International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:1, No:4, 2007

Customer Segmentation in Foreign Trade based on Clustering Algorithms

Case Study: Trade Promotion Organization of Iran

Samira Malekmohammadi Golsefid, Mehdi Ghazanfari, Somayeh Alizadeh

***Abstract*—**The goal of this paper is to segment the countries based on the value of export from Iran during 14 years ending at 2005. To measure the dissimilarity among export baskets of different countries, we define Dissimilarity Export Basket (DEB) function and use this distance function in K-means algorithm. The DEB function is defined based on the concepts of the association rules and the value of export group-commodities. In this paper, clustering quality function and clusters intraclass inertia are defined to, respectively, calculate the optimum number of clusters and to compare the functionality of DEB versus Euclidean distance. We have also study the effects of importance weight in DEB function to improve clustering quality. Lastly when segmentation is completed, a designated RFM model is used to analyze the relative profitability of each cluster.

***Keywords*—**Customers segmentation, Customer relationship management, Clustering, Data Mining

I. Introduction

THE concept of segmentation is central to customer relationship management(CRM). Segmentation means partitioning a population of customers into different segments, considering the most within-segment homogeneity and between segment heterogeneity. Segmentation is valuable because it allows the end user to look at the entire data base from a much higher level. It also allows company to differentially treat consumers in different segments.

One-to-one marketing is the ideal marketing strategy, in which every marketing campaign or product is optimally targeted for each individual customer; but this is not always possible. Thus, segmentation is required to distinguish similar clients and put them together in a segment. Doubtlessly using segmentation to understand customer's needs is much easier, faster and more economical than uniquely investing to understand them particularly[1].

There are different methods for segmentation. Hyunseok Hwang ,Taesoo Jung and Euiho Suh in 2006 introduced a framework for analyzing customer value and segmenting customers based on their current value, potential value, and customer loyalty[6]. C.-Y. Tsai, C.-C. Chiu in 2004 developed a market segmentation methodology based on product specific variables such as purchased items and the associative monetary expenses from the transactional history of customers to address the unreliable results of segmentation based on general variables like customer demographics [7]. H.W. Shina and S.Y. Sohn in 2004 used three clustering methods (K-means, self-organizing map, and fuzzy K-means) for segmentation to find properly graded stock market brokerage commission rates based on transactional data[8]. Jedid-Jah Jonkera, Nanda Piersmab and Dirk Van den Poelc, in 2004 presented an approach to segment customers based on R(Recency), F(Frequency), and M(Monetary value) variables[9].

In this study, we segment customers of Trade Promotion Organization of Iran using a proposed distance function which measures dissimilarities among export baskets of different countries based on association rules concepts. Later, in order to suggest the best strategy for promoting each segment, we analyze each cluster using RFM model. Variables used for segmentation criteria are "the value of the group-commodities", "the type of group-commodities" and "the correlation between export group-commodities".

This paper is organized as follows. In section II, the countries segmentation methodology developed in this research is described. In section III, RFM model is applied to analyze the value of each segment. Subsequently, in section □, the proposed countries segmentation Methodology is put into practice using Trade Promotion Organization of Iran (TPO) data bases. Finally in section □, the implication of the results is discussed and further study areas are suggested.

II. Countries Segmentation Methodology

After data preparation which is the first essential part of data mining procedures we propose Dissimilarity Export Basket (DEB) function based on association rules concepts for clustering countries by K-means algorithm. To prove the superiority of clustering using DEB function as the distance function in K-means algorithm against using Euclidean distance function in this research, we calculate the quality of clustering. Then, we determine the optimal number of clusters and finally we improve DEB function by considering the importance of time of transactions. Fig 1 illustrates a countries segmentation methodology.

Samira Malek Mohamadi Golsefid is with the Industrial Engineering Department, Iran University of Science and Technology, Master of Information Technology Engineering, Tehran, Iran (phone: 98-912-3451904; e-mail: samira.malek@gmail.com).

