New approach to customer segmentation based on changes in customer value

Article in Journal of Marketing Analytics · September 2015

DOI: 10.1057/jma.2015.10

2 authors:

Monireh Hosseini
Khaje Nasir Toosi University of Technology

Mostafa Shabani
Aarhus University

Original Article

New approach to customer segmentation based on changes in customer value

Received (in revised form): 20th July 2015

Monireh Hosseini
is Assistant Professor of IT in the Industrial Engineering Faculty of K.N. Toosi University of Technology. Her research interests focuses on customer relationship management, customer value analytics, customer lifetime value, network models of customer value, business-to-business relationships and networks, customer-centric information systems, e-strategy and Internet marketing. She is the highly commended winner of the 2011 Emerald / EFMD Outstanding Doctoral Research Awards for the thesis 'Customer Value Optimization Using Value Network Analysis Approach'.

Mostafa Shabani
is MSc Information System graduated from the Industrial Engineering Faculty of K.N.Toosi University of Technology. His research interests focuses on data mining and business analytics. His MSc thesis is 'The use of change mining in improving customer relationship management using RFM model'.

Correspondence: Monireh Hosseini, K. N. Toosi University of Technology, No. 24, Shahid Agahi Alley, Dabestan Street, Resalat Expressway, Seyed Khandan, Tehran, 1999143344, Iran
E-mail: hosseini@kntu.ac.ir

ABSTRACT In today's fast moving world of marketing from product-orientation to customer-orientation, the management of customer treatment can be seen as a key to achieve revenue growth and profitability. Knowledge of customer behavior can help marketing managers re-evaluate their strategies with the customers and plan to improve and expand their application of the most effective strategies. B2B or business customers are more complex, their buying process is more complicated and their sales value is greater. The business marketers usually prefer to cooperate with fewer but larger buyers than the final consumer marketer. As a business transaction requires more decision makings and more professional buying effort than the consumer market does, the efficient relationship with business customers is of paramount importance. Most customer segmentation approaches based on customer value fail to account for the factor of time and the trend of value changes in their analysis. In this article, we classify customers based on their value using the RFM model and *K*-means clustering method. Then, an assessment of changes over several periods of time is carried out. The originality of this research lies in its incorporation of time and trend of customer value changes in improving the accuracy of predictions based on the past behavior of customers. For this purpose, we used the POS customer transactions.
Journal of Marketing Analytics (2015) **3,** 110–121. doi:10.1057/jma.2015.10

Keywords: segmentation; Customer lifetime value; Temporal Data Mining; RFM model

INTRODUCTION

With the development of customer-oriented behavior in business, growing attention has been paid to customers and their needs as one of vital factors to gain higher profits (Cheng and Chen, 2009). Customer relationship management (CRM) seeks to identify customer needs and facilitate the interaction between customers and businesses (Ling and Yen, 2001). Given the changing customer behavior and needs in today's market, businesses must make decisions in keeping with these changes. Most researches on customer value and clustering-based segmentation do not consider changes over time and there is a paucity of researches on customer behavior based on association rules (Chen *et al*, 2005; Böttcher *et al*, 2009; Huang, 2012). The integration of CRM and analytic structures like data mining has been one of the important issues in recent years. This vision can help businesses develop new strategies or verify and correct their current strategies (Bose and Chen, 2009; Davenport *et al*, 2010). Nowadays, with the application of data mining technology in CRM, techniques like decision trees, clustering algorithms, genetic algorithms and association rules in different areas like commerce have been used to solve customer problems and formulate new strategies (Berson *et al*, 1999; Turban *et al*, 2008; Bramer,2007). In this research, attempts have been made to apply clustering techniques and mining changes to customer value over time to bring a new approach to the customer segmentation. The rest of this article is organized as follows: a review of literature is presented in the next section and a definition of customer value analysis is provided in the following section. The section after that offers a review of customer value analysis and discuss the shortcomings of previous studies. Our proposed model is presented in the penultimate section, and the conclusions are drawn in the final section.

RELATED WORKS

With rapid changes in market, making strategies based on customer behavior change with respect to the problems caused over time has been a challenge. This study seeks to address this issue. For this purpose, in this section a review of literature on customer segmentation based on customer value, RFM model, as the customer value calculation model, and analysis of customer value changing is presented.

Customer value analysis

Customer segmentation is the process of dividing customers into groups with similar characteristics or features. Customer behavior analysis and customer segmentation are mainly based on customer demographic variables (Song and Kim, 2001). One was to implement customer segmentation is to use customer value analysis (Marcus, 1998). Customer value analysis is an analytical structure for interpreting customer behavior from the vast source of otherwise meaningless data. There are different definitions for customer value. Kotler (2001) defines it as the profit of net present value. Hwang *et al* (2004) describes customer value as the sum of revenues gained by customers over lifetime of transitions. Thus, customer values are based on past and potential profit as well as defection probability. Later, Kim *et al* (2006) proposed a new customer value model, which was based on current value, potential value and loyalty. One of the most effective customer segmentation models based on customer value is the RFM model (Verhoef *et al*, 2003), which was introduced by Bauer (1988) and later developed by Hughes (1994). In recent studies, the RFM model has been adopted in different industries and under various conditions by adding extra parameters. Stone (1995) introduced RFM model with unequal value for each *R, F*, and *M* parameter. Tsai and Chiu (2004) attempted to weigh the parameters based on the characteristics of industries. Liu and Shih (2005) proposed WRFM, which is the defining weight for each parameter, using the analytic hierarchy process. Chang and Tsay (2004) proposed LRFM model that incorporates customer retention length (*L*) to the RFM model. Chang and Tsai (2011) tried to incorporate the characteristics of purchased products into their analysis by introducing GRFM (group RFM). Later Hosseini *et al* (2010) included another factor called 'period of product activity' in the RFM model (RFML) for vehicle manufacturing industries.

