ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.10

# Enhanced K-means Clustering Algorithm for Automobile Routing

V.Praveen[1], V.Hemalatha[2], M.Poovizhi[3]
Assistant Professor[1, 2, 3]
Department of CSE
N.S.N College of Engineering &Technology, Karur, Tamilnadu, India

**Abstract:**
A solution is designed for the vehicles to minimize the cost of distribution by which it can supply the goods to the customers with its known capacity can be named as a vehicle routing problem. In Clarke and Wrights saving method and Chopra and Meindl savings matrix method mainly an efficient vehicle routing can be achieved by calculating the distance matrix and savings matrix values based on the customers location or the path where the customer's resides. The main objectives of this paper are to reduce the total distance and the total number of vehicles which is used to deliver the goods to the customers. The proposed algorithm is based on k-means clustering algorithm technique which is used in the data mining scenario effectively. The proposed algorithm decreases the total distance and the number of vehicles assigning to each route. The important thing need to consider here is that, this new algorithm can enhance the Chopra & Meindl saving matrix method and Clarke and Wright saving matrix method.

**Keywords-**k-means clustering, centroid, cluster, saving matrix, vehicle routing problem

## I. INTRODUCTION

In general, there are many practical applications which can provide efficient distribution of goods to the customers. Goods in the sense it can be any home appliance products which are used daily. The vehicle scheduling problem was first formulated in the year 1959 [8] and in that set of customers with each of its known locations and known demand for any commodity, and that required goods can be delivered to the customer from a single depot by some calculated amount of delivery vehicles with some basic conditions and constraints are specified[2]:

(i) The demands of all customers are met
(ii) Each customer is served by exactly one vehicle
(iii) For each route the total demands must not exceed the capacity of the vehicle which is already defined.

From a depot different products must be distributed to several retailers. An efficient collection (or) distribution of goods keeps transport inventories low, it saves resources and energy. Therefore vehicle routing is one of the important topics for this kind of problems. The vehicle routing problem is a common name given to a whole class of problems involving the visiting of customers by using vehicles. These problems derive their name from the basic practical problem of supplying geographically dispersed customers with goods using a number of vehicles operating from a common goods depot (or) warehouse. For a classical vehicle routing problem, the best solution is to serve the goods to the customers exactly once by starting from and ending to the depot. The main objective is to reduce the overall transportation cost by satisfying all the constraints. The cost for transporting the goods can be reduced by minimizing the total distance travelled and as well as the total number of vehicles. While comparable to the classical vehicle routing problem, the majority of the real world problems are much more complex to solve. In general the classical vehicle routing problem is based on some constraints like the total vehicle capacity or some time interval to reach the customers. Vehicle routing problem is also known as vehicle scheduling or delivery problem. For example, Garbage must be collected from households and industries to a distracting place, so for transportation that we need an efficient route to travel from one place to another. It is very much useful for the daily transportation because it reduces the cost of forming the routes based on the capacity of the vehicle.

## II. LITERATURE SURVEY

The first article for the 'Truck dispatching problem' was published by Dantzig and Ramser who presented a larger truck dispatching problem, that is referred to as D&R problem and many more publications has been made which is completely relevant to this article after it was published[1]. The Clark and Wright algorithm [9] is one of the most popular heuristic algorithms in the vehicle routing problem. Cordeau et al. [6] present that parallel version is much better because merge yielding the largest saving is always implemented, but the sequential version keeps expanding the same route until there is no longer feasible route. Chopra and Meindl [3] provide a solution for transport planning, in that they present a routing and scheduling, transportation problem for a company in which they solve with a method called savings matrix method. It can be classified into four steps which are:

(1) Identify the distance matrix for the given location,
(2) Calculating the saving matrix using the distance matrix values,
(3) Assigns, customers to vehicles or routes, and
(4) Sequence the customers within the routes.

The first two steps are explained clearly. The third step is that assign the customers to vehicles and routes by, initially each customer are assigned to a separate route. If the two routes can provide a feasible solution by which it doesn't cross the limited capacity means it can be combined. The procedure is continued until no more feasible combinations are possible. For a transportation problem, Lumsden (2006) and Jonsson (2008) present a similar explanation but it is not clear. Rand [11] made an analysis and presents an article about the different saving methods for the vehicle routing problems. In that he argues about the parallel version, because it is not always better than the sequential version. Parallel version is a heuristic and there is no assurance from the obtained results that it produces the optimal solution or near optimal solution.

Mainly Chopra and Meindl savings matrix method [4] decreases the total length of the trip about 7% and even a better solution is obtained against this method by decreasing the total distance of about 3%. Chopra and Meindl description present a clearer idea about saving methods than the original Clarke and Wright algorithm (1964), even though so many wonder that how a vehicle routing is done effectively. The method presented here is widely used in the teaching section since 2008, because it provides an excellent result and easy to implement when compared to the Clarke and Wright saving method. An important advantage for saving methods in vehicle routing problem is its simplicity and robustness. So it's a best suited one for vehicle routing problem.