Mehdi Ghazanfari., is with the Industrial Engineering Department, Iran University of Science and Technology, Associate professor, Tehran, Iran (e-mail: mehdi@iust.ac.ir).

Somayeh Alizadeh is with the Industrial Engineering Department, Iran University of Science and Technology, PHD candidate, Tehran, Iran (e-mail: s_alizade@mail.iust.ac.ir).

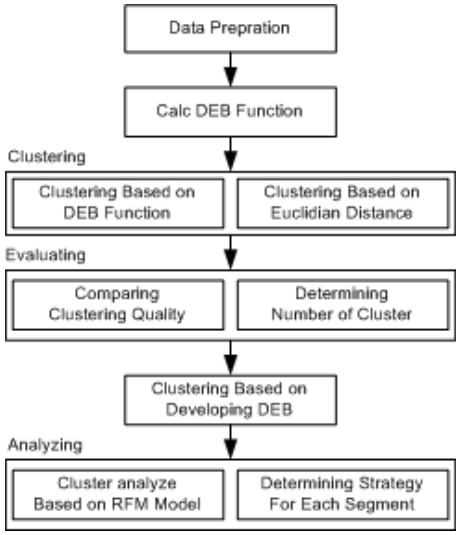


Fig. 1 The proposed Countries Segmentation Methodology

A. Data Preparation

Data preparation is the step we take to integrate the data base as per the purpose of the study and the applied algorithms requirements. The parameters which are used in the DEB function are listed below along with their meanings.

G: The set of all export group-commodities from Iran

T^0 : Export transaction data base

t^0 : consists of at least one row of T^0 database that records a Country ID, transaction time, group-commodity, monetary value.

c_i : *i*th country

g_{ia} : *a*th group-commodity which is exported to *i*th country

m_{ia} : The total monetary value of g_{ia} which is exported to the *i*th country.

To observe a country export behavior from Iran, we need to retrieve all exported group-commodities to that country along with the aggregated monetary expenses for these group-commodities during the specified 14 years period. Let $CountryCode_i$ be the country ID of a c_i; $GoodGroup_i = \{g_{ia} | g_{ia} \in G\}$ be the set of group-commodities exported to c_i: Let $moneyset_i = \{m_{ia} | a = 1,...,\|GoodGroup_i\|\}$ be the set of aggregated monetary value of the exported group-commodities. Therefore, an aggregated record that describes the export behavior of the country c_i can be represented as $t_i^c = (CountryCode_i, GoodGroup_i, moneyset_i)$ and stored in the cumulative transaction database T^c.

B. Dissimilarity Export Basket Function (DEB)

To measure the dissimilarity between two countries, first, we calculate the export association between each two group-commodities based on support concept in the association rule.

$$s(\{g_i, g_j\}) = \frac{\left\|\{t^0 \in T^0 | t^0 contains \{g_i, g_j\}\}\right\|}{\left\|T^0\right\|} \quad \text{where} \quad g_i, g_j \in G \qquad (1)$$

$s(\{g_i, g_j\})$ is the proportion of transactions containing the $GoodGroup\{g_i, g_j\}$ to all transactions in T^0. However, the support value could be very low if an itemset contains rarely co-exported group-commodities [2]. To measure the distance in DEB function, we calculate the mutual correlation between the export group-commodities. The correlation between two group commodities is defined based on coherence(Jaccard) as below[13].

$$\gamma(\{g_i, g_j\}) = \frac{s(\{g_i, g_j\})}{s(g_i) + s(g_j) - s(\{g_i, g_j\})} \qquad (2)$$

$\gamma(\{g_i, g_j\})$ is ranged from 0 to 1. If $g_i = g_j$, $\gamma(\{g_i, g_j\}) = 1$.

