A new methodology to study customer electrocardiogram using RFM analysis and clustering

Asso Hamzehei [a] , Mohammad Fathian[a*] , Hamid Farvaresh[b] and Mohammad Reza Gholamian[a]

[a]School of Industrial Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran
[b]Department of Industrial Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran

A R T I C L E I N F O	A B S T R A C T
	One of the primary issues on marketing planning is to know the customer's behavioral trends. A customer's purchasing interest may fluctuate for different reasons and it is important to find the declining or increasing trends whenever they happen. It is important to study these fluctuations to improve customer relationships. There are different methods to increase the customer's willingness such as planning good promotions, an increase on advertisement, etc. This paper proposes a new methodology to measure customer's behavioral trends called customer electrocardiogram. The proposed model of this paper uses K-means clustering method with RFM analysis to study customer's fluctuations over different time frames. We also apply the proposed electrocardiogram methodology for a real-world case study of food industry and the results are discussed in details.

1. Introduction

During the past two decades, information technology has dramatically changed the relationship between the customers and businesses by increasing new opportunities for marketing (Ngai et al., 2009). In order to better understanding the customers' needs we could gather all sales' transactions to analyze using different scientific methods. A good analysis of data could help understand customers' behavior through transforming the data into some classified information. The classified data will help us extract the necessary knowledge, which could lead for better marketing planning (Bottcher et al., 2009). Customer behavioral analysis is useful for detecting prospective customers who are wealthy and we wish to retain them as much as possible. One of the primary questions to analyze the data is to choose an appropriate tool such as data mining to get the necessary results (Yuantao & Siqin, 2008). Data mining is one of the most popular tools for predicting and analyzing company's data to discover useful trends and patterns. Data mining is the result of applying sophisticated modeling techniques from the diverse fields of statistics, artificial intelligence, and database management (Yuantao & Siqin, 2008; Han & Kambert, 2001). There are many evidences to indicate that purchasing records are

important for direct marketing. In fact, demographics and purchasing history of consumers play important role on profitability of direct marketing activities and the ability to target consumers, effectively (Kaefer et al., 2005; Kim & Street, 2004).

Customer segmentation tries to group customers with comparable desires and purchasing behaviors in accordance with consistency among members in the same group and diversity among different groups. Generally, customer's segmentation is associated with statistical segmentation, lifestyle segmentation, behavior segmentation and benefit segmentation. Through the customer segmentation, the function of behavior segmentation is to classify customers in terms of behavior pattern of the existed customers in database (Ying & Feng, 2008; Minghua, 2008; Xiang-Bin & Yi-Jun, 2006). Various groups which have been derived based on customer behavior analysis, will contribute to the effectiveness of corporate marketing activities, increase customer satisfaction and loyalty, and facilitate enterprises to make competitive advantages (Minghua, 2008). There are many works such as decision trees and clustering dedicated to study the behavior segmentation based on the purchase behavior. Customer behavior segmentation has been widely used in various industries and areas, including e-commerce, retail markets, tourism, marketing, etc.

By segmenting the consumer markets according to their behavioral patterns, one can target the acquisition and retention of highly profitable and potentially lucrative customers. One of the popular methods for analyzing customer behavior, defining customer segmentation, and market segmentation is RFM analysis. Cheng and Chen used RFM method to value customers and clustered them by K-Means (Cheng & Chen, 2008). According to Hung et al. (Huang et al, 2009), customers with the same pattern of purchasing are only clustered and RFM is used to calculate the value of each cluster. Tsai and Chiu (2004) proposed a market segmentation methodology based on product specific variables such as items purchased and the associative monetary transactional history of customers and they used RFM to analyze the relative profitability of each customer's cluster. Yeh et al. (2009) extended the traditional RFM model by including two parameters, time since the first purchase and churn probability. The organization of this paper is as follows. Section 2 briefly introduces short time series distance, K-means clustering, and RFM analysis. Section 3 describes the proposed methodology, section 4 includes the numerical results for a real-world case study, and section 5 summarizes the contribution of the paper.