## III. NUMERIC EXAMPLE

The ultimate base for the vehicle routing problem is travelling salesman problem because the constraint "visit all the customers exactly once" used in the vehicle routing problem is as same as that of travelling salesman problem. Imagine that a delivery boy must visit some 'n' number of customers and returned to the starting point after visiting all the customers exactly once and the total cost for visiting all the customers is the major problem. Laporte [12] describes the Clark and Wright algorithm. The solution is to obtain a minimum cost route to visit all the customers exactly once. Suppose when the cost for travel from city x to city y is equals to the cost of city y to city x, then the problem is symmetric. If $C_{xy} \neq C_{yx}$ means then the problem is asymmetric. Starting from the central depot (or warehouse), goods or items delivered to the customers: 0-8. Initially the distance for each customer is presented in the Table I as locations and the demands for each customer are also given. According to the existing scenario both the Clarke& Wright savings method and Chopra & Meindl savings method uses the symmetric cost for returning to the depot, i.e. the distance from 1 to 5 is the same as the distance from 5 to 1.

**Table. I. Distance and demands for customers**

| Customer | Location | Demand |
|---|---|---|
| 1 | (22,22) | 18 |
| 2 | (36,26) | 26 |
| 3 | (21,45) | 11 |
| 4 | (45,35) | 30 |
| 5 | (55,20) | 21 |
| 6 | (55,45) | 16 |
| 7 | (26,59) | 29 |
| 8 | (55,65) | 37 |

Here the location for the depot (or warehouse) is (40, 40) (x-axis and y-axis values) and obviously the demand is zero.

**Table.2. Distance matrix calculation**

| $C_{ij}$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | - | 26 | 15 | 20 | 7 | 25 | 16 | 24 | 29 |
| 1 | | - | 15 | 23 | 26 | 33 | 40 | 38 | 54 |
| 2 | | | - | 24 | 13 | 20 | 27 | 35 | 43 |
| 3 | | | | - | 26 | 42 | 34 | 15 | 39 |
| 4 | | | | | - | 18 | 14 | 31 | 32 |
| 5 | | | | | | - | 25 | 49 | 45 |
| 6 | | | | | | | - | 32 | 20 |
| 7 | | | | | | | | - | 30 |

Based on the customer location, initially distance matrix is calculated using the Equation 1 and its cost are symmetric.

**Equation 1, Distance matrix formula**

$$D(c_i, k) = \sqrt{(x_{c_i} - x_k)^2 + (y_{c_i} - y_k)^2} \quad (1)$$

Equation 1 represents the distance between the customer $c_i$ and the depot k. The calculated distance matrix values are shown in the Table II. Based on the calculated distance matrix values, the cost savings is calculated using the Equation 2.

**Equation 2, Savings matrix formula,**

$$S(c_i, c_j) = D(k, c_i) + D(k, c_j) - D(c_i, c_j) \quad (2)$$

**Table.3. Solution for the problem with four routes**

| | Trip | Total Distance | Total Demands |
|---|---|---|---|
| Route1 | 0-1-2-5-0 | 86 | 65 |
| Route2 | 0-3-7-0 | 59 | 40 |
| Route3 | 0-4-0 | 14 | 30 |
| Route4 | 0-6-8-0 | 65 | 53 |

Once the values are found, it is arranged in a non-increasing order (from largest to smallest), and then the routes are combined one by one up to the total capacity is reached. If the newly added route exceeds the total capacity, then the new link is discarded and the previously formed links are undisturbed. Based on the total capacity defined earlier, the number of vehicles may found along with the feasible route for each vehicle. The maximum vehicle capacity defined in this example is 70. Table III shows that the result with the total distance of 224 and four vehicles is needed for that transportation. Finally the search procedure applies to the final routes, it reduces the total distance of 1% compared to the Clarke and Wright saving method but the total number of vehicles is increased by one. The search procedure method doesn't produce a better solution, but Chopra and Meindl savings matrix method produces a result of about 3% decrease in the total cost of the trip. This decrease is not enough when the vehicle is used daily for delivering the goods to the customers, so for these kind of methods are proposed to solve all kinds of vehicle routing problems.

