Engineering Applications of Artificial Intelligence 47 (2016) 101–109

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A biclustering-based method for market segmentation using customer pain points

Binda Wang [a], Yunwen Miao [a], Hongya Zhao [b], Jian Jin [c], Yizeng Chen [d,*]

[a] Department of Management Science and Engineering, School of Management, Shanghai University, Shanghai, China
[b] Industrial Center, Shenzhen Polytechnic, Shenzhen, China
[c] Department of Information Management, School of Government, Beijing Normal University, Beijing, China
[d] School of Management, Shenzhen Polytechnic, Shenzhen, China

ARTICLE INFO

Keywords:
Biclustering
Customer pain points
Market segmentation
BCBimax algorithm

ABSTRACT

Market segmentation plays a crucial role in product design and development. However, conventional segmentation approaches based on one-way cluster analysis techniques have met two special challenges in practice. First, conventional approaches that derive a global result rather than a local one fail to cluster customers into such groups who have similar characteristics on a fraction of variables. Second, since there is no formal mechanism to select appropriate segmentation variables, different combination of variables will obtain different segmentation results, which makes the approaches not quite convincing. To overcome the two limitations, a novel biclustering-based market segmentation method by using customer pain points is proposed in this paper. Different from one-way algorithms clustering only rows or only columns, biclustering algorithms cluster both rows associated with customers and columns associated with customer pain points simultaneously to identify homogenous subgroups of customers with common characteristics towards a subset of segmentation variables. In addition, customer pain points are used to replace traditional segmentation variables in the presented method, which makes the results more reasonable. Subsequently, an illustrated example is studied to demonstrate the effectiveness of the presented method.

1. Introduction

As a classic market strategy, market segmentation separates a large number of customers into meaningful groups who share similar characteristics, requirements and behaviors (Smith, 1956; Dickson and Ginter, 1987; McDonald and Dunbar, 2004). It plays a crucial role in product development and has become an essential part of product innovation (Moorthy, 1984; Debo et al., 2005). According to the information of segments, companies can develop new types of products to match diverse customer needs or deploy resources effectively to manufacture a product for the most potential segment (Chan et al., 2012). Hence, segmenting markets enables companies to increase the opportunities of market success under and enhance the competitiveness under the drastic market competitions. Nevertheless, nowadays conventional market segmentation approaches face two main challenges.

The first challenge comes from the critical limitation of the standard partitioning clustering algorithms widely used in market segmentation approaches, such as hierarchical clustering, *k*-means clustering and self-organizing maps (Punj and Stewart, 1983; Wedel, 2000). These algorithms aim to divide a set of objects into groups (clusters) by finding a one-way division of data to produce clusters where customers behave similarly over all the segmentation variables (Kluger et al., 2003). However, in the practical scenarios, customers share similarly only on a small fraction of variables, such as knowledge, need, attitude, interest and loyalty status (Grover and Srinivasan, 1987; Yankelovich and Meer, 2006). In other words, these clustering methods obtain a global model rather than a local model, failing to discover subgroups of customers who have similar characteristics on partial variables, especially in high-dimensional data (Zhao et al., 2012).

The second challenge is that in order to tailor products precisely to satisfy customer requirements, selecting suitable segmentation variables is of great importance to carry out accurate market segmentation (Wedel, 2000). However, variables adopted by researchers and analysts differ largely. For example, segmentation according to customer behavior is based on variables like customer knowledge, loyalty status, attitude and response, while psychographic segmentation partitions customers into groups in accordance with their interests, opinions, lifestyles and activities (Beane and Ennis, 1987). In fact, market segmentation can also be implemented based on the customer pain point, which is ignored

* Corresponding author.
E-mail address: mfcyz@shu.edu.cn (Y. Chen).

http://dx.doi.org/10.1016/j.engappai.2015.06.005
0952-1976/

by previous literatures. In the business, customer pains are not physical but emotional. They are caused by customers' psychological gap or dissatisfaction, since customer expectations are not fulfilled through experiencing products (Lee, 2014). Pain points reflect customers' core concerns, main interests and emergent needs for products, thus identifying groups of customers who have similar pain points is more beneficial for companies to achieve accurate market segmentation and positioning (De Bonte and Fletcher, 2014). Furthermore, after segmenting markets according to customer pain points, the company can figure out key solutions for customer pains of each market niches. This can enhance product attractiveness and improve customer satisfaction effectively. Because the pains that make customers uncomfortable, annoying or frustrating towards a product, normally result from deficiencies, shortcomings, problems, or defects of the product (Homburg and Fürst, 2007). In general, people spend their money on two things, i.e., to combat pain and to pursue pleasure (Tybout and Artz, 1994). It's likely that the worse the pain or the problem, the more chance the company has of selling a solution. The better the solution the company offers, the quicker the customer will pay. Generally, there is more staying power in 'pain-solving businesses' than that in 'pleasure businesses' (Lee, 2014). In other words, if the company can develop products with a 'cure' that makes customer pains go away, it will greatly increase the success rate of new or revised products (Tuli et al., 2007).