Calculating the similarity between country c_i and c_j is the next step. To do so let the aggregated record t_i^c for country c_i be $(CountryCode_i, GoodGroup_i, moneyset_i)$ where $GoodGroup_i = \{g_{i1}, g_{i2},..., g_{is}\}$, $moneyset_i = \{m_{i1}, m_{i2},..., m_{is}\}$ and the aggregated record t_j^c for country c_j be $(CountryCode_j, GoodGroup_j, moneyset_j)$ where $GoodGroup_j = \{g_{j1}, g_{j2},..., g_{jt}\}$ and $moneyset_j = \{m_{j1}, m_{j2},..., m_{jt}\}$. The similarity function is defined as:

$$\begin{cases} Sim(c_i, c_j) = \dfrac{\sum_{a=1}^{s}\sum_{b=1}^{t} m_{ia} \times m_{jb} \times \gamma(\{g_{ia}, g_{jb}\})}{\sum_{a=1}^{s}\sum_{b=1}^{t} m_{ia} \times m_{jb}} & c_i \neq c_j \\ Sim(c_i, c_j) = 1 & c_i = c_j \end{cases} \qquad (3)$$

Accordingly, the distance between two countries or the dissimilarity between export baskets of two countries - c_i, c_j - is defined as under:

$$\begin{cases} DEB(c_i, c_j) = 1 - Sim(c_i, c_j) & c_i \neq c_j \\ DEB(c_i, c_j) = 0 & c_i = c_j \end{cases} \qquad (4)$$

The above equation considers the correlation between each two group-commodities as well as the monetary value of each country. For example, if group-commodity A is often co-exported with group-commodity B while less with group-commodity C, the co-export association between A and B should be stronger than that between A and C. Ignoring these associations and treating all group-commodities equally creates a similarity bias.

C. Clustering

The process of grouping a set of objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [14].In this study, K-means algorithm is used for country clustering. To measure the distance between countries, DEB

function and Euclidean distance function are used.

K-means Algorithm Based on Euclidian Distance

K-means clustering proceeds in the following order. Firstly, K number of observations is randomly selected among all N number of observations according to the number of clusters. They become centers of initial clusters. Secondly, for each of remaining N–K observations, find the nearest cluster in terms of the defined distance function. After each observation is assigned the nearest cluster, the center of the cluster is recomputed. Lastly, after the allocation of all observation, calculate the distance between each observation and cluster's center point and confirm whether it is allocated to the nearest cluster or not [2].

The distance (dissimilarity) between countries c_i, c_j in the terms of Euclidean distance is as per below:

$$Dist(c_i, c_j) = \sqrt{\sum_{l=1}^{L} (m_{il} - m_{jl})^2} \quad (5)$$

where L equals the total number of export group-commodities.

During the next sections we use the introduced DEB function in K-means algorithm instead of Euclidean distance and the results are compared with each other.

K-means Algorithm based on DEB function

For clustering based on DEB function, we use equation (4) as the distance meter. The distances between each two countries are calculated and stored in a distance matrix.

$$Dis = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1j} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2j} & \dots & d_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{i1} & d_{12} & \dots & d_{ij} & \dots & d_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nj} & \dots & d_{nn} \end{bmatrix}_{n \times n}$$

where n shows the number of countries, d_{ij} represents the distance between c_i and c_j, $i, j \le n$. According to equation (4), the main diagonal which represents the distance between each country with itself is zero.

$$Dis = \begin{bmatrix} 0 & d_{12} & \dots & d_{1j} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2j} & \dots & d_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{i1} & d_{12} & \dots & 0 & \dots & d_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nj} & \dots & 0 \end{bmatrix}_{n \times n}$$

D. Clustering Quality Evaluation

There are many different criteria for evaluating clusters quality. The best clustering is the one which its within-cluster homogeneity or similarity and between cluster heterogeneity or dissimilarity are the most. All clustering quality criteria are based on this concept. Some of them only satisfy one of these two parameters and some consider both for evaluation purpose [14].