## IV.PROPOSED METHOD

The first parameter is the total distance travelled to deliver the goods or some products to different distribution points. Because of routing problems, the main thing that all need to focus is on the total distance. If the total distance is reduced at some amount of range means definitely the total cost is reduced gradually in parallel. The second parameter is the total number of vehicles. This type of parameter will help for a large instance set of problems which are mainly related to real world applications. When the total distance is reduced to a somewhat minimum then the time consumption is reduced because the time consumption is calculated based on the total distance travelled to distribute the goods to the customers. In existing methods, initially the distance matrix values are calculated using the Equation 1 and using that value savings matrix values [10] are calculated to find the most efficient route for the vehicles which are going to deliver the goods to the customers and simultaneously the total load doesn't exceed the maximum vehicle capacity.

### *A. Enhanced K-means Clustering Algorithm (EKCA)*

The proposed methodology is based on k-means clustering algorithm method for solving the vehicle routing problem with multiple depots [5].

#### *A.1.1Flowchart for Enhanced K-means Clustering Algorithm*

The flowchart for the Enhanced K-means Clustering Algorithm (EKCA) is shown in the Figure 1.

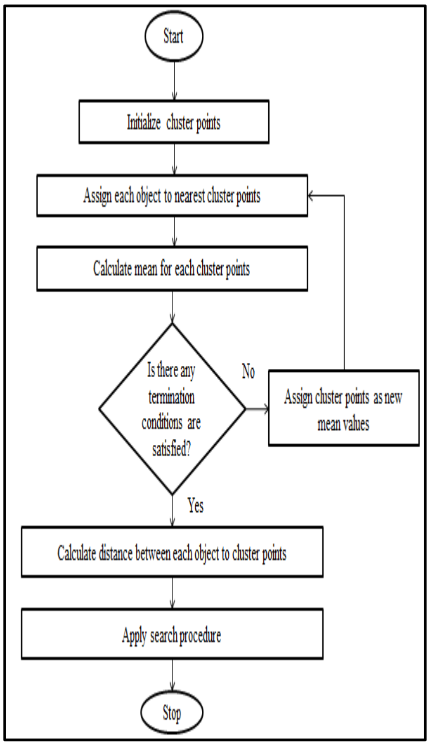


**Figure.1. Enhanced K-Means Clustering Algorithm**

#### *A.1.2 Numeric Example*

A set of 15 customers with their locations and demands are shown in the Table IV. Initially the customers are divided into two set of points. Each set of points refers to one cluster point. When the clustering is performed for a set of values, there may be two or more values can be formed. Based on the algorithm of k-means clustering, two centroid values are found with two sets of different customers.

**Table.4. Customers with location and demands**

| Customers | A(x-axis) | B (y-axis) | Demands |
|---|---|---|---|
| 1 | 40 | 40 | 12 |
| 2 | 36 | 26 | 21 |
| 3 | 21 | 45 | 25 |
| 4 | 45 | 35 | 15 |
| 5 | 10 | 10 | 16 |
| 6 | 55 | 45 | 24 |
| 7 | 26 | 59 | 12 |
| 8 | 55 | 15 | 17 |
| 9 | 40 | 30 | 20 |
| 10 | 20 | 14 | 25 |
| 11 | 44 | 11 | 15 |
| 12 | 64 | 11 | 12 |
| 13 | 19 | 24 | 16 |
| 14 | 33 | 26 | 21 |
| 15 | 15 | 12 | 13 |

**Table.5. Final centroid points with satisfied values**

| Cluster Points | Customers | (A,B) | Centroid Value 1(37,40) | Centroid Value 2(33,15) |
|---|---|---|---|---|
| 1 | 1 | (40,40) | 3 | 25.961 |
| 1 | 3 | (21,45) | 6.32 | 28.071 |
| 1 | 4 | (45,35) | 2 | 25.079 |
| 1 | 6 | (55,45) | 3.162 | 26.925 |
| 1 | 7 | (26,59) | 5 | 30.265 |
| 1 | 9 | (40,30) | 2.236 | 27.459 |
| 1 | 14 | (33,26) | 0 | 25.317 |
| 2 | 2 | (36,26) | 14.035 | 11.40 |
| 2 | 5 | (10,10) | 26.076 | 10.44 |
| 2 | 8 | (55,15) | 23.194 | 2.236 |
| 2 | 10 | (20,14) | 25 | 3.162 |
| 2 | 11 | (44,11) | 25.317 | 0 |
| 2 | 12 | (64,14) | 27.01 | 5.385 |
| 2 | 13 | (19,24) | 25.079 | 2 |
| 2 | 15 | (15,12) | 25.317 | 0 |

Each customer belongs to one centroid point, otherwise called as depot. Using the Euclidean distance formula, the distance between the customers to particular depots is calculated and finally the total distance and the total number of vehicles needed to perform efficient transportation is found. This method definitely provides a good solution compared to the previous proposals by different authors.

## V. COMPARISON RESULTS

The sample problems were solved by Net Beans IDE and the proposed Enhanced k-means clustering algorithm shows the better result compared to Clarke and Wright's saving method. The results for difference in total number of vehicles and total distance are shown in the Figure 3 and Figure 4 respectively.

