

CLASSIFYING CUSTOMERS TO IMPROVE RETAILER'S PROFIT FOR CRM

P.Umamaheswari, M.Jeya Pandian, P.Venkateswari

SRC, SASTRA Deemed University, Tamil Nadu, India

ABSTRACT

Objective: With the rapid change in the business sector and frequent changes with the buying behaviour of the customers, we need a way to establish the relationship with the customers which leads to the study of Customer Relationship Management¹ (CRM) which is very much helpful to maintain a good customer relationship. **Methodology:** Here, data mining techniques have been used combined with self-organizing maps (SOM) and clustering algorithms-K-Medoids, ROCK, Bisecting K-Means. RFM (recency, frequency, and monetary) technique is also used to categorize customers and then to implement certain measures for the respective clusters. **Findings:** The data mining techniques² used here helps to classify customers into four types which includes loyal, potential, new and lost customers. **Applications/Improvements:** by using above mentioned techniques we found that hierarchical clustering prove to be efficient as K value increases.

Keywords: SOM, RFM, K-Medoids, Hierarchical Clustering, CRM.

1. INTRODUCTION

In today's competitive era, moving in the direction of customer-oriented markets with increased access to customer's transaction data, classifying customers who are loyal and estimating how long a customer would pertain in our business plays a crucial role. Innovation is what makes us unique from the rest of our competitors. Such activities used to capture the customer needs to get better customer satisfaction and to retain our customers. In this view, CRM is a broadly recognized strategy in acquiring and retaining customers. Implementing customer relationship management strategy helps organizations to identify and segment customers and create long-term relationships with them, and as a result, they can maximize customer lifetime value and it is important that customer and company agree on product value and good customer services. Therefore, organizations should not seek to develop a product to satisfy their customers, but they should track customer buying behaviour and present specific items for each segment. That is, customer segmentation based on the purchase behaviour is important for developing outstanding marketing strategies, which in successively causes creating and maintaining a competing advantage.

A problem in applying marketing strategies is how to appropriately partition the market and package separate products and services for the targeted customer segments. Segmentation is a process of splitting merchandise into specific groups that needs unique experiences or marketing strategies.

Clustering the customers is the most important techniques used to identify these segments. Different types of clustering techniques can be used for identifying the segments.

The techniques discussed generally takes into account the identification and analysis of various customer behaviours such as purchasing power, how much the customer consumes and things that can attract the different types of customers, their experience and the services.

Current methods of customer value analysis based on past customer behaviour patterns or demographic variables, are limited for predicting future customer behaviour. So today, analysts are trying to use better patterns for analyzing customer value. RFM³ method is one of the most common methods for segmenting and identifying customer value in the organization. This method use three variables to extract the behavioural characteristics of customers and cluster customers: Recency, Frequency and Monetary(RFM). Clustering on the basis of RFM attributes provide more knowledge about the customers' original marketing behaviour levels than the other cluster analyses. This also helps us to guess the future buying behaviour of the customer that how recent the customer would possibly purchase, how often the customer can make a purchase, and then the value of the customers' purchases amount to. Here the results of this method have been presented as inputs for clustering algorithms to determine customer loyalty. By Providing expected products of customers, there is a possibility of improvement in customer relationship management and finally loyalty promotion are the results of such analyzes.

2. RELATED WORK

2.1 CRM

The major task of identifying the various kinds of customers in order to meet the demands in terms of their product/services is of utmost importance to companies that want to sustain and stay on the top of the business ladder. CRM⁴ is nothing but the adoption of IT to bring in new customers and to make old customers sustain in our company so as to continue a long-term relationship with the customers, which helps to improve customers' loyalty, retention and profitability. The plan of CRM is to make the customer to keep on loyal to the company by maintaining enduring relationships. Therefore, CRM gives an integrated and distinct view of customers done using logical tools, relationship that is one way at any rate of the channel of communication, how effective and efficient the processes are which are involved in building the customer relationship.

2.2 RFM

Recency Frequency Monetary is a behavior-based model that is used to observe the customers' behavior and then guess their activities in the database. These three measures are combined to represent a RFM code. Recency is a measure of how recent the customer has made a purchased which can be measured in terms of number of days or months. Frequency is a measure of the number of transactions made in a specific duration of time. Monetary is a measure of the money spent during the period. In most of the researches, two methods are common for classifying loyal customers⁵, one of them is in terms of demographic attributes

(i.e age, gender, etc.) and the other is in terms of interactive attributes of customers that are expressed with RFM.

There are many ways for calculating RFM scores. The approach which we have used in this paper is based on averages. Recency can be calculated by taking the average of the “day of purchase”. Frequency can be calculated by taking the sum of product the number of purchases with the respective customer.id and dividing it by the total number of purchases. Monetary can be found from the average of the amount spent by all customers⁶.

2.3 SOM

A Self-organizing Map is a data visualization technique which employs unsupervised learning. SOM maps⁷ multidimensional valued data to a lower dimensional spaces where the geometric relationship between the points are used to indicate their similarity. This reduction in dimensionality allows people to see and observe what would otherwise be, for all intents and purposes, indecipherable data. SOM¹⁰ comprises of points which are represented as nodes(also called neurons). Associated with each node is a weight vector having the same dimension as given as the input data points, and also a point on the map space. The nodes are arranged in a 2D space which can have grids of shape either as a hexagon or rectangle. SOM has a computational layer of nodes that are arranged as rows and columns and exist as a single layer. Every neuron is completely united to all the other nodes of the input data. A 1D map will have only one row or column in the computational layer.

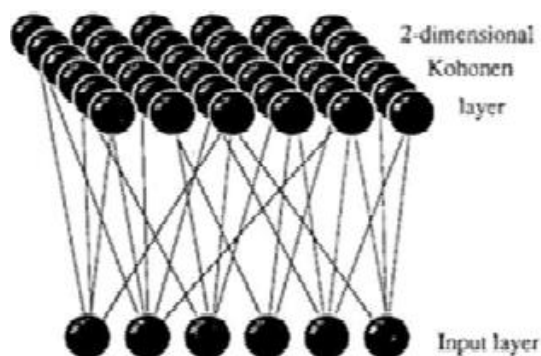


Figure 1. A simple kohonen network

Each node is associated with a specific topological position (a 2D coordinate lattice) containing vector of weights that have the dimension as the one that exist in the input vectors. Each neuron contains vector with weight W , of n dimensions.

Following are the steps involved in training SOM:

1. First, each node is initialized with a weight.
2. From the training dataset, randomly a vector is selected and then offered to the lattice.

3. Each and every node is verified to determine which one's weights are most related to the input vector. The node with the most similar weight is said to be a winning node. It is also called as the **Best Matching Unit(BMU)**.
4. The radius has been calculated for the neighbourhood of the BMU. This value is huge initially, as it will be the 'radius' of the lattice, but it will lessen for each iteration. All the nodes that are found within this radius are meant to be inside the neighbourhood of BMU.
5. The weights of the neighbouring nodes (the nodes found in step 4) are changed to make them more like the input vector and then the more nearer a node is to the BMU, the more its weights gets accustomed.
6. Do again the steps for N iterations.

2.4 K-MEANS

A commonly used partition-based clustering technique is K-means¹⁴. It is an iterative refinement technique which is usually very fast. Given an initial dataset, the centroids can be obtained by (1)based in the current centroids the data points can be assigned. (2) based on the assigned centroids choose the centroids (points that are the center of a cluster). The random partitioning method is used for initialization wherein a transaction is selected at random and centroid calculations are performed. A point is in a particular cluster if it is closer to the cluster's centroid than the remaining centroid. K-Means algorithm¹⁵ determines the best breaking of n entities into k clusters, so that the inter-distance between the clusters is minimized. The final aim is to segment the n data items in k clusters where $X_i, i=1, 2, \dots, k$ in order to minimize the within-cluster sum of squares (WCSS), defined as:

$$\sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2$$

where $\|x_i^j - c_j\|$ gives the Euclidean distance between a point in the set and the clusters' centroid value.

2.5 K-MEDOIDS

K-medoids is another partitioning technique which clusters data set having some m objects into k clusters. The value of k is taken from the result of Self Organising Maps (SOM). A medoid is an object that represents the cluster. K-medoids has two versions, namely **PAM**(Partitioning Around Medoids) and **CLARA** (Clustering Large Applications). Here we are using the former, i.e. PAM which is as follows:

- Arbitrarily choose k values of the m data object as the medoids.

- Each data object should be associated to the nearest medoid. (Closeness measure is done using Manhattan distance.)
- For every medoid n
 - For each non-medoid data object o'
 - Swap n and o' and compute the total cost

- Pick the configuration with the tiniest cost.
- Do again the same steps until there is no alteration in the medoid value.

Transactions are chosen at random and are made the centers of the clusters. Calculate the distance between each center and the other points and then each object is associated to its closest medoids. Manhattan distance is used for calculating the cost as follows:

$$\text{cost}(a, c) = \sum_{i=1}^n |a_i - c_i|$$

Where a is any data object, c is the medoid, and n is the dimension of the object .

2.6 ROCK

It is the category of agglomerative hierarchical clustering algorithm. It proceeds as follows:



Figure 2. ROCK process

The common neighbor which is between two objects is called *links*. If two points are having similar neighborhoods, then those two points are belonging to the same cluster and so they can be merged. For example, if two points l_i and l_j are neighbors and $\text{SIM}(l_i, l_j) \geq \alpha$, then these points can be grouped into a cluster. Here, SIM is the similarity function and α is the user-specified threshold. SIM can be used to measure any distance metric or even a non-metric that is normalized, so the values fall between 0 and 1. Larger values of SIM indicate that the points are similar. The number of neighbors between l_i and l_j is used to count the number of links between l_i and l_j . If the number of links between two points is more, then it may belong to the same cluster. ROCK is more robust than other standard clustering methods because it considers the neighboring data points. It focuses only on point similarity.

The concepts of neighbors and links between any two points or transactions t_i and t_j are calculated Jaccard coefficient which is defined as:

$$\text{SIM}(t_i, t_j) = \frac{|t_i \cap t_j|}{|t_i \cup t_j|}$$

3. WORKFLOW PROPOSED METHOD

The workflow for our proposed method is as follows:

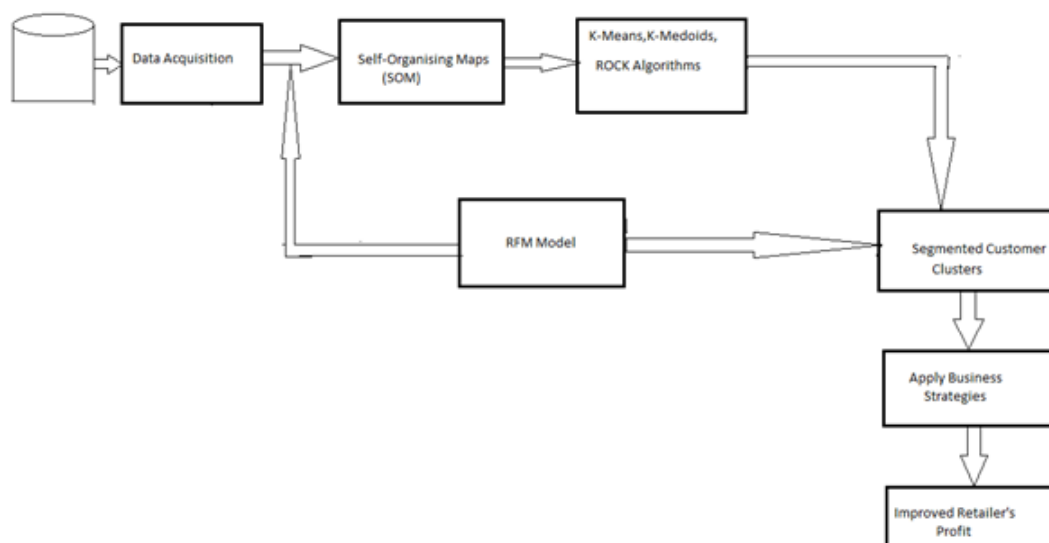


Figure 3. System Architecture-Workflow

The transaction data that we used were for duration of 3 months and was taken from a supermarket. Each transaction includes 6 columns namely customer id, purchase days, product id, price, quantity, amount. There are a total of 409 transactions as processed by 30 customers. As defined earlier, we calculated recency by taking the average of the purchase days of all these transactions and then frequency calculated by taking the sum of product the number of purchases with the respective customer.id also dividing it by the total number of purchases. We calculated the monetary by taking the average of the amount spent by the customers.

