An Obscure Method for Clustering in Android Using k-Medoid and Apriori Algorithm

Amar Lalwani, Sriparna Banerjee, Manisha Mouly Kindo, and Syed Zishan Ali(✉)

Bhilai Institute of Technology, Raipur, Chhattisgarh, India
amarlalwani0202@gmail.com,
sriparnabanerjee15@gmail.com, mouly.kindo@gmail.com,
zishan786s@gmail.com

Abstract. In today's scenario, there is quick evolution in each field which contains majority and distinctive sorts of information. In order to differentiate sample data from the other, the amalgamation of data mining techniques with other useful algorithms is done. Android development is one of the major arena where there is tremendous need to execute these calculations. Combining frequent pattern calculation with clustering is extremely efficacious for android. In this paper the work is done in two levels, initial stage concentrates on generation of clusters and final stage deals with finding the frequent patterns.

Keywords: Android · Clustering · Itemsets

1 Introduction

Clustering is a unique multi-objective optimization problem solving technique of generating set of objects that are similar to one another but dissimilar from the objects belonging to the other set. The term "clustering" is used in several research communities to describe methods for grouping of unlabeled data. Clustering is significant in various exploratory pattern- analysis, decision-making, grouping, machine-learning situations, including data mining, pattern classification document retrieval and image segmentation. Clustering techniques [1, 2, 4–6, 12] are enforced as it is tough for humans to intuitively perceive knowledge in multi dimensional space. Here a brief overview of each of the techniques is presented and later on the methodology along with the main concept will be delineated.

2 Algorithms

Finding frequent patterns among various datasets requires the implementation of Algorithms. The Algorithms is applied according to the work and objects utilized in the subject. Below are the algorithms which is utilized in this work:

S.C. Satapathy and A. Joshi (eds.), *Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 1*, Smart Innovation, Systems and Technologies 83, DOI 10.1007/978-3-319-63673-3_9

2.1 k-Medoid Clustering

Medoids [9, 10, 15] are representative objects of a data set or a cluster with a data set which are similar in concept to means or centroids, but medoids are always members of a data set.when a mean or centroid cannot be defined medoids are used such as 3-D trajectories. *k*- medoids algorithm breaks the dataset into groups and attempts to minimize the list and between points labeled to be in clusters. A point is designated as the centre of particular cluster. In k-medoids, data points are chosen as centres (medoids or exemplars) that works with an arbitrary metrics of distances between them.

Objects are taken in account to represent an individual cluster. For each of the remaining objects the value is assigned to the cluster and based on the distance between the objects the minimum cost is compared and then evaluated. This algorithm [11, 14] is very efficacious in generating clusters which is required in most of the pragmatic cases.

The objective function is defined as

$$j = \sum_{\substack{I=1 \\ p \in C_i}}^{K} \sum_{p \in C_i} |p - o_i| \tag{1}$$

where,

E: the sum of absolute error for all objects in the data set
p: the data point in the space representing an object
$o_{i:}$ is the representative object of cluster C_i

2.2 Apriori Algorithm

The apriori algorithm [3] is one of the most broadly used tools for association rule mining. It uses priori knowledge of frequent itemset property for association rule mining. The algorithm makes use of downward closer property and is a bottom up search which moves upward level wise in the lattice. The main idea is to produce candidate itemsets of a given size and then verify if the count is actually large. Here k-itemsets are used to explore (k+1)-itemset. Therefore this is an iterative process in which we can generate the candidate of any pass by joining frequent itemsets of previous pass.

3 Methodology

The foremost concept implemented in this paper is based on the detection of frequent patterns in a particular cluster. Each object in the clusters have a distinct position that may vary depending upon the points obtained and the distinct clusters are formed by applying k-medoid clustering method along with apriori algorithm.

Step 1 Initially inputs are given as coordinates for in order to create set of clusters. The points are then imported to the RStudio [7, 8] for further plotting of the graph Table 1.