RFM model

RFM model is a well-known analytical methods of customer value widely used in customer segmentation (Verhoef *et al*, 2003). This is a Behavioral Model that analyzes existing data in the database and then based on the results, tries to predict feature customer behavior (Wei *et al*, 2010). This model consists of three parameters including recency of the last purchase (R), frequency of the purchases (F) and monetary value of the purchases (M). The first factor (R) refers to the interval between the last purchase behavior of the customer and the present time. The second factor (F) refers to the number of transactions in a specific period, and the last factor (M) refers to the amount of money spent in a particular period (Hughes, 1994). The usual procedure of the RFM model involves preparing data based on each factor and then dividing them into five subsets. For the first factor, the data is organized based on the purchase date in a decreasing order, with the top subset getting a maximum rate of 5 and the last subset getting a rate of 1. This procedure is applied to customers based on two other factors. Each customer receives three different scores, which are then added and the highest scores are placed at the top 20 per cent of the data, and the groups of customer with lower scores are placed at lower categories (Hughes, 1994).

MINING CHANGES OVERTIME

As discussed earlier, data mining is the statistical analysis of data based on the information of models, which is appropriate for the time period. Recent studies suggest that data changes over time and therefore the results would not be helpful (Roddick and Spiliopoulou, 2002). Today's world is in a constant state of flux that makes the previous results out-of-date. The new approach is to mine these changes over time periods. In this approach, time is split into intervals $T = \langle t_0, \ldots, t_n \rangle$ where for each t_i, there is a M_i model. These models are based on statistical methods of data mining. Here, the main purpose is to analyze the evolution of $\langle M_1, M_2, \ldots, M_n \rangle$ that creates the final model M (Böttcher *et al*, 2008).

In this context, most approaches are in association rules. Chen *et al* (2005) attempted to integrate customer behavior variables, demographic variables and transaction database to present a method of mining changes in customer behavior. Böttcher *et al* (2009) presented a system of customer segmentation based on the discovery of frequent itemsets and the analysis of their change over time. Hu *et al* (2013) used RFM analysis in the sequential pattern mining process. As can be seen, there is a paucity of studies on customer value analysis based on clustering techniques.

SEGMENTATION OF CHANGES IN CUSTOMER VALUE

As mentioned earlier, analyzing data for a whole period of time may be troublesome. For instance, consider two customers one of whom displaying a decreasing trend in F and M factor of RFM model whereas the other demonstrates an increasing trend in these two factors. However, after using the RFM model, both would be placed in the same segment because of having a similar total value. The resulting segmentation suggests that these two customers have the same behavior, which is an incorrect result. For further clarification see Figure 1 and Figure 2.

These two figures (Figure 1, Figure 2) display two customers who are placed in the same segment after implementing the RFM model. This problem is caused because of the simultaneous analysis of the whole time.
In Figure 1 and Figure 2, customers show conflicting trends in the F and M factors, thus implementing the same strategy for these two customers will not be helpful. One of them is churning the business, and the other one brings larger profits for the business. To solve this problem, the factor of time should be incorporated in our analysis.

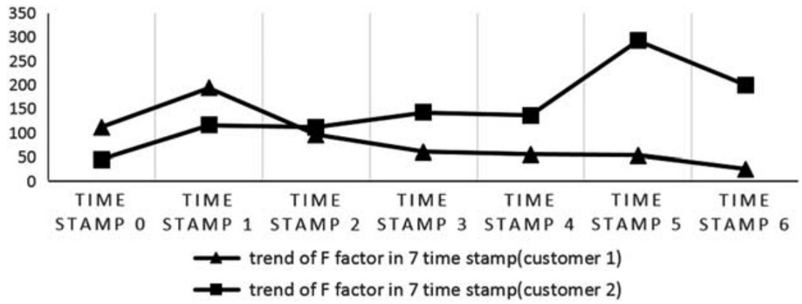

Figure 1: Trend of *F* factor in seven time stamps for two customers.

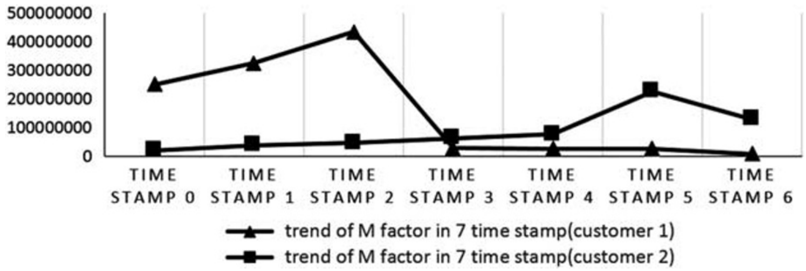

Figure 2: Trend of *M* factor in seven time stamps for two customers.

