An approach to optimised customer segmentation and profiling using RFM, LTV, and demographic features

Morteza Namvar*

No. 23, Syasat 11 Alley, Modaress Street,
Kashan, 87196-87384, Iran
E-mail: mortezanamvar@gmail.com
*Corresponding author

Sahand Khakabimamaghani

3rd Floor, No.86, Shahid Beheshti Blvd., 1st Phase,
13888-17165, Western Tehransar,
Tehransar, Tehran, Iran
E-mail: sahand.cs82@gmail.com

Mohammad Reza Gholamian

School of Industrial Engineering,
Iran University of Science and Technology (IUST),
Narmak, 16844, Tehran, Iran
E-mail: gholamian@iust.ac.ir

Abstract: Customer segmentation and profiling are increasingly significant issues in today's competitive commercial area. Many studies have reviewed the application of data mining technology in customer segmentation, and achieved sound effectives. But in the most cases, it is performed using customer data from especial point of view, rather than from systematical method considering all stages of CRM. This paper constructs a new customer segmentation method based on RFM, LTV, and demographic parameters with the aid of data mining tools. In this method, first different combinations of RFM and demographic variables are used for clustering. Second, using LTV, the best clustering is chosen. Finally, to build customer profiles each segment is compared to other segments regarding different features. The method has been applied to a dataset from a food chain stores and resulted in some useful management measures and suggestions.

Keywords: customer relationship management; segmentation; data mining; clustering; profiling.

Reference to this paper should be made as follows: Namvar, M., Khakabimamaghani, S. and Gholamian, M.R. (2011) 'An approach to optimised customer segmentation and profiling using RFM, LTV, and demographic features', *Int. J. Electronic Customer Relationship Management*, Vol. 5, Nos. 3/4, pp.220–235.

Biographical notes: Morteza Namvar received his MSc in Electronic Commerce from Iran University of Science and Technology, Tehran in 2010 and BSc in Computer Engineering from Shahed University, Tehran in 2005. He is more interested in the areas of intellectual capital, knowledge management and CRM. His researches have been published in *Management Decision* and several international conference proceedings.

Sahand Khakabimamaghani received his MS in Electronic Commerce from Iran University of Science and Technology, Tehran, in 2010. He has instructed Artificial Intelligence, Databases, Algorithms and other courses in different universities. He has been in reviewing committee of the 4th IEEE International Conference on Computer Science and Information Technology (ICCSIT 2011) and *African Journal of Marketing Management*. His publications include one book and many journal and conference papers. He is mainly interested in artificial intelligence, data mining, pattern recognition, bioinformatics and business intelligence. He is now a PhD student of Bioinformatics in the University of Tehran.

Mohammad Reza Gholamian received his BS and PhD in Industrial Engineering from Amirkabir University of Technology, Tehran in 1996 and MS in Industrial Engineering from Isfahan University of Technology, Isfahan in 1998. He is currently an Assistant Professor in the Department of Industrial Engineering at the Iran University of Science and Technology at Tehran. His research interests are in intelligent systems and multicriteria decision making. He has published four books, 50 papers in the international conferences and 17 papers in the international journals.

1 Introduction

Customer segmentation and profiling, as the primary stage of CRM (Hruschka, 1986), is an increasingly pressing issue in today's over-competitive commercial area. To this aim "more and more literatures have researched the application of data mining technology in customer segmentation, and achieved sound effectives" (Stone et al., 2006). However, most of these researches segment customers only from a single point of view, rather than using a systematical segmentation framework (Namvar et al., 2010).

Demographic variables, RFM, and LTV are the most common input variables used in the literature for customer segmentation and clustering (e.g., Hung and Tsai, 2008; Jutla et al., 2001; Lee and Park, 2005; Mo et al., 2010). In spite of usage of demographic variables in all the stages of CRM, RFM and LTV are mostly used in customer retention and development (e.g., Chan, 2008). Hence, a combination of these three variables has not been considered enough in the literature to the aim of customer segmentation and profiling.

To achieve the best customer segmentation method with the most meaningful clusters, the above three variables are combined in a very novel approach in different stages of this work (i.e., customers segmentation, segmentation evaluation and segments profiling). First, based on demographic variables and RFM different sequences for customer segmentation have been examined. Second, LTV is employed as a new criterion for choosing the best segmentation method. Finally, clusters of the selected segmentation method has been profiled and visualised by an innovative rank-based approach which compares these three variables in different clusters.

The remainder of this paper is organised as follows. Section 2 reviews the previous studies related to customer segmentation and profiling. This part presents the limitations of existing studies and explains the background reasons of this study. Section 3 describes the proposed intelligent customer profiling approach. Consequently, the results of applying the approach in the case of a food chain store will be presented. Then a discussion on the result would be presented. Finally, Section 6 concludes the paper with some general discussions and an agenda for further research.