Table 1. Number of objects

Shapes		
○	+	▲
2	-	-
-	7	-
-	-	6

Step 2 Based on this information, the values of input provided for the location of the object are assigned with two centres of the cluster. Then the cost is evaluated using the Minkoswki or Manhattan distance.

Step 3 Implementing the procedure of *k*-medoid clustering [13] method, such that the minimum cost of traversal is obtained.

Step 4 The medoids using which the minimum cost is obtained is utilised in further steps.

Step 5 Final clusters will be formed after applying Apriori algorithm to find the frequent object or itemset.

Step 6 The number of objects in the corresponding clusters is highlighted by various shapes Fig. 1.

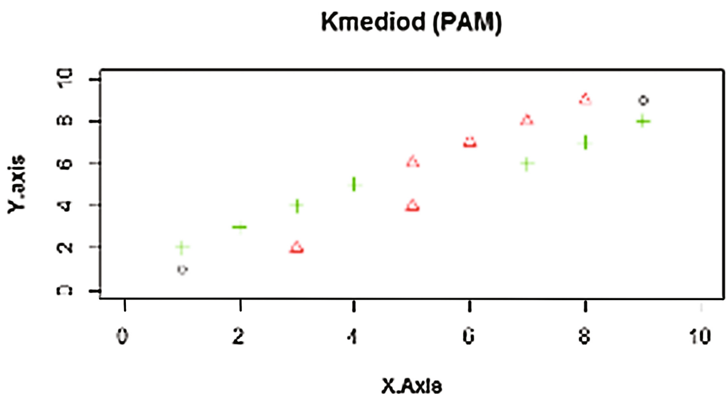

Fig. 1. Classification of objects

In the above figure there are different objects assumed, having non-identical items in the individual datasets. It is clearly displayed that the objects (triangle, star and circle) is capable of representing any living being like human, animals which is dynamic in nature Fig. 2.

Step 1. Input the number of objects to be clustered..

Step 2. Feed the coordinate values for the objects.

Step 3. Calculate the value of absolute error for all objects, then compare the value of minimum cost ‘S’. .

Step 4. If S<0 then continue else the program will terminate .

Step 5. Assign the number of clusters to be formed.

Step 6. Final clusters are plotted along with objects.

Step 7.The frequency of the objects in the respective clusters is calculated.

Fig. 2. Flow of algorithm

This concept in successfully implemented in java environment as shown below: Figs. 3 and 4