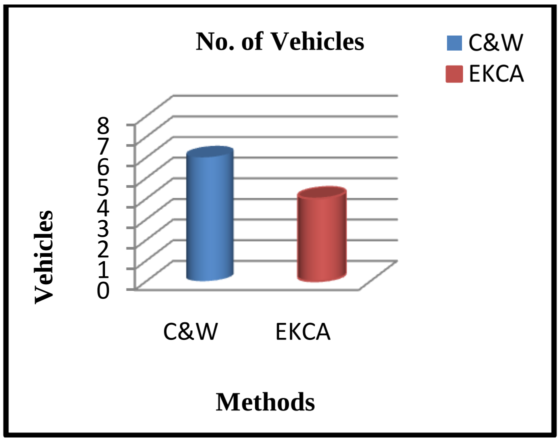


**Figure.3. Number of Vehicles**

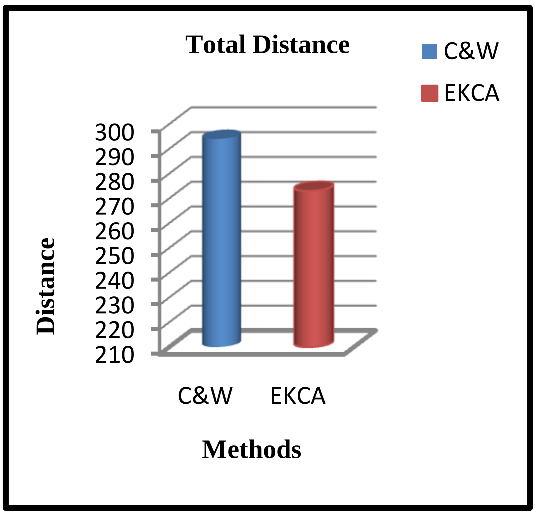


**Figure.4. Total Distance**

## VI. CONCLUSION

While calculating the savings matrix approach for the Clark and Wright method, it provides a good solution for the small instance set, but for a large instance set it doesn't yield a better result. The proposed method shows a better solution against the previously proposed methods and k-means clustering methods which are mainly used in the data mining concepts. Among all the existing methods, our enhanced k-means based clustering method can reduce the total number of vehicles of about 40% (small set of problems) and total distance of about 7% while using multi depots for delivering the products to customers.

## VII. REFERENCES


[1]. Anders Segerstedt, 2013 'A simple heuristic for vehicle routing – A variant of Clarke and Wright's saving method', Lulea University of Technology, Industrial logistics, SE-97187 Lulea, Sweden.

[2]. Behrouz Afshar-Nadjafi, Alireza Afshar-Nadjafi, 2014 'A constructive heuristic for time dependent multidepot vehicle routing problem with time-windows and heterogeneous fleet', Journal of King Saud university-Engineering sciences.

[3]. Chopra, S., Meindl, P., 2004. Supply Chain Management – Strategy, Planning and Operation, 2nd ed. Pearson Prentice hall, Upper Saddle River.

[4]. Chopra, S., Meindl, P., 2010. Supply Chain Management – Strategy, Planning and Operation, 2nd ed. Pearson Prentice hall, Upper Saddle River.

[5].Christian Prins, 'Two memetic algorithms for heterogeneous fleet vehicle routing problems', engineering applications of artificial intelligence22, pp. 916-928, 2009.

[6]. Cordeau, J-E, Gendreau, M., Laporte, G., Potvin, J-Y., Semet, F., 2002 A guide to vehicle routing heuristics. Journal of the Operational Research Society 53(5), 512-522.

[7]. Christofides, N., Mingozzi, A., Toth, P., 1979. The vehicle routing problem. In: Christofides, N., Mingozzi, A., Toth, P., Sand,C,(Eds.), Combinatorial Optimization, Wiley, Chichester, pp.315-338.

[8]. G.B.Dantzig, J.H.Ramser, "The truck dispatching problem", Management Science, Vol. 6, No. 1, pp.80-91, 1959.

[9].G.Clarke, J.Wright, "Scheduling of vehicles from a central depot to a number of delivery points", Operations Research, Vol. 12, No. 4, pp. 568-581, 1964.

[10]. Gaurav Sharma, Preeti Bansal, "Min-Min approach for scheduling in grid environment", International Journal of Latest Trends in Engineering and Technology(IJLTET), Vol 1, Issue 1, May 2012.

[11]. Graham K Rand, 'The life and times of savings method for vehicle routing problems', http://www.orssa.org.za, Volume 25(2), pp. 125-145, 2009.

[12]. Laporte, G, 'the vehicle routing problem: An overview of exact and approximate algorithms' European Journal of Operational Research 59(3), 345-358.