihood variables	84.63%	56.31%

From the analysis of Table 1 we can conclude that partial churn prediction in this context is promising. The AUC values are high both when logistic regression and random forests are used. Moreover, the PCC values exceed the proportional chance criterion proposed by Morrison (1969) of 0.50 ($= 0.49^2 + 0.51^2$). The results obtained also show that logistic regression outperforms random forests in both models tested. When comparing the results reported, focusing on the relevance of the sequence likelihood variables proposed in this study, we can conclude that these variables have a positive effect on the prediction ability. Indeed, the different classification techniques present higher performance when these variables are incorporated into the models. The importance of these variables is particularly evident when the random forests technique is used, since the AUC increases by approximately 4 percentage points. For each technique considered, the test proposed by DeLong et al. (1988) shows that the AUC values obtained when the prediction model includes the proposed behavioral variables and the AUC values obtained when these variables are not included are statistically significantly different. The PCC is also higher (approximately 2%) when the prediction models include the sequence likelihood variables.

Since logistic regression is faster than random forests and since it leads to higher levels of performance, this classification technique can be of particular interest for marketeers. However, it is important to note that the more variables there are and the richer the data set, the greater the tendency of random forests to outperform logistic regression. In fact, even when there is a large number of predictors, random forests do not suffer from overfitting problems (see Breiman, 2001, for further discussion), which is not guaranteed when using logistic regression.

5 Conclusion and issues for future research

This paper proposes a model to predict partial customer churn in the retail sector. This model contributes to the literature by including a measure of the similarity of the sequence of customers' first purchases, in terms of product category, with the sequence recorded for churners and non-churners. This sequence likelihood is modeled by using Markov-for-discrimination. Both logistic regression and random forests are used in this study.

In order to test the proposed model in a real context, we used a European retail company as case study. This fact allowed us to test the model with a large dataset. For practitioners, the ability of prediction models to handle large datasets can condition the applicability of these models in real contexts.

The results reported highlight the relevance of the proposed model, since the performance of the models including the sequence likelihood variables is higher than the performance of the models not including these variables. This supremacy is measured in terms of the area under the receiver operating characteristic curve (AUC) and percentage of correctly classified instances (PCC). Furthermore, the results reported suggest that the logistic regression technique outperforms the random forests technique.

This study is limited to analyzing sequences of what people buy in physical stores. Future studies might combine this information with the information collected via a company's online store. Further studies might also combine customer perception information, collected by means of surveys and complaints analysis.

In this study we used two years of data to identify partial defectors among new customers. Whenever more data is available, we would identify more new customers and we would be able to evaluate the defective behavior over a longer time period. This would give the opportunity to check what subsequently happens to people classified as partial churners. Moreover, when more data are available we would be able to investigate how often the retention model should be updated in order to reduce the churn rate.

Identifying customers as partial churners is just a starting point for the managerial process of retaining these customers. Further analysis should be conducted in order to know how marketing investment should be employed and who should be the recipients. For example, it may be important to decide if it is worth it sending a promotional voucher or customized advertising to each customer identified as a partial churner.

6 Acknowledgements

The funding of this research through the scholarship SFRH/BD/60970/2009 from the Portuguese Foundation of Science and Technology (FCT) is gratefully acknowledged by the first author.

References

Agresti, A. (1996). *An introduction to categorical data analysis*. Wiley-Interscience, New York.

Bejou, D., Ennew, C. T., and Palmer, A. (1998). Trust, ethics and relationship satisfaction. *International Journal of Bank Marketing*, 16:170–175.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *The Journal of Political Economy*, 100(5):992–1026.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–70.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

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. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36:4626–4636.

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.

Dick, A. S. and Basu, K. (1994). Customer loyalty: Toward an integrated conceptual framework. *Journal of the Academy of Marketing Science*, 22:99–113.

Dudoit, S., Fridlyand, J., and Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the american statistical association*, 97(457):77–87.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge University Press, Cambridge U.K.

Dwyer, F. R. (1989). Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing*, 3(4):8–15.

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

Grover, R. and Vriens, M. (2006). *The handbook of marketing research: uses, misuses, and future advances*. Sage Publications, California.

Gujarati, D. (2002). *Basic econometrics*. McGraw-Hill/Irwin, New York.

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.

Hwang, H., Jung, T., and Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26(2):181–188.

Kalbfleisch, J. and Prentice, R. L. (1980). *The statistical analysis of failure time data*. Wiley, New York.

Kamakura, W., Ramaswami, S., and Srivastava, R. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, 8(4):329–349.

Kotler, P. (1999). *Marketing management: analysis, planning, implementation, and control*. Prentice Hall, New Jersey.

