ISSN : 0974 - 7435 Volume 7 Issue 10

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 7(10), 2013 [392-397]

WK-means and branch-bound method based for cloud logistics scheduling

Xun-Yi Ren*, Da-Rong Yuan

College of Computer, Nanjing University of Posts and Telecommunications, (CHINA)

ABSTRACT

The efficient and accurate logistics scheduling problem has become the bottleneck that impedes the e-commerce development of China. Cloud-based logistics can achieve resource sharing and centralized logistics scheduling, which is expected to fundamentally solve the problems encountered in logistics scheduling. However, the current research about cloud logistics scheduling is only the beginning. Most methods based on exact algorithms and heuristic scheduling algorithms are time-consuming and inefficient while efficient scheduling algorithm is relatively scarce. This article takes cloud logistics scheduling problem as a NP-hard problem with multi-constraint and multi-objective decision making and establishes a multi-objective optimization cloud logistics scheduling model. K-means algorithm is used to cluster large and complex distribution network, but due to the load balancing problem in practical application, we use WK-means cluster which take the weight as an external constraints to balance the workload between each cluster. Large-scale VRP problem will eventually be divided into point-to-point TSP problem which we can use the branch-bound to solve and optimize. Simulation results show that the proposed scheme is more accurate and efficient than the existing typical heuristic scheduling method.

KEYWORDS

Cloud logistics scheduling;
WK-means;
Branch-bound algorithm;
Load balancing,
VRP.

INTRODUCTION

With the development of global e-commerce and population growth of online shopping, many issues of logistics scheduling such as information, resources, scheduling and transportation, have become increasingly prominent. These problems have affected the online shopping environment, e-commerce enterprises, and even the National economic development process[1].

Traditional logistics service platform often needs a lot of manpower and material resources of many types of businesses that are involved in the logistics industry chain, and because of the information isolation of each enterprise, it's difficult to seamlessly share the data and resources, resulting in the overall logistics industry chain service inefficiency. With the emergence of cloud computing, logistics service platform based on cloud computing[2] makes the resource sharing and information service level revolutionary change, powerful promotes the development of the e-commerce logistics industry, and maybe fundamentally solves the problems of various roles that are faced with during production, transportation and consumption in logistics industry chain. Therefore, all kinds of technical problems about cloud logistics have become the research focus in the industry and academia in recent years.

In many key technologies needed to be researched in cloud logistics, cloud logistics scheduling problem is one of the most core technology. Cloud logistics technology has brought new opportunities to many industries, especially logistics enterprises. But how to accurately deliver goods to every customer in the shortest time, it needs a scientific and effective logistics scheduling algorithm. Logistics scheduling is a NP-hard problem with multi-constraint and multi-objective decision making in nature. Logistics scheduling based on cloud computing is a large-scale VRP problem[3] dealing with a lot of nodes at the same time. Currently to deal with such problem, there are Exact algorithm, Classical heuristics and Meta-heuristics algorithm. The time complexity of these algorithms is mostly O (n^2). According to it, the VRP problem of 10000 nodes is expected to spend at least 100 hours. So cloud logistics scheduling problem needs to find the appropriate algorithm.

In this paper, we adopt the divide-and-conquer method. K-means algorithm is used to cluster large and complex distribution network, but due to the load balancing problem in practical application, we use WK-means cluster which take the weight as an external constraints to balance the workload betweeneach cluster. After clustering, large-scale VRP problem will eventually be divided into point-to-point TSP problem which we can use the branch-bound to solve and optimize. In order to verify the validity of the algorithm, we use real lattice data download from TSPLIB site to simulate test. While considering different nodes distribution in different regions, it will simulate two kinds of stochastic models which are the uniform distribution and Gaussian distribution. The location of nodes in the uniform distribution model is randomly generated, used to simulate normal scheduling problem. In Gaussian distribution model, the location of nodes is concentrated around one center point. It is used to simulate scheduling problems that customer nodes gather in several cities. Simulation results show that the proposed scheme is more accurate and efficient than the existing typical heuristic scheduling method.

CLOUD LOGISTICS SCHEDULING MODEL

In the background of cloud environment, we define the large-scale cloud logistics scheduling model: vehicles start from a warehouse supply point to deliver goods to n nodes, the goal of optimization is to minimize the value of delivery time and unmet cargo quantity so as to establish the model.

