Knowledge-Based Systems 44 (2013) 20–33

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Ranking and selection of unsupervised learning marketing segmentation

Germán Sánchez-Hernández [a,b], Francisco Chiclana [c,d,*], Núria Agell [a], Juan Carlos Aguado [b]

[a] ESADE Business School, Universitat Ramon Llull, Barcelona, Spain
[b] Automatic Control Department, Universitat Politècnica de Catalunya, Barcelona, Spain
[c] Centre for Computational Intelligence, Faculty of Technology, De Montfort University, Leicester, UK
[d] DMU Interdisciplinary Group in Intelligent Transport Systems, Faculty of Technology, De Montfort University, Leicester, UK

ARTICLE INFO

Article history:
Received 16 July 2012
Received in revised form 19 November 2012
Accepted 10 January 2013
Available online 10 February 2013

Keywords:
Fuzzy selection criteria
OWA operator
Classification selection
Market segmentation
Linguistic quantifier

ABSTRACT

This paper addresses the problem of choosing the most appropriate classification from a given set of classifications of a set of patterns. This is a relevant topic on unsupervised systems and clustering analysis because different classifications can in general be obtained from the same data set. The provided methodology is based on five fuzzy criteria which are aggregated using an Ordered Weighted Averaging (OWA) operator. To this end, a novel multi-criteria decision making (MCDM) system is defined, which assesses the degree up to which each criterion is met by all classifications. The corresponding single evaluations are then proposed to be aggregated into a collective one by means of an OWA operator guided by a fuzzy linguistic quantifier, which is used to implement the concept of fuzzy majority in the selection process. This new methodology is applied to a real marketing case based on a business to business (B2B) environment to help marketing experts during the segmentation process. As a result, a segmentation containing three segments consisting of 35, 98 and 127 points of sale respectively is selected to be the most suitable to endorse marketing strategies of the firm. Finally, an analysis of the managerial implications of the proposed methodology solution is provided.

1. Introduction

The use of unsupervised learning systems allows the behaviour of certain phenomena to be identified without relying on expert knowledge or information from past situations. Indeed, the main characteristic of this type of learning systems is that it works with patterns without explicitly knowing their output. Because of this, unsupervised learning systems have been considered in the literature as systems capable to capture knowledge from complex structures [1–3].

Choosing the most appropriate classification from a given set of classifications of a set of patterns is an important topic on unsupervised systems and, in particular, on clustering analysis. In most cases, the use of these techniques leads to several classifications as outputs, i.e. various classifications are compatible with the set of given patterns. For this reason, research in this area aims to develop suitable tools and models for selecting classifications [4–6].

Previous research in this direction uses selection criteria as filters: a set of criteria is applied sequentially to all the obtained classifications [6–9]. All those classifications failing to meet a particular criterion are discarded and not taken into account in the application of the subsequent criterion. The following drawback can be associated with this type of methodology: because a true–false decision is applied in the application of each criterion, this could result in classifications being discarded and not taken into account when they marginally fail to meet one particular criterion but meet other criteria with a high score. Therefore, a classification might be discarded prematurely when its ‘overall’ score, with respect to the set of criteria, would have been high. In an extreme case, this methodology could produce no result because none of the classifications meet a particular criterion, which could indicate that the criterion in particular might not have been the most adequate or taken into account.

An alternative approach to the sequential approach described above, which has been successfully applied in multi-criteria decision making (MCDM), is that of evaluating the degree up to which each criterion is met by all classifications, i.e. the use of fuzzy criteria, and, only after this, obtaining an overall aggregated value for each classification reflecting the degree up to which the whole set of criteria is satisfied by each classification. Note that the objective of the aggregation step is to combine a set of criteria in such a way that the final aggregation output takes into account all the single fuzzy criterion [10]. The final selection of classifications naturally derives from this set of overall degrees, and the drawback mentioned above does not apply.

* Corresponding author at: Centre for Computational Intelligence, Faculty of Technology, De Montfort University, Leicester, UK.

E-mail addresses: german.sanchez@esade.edu (G. Sánchez-Hernández), chiclana@dmu.ac.uk (F. Chiclana), nuria.agell@esade.edu (N. Agell), juan.carlos.aguado@upc.edu (J.C. Aguado).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.01.012

Many different families of aggregation operators have been studied [10–20]. Among them the Ordered Weighted Averaging (OWA) operator proposed by Yager [19] is one of the most widely used. Among the reasons to support this extensive use of the OWA operator is that it allows the implementation of the concept of fuzzy majority in the aggregation phase by means of a fuzzy linguistic quantifier [21] representing the proportion of satisfied criteria 'necessary for a good solution' [22]. This is done by using the linguistic quantifier in the computation of the weights associated with the OWA operator. In addition, Marichal [23] investigated the aggregation of dependent criteria and the fuzzy integral was found to be the appropriate aggregation operator in these cases. The most representative fuzzy integrals are the Choquet integral and the Sugeno integral. It is well known that the OWA operator is a particular case of Choquet integral, and consequently it is not necessary to assume independence of criteria when using the OWA operator.