Kumar, D. A. and Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. *International Journal of Data Analysis Techniques and Strategies*, 1(1):4–28.

Larivire, B. and 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 Systems with Applications*, 27(2):277–285.

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.

Martin, C., Clark, M., Peck, H., and Payne, A. (1995). *Relationship Marketing for Competitive Advantage: Winning and Keeping Customers*. Butterworth-Heinemann, Oxford.

Mavri, M. and Ioannou, G. (2008). Customer switching behaviour in greek banking services using survival analysis. *Managerial Finance*, 34:186–197.

Mcculloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysic*, 5:133, 115.

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.

Migueis, V. L., Van den Poel, D., Camanho, A. S., and Falcao e Cunha, J. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, 39(12):11250–11256.

Morik, K. and Kpcke, H. (2004). Analysing customer churn in insurance data - a case study. In Boulicaut, J.-F., Esposito, F., Giannotti, F., and Pedreschi, D., editors, *Knowledge Discovery in Databases*, volume 3202 of *Lecture Notes in Computer Science*, pages 325–336, Italy. Springer.

Morrison, D. G. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, 6(2):156–163.

Murthy, S. K. (1997). Automatic construction of decision trees from data: A Multi-Disciplinary survey. *Data mining and knowledge discovery*, 2:345–389.

Nakahara, T. and Yada, K. (2012). Analyzing consumers' shopping behavior using RFID data and pattern mining. *Advances in Data Analysis and Classification - Special Issue on Data Analysis and Classification in Marketing*. (forthcoming).

Neter, J., Kutner, M., Wasserman, W., and Nachtsheim, C. (1996). *Applied Linear Statistical Models*. McGraw-Hill/Irwin, New York, 4 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.

Novemsky, N. and Dhar, R. (2005). Goal fulfillment and goal targets in sequential choice. *Journal of Consumer Research*, 32(3):396–404.

Paruelo, J. and Tomasel, F. (1997). Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecological Modelling*, 98(2-3):173–186.

Peterson, R. A. (1995). Relationship marketing and the consumer. *Journal of the Academy of*

Marketing Science, 23:278–281.

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.

Quinlan, J. R. (1992). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco.

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

Saren, M. J. and Tzokas, N. X. (1998). Some dangerous axioms of relationship marketing. *Journal of Strategic Marketing*, 6(3):187–196.

Strandvik, T. and Liljander, V. (1994). Relationship strength in bank services. *Relationship Marketing: Theory, Methods and Applications*, pages 356–359.

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.

Wei, C. and Chiu, I. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications*, 23(2):103–112.

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

APPENDIX

Table A1
Non-churners transition matrix.