In order to describe the optimization model, we define the symbols below: D represents a set of customer demand nodes, which contains the coordinates of each node, the distance between two nodes, and each customer's order quantity.

W_k represents order quantity for note k; W_{total} represents total order quantity; m represents vehicle number for transport; $W_{\max}$ represents maximum cargo capacity for each vehicle; U_k represents unmet order quantity for node k; T_k represents the delivery time to k node; $X_{ijk} = \begin{cases} 1, (i, j) \text{ is traversed by vehicle } k \\ 0, \text{otherwise} \end{cases}$

So we can define the VRP model of cloud logistics scheduling as follows:

$$Min(\sum_{k \in D} U_k + \lambda \sum_{k \in D} T_k)$$

S.T.ÿ

1) For load balancing, we need the volume of transport satisfy $W_{\max} = \frac{W_{total}}{m}$ as much as possible;
2) Each vehicle can't exceed the maximum cargo capacityÿ $\sum_{k \in D} W_k < W_{\max}$;
3) Each node can only be accessed one time $\sum_{j \in D} X_{ij} = 1(\forall i \in D)$ $\sum_{i \in D} X_{ji} = 1(\forall j \in D)$;
4) After dischargeÿthe vehicle must leave, can't stay in the node, $\sum_{j \in D} X_{ij} = \sum_{j \in D} X_{ji} (\forall i \in D)$;

FULL PAPER

Generally described as:

$$\begin{cases} Min(\sum_{k\in D} U_k + \lambda \sum_{k\in D} T_k) \\ S.T.: W_{max} = \frac{W_{total}}{m} \\ \sum_{k\in D} W_k < W_{max} \\ \sum_{j\in D} X_{ij} = 1(\forall i \in D) \\ \sum_{i\in D} X_{ji} = 1(\forall j \in D) \\ \sum_{j\in D} X_{ij} = \sum_{j\in D} X_{ji}(\forall i \in D) \\ X_{ij} = \{0,1\}, U_k \geq 0, T_k \geq 0 \\ W_k \geq 0, W_{max} \geq 0, W_{total} \geq 0 \end{cases}$$

ALGORITHM DESCRIPTION

The proposed algorithm is divided into two stages: Using weighted K-means algorithm[4,5] divides customer nodes into k classes on the basis of the geographical location and delivery cargo quantity in the first phase. The second phase is assigning task to k vehicles according to the clustering results, aiming at this kind of point to point TSP problem[6], we use the branch-bound algorithm to optimize the path of a transport vehicle.

WK-means

After dealing with the clustering of large-scale nodes, we hope to achieve the following results:

1) There is no overlap and intersection between each divided area;
2) Nodes in one area relatively concentrated, so that each vehicle will save a lot of transportation time;
3) Every vehicle's freight amount is roughly same, should not have been a vehicle carrying supplies special phenomenon of more or less.

Input a collection D containing n customer nodes, each node in which, we assume as $d_1, d_2, d_3 \ldots\ldots d_n$, the total volume is $W_n = \sum_{i=1}^{n} weight(d_i)$, if we assign the goods to k transportation vehicles, so each vehicle need to transport the goods amount for $W_{ave} = \frac{W_n}{k}$ as much as possible.

We need to output k sets $D_1, D_2, D_3 \ldots\ldots D_k$, the conditions must be satisfied are:

1 $D_i \cap D_j = \varnothing$ ($i \neq j$ and $1 \leq i \leq k$ and $1 \leq j \leq k$)

2 For collection $D_i (1 \leq i \leq k)$, define $W_i = \sum weight(d_j)$, $d_j \in D$ ÿrepresenting a cluster of goods quantity which is also a vehicle load.

We hope $\Delta W = \sqrt{\sum_{i=1}^{k} (W_i - W_{ave})^2}$ get the minimumÿthe more small ΔW is, the more balanced the vehicle load is.

