Expert Systems with Applications 36 (2009) 5909–5915

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A case study of applying data mining techniques in an outfitter's customer value analysis

Shian-Chang Huang [a], En-Chi Chang [b], Hsin-Hung Wu [a,*]

[a] Department of Business Administration, National Changhua University of Education, No. 2, Shida Road, Changhua City, Changhua 500, Taiwan
[b] Manchester Business School, Booth Street West, Manchester, M15 6PB, UK

ARTICLE INFO

Keywords:
K-means method
Fuzzy c-means method
Bagged clustering algorithm
Value analysis
Cluster quality assessment

ABSTRACT

This study applies *K*-means method, fuzzy c-means clustering method and bagged clustering algorithm to the analysis of customer value for an outfitter in Taipei, Taiwan. These three techniques bear similar philosophy for data classification. Thus, it would be of interest to know which clustering technique performs best in a real world case of evaluating customer value. Using cluster quality assessment, this study concludes that bagged clustering algorithm outperforms the other two methods. To conclude the analyses, this study also suggests marketing strategies for each cluster based on the results generated by bagged clustering technique.

1. Introduction

Outdoor activities are gaining their popularity in Taiwan and opportunities for selling outdoor outfits are abundant. According to the Commerce Industrial Services Portal of the Ministry of Economic Affairs, Republic of China (http://gcis.nat.gov.tw/English/index.jsp), there were 52 outfitters, most of which were located in Northern and Central Taiwan areas at the end of 2006. The large number of outfitters brought competition, which led to the decrease of profits. To survive competition and sustain profits, the outfitter must identify and retain customers of high value and profit potentials. Achieving the aforementioned goals will require the outfitter to customize marketing strategies and fulfill the needs of different customers and also to allocate resources effectively and efficiently, based on a well-managed customer database.

Managing customer database is not an easy task. As the transaction record of a company becomes much larger in size as the time goes by, it might be necessary to divide all customers into appropriate number of clusters based on some similarities in these customers by using data mining techniques, particularly the clustering techniques. The values of different customer groups can then be calculated and evaluated to provide useful decisional information for management to utilize resources rationally.

A variety of clustering techniques is commonly seen in practice. This study will discuss three clustering techniques, i.e., *K*-means method, fuzzy c-means method, and bagged clustering algorithm. These three techniques bear similar philosophy. *K*-means method is the most commonly seen approach for classification (Davidson, 2002). Fuzzy c-means method, very similar to the philosophy of *K*-means method, uses membership grades for data clustering (Jain, Murty, & Flynn, 1999). Bagged clustering algorithm based on *K*-means method and hierarchical methods provides another way for clustering (Dolnicar & Leisch, 2004). Due to the similarities in philosophy, it would be of interest to use these three methods in a case study and then evaluate which clustering technique performs better under the same circumstances.

A real case of an outfitter in Taipei, Taiwan will illustrate how these three techniques can be used in managing customer data and helping draft promotion strategies. The transaction data consist of 551 customers who shopped at the outfitter's store from April 2004 to March 2006. The profile for each customer includes the membership number, gender, birth date, zip code, shopping frequency, and the total spending at the store.

This paper is organized as follows: Section 2 reviews *K*-means method, fuzzy c-means method, and bagged clustering algorithm. The clustering case study of the outfitter based on the three techniques is provided and analyzed in Section 3. Marketing implications including promotion strategies for different clusters drawn from the best technique among the three are discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Review of clustering methods

K-means method is one of the most commonly used approaches for classification and is an *exclusive clustering* algorithm, where if a certain data point belongs to a definite cluster then it could not be included in another cluster (Davidson, 2002). Fuzzy c-means method, on the other hand, is an *overlapping clustering* algorithm, where

* Corresponding author. Tel.: +886 4 7232105x7412; fax: +886 4 7211292.
E-mail addresses: hhwu@cc.ncue.edu.tw, drhhwu@yahoo.com.tw (H.-H. Wu).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.07.027

a data point can belong to many clusters with different membership grades between zero and one (Jain et al., 1999; Nascimento et al., 2000). Bagged clustering algorithm is a combination of partitioning methods such as *K*-means method and hierarchical methods that provides another way to assess and enhance the stability of a partitioning method using hierarchical clustering (Dolnicar & Leisch, 2004). That is, *K*-means method is the fundamental clustering method, whereas fuzzy c-means clustering method and bagged clustering algorithm can be viewed as improved *K*-means method for data clustering. The review of *K*-means, fuzzy c-means, and bagged methods are as follows.