2. Problem statement

2.1 Short time series distance

Möller-Levet et al. (2003) used slope to compute the distance between two short time series. The distance between two-time series x_i and v_j is the sum of the squared differences of the slopes in two time series which is as follows,

$$d_{STS} = \sqrt{\sum_{k=1}^{p}\left(\frac{v_{j(k+1)} - v_{jk}}{t_{k+1} - t_k} - \frac{x_{i(k+1)} - x_{ik}}{t_{k+1} - t_k}\right)^2}, \tag{1}$$

where t_k is the time point for data point x_{ik} and v_{jk} and to remove the effect of scale, z standardization of the series is recommended.

2.2 K-Means Clustering

In data mining, k-means clustering is a method of clustering which aims to partition N observations into K clusters in which each observation belongs to the cluster with the nearest mean. Normal evaluation of a proper K is to minimize the inner-cluster variation and maximize the among-cluster

variation. K-means clustering is sensitive to outliers, so, outliers must be deleted before completing clustering (Ying & Feng, 2008; Cheng & Chen, 2008; Farvaresh & Sepehri, 2010).

2.3 RFM Analysis

RFM analysis is the most powerful and the simplest model to extract the value of customers. RFM analysis models three dimensions of customer transactional data, recency, frequency and monetary, to classify customer's behavior. RFM analysis is also used to calculate a score for each customer (Cheng & Chen, 2008; Yeh et al., 2009) where customers with high scores are usually the most highly responsive to promotions, the most likely to repeat a behavior, the most profitable, and vice versa. The details of the definitions of RFM analysis are described as follows (Cheng & Chen, 2008; Yeh et al., 2009; Lingras et al., 2005):

- R represents recency, which is the interval between the time that the latest consuming behavior happens where shorter intervals have bigger values.
- F stands for frequency, which refers to the number of transactions in a particular period, for example, two times per one year, two times per one quarter or two times per one month.
- M also stands for monetary value, which means the amount of money is consumed in a particular period.

3. Research design

In this section, we present a customer behavior analysis methodology which is based on customer segmentation changes in the sequence of time intervals. The proposed model of this paper has three phases. In the first stage of the method, we transform the customer's behavioral trends from the past transactional data. The main works in this phase is preprocessing, setting time intervals, assigning customers via RFM to pre-determined segments, and extracting behavioral trends. Customer's clustering based on behavioral trends is performed in the second phase and the main behavioral trend is determined. This phase consists of two steps, clustering behavioral trends and identifying dominant trends. Consequently, in the third phase named clusters analysis, we analyze clusters through analyzing the clustering, labeling and profiling.

3.1 Customer's behavioral trends derivation

There are different ways of gathering the customers' information such as purchasing goods using credit cards, debit cards, etc. In fact, when a particular customer buys a product, his/her credit information along with other necessary information such as customer's demographic, customer's life styles and product characteristics are stored in a central database. Any study on the customer's intention to buy a specific product can be traced using the database and customer's behavior is derived based on monitoring the purchase trend over a predefined period of time and any supplementary data will also help us achieve more accurate results from customer's behavior analysis. Since the primary focus of this study is to find the behavioral trends of customers, we need to find out whether such trend exists or not. The following steps of the first phase, intend to extract the customers behavioral trends.

3.1.1 Preprocessing

One of the main activities in any data-mining project is to preprocess the data. The goal of this step is to prepare the data to be ready to analyze through studying the purchasing data. The preprocessing operation includes data integration and data transformation.

3.1.2 Setting time intervals

In the customer's behavior changes analysis we can divide the time into different intervals instead of considering the time as a single slope. The advantage of dividing time is positioning of every customer in each time intervals. Therefore, the changes of customer's position over the time are tractable. The problem in this part is to determine how the time intervals are defined and how many time intervals are needed. We have already explained the response for the first part of the question

but the answer for the second question depends on different factors such as the type of industry, density of the historical purchases, and the duration of existence of data. The common ways of dividing the time is based on dividing by day, week, quarter, half year, or year. As we mentioned earlier, the analyzer decide to use one of the dividing factors based on the different conditions.