Algorithm steps:

1) First of all, according to the normal K-means clustering method, we choose k customer nodesd_1, d_2, d_3, ……, d_k and form the initial cluster D_1, D_2, D_3, ……D_K;
2) For the lattice collection composed of k two-dimensional weighted nodes, based on the weight w and coordinates (x, y), we calculate the weighted center of mass and total weight of each initial cluster:
$$x_k = \frac{\sum_i x_i y_i}{\sum_i x_i}, \quad y_k = \frac{\sum_i y_i w_i}{\sum_i y_i},$$
$i \in D'_k$; Total weight: $w_k = \sum_i w_i$;
3) Weighted classification is the key idea of the clustering method. Assuming that we no longer take the ratio of 1:1 when comparing distance, but adopt the measure method based on weight. For each node, calc the cluster:$D'_k : d_{ij} = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{w_i^{\gamma}}$, γ is called weight attenuation coefficient. When = 0, it's just the partition distance formula of normal K-means cluster. Through this formula we can get k weighted distances d_{i1}, d_{i2}, d_{i3}……d_{ik}, select the smallest distance d_{if}, and then corresponding node d_i will be added to cluster D'_f, after clustering all nodes in the set D again, we get a new clustering D''_1, D''_2, D''_3……D''_K;
4) Recalculating the weighted center of mass and total weight of each new clustering according to the step 2;
5) Repeat step 3 and 4, until the clustering convergence, and output the final clustering results D_1, D_2, D_3……D_K

Branch-bound algorithm

According to the clustering results, assigning task to k s, every vehicle need to traverse all nodes in a cluster, in which there are multiple paths to choose, and how to find out a route having the minimum time and lowest costs, it can be boiled down to solving the TSP (Travelling Salesman Problem). Formalized description for TSP problem: given a weighted directed graph G (V, E), each directed edge (u, v) E, there is a weighted value C<u, v>. We need to find a shortest loop which accesses each node once and only once, namely the shortest Hamilton loop[6].

To remedy the above problem, we adopt branch-bound algorithm, pictorial instructions:

Figure 1 shows a weighted graph G = (V, E, w), $V=\{v_i | i=1,2,3\ldots\ldots n\}$ is vertex in the picture which can

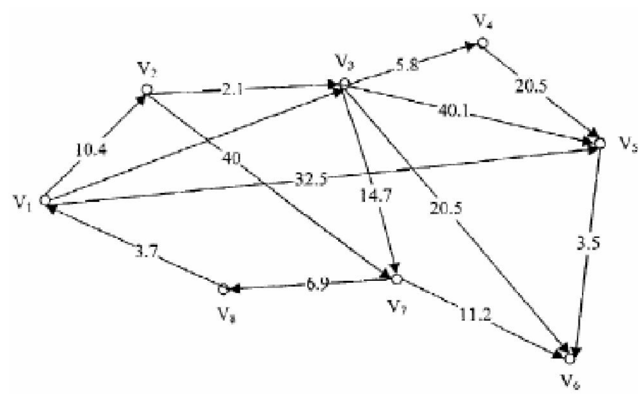

Figure 1 : Directed graph path selection problem of eight vertices

be seen as nodes in each cluster; $E=\{<v_i,v_j> | ,v_i, v_j , i, j=1,2,3\ldots\ldots n \}$ is the edge collection; w is the weight of edge, or point-to-point cost function. To find the optimal path from vertex v_i to v_j under the constraint condition of the shortest path and the least time, we convert graph into adjacency matrix which can be path, vehicle speed and transportation cost. We define S_{ij} as speed from vertex i to j, D_{ij} as distance from vertex i to j, as transport unit price from vertex i to j, W as the cost that customer can accept.

Assumptions

1) Every vehicle meets all the goods of each node;
2) Allowing delivery time delay;
3) Vehicle can only stay once at each node;
4) Each Vehicle is responsible for its own clustering task only, not allow across regions.

Upper and lower bounds function:

Lower bound: $\min F = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{D_{ij}}{S_{ij}}$;

Upper bound: $\max G = W - \sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij} D_{ij}$;

$i、j \in V，S_{ij}、D_{ij}、P_{ij} \geq 0$;

According to the upper and lower bounds function, we can find out the optimal path. When select the extensible child nodes from head node, we insert nodes which consume the least time into queue head, followed by nodes cost the least transportation, and make such nodes live, traverse in sequence until we get the optimal path.

