

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

Pierpaolo D'Urso^a, Marta Disegna^{b,*}, Riccardo Massari^a, Girish Prayag^c

^a*Dipartimento di Scienze Sociali ed Economiche, Sapienza University of Roma, P.le Aldo Moro 5, 00185 Rome, 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*

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.

Keywords: Bagged Clustering, Fuzzy C -means, Fuzzy data, Chinese consumers, Likert-type scales.

*Corresponding author

Email addresses: pierpaolo.durso@uniroma1.it (Pierpaolo D'Urso), marta.disegna@unibz.it (Marta Disegna), riccardo.massari@uniroma1.it (Riccardo Massari), girish.prayag@canterbury.ac.nz (Girish Prayag)

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 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., 37, 49, 50] and post-hoc [e.g., 29, 31, 51]. The segmentation of images is not only prevalent in tourism studies but has also been related to benefit segmentation. Researchers have used the push and pull attributes of

destinations for benefit segmentation [52]. While the push attributes generally refer to motives for travel, the pull attributes are related to the features, attractions, and other attributes of the destination itself [53]. The pull attributes have also been described as images [42, 54] and hence, segmented in an effort to evaluate perceptions of place [e.g., 14, 30, 55].

Existing studies on segmentation of destination images or pull attributes have relied on a variety of techniques including, cluster analysis [31, 56], factor-cluster analysis [29, 30, 55], factor analysis and *t*-test/ANOVA [51], factor analysis and regression [57], canonical correlation analysis [14], and multi-dimensional scaling [48] amongst many others. The use of a factor-cluster approach for segmentation in general has been heavily criticized [see 1, 58]. The *k*-means algorithm is prevalent in many of the studies utilizing cluster analysis but its limitations are rarely acknowledged. The use of a structured list of destination attributes for measurement purposes is also heavily criticized [18, 50]. Stepchenkova and Mills [45] highlight the need for newer methods to understand destination image. Using the concept of image diversity (richness, evenness and dominance indices), Stepchenkova and Li [50] explored inter-group perceptions of image based on qualitative information. While this approach is certainly useful, the value of existing segmentation approaches can be enhanced by borrowing ideas from other fields including machine learning and knowledge discovery, computational intelligence, pattern recognition, fuzzy sets, and Bayesian techniques to improve stability, robustness, and reproducibility of quantitative methods of segmentation [3].

2.2. Chinese Travellers Images of Western Europe

Understanding Chinese outbound tourists' expectations and perceptions of the west is still in its infancy [59, 60]. Chinese travel in Europe remains well ahead of economic growth, with the majority of European destinations reporting double digit increases in terms of arrivals and overnight stays [61]. While the number of arrivals is growing, Europe's share of the Chinese outbound travel market is slightly but steadily declining. A better understanding of the profile and needs of Chinese travellers, together with a critical review of legal and cultural barriers to travel are necessary to tap into this market [62]. In particular, understanding the perceived image of Europe among potential Chinese travellers is necessary for effective destination marketing. Limited studies have been devoted to understanding the image of Western Europe and/or individual countries within this region [63]. In fact, Cai et

al. [64] meta review of the Chinese outbound travel market confirms that Europe is an under-researched context. In contrast, a prolific trade literature on the Chinese market has emerged in recent years from various sources (e.g., VisitBritain, Euromonitor, European Travel Commission, TUI Think Tank, Financial Times, etc.), but does not necessarily assess perceived images of the region but rather focuses on perceptions of individual countries.

A survey of Chinese middle income outbound tourists by the Financial Times in 2012, found the most popular intended travel destination in 2013 would be France, UK, Italy, Germany, and Switzerland. France and Italy are closely associated with romance and lifestyle while Germany is perceived as the gateway to Europe. The Chinese outbound market highly values scenic beauty, safety, value for money, infrastructure, quality food, and quality accommodation when travelling to western countries [65]. For a pleasure trip, they also typically like to visit famous attractions, experience different cultures, and obtain good service in hotels/restaurants [66]. Food related attributes such as variety and diversity of food and tourists' own food culture have an impact on Chinese tourists' evaluations of their travel dining experience [67] and they expect convenient transportation and opportunities for shopping [60].

3. The case study

3.1. Data Collection and Sample Characteristics

Data for this study were collected from a survey in Beijing as part of a larger study on Chinese perceptions of Western Europe. Beijing is known for its high propensity to travel and its trend setting status in lifestyle [68]. Two trained interviewers approached potential respondents outside convenient locations such as high street shopping centers, leisure centers, tourist attractions, and local universities, following a procedure adapted from the study of Hsu et al. [68] on the Chinese market. A screening question, adapted from Li et al. [14] and Pan and Li [40], on the potential respondent intention to travel to Western Europe was used to identify the correct target population of 18 to 44 years old. Intention to travel may not accurately reflect actual behaviour [69], but can be used as a reliable indicator to understand tourism outbound markets [14]. The target population of 18 to 44 years old is not only the largest group, but also has the highest propensity for outbound travel [70]. Within this group, the 30 to 44 years old is a well-educated segment in their prime earning years [71]. The younger segment of

the target population is also more autonomous [65] and is already travelling as students in Western Europe [70]. Likewise, the 21 to 35 years old are well educated and part of an emerging independent travel segment [72]. Hence, the focus on the age group of 18 to 44 years old potentially captures visitors with diverse travel orientations (group vs. independent travel) and perceptions. Of the 600 questionnaires distributed, 328 (54.6%) were usable. The demographic profile of the sample indicates more female (57%) than male respondents, mostly single (60%), less than 26 years old (51%), with some university/college degree (59%) or postgraduate degree (36%), and earning an average monthly income of less than 7,000 Chinese yuan (RMB)(67%). Of the respondents, 54% had a full time job while 43% described themselves as students. Respondents will travel for holiday (84%) and studying purposes (19%) mostly. First-time visitors (75%) to Western Europe would constitute the majority.

3.2. Survey Instrument

To capture Chinese travellers' perceptions, 21 image attributes measured the tourism product generally offered to Chinese travellers such as attractive scenery and natural attractions and cultural/historical attractions, and the more general images of Western Europe such as cities with modern technology, quality accommodation, and quality tourist services [60, 63, 65, 66, 73]. The items were measured on a bipolar 7-point Likert scale anchored on [1] *Offers very little* and [7] *Offers very much*. Socio-demographics, travelling characteristics, and information sources most likely to use to plan a trip to Western Europe were also measured. The questionnaire (available in two languages) was pilot tested in Beijing among 20 respondents from the targeted group. Figure 1 displays the percentage distribution for each image attribute measured. Typically, "attractive scenery and natural attractions" is the perceived image offered the most by Western Europe to Chinese travellers. Attributes such as "Festivals, events, and shows", "Quality shopping experiences", and "Language barriers" are perceived as "lesser" offered by Western Europe.

4. The methodology

Clustering could be subject to several sources of uncertainty concerning, amongst others, the imprecision/vagueness of observed features and the assignment of units to clusters [8, 17].

The first source of uncertainty is related to the ordinary use of linguistic expressions, such as Likert-type scales, to capture individual perceptions and opinions of human feelings in general. The interpretation of the meaning of each linguistic expression is subjective, vague and uncertain, since it depends on the characteristics and personal knowledge of the respondent [16, 20, 21, 74]. By using Likert-type scales respondents are forced to automatically convert their opinions to scores and these conversions can be inaccurate, thus causing loss of information. Therefore, the use of such scales incorporates a certain degree of imprecision, ambiguity, and uncertainty, due to the subjective meaning of each value on a rating scale [16, 20, 21, 75].

Generally it is difficult to manage uncertain and/or vague data through traditional methods [76]. For this reason, fuzzy sets, proposed by [77], are a suitable solution to cope with this source of uncertainty [78]. A fuzzy set is defined by a function that assigns to each unit a membership degree. This membership degree indicates how much the unit is close, similar, or compatible with the concept expressed by the fuzzy set. Fuzzy numbers are convex and normalized fuzzy sets with a piecewise continuous membership function defined in \mathbb{R} that maps an interval to $[0, 1]$. The use of fuzzy sets and fuzzy numbers has gain attention in the literature mainly for the following reasons: 1) they are able to capture and measure the uncertainty of individual evaluations (20, 23, 74); 2) fuzzy numbers have a very intuitive meaning and it is more comprehensive than other methods [79]; 3) fuzzy sets can better describe complex processes of the real-life than traditional statistical methods [79]; 4) they can be adapted to a wide range of imprecise data due to the richness of the existing fuzzy scales [23, 78, 79]. As a consequence, when Likert-type scales, or any other linguistic variables, are used in a questionnaire it is useful to formalize them in terms of fuzzy numbers, in order to reduce the imprecision/vagueness of the observed data. In fact, when this kind of linguistic expressions are adopted it is not possible to completely eliminate the source of uncertainty they produce, while it is possible to reduce it by considering a range of possible values (i.e. the fuzzy number), rather than only one score, per each individual answer. Note that, in order to avoid this problem, when designing the questionnaire a fuzzy rating scale could be adopted [23, 80, 81].

The second source of uncertainty can be coped with by adopting a fuzzy clustering approach. Units are assigned to each cluster with a membership degree, ranging between 0 and 1. The greater the membership degree of the unit to a given cluster, the greater is the confidence in assigning the

unit to that cluster. This approach has the advantage of capturing the vague (fuzzy) behaviour of particular units [82]. This is not unreasonable in market segmentation given that customers may share some characteristics to more than one segment [83]. Hence, assigning a customer to only one cluster entails a loss of information [84]. In addition, a fuzzy clustering algorithm is less affected by local optima issues [16], in comparison to hard algorithms. It is also computationally more efficient, since significant changes of cluster membership rarely occur in the classification procedure [17].