2 Literature review

Although customer segmentation and market segmentation have many similarities, there are some critical differences regarding input variables used in their clustering mechanisms (Namvar et al., 2010). Market segmentation usually aims at acquiring new customers, and deals with the first step of CRM (i.e., customer acquisition) using socio-demographic data, but, customer segmentation is used at all steps of CRM using both socio-demographic and transactional data. "We can imagine that customer cultivation and retention are more important than customer acquisition, because lack of information on new customers makes it difficult to select target customers and this will cause inefficient marketing efforts" (Hwang et al., 2004).

Chan (2008) has classified existing customer segmentation methods into methodology-oriented and application-oriented approaches. Most of methodology-driven studies modify some data clustering techniques, such as SOM, or use a combination of two or more data mining techniques to achieve more accurate clusters or segments (e.g., Boone, 2002; Lee et al., 2004; Jonker et al., 2004; Huang et al., 2007; Kim and Ahn, 2008; Wang, 2010; Mahdavi et al., 2010). "On the other hand, application-oriented researches must search for the optimum method for solving segmentation problems in specific applications" (Chan, 2008). They usually define and create new variable for clustering procedure or use different variables in sequential clustering steps (e.g., Hwang et al., 2004; Kim et al., 2006; Hsieh, 2004; Chang et al., 2007; Chan, 2008; Sheu et al., 2009; McCarty and Hastak, 2007; Stone et al., 2006; Lee and Park, 2005; Cheng and Chen, 2009; Kim et al., 2005).

In the literature of the latter approach, application-oriented customer segmentation, LTV has an important role. For example, Hwang et al. (2004) suggested an LTV model considering past profit contribution, potential benefit, and defection probability of a customer for wireless telecommunication customers segmentation. Kim et al. (2006) proposed a framework for analysing customer value and customer segmentation based on their value. Then, the framework is illustrated through a case study on a wireless telecommunication company and strategies are offered according to customer segments.

Another important input variable for application-oriented customer segmentation is RFM. For instance, Hsieh (2004) used a self-organising map (SOM) neural network to identify groups of customers based on repayment behaviour and recency, frequency, and monetary behavioural scoring predictors. He also classified bank customers into three major profitable groups of customers. The resulting groups of customers were then profiled by customer's attributes determined by using an Apriori association rule inducer. Lately, McCarty and Hastak (2007) investigated RFM, CHAID, and logistic regression as analytical methods for direct marketing segmentation, using two different datasets. Finally, Cheng and Chen (2009) proposed a new procedure, joining quantitative value of RFM attributes and K-means algorithm into rough set theory, to extract meaning rules.

Besides, a combination of above mentioned input variables has been utilised by researchers too. For example, Chan (2008) presented a novel approach that combines customer targeting and customer segmentation for campaign strategies. This investigation identified customer behaviour using a RFM model and then used an LTV model to evaluate proposed segmented customers.

Some authors have used a combination of other different variables and measures to cluster customers. For instance, Lee and Park (2005) aimed at providing an easy, efficient, and more practical alternative approach based on the customer satisfaction survey for the profitable customers segmentation. The authors presented a multi-agent-based system, called the survey-based profitable customers segmentation system that executes the customer satisfaction survey and conducts the mining of the customer satisfaction survey, socio-demographic and accounting database through the integrated uses of business intelligence tools such as data envelopment analysis (DEA), SOM neural network, and C4.5 for the profitable customers segmentation. Chang et al. (2007) proposed an anticipation model for potential customers in purchasing behaviour. Their model is inferred from past purchasing behaviour of loyal customers and the web server log files of loyal and potential customers by means of clustering analysis and association rules analysis. In the same year, Stone et al. focused on proposing a customer segmentation framework based on data mining and constructed a new customer segmentation method based on survival character. Their new customer segmentation method consists of two steps: first, with K-means clustering arithmetic, customers are clustered into different segments by similar survival characters (i.e., churn trend). Second, each cluster's survival/hazard function is predicted by survival analysing, then, the validity of clustering is tested and customer churn trend is identified.

Sheu et al. (2009) investigated integrating data mining and experiential marketing to segment online game customers. The results can help the firms to predict and understand the new consumer's purchase behaviour. Kim et al. (2005) determined and characterised groups of retail customers, based on their perception of commitment to the retailer and the degree to which they used its technological equipment.

Table 1 Summarisation of input variables used in segmentation models

Input variables used	*Authors*
Demographic	Hung and Tsai (2008), Jutla et al. (2001), Lee and Park (2005) and Mo et al. (2010)
RFM	Cheng and Chen (2009)
LTV	Kim et al. (2006)
Demographic + RFM	Hsieh (2004), McCarty and Hastak (2007) and Wu and Chou (2010)
Demographic + LTV	Hwang et al. (2004)
LTV + RFM	Chan (2008)
Demographic + RFM + LTV	-
Other	Chang et al. (2007), Lee et al. (2004), Sheu et al. (2009), Stone et al. (2006), Huang et al. (2007), Kim and Ahn (2008) and Gil-Saura and Ruiz-Molina (2009)

In addition, as mentioned before, some authors have focused on the methodology-oriented segmentation procedure. For example, Lee et al. (2004) developed a new methodology for cross-national market segmentation. The authors proposed a two-phase approach integrating statistical and data mining methods. The first phase is conducted by a statistical method (multi-group confirmatory factor analysis) to test the difference between national clustering factors. The second phase is conducted by a data mining method (a two level SOM) to develop the actual clusters within each nation. Huang et al. (2007) used support vector clustering for marketing segmentation. Kim and Ahn (2008), also, proposed a novel clustering algorithm based on genetic algorithms to effectively segment the online shopping market. At the same time, Hung and Tsai proposed a market segmentation approach, namely the hierarchical self-organising segmentation model, for market segmentation of real world multimedia on demand in Taiwan. Table 1 categorises segmentation models proposed by different authors according to their input variables.