Algorithm steps

1) To represent all the nodes of each clustering in the form of adjacency matrix;
2) Looking for the child nodes from head node, and insert head node into the heap;
3) To determine whether the heap is empty, if empty it returns transport route; if not null, based on the upper and lower bounds function, we insert nodes which consume the least time into the minimum heap, insert nodes which cost the least transportation into max heap, and remove child nodes which do not conform to the constraints;
4) Looking for the target node, if found the program terminates and return route. Otherwise it turn back to step 3.

SIMULATION RESULTS

In order to verify the validity of the algorithm, considering different nodes distribution in different cities, we will simulate two kinds of stochastic models, the uniform distribution and Gaussian distribution. And we compare the simulation results with original K-means method and Two-level Genetic Algorithm TGAC[7] All of them are simulated on the MATLAB language in 2 CPU, 1.86 GHz 2 GB environment. Nodes distribution data are derived from TSP testing case of TSPLIB site.

As shown in figure 2(a), the location of nodes in the uniform distribution model is randomly generated, used to simulate normal scheduling problem. In Gaussian distribution model, the location of nodes is concentrated around one center point, as shown in figure 2(b), it is used to simulate scheduling problems that customer nodes gather in several cities.

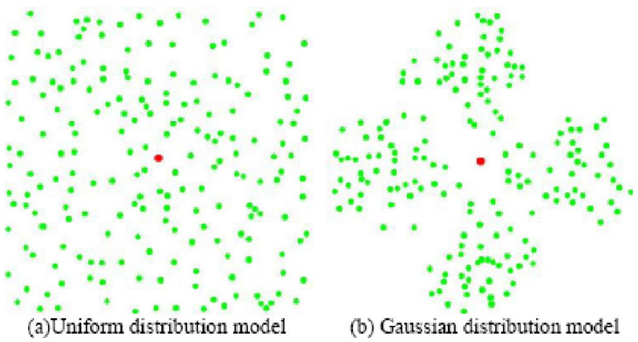

Figure 2 : Customer nodes distribution model

Compared with normal K-means cluster

Figure 3 shows the simulation results of WK-means and normal K-means based on uniform distribution model. From the picture, for uniform distribution model, using the WK-means clustering method is superior to K-means cluster. Without considering the load balance, the normal K-means method is easy to cause uneven classification that a cluster contains more nodes, another is less. Thus it will lead to the on-board resources squandered. This is clearly not what we want.

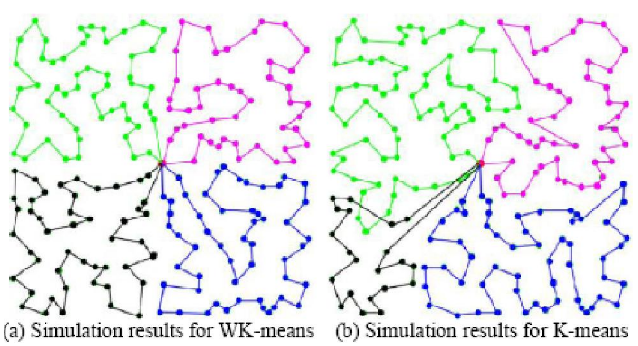

Figure 3 : Simulation results based on uniform distribution model

Figure 4 shows simulation results based on Gaussian distribution model. Due to the node distribution is relatively concentrated, so the clustering results of two methods are basically identical, finally the optimal paths are almost the same.

Compared with TGAC algorithm

Figure 5 and figure 6 show the simulation results compared with TGAC algorithm[7] based on uniform distribution model and Gaussian distribution model. It can be seen from figure, the optimal path graph calculated by TGAC algorithm is chaos, crossed and uneven, many adjacent nodes are assigned in different transport tasks, especially in large-scale logistics scheduling based on cloud computing, this disadvantage will be more obvious. TGAC algorithm can't meet our optimization goal, distribution model.

