Expert Systems with Applications 37 (2010) 8395–8400

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques

Chih-Hsuan Wang *

Ming Chuan University, Department of Business Administration, Taiwan

ARTICLE INFO

Keywords:
Robust classification
Robust segmentation
Kernel induced fuzzy clustering

ABSTRACT

To understand customers' characteristics and their desire is critical for modern CRM (customer relationship management). The easiest way for a company to achieve this goal is to target their customers and then to serve them through providing a variety of personalized and satisfactory goods or service. In order to put the right products or services and allocate resources to specific targeted groups, many CRM researchers and/or practitioners attempt to provide a variety of ways for effective customer segmentation. Unfortunately, most existing approaches are vulnerable to outliers in practice and hence segmentation results may be unsatisfactory or seriously biased. In this study, a hybrid approach that incorporates kernel induced fuzzy clustering techniques is proposed to overcome the above-mentioned difficulties. Two real datasets, including the WINE and the RFM, are used to validate the proposed approach. Experimental results show that the proposed approach cannot only fulfill robust classification, but also achieve robust segmentation when applied to the noisy dataset.

1. Introduction

Today, the mass marketing approach cannot satisfy customers' needs and their diverse preferences and most companies need to contact, serve, and manage their customers through providing a variety of attractive, personalized, and satisfactory goods or service. Market segmentation assumes that groups of customers with similar needs and purchasing patterns are likely to demonstrate a homogeneous response to marketing programs that target specific customer groups (Tsai & Chiu, 2004). With proper market segmentation, companies can put the right products or services to a targeted customer group and hence improve the efficiency of their marketing strategies. In order to understand their customers more clearly, companies may integrate an abundance of data collected from multiple channels. Typical ways include web browsing, purchasing pattern, complaints demographics and psychographic behavior.

According to the so-called "20–80" rule, a dramatic business improvement is often achieved by identifying the 20% of core customers and by maximizing the attention applied to them since they will account for the 80% of contribution of company's profit. Therefore, satisfying existing customers' needs and build close relationships with them will be very imperative in modern electronic commerce. Owing to the rapid development of data warehousing and data-mining techniques, it is less costly to "up-sell" or to "cross-sell" the existing customers. However, acquiring new customers is still difficult and expensive. Based on this perspective, companies need to understand their customers by analyzing customer information, to differentiate between various groups, to identify the most or the least valuable customers, and to increase customer loyalty through providing customized products and services (Ha, 2007).

One of the critical and challenging issues for successful market segmentation is the selection of the segmentation variables (Tsai & Chiu, 2004). In general, segmentation variables can be roughly classified into customer related variables (i.e. demographics, lifestyles) and product specific variables (i.e. purchasing behavior, transaction records). In spite of various types of segmentation variables, practical marketers continue to use RFM (recency, frequency, and monetary) models since it is easy to be implemented and to be understood by decision makers (McCarty & Hastak, 2007). Specifically, "recency" denotes the length of time period since the last purchase, "frequency" means the number of purchases within a certain period, and "monetary" represents the amount of money spent during a certain period. There are a variety of ways of applying RFM model on customer segmentation, including *K*-means or fuzzy *C* means (FCM), artificial neural network (ANN), and decision tree (DT). Unfortunately, most of the above-mentioned methods still have the following flaws:

* Current address: Department of Industrial Engineering & Management, National Chiao Tung University, Taiwan. Tel.: +886 3 5714261.
E-mail addresses: jswang@mail2000.com.tw, chihswang@gmail.com.

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.05.042

- The adverse effect of outliers is usually omitted or rarely investigated.
- Segmentation results are very vulnerable to outliers or noisy data.
- The determination of the number of clusters is ambiguous or inconsistent.

Therefore, this inspires us to develop a hybrid approach that is capable to quickly detect outliers and to segment customers more effectively. The remainder of this paper is organized as follows. Section 2 discusses related work of outlier detection and robust segmentation. Section 3 reviews possibilistic clustering and probabilistic clustering techniques. Experimental results collected from two real datasets are illustrated in Section 4 and conclusions are drawn in Section 5.