Michaud has presented a method called "intraclass" method for comparing the different clustering techniques performances in 1997. Intraclass inertia is a measure of how compact each cluster is when the number of cluster is fixed. Usually the variables are scaled to be in the same range [15]. Since we have discrete data in this project, we replace the means of cluster with its center. In order to calculate the compactness of each cluster, the intraclass inertia I_n of cluster n is defined as

$$I_n = \sum_{c_i \in O^n{}_j} Dist(c_i, c^n) \quad (6)$$

Finally, the intraclass inertia F(K) for a given K clusters is defined as:

$$F(K) = \frac{1}{K} \sum_{n=1}^{K} \sum_{c_i \in O^n} Dist(c_i, c^n) \quad (7)$$

One can see that F(K) is the average squared Euclidean distance between each observation and its cluster mean. The lower is the F(K), the clusters are more compacted and the clustering is better[8].

Since the introduced distance function (DEB) is not available in ready packages of k-means algorithms in data mining software, a program is written by Visual Basic Language to calculate the Distance Matrix. Then clustering is done based on the Dis matrix using R software. Also some other subprograms are written to calculate clustering quality function, clusters intraclass inertia and optimum number of clusters. The relation between the sub programs, data mining software and data bases which are used in this study are shown in Fig2.

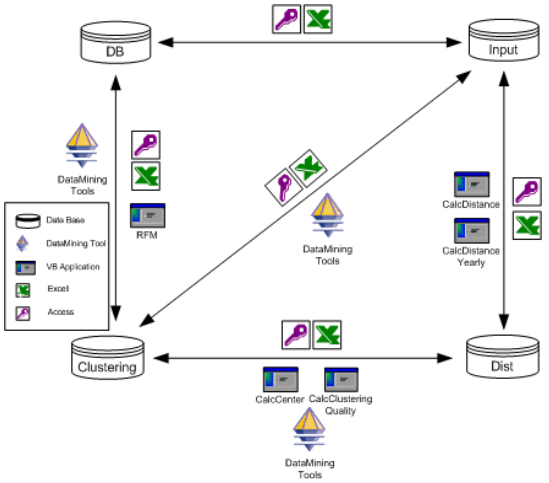


Fig. 2 The relation between the sub programs, data mining software and data bases which are used

E. Optimal Number of Clusters

The objective of DEB function in K-means algorithm is to maximize the sum of the similarities between a cluster center and all countries in the same cluster, and to minimize the sum of the similarities between two cluster centers in different clusters. Let $O = \{c^n | n = 1, \dots, K\}$ be the set of K cluster centers, c^n be the cluster center of the nth cluster, $O^n = \{ c_i | i = 1, \dots, \| T^c - O \| \}$

be the set of remaining countries that were not selected as a cluster center and T^c be the aggregated records which clustering is performed on it. Therefore, the quality of the clustering result with K clusters can be defined as Eq. (8):

$$\rho(K) = \frac{1}{K}\sum_{n=1}^{K}\left(\underset{1\le m\le K, m\ne n}{Min}\left\{\frac{\eta_n + \eta_m}{\delta_{nm}}\right\}\right) \quad (8)$$

$$\eta_n = \frac{1}{\left\|O^n\right\|}\sum_{c_i \in O^n} Sim(c_i, c^n) \quad (9)$$

$$\eta_m = \frac{1}{\left\|O^m\right\|}\sum_{cj \in O^m} Sim(c_j, c^m) \quad (10)$$

$$\delta_{nm} = Sim(c^n, c^m) \quad (11)$$

Eq. (9) defines η_n as the average of similarities between cluster center c^n and all customers in cluster O^n Eq. (10) states that η_m is the average of the similarities between cluster center c^m and all customers in cluster O^m: Eq. (11) defines δ_{nm} as the similarity between c^n and c^m.

$\rho(K)$ is calculated considering the Clustering Quality Function which is presented in equation (8) between the lower boundary s and the higher boundary t. The K value which results the maximum $\rho(K)$ amount, is the optimal number of clusters.