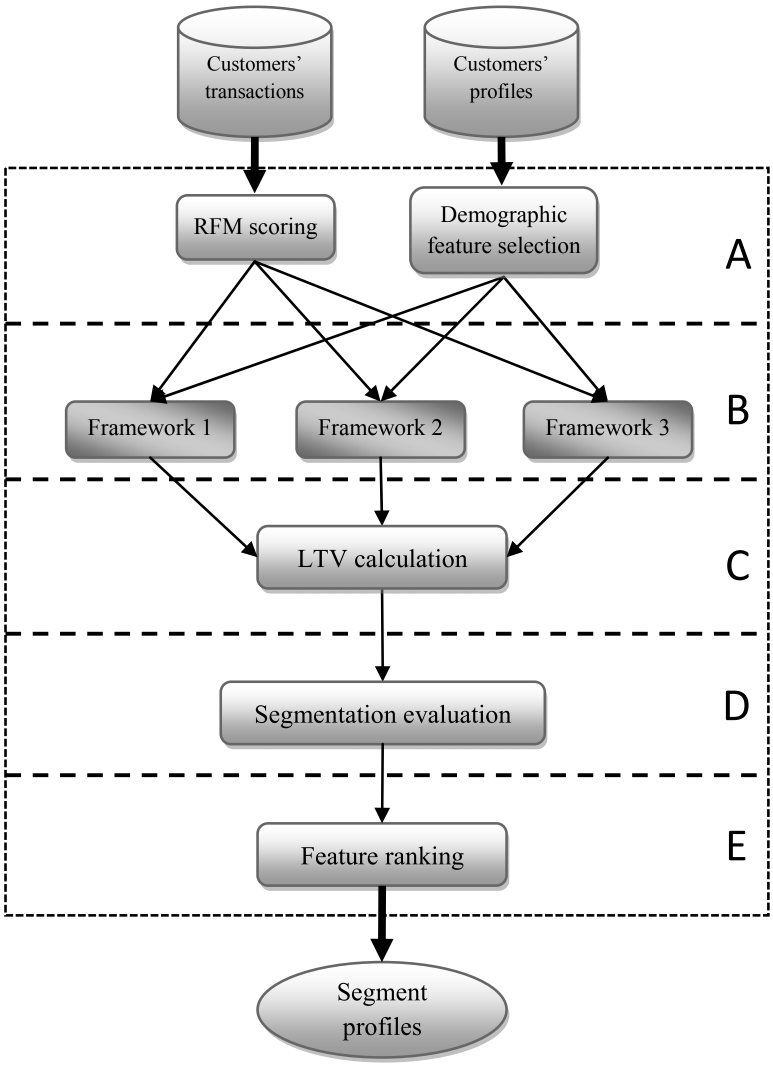
3 Proposed approach

The design structure of this research is illustrated in the Figure 1. It is consisted of three chief phases: data preparation, data modelling, and model evaluation. These three consequent phases are oriented from the core of every data mining methodology, such as CRISP-DM (Shearer, 2000). But, the proposed methodology may have broken some of these phases into two or three steps in order to make them easier to conduct and understand.

In a brief description, the initial step A prepares data for segmentation phase. As in every other data mining processes, here the first step is to prepare the data at hand. Data preparation involves tasks such as filling missing data, removing outliers, feature extraction, and feature selection. In the proposed approach, it is assumed that all the required tasks except feature selection and extraction are conducted for the data in 'Customers' Transactions' and 'Customers' Profiles' databases. So, these two more significant and innovative points are used here. In order to find the significant demographic features for demographic segmentation it is proposed that first customers are clustered using all of their demographic features. This action reveals the unimportant demographic features which have not significant effects on the results of segmentation, so these features might be removed from the set of current features used in clustering. Another significant point was the coefficients of recency, frequency, and monetary variables in calculating the RFM scores. As a result of this step, determinant demographic features of customers and their RFM values are extracted and prepared for next step.

After data preparation, in the segmentation step, step B, three different frameworks are proposed by using customer's demographic and RFM variables from previous step to generate different customer clusters. The aim of this step is to model data and find the best sequence of actions for segmentation among all possible sequences. We propose three frameworks for customer segmentation. The first framework consists of two chief phases: first, RFM-based segmentation and, next, demographic segmentation. The second framework is similar to the first one, but with a reverse approach. Unlike the first two frameworks, the third one consists of a single phase, in which all customers are segmented based on a combination of their significant demographic features and RFM scores. Accordingly, in step D, the best applicable segmentation for the data at hand could be identified based upon an evaluation criterion described in step C and the resulted models in this step.