4.1. Fuzzification of the Likert variables

As underlined in the previous section, the image attributes observed in our case study are measured on bipolar 7-point Likert scales, therefore the recoding of these variables into *LR* fuzzy variables is necessary. In general, a class of trapezoidal *LR* fuzzy data can be defined as follows:

$$\tilde{\mathbf{X}} \equiv \{\tilde{x}_{ij} = (m_{1ij}, m_{2ij}, l_{ij}, r_{ij})_{LR} : i = 1, \dots, N; j = 1, \dots, J\}, \quad (1)$$

where $\tilde{x}_{ij} = (m_{1ij}, m_{2ij}, l_{ij}, r_{ij})_{LR}$ denotes the *LR* fuzzy variable j observed on the i th object; m_{1ij} and m_{2ij} indicate the left and right centers; l_{ij} and r_{ij} represent the left and right spreads.

In this study we have adopted the fuzzy coding suggested by Kazemifard et al. [85]. Figure 2 represents the corresponding membership function.

It is important to remark that (1) the choice of the fuzzy coding of the Likert-type scales and (2) the analysis of the robustness and stability of the results obtained from a fuzzy data analysis are two important research topics widely discussed in the literature [80].

Moreover, the *elicitation* and *specification* of the membership functions are two critical issues in the conversion of Likert-type scales into fuzzy data.

- *Elicitation*: the choice of the membership functions is subjective and must be determined by experts in the specific problem area since the membership functions are context-sensitive. The choice of the membership function should be made in such a way that the function is able to capture the individual judgements of the person involved. “Suppose that an interviewer asks how a person judges her (his) health. [...] If a person is optimistic and has never had considerable health diseases, it is plausible that she (he) feels ‘very well’. The term ‘very well’ can be fuzzified as a number in [0.85, 1.00] in the scale [0,1]. Conversely, another optimistic person who recently had a critical surgery operation,

may still answer ‘very well’ but now the term could be fuzzified as a fuzzy number in $[0.60, 0.75]$. Similarly, if a person says ‘wait for me about 10 minutes’, the fuzzification of ‘about 10’ may depend on the nationality. Specifically, this usually means from 5 to 15 minutes, but, for instance, if she (he) is Italian, the appropriate fuzzy coding could be from 10 minutes to half an hour or more. Therefore, if possible, the process of fuzzification should be constructed ad hoc of each person to be analysed. Unfortunately, it is sometimes hard to adopt an ad hoc fuzzification procedure. In these cases, one has to choose fuzzy numbers such that they capture the approximate reasoning of the whole of the persons involved” [86]. However, to this purpose, we remark that the consistency, robustness, and stability of the results obtained from a fuzzy data analysis and the connected sensitivity analysis are open problems extensively discussed in the literature.

- *Specification*: in the analysis of multivariate fuzzy data, particular attention must be paid to the *specification* of the membership functions since p variables must be simultaneously analysed. In this context, the *conjunctive* and the *disjunctive* approach can be distinguished [87]. In the *conjunctive* approach, the fuzzy relationship defined on the Cartesian product of the reference universes of the J variables is taken into account. In other words, this approach focuses on studying a fuzzy variable which is the resultant of the J original variables. In this perspective, *non interactive* and *interactive* variables can be distinguished. From a statistical point of view, the adoption of the *conjunctive* approach to the multidimensional fuzzy variables involves a specific interest in studying the fuzzy relationship looked at as a “variable” in itself, which could be observed on the N objects. Conversely, in the *disjunctive* approach, the attention is paid on the set of the J “juxtaposed” variables, observed as a whole in the group of N objects. In this case, we have J membership functions and the investigation of the links among the J fuzzy variables is carried out directly on the matrix of fuzzy data concerning the NJ -variate observations [16, 87].

In our empirical analysis, we have adopted a disjunctive approach for the specification of the membership function. In general, the choice of the shape of the membership functions (elicitation) could be carried out prior to the choice of the squared distance measure used in the clustering algorithm. In

our case, the coding of the Likert variables into fuzzy variables was based on (1). The dissimilarity between two units is then measured by comparing the fuzzy data observed on each unit, i.e. considering the distances for the centers and the spreads of the fuzzy data and using a suitable weighting system for such distance components. By considering the i th and i' th units, Coppi et al. [17] proposed the following squared (Euclidean) distance measure that we have adopted in this analysis:

$$d_F^2(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_{i'}) = [w_M^2 (\|\mathbf{m}_{1i} - \mathbf{m}_{1i'}\|^2 + \|\mathbf{m}_{2i} - \mathbf{m}_{2i'}\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{l}_{i'}\|^2 + \|\mathbf{r}_i - \mathbf{r}_{i'}\|^2)], \quad (2)$$

where $\tilde{\mathbf{x}}_i \equiv \{\tilde{x}_{ij} = (m_{1ij}, m_{2ij}, l_{ij}, r_{ij})_{LR}\}$ denote the fuzzy data vector for the i th object; $\mathbf{m}_{1i} \equiv (m_{1i1}, \dots, m_{1ij}, \dots, m_{1iJ})'$, $\mathbf{m}_{2i} \equiv (m_{2i1}, \dots, m_{2ij}, \dots, m_{2iJ})'$, $\mathbf{l}_i \equiv (l_{i1}, \dots, l_{ij}, \dots, l_{iJ})'$ and $\mathbf{r}_i \equiv (r_{i1}, \dots, r_{ij}, \dots, r_{iJ})'$; $\|\mathbf{m}_{1i} - \mathbf{m}_{1i'}\|^2$ and $\|\mathbf{m}_{2i} - \mathbf{m}_{2i'}\|^2$ are the squared Euclidean distances between the left and right centers, respectively; $\|\mathbf{l}_i - \mathbf{l}_{i'}\|^2$ and $\|\mathbf{r}_i - \mathbf{r}_{i'}\|^2$ are the squared Euclidean distances between the left and right spread, respectively; $w_M, w_S \geq 0$ are suitable weights for the center component and the spread component of (2) constrained by the following conditions: $w_M + w_S = 1$ (*normalization condition*) and $w_M \geq w_S \geq 0$ (*coherence condition*) [17].

The squared distance measure (2) is defined by considering only the centers and the spreads of the fuzzy data. As a consequence, the *a priori* choice of the shape of the membership function is not made before the use of (2) and, consequently, before the application of the clustering algorithm. Thus, with respect to the membership function of the fuzzy data, the adopted (squared) distance measure and, consequently, the clustering algorithm are, as it were, of “free shape”.

4.2. The fuzzy C -means algorithm for fuzzy data (FCM-FD)

The fuzzy C -means algorithm for fuzzy data (FCM-FD) introduced by Coppi et al. [17] allows us both to use fuzzy data as segmentation variables and to assign units (on the basis of their membership degree) to more than one cluster at the same time, effectively addressing the two issues mentioned at the beginning of the section. Schematically, the objective function to be minimized is the following:

$$\left\{ \begin{array}{l} \min : \sum_{i=1}^N \sum_{c=1}^C u_{ic}^\alpha d_F^2(\tilde{\mathbf{x}}_i, \tilde{\mathbf{h}}_c) = \\ \sum_{i=1}^N \sum_{c=1}^C u_{ic}^\alpha [w_M^2 (\|\mathbf{m}_{1i} - \mathbf{h}_c^{M_1}\|^2 + \|\mathbf{m}_{2i} - \mathbf{h}_c^{M_2}\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)] \\ s.t. : \sum_{c=1}^C u_{ic} = 1, \quad u_{ic} \geq 0, \\ w_M \geq w_S \geq 0, \quad w_M + w_S = 1 \end{array} \right. \quad (3)$$

where: $\alpha > 1$ is an exponent that controls the fuzziness of the obtained partition; u_{ic} indicates the membership degree of the i th unit in the c th cluster; $d_F^2(\tilde{\mathbf{x}}_i, \tilde{\mathbf{h}}_c)$ represents the suggested dissimilarity measure between the i th unit and the prototype of the c th cluster; the fuzzy vector $\tilde{\mathbf{h}}_c \equiv \{h_{cj} = (h_{cj}^{M_1}, h_{cj}^{M_2}, h_{cj}^L, h_{cj}^R); j = 1, \dots, J\}$ represents the fuzzy prototype of the c th cluster ($c = 1, \dots, C$); $\mathbf{h}_c^{M_1}$, $\mathbf{h}_c^{M_2}$, \mathbf{h}_c^L and \mathbf{h}_c^R represent respectively, the left and right centers and the left and right spreads of the c th fuzzy prototype.