3.1.3 segmenting customers via RFM to pre-determined segments

In this step of the methodology, customers are segmented into a range of segments for each time interval by scoring RFM. Classification of customers into five clusters is executed in terms of average purchase payments (monetary) and purchase frequency. In this type of segmentation, the recency is hidden in the time frames. In each interval, we calculate the amount of M (monetary) and F (frequency). Since the variables F and M are highly correlated, we calculate two new variables (components) by means of principal component analysis (PCA) and it is expected that the new variable (component) describe most of the variance of the original space. Then the segmentation of the customers is executed based on the first principal component. Table 1 explains how the customers are assigned into five segments.

Table 1
Decision rules for segmentation of customers in each time interval

Segment num.	Distribution of customers
1	$FM = 0$
2	$FM < \mu\,(FM)$
3	$\mu\,(FM) < FM < \mu\,(FM) + \sigma\,(FM)$
4	$\mu\,(FM) + \sigma\,(FM) < FM < \mu\,(FM) + 2\,\sigma\,(FM)$
5	$\mu\,(FM) + 2\,\sigma\,(FM) < FM < \mu\,(FM) + 3\,\sigma\,(FM)$

In Table 1 FM is the new dimension that is the result of PCA on frequency and Monetary, μ and σ are mean and standard deviation of FM, respectively.

3.1.4 Indentifying customers' behavioral trends
Now, we have a sequence of customer's segments in each time interval. Changes of customers among the time intervals show the dynamic behavior of customers. Table 2 represents the sequence of three randomly selected customers, label of segment, among five time intervals.

Table 2
Customer segments in five sequential time interval

Customer	Intervals				
	First	Second	Third	Fourth	Fifth
x_1	2	3	2	4	4
x_2	4	2	3	4	0
x_3	0	1	1	2	0

3.2 Customer's clustering based on behavioral trends

3.2.1 Clustering behavioral trends
This research exploits the K-means clustering method to cluster customer's extracted trends of behavior. The first outcome in this part of methodology is diagnosing outlier trends and removing them from the process of analysis. The second part is to achieve different clusters of trends. These clusters represent different groups of behaviors among interaction of customers with business. In fact, the foundation of this study is based on the extracted clusters. Therefore, the accuracy of clustering of customer trends has a great effect on the consequential results of the investigation.

Most of data clustering techniques use cluster distance to analyze the data. In this research, however, we use cluster behavior instead of distance for our analysis. Therefore, we replace the Euclidean distance function with a fitted one. One of the popular options is to use short time series distance (STSD) function. In our study, since K-means clustering could not determine the best number of clusters, we replaced it with Dunn's index which is an indicator of clustering quality. The Dunn's index is defined as follows,

$$D_{nc} = \min_{t=1,\cdots,nc} \left\{ \min_{j=t+1,\cdots,nc} \left(\frac{d(c_t,c_j)}{\max\limits_{k=1,\cdots,nc} \mathrm{diam}(c_k)} \right) \right\}, \tag{2}$$

where

$$d(c_i,c_j) = \min_{x\in c_i, y\in c_j} d(x,y), \tag{3}$$

$$\mathrm{diam}(c) = \max_{x,y\in c} d(x,y). \tag{4}$$

$D(c_i,c_j)$ is the distance between two clusters c_i and c_j. diam(c) is the maximum diameter of cluster c. This index tries to find the best number of clusters with numerous iterations. In each iteration, a specific number of clusters will be used and a number of clusters with the highest amount of Dunn's index will be used in the k-means clustering.