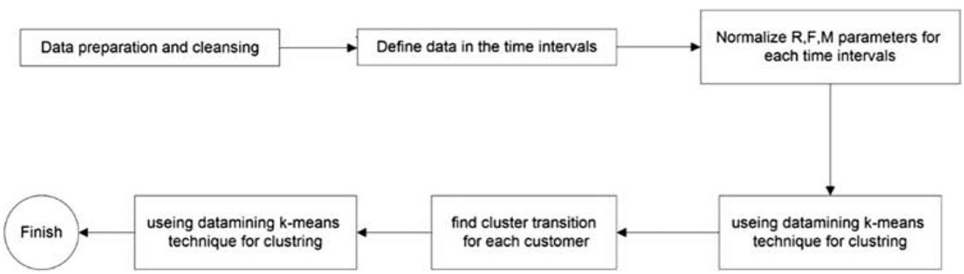

Figure 3: Proposed model flowchart.

PROPOSED MODEL

The basic RFM customer segmentation contains some problems that have been discussed in the section 'Mining changes over time'. In this section, we represent the model and describe its algorithm. To overcome the problems of the RFM model, first we need to split time into separate parts and then based on the RFM model calculate customer value for each time interval. In doing so, we obtain a value trend for each customer. Finally, we segment customers by using *K*-means clustering data mining technique based on the value trends of these customers. Figure 3 represents the steps of our proposed model.

Our proposed model consists of the flowing six steps:

1. data preparation and cleansing;
2. incorporating data into time intervals;
3. normalizing *R, F* and *M* parameters for each time interval;
4. using data mining *K*-means technique for clustering;
5. finding cluster transition for each customer and 6. using data mining *K*-means technique to cluster the data obtained in the section 'Proposed model'.

All procedures are described separately below:

Step 1 Data preparation and cleansing:

In Step 1, the data is cleaned by removing irrelevant values, noisy and incorrect data as well as the fields of data inappropriate to our research and we prepare data structure for implementing RFM analysis like choosing *R, F* and *M* parameters.

Step 2 Incorporating data into time intervals:

As mentioned earlier, we try to include time into the RFM model. For this purpose, we split the time period into z equal time intervals (equation (1)).

$$T = \sum_{i=1}^{i=z} T_i,\ T = \{T_I, T_2, \ldots, T_z\}, |T_1| = |T_2| = |T_Z| \qquad (1)$$

Step 3 Normalizing RFM model parameters for each time interval:

To control the effect of each parameter on other parameters, we normalize *R, F* and *M* parameters between (0, 1) (Bramer, 2007). Here, because of the independency of time intervals in analysis, each time interval must be normalized separately. If the whole time is normalized and then split into time intervals, there would be some values that their influence exceeds time intervals and affect the normalization of whole data. Given the importance of the independent of analysis of each time interval, this will be inaccurate.

Step 4 Using *K*-means data mining technique to classify customers in each time interval:

After preparing data in Step 3 using the data mining *K*-means technique (Han et al., 2011), we try to segment customers based on the RFM model.

Step 5 Discovering each customer's value changing trend in the whole time intervals:

After implementing RFM model for all time intervals, a changing value trend for each customer is obtained, as demonstrated in Figure 4. It shows the schematic view of the value changing trend for customer X_i by calculating the customer cluster value for each interval $\{T_I, T_2, \ldots, T_Z\}$. The result is a set of data that shows the value changing trend of the customer throughout the analysis period.

In Figure 4, $D_{T_j X_i}$ is the value of customer X_i in time interval T_j and $[D_{T_1 X_i}, D_{T_2 X_i}, D_{T_3 X_i}, \ldots, D_{T_Z X_i}]$ is the customer X_i value changing trend in time

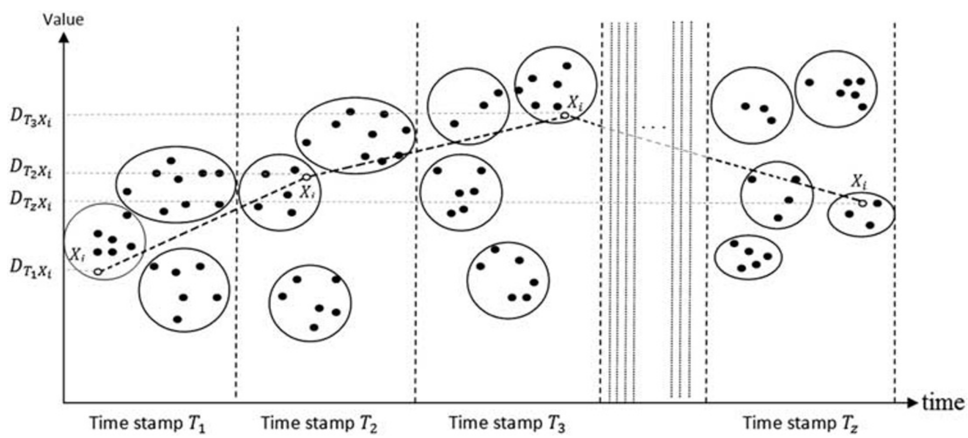

Figure 4: Customer values in *Z* time interval.

periods $\{T_I, T_2, \ldots, T_Z\}$. The collection of this type of data produces new data sets that will be used in the next step.

