Customer segmentation based on a collaborative recommendation system: Application to a mass retail company

Virginie Vandenbulcke, Fabian Lecron, Caroline Ducarroz, François Fouss

Abstract

Identifying customer segments has been arousing interest in the literature for decades, for various reasons. This paper introduces a different way of segmenting customers in the retail sector based on shopping behaviors: unlike traditional criteria, our segmentation criterion lies on items that we recommend shoppers to buy in the future, these recommendations being defined on what other shoppers, who act similarly in terms of shopping basket, buy. Our approach is thus made of two steps: first, products to be recommended to each specific customer are determined and second, customers are segmented following these recommendations. A first analysis on real data has been performed and preliminary results are shown in this paper. A possible application lies in the area of targeted promotional material, for example advertising folders or brochures adapted for each segment.

Keywords

Customer segmentation – Recommender system – Targeted communication - Retailing

Track

Retailing, Channel Management and Logistics
1. Introduction

Identifying segments of customers has been arousing interest in the literature for decades (see, for example, Smith, 1956; Dhalla and Mahatoo, 1976; Beane and Ennis, 1989; Kamakura and Russell, 1989; Van Raaij and Verhallen, 1994; Lambin et al., 2005), for various reasons. From a managerial point of view, it has been shown that moving from mass marketing to a more targeted marketing (i.e., based on segments) is important and valuable (Rossi et al., 1996; Kolyshkina et al., 2010) and that business performance can be improved by applying customer segmentation (Rossi et al., 1996; Joh et al., 2003). Furthermore, when applied in the retail sector, segmentation can help managers to address the right subdivision of the market (Boone and Roehm, 2002), increase loyalty (Joh et al., 2003; Mägi, 2003; Danaher et al., 2003), gather additional market information (e.g., discovering unexpected segments (Kotler et al., 2009)), or increase unplanned purchases. Note that segmentation is also beneficial for the customer, e.g., time saving, less product research, better decision-making, etc. (see, for details, Elrod and Winer, 1982; Kotler et al., 2009; Claeyssen et al., 2011). As a matter of fact, giving the right information at the right place to the right customer turns out to be a competitive advantage for a company (Jackson, 2007).

Traditional criteria used to segment a market are mostly based on socio-demographic characteristics (see, for example, Sexton, 1974; Williams et al., 1978; Rossi et al., 1996; Joh et al., 2003; Lambin et al., 2005) or on individual shopping behavior, usually studied through the customer’s shopping history (e.g., via loyalty cards) (Bell and Lattin, 1998; Kim et al., 1999; Popkowski et al., 2000; Kolyshkina et al., 2010).

This paper proposes a different way of segmenting customers based on shopping behaviors: unlike traditional criteria, our segmentation criterion lies on items that we recommend shoppers to buy in the future, these recommendations being defined on what other shoppers, who act similarly in terms of shopping basket, buy. More precisely, our approach is made of two steps: first, products to be recommended to each specific customer are determined (based on products that other customers – whose shopping basket is similar to this specific customer – buy), and second, customers are segmented following these recommendations (see Figure 1).

![Figure 1: Framework overview](image_url)

This segmentation is thus based on a “recommended” shopping basket for each customer. The segmentation base is not a prediction of shoppers’ future behavior: it is a recommendation about their future behavior. Managerial implications can be derived from this segmentation, and as such new marketing strategies in the offline retail sector can be developed or existing strategies improved. The goal therefore is to convey appropriate information to the appropriate offline retail customer using a new way of segmenting these shoppers. One possible application lies in the area of targeted promotional material, for example advertising folders or brochures adapted for each segment.

Section 2 introduces the literature review and presents the positioning of this paper. In section 3, the data, model and methodology are described. Finally, section 4 gives results and discussion. Concluding remarks and possible extensions are discussed in section 5.

2. Literature Review and Positioning of the Paper

Segmentation divides a population into different groups based on similar characteristics, behaviors, preferences, etc. (see Aurier, 1989; Boone and Roehm, 2002; Kotler et al., 2009).
The criteria used in the retail literature to build customer segments can nicely be classified by the answer they provide to the following questions on the shopper.

**Who is he?** Such segmentations are based on socio-demographic characteristics, e.g., sex, age, wage, status, job, domicile, etc. (Sexton, 1974; Wedel and Steenkamp, 1991; Kotler et al., 2009).

**When does he shop?** The segmentation criterion is the shopping frequency (Kim and Park, 1997; Popkowski and Timmermans, 1997).

**How much does he spend?** This question is linked to the total amount of the customer’s shopping basket, i.e., the amount of money spent by customers per shopping trip (Kahn and Schmittlein, 1989; Kahn and Schmittlein, 1992; Popkowski and Timmermans, 1997; Walters and Jamil, 2003; Boztug and Reutterer, 2008).

**How does he shop?** This is a reference to the customer’s loyalty to the stores he frequents (Popkowski and Timmermans, 1997; Popkowski et al., 2000; Danaher et al., 2003; Joh et al., 2003), as well as the type of store (EDLP vs. HILO) he chooses (Bell and Lattin, 1998).

