Expert Systems with Applications 40 (2013) 2224–2232

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Visitors of two types of museums: A segmentation study

Juan Gabriel Brida, Marta Disegna *, Raffaele Scuderi

School of Economics and Management, Free University of Bozen-Bolzano, Universitätsplatz 1 – piazza Università 1, 39100 Bozen-Bolzano, Italy

ARTICLE INFO

Keywords:
Bagged Clustering
Logit models
Museum
Segmentation
Motivation

ABSTRACT

Market segmentation comprises a wide range of measurement tools that are useful for the sake of supporting marketing and promotional policies also in the sector of cultural economics. This paper aims to contribute to the literature on segmenting cultural visitors by using the Bagged Clustering method, as an alternative and effective strategy to conduct cluster analysis when binary variables are used. The technique is a combination of hierarchical and partitioning methods and presents several advantages with respect to more standard techniques, such as *k*-means and LVQ. For this purpose, two ad hoc surveys were conducted between June and September 2011 in the two principal museums of the two provinces of the Trentino-South Tyrol region (Bolzano and Trento), Northern Italy: the South Tyrol Museum of Archaeology in Bolzano (ÖTZI), hosting the permanent exhibition of the "Iceman" Ötzi, and the Museum of Modern and Contemporaneous Art of Trento and Rovereto (MART). The segmentation analysis was conducted separately for the two kinds of museums in order to find similarities and differences in behaviour patterns and characteristics of visitors. The analysis identified three and two cluster segments respectively for the MART and ÖTZI visitors, where two ÖTZI clusters presented similar characteristics to two out of three MART groups. Conclusions highlight marketing and managerial implications for a better direction of the museums.

1. Introduction

Museums are the most popular cultural attractions, usually followed by art galleries and monuments (McKercher, 2004). For a long time visitors of cultural attractions were treated as a homogeneous mass of people. The tendency of the recent tourism literature is instead to consider them as a heterogeneous market with different characteristics, perceptions and needs (Hughes, 2002). Brida, Disegna, and Osti (2012) showed that visitors of Christmas Markets in Northern Italy clustered into three groups according to a set of motivational factors that drove them to make the visit. Other studies showed that tourists who visited art museums presented different socio-demographic characteristics (in particular regarding the level of education, income and occupation) than those who engaged in festivals, musical activities, theme parks, amusements parks, local fairs, and events (Bennett, 1994; Kim, Cheng, & O'Leary, 2007; Schuster, 1991).

Most research on tourism considered different types of museums (like art museum, stamps, history, science, and even children's museums) as a unique cultural attraction with the same "label". However, MacDonald and Alsford (1995) suggested they are heterogeneous, by affirming that "all museums are products of their particular cultural and historical experiences". Each museum exhibits its peculiarity by offering visitors different kinds of involvements (Dicks, 2003) and experiences, which are suitable for different kinds of tourists. Furthermore, an art museum, a history museum, an opera, or an outdoor festival might produce different experiences in visitors (Stylianou-Lambert, 2011). For these reasons research should analyse cultural attractions, and in particular museums, separately according to the subject matter and the experiences that they offer (Stylianou-Lambert, 2011).

Profiling museum visitors by taking into consideration also the different characteristics of the museums can be of crucial importance for managers and marketing analysts. Identifying homogeneous clusters of consumers-visitors can be in fact an essential step for planning and developing appropriate strategies, in order to satisfy the needs of each segment of guests. In this context clustering proposes a set of widely used unsupervised techniques with the aim to discover hidden associations among statistical units and identifying segments (Saarenvirta, 1998). Given a set of selected segmentation variables, these methodologies aggregate the units in groups, in such a way that each aggregation contains the most similar units, and at the same time it is dissimilar from the remainder. The supervision means that "membership of data points which can illustrate the general structure of the group is required in order to derive the classification rules". Therefore the absence of supervision implies that there is no rule for initiation of classification (Budayan, Dikmen, & Birgonul, 2009). This implies that the

* Corresponding author.
E-mail addresses: JuanGabriel.Brida@unibz.it (J.G. Brida), marta.disegna@unibz.it (M. Disegna), Raffaele.Scuderi@unibz.it (R. Scuderi).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.10.039

empirical distribution and characteristics of the data will determine the cluster membership.

Since the introduction of market segmentation in the late 1950s, the number and type of approaches for segmentation has grown enormously (Dolnicar & Leisch, 2004; Lia, Chu, & Hsiao, 2009). Unfortunately, as emphasized by many researchers, no absolutely "correct" way to segment a market exists in the literature (Beane & Ennis, 1987; Brida, Disegna et al., 2012; Dolnicar et al., 2008; Kotler, Bowen, & Makens, 2010). On the contrary, the researcher intervenes in different moments of the estimation process, which of course involves the final results. This implies that "clustering is exploratory data analysis and different methods present different views of data" (Leisch, 2006). The degrees of freedom in the clustering algorithm concern, among other things, the variables selection, the choice of a measure of dissimilarity between units, the final number of clusters, the test of the clustering solution as not purely random, the interpretation of final results for addressing management and marketing. Moreover, one has to bear in mind that "in the case of no clear cluster structure there is no "correct" solution" (Leisch, 2006).