Step 6 Using data mining *K*-means technique for customer classification based on the data generated from Step 5:

On the basis of the data set obtained in Step 5 (Table 1), we will use *K*-means technique to cluster the changing value trends of customers. In doing so, the customers with the same value changing trend will be in the same segment.

EMPIRICAL CASE STUDY

The data consists of 500 000 POS purchase transactions[1] in 7 months of B2B customers. A sample of the data is illustrated in Table 2.

The implementation steps of our model are described below:

Step 1 Data preparation and cleansing:

First, the data is cleaned by deleting irrelevant values, noisy and incorrect data as well as the fields of data inappropriate to our research. Then, the data is organized in such a way to extract RFM model parameters. According to Table 2, terminal_ID is considered as customer_ID, transaction count is considered as parameter *F*, price is considered as parameter *M* and date of last purchase is considered as *R*. Table 3 shows the final data ready for the analysis.

Step 2 Incorporating data into time intervals:

The data are divided into 1-month periods based on their type. Here, the important part is *R* factor that determines the interval between the last day of the month and the last purchase in the same month.

Step 3 Normalizing RFM model parameters for each time period:

After dividing data into the seven 1-month periods, each time interval is normalized. Table 4 shows the data after normalization for the first time period.

Step 4 Using *K*-means data mining technique to classify customers in each time interval:

Table 1: Customer value in *Z* time interval

Customer ID	*Value in time stamp 1*	*Value in time stamp 2*	*Value in time stamp 3*	*Value in time stamp z–1*	*Value in time stamp z*
X_1	$D_{T_1}x_1$	$D_{T_2}x_1$	$D_{T_3}x_1$	$D_{T_{z-1}}x_1$	$D_{T_z}x_1$
X_2	$D_{T_1}x_2$	$D_{T_2}x_2$	$D_{T_3}x_2$	$D_{T_{z-1}}x_2$	$D_{T_z}x_2$
X_3	$D_{T_1}x_3$	$D_{T_2}x_3$	$D_{T_3}x_3$	$D_{T_{z-1}}x_3$	$D_{T_z}x_3$

Table 2: Database fields

Field name	*Field ID*
Terminal ID	1
Install date	2
Terminal status	3
Guild	4
Number of transactions	5
Amount	6
PSP bank	7
Date of Last Transaction	8

Table 3: Data ready for analysis

M	*F*	*R*	*Customer ID*
1895200	10	2	33 379 787
12494250	17	5	33 370 246
2762500	18	6	33 336 646
2927950	62	1	33 374 516
17250500	10	1	33 445 126
1436800	14	1	33 364 939

M: Unit: IRR, mean value for all periods (all customers): 344 194 332 726 IRR.
R: Unit: Days, interval between customer's last purchase and the last day of the time period.
F: Unit: number of transactions.

Table 4: Normalized data for first time period

M	*F*	*R*	*Customer ID*
0.000773	0.001319	0.066	33 379 787
0.000171	0.002242	0.1666	33 370 246
0.000181	0.003693	0.2	33 336 646
0.001067	0.008177	0.0333	33 374 516
0.000089	0.001319	0.0333	33 445 126
0.000083	0.001846	0.0333	33 364 939

We use *K*-means data mining technique to cluster customers in each time period. Table 5 shows the results of customer clustering for two consecutive time periods. The important point is that the cluster number is similar in two time intervals. Also, given the independent analysis of each time interval, there is no relationship between them.

Step 5

Discovering each customer's value changing trend in all time intervals:

Table 5: Results of customers clustering for two consecutive time periods

Cluster ID in second time interval	*Cluster ID in first time interval*	*Customer ID*
3	3	33 015 992
4	6	33 422 651
4	6	81 012 799
2	2	33 264 429
4	6	81 003 077
7	7	81 021 418

After splitting data into seven time intervals and implementing *K*-means technique, we will obtain seven values for each customer that represent customer value in each time interval. Table 6 shows some of customer value changing trends. This data are used in Step 6 of the analysis.

Step 6 Using data mining *K*-means technique for customer classification based on the data derived from Step 5:

Using the dataset derived from Step 5 and data mining *K*-means technique with 11 clusters, the customers with similar value changing trends will be in the same segment. Davies–Bouldin Index (DBI) was used for clusters number evaluation where smaller values indicate more efficient clustering. Small DBI values are achieved for a solution with low intra-segment

Table 6: Customers value changing trends

Customer value in time interval 7	*Customer value in time interval 6*	*Customer value in time interval 5*	*Customer value in time interval 4*	*Customer value in time interval 3*	*Customer value in time interval 2*	*Customer value in time interval 1*	*Customer ID*
0.1728	0.113953	0.8054	0	1	0.2151	1	33 372 058
0.1728	0.113953	0.8054	0	1	0.2151	1	33 076 017
0.1728	0.113953	0.8054	0	1	0.2151	1	33 015 992
0.295	0.1226	0.3030	0.070	0.06481	1	0.365	33 422 651
0.174	0	0.01510	0.0706	0.405	1	0.36	81 012 799
1	0.215	1	0	0.8054	0.1139	0.172	33 264 429
0.295	0.122	0.015102	0.0706	0.4053	1	0.365	81 003 077
0.0983	1	0.03742	0.1242	0.4053	0.0230	0.0115	81 021 418