By solving the minimization problem (3) by means of the Lagrangian multiplier method, we obtain the following iterative solutions [17]:

$$u_{ic} = \frac{[w_M^2 (\|\mathbf{m}_{1i} - \mathbf{h}_c^{M_1}\|^2 + \|\mathbf{m}_{2i} - \mathbf{h}_c^{M_2}\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)]^{-\frac{1}{\alpha-1}}}{\sum_{c'=1}^C [w_M^2 (\|\mathbf{m}_{1i} - \mathbf{h}_{c'}^{M_1}\|^2 + \|\mathbf{m}_{2i} - \mathbf{h}_{c'}^{M_2}\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{h}_{c'}^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_{c'}^R\|^2)]^{-\frac{1}{\alpha-1}}}, \quad (4)$$

$$\mathbf{h}_c^{M_1} = \frac{\sum_{i=1}^N u_{ic}^\alpha \mathbf{m}_{1i}}{\sum_{i=1}^N u_{ic}^\alpha}, \quad \mathbf{h}_c^{M_2} = \frac{\sum_{i=1}^N u_{ic}^\alpha \mathbf{m}_{2i}}{\sum_{i=1}^N u_{ic}^\alpha}, \quad \mathbf{h}_c^L = \frac{\sum_{i=1}^N u_{ic}^\alpha \mathbf{l}_i}{\sum_{i=1}^N u_{ic}^\alpha}, \quad \mathbf{h}_c^R = \frac{\sum_{i=1}^N u_{ic}^\alpha \mathbf{r}_i}{\sum_{i=1}^N u_{ic}^\alpha}, \quad (5)$$

$$w_M = \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^\alpha (\|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^\alpha (\|\mathbf{m}_{1i} - \mathbf{h}_c^{M_1}\|^2 + \|\mathbf{m}_{2i} - \mathbf{h}_c^{M_2}\|^2 + \|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)}$$

$$(w_S = 1 - w_M)$$

(6)

A crucial assumption of the clustering model (3) is that the prototypes are of *LR* fuzzy type, inheriting their typology by the observed data [17]. Coppi

et al. [17] remarked that prototypes are weighted means of the observed units, in which the system of weights is provided by the membership degrees. In such way, the extent to which each unit belongs to a given cluster is incorporated in the definition of the prototypes.

FCM-FD has several advantages, inherited by FCM, with respect to standard (hard, or crisp) k -means clustering. [16, 17, 88]. FCM-FD has a minor tendency of hitting local optima. Empirical studies have shown that, due to the stability of the results, fuzzy clustering is an effective starting point for traditional clustering [89]. It is also computationally more efficient, since the convergence is speedy [90], and could be further accelerated by adopting suitable techniques [91].

4.3. The Bagged fuzzy C -means algorithm for fuzzy data (BFCM-FD)

The BC method combines partitioning and hierarchical clustering procedures [12]. Firstly, B bootstrap samples are drawn with replacement from the same data set. Secondly, a partitioning method (chosen by the researcher) is applied to the B bootstrap sample obtaining $(B \cdot C)$ centres, where C is the initial number of centres selected for the partitioning method. Finally, a hierarchical method is applied to the $(B \cdot C)$ centres in order to group the centres and, indirectly, the units closest to them.

Partitioning methods suffer typically from the influence of the random initialization on the local minimum where the algorithm converges. The rationale of BC is that by repeatedly training the partitioning method on new data sets, given by the bootstrap replicates, different solutions are obtained which are independent from random initialization. Then, the “correct” solution is obtained by combining the solutions of the partition method into a new data set on which a hierarchical method is applied [12]. For more technical information regarding this algorithm see [12, 92, 93].

This method has many advantages both respect to the hierarchical and non-hierarchical methods. In particular, as stated in section 1, k -means algorithm suffers of many disadvantages that the BC method overcomes [12, 92, 93]. Firstly, the BC is more stable than the k -means since it depends less on the starting solution; secondly, the final number of clusters is obtained at the end of the entire procedure and the starting value of C does not affect the result; thirdly, the BC method is able to discover the niche segments.

In this section we illustrate how the fuzzy theory is integrated in the BC method, thus creating a novel segmentation method. In practice, we make use of the FCM-FD, illustrated in the previous section, in the partitioning step of

the BC procedure, thus obtaining the Bagged fuzzy C -means algorithm for fuzzy data (BFCM-FD). BFCM-FD procedure can be summarized as follows:

1. first, B bootstrap samples of N units are drawn with replacement from the original fuzzy data matrix $\tilde{\mathbf{X}}$,

$$\tilde{\mathbf{X}}^1, \dots, \tilde{\mathbf{X}}^b, \dots, \tilde{\mathbf{X}}^B;$$

2. then, the FCM-FD algorithm (3) is run on each bootstrap sample, thus obtaining $(B \cdot C)$ fuzzy prototypes:

$$\begin{aligned} & \{\tilde{\mathbf{h}}_1^1, \dots, \tilde{\mathbf{h}}_c^1, \dots, \tilde{\mathbf{h}}_C^1\}, \dots, \\ & \{\tilde{\mathbf{h}}_1^b, \dots, \tilde{\mathbf{h}}_c^b, \dots, \tilde{\mathbf{h}}_C^b\}, \dots, \\ & \{\tilde{\mathbf{h}}_1^B, \dots, \tilde{\mathbf{h}}_c^B, \dots, \tilde{\mathbf{h}}_C^B\}, \end{aligned}$$

where C ($c = 1, \dots, C$) is the number of fuzzy prototypes and $\tilde{\mathbf{h}}_c^b$ is the c th fuzzy prototype of the b th bootstrap sample $\tilde{\mathbf{X}}^b$;

3. all the fuzzy prototypes are arranged in a new data set $\tilde{\mathbf{H}}_{B \cdot C}$, where the b th row represent the fuzzy prototypes detected on the b th bootstrap sample;
4. then, a distance matrix between the fuzzy prototypes in $\tilde{\mathbf{H}}_{B \cdot C}$ is computed, by using the distance for fuzzy data (2);
5. successively, a hierarchical cluster algorithm is run on $\tilde{\mathbf{H}}_{B \cdot C}$, in order to produce a family of partitions of the prototypes. The result is represented with a dendrogram and the best partition of P final clusters is obtained investigating the graphic, or by means of suitable criteria (see below);
6. finally, the membership degree of unit i to each final cluster p ($p = 1, \dots, P$), \hat{u}_{ip} , is obtained selecting the maximum membership degree of the unit to all the fuzzy prototypes classified in the cluster p .

As for the detection of the best partition in the dendrogram, we make use of the Average Silhouette width (I_S) criterion proposed by Rousseeuw [94]. In our context, let us consider a fuzzy prototype $\tilde{\mathbf{h}}_c$ belonging to the generic cluster p , $\tilde{\mathbf{h}}_{c[p]}$. Let the average distance of the c -th prototype to all other prototypes belonging to cluster p be denoted by a_{cp} . Also, let the average distance of this prototype to all prototypes belonging to another cluster p' ($p' \neq p$), be called $d_{cp'}$. Finally, let b_{cp} be the minimum $d_{cp'}$ computed

over p' ($p' \neq p$), which represents the dissimilarity of the prototype $\tilde{\mathbf{h}}_{c[p]}$ to its closest neighbouring cluster. Then, the silhouette width of the prototype is defined as follows:

$$S_c = \frac{b_{cp} - a_{cp}}{\max\{a_{cp}, b_{cp}\}}, \quad (7)$$

where the denominator is a normalization term.

The higher S_c , the better the assignment of c th prototype to the p th cluster. The I_S , defined as the average of S_c over all the prototypes, is:

$$I_S = \frac{1}{B \cdot C} \sum_{c=1}^{B \cdot C} S_c. \quad (8)$$

Finally, the best partition is achieved when the silhouette is maximized, which implies minimizing the intra-cluster distance (a_{cp}) while maximizing the inter-cluster distance (b_{cp}).

4.4. Some remarks about the algorithm

The proposed algorithm inherits the advantages of both FCM-FD and BC shown in sections 4.2 and 4.3. In particular, we can observe that:

- due to BC, final results are stable, robust and the a-priori definition of the number of clusters is not required.
- with respect to traditional BC which makes use of hard k -means, the adoption of FCM-FD in the partitioning step ensures that results are less affected by local optimal and that the convergence at each bootstrap replicates is faster than with hard k -means [16, 17, 95].
- in comparison with BC, BFCM-FD allows units to belong to more than one cluster simultaneously (via FCM-FD). Note that \hat{u}_{ip} is not proper as a membership degree (in a fuzzy approach point of view) since $\sum_{p=1}^P \hat{u}_{ip} \neq 1$. By simply scaling \hat{u}_{ip} by $\hat{u}_i = \sum_{p=1}^P \hat{u}_{ip}$, we could obtain the normalized value $\tilde{u}_{ip} = \hat{u}_{ip} / \hat{u}_i$, such that $\sum_{p=1}^P \tilde{u}_{ip} = 1$. However, for the i th unit \tilde{u}_{ip} and \hat{u}_{ip} differ only by a multiplicative constant. Since we are interested in the ordering induced by the values \hat{u}_{ip} , and since this ordering is invariant with respect to the scaling operation, in the following we make use of \hat{u}_{ip} .

- the complexity of BFCM-FD is due to the combination of several technique. First, bootstrap is applied to FCM-FD, hence multiplying B times the complexity of the partitioning method. The complexity of FCM-FD depends on the number of units, of the fuzzy variables, of the centroids and on the number of iterations. However, the number of iterations is typically small since the convergence is often obtained in the first iteration [90].

To show how the number of units and bootstrap replicates affect the computational time, we have conducted a study based on simulated data sets. We run BFC-FD by increasing the size of the data sets (from 100 to 1000, step 100) and the number of bootstrap (100, 250 and 500). Each data set is composed by two fuzzy variables and has been constructed in such a way that two clusters are generated in each simulation. Results are illustrated in Figure 3. As it can be seen, the computation time seems to be more affected by the number of bootstrap replicates than by the size of the data set.

5. Results

5.1. Identified Clusters

In this section we illustrate the main findings of the BFCM-FD procedure applied to the described sample of Chinese travellers.

As for the parameters of the BFCM-FD procedure, we have set $\alpha = 2$, as usual in the related literature, $C = 10$ and, as described in section 4.2 we have let the algorithm to determine the values of w_M and w_S . Note that setting the value of the fuzzification parameter α is mainly a heuristic issue [88]. Also the value of C is a minor issue for BC model, as illustrated in section 4.3.

