*2012 International Conference on Computer Technology and Science (ICCTS 2012)*
***IPCSIT vol. 47 (2012) © (2012) IACSIT Press, Singapore***
*DOI: 10.7763/IPCSIT.2012.V47.19*

# A Noble Approach of Clustering the Users in M-Commerce for Providing Segmented Promotion of Goods & Services Using K-means Algorithm

Dhrubasish Sarkar[+]

Dept. of Computer Applications, NSHM College of Management & Technology Kolkata,West Bengal, India

**Abstract.** In recent years, the mobile commerce is becoming more and more popular because of its portability, flexibility and mobility. Mobile commerce extends the 'anytime access' paradigm offered by the electronic commerce to that of 'anytime and from anywhere access'. In Mobile E-commerce applications, the user clustering plays an important role for segmented promotion. This paper incorporates the K-means clustering algorithm model for the purpose of segmented promotion in M-Commerce.

**Keywords:** Clustering, K-Means, M-Commerce, E-Commerce, Segmented Promotion

## 1. Introduction

The current scenario depicts a competitive market, where today's customer gives more preference to the level of technology used with respect to a new product, features such as durability, faster accessibility and most importantly the quality of information delivered at the user end. The consumption pattern of the goods and services are factor dependent such as: financial status, buying habits, culture, and media exposure etc. The tastes and preferences of the current generation are changing rapidly; changing consumption patterns trigger changes in shopping styles of customers and also the factors that drive people into stores [1]. Today's consumers are not only price sensitive but the eagerness to pay more depends on better facilities and ambience. The customer buying patterns also depends on change in marketing strategies of companies with change in consumer buying behaviour. With change in consumer buying behaviour the companies also made necessary changes in their marketing strategies utilizing the CRM and IT to create profitable, long-term relationships with the customers.

## 2. Problem Statement

Market segmentation is an appropriate method for realizing needs of different groups and providing people a chance to be observed and to realize their needs. This method facilitates one-to-one marketing [2]. The objective of this research is from a two-fold angle, firstly identifying the target audience and making sure error-free delivery of information to the consumer and secondly enabling a proper search engine which can cater to the exact query statement of the consumer. This paper incorporates the K-means clustering algorithm model for the purpose of segmented promotion in M-Commerce.

## 3. Review of Literature

With the development of wireless mobile communication technology and mobile device, the mobile e-commerce has become more popular [3]. Mobile Commerce can be defined as any electronic commerce activity conducted over the wireless network through mobile devices. According to Tarasewich, Nickerson

[+] Corresponding author.
*E-mail address*: dhrubasish@inbox.com

and Warkentin (2002) – mobile commerce includes "all activities related to a (potential) commercial transaction conducted through communications networks that interface with wireless (or mobile) devices". Compared with the traditional e-commerce, mobile e-commerce has more advantages such as mobility, remote accessibility and portability [4].

Clustering techniques are mostly unsupervised methods that can be used to organize data into groups based on similarities among the individual data items. Most clustering algorithms do not rely on assumptions common to conventional statistical methods, such as the underlying statistical distribution of data, and therefore they are useful in situations where little prior knowledge exists. The potential of clustering algorithms to reveal the underlying structures in data can be exploited in a wide variety of applications, including classification, image processing, pattern recognition, modelling and identification.

## 4. Overview of Clustering Methods

Clustering methods can be classified in a number of ways [5]. Since clusters can formally be seen as subsets of the data set, one possible classification of clustering methods can be according to whether the subsets are crisp (hard) or fuzzy.

Hard clustering methods are based on classical set theory, and require that an object either does or does not belong to a cluster. The hard clustering means partitioning the data into a specified number of mutually exclusive subsets.

Fuzzy clustering methods, however, allow the objects to belong to several clusters simultaneously, with different degrees of membership. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership.

In order to determine the prospective consumers, the clustering technique can be used. In the current paper, the K-means algorithm has been proposed for clustering the users because of its computational simplicity [6]. To achieve the desired objective firstly the database of the mobile users with relevant information has been collected and further applied the K-Means clustering technique to do some predictive analysis which can further help in focusing product marketing in M-Commerce.

## 5. K-Means Clustering Algorithm

The K-Means algorithm (MacQueen, 1967) uses the reciprocal of Euclidean distance to compute the similarity between the input vector and the clustering centre [7].

The formula of Euclidean distance is as follows:

$$d_w(x_k, c_i) = \sqrt{\sum_{j=1}^{m} (x_{kj} - c_{ij})^2}$$

Where $x_{kj}$ denotes the input vector, $c_{ij}$ denotes the clustering centre.