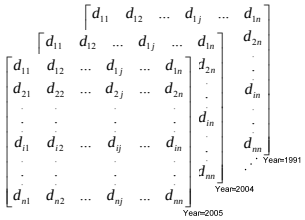
$$\hat{K} \equiv \arg\underset{s\le K\le t}{Max}\{\rho(K)\} \quad (12)$$

Using Eq. (12), an optimal value for K can be objectively determined for country segmentation [7].

F. DEB Function Improvement

By aggregating the records of the specified 14 years period, we have actually ignored the importance of time and therefore the correlation between the exported items in different years is ignored as well.

In order to address the said issue, we have calculated the distance matrix of each year separately and considered importance weight in our calculation as shown in equation (13) to acquire the total distance matrix for the whole period of 14 years. Importance weights are obtained based on experts' comments and by using AHP method.



$$Dis = \sum_{y=1}^{Y} w_y D_y \quad (13)$$

D_y represents distance matrix of y year, w_y is the yth year coefficient and $1 \le y \le 14$.

III. A RFM MODEL FOR PROFITIBILITY EVALUATION

The RFM model measures the customer value based on Recency (R), Frequency (F), and Monetary (M) criteria [16]. Recency measures the interval between the most recent time we had export to each of the countries and the analyzing time. Frequency measures the export frequency within a specified period. Monetary measures the total monetary value within a specified period. This model is used to analyze the relative profitability for each country cluster from the segmentation result after executing the proposed DEB algorithm. With this model, an enterprise can quickly find the target clusters and adjust its marketing programs and business initiatives to provide the right products, services and resources to the target clusters. Based on this scheme, the value of a country can be represented as:

$$V(c_i) = W^R \times R(c_i) + W^F \times F(c_i) + W^M \times M(c_i) \quad (14)$$

where $R(c_i)$, $F(c_i)$, and $M(c_i)$ represent the scores for country c_i in terms of the R, F, and M criteria, respectively. W^R, W^F, and W^M represent the importance weights for the R, F, and M criteria, respectively. In addition, $W^R + W^F + W^M = 1$.

The scores can vary depending on the types of applications and scoring approaches [16]. The scores retrieved from the original transaction database are normalized before calculating the value of a country. Therefore, the $R(c_i)$ ، $F(c_i)$ and $M(c_i)$ scores can be redefined as follows:

$$R(c_i) = \frac{Q^R - Q_{Min}{}^R}{Q_{Max}^R - Q_{Min}^R} \quad (15)$$

$$F(c_i) = \frac{Q^F - Q_{Min}{}^F}{Q_{Max}^F - Q_{Min}^F} \quad (16)$$

$$M(c_i) = \frac{Q^M - Q_{Min}{}^M}{Q_{Max}^M - Q_{Min}^M} \quad (17)$$

where Q^R ، Q^F and Q^M represent the original values for a country c_i according to the definition of R, F, and M. $Q_{Min}{}^R$ ، $Q_{Min}{}^F$ and $Q_{Min}{}^M$ represent the minimum values of R, F and M and $Q_{Max}{}^R$ ، $Q_{Max}{}^F$ and $Q_{Max}{}^M$ represent the maximum values of the same.

The profitability of the nth country cluster O^n can be acquired by calculating the average for all country values in the cluster. This can be defined as Eq. (18):

$$V(O^n) = W^R \times R(O^n) + W^F \times F(O^n) + W^M \times M(O^n) \quad (18)$$

$$R(O^n) = \frac{\sum_{c_i \in O^n} R(c_i)}{\left\|O^n\right\|} \quad (19)$$

$$F(O^n) = \frac{\sum_{c_i \in O^n} F(c_i)}{\|O^n\|} \quad (20)$$

$$M(O^n) = \frac{\sum_{c_i \in O^n} M(c_i)}{\|O^n\|} \quad (21)$$

where $R(O^n)$ ، $F(O^n)$ and $M(O^n)$ represent the scores for the nth cluster O^n in terms of R, F, and M respectively. After the profitability for all clusters is known, the clusters are ranked and the most important one is identified. This is helpful for an enterprise for planning and determining long time and short time strategies to offer a better service to specific customer clusters.