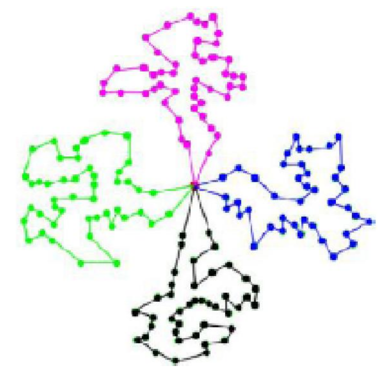

Figure 4 : Simulation results based on Gaussian distribution model

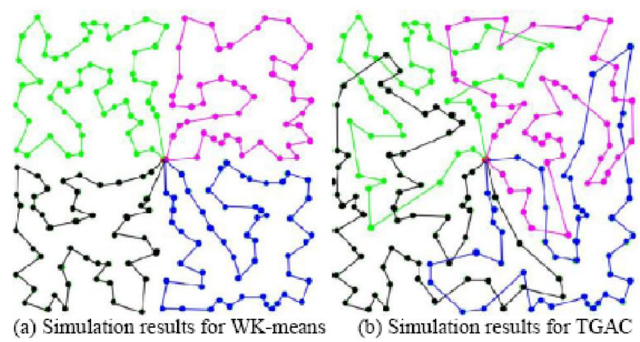

Figure 5 : Simulation results based on uniform distribution model

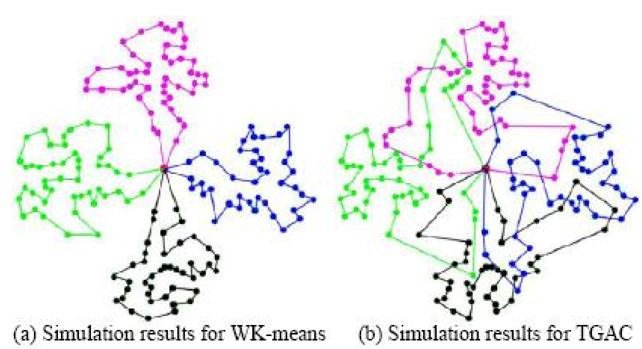

Figure 6 : Simulation results based on gaussian

CONCLUSION

For large-scale logistics scheduling problems, divide-and-conquer method can handle but low precision, while branch-bound method can get exact solution but high time complexity. Based on this, this paper designs an effective large-scale logistics scheduling algorithm based on clustering and branch-bound method. Since the original K-means clustering method does not take into account the load balancing problem in practical application, this paper adopt a kind of WK-means cluster with weight as an external constraints. According to the actual situation in logistics, each customer is assigned to a respective weight, and when clustering, the weight of each customer is seen as a constraint condition, so that the weight of each region is roughly equal. Finally large-scale logistics problem is effectively divided to the optimal scale for using branch-bound method. The solution lowers the time complexity while significantly improves the precision of dealing with large-scale logistics scheduling problems. It overcomes the disadvantage of time complexity and precision that general algorithms can't meet at the same time. Finally the simulation results prove that the algorithm has obvious advantages compared with other algorithms, especially the more nodes, the more obvious advantages.

ACKNOWLEDGEMENTS

This work is supported by the National Grand Fundamental Research 973 Program of China under (2011CB302903), National Natural Science Foundation of China (61073188, 61073115), Key University Science Research Project of Jiangsu Province (11KJA520002), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (yx002001).

REFERENCES

[1] Chen Tianming, Xu Qun; Break through the logistics bottlenecks under cloud computing. Economic Outlook the Bohai Sea., **10**, 57-59 **(2012)**.

[2] Wang Xun, Li Wenfeng, Zhong Ye, ZHAO Wei; Research on cloud logistics-based one-stop service platform for logistics center. Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on. IEEE., 558-563 **(2012)**.

[3] P.Toth, D.Vigo; The vehicle routing problem. Philadelphia: SIAM Publishing, **(2002)**.

[4] Joshua, Huang Zhexue, K.N.Michaelg et al.; Automated variable weighting in k-means type clustering. Pattern Analysis and Machine Intelligence, IEEE Transactions on., **05**, 657-667 **(2008)**.

[5] Yuan Fuyong, Zhang Xiaocai, Luo Sibiao; Accurate property weighted K-means clustering algorithm based on information entropy. Computer Applications., **06**, 1675-1677 **(2011)**.

[6] Chen Tao, Zhang Sifa; Branch and Bound Method to Solve Realistic TSP. Computer Engineering and Design., **10**, 2431-2433 **(2009)**.

[7] Ding Chao, Cheng Ye, HE Miao; Two-level Genetic Algorithm for Clustered Traveling Salesman Problem with Application in Large-scale TSPs. TsingHua Science And Technology., **20**, 459-465 **(2007)**.