Usually, standard deviation is used as the standard measure function. The formula is as follows:

$$E = \sqrt{\sum_{j=1}^{k} \sum_{i \in c_j} \left| i_l - w_j \right|^2}$$

Where $i_l$ denotes the input vector, $w_j$ denotes the clustering center.

### 5.1. K-Means Algorithm

The general K-Means algorithm [8] is composed of the following steps:

1. *Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.*
2. *Assign each object to the group that has the closest centroid.*
3. *When all objects have been assigned, recalculate the positions of the K centroids.*
4. *Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.*

As mentioned above, the main reason to use K-Means clustering algorithm for analysis is because it is simple, reliable and the large data sets can be efficiently classified very fast [7] which are the primary factors for any mobile application.

## 6. Experiments & Results

In this paper, approx 500 mobile users' data has been used for clustering. The user profile includes id, age, gender, domicile, income, married and family_size.

| UserID | Age | Gender | Domicile | Income | Married | Family_Size |
|---|---|---|---|---|---|---|
| M16603 | 48 | Female | Metro | 18000 | No | 1 |
| M16604 | 40 | Male | Urban | 30000 | Yes | 3 |
| M16605 | 51 | Female | Metro | 17000 | Yes | 2 |
| M16606 | 23 | Female | Urban | 20000 | Yes | 3 |
| M16607 | 57 | Female | Rural | 51000 | Yes | 1 |

Table – 1. User Profile (sample data)

The details of the selected attributes are mentioned below.

*Attribute Name: Age, Type: Numeric, Distinct Value: 50*
Min – 18 Max – 67 Mean – 42.714 StdDev – 14.535

*Attribute Name: Gender, Type: Nominal, Distinct Value: 2*
Female – 245 Male – 255

*Attribute Name: Domicile, Type: Nominal, Distinct Value: 3*
Metro – 229 Urban – 189 Rural – 82

*Attribute Name: Income, Type: Numeric, Distinct Value: 58*
Min – 5000 Max – 63000 Mean – 27764 StdDiv – 12892.041

*Attribute Name: Married, Type: Nominal, Distinct Value: 2*
No – 172 Yes – 328

*Attribute Name: Family_Size, Type: Numeric, Distinct Value: 5*
Min – 1 Max – 5 Mean – 2.126 StdDiv – 0.992

For the K-Means clustering, the WEKA software has been used. WEKA SimpleKMeans algorithm automatically handles a mixture of categorical and numerical attributes. Furthermore, the algorithm automatically normalizes numerical attributes (like age, income, family_size etc.) when doing distance computations [9]. The user.csv file has been converted [10] into the user.arff and then the user.arff has been clustered using the WEKA. The clustered data in usercluster.arff can be used for the segmented promotion in M-Commerce.