IV. Case Study

To demonstrate the performance of the proposed countries segmentation methodology, we use the export data of the specified 14 years period from the TPO.

According to retrieved information from TPO databases, Islamic republic of Iran exports goods and services to total 210 countries. These goods and services are of a wide range which is about 16000 types. In this study, goods and services are categorized to 99 export group-commodities based on HS code system. There were 222078 transactions generated jointly by 210 countries in transaction data base containing 99 export group-commodities.

A. Clustering

When data preparation, data integration and extracting data with appropriate format are done, we started clustering the countries using K-Means algorithm. The clustering is done based on two different distance functions, Euclidean distance function (equation 5) and DEB function (equation 4). The intraclass inertia is calculated using equation 7 for both clustering methods. The lower is the result of eq 7, the compactness is more and consequently the clustering is better. The compactness of the clustering based on DEB function is more than the one based on Euclidean distance. The importance weights which are resulted from AHP method are applied in equation 13 to improve the DEB function. As it is shown in Table I the improved DEB function resulted the best clustering.

TABLE I
THE CLUSTERING COMPACTNESS COMPARISON RESULT USING EUCLIDEAN, DEB AND IMPROVED DEB DISTANCE FUNCTIONS IN K-MEANS ALGORITHM

Clustering Distance Function	Cluster Intraclass
Euclidean Distance Function	1.2949
DEB Function	0.1200
Improved DEB Function	0.0742

To determine the optimum number of clusters, we used equation 9 to calculate $\rho(K)$ for different K values. K shows the optimum number of clusters, where $\rho(K)$ is maximum.

The quality of clustering for different K values, $\rho(K)$, is calculated and tabulated in Table II for $2 \le K \le 9$.

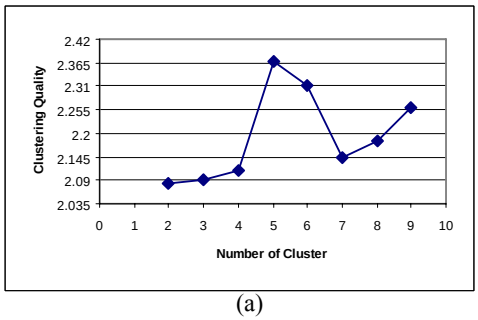
TABLE II
CLUSTERING QUALITY COMPARISON FOR $2 \le K \le 9$

Number of Cluster	DEB Function	Euclidean Distance Function
2	2.0805	1.9904
3	2.0908	1.9903
4	2.1118	1.9905
5	2.3699	1.9921
6	2.3104	1.9913
7	2.1428	1.9908
8	2.1821	1.9903
9	2.2616	1.9921

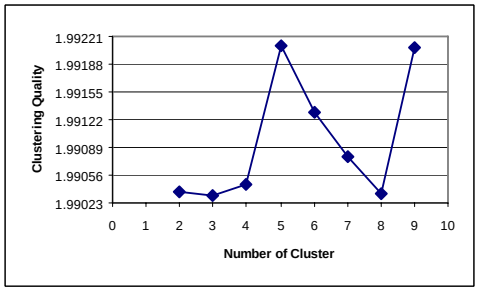
As demonstrated in Fig 3, the highest $\rho(K)$ amount is happened at K=5 and therefore the optimal number of clusters is 5.

B. Clusters Analysis by RFM Model

To analyze the clusters using RFM model, R, F and M criteria should be calculated for all countries and clusters. $R(C_i)$ is calculated using equation 15 and is between 1 and 14. When $R(C_i)$= 14, it means that Iran has exported goods/services to that country very recently. $F(C_i)$ is calculated using equation 16 and similarly the values are between 1 and 14. In order to calculate each country monetary value, we need to calculate the total export volume to that particular country within the specified period. Using equation 17, $M(C_i)$ is acquired. The weight for the R, F, M criteria were set as $W^R = 0.2, W^F = 0.5, W^M = 0.3$.



