Expert Systems with Applications 41 (2014) 4186–4196

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Integrating Fuzzy C-Means and TOPSIS for performance evaluation: An application and comparative analysis

Chunguang Bai [a,*], Dileep Dhavale [b], Joseph Sarkis [c,1]

[a] School of Management Science and Engineering, Dongbei University of Finance & Economics, Jianshan Street 217, Dalian 116025, PR China
[b] Graduate School of Management, Clark University, 950 Main Street, Worcester, MA 01610-1477, USA
[c] School of Business, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, USA

ARTICLE INFO

Keywords:
Performance evaluation
Fuzzy C-Means
TOPSIS

ABSTRACT

In this paper we introduce a multi-method multiple criteria approach for evaluating the performance of organizations. Performance analysis may include both strategic and operational performance, as well as financial and other less tangible factors. This paper introduces the use of Fuzzy C-Means and TOPSIS for organizational performance evaluation purposes. Using real company data and balanced scorecard accounting and performance dimensions the methodology is applied and evaluated. The predictive abilities of the technique from an organizational performance evaluation approach are evaluated using this data. One of the results from the illustrative application is that economic performance evaluation is not the best predictor of overall viability of some organizations, especially e-commerce based organizations.

1. Introduction

The field of discrete alternative multiple criteria decision analysis (MCDA) and choice have been a mainstay of modeling for a number of years. MCDA has had numerous practical applications in addition to an extensive theoretical history. Managers and decision makers within organizations and people in their everyday lives face a broad variety of decisions that require the consideration and tradeoffs associated with multiple attributes, criteria, and factors. Areas in environmental, economic, operational, strategic, marketing, engineering, design, educational, and psychological disciplines and others have come to rely on some of the latest developments in MCDA. Our research seeks to advance the understanding and development of models in MCDA to make decision making more effective, efficient, and reliable. Using clustering approaches for MCDA is one way of accomplishing this task. We seek to introduce a unique multi-stage approach to address this issue.

In this paper, we introduce the integration of Fuzzy C-Means (FCM), and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) for discrete alternative multiple criteria ranking. We have combined the strong points of these two methodologies to construct an integrated approach to rank data objects based on multiple criteria. The combined method minimizes computational effort needed to arrive at rankings since it develops them in two stages (achieving an 'efficiency' target for a good MCDA model). First, the FCM algorithm is used categorize similar data objects into clusters. Then TOPSIS is utilized to develop rankings for the clusters and also separate rankings for the data objects within the clusters. To examine the validity of this tool, a comparative analysis will be completed to evaluate the integrated FCM/TOPSIS approach against a novel methodology based on Bayesian inference and a latent class clustering model (Mistry, Sarkis, & Dhavale, 2014). This research aspect will determine the effectiveness and reliability of the proposed FCM/TOPSIS technique. This latent class technique uses Markov Chain Monte Carlo simulation to extract information about the objects that is already present in the multi-characteristic data, but is not readily apparent. As part of the research we found some divergences in the solutions when comparing both approaches using real world data on electronic commerce organizations.

A secondary research goal and contribution of this paper is an evaluation of applying the proposed tool to identify characteristics of company resilience and viability. Specifically, in this paper, we apply FCM/TOPSIS and latent class model to real-world archival data to determine the viability of electronic commerce organizations. The perspectives defined in the well-known Balanced Scorecard method are used to develop the multi-criteria objectives. A unique evaluation technique called a displacement index is used in the comparative analysis and validation/reliability of the joint FCM/TOPSIS methodology. In this research we examine which dimensions of the Balanced Scorecard method would be better predictors of resilience.

* Corresponding author. Tel.: +86 13664228458; fax: +86 (411) 87403733.
E-mail addresses: chunguang.bai@dufe.edu.cn (C. Bai), ddhavale@clarku.edu (D. Dhavale), jsarkis@WPI.edu (J. Sarkis).
[1] Tel.: +1 (508) 831 4831.

0957-4174/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.12.037

This paper contributes to the literature first by introducing the integration of FCM/TOPSIS and secondly, evaluating this approach using real world information and comparing it against a benchmark provided by a robust statistical technique based on Bayesian inference. We also provide a variety of analyses that allow us to evaluate how well various factors or perspectives may perform in this environment. Additionally, potential applications and future research avenues are identified to help build upon this novel approach.

To help accomplish the goals of this paper, we begin by providing a brief background on the use of clustering approaches for MCDA, along with background on FCM/TOPSIS techniques. Then data for an illustrative application of the approach is presented. The section following contains comparative analysis and discussion. The paper concludes with a summary of findings, limitations and direction for future research.

2. Background

Many approaches exist for MCDA (e.g. see Guitouni & Martel, 1998) and can be categorized on a variety of dimensions including explicit versus implicit alternative valuation, compensatory versus non-compensatory, discrete versus continuous alternative, interactive versus non-interactive, objective versus subjective weighting (Bai & Sarkis, 2012; Wang, Jing, Zhang, & Zhao, 2009; Yoon & Hwang, 1995). Applications of MCDA approaches have been completed as stand-alone tools, but many times multi-stage multi-methodology approaches have proven to be more useful.