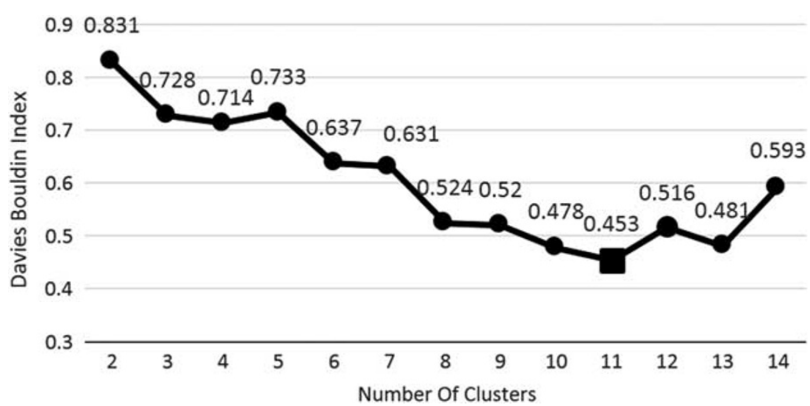

Figure 5: Optimum number of clusters.

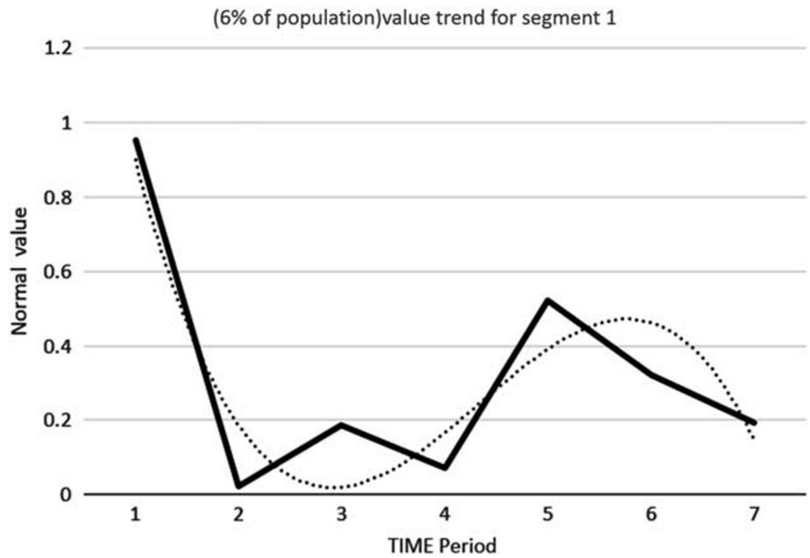

Figure 6: Value trend for Segment 1.

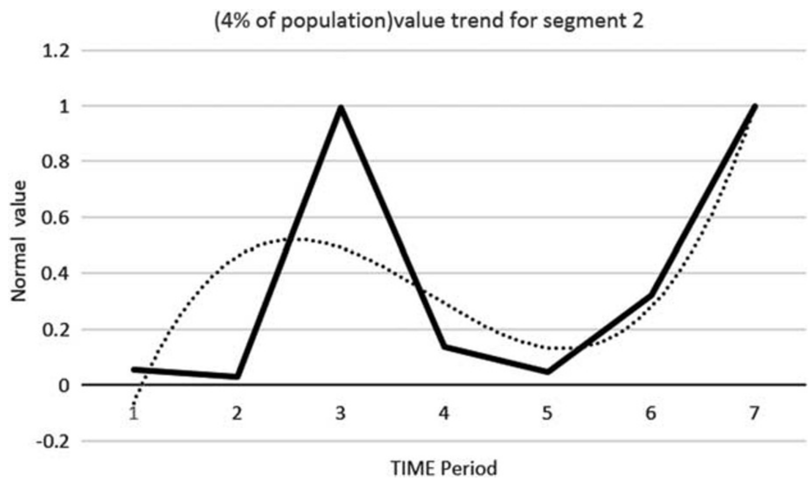

Figure 7: Value trend for Segment 2.

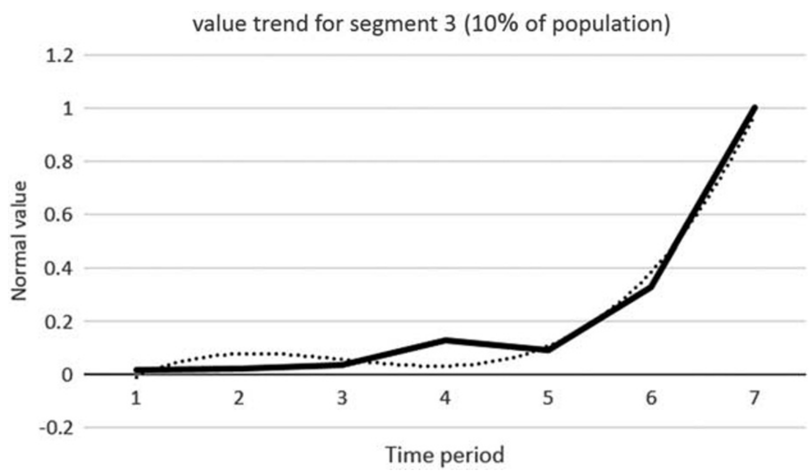

Figure 8: Value trend for Segment 3.