3.2.2 Identifying dominant trends

In this study, we have realized that dominant trends are frequent fluctuation trends among many of customers. In order to recognize such trends, clusters need to have a significant number of customers to be regarded as dominant trends. In our study, the customers' behaviors are diversified and we are faced with a high number of small clusters. Therefore, we have merged some of small clusters into larger ones by means of Cosine similarity measure. We also determine a predetermined number for the final clustering number, which is resulted from the cluster visual exploration and consulting domain experts. As pointed out earlier, Cosine similarity can be used to find out resemblance of different located trends in a cluster.

3.3 Clusters analysis

3.3.1 Analyzing clusters

As we explained earlier, a comprehensive study on customer's fluctuations across RFM segments over time can reveal useful information on customer's behavior. The acquired dominant trend in the previous step is used in this section as a representative of accepted clusters to analyze. If there is an expansion trend on desired clusters it is a positive indication for marketing managers. They can further analyze the cause of changes between the sequences of periods. Expansion of undesirable clusters, on the other hand, can be used to take preventive measures. The study of changing the nature of cluster memberships of individuals is useful for target marketing.

3.3.2 Labeling and profiling

In the last step of our methodology, we associated a label for each detected cluster and provided a profile of the clusters' members. Labeling of each cluster will be determined based on the shape of fluctuations (increase, decrease, and being stable during the periods). Furthermore, specifying profile of clusters is another aspect of this step. In our study, profile is behavior-based, involving what the customer is actually doing such as whether they have multiple visits, multiple purchases and the amount of money they spend on their regular purchases. These are the questions, which would be focused by looking at the behavior.

4. Case study

The dataset we used in this research is associated with a food mart located in United States with two years of information. It includes customer information about customer spending, visits, sold product categories, and other transactional data. Customers were segmented in three-month intervals, eight times during 24 months, based on their transactional data from January 1997 to December 1998. When we split data into different intervals, we need to make sure that there are enough data for each interval. In addition, determining the time interval duration depends on the type of industries. The proposed study of this paper considers three months for the length of each time interval to create new segmentation (i.e., one quarter).

Data preparation has a significant effort before mining can be applied. Data preparation in this study consisted of two steps of data cleaning and data aggregation. We needed to separate the customers with two years continuous transactions in the data set. In our survey, there were 10281 customers. More specifically, in the data cleaning step, customers who did not have any purchase transaction in first year or customers who had no purchase transaction in three quarters out of 8 quarters were removed from the dataset. Consequently, a dataset including 3141 customers was obtained. In the second step, the purchased items were aggregated such that each record of new dataset represents a bill including all purchased items in a same day. Then, the F and M values as RFM measures were calculated for each time intervals. For assigning customers to the segments, we projected F and M dimensions into one new dimension using PCA method. The new dimension, representing customer's purchasing behavior aggregately, was utilized to assign customers into five segments for each time interval. In other words, each customer is assigned to one of pre-determined segments based on the resulted new dimension for each time interval. As noted previously, each segment was defined based on the value of R and M. The results indicate that the first segment contains customers not having any purchase in the corresponding time frame; the second segment includes customers with low monetary and frequency; the third segment contains customers with medium value of the new dimension and the forth segment shows customers with high amount of monetary and frequency. In the last segment, the golden and very high spending customers are located.
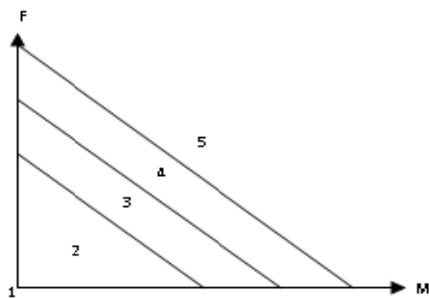