One promising and relatively novel approach for MCDA is through the use of clustering techniques. Clustering of objects involves creating subsets such that objects in a given subset (or a cluster) are more similar to each other based on multiple criteria defined by a decision maker compared to objects that do not belong to that cluster (Hartigan, 1975; Izakian & Abraham, 2011; Khan & Ahmad, 2004). The concept of clustering is very useful in dealing with multi-characteristic data objects that a user wants to categorize based one or more the characteristics. Literature is replete with practical application of this concept. For example, it was applied to robotic learning (Alpaydin, 2004), data search and filtering (Tan, Steinbach, & Kumar, 2005) and pattern recognition (Webb, 2002).

Hathaway and Bezdek (1993) classify different types clustering methods as hard, fuzzy, probabilistic and possibilistic. One shortcoming of applying traditional or hard clustering techniques, such as K-Means, is that they are deterministic in nature; a data object either belongs to a cluster with probability of 1 or does not ($p = 0$). Hard clustering algorithms work well if there are known and well-defined boundaries for the data objects. These algorithms are mainly used to classify the data objects into clusters rather that to extract hidden unknown information about relationships amongst the objects.

Alternatively, it is possible to explore the hidden relationship using algorithms that do not assign the binary probabilities but allow each data object to belong to a set of clusters based on a discrete, empirical probability distribution. The concept of uncertainty in belonging to a cluster and thus exploring and extracting the hidden relationships amongst the data objects can be achieved using fuzzy theory. Fuzzy mathematical theory was first introduced by Zadeh (1965). The model provides a mechanism to represent and manipulate uncertainty and ambiguity of cluster relationships and provides an iterative algorithm to optimize assignment of data objects to clusters. The level of membership in a fuzzy cluster depends on the closeness of a data object to the cluster center. This degree of membership is quantified by a value in the interval $[0,1]$ and indicates the strength of the association between that object and a particular cluster centroid. The value may also be viewed as probability of a given data object belonging to a certain cluster. One of the well-known, often-used fuzzy clustering methods is the Fuzzy C-Means (FCM) algorithm, which was developed by Bezdek (1981). FCM clustering has been widely applied in fields such as astronomy, geology, medical imaging, target recognition, and image segmentation (Bezdek & Pal, 1992; Chuang, Tzeng, Chen, Wu, & Chen, 2006; Tizhoosh, 1998).

Several enhancements to the original FCM method such as integration with genetic algorithms, simulated annealing, ant colony optimization and fuzzy particle swarm optimization and their various amalgamations have been proposed to overcome some inherent limitations of the original method (Celikyilmaz & Burhan Türkşen, 2008a, 2008b; Izakian & Abraham, 2011). The two major deficiencies are the tendency of FCM iterative algorithm to become trapped in a local minimum and the slow convergence towards the optimal values. Another minor issue with the algorithm results from the initial values assigned to the membership functions. The initialization of these values by a user determines the direction of the iterative search for minima, thus different initializations may result in different minima of the multi-modal objective function (Kannan, Devi, Ramathilagam, & Takezawa, 2013; Siang Tan & Mat Isa, 2011).

Although researchers have proposed many algorithms, unfortunately they still provide tenuous results for data points that are far from all cluster centroids. These data points, the outliers, can belong to many clusters without indication of strong membership in any one cluster. This phenomenon sometimes is referred to as very fuzzy partitions (Bandyopadhyay, 2005; BaşKıR & Türkşen, 2013; Honda & Ichihashi, 2005). With these weaknesses in mind, we now present an overview of FCM with additional insights into the procedure and developments.

2.1. Overview of Fuzzy C-Means

In this section, we provide a brief overview of the FCM algorithm. The fuzzy clustering, meaning non-deterministic clustering, of objects is described by a matrix U with n rows (number of objects) and c columns (number of clusters). The element u_{ik} indicates the degree of association or membership of the ith object with the kth cluster. There is no certainty of an object belonging to a cluster, only a probability that it may do so. Thus the algorithm assigns data objects to each cluster in a stochastic fashion. It partitions a set of n objects each with P characteristics or data vectors, into c clusters. The centroid or cluster center of each cluster, v_i, $V = \{v_1, v_2, \ldots, v_c\}$, is computed using the degree of membership as a weight. Based on this description, we can further quantify the relationships as follows:

$$u_{ik} \in [0,1] \quad \forall i = 1,2,\ldots,n; \quad \forall k = 1,2,\ldots,c; \tag{1}$$

$$\sum_{k=1}^{c} u_{ik} = 1, \quad \forall i = 1,2,\ldots,n; \tag{2}$$

$$0 \leqslant \sum_{i=1}^{n} u_{ik} \leqslant n \quad \forall k = 1,2,\ldots,c; \tag{3}$$

The FCM algorithm seeks to minimize the objective function (4):

$$\min J(U,V) = \sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik}^{m}(\|x_i - v_k\|_A), \tag{4}$$

where m ($m > 1$) is a scalar termed the weighting exponent and controls the amount of fuzziness of the resulting clusters, and is the Euclidian distance from object x_i to the cluster center v_k. It is possible to use distances other that Euclidian distances when forming