2009 Third International Symposium on Intelligent Information Technology Application

Balanced K-Means Algorithm for Partitioning Areas in Large-Scale Vehicle Routing Problem

Ruhan He
College Computer Science,
Wuhan University of Science and Engineering,
Wuhan, 430073, P. R. China.
E-mail: heruhan@gmail.com

Jiaxia Sun
School of Information Engineering,
Henan Institute of Science and Technology,
Xinxiang, 453003, Henan Province, P. R. China.

Weibin Xu
Hubei Provincial Tobacco Monopoly Bureau,
Wuhan, 430030, P. R. China.

Bingqiao Zu
Hubei Provincial Tobacco Monopoly Bureau,
Wuhan, 430030, P. R. China.

***Abstract*—We present a new and effective algorithm, balanced k-means, for partitioning areas in large-scale vehicle routing problem (VRP). The algorithm divides two-stage procedures. The traditional k-means is used to partition the whole customers into several areas in the first stage and a border adjustment algorithm aims to adjust the unbalanced areas to be balanced in the second stage. The objective of partitioning areas is to design a group of geographically closed customers with balanced number of customers. The presented algorithm is specifically designed for large-scale problems based on decomposition strategy. The computational experiments were carried out on a real dataset with 1882 customers. The results demonstrate that the suggested method is highly competitive, providing the balanced areas in real application.**

Keywords-Vehicle Routing Problem (VRP); K-Means; Partitioning Area

I. INTRODUCTION

The Vehicle Routing Problem (VRP) is one of the core operations research and combinatorial optimization problem classes with numerous applications. The basic VRP may be briefly described as follows. Given on or more depots, a fleet of vehicles, homogeneous or not, and a set of customers with known or forecast demands, find a set of closed routes, originating and, generally, ending at one of the depots, to service all customers at minimum cost, while satisfying vehicle and depot capacity constraints. Other characteristics and requirements may be considered, such as service and travel time restrictions, multiple commodities with different transportation requirements, time-dependent and uncertain demands or travel times, etc., yielding a rich set of VRP variants [1-5].

In the VRP, the objective is to construct a minimum cost set of routes serving all customers where the demand of each customer is not greater than the vehicle capacity and where each customer is visited exactly once. Vehicle routing problems have been the object of numerous studies and a very large number of papers propose solution methods. Because most VRP variants are NP-Hard, exact solution methods are confined to limited-size problem instances. Heuristics, meta-heuristics principally, are thus proposed in most cases. Contributions are continuously being made to VRP methodology and practice alike. Yet, despite the progress the field has seen in recent years, many challenges still stand and new ones are emerging.

We have faced a practical problem as follows: there are large scale customers in a big city along with its suburb to be delivered the cigarettes by a week (only five days worked). Therefore, the first step is to partition the city to be five areas, and we can server one area with one day. Obviously, the number of the final areas is determinate. K-means clustering algorithm is chosen for the partitioning naturally. But the distribution of the customer is extremely imbalance. The center of the city is very concentrated while the suburb is relative sparse. Therefore, the result of k-means clustering is extremely imbalance for each area.

For the large-scale VRP, good solutions are very hard to find and practically impossible to solve exactly. Meta-heuristics are used to face this kind of problems in order to find high quality solutions with a reasonable computational effort [6]. The practical solving limit of many algorithms is few hundreds of customers that have to be serviced with a fleet of few dozens of vehicles. So, real-life problem instances which can involve several thousands of customers clearly cross this borderline. One possible approach is to use decomposition strategies like the POPMUSIC framework [7] that try to overcome size restrictions. Decomposition strategies were recently also successfully applied to large scale real world problems [8].

For real large-scale VRP, we adopt a decomposition strategy to divide the big problem into several sub-problems. The first of sub-problem is to partition the total customers into five areas.

In this paper, based on decomposition strategy, we propose a new and effective algorithm, balanced k-means, for the partitioning areas in the large scale VRP. The algorithm is divided two-stage procedures. The k-means is used to partition the whole customers into several areas in the first stage and a border adjustment algorithm aims to adjust the unbalanced areas to be balanced in the second stage. The objective of partitioning areas is to design a group of geographically closed customers with balanced number of customers. The presented algorithm is specifically designed for large-scale problems.