Fig. 1. Customer values for five intervals   **Fig. 1.** Two customer behavior trends
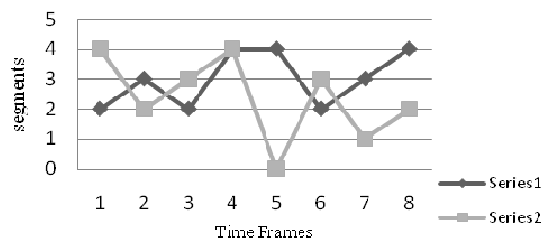
The process has been performed through all time frames. Consequently, eight segments are associated to each customer so that each segment indicates the relative position of the customer in the corresponding period. Therefore, a time based serial of segments is obtained for each customer showing relative position of customer in a sequence of eight segments (see Fig. 1 and Fig. 2). These sequences describe customers' behavioral fluctuations and provide a basis for customer behavior analysis in this study. Table 3 shows the distribution of customers in the segments along with the periods.

Table 3

Distribution of customers in the segments along the time frames

| Segment	Time frame							
	1	2	3	4	5	6	7	8
1	930	996	964	837	1002	1111	1029	833
2	1224	1271	1289	1349	1473	1426	1463	1567
3	556	517	531	577	500	429	505	548
4	246	212	208	225	128	126	106	133
5	185	145	149	153	38	49	38	60
Sum	3141	3141	3141	3141	3141	3141	3141	3141

The proposed methodology tries to find the patterns of fluctuations of customer purchasing behavior over the time. These patterns are the foundation of our targeted customer behavior analysis. In this research, the term of electrocardiogram is used for customer behavioral trends. Identifying dominant fluctuations patters of customer behavior can be considered as a good basis for tuning the company relationships with customer. It is evident that there are some drivers/factors or reasons behind these fluctuations. Studying the roots and reasons for these fluctuations are out of the scope of the current study. For instance, there may be business, customer or even macroeconomic related drivers. As an important point regarding the fluctuations drivers, note that such analysis reveals some hidden factors having significant effects the customer churn or loyalty. Furthermore, identifying such drivers can be beneficial to improve customer relationship and promote customers. For clustering customers' behaviors in this research, we used K-means with short time series distance function. In K-means clustering, the user has to specify the number of clusters. We used Dunn's index for defining the number of clusters in our study. For starting the clustering we selected the preliminary centers of clusters, randomly. Since the result of clustering depends on the preliminary center, we repeated the procedure ten times. We identified the best clustering number from all ten clusters by comparing their Dunn's index. Table 4 shows the amount of Dunn's index for the results related to the clustering with highest Dunn's index where the highest Dunn index was 0.5423 which was associated with clusters number eleven. Table 5 also shows the distribution of customers in eleven clusters.

Table 3

Quality of Dunn's Index for different number of clusters in K-Means clustering

Dunn's Index	Cluster number	Dunn's Index	Cluster number
0.346	2	0.4851	12
0.3974	3	0.3974	13
0.3244	4	0.3244	14
0.3203	5	0.2554	15
0.3735	6	0.433	16
0.4189	7	0.3974	17
0.4286	8	0.3262	18
0.4082	9	0.3536	19
0.4758	10	0.281	20
0.5423	11		