Figure 1 Research methodology



In step C, LTV is calculated for all customers. It is proposed to use the LTV calculation method suggested by Hwang et al. (2004), since it is a relatively comprehensive method that considers potential benefit and customer loyalty in addition to past value of customers. The calculated LTV in this step will be utilised in the next step, step D, to compare values of the resulted clusters in previous step, step B.

In step D, as the most important feature for customer valuation, LTVs calculated in the previous step are utilised for determining the best segmentation framework from step B. In other words, this step determines the segmentation framework which distributes customers to the most suitable segments regarding their values. For choosing the best segmentation framework amongst the above methods, the quality of each segmentation method should be measured. In most previous studies, the quality of segmentation is measured based on within-segment and inter-segment heterogeneity (Wedel and Kamakura, 2000). Recently, marketers are concerned with and interested in maximising the net value of targeted customers, rather than caring about within-segment homogeneity or targeting rate (Hwang et al., 2004; Chan, 2008). Here by using LTV scores of segments, which are resulted from step C, first the quality of segmentation methods are assessed, and then measuring the value of each segment is possible. This enables us to choose the best segmentation and base the LTV analysis upon it for finding the valuable customers, who will probably make more net value for business.

The last step, step E, finalises the methodology by analysing and comparing different segments generated in step D through a novel rank-based method. After the segments are identified, they should be analysed and compared regarding the values of different features. In the proposed approach, instead of the exact values of numerical variable in interpreting the segmentation results, their rankings amongst the other segments are indicated and used for comparing segments.

4 Results

This section introduces the data set used for experiencing the proposed frameworks and discusses the results of these experiments. The dataset used for experiencing the proposed frameworks belongs to a brick and mortar food chain store. The chain store marketing managers are aimed at preparing a map of their customers, which could help them in designing marketing strategies, especially customer retention and development strategies. Thus, they want to know the true segmentation of their existing customers.

In this store, each customer has a membership card and a socio-demographic profile stored in the store's database. The data stored about customers are described in the corresponding section. Furthermore, each time a consumer refers to one of the chain stores and purchases an item its transactions are stored in the store's databases. A transaction involves the time of purchase, purchased item, the quantity purchased, and the monetary value of transaction.

The data used in current research are related to 23 months of customer interactions and also include customer demographic information. The number of involved customers is 7,954. Furthermore, the time span of 23 months is divided into two parts consisted of first 17 months and last six months. The first part provides the independent features and the other part supports the target feature for all the prediction models generated throughout the frameworks' steps. Moreover, the RFM scores and current customer LTVs are calculated as detailed before based on the data related to the first part of the time span.

4.1 Demographic feature selection and RFM scoring

4.1.1 Demographic feature selection

In this step and other consequent steps, it is proposed to use Two Step [i.e., an improved version of BIRCH clustering algorithm which is proposed by SPSS (Zhang et al., 1996)] as the clustering algorithm, because this clustering method does not require knowing the number of clusters prior to performing the clustering. In order to find the significant demographic features for demographic segmentation it is proposed that first customers are clustered using all of their demographic features. This action reveals the unimportant demographic features which have not significant effects on the results of segmentation, so these features might be removed from the set of current features used in clustering. Then, the remaining set of demographic features is used for a second round of clustering with Two Step algorithm. Again some of the input features might be identified as insignificant. Accordingly, those features are removed and the process continues with the remaining features. This iteration is repeated until no feature is identified insignificant in the last Two Step clustering round. Finally, the remaining set of features is selected for demographic segmentation in any phase of the proposed frameworks. This suggested process is similar to Recursive Feature Elimination for classifiers (e.g., Lessman and Voβ, 2009), but it is applied to clustering instead of classification.

At result, six features were identified as the most influencing ones in customer demographic segmentation. Those were ‘annual income’, ‘number of cars owned’, ‘house-ownership situation’, ‘membership card’, and ‘occupation’. Features such as ‘annual income’ and ‘occupation’ are widely used in other similar researches too (e.g., Kim and Ahn, 2008; Cheng and Chan, 2009). But, others are specific to this case and this dataset (e.g., ‘membership card’). In the feature selection process, other popularly used features like ‘marital status’, ‘gender’, and ‘city’ where identified not as significant as the selected ones in this study.