	Groce.	Hygie.	Dai.	Drinks	Frui.	Bake.	Delic.	Cul.	House	Bu.	Fishe.	Take.	Brico.	Lei.	Stor.	Pets.	Women.	Men.	Baby.	Child.
Grocery	0.0%	8.6%	7.8%	8.3%	8.1%	8.4%	8.2%	7.0%	6.6%	5.9%	5.1%	4.3%	4.2%	4.1%	3.1%	2.8%	2.3%	1.9%	1.7%	1.7%
Hygiene/cleanli.	6.4%	0.0%	7.3%	8.3%	8.2%	8.4%	8.1%	7.2%	6.9%	6.1%	5.0%	4.6%	4.5%	4.3%	3.5%	3.1%	2.4%	2.0%	1.9%	1.7%
Daily/frozen	5.3%	8.0%	0.0%	8.1%	7.9%	8.4%	8.1%	7.4%	6.9%	6.4%	5.4%	4.9%	4.6%	4.3%	3.5%	3.2%	2.4%	1.9%	1.7%	1.7%
Drinks	4.9%	8.4%	7.0%	0.0%	8.2%	8.3%	8.1%	7.5%	6.9%	6.4%	5.5%	5.0%	4.8%	4.4%	3.5%	3.4%	2.4%	2.0%	1.8%	1.7%
Fruits/vegeta.	5.4%	8.3%	6.3%	7.7%	0.0%	8.2%	7.6%	7.3%	7.1%	6.8%	6.0%	5.2%	4.8%	4.3%	3.5%	3.5%	2.5%	2.0%	1.8%	1.7%
Bakery	5.5%	8.6%	6.9%	7.8%	7.7%	0.0%	7.5%	7.4%	7.1%	6.6%	5.6%	5.3%	4.6%	4.2%	3.6%	3.6%	2.5%	2.1%	1.8%	1.7%
Delicate.	4.5%	8.1%	5.6%	7.7%	6.8%	7.6%	0.0%	7.9%	7.7%	6.8%	6.0%	5.6%	5.2%	4.5%	3.8%	3.9%	2.6%	2.0%	1.8%	1.7%
Culture	7.6%	8.1%	8.0%	8.4%	7.3%	8.0%	7.5%	0.0%	6.7%	5.4%	4.2%	4.4%	4.4%	4.5%	3.6%	3.3%	2.7%	2.0%	1.8%	2.0%
House	5.5%	7.1%	7.2%	7.4%	7.4%	7.7%	7.4%	7.5%	0.0%	5.9%	5.0%	4.5%	5.1%	4.9%	4.3%	3.9%	2.9%	2.2%	2.2%	2.0%
Meat	4.1%	7.8%	5.4%	7.0%	6.1%	7.5%	6.8%	8.0%	7.6%	0.0%	6.7%	5.8%	5.3%	4.6%	4.2%	4.2%	2.7%	2.4%	1.9%	1.7%
Fish	4.9%	7.8%	5.9%	7.2%	5.0%	7.6%	7.0%	7.0%	7.4%	6.7%	0.0%	5.7%	5.5%	4.9%	4.2%	4.7%	2.6%	2.4%	1.8%	1.7%
Takeaway	4.3%	7.9%	6.2%	6.6%	7.3%	6.5%	6.4%	7.5%	7.4%	6.7%	5.9%	0.0%	5.1%	4.7%	4.3%	4.4%	3.0%	2.3%	1.8%	1.9%
Bricolage/auto	6.2%	5.9%	7.0%	7.5%	7.4%	7.8%	7.7%	7.0%	6.6%	5.7%	5.0%	4.5%	0.0%	4.7%	3.9%	4.2%	2.8%	2.5%	1.8%	1.6%
Leisure	7.3%	7.6%	7.8%	7.7%	7.1%	7.5%	7.1%	7.1%	6.5%	4.9%	4.5%	4.0%	4.5%	0.0%	3.5%	3.3%	2.7%	2.1%	2.3%	2.5%
Storage	5.8%	5.9%	6.9%	7.1%	6.8%	7.3%	7.4%	7.1%	6.3%	5.8%	4.7%	4.7%	5.4%	4.5%	0.0%	4.5%	3.3%	2.5%	1.9%	2.1%
Pets/plants	3.8%	4.9%	5.0%	6.8%	6.9%	6.9%	7.1%	7.9%	7.2%	6.4%	5.7%	5.9%	6.6%	4.2%	6.4%	0.0%	2.7%	2.3%	1.8%	1.6%
Women-textile	6.4%	7.1%	7.4%	7.5%	6.9%	7.4%	6.8%	6.7%	6.3%	5.0%	4.0%	4.5%	4.4%	4.8%	4.1%	3.5%	0.0%	2.9%	2.0%	2.3%
Men-textile	6.7%	7.2%	7.5%	7.2%	6.6%	7.2%	6.7%	6.3%	6.8%	4.8%	4.4%	4.1%	4.7%	4.4%	3.7%	3.7%	3.4%	0.0%	2.1%	2.4%
Baby-textile	7.8%	7.5%	7.5%	7.8%	6.6%	6.9%	6.6%	6.2%	6.5%	4.6%	3.7%	3.4%	4.1%	5.7%	3.6%	2.3%	3.5%	2.4%	0.0%	3.1%
Child-textile	7.5%	7.9%	7.2%	7.3%	6.3%	7.0%	6.0%	7.0%	6.2%	4.3%	3.5%	4.1%	3.8%	6.4%	3.5%	2.6%	3.8%	2.7%	2.8%	0.0%

Table A2
Churners transition matrix.