Table 4

Distribution of customers in eleven clusters

num. of cluster	1	2	3	4	5	6	7	8	9	10	11
Num. of Customers	240	249	175	136	250	284	251	238	265	197	224
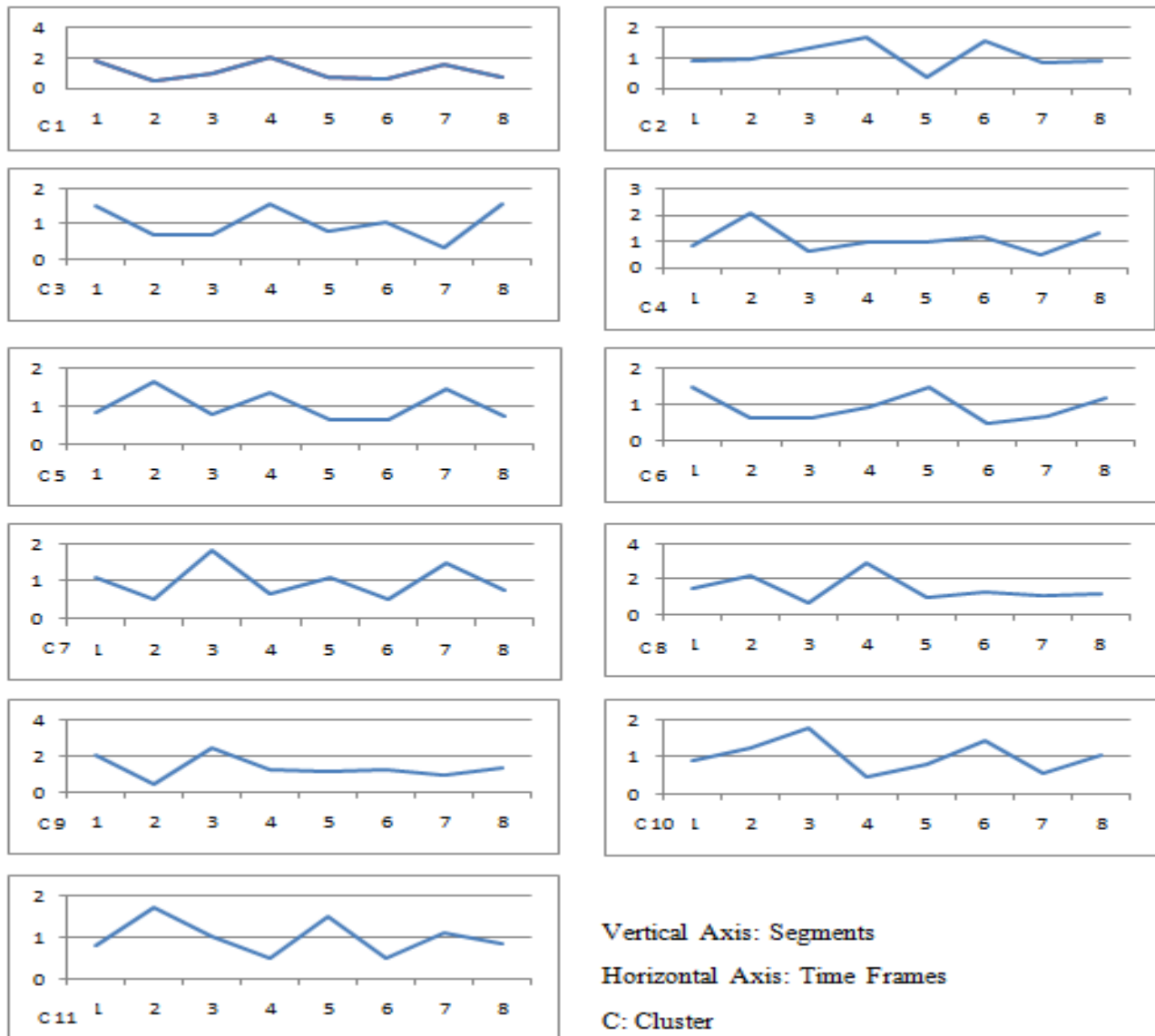
Fig. 3. Electrocardiogram of customers in eleven clusters

Fig 3 shows the electrocardiogram of customers in eleven clusters. There are three events when we perform an analysis on clusters and extract prominent trends' behaviors. These events include when customers' behaviors remain plateau, when customers' behaviors fall down, and when customers' behaviors rise across the time from monetry and frequency point of view. Companies could make decision and define suitable strategies based on customer's behaviour. In clusters where customers follow a constant behavior in a period of time, a suitable promotions could be used to improve their behaviors. In clusters with continous declining frequency and monetary, we may come to a conclusion that these customers are going to churn and switch to competitors which is considered to be a red alarm. Clusters with improve in their behaviors are good candidates to make them loyal customers. With concerning prominent behavioral trends, in the clusters 8 and 9 a constant behavior is obvious after time frams of 4 and 5. They are good candidates to consider for possible promotion strategies to improve and to grow their value. In the behavioral trends in clusters 10 and 11, it is obvious that customers follow a declining trend. The average results for frequencies and monetaries of customers in the first year instead of the second year show this event. There is a steady increase in customer value for customers located in clusters 5 and 6 in time frame 3, 4 and 5. This trend declines shortly later and will rose again during the last time frame. The company should not only prevent the declining trend for this group of customers but also we must guide them to become loyal customers.