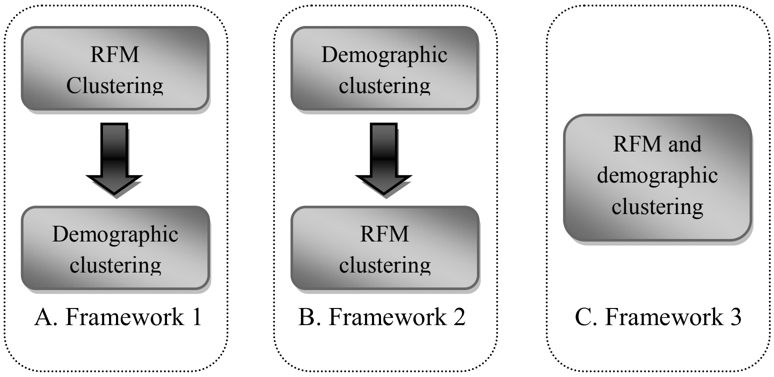
4.1.2 RFM scoring

Another significant point was the coefficients of recency, frequency, and monetary variables in calculating the RFM scores. Firstly, applying the binning technique the recency, frequency, and monetary values of all customers were divided to 5 bins. Consequently, the amounts of recency, frequency, and monetary scores ranged from 1 to 5. Secondly, for calculating the RFM score of an individual customer, its weighted recency, frequency, and monetary scores were summed in order to find its RFM score. Here, the coefficients used in calculating RFM scores are set equal to 1 to let all the three factors have the same weights and be equally significant.

4.2 Segmentation frameworks

After data preparation, the next step is to model the data. Unlike the previous similar works in retailing industry, such as Kim and Ahn (2008) and Boone and Roehm (2002), which have improved segmentation by modifying applied clustering algorithms, the focus of this research is on breaking segmentation process into several steps and examining the sequence of using different input features in the steps of segmentation process. The three proposed frameworks differ regarding this aspect. Besides, as described before and discussed later, the segmentation evaluation criterion is different in this research compared with the previous ones. In the following the process of generating these three frameworks for customer segmentation is explained.

Figure 2 Proposed frameworks



1 The first framework consists of two chief phases: RFM-based segmentation and demographic segmentation. In the first phase after calculating RFM scores of all customers, Two Step clustering algorithm is used for segmenting them based on their scores. Then for each RFM segment resulted in the first phase, a demographic segmentation is done using the selected features. Once more, Two Step clustering algorithm is exploited here. Finally, a set of behaviourally (i.e., RFM) and demographically distinct segments would be generated which are ready for LTV analysis in the next steps described later.

2 The second framework is similar to the first one, but with a reverse approach. Unlike the first framework, in the second one, first customers are clustered regarding their significant selected demographic features and then each demographic segment is further clustered behaviourally based upon RFM scores of individual customers. Consequently a set of distinct segments would be at hand for LTV analysis in the later steps.

3 Unlike the first two frameworks, the third one consists of a single phase, in which all customers are segmented based on a combination of their significant demographic features and RFM scores with Two Step clustering method. The above three frameworks are illustrated in Figure 2.

These proposed frameworks differ in putting more significance to various groups of features by letting them be involved in the segmentation earlier than other features. Using these three different frameworks makes it possible to examine all applicable sequences of inputting features. Accordingly, in the next steps, the best applicable segmentation for the data at hand could be identified based upon an evaluation criterion described later.

With RFM scores and demographic features at hand from previous step, it was possible to conduct all the three proposed segmentation frameworks. The first framework resulted in 20 segments while the second one could distinguish 11 different clusters. The third and the simplest approach detected only three segments. In the next Section (4.3), the LTV metric for evaluating these three frameworks is explained. Then, in the Section 4.4 evaluating these frameworks by the calculated LTVs from Section 4.3 will be discussed.

4.3 Customer lifetime value measurement

After segmenting customers, Lifetime values of customers were calculated through the formula proposed by Hwang et al. (2004). They proposed a new LTV model considering churn rate of a customer. According to authors, in order to calculate customer's LTV, current value and potential value of customers should be calculated separately. Furthermore, potential value of customers is predicted using equation (1).

$$Potential\ value_i = \sum_{j=1}^{n} Prob_{ij} \times Profit_{ij} \tag{1}$$

where $Prob_{ij}$ is the probability that customer i would use the service j among n-optional services and $Profit_{ij}$ means the profit that a company can get from the customer i who uses the optional service j.

It is notable that for constructing prediction models and calculating predictions (i.e., churn probability and potential value), considered as requirements for LTV calculation, artificial neural networks are utilised. As stated earlier, the dependent variables are provided by the first 17 months and the target features are calculated regarding the last 6 months of data at hand. Tables 2 and 3 illustrate the input features and their importance in prediction models of $Prob_{ij}$ and churn probability respectively.

Finally, at the end of this section, by constructing two prediction model (i.e., potential value prediction and churn prediction), the LTV of each customer was calculated by the proposed formula by Hwang et al. (2004). The resulted LTVs for each of customers will be used in the next Section (4.4) to evaluate and three constructed frameworks in the previous Section (4.2).

