



## Customer data mining for lifestyle segmentation

V.L. Miguéis\*, A.S. Camanho, João Falcão e Cunha

Faculdade de Engenharia da Universidade do Porto, Portugal

### ARTICLE INFO

**Keywords:**  
Retailing  
Clustering  
Segmentation  
Lifestyle

### ABSTRACT

A good relationship between companies and customers is a crucial factor of competitiveness. Market segmentation is a key issue for companies to develop and maintain loyal relationships with customers as well as to promote the increase of company sales. This paper proposes a method for market segmentation in retailing based on customers' lifestyle, supported by information extracted from a large transactional database. A set of typical shopping baskets are mined from the database, using a variable clustering algorithm, and these are used to infer customers lifestyle. Customers are assigned to a lifestyle segment based on their purchases history. This study is done in collaboration with an European retailing company.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

The recent economic and social changes that occurred in Europe transformed the retailing sector. In particular, the relationship between companies and customers changed significantly. In the past, companies focused on selling products and services without searching detailed knowledge concerning the customers who bought the products and services. With the proliferation of competitors, it became more difficult to attract new customers, such that companies had to intensify efforts to keep current consumers. The evolution of social and economic conditions also changed lifestyles, and as a result customers are less inclined to absorb all the information they receive from the companies. This context led companies to evolve from product/service-centered strategies to customer-centered strategies. The establishment of loyalty relationships with customers also became a main strategic goal. Indeed, companies wishing to be at the leading edge have to continually improve the service levels in order to ensure a good business relationship with customers.

Some companies invested in building databases that are able to collect a big amount of customer-related data. For each customer, millions of data objects are collected, allowing the analysis of the complete purchasing history. However, the information obtained is seldom integrated in the design of business functions such as marketing campaigns. In fact, in most companies the information available is not integrated in procedures to aid decision making. The overwhelming amounts of data have often resulted in the problem of information overload but knowledge starvation. Analysts are not being able to keep pace to study the data and turn it into useful knowledge for application purposes.

Data mining (DM) techniques are rising as tools to analyze data resulting from customers' activity, stored in large databases. They can be applied in order to detect significant patterns and rules underlying consumer behavior. However, the use of DM in marketing is still incipient and most companies still use mass strategies to instigate customers loyalty. The marketing segmentation of customers or the identification of customer groups with similar behavior patterns is often done in an ad-hoc way, which constitutes the basis for the definition of customized promotions.

This paper proposes a method for customers segmentation, informed by the nature of the products purchased by customers. This method is based on clustering techniques, which enable segmenting customers according to their lifestyles.

The structure of the remainder of the paper is as follows. Section 2 includes a review of segmentation approaches. Section 3 introduces the company used as case study. Section 4 includes a presentation of the methodology, and Section 5 presents the data and discusses the results. Section 6 suggests marketing actions based on the lifestyle segmentation. The paper finishes with the conclusion.

### 2. The evolution of segmentation approaches

Segmentation approaches were initially based on geographic criteria, such that companies would cluster customers according to their area of residence or work. This was followed by segmentation based on socioeconomic indicators, such that customers would be grouped according to age, gender, income or occupation. Marketing segmentation research gained momentum in the 1960s. Twedt (1964) suggested the use of segmentation models based on volume of sales, meaning that marketing efforts should focus on customers engaged in a considerable number of transactions. This approached, called "heavy half theory", highlighted that one half of

\* Corresponding author. Tel.: +351 225082132.

E-mail address: [vera.migueis@fe.up.pt](mailto:vera.migueis@fe.up.pt) (V.L. Miguéis).

the customers can account for up to 80% of total sales. Frank, Massy, and Boyd (1967) criticized this segmentation arguing that this assumes that the heavy purchasers have some socioeconomic characteristics that differentiate them from other purchasers, what was rejected by the regression analysis carried out. Subsequently Haley (1968) introduced a segmentation model based on the perceived value that consumers receive from a good or service over alternatives. Thus, the market would be partitioned in terms of the quality, performance, image, service, special features, or other benefits prospective consumers seek. These models triggered further research that allowed to obtain sophisticated lifestyle-oriented approaches to segment customers. The lifestyle concept, introduced in the marketing field by Lazer (1964), is based upon the fact that individuals have characteristic patterns of living, which may influence their motivation to purchase products and brands. During the 1970s, the validity of the multivariate approaches used to identify the variables that affect deal proneness was criticized (see Green & Wind, 1973), which motivated the development of enhanced theoretical models of consumer behavior (e.g., Blattberg, Buesing, Peacock, & Sen, 1978). One decade later, Mitchell (1983) developed a generalizable psychographic segmentation model that divided the market into groups based on social class, lifestyle and personality characteristics. However, practical implementation difficulties of this complex segmentation model was widely noted during the 1990s in, for example, Piercy and Morgan (1993); Dibb and Simkin (1997).