The most popular clustering techniques are partitioning and hierarchical methods. The standard partitioning procedures aim to group the observations around a centre in order to find a segmentation of a set of units in an *a priori* fixed number of clusters. In the marketing and tourism literature *k*-means is the most commonly used algorithm that falls into this category. Hierarchical methods instead obtain the final clusters solution by repeatedly joining the "closest" clusters composed of one or more observations (agglomerative clustering), or repeatedly splitting the "further" clusters (divisive clustering). This study instead makes use of Bagged Clustering, which combines both hierarchical and partitioning methods. It was proposed by Leisch (1999) and has the advantages of overcoming many of the limitations of the two methods. This method has been used successfully in the past by Leisch himself or his research team, for the sake of tourism market segmentation (Dolnicar & Leisch, 2000, 2003; Dolnicar et al., 2008), but it has been applied infrequently by other researchers in the same field or in others (Huang, Chang, & Wu, 2009). Its application to the field of culture aims to study the profiles of tourists with respect to their motivations in visiting two different types of museums. This can shed light on investigating whether museums offering different experiences are visited by heterogeneous types of tourists, or on the contrary if segments with common characteristics can be detected. This objective is pursued by using a dataset from ad hoc surveys. These were conducted from June to September 2011 in the two main museums of Trento and Bolzano, the two provinces of the Trentino-South Tyrol region. The South Tyrol Museum of Archaeology (shortened to ÖTZI) is located in the Province of Bolzano and hosts the permanent exhibition of the mummy Ötzi, "the Iceman", whereas the Museum of Modern and Contemporary Art (shortened to MART) is places in the province of Trento and owns one of the most important collections in Italy for what concerns this artistic period.

The article will first proceed by outlining the research objectives, overviewing the clustering technique adopted, presenting the sample and questionnaire employed, and discussing the clustering results combined with binary and multiple Logit analysis. Both academic and practical implications, limitations of the research and future perspectives are provided.

2. Research objective

The focus of this paper is to find and describe groups of visitors with similar motivational characteristics in visiting an archaeological and a modern and contemporary art museum. This work constitutes a first attempt in studying whether there can be detected heterogeneous profile of visitors in two different types of museums.

The set of motivational factors for segmentation were measured as binary variables, i.e. "Yes/No", in the ad hoc survey used in this study. When binary data are used for the sake of clustering observations, it is a common practice in literature to use one of the following approaches: applying a hierarchical clustering method using a dissimilarity measures, such as Jaccard, Russell/Rao, Matching, or Dice, computed on the original data (Fingh, 2005; Řezanková, 2009); applying the *k*-means method using the Euclidean measure on the original data (Leisch, 2006); transforming the binary variables into continuous ones through a Factor Analysis, like the Correspondence Analysis, and then use the results as input of a clustering method using the Euclidean distance (Bouguila, 2010). When *k*-means is applied on the original data, the centres give the conditional marginal probabilities of observing a "1" (i.e., "YES" in one of the segmentation variables) given the cluster membership. It is important to underline that the dissimilarity measures for binary data have been extensively used and analysed with hierarchical methods but not with partitioning one, in which these types of measures are less common. In this context the Bagged Clustering method can be viewed as a useful solution for two reasons: it allows segmenting visitors by using the original binary data, and it overcomes the main limitations of the traditional segmentation methods.

3. Methodology

In this study, the Bagged Clustering method proposed by Leisch (1999) was adopted. This method is a combination of partitioning and hierarchical procedures and consists of the following steps:

1. First of all, B bootstrap sample $X_N^1, \ldots, X_N^B$ were constructed by drawing with replacement from the original sample X_N, were N is the sample size.
2. A partitioning method, called base method, is chosen by the researcher (e.g., *k*-means) and applied to each bootstrapped sample. From this procedure, $B \times K$ centres $c_1^1, \ldots, c_K^1, c_1^2, \ldots, c_K^2, c_1^B, \ldots, c_K^B$ are obtained, where K is the number of centres used in the base clustering method and c_j^i is the jth centre $(j = 1, \ldots, K)$ of X_N^i, which is the ith bootstrap sample $(i = 1, \ldots, B)$.
3. All the centres are combined into a new dataset $C_{B\times K}$.
4. A hierarchical cluster algorithm is applied to the $C_{B\times K}$ dataset in order to produce a partition of the centres.
5. The final outcome is displayed through the usual dendrogram of classical hierarchical methods, where the best partition of centres is obtained by simply investigating it. Finally, the partition of the original observations results from assigning the $x \in X_N$ observations to their closest centres. In this way each observation is assigned to the cluster containing the centre to which it is associated.

Fig. 1 schematically represents the steps that characterize the Bagged Clustering.

This method can be interpreted as both a complexity-reducing pre-processing stage for the hierarchical methods and a combination procedure of several partitioning results (Kang, Hua-Xiang, & Ying, 2008; Leisch, 1999). It has a better performance in comparison to other standard clustering methods for both continuous and binary data sets (Leisch, 1999). Furthermore, the Bagged Clustering technique overcomes many limitations of both partitioning and hierarchical algorithms. Partitioning methods are more flexible and perform better with large dataset than hierarchical methods (Everitt, Landau, Leese, & Stahl, 2011). The latter have the