Table 2 Importance of features used for predicting $Prob_{ij}$

Feature	*Importance*
City	0.25
State/province	0.14
Frequency	0.08
Occupation	0.08
Product ID	0.07
Member card	0.06
Education	0.05
Num of children at home	0.05
Gender	0.03
Marital status	0.03
House-owner	0.03
Total children	0.03
Tenure	0.03
Age	0.02
Num of cars owned	0.02
Annual-income	0.02
Monetary	0.01

Table 3 Importance of features used in churn prediction model

Feature	*Importance*
State/province	0.22
Frequency	0.19
City	0.19
Tenure	0.07
Education	0.06
Occupation	0.05
Member card	0.05
Monetary	0.03
Number of cars owned	0.03
Gender	0.02
Age	0.02
Total children	0.02
House-ownership status	0.02
Annual-income	0.01
Number of children at home	0.01
Marital status	0.01

4.4 Segmentation evaluation

This section provides the results of implementing all the proposed segmentation frameworks. After finding the important demographic features (see Section 4.1.1), computing the RFM scores (see Section 4.1.2), implementing all the proposed frameworks (see Section 4.2), and finally calculating the customers' LTV (see Section 4.3), it was possible to evaluate the three generated segmentations regarding a novel LTV criterion.

For measuring the value of each segment simply the average of LTVs of all customers belonging to that segment is calculated. But, for estimating the quality of segmentation it is supposed to use the equation (2) in which C is the number of segments, $AveLTV_i$ is the average of LTVs of all customers in i^{th} segment, $ALTV$ is the average of all $AveLTV_i$, and $SDevLTV_i$ is the standard deviation of LTVs of all customers in i^{th} segment.

$$Quality = \sqrt{\frac{\sum_{i=1}^{C}\left(AveLTV_i - ALTV\right)^2}{C}} \Bigg/ \frac{\sum_{i=1}^{C} SDevLTV_i}{C} \tag{2}$$

The dividend of this fraction indicates the inter-segment heterogeneity and the subordinator measures the average of within-segment heterogeneity. So when the inter-segment heterogeneity grows and the within-segment heterogeneity shrinks, the segmentation quality increases. While this simple method calculates segmentation quality easily and differently, but the rationale on which it is based is the same as other conventional clustering separation indexes, such as Davies and Bouldin (1979).

Using this novel criterion, equation (2), the qualities of all three suggested frameworks were calculated. Regarding the resulted values indicated in Table 4, the first framework was detected as the best approach for customer segmentation for the case under study with a remarkable distance with the remaining frameworks.

Table 4 Comparison of three proposed clustering methods

Method	*Framework 1*	*Framework 2*	*Framework 3*
Quality (proposed criterion)	*1.33*	0.99	0.007

4.5 Feature-based segment ranking

In Table 5, it could be seen that the first framework has resulted in 20 segments. Using the proposed ranking method, differences between segments could easily be distinguished. For categorical features the category with the highest frequency (i.e., the mode) represents the corresponding segment and the average is used for the numerical features. Then, the rank for this type of features indicates their levels. The smaller (i.e., higher) rank shows a more advanced level. For example, partial higher education is ranked lower than higher education, because its level is inferior. Furthermore, if the value of a numerical variable is the same for two different segments, their ranks would be equal.

Table 5 Segments resulted from implementing the selected framework (the first one)

Segment no.	*LTV*	*Annual income*	*No of cars owned*	*House owner %*	*Education*	*Member card*	*Occupation*	*RFM*	*Size*
1	1	10	8	20	2	1	2	3	18
2	2	15	14	19	3	2	2	2	20
3	3	3	3	1	1	1	1	1	17
4	4	4	8	6	1	1	1	4	5
5	5	11	8	14	2	1	2	6	8
6	6	14	16	13	3	2	2	5	15
7	7	6	3	5	1	1	1	7	2
8	8	9	7	16	2	1	2	7	3
9	9	14	12	11	3	2	2	8	13
10	10	13	10	10	3	1	3	9	7
11	11	7	2	8	1	1	1	9	9
12	12	5	1	4	1	1	1	10	1
13	13	14	11	12	3	2	3	11	12
14	14	10	9	15	2	1	3	11	4
15	15	12	5	9	2	1	2	14	10
16	16	1	4	2	1	1	1	12	6
17	17	14	13	18	3	2	3	13	16
18	18	14	15	7	3	2	3	16	19
19	19	8	7	17	2	1	2	17	14
20	20	2	6	3	1	1	1	15	11

For example, segment number 1 has the highest value (i.e., LTV), a relatively low annual income, a moderate number of cars, a moderate level of education and occupation, and a small size. Furthermore, most of its members have a high level membership card and it includes less house owners in comparison to other segments. These show the features of the most valuable segment and the targeting strategies should consider these specifications.

5 Discussion

According to these results, for the data at hand, it is better to, first, segment customers based on their RFM scores and then further cluster each RFM segment using their demographic features. This means that splitting segmentation process into two steps with different input features results in a better distinguishing of customer groups in some cases.