More recently, the marketing literature raised the concern that customers are abandoning predictable patterns of consumption. The diversity of customer needs and buying behavior, influenced by lifestyle, income levels or age, makes past segmentation approaches less effective. Therefore, current models for marketing segmentation are often based on customer behaviour inferred from transaction records or surveys. The resulting data is then explored with data mining techniques, such as cluster analysis. Examples of applications of data mining for segmentation purposes using survey results include Kiang, Hu, and Fisher (2006). In the context of long-distance communication services, the clients were segmented using psychographic variables, based on data of a survey composed by 68 attitude questions. Min and Han (2005) clustered customers with similar interests in movies based on data containing explicit rating information for several movies provided by each customer. The rating information allowed to infer the perceived value of each movie for each customer. Helsen and Green (1991) also identified market segments for a new computer, system based on the use of cluster analysis with data from a customer survey. The segmentation was supported by the rate of importance given to the product attributes.

### 3. Description of the company used as case study

This paper describes an application driven methodology for effectively segmenting customers of an European retailing company. The retailing company used as case study has a chain of food-based stores, i.e. hypermarkets, large supermarkets and small supermarkets. These formats differ essentially by the sales area of the store and by the range and price of products offered. The establishment of loyal relationships with customers is a main strategic goal for this company. The development of the company information system and the implementation of a loyalty program, supported by a loyalty card, have enabled collecting data on each customer profile (e.g. customer name, address, date of birth, gender, number of people in the household, the telephone number and the number of one identification document) and transactions (date, time, store, products and prices). Currently, approximately 80% of the total number of transactions is done by customers using the loyalty card.

The company classifies each product according to the business unit, the category, the subcategory, the product description, the brand and the position of the brand based on its value. For example, a product can be classified in 12 business units (e.g. Drinks, Grocery, Fishery), in 116 categories (e.g. Beers, Desserts, Frozen), in 803 subcategories (e.g. Beers with alcohol, Fruit syrup, Frozen shellfish) and in 5 positions of the brand (i.e. Premium, Sales leader, Secondary, Own brand and Economic). The total number of different products currently commercialized by the company is about 1,363,409.

At present, the company customers are segmented in two ways. One of them consists on 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 by the company “frequency and monetary value” (FM) model. According to the values of these two variables, the company specifies 8 groups of customers. Each client integrates one of these groups, according to the average number of purchases done in a 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, Camanho, & Cunha, 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 of products they purchase. Each segment of clients is defined by using a clustering algorithm, based on the similarity between the products purchased by the client and the categories of products included in pre-defined baskets, evaluated in percentual terms. The insights provided by this segmentation method are currently not fully explored by the company. However, it is expected that in the near future, the information provided by this segmentation method can be used to guide decisions concerning the variety of products available in each store, as well as their prices.

### 4. Methodology

The methodology followed in this paper aims to segment customers from the retailing company according to their lifestyle. To achieve this purpose we first identify typical shopping baskets, by considering the products more frequently purchased together. In the context of this analysis, a shopping basket is defined as the set of distinct products bought by a customer over the period considered. Customers’ lifestyle is inferred by analyzing the products included in the typical shopping baskets. Customers are then assigned to the lifestyle segments by considering the history of their purchases.

Clustering analysis is a widely used data mining technique that maps data items into unknown groups of items with high similarity (i.e., clusters). There is a large variety of clustering algorithms available (see Jain, Murty, & Flynn, 1999 for an overview). Most clustering algorithms can be classified in partitional or hierarchical. A partitional clustering is a division of the data items into non-overlapping groups, such that each item belongs to exactly one cluster. Partitional techniques require the prior specification of the number of clusters. Despite this limitation, partitional techniques have the advantage of allowing the optimization of a criterion related to similarity of objects within clusters or dissimilarity between clusters. Hierarchical algorithms can be classified as agglomerative or divisive. An agglomerative hierarchical clustering starts with clusters containing single items and then merges them until all items are in the same cluster. In each iteration the two

Download English Version:

<https://daneshyari.com/en/article/10322441>

Download Persian Version:

<https://daneshyari.com/article/10322441>

[Daneshyari.com](https://daneshyari.com)