The rest of the paper is organized as follows: The related work is given in Section 2. The proposed method is presented in Section 3. The experimental results are presented in Section 4. Finally, we conclude the paper.

II. RELATED WORK

There are literally hundreds of papers discussing vehicle routing problems and their variations [1-5]. Almost all papers focus on problems of relatively small size. Unfortunately, many of the proposed techniques do not scale well and some recent papers specifically address large-scale problems. Bouthillier et al [9] use parallel computation for scalability. Their main contribution is an architecture allowing different search strategies to run in parallel and to communicate their progress. To date, the most successful approach for solving large-scale VRPTWs is an advanced evolutionary technique [10] building upon the success of earlier algorithms (e.g., [11]). The main innovation in [10] are the incorporation of sophisticated diversification schemes (e.g., using guided local search) into an evolutionary framework.

For large-scale VRP, the recent work in[12] who use active guided evolution strategies as well as the Variable Neighbourhood Search (VNS) approach [13] shows that even large instances may be solved in an efficient way. The 2-phase hybrid meta-heuristic [14] also proved to successfully solve problems from small sizes up to 1000 customers. Another possible approach is to use decomposition strategies like the POPMUSIC framework [7] that try to overcome size restrictions, by intelligently splitting the problem into sub problems and solving them separately. Decomposition strategies were recently also successfully applied to large scale real world problems [8].

There are many clustering algorithms. One commonly-used clustering algorithms is k-means algorithm [15]. The K-means clustering algorithm has time complexity of $O(RKN)$ where K is the number of desired clusters and R is the number of iterations needed to complete the clustering.

In this paper we present a balanced k-means algorithm with border adjustment that keeps the balance of each cluster for large-scale VRP by decomposition strategies.

III. THE PROPOSED APPROACH

In this section, a balanced k-means clustering algorithm for partitioning areas in large VRP is presented.

The objective of the proposed algorithm is to obtain the balanced areas. Each area has almost the same customers. The difference of any two areas is under a threshold T.

The distance between a point and an area (or cluster) centroid is Euclidean distance, which is shown in (1):

$$d(p,C)=\sqrt{(p_x-c_x)^2+(p_y-c_y)^2} \quad (1)$$

where p is a data point and C represents a area.

Based on (1), we can get the difference between a point in one area and another area:

$$diff(p,B)=d(p,B)-d(p,A) \quad (2)$$

where p belong to area A, and A, B represent any two different areas.

According to the data point with minimum *diff(p, B)* is between the border of areas A and B, we can adjust the data points on the border. If area A has two many data points, then we should move some points to area B.

Therefore, the balanced k-means algorithm is described as Table 1. The proposed algorithm can be divided three parts. The first part is to call the standard k-means algorithm to get the initial clustering results. Then, we can evaluate the balance for each cluster. If it is balanced, then we need not to do the next steps. Otherwise, we continue the second part, i.e. planning an adjustment among areas. Finally, we perform the third part, i.e. call the border adjustment function. We repeat the second part and third part until getting the balanced results.

The problem with k-means algorithm is that the user must know what k value to use initially. The need of practical application resolves this problem, which is also one of the reasons that we choose k-means algorithm.

TABLE I. THE BALANCED K-MEANS ALGORITHM

```
/* The main procedure */
Function Main()
   Load the data;
   Set the value of k and the initial centroid of clusters;
   Call K-Means(X, maxiter, T, K, nc);
   Get the clustering results of k-menas;
   While max(cluster)-min(cluster) > T
        Plan the steps of adjustment among clusters;
        Call borderadjust(m, A, B);
   End while
   Get the balanced clusters.
End function
/* The sub-procedure: Standard K-Means function. */
Function K-Means(X, maxiter, T, K, nc)
   While iter < maxiter
        For each point p in X
          Calculate the distance of p and each cluster C
          based on (1);
          The p point belong to the minimum distance cluster;
        End for
        Re-compute the cluster centroids nc;
        Iter = iter + 1;
        If max(cluster) – min(cluster) < T
             Break;
        End if
   End while
End function
/* The sub-procedure: The border adjustment function*/
Function borderadjust(m, A, B)
     For each point p in area A
          Calculate diff(p, B) based on (2);
     End for
     Sort all the diff(p, B) ascending;
      Move the first m point in area A based on sorted
     diff(p, B) to area B;
End function
```