Table 6
Mean of the customers new dimention in each time frams according to clusters

	Time Frame							
Cluster	1	2	3	4	5	6	7	8
1	1.741071	0.495536	0.964286	1.982143	0.758929	0.607143	1.495536	0.767857
2	0.919192	0.959596	1.30303	1.686869	0.343434	1.570707	0.858586	0.873737
3	0.826415	2.041509	0.645283	0.969811	0.950943	1.158491	0.490566	1.283019
4	1.104603	0.527197	1.820084	0.669456	1.125523	0.548117	1.481172	0.753138
5	1.54	0.708	0.7	1.572	0.82	1.06	0.344	1.556
6	1.471831	0.619718	0.633803	0.950704	1.485915	0.485915	0.665493	1.18662
7	0.868526	1.633466	0.816733	1.334661	0.673307	0.665339	1.458167	0.74502
8	1.510949	2.124088	0.664234	2.875912	0.934307	1.218978	1.051095	1.19708
9	2.08427	0.404494	2.449438	1.230337	1.078652	1.247191	0.955056	1.370787
10	0.884	1.22	1.768	0.476	0.808	1.412	0.592	1.044
11	0.779167	1.695833	1.025	0.5	1.470833	0.495833	1.104167	0.841667

Table 7
The mean of Monetary anf Frequency for eleven clusters in eight different time frames(TF)

Cluster		TF1	TF2	TF3	TF4	TF5	TF6	TF7	TF8
1	M	53.66	13.96	28.92	65.81	28.41	24.71	58.00	32.21
	F	1.86	1.86	1.08	2.24	0.99	0.81	2.00	1.06
2	M	26.05	29.44	41.55	55.55	16.43	63.24	36.34	35.78
	F	1.01	1.05	1.44	1.92	0.51	2.02	1.09	1.20
3	M	26.05	62.81	19.82	30.70	40.52	45.16	20.48	52.85
	F	0.92	2.14	0.70	1.12	1.29	1.47	0.65	1.68
4	M	36.18	14.44	58.71	20.67	45.35	22.53	59.56	33.09
	F	1.24	0.60	2.08	0.79	1.50	0.71	1.89	1.02
5	M	45.34	22.99	21.43	50.72	33.77	47.13	14.53	66.50
	F	1.68	0.76	0.80	1.82	1.07	1.39	0.48	2.05
6	M	44.65	17.15	19.61	27.74	58.89	17.88	26.21	49.04
	F	1.58	0.68	0.70	1.12	1.90	0.61	0.85	1.56
7	M	27.70	51.25	27.49	44.37	29.21	28.08	59.04	31.02
	F	0.98	1.75	0.94	1.51	0.90	0.89	1.90	0.98
8	M	45.08	71.16	19.12	105.79	38.12	51.34	45.58	51.16
	F	1.67	2.28	0.72	3.34	1.22	1.64	1.45	1.72
9	M	66.65	11.95	83.14	39.21	45.00	53.24	42.43	58.21
	F	2.29	0.49	2.78	1.37	1.44	1.71	1.30	1.91
10	M	24.89	36.62	57.10	14.19	34.67	59.16	23.79	41.17
	F	0.96	1.32	1.97	0.57	1.10	1.80	0.75	1.35
11	M	23.10	52.09	28.70	14.38	56.23	20.55	43.03	33.05
	F	0.88	1.82	1.10	0.59	1.88	0.68	1.40	1.08

There is a fluctuation for customers in clusters 4 and 7 during the time horizon and a special promotion for stabilizing these customers is necessary. There is also an intensive growth for customers in cluster 2 but it will derease at the end. Therefore, we need to perform more analysis to find out the reasones of such deacrese to prevent the same events in the future. Table 6 presents the mean of the new dimention of the customers in each time frams according to the clusters. In addition, in Table 7, the amounts of monetary and frequncy for all customers are shown seperatedly in different time frames.