2. Related works

In contrast to traditional data-mining task that aims to search for a general pattern for the majority of input data, novelty detection attempts to find the rare class whose behavior is very exceptional when compared to the rest of input data (He, Xu, Huang, & Deng, 2004). Novelty detection, or so-called outlier detection, is the identification of "novel" or "unknown" events that an expert system is not aware of during training or testing. It is very fundamental to a classification or identification system since outliers may indicate abnormal running conditions and lead to significant performance degradation. By contrast, clustering is an unsupervised process of dividing patterns into groups and make objects within a cluster show relatively high intra-similarity whereas objects between different clusters have very low inter-similarity (Jain, Murty, & Flynn, 1999).

Traditional clustering techniques could handle well-separated groups, but they could not treat with the overlapped or diverse clusters very well. Thus, support vector clustering and kernel based fuzzy clustering are further employed for segmenting complex customer profiles (Huang, Tzeng, & Ong, 2007; Wang, 2009). However, for most proposed schemes, the performance evaluation with respect to outliers is very scarce. In real implementation, outliers or noisy data samples often lead to biased clustering results and hence managerial insights are difficult to be obtained. Hence, under the noisy environment, developing a robust clustering approach becomes very imperative for successful market segmentation.

2.1. Outlier detection

An outlier is one that appears to obviously deviate from the others of the sample in which it occurs or an observation which appears to be inconsistent with the remainder of the dataset (Barnett & Lewis, 1994). They also think that outliers may be considered as noisy points lying outside a set of defined clusters or may be defined as points that lie outside of the set of clusters but are also different from the noise. An outlier may also denote an anomalous object or an intruder inside the system with malicious intentions. Detecting fraudulent usage of credit cards or mobile phones (fraud detection) and discovering potential criminal activities in electronic commerce (intrusion detection) are two typical applications. Besides, for loan application evaluation or public health benefit payments, an outlier identification system is helpful to detect any anomalies or abuse of social resource before any approval or payment.

In recent years, outlier detection has attracted much attention from both statistics community and data-mining research community and many techniques are proposed to fulfill this task. Three fundamental approaches are well reviewed (Hodge & Austin, 2004):

- *Determine outliers without any prior knowledge of the data:* it is analogous to the unsupervised clustering. This approach processes the data as a static distribution and flags the most remote points in the dataset as potential outliers.
- *Model both normality and abnormality:* this is analogous to supervised classification and requires pre-labeled data, tagged as normal or abnormal. However, the supervised classification is limited to known classes but new examples derived from a previously unseen region may be classified incorrectly.
- *Model only normality:* this is analogous to the semi-supervised paradigm as only the normal class is taught and the system needs to learn to recognize abnormality. This technique is usually named as novelty detection since it aims to define the boundary of normality instead of estimating the density of the dataset.

In addition, a state-of-the-art review respectively based on statistical approaches and neural network approaches are presented (Markou & Singh, 2003a, 2003b). To our best knowledge, most proposed schemes are based on supervised or parametric approaches. That means they need to rely on labeled training data or known data distribution. However, in real application, either labeled training data or its underlying distribution may be unknown or difficult to obtain in advance. As a matter of fact, an unsupervised and "distribution-free" RPCM is proposed to identify outliers in this study. Further details will be illustrated in Section 3.1.

2.2. Robust segmentation

Customer segmentation has become an important research issue in the field of electronic commerce because the identification of valuable segments can give market researchers the basis for effective targeting and predicting of potential customers (Kuo, Ho, & Hu, 2002). In particular, a popular data-mining technique called "clustering" is widely used for market or customer segmentation. There are many clustering algorithms proposed to deal with different problems, including partitioning clustering, hierarchical clustering, neural network based clustering, mixture model based clustering and kernel based clustering (Xu & Wunsch, 2005). Among those proposed schemes, clustering techniques involving K-means or fuzzy C means are relatively popular due to its short computation time and easy accommodation. A fuzzy clustering is adopted to group users of the on-line music industry and internet portals (Ozer, 2001, 2005). Traditional clustering techniques could handle well-separated groups but they could not treat with overlapped or diverse clusters very well. Hence, support vector clustering and kernel based fuzzy clustering are further employed for segmenting complex customer profiles (Huang et al., 2007).