IV. EXPERIMENTS

An experiment has been conducted on a real data set, which is the customers (the cigarette tradesmen) in Xiaogan city, Hubei province, P. R. China. There are 1882 customer distribute the centre and suburb of the city, which can be seen in Fig. 1. We set $K=5$ for the matter of fact. The delivery is one day for one area. Then, one week (five worked days) can sever the whole customer.

Fig. 2 is the clustering results for standard k-means algorithm. The imbalance among five areas is obvious. Even we can alleviate the degree of imbalance by adjusting the initial cluster centroids somewhat. However, the centre of the city is dense while the suburb is sparse. Therefore, the k-means clustering results can be balanced. The Fig. 2 shows the imbalance. The blue area has too many customers. The detailed number of customers for each customer can be seen in the adjust step 0 in Table 2. The blue area has over half of all customers (1124 customers) and the max difference between areas is 1006.

The result in Fig. 2 is not satisfactory obviously. Therefore, we adjust the data point on the area border. We can plan an adjustment strategy based on Fig. 2. We move 300 data points from blue area to yellow area firstly, which can be seen in Fig. 3. Then, we move 100 data points from yellow area to green area, as shown in Fig. 4. There are 460 data points moved from blue area to black area as shown in Fig. 5. Finally, we move 250 data points from black area to red area, which can be seen in Fig. 6. The changes of the number of data points for each area can be seen in Table 2.

The max difference among areas in Fig. 6 is only 18. The five areas are relatively balanced on customer number in Fig. 6, which demonstrates the effectiveness of the border adjustment algorithm.

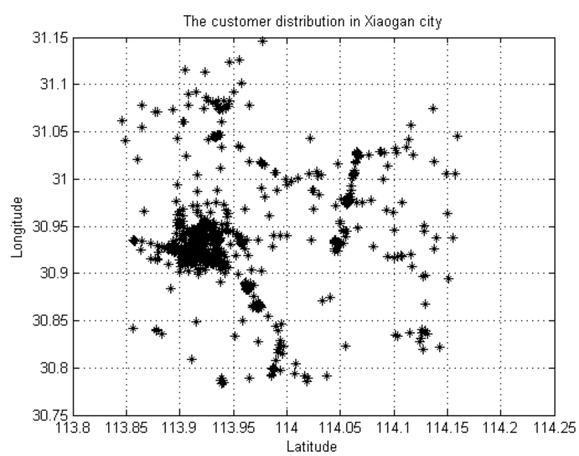

Figure 1. The customer distribution in Xiaogan city.

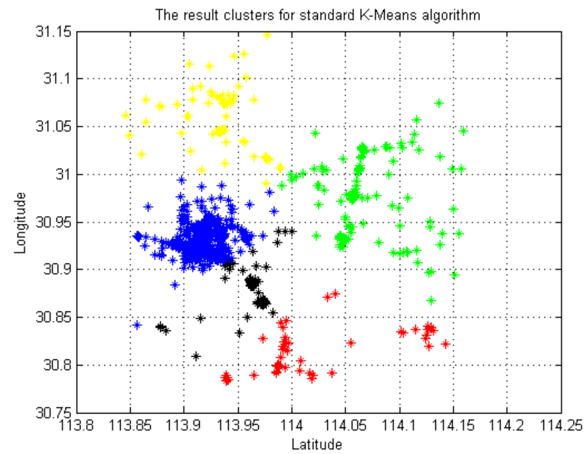

Figure 2. The results for K-Means clustering algorithm.

