open
proceedings

Understanding Customer Attrition at an Individual Level: a New Model in Grocery Retail Context

Clément Gautrais
University of Rennes 1
clement.gautrais@irisa.fr

Peggy Cellier
INSA Rennes
peggy.cellier@irisa.fr

Thomas Guyet
Agrocampus Ouest
thomas.guyet@irisa.fr

René Quiniou
Inria Rennes
rene.quiniou@inria.fr

Alexandre Termier
University of Rennes 1
alexandre.termier@irisa.fr

ABSTRACT

This paper presents a new model to detect and explain customer defection in a grocery retail context. This new model analyzes the evolution of each customer basket content. It therefore provides actionable knowledge for the retailer at an individual scale. In addition, this model is able to identify customers that are likely to defect in the future months.

CCS Concepts

•Information systems → *Data mining;*

Keywords

Data mining; Attrition modeling; Grocery retail

1. INTRODUCTION

In grocery retail context, customer defection is partial [1], in the sense that a customer will usually lower his purchases, instead of totally leaving the store. Moreover, no contract binds the customer to the retailer, so customer defection is not clearly signaled through contract ending, like it is in other businesses such as banks or phone operators. Customer defection (also called attrition) is therefore difficult to detect because there is no clear definition of when a customer is defecting.

Attrition in grocery retail environments has mainly been studied through the RFM model [3], based on behavioral variables (recency, frequency and monetary value). RFM demonstrated good performances in partial attrition detection [1] but is limited to building groups of customers which provide few explanations of attrition causes.

Building comprehensive attrition models is of interest because retailers want to lower their retention marketing expenses, by deploying accurate targeted marketing. Models using first and last sequences of purchased products have been proposed [2] and improved attrition detection, while providing more information about attrition causes. Nevertheless, these models do not explain attrition at an individual level.

Understanding attrition at the customer level is necessary to do efficient targeted marketing. This work presents a model of customer stability that allows for analyzing the evolution of individual customer's purchases to understand attrition causes, at an individual level.

2. ATTRITION MODEL

We want to characterize important items, that are bought by a given customer during successive periods. Moreover, we want to detect an evolution in this item set to model and understand customer stability.

Let $\mathcal{I} = \{i_1...i_n\}$ be the set of all items. The purchases of customer i can be represented by a chronologically ordered list $\mathcal{D}_i = \langle (b_1, t_1), \ldots, (b_N, t_N) \rangle$, with $b_j \subset \mathcal{I}$ being the content of basket j and t_j its timestamp.

Let w be a window. We divide $\mathcal{D}_i$ in consecutive non overlapping windows of time span w to define the windowed database of customer i, denoted $\mathcal{D}_i^w$, as an ordered list of tuples (t_k^B, t_k^E, u_k). k is the window number and $\mathcal{D}_i^w$ is ordered in chronological order on t_k^B. u_k is the set of all products bought during window k, delimited by t_k^B and t_k^E.

Let $p \in \mathcal{I}$ be an item, $c(k)$ be the number of windows prior to window k that contain p, $c(k) = |\{u_v|v < k, p \in u_v\}|$ and $l(k)$ be the number of windows prior to window k that do not contain p, $l(k) = |\{u_v|v < k, p \notin u_v\}|$.

We define the **significance** of p in window k as $S(p,k) = \alpha^{c(k)-l(k)}$ if $c(k) > 0$ and $S(p,k) = 0$ otherwise. α is a parameter of the method. The usual expected behavior is to increase the item significance when incrementing $c(k)$. Therefore, we generally fix $\alpha > 1$.

We define the **stability of customer** i in window k as $Stability_i^k = \sum\limits_{p \in u_k} S(p,k) / \sum\limits_{p \in \mathcal{I}} S(p,k)$.

If all products are contained in window k, the stability of the customer is equal to 1. This stability decreases when products are not contained in window k. This decrease is proportional to the significance of missing products. The more significant a product is, the more the stability will decrease if this product is not present in window k.

When the stability of some customer decreases, we can identify which product mainly caused this decrease. This product is defined as $\arg\max_{p \notin u_k} S(p,k)$, which is the most significant product that was not bought in window k. This
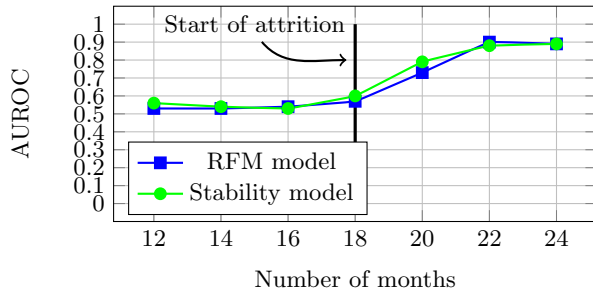
Figure 1: Performance of the attrition detection.