Recently, soft computing based methods including self-organized feature maps (SOM) and adaptive resonance theory (ART) are quite popular to be applied to many problems (Lee, Suh, Kim, & Lee, 2004; Shin & Sohn, 2004; Vellido, Lisboa, & Meehan, 1999). A SOM based approach is presented to segment the on-line shopping market (Vellido et al., 1999). A two-stage method that combined SOM with K-means clustering is introduced (Kuo et al., 2002; Lee et al., 2004). In particular, SOM is used to determine the number of clusters and K-means was employed to find the final solutions. Three clustering algorithms, including K-means, FCM and SOM, are simultaneously used to segment Korean stock trading customers and concluded that FCM is the most robust approach (Shin & Sohn, 2004). Besides, a laddering technique with ART2 network is used to acquire customer requirements (Chen, Khoo, & Yan, 2002).

However, clustering methods need to be robust against outliers or noise if they are to be useful in practice (Davé & Krishnapuram, 1997; Davé & Sen, 2002; Lin & Chen, 2004). Robustness means the performance of an algorithm should not deteriorate drastically due to noise or outliers. In fact, outliers or commonly referred to novelty instances, often exist in many real databases and hence results in unsatisfied segmentation results. The purpose of this paper attempts to facilitate the research gap between outlier detection and robust segmentation by incorporating two robust clustering methods. In particular, robust possibilistic clustering method (RPCM) is proposed to detect outliers and robust fuzzy clustering method (RFCM) is used to segment customers.

3. Proposed techniques

Robust clustering techniques involving RPCM (see Section 3.1) and RFCM (see Section 3.2) are respectively proposed to detect outliers and to segment customers for the purpose of target marketing. In addition, the number of clusters is determined by examining significant eigenvalues of the affinity matrix (see Section 3.3).

3.1. Outlier detection using RPCM

The main idea originates from possibilistic C means (PCM) proposed by Kirishnapuram and Keller (1993). It can be reconsidered to find one single cluster instead of searching for multi-clusters. In contrast to fuzzy C means (FCM) (Bezdek, 1981), the membership of each data instance can be interpreted as the "typicalness" degree instead of the "belongness" membership because the separation constraint during various clusters was removed. Assume an input dataset X within p dimension, such as $X = \{x_1, x_2, \ldots, x_n\} \subset R^P$, the objective function, typicalness updating and one common centroid can be shown below.

$$J_{PCM} = \sum_{j=1}^{n} u_j^m \|x_j - a\|^2 + \eta \sum_{j=1}^{n} (1 - u_j)^m, \tag{1}$$

$$a = \frac{\sum_{j=1}^{n} u_j^m x_j}{\sum_{j=1}^{n} u_j^m}, \tag{2}$$

$$u_j = [1 + (\|x_j - a\|^2/\eta)^{1/(m-1)}]^{-1}. \tag{3}$$

Here, $m > 1$ is known as the fuzzifier and η is a regularization parameter. Based on the objective function, the first term requires that the distance from the input data x_j to the common centroid a be as low as possible whereas the second term forces its typicalness u_j as large as possible to avoid the trivial solution. Through iterative optimization, the common centroid and typicalness updating can be easily obtained. Intuitively, those data points with low typicalness will be considered as potential outliers.

In order to enhance the robustness against noise or outliers, robust possibilistic clustering method (RPCM) is further proposed. Instead of using Euclidean distance between the data instance and the common centroid, RPCM uses a kernelized distance to reconstruct the objective function and makes the algorithm insensitive to noisy data. Their mathematical forms are listed as follows:

$$J_{RPCM} = \sum_{j=1}^{n} u_j^m \|\phi(x_j) - \phi(a)\|^2 + \eta \sum_{j=1}^{n} (1 - u_j)^m, \tag{4}$$

$$\|\phi(x_j) - \phi(a)\|^2 = K(x_j, x_j) + K(a, a) - 2K(x_j, a), \tag{5}$$

where ϕ is a nonlinear mapping from the input space to the feature space, $K(x_j, a) = \exp(-\|x_j - a\|^2/\beta)$ represents the Gaussian kernel and β denotes the kernel widths. At last, the typicalness function and the common centroid can be iteratively obtained as:

$$u_j = [1 + (2(1 - K(x_j, a))/\eta)^{1/(m-1)}]^{-1}, \tag{6}$$

$$a = \frac{\sum_{j=1}^{n} K(x_j, a) u_j^m x_j}{\sum_{j=1}^{n} K(x_j, a) u_j^m}. \tag{7}$$

Obviously, the proposed RPCM demonstrates the following two advantages. RPCM is able to compute the outlier possibility of each instance in a "continuous" manner. In other words, the possibilistic membership of RPCM can be regarded as a measure of f possibility of potential outliers. Moreover, RPCM is easy and fast to be implemented empirically since it dose not need to solve quadratic optimization or statistical testing.