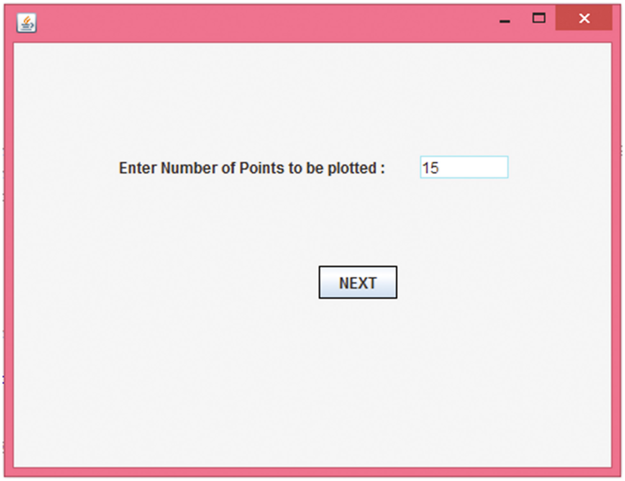

Fig. 3. Coordinate points

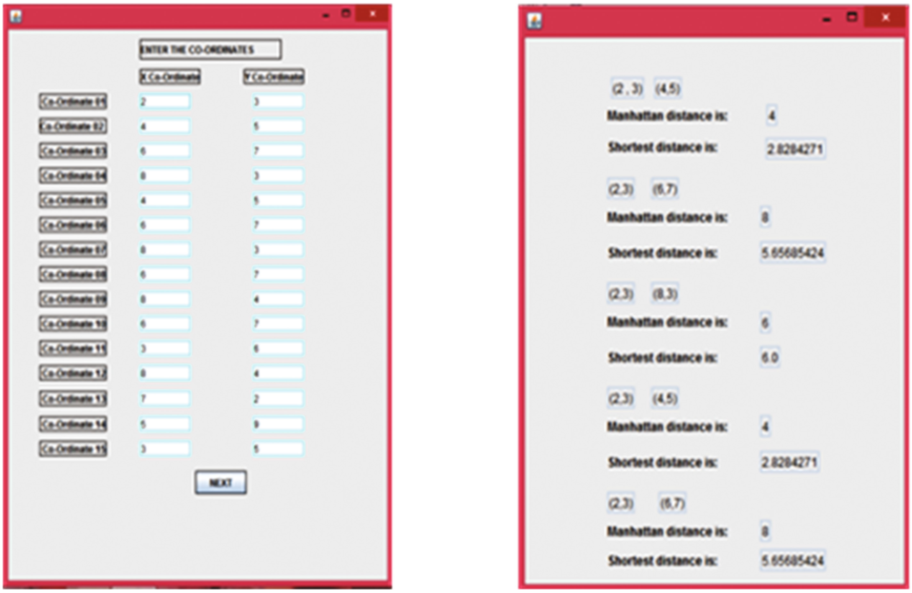

Fig. 4. Feeding coordinate values and calculating distance

4 Conclusion

As observed through the experimental analysis, a procedure for clustering in android has helped in identifying the maximum number of itemsets and their existence in an individual cluster. One of the major impact of this work is that it helps in representing the filtered objects in the cluster which can be beneficial for Android and other platforms. This work is carried out by implementing distinctive shapes and colors representing various objects that gives effective results which can be applied in Google Maps.

References

1. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 881–892 (2002)
2. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. KDD. Workshop. Text. Min. **400**, 109–110 (2000)
3. Hegland, M.: The Apriori Algorithm CMA. Australian National University, Canberra (2005)
4. Han, J., Kamber, M.: Data Mining Concepts and Techniques. Morgan Kaufmann Publishers, Burlington (2006)
5. Singh, R.V., Bhatia, M.S.: (2011) Data clustering with modified k-means Algorithm. In: Proceedings of the IEEE, pp. 717–721, June 2011. ISBN: 978-1-4577-0588-5
6. Xu, R.: Survey of clustering algorithms. IEEE Trans. Neural Netw. **16**(3), 45–678 (2005)

7. Matloff N The Art of R Programming. (http://heather.cs.ucdavis.edu/~matloff/132/NSPpart.pdf)
8. Grolemund G Hands-on programming with R, Garrett (http://droppdf.com/v/KoQlw)
9. Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New Jersey (2005)
10. Velmurugan, T., Santhanam, T.: Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points. J. Comput. Sci. **6**(3), 363–368 (2010)
11. Patel, V.R., Mehta, R.G.: Clustering algorithms: A Comprehensive Survey. In: Proceedings of the International Conference on Electronics Information and Communication Systems Engineering, Jodhpur (2011)
12. Oyelade, O.J., Oladipupo, O.O., Obagbuwa, I.C.: Application of k-Means Clustering algorithm for prediction of Students' Academic Performance. Int. J. Comput. Sci. Inf. Secur. **7**, 1002 (2010)
13. Ng, R.T., Han, J.: CLARANS: a method for clustering objects for spatial data mining. IEEE Transact. Knowl. Data Eng. **14**(5), 1003–1016 (2002)
14. Sharmila, M.R.C.: Performance evaluation of clustering algorithms. Int. J. Eng. Trends Technol. **4**(7), 2231 (2013)
15. Revathi, NalinI, s: T performance comparison of various clustering algorithm. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **3**(2), 67–72 (2013)