2.1. K-means method

K-means method, a non-hierarchical method, is a very popular approach for classification because of its simplicity of implementation and fast execution and has been widely used in market segmentation, pattern recognition, information retrieval, and so forth (Cheung, 2003; Davidson, 2002; Kuo et al., 2002). The commonly used distance in *K*-means method is Euclidean distance (Davidson, 2002; Yoon & Hwang, 1995). The formula of *K*-means method is as follows, where the distance between two points X_r and X_s is given by the square root of the sum of the squared distance over each coordinate, and $X_r = (x_{r1}, x_{r2}, x_{r3}, \ldots, x_{ri}, \ldots, x_{rn})$ and $X_s = (x_{s1}, x_{s2}, x_{s3}, \ldots, x_{si}, \ldots, x_{sn})$, and each c_i in Eq. (1) represents the weight. If the weights are normalized, then $\sum_{i=1}^{n} c_i = 1$ (Buttrey & Karo, 2002)

$$d(X_r, X_s) = \left[\sum_{i=1}^{n} c_i (x_{ri} - x_{si})^2\right]^{1/2}. \quad (1)$$

K-means method consists of the following two major steps (Davidson, 2002). First, the assignment step where the instances are placed in the closest class. Second, the re-estimation step where the class centroids are recalculated from the instances assigned to the class. However, one of the major problems of *K*-means method is to select the best value of K (Buttrey and Karo, 2002; Jain et al., 1999). Kuo et al. (2002) have pointed out that non-hierarchical methods, such as *K*-means method, can have higher accuracy if the starting point and the number of clusters are provided. Punj and Steward (1983) suggested a two-stage method by deploying Ward's minimum variance method to determine the number of clusters for *K*-means method. On the other hand, Kuo et al. (2002) have proposed a modified two-stage method by applying self-organizing feature maps to determine the number of clusters for *K*-means method. The reason is that self-organizing feature maps can converge very fast since it is a kind of learning algorithm that can continually update or reassign the observations to the closest cluster. Therefore, this study uses self-organizing feature maps to determine the number of clusters for *K*-means method.

2.2. Fuzzy c-means clustering method

Nascimento et al. (2000) discussed that the major task of partitioning methods is to partition a set of entities into a number of homogeneous clusters with respect to a suitable similarity measure. However, due to the fuzzy nature, fuzzy c-means clustering method has been developed that a data point can belong to many clusters with different membership grades between zero and one. That is, the philosophy of fuzzy clustering method is that each point has a degree of belonging to clusters expressed by fuzzy logic rather than belonging completely to just one cluster. The points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster. For each point x, a coefficient is given by the degree of being in the kth cluster $u_k(x)$. Typically, the sum of those coefficients is defined to be one, i.e., x $\sum_{k=1}^{\text{number of clusters}} u_k(x) = 1$ (Baraldi & Blonda, 1999; Yang, 1993).

The centroid of a cluster in fuzzy c-means method is the mean of all points weighted by their degree of belonging to the cluster, expressed as follows:

$$\text{center}_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}, \quad (2)$$

where the degree of belonging, $u_k(x)$, is related to the inverse of the distance to the cluster by the following equation:

$$u_k(x) = \frac{1}{d(\text{center}_k, x)}. \quad (3)$$

When the coefficients are normalized and fuzzified with $m > 1$, depicted in Eq. (4), the sum of $u_k(x)$ is one.

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\text{center}_k, x)}{d(\text{center}_j, x)}\right)} 1/m - 1. \quad (4)$$

If $m = 2$, the coefficient is equivalently to be normalized linearly, and the sum of $u_k(x)$ is one. When m is close to one, the point which is the closest to the cluster center is given much higher weight than the others. That is, fuzzy c-means method is very similar to *K*-means method, and the procedures are summarized below (Yang, 1993):

1. Choose a number of clusters.
2. Assign randomly to each point coefficient for being in the clusters.
3. Repeat the above procedures until the clustering results have been converged. The change of coefficients between two iterations is less than a given sensitivity threshold.
4. Use Eq. (2) to calculate the centroid for each cluster.
5. For each point, use Eq. (3) to compute its coefficients of being in the clusters.

2.3. The bagged clustering algorithm

Dolnicar and Leisch (2004) stated that the current popular clustering techniques fall into one of the two major categories, i.e., partitioning methods such as *K*-means or its online variant (learning vector quantization) and hierarchical methods resulting in a dendrogram. Bagged clustering algorithm, on the other hand, is a combination of both methods and the central idea is to stabilize partitioning methods such as *K*-means method by repeatedly running the clustering algorithm and combining the results (Dolnicar & Leisch, 2001). Thus, a collection of training sets by sampling from the empirical distribution of the original data can be obtained.

Bagged clustering algorithm explores the independent solutions from several runs of any partitioning clustering algorithm using hierarchical clustering. It can also be seen as an evaluation of the partitioning clustering algorithm by means of the bootstrap, which allows the researcher to identify structurally stable centers repeatedly. The procedure of using bagged clustering algorithm consists of the following five steps (Dolnicar & Leisch, 2001, 2004):

1. Given a data set X_N of size N. Construct B bootstrap training samples $X_N^1, X_N^2, \ldots, X_N^B$ of size N by drawing with replacement from the original sample X_N.
2. Use any partitioning method such as *K*-means method on each set to generate $B \times K$ centers, namely $c_{11}, c_{12}, \ldots, c_{1k}, c_{21}, \ldots, c_{BK}$, where K is the number of centers used in the partitioning method and c_{ij} is the jth center found by using X_N^i.
3. Integrate all centers into a new data set of $C^B(K)$, where $C^B(K) = \{c_{11}, c_{12}, \ldots, c_{BK}\}$.