The result of the BFCM-FD procedure is presented in Figure 4. The dendrogram (Figure 4, bottom panel) and the best partition of the units are obtained using the Average Silhouette width criterion, described in the previous section. The peaks in the Silhouette series (Figure 4, top panel) suggest that the Chinese travellers can be segmented into two groups followed by the four-clusters solution. In the following, the four-clusters solution will be considered since it allows us to obtain a more precise and detailed characterization of the market segments in comparison to the two-clusters solution.

The weighted mean values of the image attributes are graphically displayed in Figure 5. The weighted mean value of the j th original segmentation variable (\mathbf{x}_j) is calculated as follows:

$$wm_{jp} = \frac{\sum_{i=1}^N x_{ij} \hat{u}_{ip}}{\sum_{i=1}^N \hat{u}_{ip}} \quad (9)$$

The analysis of these values suggest that cluster 4, is a niche cluster ($N = 43$) of “Admirers”. This cluster comprises Chinese travellers who believe more than the other travellers that Western Europe offers all the image attributes considered. However, they rated “festivals, events and shows” and “language barriers” lower and “attractive scenery and natural attractions” higher than the other attributes. At the opposite end, cluster 3 ($N = 82$) groups the “Apathetics”. These are Chinese travellers who perceive more than the other travellers that Western Europe has little to offer on image attributes such as “easy visa procedures”, “quality shopping experiences”, “cities with modern technology”, “festival, events and shows”, and “quality food”. Yet, this group has somewhat positive perceptions of “attractive scenery and natural attractions”, “clean and unpolluted environment” and “friendly attitude towards visitors”. Cluster 1 ($N = 93$), generally has positive perceptions of most attributes with “attractive scenery and natural attractions”, “clean and unpolluted environment”, “safety and security of tourists” and “friendly attitude towards visitors” rated the highest and “language barriers” rated the lowest. Consequently, this cluster was labelled “Enthusiasts”. Finally, cluster 2 ($N = 108$), the largest cluster, grouped travellers who rated most of the image attributes as neither offering much nor offering little. This cluster was named the “Moderates”.

5.2. Cluster Profiling

To further understand the other specific characteristics of the identified clusters, the socio-demographic (gender, age, income etc.) and travel characteristics of a possible trip to Western Europe (purpose, duration, destination, information source) were used to profile the clusters. In Table A.1 (Appendix A) is reported the complete list of the profiling variables with a brief description of each, while Table 1 presents their percentage values in the whole sample and in each cluster identified.

The socio-demographic characteristics reveal that the percentages of women and of people having a University degree or less are higher in clusters 1 (“Enthusiasts”) and 4 (“Admirers”) compared to clusters 2 (“Moderates”) and 3

("Apathetics"). The "Admirers" have the lowest level of income given that this cluster has the highest percentage of respondents who stated their individual monthly income is equal to 7,000 RMB or less. However, the "Admirers", compared to the other groups, have the highest proportion of respondents having a partner and/or children (46.7% are single), are generally older (i.e. more than 26 years old), and the highest proportion of respondents employed full time (62.2%). An examination of the trip characteristics reveals that the "Enthusiasts" have the highest proportion (83.7%) of first-time visitors to Western Europe while the "Moderates" have the lowest (72.9%). The "Admirers" have the highest proportion (59.1%) of travellers preferring to stay in luxury hotels (i.e. hotel from 3 to 5 stars). Furthermore, the "Admirers" have the highest proportion of travellers staying less than two weeks on their next trip (64.4%) compared to the "Enthusiasts" (58.7%) and "Moderates" (62.9%). Regarding the main purpose of travel, the "Admirers" have the highest proportion (93.3%) of respondents travelling for holiday purposes while the "Moderates" have the highest proportion (9.3%) of those travelling for study purposes. The countries that Chinese travellers would most likely to visit on their next trip to Western Europe are France (72.6%), UK (55.5%), Italy (54.6%), Switzerland (53.1%) and Greece (50.3%). However, differences exist between the clusters. For example, the "Enthusiasts" have the highest proportion of respondents wanting to visit UK (64.5%). The "Moderates" have the highest proportion of respondents wanting to visit Netherlands (33.3%) and Spain (45.4%). The "Apathetics" have the lowest proportion of respondents wanting to visit UK (46.3%), Portugal (4.9%), Switzerland (46.3%), Germany (29.3%), Austria (14.6%) and Greece (37.8%). The "Admirers" have the highest proportion of respondents wanting to visit Belgium (20%), Portugal (13.3%), France (88.9%), Switzerland (68.9%), Germany (53.3%) and Greece (66.7%). Finally, regarding the information source that travellers are likely to use to plan their next trip to Western Europe, the "Enthusiasts" (51.6%) and "Admirers" (46.7%) have the highest proportion of respondents that prefer to use a travel agency. The "Moderates" have the highest proportion of respondents that will use guidebooks (44.4%). The "Enthusiasts" (84.9%) and "Apathetics" (76.8%) have the highest proportion of respondents that will use search engines on the Internet but the latter has also the lowest proportion of respondents (35.4%) that will use travel forums and blogs as a source of information.

6. Implications

The main objective of this study was to apply a novel segmentation method, BFCM-FD, to understand Chinese travellers' images of Western Europe. The results reveal the existence of four clusters that can be differentiated on their images and socio-demographic characteristics. The results have both theoretical and managerial implications.

6.1. Theoretical Implications

The most commonly used method (cluster analysis) and algorithm (k -means) for segmenting markets have been heavily criticized [1, 8, 9, 58]. These criticisms pertain mainly to reproducibility of clusters, stability of clusters, sub-optimal procedures in assigning units to clusters, and selecting the number of clusters. BFCM-FD overcomes many of these limitations. BFCM-FD is reproducible, inheriting this feature from BC [12]. In addition, BFCM-FD, inherits from FCM-FD the capability of providing the best performance in stability criterion in comparison to hard methods [96]. BFCM-FD allows respondents to belong to more than one cluster. Hence, it is more robust than hard and overlapped clustering methods. BFCM-FD, identifies the typical member of a segment, the strength of the membership, and the intersection of the segments. Using the BC approach, and therefore also the BFCM-FD method, it is not necessary to impose the number of clusters in advance unlike hard partitioning methods. By adopting FCM-FD, BFCM-FD thus inherits the benefits connected to both fuzzy clustering [88, 97, 98] and fuzzy formalization of imprecise information. Overall, BFCM-FD offers a rigorous, visually simple, and alternative way of segmenting tourism markets and allows for the identification of niche markets.

6.2. Managerial Implications

The results have several managerial implications for tapping into the Chinese outbound market. First, the study confirms increasing heterogeneity in the Chinese outbound market as suggested by others [e.g., 99]. The identified segments perceive Western Europe as offering much of attractive scenery and natural attractions, clean and unpolluted environment, safety and security, and cultural attractions. The positive perceptions of these attributes suggest that projected images of Western Europe in China have set realistic expectations. Europe is marketed in China using the region's rich cultural background and unique landscapes and these attributes should feature

prominently in future marketing campaigns. Similar to other international travellers to Europe, the Chinese outbound market values safety and security. Individual countries within Western Europe should ensure that tourists feel safe. Chinese tourists are attracted by the perceived “cleanliness” of Europe compared to China and the region’s perceived pristine environment can be effectively used for destination advertising and promotion in China. Overall these attributes contribute to a positive destination image that may influence destination preference, tourists’ intention to visit, and recommendation behaviour [18].

Shopping is an important pull factor of tourist destinations for the Chinese outbound market [55, 60, 100]. However, this attribute is not necessarily a strength of Western Europe. Clusters 2 (“Moderates”) and 3 (“Apathetics”) rated Western Europe as offering less of “quality shopping experiences” in comparison to other clusters. Plausible explanations for this occurrence can be found in studies of Chinese visitors to the United States [60, 100] and Singapore [55]. For example, Li et al. [60] found that Chinese tourists did not want to visit regular shops or undertake “forced” shopping but rather preferred shopping areas with local flavour and the availability of international brands at good prices. Xu and McGehee [100] found that Chinese visitors were disappointed when they found that the international brands bought were made in China or in other Asian countries. However, some visitors were interested in purchasing products made in China but unavailable in China given that such products are perceived to be of higher quality. Both Kau and Lim [55] and Xu and McGehee [100] found that prices and the lack of communication in Chinese language were major sources of dissatisfaction with the shopping experience. Hence, for perceptions of high quality shopping experiences for Chinese visitors, Western Europe service providers must emphasize products made in Europe, offer customer assistance in the Chinese language in shopping malls, and provide value for money. It may also be necessary to create awareness of the shopping infrastructure in Western Europe, timing of sales promotion, and shopping festivals that can contribute to increase the perceived quality of the shopping experience. Signage in shopping malls in Chinese language can also improve the shopping experience [100].

The attribute rated the least favourably by all the segments is “festivals, events and shows”. The new generation of Chinese travellers will not necessarily follow the classic cultural-historical itineraries currently offered in Europe, but more likely to follow personal trails mounted from movies, music, sports or culture, and personal idols. Hence, marketing to the younger

generation of Chinese tourists will require the promotion of festivals, events and shows that are relevant to this generation such as the Cannes Film Festival, shooting location of popular movies, and the home/second home of popular Chinese celebrities. Movies and music are likely to influence Chinese travellers to visit particular countries. Hence, the European Tourism Council mandate of marketing Europe in China should seek to address the “less” favourable perception of “festivals, events and shows” of Western Europe. Li et al. [60] confirm that Chinese visitors to the US are keen to experience local culture and customs through festivals, events and shows. Given that this attribute was rated significantly different by Cluster 4 (“Admirers”) and cluster 3 (“Apathetics”), suggests that potential visitors from the same outbound market may hold very different images of a country/region [56].