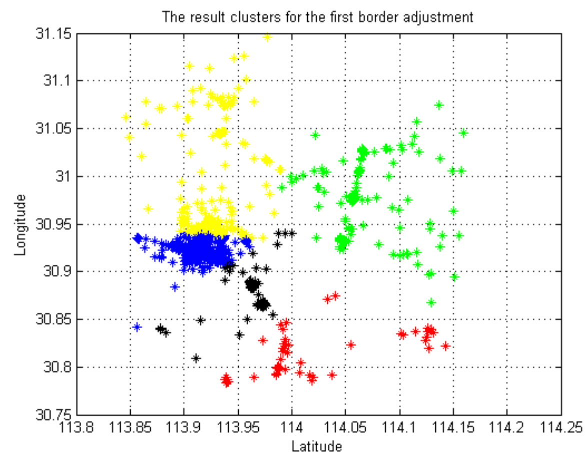

Figure 3. The first border adjustment (from blue area to yellow area).

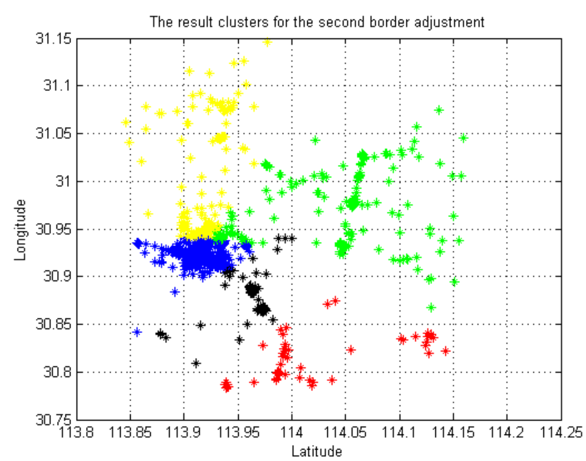

Figure 4. The second border adjustment (from yellow area to green area).

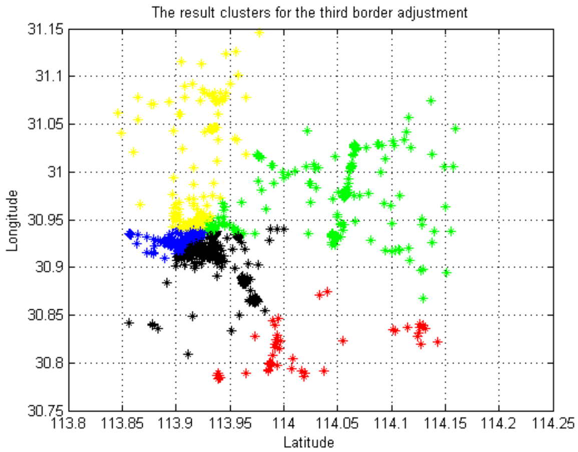

Figure 5. The third border adjustment (from blue area to black area).

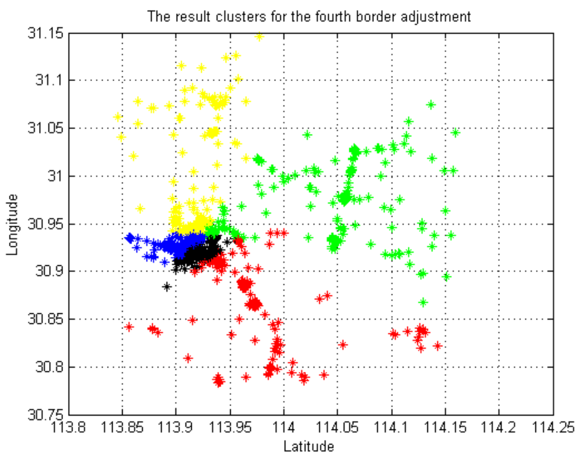

Figure 6. The fourth border adjustment (from black area to red area).