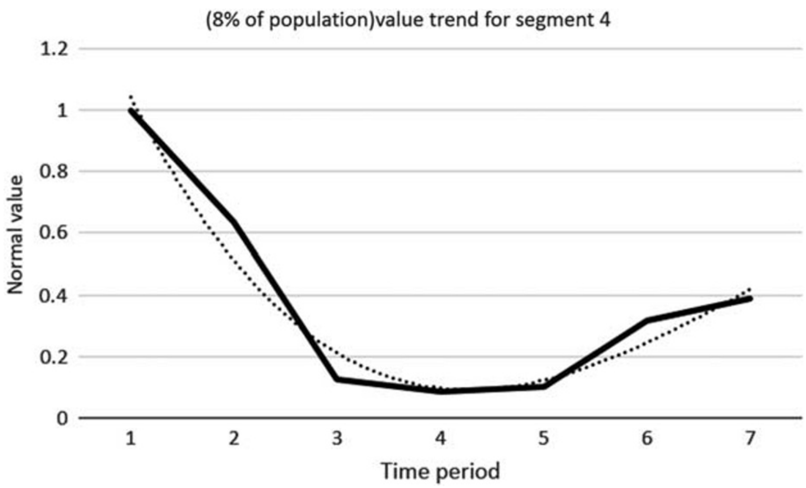

Figure 9: Value trend for Segment 4.

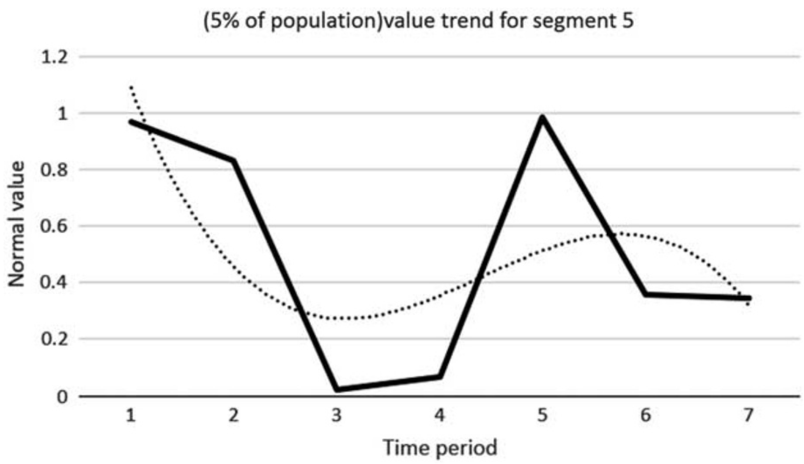

Figure 10: Value trend for Segment 5.

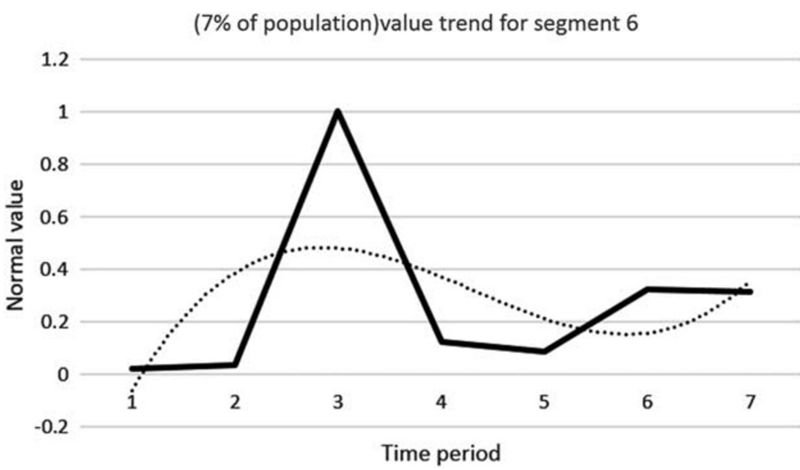

Figure 11: Value trend for Segment 6.

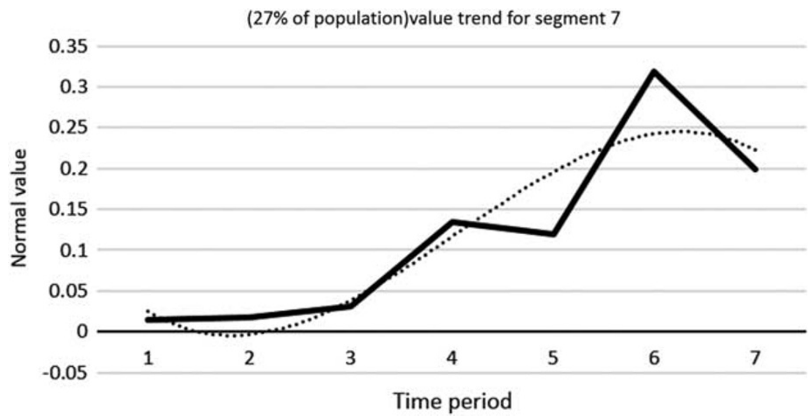

Figure 12: Value trend for Segment 7.