**Why does he shop?** Segmentation is based on his shopping motivations (Reynolds and Beatty, 1999), needs and desires (Reynolds and Beatty, 1999; Rohm and Swaminathan, 2004), and expectations towards a product or a product category (Desmet and Hekkert, 2007; Kotler et al., 2009) – perceived value, perceived benefits, and advantages sought after (Petrick, 2002; Kotler et al., 2009).

**What does he buy?** An important stream of research focuses on segmenting customers based on brand preferences, that is, brand choice probabilities (see, for example, Elrod and Winer, 1982; Kamakura and Russel, 1989; Danaher et al., 2003; Bodapati, 2008). Also of interest is the segmentation of customers based on their price and sales promotion elasticities (Allenby et al., 1998), or on their choice behavior across different product categories (Russel et al., 1997).

Our paper proposes a different way of segmenting customers in the offline retail sector. This approach goes beyond traditional segmentation criteria since an extra dimension is introduced. This extra dimension lies in recommendations that can be made to a specific customer about his future shopping basket, based on other customers’ shopping basket. It thus goes further in the question: **What do other shoppers – with purchasing behavior close to that of the studied shopper – buy?**

Our approach, by bringing together the areas of recommendation and segmentation, goes one step beyond existing literature. Though previous articles on segmentation also use, as we do, customers’ purchases to perform segmentation, none of them, to the best of our knowledge, uses recommendations (based on customers’ shopping baskets) to form segments of customers. Also, when segmenting, they do not usually incorporate information about all purchases made by customers, as pointed out by Joh et al. (2003), but only the single last purchase (see for instance, Rossi et al., 1996; Erdem, 1996; Claeyssen et al., 2011).

Including a recommendation dimension looks promising. Recent literature on recommendations shows that recommendations in a computer-mediated shopping environment are often considered as having an influence on customers’ consideration sets, – a consideration set comprises all the products out of which a customer will pick products to buy – in size as well as quality (Senecal and Nantel, 2004; Hostler et al., 2011). We posit that this is also the case in offline settings (as do Mild and Reutterer (2003), who focus on the offline retail sector).

However, we assume that these recommendations should not be used *per se* in the mass offline retail sector: forming customer segments makes much more sense in the mass offline retail sector. While today, personalization and customization are natural givens in the one-to-one online business (examples such as Amazon, Netflix, recently eBay, etc. show that the recommendation technology is readily implemented – Senecal and Nantel, 2004; Ansari et al., 2000; Hostler et al., 2011; Stenger and Bourliatass-Lajointe, 2011), most traditional mass offline retailers – that is, retailers that are not exclusively online retailers – use mass mailing or
segmented mailing for their promotional material. Due to the major differences between the offline and online sectors, it is commonly accepted that the one-to-one approach used in the online sector is infeasible both financially and operationally in the offline retail sector (Kotler et al., 2009) – e.g., it would take tremendous efforts, time and costly investments. Privacy issues are a concern as well, as some people dislike the intrusion of personalization systems into their private life (Acquisti and Varian, 2005; Stenger et al., 2011).

For the offline sector, some authors consider the concept of “mass personalization” to include global marketing strategies, where customers feel directly addressed or feel that the retailers have their best interest at heart (Kotler et al., 2009; Claeyssen et al., 2011). This can be achieved by using targeted segmentation. The goal of these efforts – to personalize the relation and communication with customers or with a group of customers – refers directly to the Customer Relationship Management (CRM). For Bodapati (2008), the main idea behind the concept of CRM is add-selling, i.e., to increase revenue. This means that retailers try to sell more products to their existing customer base rather than acquiring new customers.

Our approach allows marketing managers to form segments of customers based on recommendations, which themselves use similarities between customers in terms of their purchase history in the offline mass retail sector. As such, the goal of our paper is to propose a new way of segmenting shoppers, by conveying the appropriate information to the appropriate offline retail customers.

3. Methodology

This section first describes the data used in our analyses and then provides some guidance on our two-step approach described in Figure 1. Note that, since our objective is not to develop a new state-of-the-art algorithm nor for recommendation, nor for clustering, details of the applied algorithms (which are common in recommendation and clustering, respectively) are not provided here, while all the important references are provided.

**Data.** The data were provided by one of the major Belgian mass retailers. Due to a confidentiality clause, details on this company cannot be published. The database size is considerable. Over a period of 12 months from June 2009 to May 2010, 2,169,046 shoppers were regarded as 56,976 products. Our data are structured as illustrated in Table 1. In this table, “Number_of_visits” refers to the number of times that the $i^{th}$ customer bought the $j^{th}$ product, during the considered period.

<table>
<thead>
<tr>
<th>ID_product</th>
<th>ID_user</th>
<th>Number_of_visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>340</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>367</td>
<td>5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 1: Structure of the data file used in this work

Because of the considerable quantity of data, we decided to work on a sample of 10,000 shoppers and on the 4349 “micro-segments” of products (defined by the retailer) rather than to work with all individual products. A “micro-segment” refers to a specific product (within a specific brand), but no distinction is drawn between similar products with different packaging (e.g., 1L vs 1.5L).