(a)



(b)

Fig. 3 The clustering quality result using (a) DEB and (b) Euclidean distance functions.

TABLE III
THE RFM ANALYSIS RESULT

Cluster Number	$R(O^n)$	$F(O^n)$	$M(O^n)$	$V(O^n)$	Number of Cluster Members
4	0.9483	0.7532	0.0547	0.5827	101
5	0.9517	0.6054	0.0219	0.4996	23
1	1.0000	0.5846	0.0004	0.4924	5
3	0.8861	0.5901	0.0089	0.4749	79
2	1.0000	0.2692	0.0070	0.3367	2

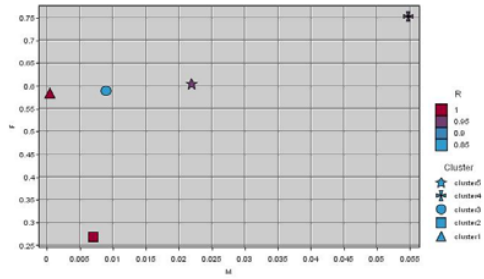


Fig. 4 The two-dimensional RFM model graph.

According to experts, frequency of export is of higher importance comparing to Recency and Monetary; therefore F value is greater than R and M in this case study. After a series of calculations using equations (18) and (19), the value of each cluster is determined and the results are summarized in Table III.

To determine customer relationship management strategy, we analyzed clusters using RFM model as follows:

The relationship between the countries of cluster 1 with Iran was comparatively longstanding during the specified period but the total monetary values of exported commodities were very low. Another feature of the objects of this cluster is that very recently they had relationship with Iran. Therefore, these are rather loyal customers with high average cluster value ($V(O^1)$) who we have to increase the volume of export to them and if we fail to do so, we should try to minimize the costs of export.

The objects of cluster 2 are those we have very recent relationship with them, the frequency of export to these countries is very low and we have not a longstanding relationship with them in our records. Hence, these are the new customers, we have to know them more closely in order to grow the volume of export to them and make the relationship long term and more beneficial.

The members of cluster 3 are those we had previously relationship with them but no transaction is done recently. In other words, these are inactive customers who are about churn reduction and are unfortunately great in number. We have to investigate the causes and determine suitable strategy to avoid losing them.

Cluster 4 is the most valuable cluster, whose members made the highest monetary value. Iran has long term relationship with these countries which was continued to the very end of the specified period of the analyzing. So these are loyal and active customers of ours with the highest profitability level whom we should keep and retain and avoid losing any of them.

The members of cluster 5 are very similar to members of cluster 1 in view of Recency and frequency but with a dramatic difference in monetary value which has made them the second important cluster in this research. Considering our relationship with these countries it is mostly possible that we can increase the volume of export applying appropriate strategies.

To provide a clear view for marketing programs, a 2-dimensional and 3-dimentional RFM model graph is depicted in Fig 4 and Fig 5, respectively.

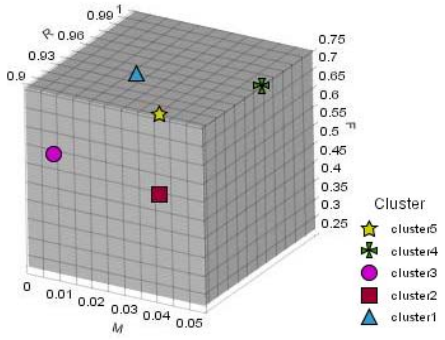


Fig. 5 The three-dimensional RFM model graph.

V. CONCLUSION

Since the increased importance is placed on customer satisfaction in today's business environment, many organizations are focusing on the notion of customer loyalty and profitability for increasing market share and customer satisfaction. CRM is emerging as core competence of an organization. Customer segmentation is a known approach to understanding the clients which will be lead to determining appropriate marketing policy.