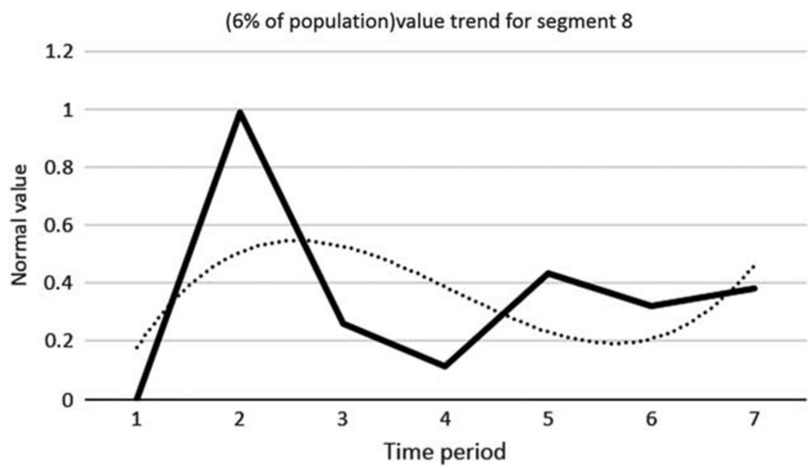

Figure 13: Value trend for Segment 8.

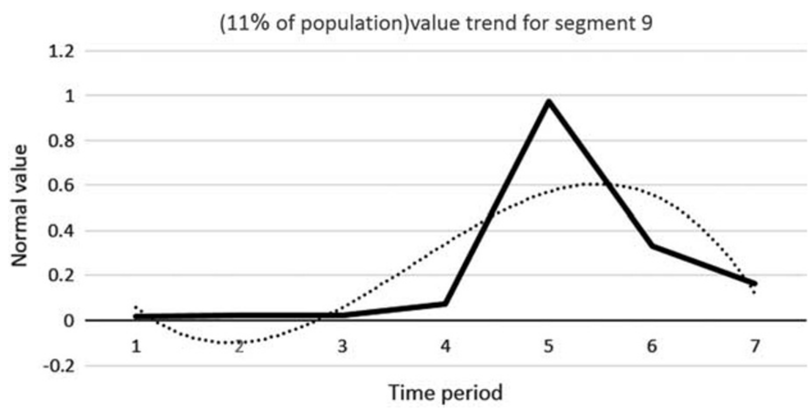

Figure 14: Value trend for Segment 9.

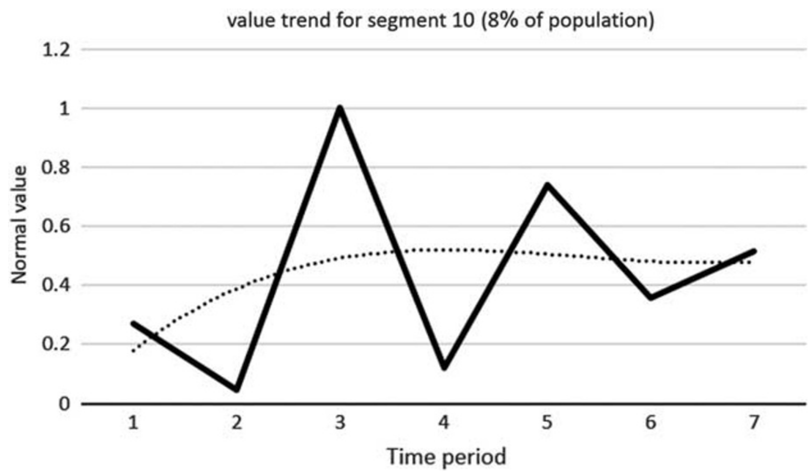

Figure 15: Value trend for Segment 10.

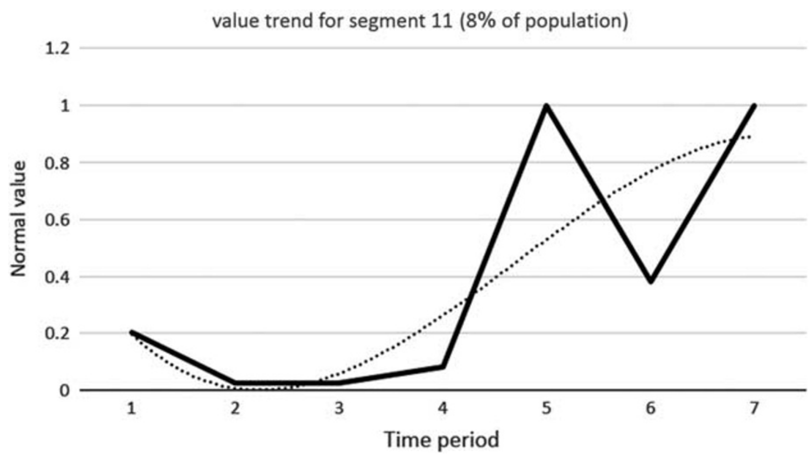

Figure 16: Value trend for Segment 11.

variance and high inter-variance segments. As shown in Figure 5, the result of optimum number for clusters is 11.

