Knowledge-Based Systems 73 (2015) 335–346

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# Bagged fuzzy clustering for fuzzy data: An application to a tourism market

Pierpaolo D'Urso [a,1], Marta Disegna [b,*,1], Riccardo Massari [a,1], Girish Prayag [c,1]

[a] Dipartimento di Scienze Sociali ed Economiche, Sapienza University of Roma, P.le Aldo Moro 5, 00185 Roma, Italy
[b] Faculty of Economics and Management, Free University of Bozen-Bolzano, Universitätsplatz 1 - piazza dell'Università, 1, 39100 Bozen-Bolzano, Italy
[c] Department of Management, Marketing and Entrepreneurship, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

ARTICLE INFO

*Article history:*
Received 26 June 2014
Received in revised form 22 September 2014
Accepted 24 October 2014
Available online 4 November 2014

*Keywords:*
Bagged clustering
Fuzzy C-means
Fuzzy data
Chinese consumers
Likert-type scales

ABSTRACT

Segmentation has several strategic and tactical implications in marketing products and services. Despite hard clustering methods having several weaknesses, they remain widely applied in marketing studies. Alternative segmentation methods such as fuzzy methods are rarely used to understand consumer behaviour. In this study, we propose a strategy of analysis, by combining the Bagged Clustering (BC) method and the fuzzy C-means clustering method for fuzzy data (FCM-FD), i.e., the Bagged fuzzy C-means clustering method for fuzzy data (BFCM-FD). The method inherits the advantages of stability and reproducibility from BC and the flexibility from FCM-FD. The method is applied on a sample of 328 Chinese consumers revealing the existence of four segments (Admirers, Enthusiasts, Moderates, and Apathetics) of the perceived images of Western Europe as a tourist destination. The results highlight the heterogeneity in Chinese consumers' place preferences and implications for place marketing are offered.



## 1. Introduction

Segmentation is critical for developing customer centric marketing and tourism strategies. Effective segmentation leads to competitive advantage, recognition and exploitation of new market opportunities, selection of the appropriate target market, enhanced differentiation and positioning, and increased profitability [1]. Despite the strategic and tactical benefits of market segmentation, there is much controversy surrounding the most commonly used methods and algorithms to segment consumer markets. Cluster analysis remains the most popular method [2–4]. The basic idea of cluster analysis is to divide a heterogeneous consumer market into homogeneous sub-groups [5]. This approach is typically representative of data driven segmentation methods [2]. Cluster analysis has been criticized for its overestimation of the validity of the segmentation results [2] and the resulting clusters have been termed "convenient fictions" [6], a marketing term that refers to the fact that no "natural groupings" could exist, and some information is inevitably lost when objects are grouped. Information loss is not problematic *per se*, but it can result in the wrong conclusions [7]. Every clustering algorithm has advantages and drawbacks and has to be chosen with awareness of its characteristics and limitations [1,2].

Clustering methods are generally split into three groups: non-overlapping (hard), overlapping, and fuzzy algorithms. In hard algorithms, each element to be grouped belongs to a single segment [1]. In overlapping algorithms, an object may belong to more than one cluster [4]. Similar to overlapping algorithms, fuzzy algorithms allow objects to belong to more than one cluster and, in addition, assign to each object a degree of membership to each segment [1,7].

Hierarchical (agglomerative) and non-hierarchical (iterative partitioning) methods are two common hard algorithms that permeate the marketing and tourism literature [1,2,8]. Ward's method remains popular among agglomerative hierarchical algorithms [1,2]. However, hierarchical methods have some drawbacks. Hierarchical methods typically become difficult with increasing sample sizes [2]. The application of hierarchical methods is not always justified in market segmentation given that it presupposes an underlying hierarchy among the objects or respondents to be clustered [1].

Among partitioning methods, *k*-means remains the most popular in marketing and tourism studies [1,2] but it suffers from: (1) identifying equally sized clusters, when in reality such patterns

* Corresponding author at: Faculty of Economics and Management, Free University of Bozen-Bolzano, Universitätsplatz 1 - piazza dell'Università, 1, 39100 Bozen-Bolzano, Italy. Tel.: +39 0471 013293; fax: +39 0471 013009.

*E-mail addresses:* pierpaolo.durso@uniroma1.it (P. D'Urso), marta.disegna@unibz.it (M. Disegna), riccardo.massari@uniroma1.it (R. Massari), girish.prayag@canterbury.ac.nz (G. Prayag).

[1] They all have contributed equally to the work.

http://dx.doi.org/10.1016/j.knosys.2014.10.015
0950-7051/

rarely exist in empirical data; (2) the clustering solution is dependent on the starting solution, and the possibility of building a marketing strategy based on weak data analysis is high; (3) the outcome of cluster analysis is much dependent on the characteristics of the data set, but such characteristics are not always accounted for; (4) repeated computations typically lead to different grouping of respondents, suggesting that solutions may be unreproducible; (5) the lack of published rules about how large the sample size should be in relation to the number of segmentation variables used leads to deceptive and uncritical partitioning exercises [3,8,9].

More recently, "ensemble methods" [10], such as the voting approach [11] and Bagged Clustering (BC) [12], have been successfully applied to enhance the performance of unstable or weak clustering algorithms. The voting algorithm combines partitions sequentially two at a time, obtaining a fuzzy partition of the data. The key idea of the BC is to repeatedly run a partitioning algorithm (like the *k*-means) on the entire sample, and then to combine the results through a hierarchical algorithm. This ensemble method is able to avoid the problem of local minima of partitioning algorithms and find a partition not affected by the randomness of initialization or the clustering process itself [11,12].