TABLE II. THE NUMBER OF CUSTOMER FOR EACH AREA (CLUSTER)

Adjust step	Red area	Blue area	Green area	Black area	Yellow area	Max diff
0	118	1124	268	160	152	1006
1	118	824	268	160	452	706
2	118	824	368	160	352	706
3	118	364	368	620	352	502
4	368	364	368	370	352	18

V. CONCLUSION AND FUTURE WORK

For real large-scale VRP, we adopt a decomposition strategy to divide the big problem into several sub-problems. The first of sub-problem is to partition the total customers into five areas. Therefore, a new and effective algorithm, balanced k-means, for the partitioning areas in the large scale vehicle routing problem is proposed in this paper. The algorithm divides two-stage procedures. The traditional k-means is used to partition the whole customers into several areas in the first stage and a border adjustment algorithm aims to adjust the unbalanced areas to be balanced in the second stage. The objective of partitioning areas is to design a group of geographically closed customers with balanced capacity and the number of customers. The presented algorithm is specifically designed for large-scale problems.

Future work includes considering multiple constrains for balance in the algorithm, such as customer demand, and performing extensive experiments with larger dataset..

ACKNOWLEDGMENT

This work is supported by Hubei Provincial Department of Education (Grant No: Q20091704).

REFERENCES

[1] J.-F. Cordeau, G. Desaulniers, J. Desrosiers, M. Solomon, and F. Soumis, "The VRPTW. The Vehicle Routing Problem.," SIAM Monographs on Discrete Mathematics and Applications, pp. 157-194., 2001.

[2] O. Braysy and M. Gendreau, "VRPTW Part I: Route Construction and Local Search Algorithms," Transportation Science, vol. 39, pp. 104-118, 2005.

[3] O. Braysy and M. Gendreau, "VRPTW Part II: Metaheuristics," Transportation Science, vol. 39, pp. 119-139, 2005.

[4] L. C. YEUN, W. R. ISMAIL, K. OMAR, and M. ZIROUR, "VEHICLE ROUTING PROBLEM: MODELS AND SOLUTIONS," Journal of Quality Measurement and Analysis, vol. 4, pp. 205-218, 2008.

[5] J.-Y. Potvin, "A Review of Bio-Inspired Algorithms for Vehicle Routing," Technique Report: CIRRELT-2008-30, 2008.

[6] L. Bianchi, M. Birattari, M. Chiarandini, et al, "Hybrid Metaheuristics for the Vehicle Routing Problem with Stochastic Demands," Dalle Molle Institute for Artificial Intelligence, Galleria 2, 6928 Manno, Switzerland Technical Report No. IDSIA-06-05, 2005.

[7] A. Ostertag, K. F. Doerner, R. F. Hartl, et al, "POPMUSIC for a Real World Large Scale Vehicle Routing Problem with Time Windows," Journal of the Operational Research Society, 2008.

[8] T. Flaberg, G. Hasle, O. Kloster, and A. Riise, "Towards solving hugescale vehicle routing problems for household type applications," presented at Network Optimization Workshop, Saint-Remy de Provence, France, 2006.

[9] A. L. Bouthillier, T. Crainic, and P. Kropf, "A Guided Cooperative Search for the VRPTW," IEEE Intelligent Systems, vol. 20, pp. 36-42, 2005.

[10] D. Mester and O. Braysy, "Active Guided Evolution Strategies for Large Scale VRPTWs," C&OR, vol. 32, pp. 1593-1614, 2005.

[11] O. Braysy, W. Dullaert, and M. Gendreau, "Evolutionary Algorithms for the VRPTW," Journal of Heuristics, vol. 20, pp. 587-611, 2004.

[12] D. Mester and O. Braysy, "Active-guided evolution strategies for largescale capacitated vehicle routing problems," Computers & Operations Research, vol. 34, pp. 2964-2975, 2007.

[13] J. Kytojoki, T. Nuortio, O. Braysy, and M. Gendreau, "An efficient variable neighborhood search heuristic for very large scale vehicle routing problems," Computers & Operations Research, vol. 34, pp. 2743-2757, 2007.

[14] J. Homberger and H. Gehring, "A two-phase hybrid metaheuristic for the vehicle routing problem with time windows," European Journal of Operational Research, vol. 162, pp. 220-238, 2005.

[15] P. S. Bradley and U. M. Fayyad, "Refining initial points for K-means clustering," presented at the Fifteenth International Conference on Machine Learning (ICML98), 1998.