To overcome the two aforementioned problems, a novel biclustering-based market segmentation method by using customer pain points is proposed in this paper. As opposed to clustering only rows or only columns in traditional one-way algorithms, a biclustering algorithm is applied to cluster both rows associated with customers and columns associated with customer pain points simultaneously in the presented method. Biclustering retrieves subgroups of objects that are similar in one subgroup of customer pain points and different in the remaining ones. This technique helps to identify homogenous subgroups of customers with common characteristics towards a subset of customer pain points.

The rest of the paper is organized as follows. In Section 2, the methodology of implementing biclustering technique for market segmentation is introduced. In Section 3, a flowchart of the biclustering-based method of market segmentation using customer pain points is formulated. In Section 4, an illustrated example is given to demonstrate the effectiveness of the proposed method. Finally, conclusions are summarized in Section 5.

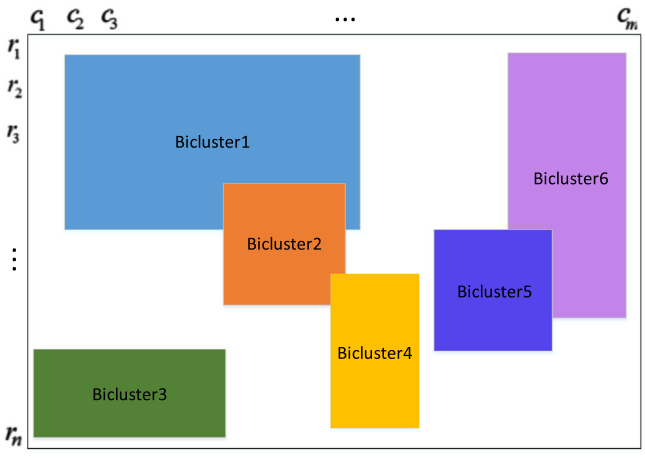

Fig. 1. Representation of the biclustering model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Methodology

2.1. The basic idea of biclustering technique

The idea of biclustering method was first put forward by Hartigan (1972). But no one applied this method for almost 30 years. On the one hand, the detection of biclusters is an NP-hard problem and the computational complexity is very high (Zhao et al., 2007, 2009). On the other hand, there were few needs for using biclustering to solve clustering problems in early years. It is the development of computer and information technologies that makes it possible to solve the biclustering problems with high-dimensional data efficiently. Until the beginning of 21th century, the analysis of microarray data brought biclustering back into focus. Because the main goal of analyzing microarray data is to classify genes based on their expression under various experiment conditions, or to classify conditions in terms of the expression of many genes, but traditional clustering methods are unfit for clustering gene expression data since (1) gene expression data is high dimensional with large space and lots of redundant or irrelevant information; (2) a large number of genes are co-regulated under limited conditions; (3) a gene may be involved in more than one biological process (Turner et al., 2005). Thus, biclustering was proposed by Cheng and Church (2000) to discover subgroups of genes that share similar transcriptional behaviors over a subset of conditions in a microarray experiment, as represented in Fig. 1.

Now besides in biological data analysis, the biclustering technique has been applied successfully in other interesting fields, including information retrieval, text mining (de Castro et al., 2007), market data analysis (Dolnicar et al., 2012), recommendation systems (Inbarani and Thangavel, 2011), financial forecasting and trading (Huang, 2011), and collaborative filtering (Symeonidis et al., 2008). It has become a powerful tool of data mining, particularly in analyzing high-dimensional datasets.

In reality, biclustering technique itself is not a specific algorithm, but the task to solve the problem of two-way partition of data. It can be achieved by a variety of algorithms that differ significantly in (1) the types of data they can deal with (e.g., binary data, ordinal data, and metric data), (2) the types of biclusters to be identified (see Figs. 2 and 3 the structures (similarity) of biclusters to be obtained, as shown in Fig. 3 (Dolnicar et al., 2012). Therefore, it is crucial to select an appropriate algorithm to handle the data with given type and structure at hand. For instance, the Plaid Models proposed by Lazzeroni and Owen (2002) is suitable for discovering arbitrarily positioned overlapped biclusters with coherent values, while the OPSMs algorithm raised by Ben-Dor et al. (2003) can find single bicluster or arbitrarily positioned overlapped biclusters with coherent evolution.

2.2. The procedure of BCBimax algorithm in market segmentation

In this subsection, the BCBimax algorithm of biclustering technique is introduced to conduct market segmentation using customer pain points. The BCBimax method that derives from the Bimax method was first put forward by Prelić et al. (2006) to serve as a reference method or a baseline for comparison of main biclustering algorithms employed in gene expression data. The original model assumes that per gene has two possible expression levels, i.e., expression and no expression in regard to a control condition. Then the gene expression values can be transformed respectively to 1 or 0 by using a 2-fold change cutoff of log2 expression values. Thus, the gene expression matrix is represented by a binary matrix. The objective of the Bimax model is to search for sub-matrices with all of expression values equaling 1 in a binary matrix, as shown in Fig. 4. The found sub-matrices exhibit that their rows (genes) behave consistent patterns over its