Expert Systems with Applications 39 (2012) 11349–11356

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Segmenting visitors of cultural events by motivation: A sequential non-linear clustering analysis of Italian Christmas Market visitors

Juan Gabriel Brida [1], Marta Disegna *,[1], Linda Osti [1]

School of Economics and Management, Free University of Bolzano, Piazza dell'Università, 39100 Bolzano (BZ), Italy

ARTICLE INFO

Keywords:
Self-organizing map
Silhouette index
Cultural event
Motivation
Segmentation
Authenticity

ABSTRACT

Considering the importance of market segmentation as a marketing tool to determine promotional policies, this paper aims to contribute to the tourism literature using the two-level approach proposed by Vesanto and Alhoniemi (2000) as an alternative and effective method to conduct cluster analyses.

For this purpose, an empirical study was conducted interviewing tourists who visited three different Christmas Markets in Northern Italy. The two-level approach is based on two clustering techniques used in sequence: a Self-Organizing Map (SOM) followed by a clustering algorithm. The Silhouette index (Rousseeuw, 1987) is used as a guideline during the second level in the selection process of both the best clustering techniques (between hierarchical and non-hierarchical) and the best partition.

The analysis identified three cluster segments and this paper demonstrates the suitability of the clustering method adopted. In the discussion of the results, marketing and managerial implications are also highlighted.

1. Introduction

The management and marketing of events has been acknowledged as an area of extreme importance to academic researchers and event practitioners (Allen, O'Toole, Harris, & McDonnell, 2008; Getz, 2008) and, in this context, it is necessary to develop a profile for attendees through market segmentation that can guide marketing and managerial decision making (Lee, Lee, & Wicks, 2004). A review of the marketing literature indicates that there is no one correct way to segment a market (Beane & Ennis, 1987; Dolnicar, 2008; Kotler, Bowen, & Makens, 2010). However, clustering is one of the most important techniques used to identify segments (Saarenvirta, 1998). Boone and Roehm (2002) point out that there are over 50 clustering methods that can be applied to deal with market segmentation problems, so marketing strategists often face the problem of selecting the most appropriate technique or methodology. When discussing the need for an in-depth knowledge of tourists' behaviour patterns and therefore the need to use clustering techniques, Bloom (2004) points out the necessity to overcome the limitations deriving from linear techniques; hence he proposes the application of non-linear techniques such as Neural Networks (NNs) algorithms. NNs have been used to segment markets in different fields including the tourist market. In tourism research Mazanec (1992) appears to be one of the first scholars to use NNs applying this technique to a market segmentation analysis of Austrian tourists in the "Euro-Sports Region". Dolnicar (1997) used a further form of non-linear techniques – the Kohonen Self-Organizing Map (Kohonen, 1984) – to identify the characteristics of summer tourists visiting Austria. This last method was also used for example: to identify strategic groups of UK hotels (Curry, Davies, Phillips, Evans, & Moutinho, 2001); to segment senior travellers in Western Australia (Kim, Wei, & Ruys, 2003); to segment the international tourist market in Cape Town, South Africa (Bloom, 2005).

According to Bloom (2004), non-linear techniques, and in particular the Kohonen Self-Organizing Map (SOM), have overcome some of the limitations of the hierarchical clustering methods as they are more robust and are not significantly affected by missing data; they do not require any prior assumptions about the underlying distribution of the data; and they are appropriate also when the sample size is large (Kuo, Ho, & Hu, 2002). In comparison to the most popular non-hierarchical clustering methods, i.e., the *k*-means algorithm, the SOM algorithm does not require knowledge of the precise number of segments being targeted, i.e., the network assigns an incoming signal to the segment having the nearest weight vector only if the distance falls within a predetermined limit, otherwise a new segment is created (Venugopal & Baets, 1994). Furthermore, SOM has been extensively used in several different fields (like engineering, natural sciences, medicine, humanities, economics, mathematics, physics, and chemistry as

* Corresponding author. Tel.: +39 0471 013293; fax: +39 0471 013009.
E-mail addresses: JuanGabriel.Brida@unibz.it (J.G. Brida), marta.disegna@unibz.it (M. Disegna), Linda.Osti@unibz.it (L. Osti).

[1] Note: Authors are listed in alphabetical order since they all have contributed equally to the work.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.041

suggested by (Oja, Kaski, & Kohonen, 2003) also thanks to its data visualization capabilities (Churilov, Bagirov, Schwartz, Smith, & Dally, 2005; Länsiluoto & Eklund, 2008). Although the SOM algorithm has been used in several different cases, it is less often used in the field of tourism and, even less, for the segmentation of events' visitors.

As underlined by Vesanto and Alhoniemi (2000), the visualization of the SOM results can only be used in order to obtain qualitative information. To produce quantitative descriptions on the observations, interesting groups of map units must be selected from the SOM. In order to reach this purpose, Vesanto and Alhoniemi (2000) proposed a two-level approach where the observations were first clustered through the SOM, and then, the weight vectors of the SOM were clustered again using a further suitable technique. The two-level approach proposed by Vesanto and Alhoniemi (2000) should not be confused with the two-stage method proposed by Kuo et al. (2002) in which the SOM is used only to determine both the number of clusters and the starting point in order to employ the *k*-means algorithm. The Vesanto and Alboniemi approach is recent and innovative with promising potentials and is currently being applied to different research fields among which biological process (Aguado, Montoya, Borras, Seco, & Ferrer, 2008), environmental sustainability and green marketing (Mostafa, 2009, 2010), biomedicine (Wang, Delabie, Aasheim, Smeland, & Myklebost, 2002), and acoustic emission (Godin, Huguet, & Gaertner, 2005). However it has not yet been used in tourism and event marketing.