5. Conclusion

This paper has presented a new methodology to measure customer's behavioral trends called customer electrocardiogram. The proposed model of this paper used clustering method with RFM

analysis to study customer's fluctuations over time. The method was also applied for a real-world case study of food industry, and the results have been discussed in details. The preliminary results of the implementation of our proposed method for a real-world case study indicates that the method could analyze customer's behavior based on pending potential and frequency of purchasing behavior. This could help detect dominant fluctuation trends of customers' behavior over time. It seems that the method can provide more findings about customer behavior fluctuations. Mining transactional data to detect or nominate some drivers to fluctuations are very demanding in customer relationship management and we leave it as further research. Future research can include dynamic clustering in segmenting customers for each period. It can predict near future fluctuations and give solutions to prevent customer churn.

References

Bottcher, M., Spott, M.D. Nauck, & Kruse, R. (2009). Mining changing customer segments in dynamic markets. *Expert Systems with Applications*, 36, 155–164.

Cheng, C., & Chen, Y. (2008). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 4176–4184.

Farvaresh, H., & Sepehri, M. M. (2010). A data mining framework for detecting subscription fraud in telecommunication. *Engineering Applications of Artificial Intelligence*, 24(1), 182-194.

Han, J., & Kambert, M. (2001). Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco.

Huang, S., Chang, E. & Wu, H. (2009). A case study of applying data mining techniques in an outfitter's customer value analysis. *Expert Systems with Applications*, 3957-4174.

Kaefer, F., Heilman C.M., & Ramenofsky, S. D. (2005). A neural network application to consumer classification to improve the timing of direct marketing activities. *Computer & Operations Research*, 32(10), 2595–2615.

Kim, Y. & Nick Street, W. (2004). An intelligent system for customer targeting: A data mining approach. *Decision Support Systems*, 37, 215–228.

Lingras, P., Hogo, M., Snorek M., & West, C. (2005). Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences*, 172 (1–2), 215–240.

Minghua H., (2008). Customer Segmentation Model Based on Retail Consumer Behavior Analysis.. IITAW '08. International Symposium on Intelligent Information Technology Application Workshops, 914-917.

Möller-Levet, C.S., Klawonn, F., Cho, K.-H., & Wolkenhauer, O. (2003). Fuzzy clustering of short time series and unevenly distributed sampling points. Proceedings of the 5th International Symposium on Intelligent Data Analysis, Berlin, Germany, August 28–30.

Ngai, E. W. T., Xiu L., & Chau, D. C. K. (2009). Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. *Expert Systems with Applications*, 36, 2592–2602.

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

Weiwen, X., Liang, C., Zhiyong, Z. & Zhuqiang, Q. (2008). RFM Value and Grey Relation Based Customer Segmentation Model in the Logistics Market Segmentation. *International Conference on Computer Science and Software Engineering*, 5, 1298 -1301.

Xiang-Bin, Y. & Yi-Jun, L. (2006). Customer Segmentation based on Neural Network with Clustering Technique, Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 265-268.

Yeh, C., Yang, K. & Ting, T. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, 36, 5866–5871.

Ying, L. & Feng, L. (2008). Customer segmentation analysis based on SOM clustering. IEEE International Conference on Service Operations and Logistics, and Informatics, 1, 15-19.

Yuantao, J. & Siqin, Y. (2008). Mining E-Commerce Data to Analyze the Target Customer Behavior. First International Workshop on Knowledge Discovery and Data Mining, 406-409.