3.2. Robust segmentation using RFCM

FCM (fuzzy C means) can be regarded as a soft extension of hard K-means. FCM assumes that the number of clusters c, is known as a priori, and partitions a dataset X within p dimension, such as $X = \{x_1, x_2, \ldots, x_n\} \subset R^P$, into c fuzzy subsets through minimizing an objective function. The objective function, which is based on the Euclidean distance between the input data x_j and the cluster centeroid c_i, can shown as follows:

$$J_{FCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m \|x_j - c_i\|^2, \tag{8}$$

subject to the probability constraints $\sum_{i=1}^{c} u_{ij} = 1,\ 1 \leqslant j \leqslant n$ and $0 < \sum_{j=1}^{n} u_{ij} < n,\ 1 \leqslant i \leqslant c$. Similarly, the membership function u_{ij} and different cluster center c_i are respectively updated through an alternative optimization from Eqs. (9) and (10):

$$u_{ij} = \frac{\|x_j - c_i\|^{-2/(m-1)}}{\sum_{k=1}^{c} \|x_j - c_k\|^{-2/(m-1)}}, \tag{9}$$

$$c_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m}. \tag{10}$$

Here, $m > 1$ is known as the fuzzifier and $m = 2$ is usually adopted.

Obviously, FCM is not robust to tolerate noise or outliers because of assigning relatively high membership values to outliers across c various clusters. Hence, robust fuzzy clustering method (RFCM) using a kernelized distance is proposed to effectively segment customers. The objective function of RFCM and its kernel induced distance measure between the input data x_j and the cluster center c_i can be respectively shown below:

$$\begin{aligned} J_{RFCM} &= \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m \|\phi(x_j) - \phi(c_i)\|^2 \\ &= 2 \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m (1 - K(x_j, c_i)), \end{aligned} \tag{11}$$

$$\|\phi(x_j) - \phi(c_i)\|^2 = K(x_j, x_j) + K(c_i, c_i) - 2K(x_j, c_i). \tag{12}$$

By iteratively minimizing the objective function under the probability constraints $\sum_{i=1}^{c} u_{ij} = 1,\ 1 \leqslant j \leqslant n$ and $0 < \sum_{j=1}^{n} u_{ij} < n,\ 1 \leqslant i \leqslant c$, its membership function and cluster center can be respectively obtained as:

$$u_{ij} = \frac{(1 - K(x_j, c_i))^{-1/(m-1)}}{\sum_{k=1}^{c} (1 - K(x_j, c_k))^{-1/(m-1)}}, \tag{13}$$

$$c_i = \frac{\sum_{j=1}^{n} K(x_j, c_i)(u_{ij})^m x_j}{\sum_{j=1}^{n} K(x_j, c_i)(u_{ij})^m}. \tag{14}$$

Apparently, the estimation of cluster centers is weighted by the kernel function and hence the effect of outliers will be significantly decreased.

3.3. Cluster validity consideration

In general, to determine the number of clusters in advance is very challenging especially when the dataset includes diverse clusters or outliers. Most existing methods treat this problem as a measure of "cluster validity" and test various numbers of clusters within a specific range. Based on various indices for "cluster validity", the determination of the optimal number of clusters is often inconsistent among different approaches. Examining the largest eigenvalues of the affinity matrix is a good way to roughly estimate the number of clusters. If the datasets consist of clearly separated, there should be a significant drop between dominant and non-dominant eigenvalues derived from the affinity matrix. An alternative approach which relies on the structure of both eigenvalues and eigenvectors for more complex datasets is suggested (Girolami, 2002). He considered dominant terms of the following:

$$\sum_{i=1}^{n} \lambda_i \left[1_n^T u_i\right]^2, \tag{15}$$

where 1_n is a notation for a n-dimensional vector with all components equal to $1/n$, and λ_i, u_i are associated eigenvalues/eigenvectors of the affinity matrix A. In simple words, there will be N dominant terms contributed to the summation $\sum_{i=1}^{n} \lambda_i \left[1_n^T u_i\right]^2$ if there are N distinct clusters embedded in the datasets. In this study, the number of customer groups will be estimated by an eigen-decomposition consisting of both eigenvalues and eigenvectors of its kernel affinity matrix.