The data presented and analysed in this paper consist of 1193 responses by tourists visiting Christmas Markets in three cities of the Northern Italy: Bolzano, Brunico, and Trento. The sample size justifies the choice of the two-level approach in order to segment the tourists.

The first and foremost Christmas Market was held in Berlin in the 18th century and its primary function was to create a place for families to purchase children's presents. Trentino–South Tyrol (the region of our investigation) has been the first area of Italy to organise and celebrate Christmas Markets according to German culture. In fact, historically Trentino–South Tyrol was part of Austria till 1919, and still today around three quarters of the inhabitants of South Tyrol speak German as their first language and belong to the German/Austrian culture. In general Christmas Markets are open for the four weeks of advent from the end of November leading to Christmas Eve. The market stalls sell typical local products including food and drink, Christmas decorations, small gifts and presents, and local artefacts. The term "market" within the name of this event, may bring into people's mind the idea of a trade and shopping event. However, Christmas Markets are much more than just a shopping experience. In fact, shopping for Christmas presents has become just a simple side product of the event. The entire organisation of Christmas Markets in the 1970s, 1980s and 1990s has been about creating an event where people can experience the Christmas atmosphere according to the German tradition.

The paper will first present the research objectives followed by a presentation of the SOM algorithm and the internal validity measures. The paper continues with the outcomes of the empirical results and finishes with a discussion of the results, including practical implications of the two-level approach, limitations of the research and future perspectives.

2. Research objective

Profiling cultural event tourists is very important because one of the essential steps in marketing planning is to segment the market and develop strategies to satisfy the needs of the principal segments.

Through the implementation of the two-level approach proposed by Vesanto and Alhoniemi (2000) we would like to contribute to the tourism literature with an alternative and effective method to conduct cluster analyses.

Taken a dataset of 1193 tourists collected on three Christmas Markets in Northern Italy, we first conducted a SOM and then clustered the weight vectors of the SOM with another clustering technique. To find the best clustering techniques, between hierarchical and non-hierarchical, and to select the best number of clusters for the second level of the process we have used the Silhouette index (Rousseeuw, 1987), an internal validity index. As underlined by Vesanto and Alhoniemi (2000), in practice the value of the internal validity index must be interpreted as a guideline rather than an absolute truth. The use of this kind of index is a novelty in tourism study where it is common practice to decide:

(1) the clustering technique on the basis of the trade-off between advantages and disadvantages of hierarchical and non-hierarchical techniques (Pérez & Nadal, 2005), simply considering the sample size (Brida, Osti, & Barquet, 2010), or through unknown reasons (Denicolai, Cioccarelli, & Zucchella, 2010; Devesa, Laguna, & Palacios, 2010; Dey & Sarma, 2010; Lo, McKercher, Lo, Cheung, & Law, 2011; Wolf, Hagenloh, & Croft, 2011);
(2) the number of clusters of the non-hierarchical algorithm on the basis of practical and subjective preference (Albalate & Bel, 2010; Choi, 2011; Konu, Laukkanen, & Komppula, 2011; Pérez & Nadal, 2005) or deriving this information from a hierarchical cluster method (Bigné & Andreu, 2004; Chen & Hsu, 1999; Claver-Cortés, Molina-Azorín, & Pereira-Moliner, 2007; Punj & Steward, 1983).

3. Methodology

The SOM algorithm is a feedforward NN, which produces a non-linear projection of the observed points from the input layer to the output layer (the map) of lower dimensions. In the input layer the *j*-th neuron is described by a *p*-dimensional vector; the output layer consists of *M* output nodes, or neurons, usually structured in a bi-dimensional grid with a hexagonal or rectangular lattice. Two vectors describe each output node: a *p*-dimensional vector called weight vector, or prototype, which represents all training sample, or neurons in input layer, associated with it; a *q*-dimensional vector (with $q < p$) $\mathbf{r}_i$ of coordinates that identify its position in the output layer.

Schematically the SOM procedure consists of the following five stages:

(1) *Initialization*: the *M* weight vectors are initialized (Kohonen, 1984).
(2) *Competition*: each training sample, $\mathbf{x}_j$, is compared with each weight vector, $\mathbf{m}_i$, to find the output node, $c(j)$, that is more close, in terms of distance, to $\mathbf{x}_j$. The distance measure typically used is the Euclidian distance. The winning node $c(j)$, which is called Best-Matching Unit (BMU), is identified through the following rule:

$$c(j) = \arg\min_{1 \leqslant i \leqslant M} \{||\mathbf{x}_j - \mathbf{m}_i||\}, j = 1, \ldots, N \quad (1)$$

(3) *Cooperation*: in this stage it is necessary to define a neighbourhood function that allows to identify the output nodes close to the BMU, $c(j)$, to be updating in the next step.

The most widely used neighbourhood function in the literature, $h_{c(j),i}$, is the Gaussian function: