

APPLICATION OF ENHANCED CLUSTERING TECHNIQUE USING SIMILARITY MEASURE FOR MARKET SEGMENTATION

M M Kodabagi¹, Savita S Hanji², Sanjay V Hanji³

^{1,2}Department of Computer Science and Engineering,
Basaveshwar Engineering College, Bagalkot.
malik123_mk@rediffmail.com
savitawali@gmail.com

³Department of Management Studies, Basaveshwar Engineering College,
Bagalkot.
sanjayhanji_94@rediffmail.com

ABSTRACT

Segmentation is one of the very important strategic tools used by the marketer. Segmentation strategy is based on the concept that no firm can satisfy all needs of one customer or one need of all the customers. The customers are too numerous and diverse in their buying requirements, hence the marketers or companies cannot cater to the requirements of all customers that too in a broad market such as two-wheelers. Cluster analysis is a class of techniques used to identify the group of customers with similar behaviors given a large database of customer data containing their properties and past buying records. Clustering is one of the unsupervised learning method in which a set of data points are separated into uniform groups. The k-means is one of the most widely used clustering techniques used for various applications. The main drawback of original k-means clustering algorithm is dead centers. Dead centers are centers that have no associated data points. The original k-means clustering algorithm with Euclidian distance treats all features equally and does not accurately reflect the similarity among data points. In this paper, an attempt has been made to apply enhanced clustering algorithm which uses similarity measure for clustering (segmentation) of two-wheeler market data. The enhanced clustering algorithm works in two phases; Seed Point Selection and Clustering. The method adapts new strategy to cluster data points more efficiently and accurately, and also avoids dead centers. The enhanced clustering algorithm is found to be efficient in meaningful segmentation of two-wheeler market data. The results of market segmentation are discussed.

KEYWORDS

Enhanced Clustering, Market Segmentation, Two-wheelers, Similarity Measures, Seed Point Selection.

1. INTRODUCTION

In the last few years, the Indian two-wheeler industry has seen spectacular growth. The country stands next to China and Japan in terms of production and sales respectively. Majority of Indians, especially the youngsters prefer motorbikes rather than cars. Capturing a large share in the two-wheeler industry, bikes and scooters cover a major segment. Bikes are considered to be the favorite among the youth generation, as they help in easy commutation. Large varieties of two

wheelers are available in the market, known for their latest technology and enhanced mileage. Indian bikes, scooters and mopeds represent style and class for both men and women in India [1].

The domestic two-wheeler industry recorded sales volumes of 13.8 million units in 2012-13, a growth of 2.9% over the previous year. India's per capita real GDP growth at 8.6% (CAGR) over the six year period 2005-2011 had contributed substantially towards raising the standard of living of households, which in turn had been one of the key drivers of growth for the country's automobile industry [2].

The general downtrend in the economy appears to have no impact on the two-wheeler industry with many manufacturers reporting robust sales numbers for December 2013 [3].

The growing middle class population, prosperous rural India and the rarity of reliable public transport system is leading to a large number of two wheelers added to the roads every day. Indian roads in most cities, villages and towns are narrow. Two-wheelers allow people to navigate such roads easily. Fuel-efficiency is a huge advantage. With the cost of petrol increasing steadily, two-wheeler makes the daily travel both affordable and convenient. Easy availability of auto finances at attractive schemes has made a two-wheeler a must in most urban and rural homes [4].

A company cannot service all customers in a broad market such as two-wheelers. The customer are too numerous and diverse in their buying requirements. A company needs to identify the market segments which it can serve effectively. Many companies are embracing target marketing. Here sellers distinguish the major market segments, target one or more of these segments, and develop products and marketing programs tailored to each. Instead of scattering their marketing effort, they focus on the buyers they have the greatest chance of satisfying. A market segment consists of a group of customers who share a similar set of needs and wants. The marketer's task is to identify the segments and decide which one(s) to target. The company can create more fine-tuned product or service offering and price it appropriately for the target segment. The company can more easily select the best distribution and communications channels, and it will also have a clearer picture of its competitors, which are the companies going after the same segment [5].

Cluster analysis is a class of techniques used to classify objects into relatively homogeneous groups called *Clusters* [6]. Clustering is the process of organizing objects into groups whose members are similar in some way. Clustering can be applied in many fields: Marketing (finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records), Biology (classification of plants and animals given their features), and Libraries (book ordering), Insurance (identifying groups of motor insurance policy holders with a high average claim cost, identifying frauds), city planning (identifying groups of houses according to their house type, value and geographical locations), Earthquake studies (clustering observed earthquake epicenters to identify dangerous zones), WWW (Classification; Clustering weblog data to discover groups of similar access patterns).

The notion of what constitutes a good cluster depends on the application and there are many methods for finding clusters subject to various criteria, both ad hoc and systematic. These include approaches based on splitting and merging such as ISODATA, randomized approaches such as CLARA, CLARANS, and methods based on neural nets, and methods designed to scale large databases, including DBSCAN, BIRCH, and ScaleKM.

Among clustering formulations the most widely used and studied is *k-means* clustering. Given a set of n data points in real d -dimensional space, and an integer k , the problem is to determine a set of k points in d -dimensional space, called centers, so as to minimize the mean squared distance

from each data point to its nearest center. This measure is often called the squared-error distortion and this type of clustering falls into the general category of variance based clustering.

As with any technology there are challenges while clustering of data. The clustering algorithm works on the assumption that the initial centers are provided. The search for the final clusters or centers starts from these initial centers. Without the proper initialization the algorithm may generate a set of poor final centers and this problem can become serious if the data are clustered using an on-line *k-means* clustering algorithm. In general, there are three basic problems that normally arise during clustering namely dead centers, local minima and centers redundancy [7]. Dead centers are centers that have no members or associated data. These centers are normally located between two active centers or outside the data range. The problem may arise due to bad initial centers, possibly because the centers have been initialised too far away from data. Therefore, it is a good idea to select the initial centers randomly from the data or to set them to some random values within the data range. However, this does not guarantee that all the centers are equally active. Some centers may have too many members and be frequently updated during the clustering process whereas some other centers may have only a few members and are hardly ever updated.

Hence in the current work, a new clustering algorithm [8] that uses similarity measure for clustering of data is used to cluster two-wheeler market data. The enhanced clustering algorithm works in two phases; Seed Point Selection and Clustering. The method adapts new strategy to cluster data points more efficiently and accurately, and also avoids dead centers. The results of clustered data is analysed and found to be more accurate.

2. LITERATURE REVIEW

Some of the related works on clustering and its applications are summarised in the following: The distinguishing feature to demonstrate a practical procedure for conducting hybrid market segmentation is presented in [9]. In this hybrid segmentation is used as a combination of demographic, psychological, psychographic, socio-cultural, and benefit sought from the product segmentation types.

A system to analyze the performance of students using k-means clustering algorithm coupled with deterministic model is proposed in [10]. The result of analysis will assist the academic planners in evaluating the performance of students during specific semester and steps that need to be taken to improve students' performance from next batch onwards.

A hybrid procedure based on Decision Tree of Data mining method and Data Clustering that enables academicians to predict student's GPA is presented in [11]. Based on predicted students GPA instructor can take necessary step to improve student academic performance.

A hybrid procedure based on Neural Networks (NN) and Data Clustering that enables academicians to predict students' GPA is proposed in [12]. This procedure predicts students' GPA according to their foreign language performance at a first stage then classifies the student in a well-defined cluster for further advising and follows up by forming a new system entry. This procedure has mainly a twofold objective. It allows meticulous advising during registration and thus, helps maintain high retention rate, acceptable GPA and grant management. Additionally, it endows instructors an anticipated estimation of their students' capabilities during team forming and in-class participation. The results demonstrated a high level of accuracy and efficiency in identifying slow, moderate and fast learners and in endowing advisors as well as instructors an efficient tool in tackling this specific aspect of the learners' academic standards and path.

A system for analyzing students' results based on cluster analysis and uses standard statistical algorithms to arrange their scores data according to the level of their performance is described in [13]. The system also implemented k-mean clustering algorithm for analyzing students' result data. The model was combined with the deterministic model to analyze the students' results of a private Institution in Nigeria which is a good benchmark to monitor the progression of academic performance of students in higher Institution for the purpose of making an effective decision by the academic planners.

Data clustering technique named k-means clustering is applied to analyze student's learning behaviour is presented in [14]. The student's evaluation factor like class quizzes, mid and final exam assignment are studied. It is recommended that all these correlated information should be conveyed to the class advisor before the conduction of final exam. This study will help the teachers to reduce the drop out ratio to a significant level and improve the performance of students.

A heuristic method to find better initial centroids as well as more accurate clusters with less computational time is proposed in [15]. Experimental results show that the proposed algorithm generates clusters with better accuracy thus improve the performance of *k-means* clustering algorithm.

A new method to compute initial cluster centers for *k-means* clustering is presented in [16]. The method is based on an efficient technique for estimating the modes of a distribution. The new method is applied to the *k-means* algorithm. The experimental results show better performance of the proposed method.

K-means clustering algorithm based on coefficient of variation (*CV-k-means*) is proposed in [17]. The *CV-k-means* clustering algorithm uses variation coefficient weight vector to decrease the affects of irrelevant features. The experimental results show that the proposed algorithm can generate better clustering results than *k-means* algorithm do.

The empirical study to provide up-to-date assessment of cluster analysis application in marketing research and to examine the extent to which some of the ubiquitous problems associated with its usage have been addressed by marketing researchers is presented in [18]. Therefore, more than 200 journal articles published since 2000 in which cluster analysis was empirically used in a marketing research setting were analyzed.

A study in revealing typical patterns of data driven segmentation is presented in [19]. It also provides a critical analysis of emerged standards and suggests improvements.

A generalize discriminative clustering to structure and complex output variables that can be represented as graphical models is presented in [20].

It is observed from the literature survey that existing clustering algorithms based on *k-means* use kd-tree data structure. The most obvious source of inefficiency in these algorithms is that they pass no information from one stage to the next. Presumably, in the later stages, as the centers are converging to their final positions, one would expect that the vast majority of the data points have the same closest center from one stage to the next. These algorithms perform badly with increases in dimensionality. This is because the most important data structure used in the algorithm, the kd-tree, does not scale well with increases in dimension.

A good algorithm should exploit coherence to improve the running time and has to repeat number of times producing different results in different independent runs. The method should be scalable

and can be coupled with a scalable clustering algorithm to address the large-scale clustering problems as in data mining.

The standard *k-means* clustering algorithm [21] generates null clusters and also it treats all features equally and does not reflect the similarity among data. It does not maintain specified gap between cluster centroids. Hence, the current work a new enhanced *k-means* clustering algorithm [19] is taken for segmenting two-wheeler market data.

3. METHODOLOGY FOR TWO-WHEELER MARKET DATA COLLECTION

The demographic and customer need based variables were taken for the study. These variables were used to segment the two-wheeler market on the bases of customer demographics and their needs. The sample size involved was 200 respondents. Simple random sampling technique was employed for selecting the samples. The data collection tool was a well structured questionnaire which followed after several pre-tests. The questionnaire consisted mix of interval scales and nominal scales. The interval scale was used to collect the customer benefit sought (need based) segmentation data such as style, power, mileage, price, features, and low maintenance. A 5 point Likert-type Scale was used to set the interval data. The nominal scale was used to describe the demographic profiles of respondents such as age, gender, qualification, occupation, marital status, and monthly income. Data was collected from northern districts of Karnataka, India. The two wheeler market segmentation was done using newly defined similarity measure clustering technique which is discussed below.

4. CLUSTERING TECHNIQUE USING SIMILARITY MEASURE

The enhanced clustering method uses newly defined similarity measure for the purpose of clustering data points in d -dimensional space. The methodology basically has two subsystems namely Seed Point Selection and Clustering. The block diagram of proposed model is given in figure 1. The subsystems of method are explained in detail in the following section.

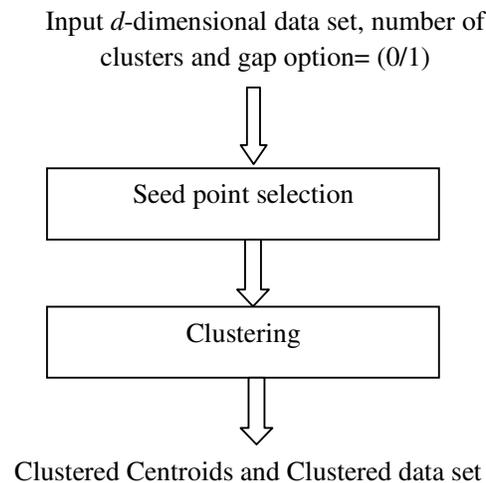


Figure 1. Block Diagram of Enhanced Clustering Method

4. 1. Seed Point Selection

In this phase, seed points i.e. initial cluster centroids are selected based on the value k (number of clusters specified by user). The minimum distance between cluster centroids can be specified by user which is optional. If user has specified the minimum distance between cluster centroids, the first k data points from the given data set that satisfy following two conditions are taken as seed points:

1. The data points should be unique.
2. The distance between data points should be equal to or greater than minimum distance specified by user.

If user has not specified the minimum distance between cluster centroids, then the first k different data points from the given data set are taken as seed points.

The distance between two data points $d_1(x_1, x_2, \dots, x_n)$ and $d_2(y_1, y_2, \dots, y_n)$ in d -dimensional space is evaluated using Euclidian Distance as described in equation (1):

$$\text{Euclidian distance} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

where x_i and y_i are attributes of data points d_1 and d_2 in i^{th} dimension respectively. n is number of data points.

Procedure for Seed Point Selection is given in Algorithm1 below:

Algorithm1: Seed Point Selection Algorithm

Input: Data set, k value and minimum distance between cluster centroids dc (Optional).

Output: Array containing Seed Points.

Step 1: Initialize $i=1$;
Store i^{th} data point from given data set in array containing seed points.

Step 2: Repeat till k seed points are selected
Increment i ;
Select i^{th} data point d_i from given data set;
Compare d_i with all seed points stored in array containing seed points.
if d_i is equal to any of the seed point in array containing seed points
Repeat Step 2;
else if dc is specified
for each seed point s in array containing seed points
Compute distance g between d_i and s
if g is equal to or greater than dc
Store the data point d_i as seed point in array containing seed points.

```

else
    Repeat Step 2
end
end
end
else
    Store the data point  $d_i$  as seed point in array containing seed points.
End
STOP

```

4.2. Clustering

In this phase, clustering of data points is carried out using newly defined similarity measure. Each data point is taken from given data set and a similarity measure is calculated between taken data point and all seed points. This process is repeated for all n data points and the result is stored in $k \times n$ sized matrix which is called Similarity Measure Matrix. Once similarity measure matrix is built, based on the value of similarity measure, each data point is assigned the group. In similarity measure matrix, for each data point similarity measures are calculated with respect to all k cluster centroids. Higher the similarity measure value more closely is the data point associated to the cluster centroid. The data point belongs to that centroid which has maximum similarity measure. Given a set of n data points (d_1, d_2, \dots, d_n), where each data point is a d -dimensional real vector, *k-means* clustering aims to partition the n data points into k sets ($k \leq n$). The similarity measure between cluster centroid c (x_1, x_2, \dots, x_m) and the data set s (y_1, y_2, \dots, y_m) is calculated using formula depicted in equation (2).

$$\text{Similarity measure} = 1 - \sum_{i=1}^m \frac{(x_i - y_i)}{(x_i + y_i)} \quad (2)$$

Once clustered data matrix is built, the new centroids are evaluated for clustered data. The centroid of m data points i.e. $d_1, d_2, d_3, \dots, d_m$ each with n dimensions is given using following formula:

$d_{11}, d_{12}, d_{13}, \dots, d_{1n}$ are feature of d_1 .

$d_{21}, d_{22}, d_{23}, \dots, d_{2m}$ are feature of d_2 and soon for n data points.

$$\text{Centroid} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n d_{ij} \quad (3)$$

This process is repeated till convergence is reached. Clustering will reach *convergence point* when clustered data matrix of successive loops are same i.e. data points does not change their cluster in further clustered data matrix evaluation. Suppose G_n is the clustered data at n^{th} iteration and G_{n+1} is the clustered data matrix at $n+1$ iteration. The convergence data points will be reached when:

$$G_n = G_{n+1} \quad (4)$$

Algorithm

Procedure for Clustering is given in Algorithm2 below:

Algorithm2: Clustering Algorithm

- Description: The proposed clustering algorithm.
- Input: The proposed clustering algorithm takes three inputs.
1. Set of n data points in d -dimensional space.
 2. Number of clusters (k).
 3. Gap option set to either 0 or 1. 1 is set to cluster data points maintaining minimum distance between from clustered centroids and 0 to cluster data points without maintaining minimum distance between clustered centroids.
- Output: Centroids of final clusters and clustered data points.
- Step 1: Initialize clustered data matrix to zero matrix. If third parameter is set to 1, accept minimum distance between cluster centroids from user.
- Step 2: Select Seed Points based on the third parameter. If third parameter is set to 1, then select first k data points from given data set that satisfy two conditions described in Section 2.1.1 else directly select first k unique data points from given data set.
- Step 3: For each data point, calculate the Similarity Measures to all the centroids.
- Step 4: Assign each data point to the centroid that has maximum similarity measure (i.e Building clustered data matrix).
- Step 5: Check the Convergence. If convergence is reached, go to Step 8 else go to Step 6.
- Step 6: Calculate new mean for newly clustered data points to be centroids of clusters. If third parameter is set to 1 then check the newly calculated centroids maintain minimum distance between cluster centroids. If cluster centroids fails to maintain minimum then go to Step 8 else go to Step 7
- Step 7: Repeat Step 3 to Step 5.
- Step 8: STOP.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The enhanced clustering algorithm was tested for two-wheeler market. It is observed that the enhanced clustering mechanism is efficient when compared to original *k-means* clustering algorithm. Experimental results showed that the enhanced clustering algorithm solved the problem dead centers that were found in original *k-means* clustering algorithm as shown in figure 2 given below;

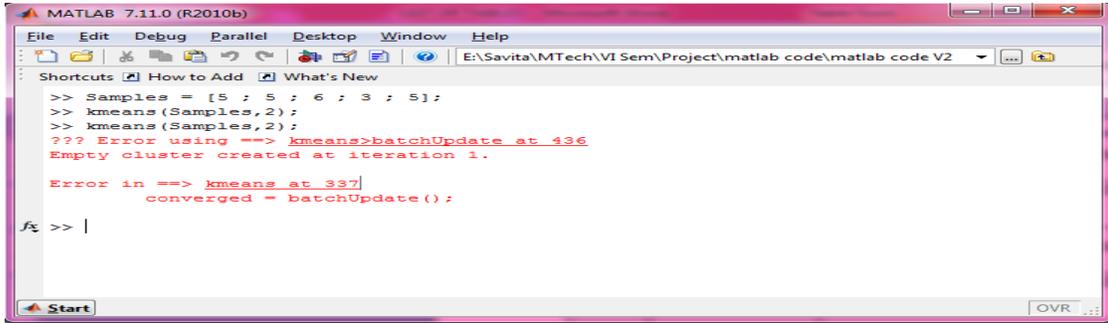


Figure 2: Error in Original k-means clustering algorithm

Table 1. Cluster solutions for the two-wheeler data

Final Cluster Centers					
S.No	Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	Age	1.5	1.2	2.2	1.8
2	Gender	1.2	1.1	1.1	1.2
3	Qualification	1.7	1.4	2.0	1.9
4	Occupation	1.9	2.7	1.5	2.2
5	Marital Status	1.7	1.9	1.2	1.5
6	Monthly Income	1.2	1.1	1.7	1.3
7	Style	4.2	4.5	3.5	2.2
8	Power	4.3	4.2	4.1	1.9
9	Mileage	3.3	3.0	3.9	3.9
10	Price	2.6	3.4	3.6	4.2
11	Features	2.7	4.1	4.0	3.9
12	Maintenance	2.8	4.2	4.4	4.3

The table 1 shows four cluster solutions for the two-wheeler data. The demographic and customer need based variables along with their final cluster centers for each of the clusters (or segments) is shown. After analysing the results, more detailed interpretation of the final cluster centers for each of the clusters is displayed in table 2.

Table 2. Interpretation of Final Cluster Centers

Final Cluster Centers					
S.No	Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	Age	20 to 30 years	20 to 25 years	25 to 35 years	25 to 35 years
2	Gender	Majorly Male	Majorly Male	Majorly Male	Majorly Male
3	Qualification	Many are Graduates & Some are Undergraduates	Both Graduates & Undergraduates	Graduates	Graduates
4	Occupation	Job-Holders	Both Job-Holders & Students	Both Job-Holders & Businessmen	Job-Holders

5	Marital Status	Many Unmarried and Some Married	Unmarried	Many are Married & Some are Unmarried	Both Married & Unmarried
6	Monthly Income	0 to Rs.15000/-	0 to Rs.15000/-	Most have Rs.15000 to Rs.30000 & few have 0 to Rs.15000.	Most have 0 to Rs.15000 & few have Rs.15000 to Rs.30000
7	Need for Style	Agree	Somewhat Strongly Agree	Somewhat Agree	Disagree
8	Need for Power	Agree	Agree	Agree	Disagree
9	Need for Mileage	Somewhat Agree	Neither Agree Nor Disagree	Agree	Agree
10	Need for Low Price	Somewhat Agree	Somewhat Agree	Somewhat Agree	Agree
11	Need for More Features	Somewhat Agree	Agree	Agree	Agree
12	Need for Low Maintenance cost	Somewhat Agree	Agree	Somewhat Strongly Agree	Somewhat Strongly Agree

Cluster 1: The first cluster or the segment consists of people who are between age limits of 20 to 30 years and majority of them are male. Most of the people in this segment are graduates along with few are undergraduates and the segment is solely dominated by job holders. Here many people are unmarried with few are married ones, and their monthly income ranges to a maximum of Rs.15,000. The first thing that these people look for in a two-wheeler is its style and power followed by price, features and maintenance cost. Most of the people in this segment are unmarried young graduates and have a job so they earn their money but with no commitments. Hence these people prefer two-wheelers with style and power, and give little less importance to price of the two-wheelers, their features and maintenance cost.

Cluster 2: The second cluster or the segment comprises of people who are between 20 to 25 years of age; majority are male; same number of graduates and undergraduates are found in this cluster; and it is dominated by both jobholders and students. Mainly unmarried youths constitute this segment with monthly income up to Rs.15000. These people give very high importance to style followed by power, features and maintenance cost. They are not much bothered about the mileage part of the two-wheeler. Since most of these people are unmarried youths who are either undergraduate students with some pocket money and/or fresh graduated jobholders with some basic start-up salaries, these people assign high priorities to style followed by power, features, and low maintenance but don't care much about mileage as they don't seem to have any major commitments.

Cluster 3: The third cluster or the segment includes people who are between 25 to 35 years of age; majority are male; graduates; with both jobholders and businessmen; many married and some unmarried people. Most of these people have their monthly income range between Rs.15000 to Rs.30000 and some have income up to Rs.15000. These people pay more importance to low maintenance cost of the two-wheeler followed by power, mileage, and features. They are not much bothered about initial purchase price of the two-wheeler. This trend in their preferences

is quite obvious because these people are little more elderly than previous segments, graduates with all earning class and most of them are married along with some unmarried people. This segment has more number of people whose earning is up to Rs.30000 per month. Therefore their commitments makes them to be conscious about maintenance cost and mileage of the two-wheeler, their age drives them to look for power and features in the two-wheelers and their income substantiates their buying power, so price of the two-wheeler is not contemplated.

Cluster 4: The fourth cluster or the segment holds people who are between 25 to 35 years of age, majority are males; graduates who are doing jobs, and there are both married and unmarried people. Most of them have their monthly income ranging from 0 to Rs.15000 along with some people who have their income between Rs.15000 to Rs.30000. These people give more weight-age to maintenance cost of the two-wheeler followed by mileage, price and features of the two-wheeler. They completely ignore styling and power of the two-wheeler. These people are graduates who are both married and unmarried people and seem to be more matured when its compared to other segments. Many don't enjoy huge salaries so they are more concerned about the maintenance cost, mileage, price of the two-wheeler. They also lookout for good features in the two-wheelers but not much bothered about style and power in the two-wheelers as long as they are economical.

Table 3 shows the number of data points which actually belong to each of the four clusters. Out of total 200 data points (respondents), 23 belong to Cluster 1, 66 belong to Cluster 2, 59 belong to Cluster 3 and 52 belong to Cluster 4.

Table 3. Clustered Data Set

Number of data points in each cluster	
Cluster	Number of data points
1	23
2	66
3	59
4	52

6. CONCLUSION AND FUTURE SCOPE

In this paper, the two-wheeler market segmentation is carried using enhanced clustering algorithm that uses similarity measure for clustering of data. The enhanced clustering algorithm works in two phases; Seed Point Selection and Clustering. The method adapts new strategy to cluster data points more efficiently and accurately, and also avoids dead centers. The enhanced clustering algorithm is found to be efficient in meaningful segmentation of two-wheeler market data. The two-wheeler market was divided into four segments. The results of market segmentation are discussed. Each of the segments reveal some meaningful information which will be helpful for marketing departments of two-wheeler companies to decide which segments to target for their two-wheeler products and which segments to ignore. In that way, the companies will be in a better position to design proper positioning strategies for their products in the selected segments. These segments will also help the marketers to device their marketing mix strategies i.e., their product, price, place and promotion decisions. Future scope exists to incorporate such new clustering mechanisms to improve efficiency of clustering in various fields.

REFERENCES

- [1] Automotive Industry in India, *Auto Guide*, (1996-2014). Viewed on January 21, 2014. <http://auto.indiamart.com/two-wheelers>.
- [2] Indian Two Wheeler Industry, *ICRA Research Services*, (2013). Viewed on December 21, 2013. <http://icra.in/Files/ticker/SH-2013-Q2-1-ICRA-Two-Wheeler.pdf>.
- [3] Two Wheeler Makers Report Robust Sales in December, *The Hindu*, (2014). View on January 10, 2014. <http://www.thehindu.com/business/Industry/twowheeler-makers-report-robust-sales-in-december/article5530929.ece>.
- [4] Bagirathi Iyer, (June 2011), A Study on Reasons for Low Acceptability of Battery Operated 2 Wheelers (E-Bikes) in Nagpur Region with Special Reference to Electrotherm India Limited's Yobykes, *International Journal of Research in Finance & Marketing*, Volume1, Issue2.
- [5] Philip Kotler, (2003), *Marketing Management*, 11th Edition, Pearson Education, Delhi, India, pp: 278-299.
- [6] Naresh K. Malhotra and Satyabhushan Dash, (2007), *Marketing Research- An Applied Orientation*, Fifth Edition, Pearson Education, New Delhi, India, pp. 636-661.
- [7] M.N. Tuma, S. Scholz, and R. Decker, (2009), The Application of Cluster Analysis in Marketing Research: A Literature Analysis, *Business Quest Journal*, vol. 14.
- [8] M M Kodabagi, Savita S Hanji, Ravatappa A.B, (2013) "This is my paper", A Novel Clustering Technique Using Similarity Measure, *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, Special Issue IDEAS-2013, ISSN: 2278-621X.
- [9] H. N. Ramesh, Vinod S., Sanjay Hanji, (Jan-June 2012) "This is my paper", Hybrid Segmentation At Two Wheeler Market In India, *IEMS Journal of Management Research*, ISSN : 2249-569X, Volume1, Issue-1.
- [10] Rakesh Kumar Arora, Dr. Dharmendra Badal, (April - June 2013), Evaluating Student's Performance Using k-Means Clustering, *International Journal of Computer Science And Technology(IJCST)*, ISSN(ONLINE) : 0976-8491 ISSN(PRINT) : 2229-4333 Vol. 4, Issue-2.
- [11] Hedayetul Islam Shovon, Mahfuza Haque, (2012), An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree, *International Journal of Advanced Computer Science and Applications(IJACSA)*, Vol. 3, No. 8.
- [12] Chady El Moutary, Marie Khair, Walid Zakhem, (2011), Improving Student's Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education, *The Research Bulletin of Jordan ACM*, Vol. I (III).
- [13] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C, (2010), Application of k-Means Clustering algorithm for prediction of Students' Academic Performance, *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 7.
- [14] Md. Hedayetul Islam Shovon, Mahfuza Haque, (July 2012), Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm, *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277 128X Vol 2, Issue 7.
- [15] Mahmud M.S, Rahman M.M, Akhtar M.N, (2012), Improvement of K-means clustering algorithm with better initial centroids based on weighted average, *Electrical & Computer Engineering (ICECE), 7th International Conference*, Pages:647-650.
- [16] Xiaoping Qin, Shijue Zheng, (2009), A New Method for Initialising Clustering Algorithm, *Knowledge Acquisition and Modelin., KAM '09. Second International Symposium on Nov 30 2009* Volume: 2, Pages:41-44.
- [17] Shuhua Ren, Alin Fan, (2011), K-means clustering algorithm based on coefficient of variation, *Image and Signal Processing (CISP), 4th International Congress*, Volume:4, Pages :2076-2079.
- [18] M. Y. Mashor, (2000), Hybrid training algorithm for RBF network, *International Journal of The Computer, The Internet and Management*, vol. 8, no. 2, pp. 50-65.
- [19] Sara Dolnicar, (2003), Using Cluster analysis for market segment-typical misconceptions, established methodological weakness and some recommendations for improvements. *Australasian Journal of Market Research*.
- [20] Peter Haider, Luca Chiarandini, Ulf Brefeld, (2012), *Discriminative for Market Segmentation, KDD'12*, August 12-16, Beijing, China.
- [21] Kardi Teknomo, *K-means clustering*, viewed website on June 2013. <http://www.planetsourcecode.com/vb/scripts/ShowCode.asp?txtCodeId=26983&lngWId=1>.

AUTHORS

M M Kodabagi

Prof M M Kodabagi is currently working as Assistant Professor in Department of Computer Science & Engineering at Basaveshwar Engineering College, Bagalkot. He has more than 10 years of teaching experience in reputed institutes. He has also worked as Scientist 'B' for Electronics Radar Development Establishment, Defence Research & Development Organisation, Bangalore. He has conducted many workshops and also delivered invited lectures at various engineering colleges. He has published many papers in international journals.



Savita S Hanji

Prof. Savita S. Hanji is working as Assistant Professor in Department of Computer Science & Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India. Her areas of interest include Database Management System, Web Developments, Digital Image Processing, Data Mining and Clustering. She has conducted Oracle Certification course and certified many students. She is also certified in Cambridge University for Innovative Teaching Methodologies.



Sanjay V. Hanji

Prof. Sanjay Hanji is working as Associate Professor in Department of Management Studies, Basaveshwar Engineering College, Bagalkot, Karnataka, India. He is a marketing faculty and his subject areas of interest includes marketing research, business statistics, management information systems, consumer behavior, strategic brand management, integrated marketing communications, service marketing, sales and distribution management, and retailing management. He is pursuing his PhD in Business Administration under Visvesvaraya Technological University, Belgaum, Karnataka, India. He has published many papers in national and international journals.