Table 1
Misclassification counts for the *WINE* dataset.

	Type A	Type B	Type C	Total	Error (%)
FCM	1	12	0	13	7.3
RFCM	1	6	0	7	3.9

4. Experimental results

In this study, two real datasets are used to validate the proposed method: the first is the *WINE* dataset downloaded from http://archive.ics.uci.edu/ml/datasets/Wine and the other is the RFM dataset provided by Taiwan Toyota automobile dealer. In addition, two kinds of evaluation metrics are used to test the performance of the proposed approach: "misclassification error" is used for the *WINE* dataset, and total "within-variance" (see Eqs. (17) and (18)) is used in the *RFM* dataset.

4.1. WINE dataset

The *WINE* dataset consists of 13 features belonging to three physical classes. This dataset was obtained by chemical analysis of wine produced by three different cultivators of Italy. Specifically, it contains 178 samples, with 59 in class 1, 71 in class 2, and 48 in class 3. Besides, the feature variances span a wide range and indicate that outliers are very likely to exist within the dataset. In this study, those potential outliers are intentionally kept to test robust performance of the proposed RFCM. For the problem of supervised classification, a comparison between FCM and RFCM is shown in Table 1. Obviously, the total error count for FCM is 13 whereas for RFCM is only 7. Therefore, RFCM is more capable to handle the noisy dataset than FCM since RFCM can significantly reduce the effect of outliers.

4.2. RFM dataset

Taiwan Toyota automobile retailer provided a motor-maintenance dataset composed of 162 distribution centers that lasted from January 2006 to December 2006. Meanwhile, three features involving R (recency), F (frequency), M (monetary), are used to segment customers and are standardized by the following form (see Eq. (16)).

$$X_S = (X - X_{\min})/(X_{\max} - X_{\min}), \tag{16}$$

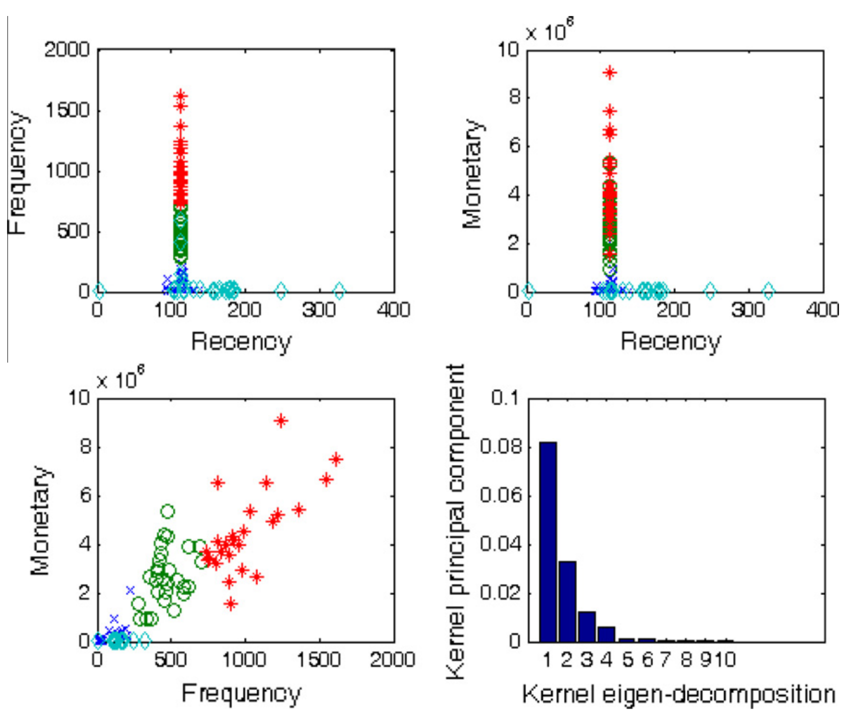

Fig. 1. Customer segmentation using RFM variables.