Visa requirements continue to be perceived as a significant constraint of travel for the Chinese outbound market [65, 101]. Except for Cluster 3 (“Apathetics”), all the other clusters perceived Western Europe as offering much of easy visa procedures. In the last few years, countries such as France and Germany have taken active steps to ease visa procedures for Chinese tourists but the tough visa requirements of countries such as UK has kept Chinese visitors away, often leading to significant loss in retail revenues. Likewise, three of the four clusters of Chinese visitors perceived Western Europe as offering much of language barriers, meaning that significant language barriers exist.

While variety of food was rated positively by all the segments, Western Europe is perceived as offering much of quality food by only three of the four segments. In general, considerable differences exist between cultures in terms of the perception of attractiveness of food from other cultures [67]. For Chinese consumers, food that is different in taste, culture and quality is fashionable and desirable [102]. However, not all Chinese visitors are eager to try local food. They prefer familiar flavours and cooking methods [103]. The findings also suggest preferences for countries such as France, Italy and UK which conform to findings from the European Travel Commission. Countries such as Spain, Austria, Ireland, and Portugal may need to more pro-actively market to Chinese visitors. These countries have the lowest preference for travel among potential Chinese visitors. Surprisingly, Clusters 1 (“Enthusiasts”) and 4 (“Admirers”) are most likely to use travel agencies as an information source for travel to Western Europe. This contrasts to other studies [e.g., 65] suggesting that Chinese visitors are most likely to gather information from TV programs and friends. However, Cluster 1 (“Enthusi-

asts”) has a high proportion of 18 to 25 years old and they are likely to use the Internet as a source of travel information. Communicating with less experienced travellers is about building product awareness, thereby focusing on the cognitive attributes of a destination, whereas the communication strategy for more experienced travellers requires a focus on the emotional and unique aspects of a destination [50]. Overall, the findings confirm the need for more and more fine-tuned segmentation of the Chinese market. Indeed, different segments of Chinese visitors may hold different images of Western Europe and have different preferences for countries they want to visit. Understanding the expectations of the Chinese outbound market is critical for service provision [60] and identifying the image of a region/country is critical for destination benchmarking and competitiveness analysis [50]. Hence, the results can be used to monitor the evolution of the image of Western Europe in China and assist destination marketers in selecting the appropriate image associations for destination differentiation and positioning purposes.

7. Conclusions

The Chinese outbound market is undoubtedly a growth market for Western Europe. In this study a novel segmentation method, the BFCM-FD, was adopted in order to cluster Chinese visitors by their perceived image of Western Europe as destination. The results suggest that heterogeneity in the perceived image of Western Europe exist and that appropriate marketing strategies for each segment of potential visitors are required and must be implemented. Although, the BFCM-FD method overcomes many of the limitations of traditional hard (both hierarchical and non-hierarchical) algorithms, this study is not without limitations.

First, as many clustering methods available in the literature, the BFCM-FD procedure could be afflicted by outlier sensibility. In particular, FCM-FD algorithm is sensible to the presence of outliers in the data [104]. In order to neutralize and smooth the disruptive effects of possible outliers in the BFCM-FD clustering procedure it would be useful to consider robust fuzzy clustering algorithms for fuzzy data [see, e.g., 95].

Second, the sample of Chinese visitors was identified from Beijing and the results are pertinent to the outbound market from this city only. Research on the Chinese outbound market [14, 59, 60] suggests that Shanghai and Guangzhou are also important generating markets and this study should be replicated in other locations within China. Third, the list of image at-

tributes has mostly cognitive images. Stepchenkova and Li [50] suggest that the Chinese outbound market is driven by cognitive and affective images, with less experienced travellers associating mainly cognitive images with a destination. Hence, it would be worthwhile for future studies to extend the list of attributes for Western Europe to include affective images and also to assess the image of individual countries such as Spain, Italy and France for competitive analysis. Fourth, this study focused on segmenting destination image but there is a need for novel qualitative and quantitative methods to understand the image construct and competitive images of destinations [44, 50]. Quantitative methods such as Bray-Curtis dissimilarity index and Electre II methods may be useful in achieving that [46].

As a final remark, in the specific case of imprecision associated with the use in the evaluation process of linguistic term-based scales we can distinguish two fuzzy approaches: (a) *fuzzy rating scale* [80, 105]; (b) *Likert or associated fuzzy conversion scales* [80]. Indeed, since nowadays an increasing amount of data arises from human judgements, perceptions and evaluations, it seems necessary to find suitable instruments able to handle and capture the imprecision contained in them. In this study, a fuzzy transformation of the Likert-type variables is proposed, i.e. a fuzzy version of Likert-type scales is employed. In this regards, we remark that an interesting proposal in the tourism context have been suggested by Benítez, et al. [20]. However, in general, robustness and stability of the results obtained from a fuzzy data analysis are still open problems. We will investigate in deep this important topics in our future studies.

- [1] M. N. Tuma, R. Decker, S. W. Scholz, A survey of the challenges and pitfalls of cluster analysis application in market segmentation, *International Journal of Market Research* 53 (3) (2011) 391–414.
- [2] S. Dolnicar, A review of data-driven market segmentation in tourism, *Journal of Travel & Tourism Marketing* 12 (1) (2002) 1–22.
- [3] A. K. Jain, Data clustering: 50 years beyond k -means, *Pattern Recognition Letters* 31 (2010) 651–666.
- [4] M. Wedel, W. A. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, 2nd Edition, *Archives of Psychology*, Dordrecht: Kluwer, 2000.
- [5] G. Punj, D. W. Stewart, Cluster analysis in marketing research: review and suggestions for application, *Journal of Marketing Research* 20 (2) (1983) 138–148.
- [6] T. Babinec, Editorial: Some musings on cluster analysis, *Journal of Targeting, Measurement and Analysis for Marketing* 10 (4) (2002) 301–303.
- [7] N. Franke, H. Reisinger, D. Hoppe, Remaining within-cluster heterogeneity: a meta-analysis of the “dark-side” of clustering methods, *Journal of Marketing Management* 25 (3/4) (2009) 273–293.
- [8] S. Dolnicar, Using cluster analysis for market segmentation—typical misconceptions, established methodological weaknesses and some recommendations for improvement, *Australasian Journal of Market Research* 11 (2) (2003) 5–12.
- [9] S. Dolnicar, S. Kaiser, K. Lazarevski, F. Leisch, Biclustering: overcoming data dimensionality problems in market segmentation, *Journal of Travel Research* 51 (1) (2012) 41–49.
- [10] A. Strehl, J. Ghosh, Cluster ensembles – a knowledge reuse framework for combining multiple partitions, *Journal of Machine Learning Research* 3 (2002) 583–617.
- [11] E. Dimitriadou, A. Weingessel, H. Kurt, A combination scheme for fuzzy clustering, *International Journal of Pattern Recognition and Artificial Intelligence* 16 (7) (2002) 901–912.

- [12] F. Leisch, Bagged clustering, Working paper 51, SFB Adaptive Information Systems and Modelling in Economics and Management Science, WU Vienna University of Economics and Business (1999).
- [13] A. Chaturvedi, J. D. Carroll, P. E. Green, J. A. Rotondo, A feature-based approach to market segmentation via overlapping k -centroids clustering, *Journal of Marketing Research* 34 (3) (1997) 370–377.
- [14] X. Li, F. Meng, M. Uysal, B. Mihalik, Understanding China’s long-haul outbound travel market: An overlapped segmentation approach, *Journal of Business Research* 66 (2013) 786–793.
- [15] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, Norwell, MA, USA, 1981.
- [16] P. D’Urso, Clustering of fuzzy data, in: J. V. De Oliveira, W. Pedrycz (Eds.), *Advances in fuzzy clustering and its applications*, J. Wiley and Sons, 2007, pp. 155–192.
- [17] R. Coppi, P. D’Urso, P. Giordani, Fuzzy and possibilistic clustering for fuzzy data, *Computational Statistics & Data Analysis* 56 (4) (2012) 915–927.
- [18] S. Dolnicar, B. Grün, Validly measuring destination image in survey studies, *Journal of Travel Research* 52 (1) (2013) 3–14.
- [19] M. G. Gallarza, I. G. Saura, H. C. Garcia, Destination image: towards a conceptual framework, *Annals of Tourism Research* 29 (1) (2002) 56–78.
- [20] J. M. Benítez, J. C. Martín, C. Román, Using fuzzy number for measuring quality of service in the hotel industry, *Tourism Management* 28 (2) (2007) 544–555.
- [21] P. D’Urso, L. De Giovanni, M. Disegna, R. Massari, Bagged clustering and its application to tourism market segmentation, *Expert Systems with Applications* 40 (12) (2013) 4944–4956. doi:<http://dx.doi.org/10.1016/j.eswa.2013.03.005>.
- [22] T.-H. Hsu, L.-Z. Lin, Using fuzzy set theoretic techniques to analyze travel risk: an empirical study, *Tourism Management* 27 (2006) 968–981.