	Groce.	Hygie.	Dai.	Drinks	Frui.	Bake.	Delic.	Cul.	House	Bu.	Fishe.	Take.	Brico.	Lei.	Stor.	Pets.	Women.	Men.	Baby.	Child.
Grocery	0.0%	8.7%	7.6%	8.1%	8.0%	8.7%	8.5%	7.0%	6.4%	6.3%	5.0%	4.7%	4.3%	3.3%	3.2%	3.4%	2.1%	1.7%	1.6%	1.6%
Hygiene/cleanliness	6.7%	0.0%	7.5%	8.2%	8.2%	8.7%	8.4%	7.2%	6.9%	6.1%	5.0%	4.8%	4.6%	3.4%	3.6%	3.7%	2.3%	1.8%	1.6%	1.6%
Daily/frozen	5.6%	8.2%	0.0%	8.1%	7.9%	8.6%	8.3%	7.3%	6.8%	6.7%	5.4%	5.1%	4.5%	3.3%	3.4%	3.8%	2.2%	1.7%	1.6%	1.6%
Drinks	5.0%	8.2%	6.6%	0.0%	7.8%	8.2%	8.2%	7.7%	6.9%	6.7%	5.3%	5.4%	4.9%	3.5%	3.8%	4.1%	2.4%	1.9%	1.6%	1.6%
Fruits/vegetables	5.4%	8.1%	6.2%	7.8%	0.0%	8.4%	7.8%	7.4%	6.8%	7.0%	6.0%	5.2%	4.9%	3.5%	3.9%	4.1%	2.3%	1.9%	1.5%	1.6%
Bakery	5.4%	8.5%	6.9%	7.6%	7.5%	0.0%	7.8%	7.4%	6.9%	6.9%	5.7%	5.7%	5.0%	3.5%	4.0%	4.0%	2.3%	1.8%	1.6%	1.6%
Delicatessen	4.6%	8.0%	5.5%	7.4%	7.0%	7.4%	0.0%	7.9%	7.3%	7.4%	6.2%	6.0%	5.5%	3.7%	4.2%	4.4%	2.5%	1.9%	1.6%	1.5%
Culture	6.6%	7.5%	7.4%	7.5%	7.1%	8.0%	7.6%	0.0%	6.7%	5.7%	4.6%	5.3%	5.0%	4.0%	4.3%	4.3%	2.8%	2.1%	1.6%	2.1%
House	5.3%	6.2%	6.5%	7.2%	6.9%	7.5%	7.2%	7.7%	0.0%	6.1%	5.1%	5.3%	5.3%	4.1%	5.3%	4.9%	3.2%	2.4%	1.7%	2.0%
Meat	3.8%	7.4%	5.1%	6.9%	6.0%	7.1%	6.7%	8.0%	7.7%	0.0%	7.1%	6.2%	5.6%	4.1%	4.8%	5.2%	2.6%	2.3%	1.7%	1.7%
Fish	4.5%	6.8%	5.1%	6.6%	4.5%	7.2%	6.5%	7.8%	7.8%	7.5%	0.0%	6.0%	6.1%	4.7%	5.1%	5.2%	2.7%	2.4%	1.8%	1.6%
Takeaway	4.1%	7.4%	5.4%	6.2%	6.9%	6.0%	6.2%	8.1%	7.5%	6.9%	6.2%	0.0%	5.7%	4.2%	5.1%	5.4%	3.1%	2.3%	1.6%	1.8%
Bricolage/auto	5.2%	5.4%	5.9%	7.0%	6.9%	7.4%	7.1%	7.6%	6.9%	6.2%	5.4%	5.4%	0.0%	4.3%	5.2%	5.3%	2.7%	2.3%	1.6%	1.8%
Leisure	6.4%	6.4%	7.0%	7.1%	6.6%	7.5%	7.1%	7.2%	6.5%	5.2%	4.6%	4.3%	5.1%	0.0%	4.3%	4.4%	3.0%	2.4%	2.1%	2.8%
Storage	5.0%	4.7%	5.8%	6.0%	6.4%	6.7%	6.8%	7.3%	7.0%	5.8%	5.6%	6.1%	5.9%	4.4%	0.0%	6.1%	3.6%	2.8%	1.9%	1.9%
Pets/plants	4.0%	4.8%	5.4%	6.1%	6.3%	6.5%	6.6%	8.1%	7.1%	6.4%	6.1%	6.1%	6.5%	4.2%	6.6%	0.0%	3.1%	2.6%	1.7%	1.9%
Women-textile	5.4%	6.0%	6.7%	6.4%	6.6%	7.2%	6.7%	6.9%	6.6%	5.2%	4.7%	5.1%	5.0%	4.5%	4.6%	4.5%	0.0%	3.2%	2.1%	2.7%
Men-textile	5.8%	6.4%	6.1%	6.3%	6.2%	6.8%	6.0%	6.7%	6.6%	5.3%	5.1%	4.6%	5.4%	4.3%	4.7%	4.7%	4.2%	0.0%	2.3%	2.5%
Baby-textile	6.9%	7.0%	7.2%	6.7%	6.4%	6.6%	6.3%	6.5%	6.7%	5.0%	4.1%	4.3%	4.3%	5.5%	4.0%	3.0%	3.8%	2.7%	0.0%	3.1%
Child-textile	6.4%	6.3%	6.5%	6.4%	6.0%	7.3%	6.3%	7.5%	6.4%	4.9%	4.4%	4.6%	3.8%	5.6%	4.1%	3.0%	4.7%	3.0%	2.6%	0.0%