4. RESULTS AND EVALUATION

The initial statistics for the R (recency), F (frequency) and M (monetary) values are calculated and listed in Table 1. Then the minimum and maximum recency values are 1 and 90 respectively. The extremities for the frequency are calculated to be 1 and 59. For monetary the minimum and maximum values are 11 and 1800 in Indian rupees.

SOM technique is applied on the dataset and approximate k value is found. Since every iteration of the algorithm yields different values, the maximum repeated value is chosen as the final approximate k value. The result of it is shown in Figure 4.

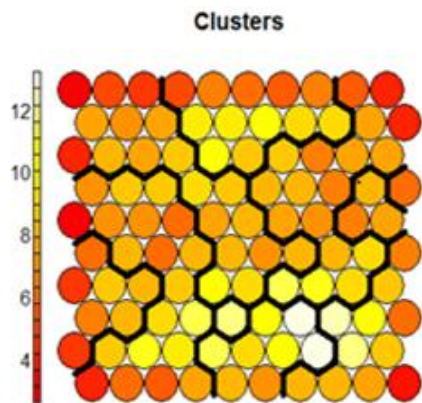


Figure 4. Ten clusters developed by SOM method.

The k value obtained from SOM is given as the input to further clustering algorithms such as k-means, k-medoids and ROCK.

The RFM strategy is applied again to the clusters obtained from the clustering algorithms.

	Minimum	Maximum	Average
Recency	1	90	46.44
Frequency	1	59	17.78
Monetary	11	1800	385.11

Table 1. The initial values of Recency, Frequency, and Monetary

For each cluster average R (recency), average F (frequency), average M (monetary) is calculated. Table 2 summarizes the average R, F, M and the symbol R, F, and M for those above the average and then a comparison is done between the initial values and the values obtained after clustering.

CLUSTER	AVG.R	AVG.F	AVG.M	RFM
1	49.75	17.82	1018.79	R F M
2	46.34	18.25	226.64	_ F _
3	45.65	16.50	922.38	_ _ M
4	47.23	17.45	381.83	R _ _
5	49.38	16.00	639.26	R _ _
6	50.11	19.15	639.26	R F M
7	44.34	17.49	59.61	_ _ _
8	41.91	18.64	1237.14	_ F M
9	41.91	16.75	558.5	R _ M
10	38.73	2.73	1780.8	_ F M

Table 2. A comparison between initial and final RFM values.

By observing Table 2, we can notice that if the values of average R, F, M exceeds the initial average, it is represented by either R/F/M. Clusters 1 and 6 have larger R, F and M values than the initial R, F, M values, indicating that they are *Loyal Customers* and they contribute more towards the supermarket. For this type of customers, the company can send regular catalogues, understand their purchasing behaviour and their needs through the transaction they make and can often send them frequent messages to provide them information about the offers available for the new products. This can increase their loyalty.

Cluster 8 and 10 shows an increase in F and M value. This means that these customers are highly capable of purchasing but have not shown up recently. Hence they can be categorised as *Potential Customers*. They purchase frequently and pay a large amount so that the management should take care that they visit regularly. For these types of customers, the company should introduce more discount policies and fulfil their daily needs.

Clusters 4 and 5 have a greater R values, but their average values of F and M are smaller than the initial values of F and M. They are considered as *New Customers*. They purchase more recently and are able to be loyal clients for a long span of time. The supermarket could enhance relationship between these customers by often contacting with these customers and in search of ways to meet out their demands so that they visit more frequently and purchase in bulk. For example, we can offer special discounts to these customers so that they are encouraged to purchase more often. There can be another version of *New Customers*. Cluster 9 has a greater R and M value, but smaller F value. This means that these types of customers have recently visited and have spent more money. Company must introduce new schemes such that they visit the store frequently.