- [23] B. Sinova, M. Á. Gil, A. Colubi, S. Van Aelst, The median of a random fuzzy number. the 1-norm distance approach, *Fuzzy Sets and Systems* 200 (2012) 99–115.
- [24] T.-K. Hsu, Y.-F. Tsai, H.-H. Wu, The preference analysis for tourist choice of destination: a case study of taiwan, *Tourism Management* 30 (2009) 288–297.
- [25] L.-Z. Lin, W.-C. Chen, T.-J. Chang, Using FQFD to analyze island accommodation management in fuzzy linguistic preferences, *Expert Systems with applications* 38 (2011) 7738–7745.
- [26] E. Ngai, F. Wat, Design and development of a fuzzy expert system for hotel selection, *Omega* 31 (4) (2003) 275 – 286. doi:[http://dx.doi.org/10.1016/S0305-0483\(03\)00050-1](http://dx.doi.org/10.1016/S0305-0483(03)00050-1).
- [27] Y.-Y. Wu, H.-L. Hsiao, Y.-F. Ho, Urban ecotourism: defining and assessing dimensions using fuzzy number construction, *Tourism Management* 31 (2010) 739–743.
- [28] R. Ghaemi, N. Sulaiman, H. Ibrahim, N. Mustapha, A survey: Clustering ensembles techniques, *World Academy of Science, Engineering and Technology* 29 (2009) 636–645.
- [29] B. Leisen, Image segmentation: the case of tourism destination, *Journal of Services Marketing* 15 (1) (2001) 49–66.
- [30] G. Prayag, Images as pull factors of a tourist destination: A factor-cluster segmentation analysis, *Tourism Analysis* 15 (2) (2010) 213–226.
- [31] G. Prayag, Paradise for who? Segmenting visitors' satisfaction with cognitive image and predicting behavioral loyalty, *International Journal of Tourism Research* 14 (1) (2012) 1–15.
- [32] J. E. Bigné, I. S. Garcia, S. S. Blas, The functional-psychological continuum in the cognitive image of a destination: A confirmatory analysis, *Tourism Management* 30 (6) (2009) 715–723.
- [33] C.-C. Chen, Y.-H. Lin, J. F. Petrick, Social biases of destination perceptions, *Journal of Travel Research* 52 (2) (2012) 240–252.

- [34] Y. Kanemasu, Social construction of touristic imagery: case of Fiji, *Annals of Tourism Research* 43 (2013) 456–481. doi:<http://dx.doi.org/10.1016/j.annals.2013.06.004>.
- [35] A. D. A. Tasci, W. C. Gartner, Destination image and its functional relationships, *Journal of Travel Research* 45 (2007) 413–425.
- [36] J. Crompton, An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image, *Journal of Travel Research* 17 (1979) 18–23.
- [37] P. Fakeye, J. Crompton, Image differences between prospective, first-time, and repeat visitors to the Lower Rio Grande Valley, *Journal of Travel Research* 29 (1991) 10–16.
- [38] F. Lawson, M. Bond-Bovy, *Tourism and Recreational Development*, Architectural Press, London, 1977.
- [39] J. E. Bigné, M. I. Sanchez, J. Sanchez, Tourism image, evaluation variables, and after purchase behaviour: Inter-relationship, *Tourism Management* 26 (6) (2001) 607–616.
- [40] B. Pan, X. Li, The long tail of destination image and online marketing, *Annals of Tourism Research* 38 (1) (2011) 132–152.
- [41] S. Pike, C. Ryan, Destination positioning analysis through a comparison of cognitive, affective and conative perceptions, *Journal of Travel Research* 42 (2004) 333–342.
- [42] S. Baloglu, K. W. McCleary, A model of destination image formation, *Annals of Tourism Research* 26 (4) (1999) 868–897.
- [43] C. M. Echtner, J. R. B. Ritchie, The meaning and measurement of destination image, *Journal of Tourism Studies* 2 (2) (2003) 2–12.
- [44] K. Lai, Y. Li, Core-periphery structure of destination image: concept, evidence and implication, *Annals of Tourism Research* 39 (3) (2012) 1359–1137.
- [45] S. Stepchenkova, J. E. Mills, Destination image: A meta-analysis of 2000–2007 research, *Journal of Hospitality Marketing & Management* 19 (2010) 575–609.

- [46] L. Andrades-Caldito, M. Sánchez-Rivero, J. I. Pulido-Fernández, Differentiating Competitiveness through Tourism Image Assessment: An Application to Andalusia (Spain), *Journal of Travel Research* 52 (1) (2013) 68–81.
- [47] M. F. Cracolici, P. Nijkamp, The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions, *Tourism Management* 30 (3) (2009) 336–344.
- [48] W. C. Gartner, Tourism image: Attribute measurement of state tourism products using multidimensional scaling techniques, *Journal of Travel Research* 28 (2) (1989) 16– 20.
- [49] C. B. Castro, E. M. Armario, D. M. Ruiz, The influence of market heterogeneity on the relationship between a destination’s image and tourists’ future behaviour, *Tourism Management* 28 (1) (2007) 175–187.
- [50] S. Stepchenkova, X. Li, Chinese Outbound Tourists’ Destination Image of America: Part II 51 (6) (2012) 687–703.
- [51] A. Beerli, J. D. Martin, Factors influencing destination image, *Annals of Tourism Research* 31 (3) (2004) 657–681.
- [52] D.-B. Park, Y.-S. Yoon, Segmentation by motivation in rural tourism: A Korean case study, *Tourism Management* 30 (1) (2009) 99–108.
- [53] D. B. Klenosky, The “pull” of tourism destinations: A means-end investigation, *Journal of Travel Research* 40 (4) (2002) 396–403.
- [54] G. Prayag, C. Ryan, The relationship between the push & pull attributes of a tourist destination: The role of nationality. an analytical qualitative research approach, *Current Issues in Tourism* 14 (2) (2011) 121–143.
- [55] A. K. Kau, P. S. Lim, Clustering of chinese tourists to singapore: an analysis of their motivations, values and satisfaction, *International Journal of Tourism Research* 7 (4-5) (2005) 231–248.
- [56] G. Prayag, S. Hosany, When Middle East meets West: Understanding the motives and perceptions of young tourists from United Arab Emirates, *Tourism Management* 40 (2014) 35–45.

- [57] S. Jang, C.-M. E. Wu, Seniors' travel motivation and the influential factors: An examination of Taiwanese seniors, *Tourism Management* 27 (2) (2006) 306–316.
- [58] S. Dolnicar, B. Grün, Challenging factor-cluster analysis, *Journal of Travel Research* 47 (Aug) (2008) 63–71.
- [59] X. Li, R. Harrill, M. Uysal, T. Burnett, X. Zhan, Estimating the size of the Chinese outbound travel market: A demand-side approach, *Tourism Management* 31 (2) (2010) 250–259.
- [60] X. Li, C. Lai, R. Harrill, S. Kline, L. Wang, When East meets West: An exploratory study on Chinese outbound tourists' travel expectations, *Tourism Management* 32 (4) (2011) 741–749.
- [61] European Travel Commission, European tourism needs joint efforts to stimulate growth, 2013, press Release on 24.07.13 by ETC, Brussels, Belgium.
- [62] European Travel Commission, Europe: A unique destination for Chinese travellers, 2012, press Release on 27.11.12 by ETC, Brussels, Belgium.
- [63] M. A. Corigliano, The outbound Chinese tourism to Italy: The new graduates' generation, *Journal of China Tourism Research* 7 (4) (2011) 396–410.
- [64] L. A. Cai, M. Li, B. J. Knutson, Research on China Outbound Market: A Meta-Review, *Journal of Hospitality & Leisure Marketing* 16 (1-2) (2008) 5–20.
- [65] B. Sparks, G. W. Pan, Chinese outbound tourists: Understanding their attitudes, constraints and use of information sources, *Tourism Management* 30 (4) (2009) 483–494.
- [66] X. Yu, B. Weiler, Mainland Chinese pleasure travellers to Australia: A leisure behaviour analysis, *Tourism, Culture & Communication* 3 (2001) 81–91.
- [67] R. C. Y. Chang, J. Kivela, A. H. N. Mak, Attributes that influence the evaluation of travel dining experience: When East meets West, *Tourism Management* 32 (2) (2011) 307–316.

- [68] C. H. C. Hsu, L. A. Cai, M. Li, Expectation, motivation, and attitude: A tourist behavioral model, *Journal of Travel Research* 49 (2010) 282–296.
- [69] B. McKercher, T. S. M. Tse, Is intention to return a valid proxy for actual repeat visitation?, *Journal of Travel Research* 51 (6) (2012) 671–686.
- [70] Euromonitor, Impact of Chinese travellers in Europe, 2011, report by Euromonitor International.
- [71] T. S. M. Tse, J. S. P. Hobson, The forces shaping China’s outbound tourism, *Journal of China Tourism Research* 4 (2008) 136–155.
- [72] G. Chen, J. Bao, S. S. Huang, Segmenting chinese backpackers by travel motivations, *International Journal of Tourism Research* Article first published online: 8 JAN 2013. doi:10.1002/jtr.1928.
- [73] S. S. Kim, Y. Guo, J. Agrusa, Preference and positioning analyses of overseas destinations by mainland Chinese outbound pleasure tourists, *Journal of Travel Research* 44 (2) (2005) 212–220.
- [74] R. Coppi, P. D’Urso, Fuzzy k -means clustering models for triangular fuzzy time trajectories, *Statistical Methods and Applications* 11 (2002) 21–24.
- [75] H.-K. Chiou, G.-H. Tzeng, D.-C. Cheng, Evaluating sustainable fishing development strategies using fuzzy MCDM approach, *Omega* 33 (3) (2005) 223 – 234. doi:http://dx.doi.org/10.1016/j.omega.2004.04.011.
- [76] T.-Y. Chou, C.-L. Hsu, M.-C. Chen, A fuzzy multi-criteria decision model for international tourist hotels location selection, *International Journal of Hospitality Management* 27 (2008) 293–301.
- [77] L. Zadeh, Fuzzy sets, *Information and control* 8 (1965) 338–353.
- [78] Y. Wang, X. Ma, Y. Lao, Y. Wang, A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization, *Expert Systems with Applications* 41 (2) (2014) 521–534.