While the proposed process seems close to Chan's (2008) suggestions amongst others at the first sight, it is fundamentally different. Chan segments customers into eight clusters based on only their RFM using GA, but our method uses customer demographics beside RFM. Moreover, Chan evaluates its segmentation regarding the total LTV of all segments generated in each iteration of GA, while the proposed methodology examines inter- and intra-segment dispersion of customers' LTVs to find the best segmentation. In other words, Chan matches customers with a set of predefined marketing strategies, but the current research reveals different existing segments and their features for the marketers to support them in designing strategies.

This study proposed a new approach for customer segmentation based on three critical variables; demographic, RFM and LTV. According to our approach, for each set of data the different combination of RFM and demographic variables should be examined, and the best combination should be selected based on LTV values of segments. For this purpose, also a novel LTV-based criterion was proposed, and the three frameworks were compared regarding this criterion.

For the aim of illustrating the performance of the proposed approach another step has been taken. To compare the selected framework with the conventional methods, k-means model is chosen as the reference model. In other words, the selected framework is compared with k-means clustering algorithm with similar number of clusters. K-means is selected because "it has been used as the comparative standard in other, similar studies and is currently the most widely used and most popular segmentation technique" (Boone and Roehm, 2002).

The results of the selected framework are compared with those of the traditional k-means algorithm when $k = 20$ using both the famous Davies-Bouldin index (Davies and Bouldin, 1979) and the proposed quality measure. The remarkable distance between framework 1 and k-means with $k = 20$ could be seen in Table 6. Accordingly, the above implications are approved.

Table 6 Comparison of the selected clustering method with traditional K-means method

Method	*Framework 1*	*k-means (k = 20)*
Davies-Bouldin	*98.66*	568.92
Quality	*1.33*	0.06

6 Conclusions

A comprehensive customer segmentation method would be used in customer acquisition as well as customer retention and development. Churn avoidance, complaint management, one to one marketing, etc., would be so more accurate with thorough customer segmentation. Besides, an effective customer profiling will complement the customer segmentation in order to design marketing strategies. Undoubtedly, to achieve comprehensive customer segmentation and customer profiling, input variables and their sequence, segment's evaluation and their visualisation are so important.

LTV, RFM and demographic variables are the most popular marketing variables amongst both academics and practitioners. While these variables were used by many authors, there is a lack of clustering methods considering all of these variables (see Table 1). In this study, a comprehensive method for customer segmentation and profiling was proposed based on these variables. The novelties of this study could be viewed from three perspectives.

First, it reveals that different sequences of involving RFM and demographic variables in clustering should be examined to achieve the most suitable one for the segmentation process. However, it may vary regarding the characteristics of different studies. It could be inferred about this data that in food chain store markets, RFM features should be used before demographic features in customer segmenting. RFM features will play more important role in identifying chief segments when they are used first.

Second, due to importance of LTV in all stages of CRM, in this study the best framework was chosen based on the LTV dispersion quality. In other words, the selected framework was able to differentiate clusters better amongst the other frameworks regarding their LTVs. Third, the innovative rank-based visualisation method in this study resulted in more meaningful profiles. Designing different marketing plans would be facilitated when the comparison is based on the proposed rank-based profiling.

This study had also very interesting implications for the target business. The results revealed that the most valuable segment was a relatively small segment with a low annual income, a moderate number of cars owned, and a moderate level of education and occupation. In other words, the income of target food chain store was mostly due to their average customers. Besides, the largest segment had a low amount of average LTV (see Table 5). Hence, customer development strategies were proposed for this segment.

For future research, other types of input parameters should be considered. It is because the most valuable segment did not have any significant priority regarding these variables in comparison to other segments (see Table 5). It means that other deterministic variables in food industry, such as survey results or customer complaints, should be explored to achieve more meaningful profiles. Besides, to better visualise the segments, other profiling methods should be combined with rank-based approach to build more meaningful profiles. Finally, it is suggested that researchers should pay more attentions to the relatively unpopular Two Step clustering algorithm, because of its independency of number of clusters and also its better segmentation quality in this study.

References

Boone, D.S. and Roehm, M. (2002) 'Retail segmentation using artificial neural networks', *International Journal of Research in Marketing*, Vol. 19, No. 3, pp.287–301.

Chan, C.C.H. (2008) 'Intelligent value-based customer segmentation method for campaign management: a case study of automobile retailer', *Expert Systems with Applications*, Vol. 34, No. 4, pp.2754–2762.

Chang, H.J., Hung, L.P. and Ho, C.L. (2007) 'An anticipation model of potential customers' purchasing behavior based on clustering analysis and association rules analysis', *Expert Systems with Applications*, Vol. 32, No. 3, pp.753–764.

Cheng, C.H. and Chen, Y.S. (2009) 'Classifying the segmentation of customer value via RFM model and RS theory', *Expert Systems with Applications*, Vol. 36, No. 3, pp.4176–4184.

Davies, D.L. and Bouldin, D.W. (1979) 'A cluster separation measure', *IEEE Trans. Pattern Anal. Mach. Intelligence*, Vol. 1, No. 2, pp.224–227.

Gil-Saura, I. and Ruiz-Molina, M.E. (2009) 'Customer segmentation based on commitment and ICT use', *Industrial Management & Data Systems*, Vol. 109, No. 2, pp.206–223.