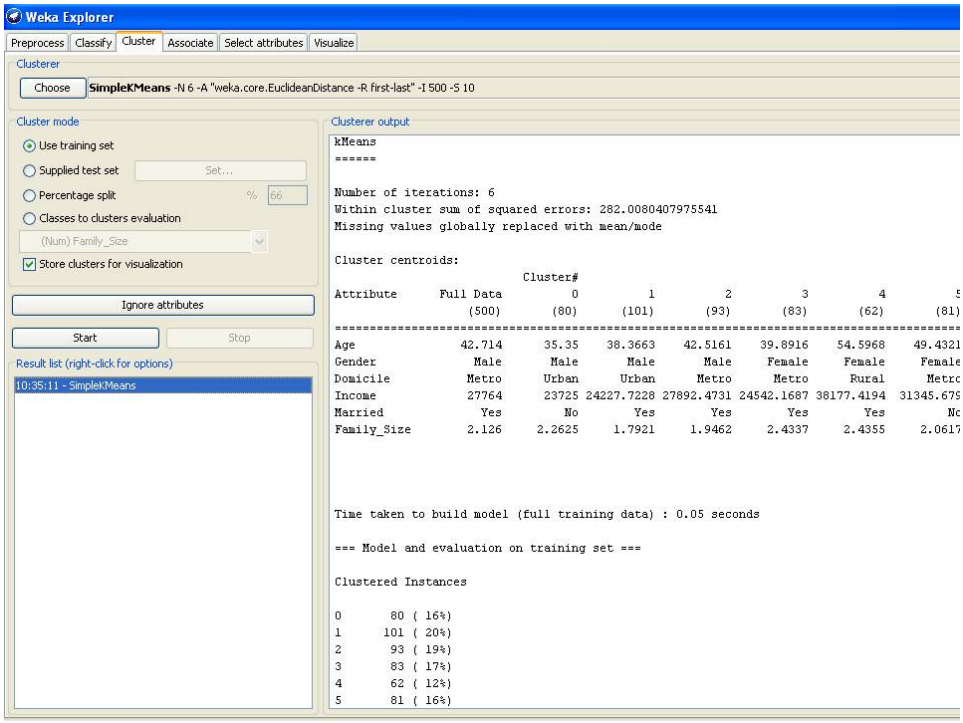


Figure 1. Clustering info

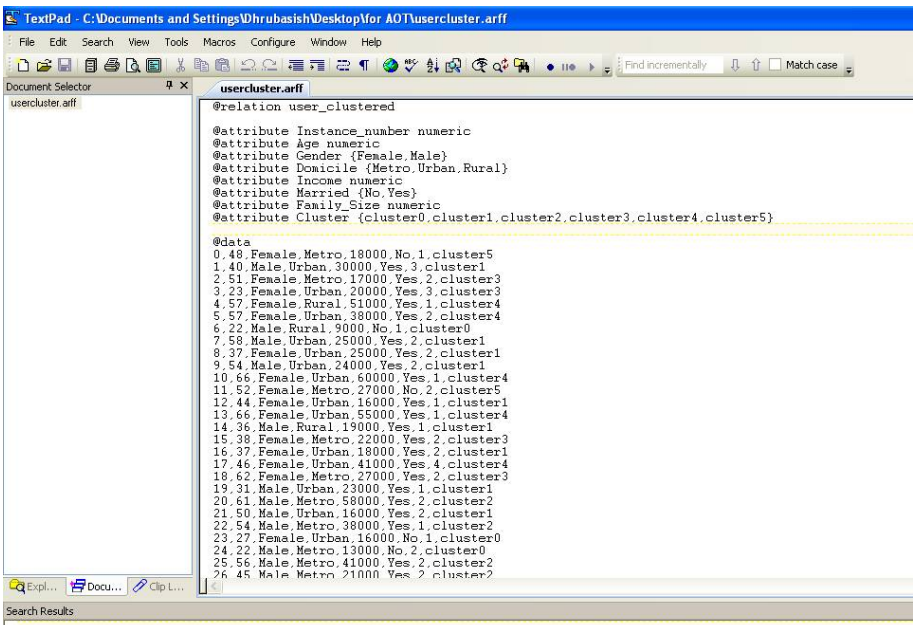


Figure 2.Clustered Data (usercluster.arff)

Clustered Instances (Total Data – 500)

| Cluster# | | Cluster# | | Cluster# | |
|---|---|---|---|---|---|
| 0 | 80 (16%) | 1 | 101 (20%) | 2 | 93 (19%) |
| 3 | 83 (17%) | 4 | 62 (12%) | 5 | 81 (16%) |

## 7. Conclusion & Future Scope

This paper proposes a recommendation algorithm based on K-means clustering where the mobile user's profile is clustered for the segmented promotion of product & services in an efficient way. The work can be further enhanced for the location based promotion, information, bill & services in M-Commerce.

## 8. References


[1] Kaur, P. and Singh, R. (2007), "Uncovering retail shopping motives of Indian youth", Young Consumers: Insight and Ideas for Responsible Marketers, Vol8, No.2, and pp.128- 138.

[2] Ahmadi P and Samsami F. "Pharmaceutical Market Segmentation using GA K-means", European Journal of Economics, Finance and Administrative Sciences ISSN 1450-2275 Issue 22 (2010).

[3] Zhang F, Liu H and Chao J. "A Two-stage Recommendation Algorithm Based on K-means Clustering In Mobile E-commerce", Journal of Computational Information Systems 6:10 (2010) 3327-3334

[4] Jie Zhang,Henry C. B. Chan. "A mobile auction service based on mobile agents: design and analysis", International Journal of Wireless and Mobile Computing, 2008, pp: 101-110

[5] Bradley, P.S., Fayyad, U.M., Reina, C.A., "Scaling EM Clustering to Large Databases", Microsoft Research Technical Report 98-35, 1998.

[6] Jigui Sun, Jie Liu. Clustering Algorithms research. Journal of Software. 2008, 19(1):48-61.

[7] Xizhao Wang, Yadong Wang, Yan Zhan. "Optimization of K-means Clustering by Weight Learning," Journal of Computer Research and Development, 2003, 40(9):869-873.

[8] http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html, accessed on Nov 15, 2011.

[9] http://maya.cs.depaul.edu/classes/ect584/weka/k-means.html, accessed on Nov 30, 2011

[10] http://slavnik.fe.uni-lj.si/markot/csv2arff/csv2arff.php, accessed on Nov 30, 2011