- [79] B. Sohrabi, I. R. Vanani, K. Tahmasebipur, S. Fazli, An exploratory analysis of hotel selection factors: A comprehensive survey of Tehran hotels, *International Journal of Hospitality Management* 31 (2012) 96–106.
- [80] S. De la Rosa de Sàa, M. Gil, G. Gonzalez-Rodriguez, M. T. López, M. Lubiano, Fuzzy rating scale-based questionnaires and their statistical analysis, *IEEE Transactions on Fuzzy Systems* (2014). doi: [10.1109/TFUZZ.2014.2307895](https://doi.org/10.1109/TFUZZ.2014.2307895).
- [81] M. A. Gil, G. González-Rodríguez, *Fuzzy vs. Likert Scale in Statistics*, in: E. Trillas, P. P. Bonissone, L. Magdalena, J. Kacprzyk (Eds.), *Combining Experimentation and Theory*, Vol. 271 of *Studies in Fuzziness and Soft Computing*, Springer, 2012, pp. 407–420.
- [82] R. Kruse, C. Döring, M.-J. Lesot, Fundamentals of fuzzy clustering, in: J. V. de Oliveira, W. Pedrycz (Eds.), *Advances in Fuzzy Clustering and Its Applications*, J. Wiley and Sons, 2007, pp. 3–30.
- [83] H. Hruschka, Market definition and segmentation using fuzzy clustering methods, *International Journal of Research in Marketing* 3 (2) (1986) 117–134.
- [84] W.-Y. Chiang, Establishment and application of fuzzy decision rules: an empirical case of the air passenger market in Taiwan, *International Journal of Tourism Research* 13 (2011) 47–456.
- [85] M. Kazemifard, A. Zaeri, N. Ghasem-Aghaee, M. A. Nematbakhsh, F. Mardukhi, Fuzzy Emotional COCOMO II Software Cost Estimation (FECSCCE) using Multi-Agent Systems, *Applied Soft Computing* 11 (2) (2011) 2260–2270.
- [86] R. Coppi, P. Giordani, P. D’Urso, Component models for fuzzy data, *Psychometrika* 71 (4) (2006) 733–761.
- [87] R. Coppi, *The fuzzy approach to multivariate statistical analysis*, Tech. Rep. 11, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma, Rome, Italy (2003).

- [88] P. D’Urso, Fuzzy clustering, in: C. Hennig, M. Meila, F. Murtagh, R. Rocci (Eds.), *Handbook of Cluster Analysis*, Chapman and Hall, 2014.
- [89] W. J. Heiser, P. J. F. Groenen, Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima, *Psychometrika* 62 (1) (1997) 63–83.
- [90] J. C. Bezdek, R. J. Hathaway, M. J. Sabin, W. T. Tucker, Convergence theory for fuzzy c-means: Counterexamples and repairs, *IEEE Transactions on Systems, Man and Cybernetics* 17 (5) (1987) 873–877.
- [91] C. Borgelt, Accelerating fuzzy clustering, *Information Sciences* 179 (23) (2009) 3985–3997.
- [92] J. G. Brida, M. Disegna, R. Scuderi, Segmenting visitors of cultural events: The case of christmas market, *Expert Systems with Applications* 41 (10) (2014) 4542–4553.
- [93] S. Dolnicar, F. Leisch, Segmenting markets by bagged clustering, *Australasian Marketing Journal* 12 (1) (2004) 51–65.
- [94] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics* 20 (1987) 53–65.
- [95] P. D’Urso, L. De Giovanni, Robust clustering of imprecise data, *Chemo-metrics and Intelligent Laboratory Systems* 136 (2014) 58–80.
- [96] Y. P. Wang, M. Gunampally, J. Chen, D. Bittel, M. G. Butler, W. W. Cai, A comparison of fuzzy clustering approaches for quantification of microarray gene expression, *The Journal of VLSI Signal Processing* 50 (2) (2008) 1–16.
- [97] B. S. Everitt, S. Landau, M. Leese, *Cluster Analysis*, 4th Edition, Wiley Publishing, 2009.
- [98] H. Hwang, W. S. Desarbo, Y. Takane, Fuzzy clusterwise generalized structured component analysis, *Psychometrika* 72 (2) (2007) 181–198.
- [99] C. Ong, H. du Cros, The post–Mao gazes: Chinese backpackers in Macau, *Annals of Tourism Research* 39 (2012) 735–754.

- [100] Y. Xu, N. G. McGehee, Shopping behavior of Chinese tourists visiting the United States: Letting the shoppers do the talking, *Tourism Management* 33 (2) (2012) 427–430.
- [101] C. Lai, X. Li, R. Harrill, Chinese outbound tourists' perceived constraints to visiting the United States, *Tourism Management* 37 (2013) 136–146.
- [102] A. Eves, L. Cheng, Cross-cultural evaluation of factors driving intention to purchase new food products – Beijing, China and South-East England, *International Journal of Consumer Studies* 31 (4) (2007) 410–417.
- [103] R. C. Y. Chang, J. Kivela, A. H. N. Mak, Food preferences of chinese tourists, *Annals of Tourism Research* 37 (4) (2010) 989–1011.
- [104] L. Á. García-Escudero, A. Gordaliza, Robustness properties of k -means and trimmed k -means, *Journal of the American Statistical Association* 94 (447) (1999) 956–969.
- [105] T. Hesketh, R. Pryor, B. Hesketh, An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences, *International Journal of Man-Machine Studies* 29 (1) (1988) 21–35.

Figures

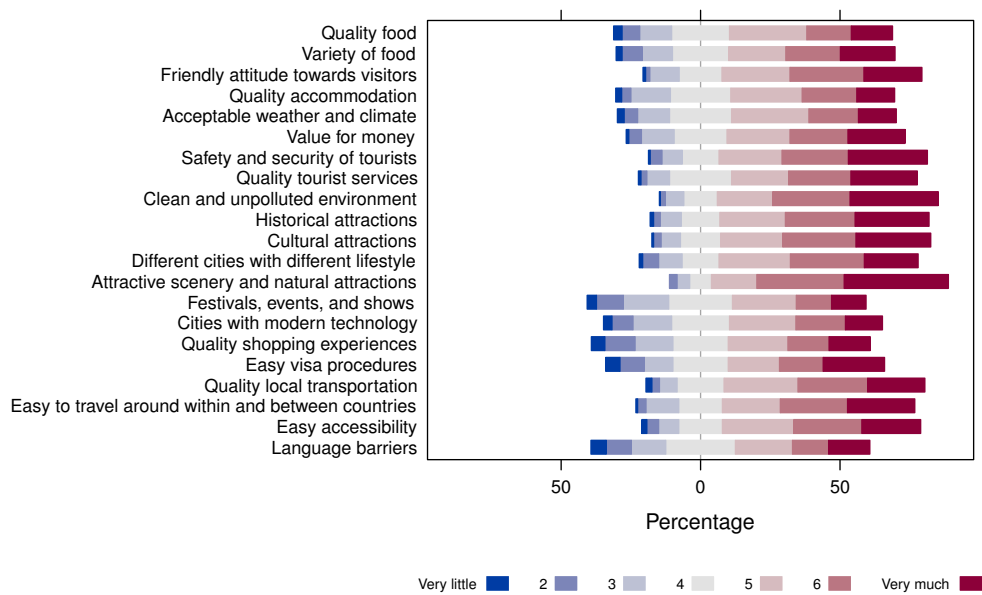


Figure 1: % distribution for each image attribute.

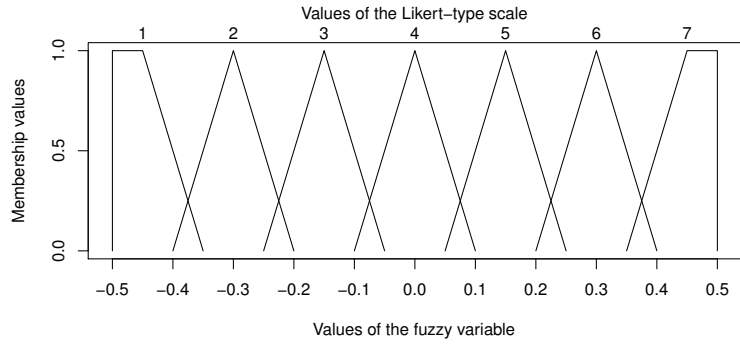


Figure 2: Fuzzy recoding of the 7-points Likert variables

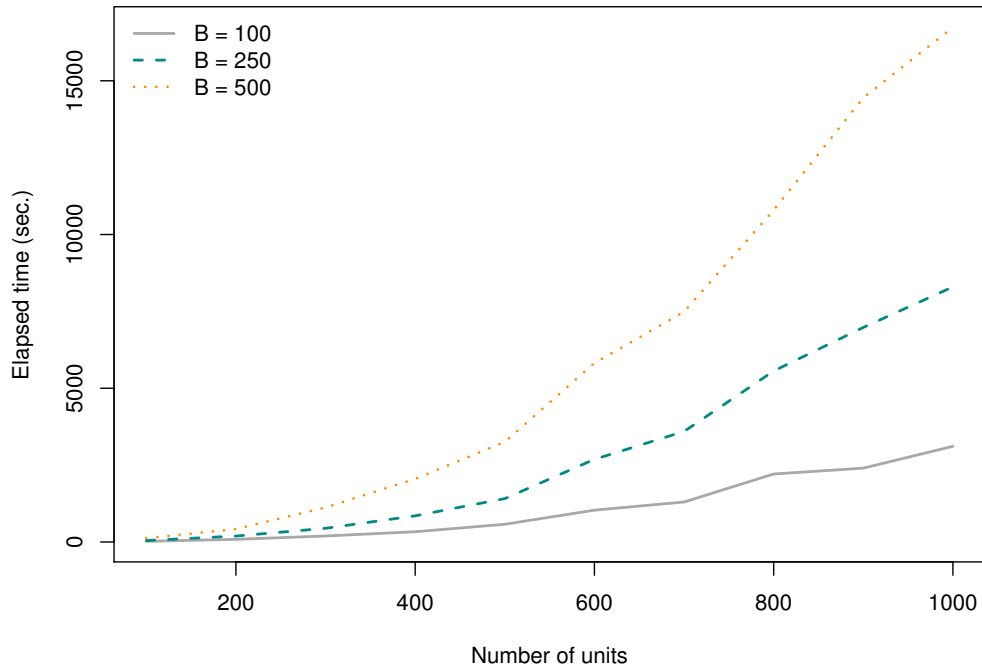


Figure 3: Time complexity for BFCM-FD.