Hruschka, H. (1986) 'Market definition and segmentation using fuzzy clustering methods', *International Journal of Research in Marketing*, Vol. 3, No. 2, pp.117–134.

Hsieh, N.C. (2004) 'An integrated data mining and behavioral scoring model for analyzing bank customers', *Expert Systems with Applications*, Vol. 27, No. 4, pp.623–633.

Huang, J.J., Tzeng, G.H. and Ong, C.S. (2007) 'Marketing segmentation using support vector clustering', *Expert Systems with Applications*, Vol. 32, No. 2, pp.13–317.

Hung, C. and Tsai, C.F. (2008) 'Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand', *Expert Systems with Applications*, Vol. 34, No. 1, pp.780–787.

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*, Vol. 26, No. 2, pp.181–188.

Jonker, J.J., Piersma, N. and Poel, D.V. (2004) 'Joint optimization of customer segmentation and marketing policy to maximize long-term profitability', *Expert Systems with Applications*, Vol. 27, No. 2, pp.159–168.

Jutla, D., Craig, J. and Bodorik, P. (2001) 'Enabling and measuring electronic customer relationship management readiness', *Proceedings of the 34th Annual Hawaii International Conference on System Sciences Organizational Systems and Technologies Track*, pp.1–10.

Kim, K.J. and Ahn, H. (2008) 'A recommender system using GA K-means clustering in an online shopping market', *Expert Systems with Applications*, Vol. 34, No. 2, pp.1200–1209.

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*, Vol. 31, No. 1, pp.101–107.

Kim, Y.S., Street, W.N., Russell, G.J. and Menczer, F. (2005) 'Customer targeting: a neural network approach guided by genetic algorithms', *Management Science*, Vol. 51, No. 2, pp.264–276.

Lee, J.H. and Park, S.C. (2005) 'Intelligent profitable customers segmentation system based on business intelligence tools', *Expert Systems with Applications*, Vol. 29, No. 1, pp.2–152.

Lee, S.C., Suh, Y.H., Kim, J.K. and Lee, K.J. (2004) 'A cross-national market segmentation of online game industry using SOM', *Expert Systems with Applications*, Vol. 27, No. 4, pp.599–570.

Lessman, S. and Voβ, S. (2009) 'A reference model for customer-centric data mining with support vector machines', *European Journal of Operations Research*, Vol. 199, No. 2, pp.520–530.

Mahdavi, I., Movahednejad, M. and Abdesh, F. (2010) 'Designing customer-oriented catalogs in e-CRM using an effective self-adaptive genetic algorithm', *Expert Systems with Applications*, Vol. 38, No. 1, pp.631–639.

McCarty, J.A. and Hastak, M. (2007) 'Segmentation approaches in data-mining: a comparison of RFM, CHAID, and logistic regression', *Journal of Business Research*, Vol. 60, No. 6, pp.656–662.

Mo, J., Kiang, M.Y., Zou, P. and Li, Y. (2010) 'A two-stage clustering approach for multi-region segmentation', *Expert Systems with Applications*, Vol. 37, No. 10, pp.7120–7131.

Namvar, M., Fathian, M., Akhavan, P. and Gholamian, M.R. (2010) 'Exploring the impacts of intellectual property on intellectual capital and company performance: the case of Iranian computer and electronic organization', *Management Decision*, Vol. 48, No. 5, pp.676–697.

Namvar, M., Gholamian, M.R. and KhakAbi, S. (2010) 'A two phase clustering method for intelligent customer segmentation', *Proceeding of First International Conference on Intelligent systems, Modeling and Simulation*, pp.215–219, Liverpool.

Shearer, C. (2000) 'The CRISP-DM model: the new blueprint for data mining', *Journal of Data Warehousing*, Vol. 5, No. 4, pp.13–22.

Sheu, J.J., Su, Y.H. and Chu, K.T. (2009) 'Segmenting online game customers – the perspective of experiential marketing', *Expert Systems with Applications*, Vol. 36, No. 4, pp.8487–8495.

Stone, M., Woodcock, N. and Wilson, M. (2006) 'Managing the change from marketing planning to customer relationship management', *Long Range Planning*, Vol. 29, No. 5, pp.675–683.

Wang, C.H. (2010) 'Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques', *Expert Systems with Applications*, Vol. 37, No. 2, pp.8395–8400.

Wedel, M. and Kamakura, W.A. (2000) *Market Segmentation: Conceptual and Methodological Foundations*, 2nd ed., Kluwer, Dordrecht.

Wu, R.S. and Chou, P.H. (2010) 'Customer segmentation of multiple category data in e-commerce using a soft-clustering approach', *Electronic Commerce Research and Applications*, Vol. 10, No. 3, pp.331–341.

Zhang, T., Ramakrishnon, R. and Livny, M. (1996) 'BIRCH: an efficient data clustering method for very large databases', *Proceedings of the ACM SIGMOD Conference on Management of Data*, pp.103–114, Montreal, Canada.