Finally, data are transposed in a matrix $M$, where each row represents a micro-segment (i.e., 4349) and each column stands for a customer (i.e., 10,000). We then measure the intensity of the link between customer $i$ and micro-segment $j$, by computing a purchase frequency (using the third column of Table 1). When customer $i$ bought micro-segment $j$, the corresponding element in the matrix $M$ (i.e., $m_{ij}$) is set to a relative purchase frequency computed by the ratio between the number of times customer $i$ bought a product belonging to micro-segment $j$.

1 Please note that other measures of the intensity of the link between a customer and a micro-segment could be considered.
and the total number of visits made by customer $i$ at the retailer’s point of sale. The resulting matrix $M$, whose elements $m_{ij}$ therefore lie between 0 and 1, is then provided as input data to an two-step algorithm.

**Methodology.** Regarding our first step (i.e., computing a list of recommended products for each customer), we determine (by considering as input data our matrix $M$) each product $j$’s nearest neighbors items, before using these neighborhoods and the shopping basket of the $i^{th}$ customer to compute recommendations (see Fouss et al., 2007). As shown by Lü et al. (2012), many recommendation algorithms have been developed during the last decades. In this paper, we apply the “commute-time kernel” (see Fouss et al., 2007 for a description of the algorithm) as it has been shown that it provides good results in a recommendation task (Fouss et al., 2012). Recommendation techniques, mainly tested on data related to movie ratings, have been here adapted to the retailing environment in two ways: (i) first, the link between a customer and a micro-segment is assessed through the purchase frequency of this micro-segment ($m_{ij}$), as opposed to most research papers in the recommendation field, which focus on explicit preference ratings and not on actual purchases, as pointed out by Moon and Russel (2008), and as opposed to Mild and Reutterer (2003), who model binary responses (choice/non-choice of customers among product categories). This is particularly important because retailers, especially in the offline mass retailing sector, do not usually have access to explicit ratings made by customers on products, but are usually able to track customers’ purchase history. Assessing purchase frequency also makes more sense in this sector since it mainly deals with mass-market products that can be bought several times (by contrast, a book is generally bought once, a movie is seen once). Using purchase frequency to measure the intensity of the link between a customer and a product thus seems to be particularly adapted to this context. (ii) Second, a trans-category view has been adopted, in that “neighbors” are designated based on their purchases in different categories. This seems to be particularly adapted in the mass retailing sector, where customers usually buy items from multiple categories during a shopping trip, and is in line with the literature, which underlines the fact that customers are permanently involved in multi-category decision making, such as grocery shopping trips (Mild and Reutterer, 2003). Regarding the second step (i.e., segmenting customers following recommendations provided in the first step), as no clustering technique is generally superior across data sets (Wedel and Kamakura, 2000), we firstly apply a hierarchical ascending method – with Ward’s minimum variance method (which does not need to determine a priori the best number of segments), and secondly, in order to validate our results, a k-means algorithm (see Tou and Gonzalez, 1974) as it is the most frequently used market segmentation technique among the clustering techniques in the literature (Boone and Roehm, 2002).

4. Results and Discussion

This section reports preliminary results for segmenting customers in the retail sector. In the literature, it is commonly accepted that a good cluster is a “set of objects in which each object is closer to every other object in the cluster than to any object not in the cluster” (see, for example, Tan et al., 2006). Therefore, clustering implies minimizing intra-cluster distances and maximizing inter-cluster distances. A hierarchical method (Ward) was first used to determine the appropriate number of clusters; the Cubic Clustering Criterion suggests a five-cluster solution (CCC = 4.99). A k-means method was then applied, and validates, since leading to a good value of CCC for a number of clusters $k=5$, this five-cluster solution. Moreover, the rand index and the adjusted rand index which are two classical measures of similarity between two clusterings, were computed, showing results (0.78 and 0.40, respectively) confirming the similarity between the clustering provided by the hierarchical method and the clustering provided by the k-means method. This number of clusters (5) seems to be manageable for retailers who want to adapt their promotional strategy, grouping together, customers with
similar recommended products. A description of each of these segments can be provided on request. However, although this description is essential for the retailer (in order to operationalize his segmentation strategy), the content of these segments is not the core part of this paper, whose goal consists in suggesting a new approach for segmenting customers.

5. Conclusion and future research

This paper proposes a new way of segmenting customers in the mass offline retail sector based on shopping behaviors. Unlike traditional criteria, our segmentation criterion relies on products that we recommend shoppers to buy in the future. To our knowledge, using clustering techniques on recommendations based on purchase histories has never been developed in the literature, despite its implications for retailers. This project, still at an early stage, requires some further steps, in particular to assess the usefulness of such a segmentation, (i) for retailers (does this segmentation allow them to communicate more effectively to their customers and in fine to strengthen the customer relationship, and (ii) for customers (do they benefit from this segmentation?). Here, (i) is essential, and could be validated through experiments; for instance, customers’ reactions to a retailer’s segmented communication (e.g., in terms of promotional folders) could be assessed in terms of products they ended up buying. Other validation techniques will be investigated as well.

6. References