In this study, DEB function is introduced based on association rules and is used to measure distances in K-means algorithm. Improvement in clustering results is observed by replacing DEB function instead of Euclidian distance function in k-means algorithm. Also, it is shown that the performance of DEB function will be increased by considering the importance of time of transactions in the period of analyzing. After segmentation, a designated RFM model is introduced to analyze the relative profitability of each country cluster and determine proper strategies to improve the whole situation. In the future, we may use descriptive and demographic data to refine DEB function.

REFERENCES

[1] Berson&Stephen Smith&Kurt Thearling , "Bulding Data Mining Application For Crm" , *Mcgraw-Hill*, 2001, Ch 13.

[2] Nong Ye, "The Handbook of Data Mining", *Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey London*, 2003, Ch2,10.

[3] Hsiao-Fan Wang & Wei-Kuo Hong, "Managing Customer Profitability In A Competitive Market By Continuous Data Mining", *Department Of Industrial Engineering And Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, Roc*, 2005.

[4] Chris Rygielski A & Jyun-Cheng Wang B, David C. Yen A, (2002), "Data Mining Techniques For Customer Relationship Management", *Technology In Society 24,* 2002, Pp483–502.
[5] Hyunseok Hwang, Taesoo Jung, Euiho Suh, "An Ltv Model And Customer Segmentation Based On Customer Value: A Case Study On The Wireless Telecommunication Industry", *Expert Systems With Applications 26,* 2004, Pp181–188.
[6] Su-Yeon Kim , Tae-Soo Jung, Eui-Ho Suh, Hyun-Seok Hwang, "Customer Segmentation And Strategy Development Based On Customer Lifetime Value: A Case Study", *Expert Systems With Applications 31*, 2006, Pp101–107.
[7] C.-Y. Tsai, C.-C. Chiu, "A Purchase-Based Market Segmentation Methodology", *Expert Systems With Applications 27*, 2004, Pp265–276.
[8] H.W. Shina, S.Y. Sohnb, "Segmentation Of Stock Trading Customers According To Potential Value",*Expert Systems With Applications 27,* 2004, Pp 27–33.
[9] Jedid-Jah Jonkera, Nanda Piersmab, Dirk Van Den Poelc, "Joint Optimization Of Customer Segmentation And Marketing Policy To Maximize Long-Term Profitability", *Expert Systems With Applications 27*, 2004, Pp159–168.
[10] Pauline A. Wilcox, Calin Gurau, "Business Modeling With Uml: The Implementation Of Crm System For Online Retailing", *Jornal Of Retailing And Consumer Services 10,* 2003, Pp181-191.
[11] Wagner A. Kamakura, Michel Wedel, Fernando De Rosa, Jose Afsonso Mazzon, "Cross-Selling Through Database Marketing: A Mixed Data Factor Analyzer For Data Augmentation And Prediction", *Intern J Of Research In Marketing 20,* 2003, Pp45-65.
[12] Arindam Banerjee And Joydeep Ghosh, "Clickstream Clustering Using Weighted Longest Common Subsequences" , *Dep Of Electrical Engineering University Of Texas At Austin,* 2002.
[13] Jiawei Han And Micheline Kamber,"Data Mining: Cluster Analysis", *Department Of Computer Science ,University Of Illinois At Urbana-Champaign,* 2006.
[14] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction To Data Mining", *Pearson Addison Wesley*, 2006.
[15] Nair, G. J., & Narendran, T. T., "Cluster Goodness: A New Measure Of Performance For Cluster Formation In The Design Of Cellular Manufacturing Systems", *International Journal Of Production Economics*, 1997, Pp49–61.
[16] Hughes, A. M., "Strategic Database Marketing: The Masterplan For Starting And Managing A Profitable, Customer-Based Marketing Program", *Probus Pub Co.,* 1994.