Fuzzy procedures allow the assignment of units to each cluster with a membership degree, relaxing the assumption of exclusiveness. A respondent can belong to several clusters without negatively impacting on the managerial usefulness. Conceptually, one consumer's higher statistical probability to belong to one segment does not necessarily mean that s/he only belongs to this segment [13]. For example, a tourist may well desire more than one attribute or benefit from a destination and hence can belong to multiple groups [14]. Among the different fuzzy clustering methods present in the literature, fuzzy *C*-means (FCM) [15] is the most popular. FCM has several advantages in comparison to hard *k*-means. In particular, FCM is less affected by local optima, and is computationally more efficient [16,17].

In the process of choosing the best algorithm, it is important for the researcher to understand that clustering performance depends strongly on the characteristics of the data to be segmented. In tourism research, information regarding attitudes, emotions, satisfaction, and other aspects that involve personal judgement is commonly captured through qualitative expressions, such as Likert scales. However, this approach has been criticized [18,19]. Using a Likert scale the researcher tries to capture a human feeling, by definition uncertain, vague, and subjective, through a linguistic expression. Therefore two important drawbacks arise: first, this type of scale entails a source of vagueness and uncertainty in evaluation since it represents subjective knowledge [20,21]; second, respondent must automatically convert an opinion on a scale and this conversion can distort the original opinion that had to be captured [22]. One way to overcome these drawbacks is to transform Likert variables into fuzzy numbers [16,23]. In the tourism field there are relevant applications of this type of transformation (see e.g., [20–22,24–27]).

In this study we propose a novel clustering method, the Bagged fuzzy *C*-means clustering method for fuzzy data (BFCM-FD), which is an ensemble method that combines BC and fuzzy *C*-means clustering method for fuzzy data (FCM-FD) to derive market segments. Note that FCM-FD inherits from FCM all the advantages above illustrated, in the case of fuzzy data. This clustering method is illustrated by analysing a sample of Chinese travellers perceptions of the image of Western Europe as a tourist destination.

The contribution of this study is threefold. First, in order to capture the ambiguity and uncertainty arising from the use of a Likert scale we propose the transformation of destination image attributes into fuzzy numbers before conducting the segmentation analysis. Second, we propose the adoption of the novel BFCM-FD that combines the strengths of BC and FCM-FD. The proposed method is less sensitive to the number and type of variables used in the clustering, inheriting this property from the BC method [28]. Furthermore, the method inherits the favourable characteristics of the FCM method mentioned earlier. In particular, the proposed method allows the attribution of a unit to more than one cluster, which is often more realistic than assigning a unit to only one cluster in tourism. Third, image segmentation studies in tourism rely heavily on cluster analysis to understand tourists' perceptions of destinations [29–31] but the stability and reproducibility of the identified clusters remain questionable. By using BFCM-FD, we obtain clusters which are stable and reproducible.

The paper is organized as follow. In Section 2, a review of the literature on destination image and image segmentation, as well as a review of Chinese travellers images of Western Europe. Section 3 describes the method used to collect the data and the survey instrument. Section 4 illustrates the various stages of data analysis, including the transformation of the Likert variables into fuzzy numbers, while Section 5 summarizes the results. Section 6 presents both the theoretical and managerial implications, while the paper concludes in Section 7 by offering some final remarks.

## 2. Theoretical background

### 2.1. Destination image

Destination image has been the subject of considerable academic interest in the last four decades. There is no accepted definition of destination image [19,32] but the literature converges around image being both a personal and social construction [33–35]. For the purpose of this study, we focus on the personal construction of destination image and define it as the sum of beliefs, ideas, and impressions that a person has of a destination [36]. Destination image is constructed on the basis of a few selected impressions among a flood of impressions [37], which may include prejudice, imaginations and emotional thoughts [38]. Destination image has direct effects on pre, during and post trip tourist behaviour [35] and has been studied from three perspectives: image components, competitive analysis, and segmentation [19,32,39,40]. Studies on the image components generally conclude a tri-component structure (cognitive, affective, and conative) prevails, whereby the cognitive component influences the affective and conative [19,41]. Alternatively, Baloglu and McCleary [42] suggest that the cognitive and affective components contribute to form an overall image of destination, also known as the composite image [39]. Likewise, Echtner and Ritchie [43] suggest a three-dimensional image model of common/unique, functional/psychological, and holistic/attribute-based that fits the multiple-attribute measurement approach commonly used in tourism studies. More recently, Lai and Li [44] propose a two dimensional model of core and periphery structure of destination image that highlights the complex, pluralistic, and constructed nature of mental structures. This approach confirms that destination image is complex, relativistic, dynamic and of multiple nature [19,45]. The second perspective of competitive analysis seeks to identify the image of a destination vis-à-vis its competitors [19,32,39] and assesses destination competitiveness [46]. Typically, a list of destination attributes is evaluated for one or more competitors and recommendations for image positioning are offered [41,47,48].

#### 2.1.1. Image segmentation

The third perspective, image segmentation is the focus of our study. Within image segmentation studies, two distinct approaches exist: *a priori* (e.g., [49,37,50]) and post hoc (e.g.,