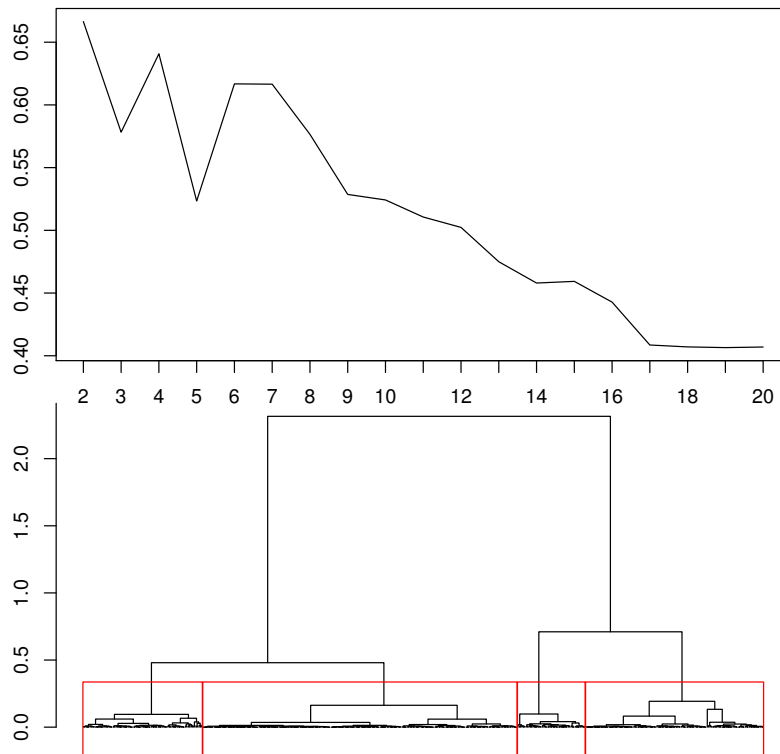


Figure 4: Values of the Silhouette index per each cluster partition from 2 to 20 (top panel) and dendrogram (bottom panel).

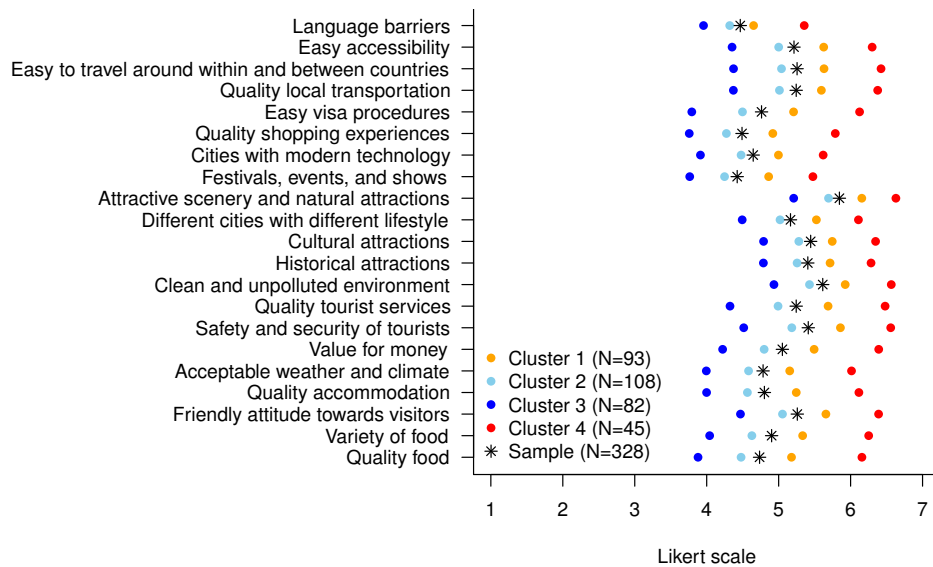


Figure 5: Weighted mean of the segmentation variable, i.e. image attributes.

Tables

Table 1: Characteristics and preferences of the travellers, and characteristics of the trip (percentage values).

Variables	Whole sample	CL1	CL2	CL3	CL4	<i>p</i> -value
<i>Socio-demographic and economic characteristics</i>						
Female	57.32	63.44	54.63	43.90	75.56	***
Individual Monthly Income	67.08	64.13	65.42	68.29	75.00	
Single	61.06	66.30	62.75	60.98	46.67	
Educational Level	62.46	69.57	58.49	56.10	68.89	
Age	51.53	56.52	51.40	52.44	40.00	
Employment Status	54.29	50.00	51.85	58.02	62.22	
<i>Trip characteristics</i>						
Preferred type of accommodation	42.64	44.57	35.19	41.46	59.09	*
Visitation status to WE	76.95	83.70	72.90	75.64	75.00	
Estimated duration of the next trip to WE	62.58	58.70	64.81	62.96	64.44	
Party group of the next trip to WE	60.37	50.00	64.49	63.29	66.67	
<i>Main Purpose of travel</i>						
VFR	3.96	2.15	4.63	3.66	6.67	
Study	19.21	21.51	20.37	19.51	11.11	
Work	5.18	4.30	9.26	2.44	2.22	
Holiday	83.54	84.95	82.41	78.05	93.33	
<i>What destinations are you most likely to visit?</i>						
UK	55.49	64.52	51.85	46.34	62.22	*
Italy	54.57	54.84	51.85	58.54	53.33	
Belgium	13.41	8.60	12.96	15.85	20.00	
Portugal	9.45	7.53	12.96	4.88	13.33	
France	72.56	74.19	65.74	70.73	88.89	**
Switzerland	53.05	51.61	52.78	46.34	68.89	
Ireland	17.99	16.13	20.37	14.63	22.22	
Netherlands	30.79	29.03	33.33	30.49	28.89	
Germany	39.63	41.94	39.81	29.27	53.33	*
Spain	39.33	37.63	45.37	32.93	40.00	
Austria	22.87	25.81	25.00	14.63	26.67	
Greece	50.30	56.99	47.22	37.80	66.67	***
<i>What information source are you likely to use to plan your trip to Western Europe?</i>						
TV or radio advertising	15.85	13.98	17.59	19.51	8.89	**
Guidebook	33.84	29.03	44.44	24.39	35.56	*
Internet search engine	77.13	84.95	75.93	76.83	64.44	
Travel agency	44.51	51.61	38.89	42.68	46.67	
Travel forums & blogs	47.56	53.76	50.00	35.37	51.11	*
Special magazine	29.88	31.18	32.41	26.83	26.67	

Notes: All Chi-square tests calculated are not significant unless indicated otherwise: *** Significant at $p \leq 0.01$, ** Significant at $p \leq 0.05$. * Significant at $p \leq 0.1$.

Appendix

A. Table

Table A.1: *Variables description*

Independent variables	Descriptions
<i>Socio-demographic and economic characteristics</i>	
Female	1= ticked; 0= not ticked
Individual monthly income	1= Individual monthly income equal to 7,000 RMB or less; 0 = otherwise
Single	1= ticked; 0= not ticked
Educational level	1 = University degree and less; 0 = Post-graduate degree
Age	1 = 18 and 25 years old; 0 = 26 years old and over
Employment Status	1 = Full-time employee; 0 = student or not employed
<i>Trip characteristics</i>	
Preferred Type of Accommodation	1= 3–5 star hotel; 0= otherwise (e.g. hostel, guest house)
Visitation Status to WE	1= First-timer in Western Europe; 0= otherwise
Estimated Duration of the Next Trip to WE	1= Less than 2 weeks in Western Europe; 0= otherwise
Party Group of the Next Trip to WE	1= Family or partner on the next trip to Western Europe; 0= otherwise
<i>Main Purpose of travel</i>	
VFR	1= visiting friends & relatives; 0= otherwise
Study	1= ticked; 0= not ticked
Work	1= ticked; 0= not ticked
Holiday	1= ticked; 0= not ticked
<i>What destinations are you most likely to visit?</i>	
UK	1= ticked; 0= not ticked
Italy	1= ticked; 0= not ticked
Belgium	1= ticked; 0= not ticked
Portugal	1= ticked; 0= not ticked
France	1= ticked; 0= not ticked
Switzerland	1= ticked; 0= not ticked
Ireland	1= ticked; 0= not ticked
Netherlands	1= ticked; 0= not ticked
Germany	1= ticked; 0= not ticked
Spain	1= ticked; 0= not ticked
Austria	1= ticked; 0= not ticked
Greece	1= ticked; 0= not ticked
<i>What information source are you likely to use to plan your trip to Western Europe?</i>	
TV or radio advertising	1= ticked; 0= not ticked
Guidebook	1= ticked; 0= not ticked
Internet search engine	1= ticked; 0= not ticked
Travel agency	1= ticked; 0= not ticked
Travel forums & blogs	1= ticked; 0= not ticked
Special magazine	1= ticked; 0= not ticked