Figures 6–16 shows the trend of these 11 customer segments.

In Figures 6–16, the customer dispersion in different segments has been shown. For instance, customers in segment 3 (Figure 8), which contains 10 per cent of the population, have the best positive type of trend without any saltation. Segment 4 (Figure 8), which includes 8 per cent of the population, consists of customers with a negative trend that shows customer churn. With these segments, business administrators can make more efficient decisions based on target segment trend.

CONCLUSION

Mining changes over different time periods provide appropriate conditions to make efficient strategies based on former behavior of customers. The basic RFM model does not consider time transition and given the lack of details about the changes in customer behavior, ineffective results are obtained, as shown in the section 'Mining changes over time'. It is obvious that dealing with customers with decreasing or increasing value trends, and similar total values require different strategies. In the proposed model, we tried to use the positive aspect of the RFM model by incorporating time into our analysis with the aim of overcoming the shortcomings of the basic RFM model.

NOTE

1. Point of sale service.

REFERENCES

Bauer, C.L. (1988) A direct mail customer purchase model. *Journal of Direct Marketing* 2(3): 16–24.

Berson, A., Smith, S. and Thearling, K. (1999) *Building Data Mining Applications for CRM.* 1st edn. New York: McGraw-Hill Professional.

Bose, I. and Chen, X. (2009) Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research* 195(1): 1–16.

Böttcher, M., Höppner, F. and Spiliopoulou, M. (2008) On exploiting the power of time in data mining. *ACM SIGKDD Explorations Newsletter* 10(2): 3–11.

Böttcher, M., Spott, M., Nauck, D. and Kruse, R. (2009) Mining changing customer segments in dynamic markets. *Expert Systems with Applications* 36(1): 155–164.

Bramer, M. (2007) *Principles of Data Mining* (Vol. 131). Berlin: Springer.

Chang, H.-C. and Tsai, H.-P. (2011) Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications* 38(12): 14499–14513.

Chang, H.H. and Tsay, S.F. (2004) Integrating of SOM and K-Mean in data mining clustering: An empirical study of CRM and profitability evaluation. *Journal of Information Management* 11(4): 161–203.

Chen, M.-C., Chiu, A.-L. and Chang, H.-H. (2005) Mining changes in customer behavior in retail marketing. *Expert Systems with Applications* 28(4): 773–781.

Cheng, C.-H. and Chen, Y.-S. (2009) Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications* 36(3): 4176–4184.

Davenport, T.H., Harris, J.G. and Morison, R. (2010) *Analytics at Work: Smarter Decisions, Better Results.* Boston: Harvard Business Press.

Han, J., Kamber, M. and Pei, J. (2011) *Data mining: concepts and techniques: concepts and techniques.* Berlin: Elsevier.

Hosseini, S.M.S., Maleki, A. and Gholamian, M.R. (2010) Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications* 37(7): 5259–5264.

Hu, Y.-H., Huang, T. C.-K. and Kao, Y.-H. (2013) Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. *Journal of Systems and Software* 86(3): 779–788.

Huang, T. C.-K. (2012) Mining the change of customer behavior in fuzzy time-interval sequential patterns. *Applied Soft Computing* 12(3): 1068–1086.

Hughes, A. (1994) *Strategic Database Marketing*, Chicago, IL: Probus Publishing.

Hwang, H., Jung, T. and Suh, E. (2004) An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert Systems with Applications* 26(2): 181–188.

Kim, S.-Y., Jung, T.-S., Suh, E.-H. and Hwang, H.-S. (2006) Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications* 31(1): 101–107.

Kotler, P. (2001) *Marketing Management.* 10th edn. Canada: Pearson Education, Canada.

Ling, R. and Yen, D.C. (2001) Customer relationship management: An analysis framework and implementation strategies. *Journal of Computer Information Systems* 41(3): 82–97.

Liu, D.-R. and Shih, Y.-Y. (2005) Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management* 42(3): 387–400.

Marcus, C. (1998) A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing* 15(5): 494–504.

Roddick, J.F. and Spiliopoulou, M. (2002) A survey of temporal knowledge discovery paradigms and methods. *Knowledge and Data Engineering, IEEE Transactions on* 14(4): 750–767.

Song, H.S. and Kim, S.H. (2001) Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications* 21(3): 157–168.

Stone, B. (1995) *Successful Direct Marketing Methods,* NTC Business Books, pp. 37–57.

Tsai, C.-Y. and Chiu, C.-C. (2004) A purchase-based market segmentation methodology. *Expert Systems with Applications* 27(2): 265–276.

Turban, E., Sharda, R., Aronson, J.E. and King, D.N. (2008) *Business Intelligence: A Managerial Approach*, Upper Saddle River, NJ: Pearson Prentice Hall.

Verhoef, P.C., Spring, P.N., Hoekstra, J.C. and Leeflang, P.S.H. (2003) The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. *Decision Support Systems* 34(4): 471–481.

Wei, J.-T., Lin, S.-Y. and Wu, H.-H. (2010) A review of the application of RFM model. *African Journal of Business Management* 4(19): 4199–4206.