Table 2
Marketing insights of four customer segments.

	Counts	Symbol	Recency	Frequency	Monetary	Strategy
Group 1	40	Diamond	High	Middle	Low	Enhancing
Group 2	28	Star	Low	High	High	Retention
Group 3	26	Circle	Low	Middle	Middle	Retention
Group 4	48	Cross	High	Low	Middle	Enhancing

Table 3
Performance evaluation for RPCM and RFCM.

		Common within variance (*COVA*)		Individual within variance (*INVA*)
With outliers	PCM	2.09×10^{16}	FCM	1.76×10^{16}
With outliers	RPCM	1.06×10^{16}	RFCM	1.47×10^{15}
Without outliers	PCM	8.4×10^{14}	FCM	3.99×10^{14}
Without outliers	RPCM	7.6×10^{14}	RFCM	3.79×10^{14}

where $X_S/X_{\max}/X_{\min}$ denote the standardized/maximal/minimal value of the corresponding feature X, respectively. Then, the standardized dataset is directed for the input of kernel eigen-decomposition to specify the number of clusters in advance. Obviously, the optimal number of segments is suggested as 4 (see Fig. 1).

Secondly, 20 outliers are successfully identified via RFCM and they are removed prior to clustering. Using "RFM" features, RFCM is adopted for customer segmentation and their marketing insights are shown in Table 2. Apparently, group 2 and group 3 are the so-called gold segments because they visit the company "recently" and purchase "regularly". By contrast, the other groups need to be enhanced to increase their purchasing frequency (for group 4) or monetary (for group 1). More importantly, higher "recency" and lower "frequency" or lower "monetary" usually indicates the higher possibility of customers' defection in the future. Hence, companies need to spend more effort to increase customers' satisfaction or loyalty since it is much easier than acquiring new customers from their business competitors.

Furthermore, to evaluate the performance of various schemes, the objective of clustering can be simply described as: to partition a set of objects into specific groups such that the data *within* the same cluster is as *homogeneous* as possible and the data *between* each cluster is as *heterogeneous* as possible. Hence, the total "within-variance" which describes how well and how compact various clusters are constructed is suggested as a performance metric (Lee et al., 2004; Shin & Sohn, 2004; Vellido et al., 1999).

Specifically, the total "within-variance" w.r.t. the common centroid (see Eq. (17) for *COVA*) or w.r.t. various individual centroids (see Eq. (18) for *INVA*) are suggested in "outlier detection" or "robust segmentation", respectively. To determine the common centroid for outlier detection, RPCM demonstrates its robust superiority over PCM owing to its less *COVA* (see Table 3). Similarly, in terms of lower *INVA*, RFCM also outperforms FCM significantly when applied to the noisy dataset but the difference between FCM and RFCM is not obvious when the outliers are removed.

$$COVA = \sum_{j=1}^{n} \|x_j - a\|^2, \tag{17}$$

$$INVA = \sum_{i=1}^{c} \sum_{j=1}^{n} \|x_j - c_i\|^2, \tag{18}$$

where x_j represents the instance composed of *RFM* features, a and c_i, respectively represent the common centroid and the centroid of ith segment of the whole dataset.

5. Conclusions

Even much work has been done in the area of customer segmentation, the evaluation of robust performance with respect to outliers has not received strong attention that it desires so far. In this paper, a hybrid approach that incorporates kernel induced fuzzy clustering techniques namely RPCM and RFCM, is presented to detect outliers efficiently and to segment customers more effectively. Based on the typicalness degree of RPCM, the outlier possibility of each instance within the whole dataset is easily obtained without the need of labeled data samples in advance. Similarly, by the aid of kernelized belongness membership, RFCM is more capable to achieve robust segmentation when applied to the noisy dataset. Two real datasets including the *WINE* and the *RFM*, are used to validate the proposed approach. More importantly, the suggested method is very promising to be applied to other business areas, such as financial fraud detection (Dorronsoro, Cinel, Sánchez, & Cruz, 1997), computer intrusion detection (Chen, Hsu, & Shen, 2005) and telecommunication churn management (Hadden, Tiwari, Roy, & Ruta, 2005).