From the application point of view, unsupervised systems have been relevant in a wide range of domains, among which it is worth mentioning: text categorisation, images recognition, telecommunications fraud detection, stock price forecasting, bioinformatics, fault diagnosis, pollution classification and clinical or socio-economic systems [24–34]. In the marketing field, finding new and creative solutions is valuable because these allow for the definition of new strategies and innovation. The use of unsupervised learning algorithms allows us to suggest segmentations that are, in principle, not trivial. In this sense, behavioural patterns of 'interesting' profiles could be established by using this type of algorithm and these may reveal new customer profiles not yet known to experts [35–39].

This paper presents a novel classification selection methodology based on a set of fuzzy criteria and the MCDM approach described above. This MCDM approach uses an aggregation function based on OWA operators defined via a linguistic qualifier to summarise the information gathered through the set of fuzzy criteria. This new methodology has been implemented in the statistical computing tool R [40] and applied to a real marketing problem.

The paper is structured as follows. In the next section five selection criteria related to market segmentation are defined, and their fuzzy nature and interpretation are considered. Following that, in Section 3, the MCDM approach is introduced and the OWA operator and fuzzy linguistic quantifier concepts are provided. A case study to select a segmentation from a real business situation is described in Section 4, and results obtained by applying the proposed new methodology are analysed. In Section 5 conclusions are drawn and suggestions for further future research work are given.

2. Fuzzy criteria for selecting classifications

The use of unsupervised learning algorithms enables to find out non trivial classifications. However, when many different classifications are obtained, how to choose the best one with respect to the proposed objective? In this section methods and criteria for the evaluation of clustering results are reviewed. Below five fuzzy indicators, adapted and extended from criteria introduced by Sánchez-Hernández et al. [8] to help solve this problem, are described and defined. For each fuzzy criterion, a membership function describing the degree up to which it is verified by a particular classification is proposed.

2.1. Clustering validation

This section reviews criteria and methods to evaluate classifications derived from the application of any of the available clustering techniques. There are mainly three types of clustering validation criteria [41,42]: internal, external and relative. An internal criterion tries of determine if the classification structure is intrinsically appropriate for the data. An external criterion of validation compares the considered classification with an *a priori* structure: either a previously known partition of the analysed dataset, typically provided by some domain experts, or an external variable not participating in the clustering process. Finally, a relative criterion measures the relative similarity between two classifications.

Several works reviewing cluster validation indexes have been published [6,43–45]. These works and other using or defining new criteria are shown in Table 1. Criteria associated with the *compactness* concept compute how closely related the individuals in a cluster are, being usually based on indexes measuring density or variance of the data; *separability* criteria determine how distinct or well-separated a cluster is from other clusters; criteria related to the *prediction strength* of the clusters usually calculate the accuracy rate of a model constructor from them [6,46]; some criteria are based on the number of important *features* [6]; criteria quantifying the achievement of *goals* can be very heterogeneous, from applying economic theories [7], being assessed by graphical visualisations [47], or checking the existence of outliers clusters or pairs of variables [6]. External criteria require the existence of an *a priori* external variable or classification defined for each of the individuals. The computation of an index associated with external criteria can be performed by any of the following indexes: Rand statistic, Jaccard coefficient, Fowlkes and Mallows index, Hubert's statistic and so on. The computation of relative criteria implies the pairwise comparison between clusters, usually performed by some domain experts.

Although there are some methods to guide the search of which comparisons should be made for minimising their number, relative criteria have not been taken into account in this work due to the usual difficulty in getting this feedback from the experts. All the analysed papers review or define criteria based on a few concepts used for clustering evaluation, while almost all concepts are covered in this work.

2.2. First criterion: useful number of classes

The usability of a classification is based on its informativeness and manageability: it is worthwhile examining classifications that have a sufficient number of classes to generate new knowledge, but are small enough to produce an easy and manageable model. For instance, in marketing environments in which these classifications are used to extract behavioural patterns to design market strategies, the number of classes distinguished is usually taken to be between three and five [53]. This is because marketing campaigns with less than three segments may not be informative; while those with more than five segments may not be manageable.

The assumption of a classification with a number of classes M between K_1 and K_2 to be considered useful for a given problem does not imply that a classification with a number of classes lower than K_1 or higher than K_2 should be automatically discarded. This is specially true in those cases when there is enough evidence to suggest that such classifications perform well with respect to the rest of criteria. A natural approach in these cases would be that of associating a value to each classification to indicate how well they fit with the criterion 'useful number of classes'. By doing this, we move from a crisp to a fuzzy interpretation of the criterion 'useful number of classes', i.e. we move from the use of a characteristic function to the use of a membership function.

Note that a classification with a single class is trivial and therefore not useful, while a classification with a number of classes between K_1 and K_2 is considered totally useful. The minimum number of classes in any classification is 1 (contains all the individuals), while the maximum is N (each class contains just 1 individual). These two classifications are not informative and therefore these