Finally, clusters 2, 3, and 7 can be categorised as *Lost Customers*. Cluster 2 has only high F value. Since R and M play a major role, these customers can be retained only if the company introduces some cheap merchandise. Cluster 3 has a greater value for M. Though these customers have spent more money, they have not visited the shop frequently. The values of R, F, M are smaller in Cluster 7 than the average. Hence these customers should be retained by giving ultra-cheap discounts so that the relationship between the shop and the customers grows stronger.

Further studies can be made so as to analyse the purchase pattern of the individual clusters and to apply the marketing strategies effectively.

5. CONCLUSION

The main aim of this paper is customer segmentation and then measuring their loyalty. First, behavioural attributes, recency, frequency and monetary were obtained using RFM model, then customers were segmented by using k-means, k-medoids, ROCK and the results were compared. Also the optimal algorithm was found out by comparing the results obtained. We found that hierarchical clustering prove to be efficient as k value increases.

Hence, we recommend next generation researches to analyze customer value based on other attributes apart from the three that we have used.

Finally, it should be noted that it is much better for the stores within the country, to try to determine and retain profitable and loyal customers using data mining techniques, rather than paying a heavy cost for attracting new customers.

6. REFERENCES

- [1] A. Bhatnagar, and S. Ghose, "**A latent class segmentation analysis of e-shoppers**", (2004). Journal of Business Research, Vol.57(7), pp.758–767.
- [2] Pavel Berkhin, "**Survey of Clustering Data Mining Techniques**", Accrue Software, Inc. 2002.
- [3] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh . "**Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study**, (2011) Procedia Computer Science, Vol.3, pp.57-63.
- [4] V. Ravi, **Advances in Banking Technology and Management: Impacts of ICT and CRM**, (2008), Information science reference, Hershey, New York, Yurchak Printing Inc.
- [5] S.Y. Kim, T.S. Jung, E.H. Suh, & H.S. Hwang, **Customer segmentation and strategy development based on customer lifetime value: A case study**, (2006), Expert Systems with Applications, Vol.31(1), pp.101-107.
- [6] N.Glady, B. Baesens, & C. Croux, **A modified Pareto/NBD approach for predicting customer lifetime value**, (2009), Expert Systems with Applications, Vol.36(2 (Part1)), pp.2062–207.
- [7] M. H. Ghaseminezhad and A. Karami, "**A novel self-organizing map (SOM) neural network for discrete groups of data clustering**", Applied Soft Computing, vol. 11, (2011), pp. 3771–3778.
- [8] S. H. Liao, P. H. Chu and P. Y. Hsiao, "**Data mining techniques and applications: A decade review from 2000 to 2011**", Expert Systems with Applications, vol. 39, (2012), pp. 11303–11311.

- [9] S. L. Shieh and I. E. Liao, "**A new approach for data clustering and visualization using self-organizing maps**", Expert Systems with Applications, vol. 39, (2012), pp. 11924–11933.
- [10] Jo-Ting Wei, Ming-Chu Lee, Hsuan-Kai Chen, Hsin-Hung Wud , **Customer relationship management in the hairdressing industry: An application of data mining techniques** , Expert Systems with Applications, An International Journal , 2013
- [11] Ranjan, J., & Bhatnagar, V. (2011). **Role of knowledge management and analytical CRM in business**: Data mining based framework. Learn Organization, 18(2), 131–148.
- [12] Abidi, S. S. R., & Ong, J. (2000). **A data mining strategy for inductive data clustering: A synergy between self-organizing neural networks and K-means clustering techniques**. Proceedings of TENCON, 2000(2), 568–573.
- [13] Beckett, A. (2000). **Strategic and marketing implications of consumer behaviour in financial services**. Service Industry Journal, 20(3), 191–208.
- [14] Chang, E. C., Huang, H. C., & Wu, H. H. (2010). **Using K-means method and spectral clustering technique in an outfitter's value analysis**. Quality & Quantity, 44(4), 807–815.
- [15] Chiu, C. Y., Chen, Y. F., Kuo, I. T., & Ku, H. C. (2009). **An intelligent market segmentation system using K-means and particle swarm optimization**. Expert Systems with Applications, 36, 4558–4565.