Acknowledgements

The authors would thank Taiwan Toyota motor dealer for providing the *RFM* dataset of her downstream distribution centers. This research is financially supported by National Science Council of Taiwan under Contract 97-2410-H-130-025.

References

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. New York: John Wiley & Sons.
Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.
Chen, W. H., Hsu, S. H., & Shen, H. P. (2005). Application of SVM and ANN for intrusion detection. *Computers & Operations Research, 32*, 2617–2634.
Chen, C. H., Khoo, C. L., & Yan, W. (2002). A strategy for acquiring customer requirement patterns using laddering technique and ART2 neural network. *Advanced Engineering Informatics, 16*, 229–240.
Davé, R. N., & Krishnapuram, R. (1997). Robust clustering models: A unified view. *IEEE Transactions on Fuzzy Systems, 5*(2), 270–293.
Davé, R. N., & Sen, S. (2002). Robust fuzzy clustering of relational data. *IEEE Transactions on Fuzzy Systems, 10*(6), 713–726.
Dorronsoro, J. R., Cinel, F., Sánchez, C., & Cruz, C. S. (1997). Neural fraud detection in credit card operations. *IEEE Transactions on Neural Networks, 8*(4), 827–834.
Girolami, M. (2002). Mercer kernel based clustering in the feature space. *IEEE Transactions on Neural Networks, 13*(3), 780–784.
Ha, S. H. (2007). Applying knowledge engineering techniques to customer analysis in the service industry. *Advanced Engineering Informatics, 21*, 293–301.
Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2005). Customer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research, 34*, 2902–2917.
He, Z., Xu, X., Huang, J. Z., & Deng, S. (2004). Mining class outliers: Concepts, algorithms and applications in CRM. *Expert Systems with Applications, 27*, 681–697.
Hodge, V. J., & Austin, J. (2004). Survey of outlier detection methodologies. *Artificial Intelligence Review, 22*, 85–126.
Huang, J. H., Tzeng, G. H., & Ong, C. S. (2007). Marketing segmentation using support vector clustering. *Expert Systems with Applications, 32*, 313–317.
Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Survey, 3*(3), 264–323.
Kirishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems, 1*(1), 98–110.
Kuo, R. J., Ho, L. M., & Hu, C. M. (2002). Integration of self-organized feature map and *K*-means algorithm for market segmentation. *Computers & Operations Research, 29*, 1475–1493.
Lee, S. C., Suh, Y. H., Kim, J. K., & Lee, K. J. (2004). A cross-national market segmentation of online game industry using SOM. *Expert Systems with Applications, 27*, 559–570.
Lin, C. C., & Chen, A. P. (2004). Fuzzy discriminant analysis with outlier detection by genetic algorithm. *Computers & Operations Research, 31*, 877–888.
Markou, M., & Singh, S. (2003a). Novelty detection: A review, Part 1: Statistical approaches. *Signal Processing, 83*(12), 2481–2497.
Markou, M., & Singh, S. (2003b). Novelty detection: A review, Part 2: Neural network based approaches. *Signal Processing, 83*(12), 2499–2521.

McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID and logistic regression. *Journal of Business Research, 20*, 656–662.

Ozer, M. (2001). User segmentation of online music services using fuzzy clustering. *Omega, 29*, 193–206.

Ozer, M. (2005). Fuzzy c-means clustering and internet portals: A case study. *European Journal of Operational Research, 164*, 696–714.

Shin, H. W., & Sohn, S. Y. (2004). Segmentation of stock trading customers according to potential value. *Expert Systems with Applications, 27*, 27–33.

Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications, 27*, 265–276.

Vellido, A., Lisboa, P. J. G., & Meehan, K. (1999). Segmentation of the on-line shopping market using neural networks. *Expert Systems with Applications, 17*, 303–314.

Wang, C. H. (2009). Outlier identification and market segmentation using kernel based clustering techniques. *Expert Systems with Applications, 36*, 3744–3750.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks, 16*(3), 645–678.