Figure 2: Defecting customer stability value example.

attrition explanation can be easily extended to a set of products.

3. EXPERIMENTS

The dataset provided by a major French retailer contains anonymized receipts of 6 millions customers, from May 2012 to August 2014. Each timestamped customer receipt describes a related basket content. A taxonomy is also provided that enables abstracting products in segments. The dataset contains 4 millions products, that are grouped into 3 388 segments.

Customer selection: Our main goal is to explain attrition. It is especially important for defecting loyal customers. The retailer provided us with the IDs of loyal customers, and of loyal customers that defected in the last 6 months. The beginning of the defection, given by the retailer, is indicated in Figure 1 by the vertical line at month 18.

3.1 Attrition prediction

The first experiment attempts to validate our model for separating loyal customers from the ones that defected during the last 6 months. To assess the relevance of our model, we use the area under the ROC curve for different window indices. We chose the AUROC because it evaluates the discrimination ability of our model. The points on these curves are obtained using different thresholds β for the customer stability. If $Stability_i^k > \beta$ the customer is considered loyal. Otherwise, the customer is considered as defecting on window k. The window length for this experiment is set to two months and the α parameter is set to 2. These values were chosen after performing a 5-fold cross-validation search. We compare our model to the standard RFM model, that uses recency, frequency and monetary variables to identify defecting customers. This RFM model is built using a logistic regression on these three types of variables. The methodology we used to compute the RFM model is similar to the one presented in [1], but we only used predictors associated to the recency, frequency and monetary variables. For each window, we compute the AUROC of our model and of the RFM model. Their evolution is plotted in Figure 1.

Figure 1 shows that our model accurately identifies customers that are in attrition in the last 6 months. This identification takes place in the first months of the customer defection. Two months after the start of attrition, our model scores an AUROC of 0.79, indicating a rather accurate detection of defecting customers. Our model and the RFM model have similar performances. This shows that our model is not only able to provide information about attrition, but is also able to detect customer attrition, with performances similar to standard attrition models such as RFM.
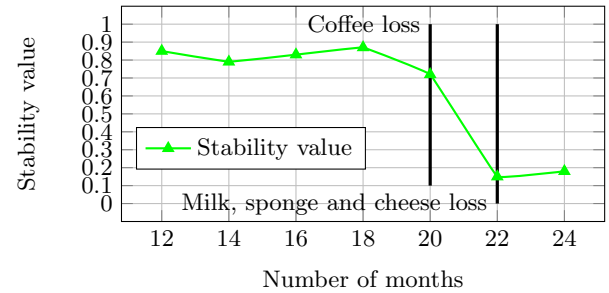
3.2 Attrition explanation

The second experiment aims to show the value of our model to explain attrition at an individual level, by illustrating that it provides actionable knowledge about the products responsible for individual customer attrition. We illustrate this on a use-case study of a defecting customer.

In Figure 2 the stability value indicates that the customer is loyal in the first months, and defecting starting from month 20. The upside of the stability value is that it can link each decrease to a loss of significant products.

For example, we can link the decrease in month 20 to the fact that the customer stopped buying coffee during this window. In month 22, the decrease is sharper because the customer lost several significant products: milk, sponge and cheese. We can perform this precise analysis for each window where our model suggests that the customer is not as loyal as he was before. This information is of interest for the retailer because he can then target his marketing on significant products that this customer is not buying anymore.

4. CONCLUSION

In this paper, we presented a new model to analyze customer attrition in a grocery retail context. This model is based on the customer basket content evolution and provides precise information about individual defecting customer. It can also reliably detect customer defection.

In the future, we plan to deepen the study of the characterization of significant products that can explain customer defection.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] W. Buckinx and D. V. den Poel. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *EJOR*, 164(1):252 − 268, 2005.

[2] V. L. Miguéis, D. Van den Poel, A. S. Camanho, and J. Falcão e Cunha. Modeling partial customer churn: On the value of first product-category purchase sequences. *ESWA*, 39(12):11250–11256, 2012.

[3] D. Shepard. The new direct marketing. *Business One Irwin, Homewood, IL*, 1990.