Vehicle Scheduling Model Based on Data Mining

Guohua Zhang[1], Ting Xie[1], Min Liu[1] & Yang Liu[1]

[1] Hunan University of Technology, China

Correspondence: Guohua Zhang, Hunan University of Technology, Zhuzhou, Hunan, 412000, China. E-mail: 9884188@qq.com

Abstract

The article presents a shortest-path model of vehicle scheduling, which based on analyzing the application of data mining in vehicle scheduling model by referring research status of data mining and describing logistics distribution process. The article also provides an algorithmic support by making the Dijkstra algorithm of the shortest path model simple and rational.

Keywords: data mining, vehicle scheduling, Dijkstra algorithm

1. Introduction

In recent years, with the deep development of economic globalization and world market integration, logistics has become an important part of business operations. As a hole in corporate "third profit source", vehicles scheduling in distribution chain has caught great attention both at home and abroad and developed rapidly. Meanwhile, as the development of science and technology, the data generated in the modern informational logistics network system shows an explosive growth. It has become a major issue that how to extract business value and effective information according to the inducing and analyzing of collected data.

Data mining solves this problem effectively. Data mining is an advanced process that extracts potential useful information, models and trends from large databases or data warehouses and find authentic and intelligible model, which is a combination of Statistics, Database Technology and Artificial Intelligence Techniques[1]. Data mining discovers unknown, valid and practical information, which is used to gain business value. The more unexpected, the more valuable the information is.

2. Data Mining and Vehicle Scheduling Research Review

2.1 Research Status of Data Mining

As a relatively new interdisciplinary, Data mining has motivational influence with the advent of massive data. By discovering useful new rules and concepts, data mining enhances data owners' abilities to understand and apply large volumes of raw data [2]. Data mining is KDD concept proposed by the 11th Joint International Conference on Artificial Intelligence. The KDD International Conference had been successfully held for 22 sessions so far, which is a platform for the display of results and academic exchanges. It is no wonder that the current research focus on data mining has been from theoretical method to practical application. The current data mining mainly includes attribute induction, sequence pattern analysis, cluster analysis, classification prediction and correlation analysis.

Although the research on using data mining in the process of logistics, distribution and vehicle scheduling is premature, it has already gained some accomplishments. The data mining algorithm can be used to optimize paths and link process in distribution in order to maximize efficiency, which based on analyzing aspects including safety monitoring and management in distribution, energy saving and emission reduction, vehicle mileage calculation and operation effect evaluation, comprehensive performance big data evaluation index, real-time fault monitoring and protection, and operation enterprise data value.

2.2 Application of Data Mining in Vehicle Scheduling

Vehicle scheduling is the central part of logistics distribution process. The logistics distribution is a logistics activity on account of customer demand, which is delivering goods to appointed places after a series of activities, such as picking, processing, packaging, dividing and assembling.

To improve the service level and reduce the delivery cost at the same time, the routes of delivery vehicles is the first problem need to be solved. Both Vehicle utilization and distribution route optimization should be considered, which means businesses should combine customer demand with the transportation route when choosing distribution strategies.

Event analysis proves that the data mining is well applied in the following aspects of vehicle scheduling:

1)Facility and equipment intellectualization: monitoring the overall status of vehicles and delivery routes through intelligent driving. Collected data analysis can provide effective data support for future decisions.

2)Operational Management datafication: It ensures data interconnection and status monitoring can be implemented in real time and provides enterprises with more reliable data services, optimize distribution programs and efficient delivery process, so that operational management can be fully evidenced.

3)Business Resource Sharing: Under the background of sharing economy, data sharing should have complementary advantages. Data should be classified for overall analysis, in order to dig out more useful information and maximize the value of means of production.

3. Vehicle scheduling model analysis

3.1 Vehicle Scheduling Problem Description

In reasonable economic premise, a distribution center provides logistics distribution services for demand of customers in a certain area. Demand for each demand point is less than the vehicle capacity. The center has enough capacity for delivering to multiple demand points. The demand and distance of each demand point and the load of each car both are known. It requires that the scheduling of vehicles should be arranged reasonably to meet the delivery tasks and minimize total distance so as to improve vehicle utilization.

3.2 Vehicle Scheduling Mathematical Model

Study of vehicle scheduling problems generally has the following assumptions:

1) Each car can load varied but given-weight articles;

2) Customer demand and demand points are known;

3) Distance from the distribution center to customer demand points is known;

4) The distribution center has sufficient supply and has required transport capacity.

Fully considering above assumptions, various mathematical models will be established for vehicle scheduling. The research of the shortest path model has been very mature. By using the graph-based shortest path concept model established for emergent situations, this article focuses on the changes of distributive importance degree, that is shown in the graph on the arc of the weight changes (larger or smaller) situation[3].The weight is determined quantitatively by the importance degree of the index and the contribution degree to the target, and finally determined by the methods of subjective weighting, objective weighting or subjective and objective weighting. In this article, the subjective weighting method is used to determine the weighting value. Based on this, the shortest path concept model is established and model changes along with the weights change are discussed.

As shown in the figure, each vertex represents each distribution point. Assuming that a number of goods need to be shipped from v_1 to v_7, each number on the path represents its weight, so the corresponding initial conceptual model w is established based on the distribution problem. When delivery cost of the distribution network changes, that is the weight (v_2, v_4) of a certain arc changes in the graph, the corresponding model w' is obtained:
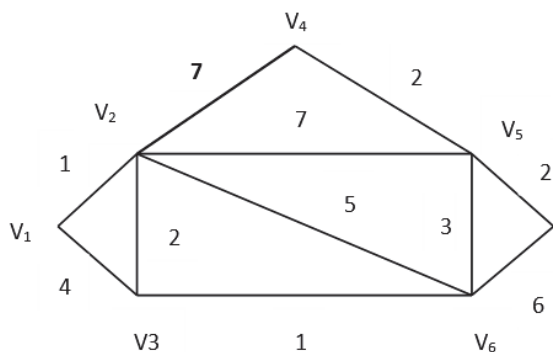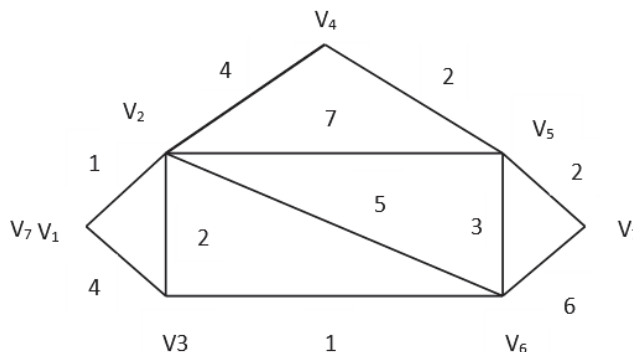
Figure 3.1 The shortest path concept model w

Figure 3.2 Change the weight of the model w '

E.D.Dijkstra algorithm simply the above two problems, causing all the calculations are limited to the associated edges. There are two shortest paths found in the original conceptual model w, which are:

$v_1 \longrightarrow v_2 \longrightarrow v_4 \longrightarrow v_5 \longrightarrow v_7$ and：$v_1 \longrightarrow v_2 \longrightarrow v_3 \longrightarrow v_6 \longrightarrow v_5 \longrightarrow v_7$

When the weights on the arc change, the only one shortest path is:

$v_1 \longrightarrow v_2 \longrightarrow v_3 \longrightarrow v_6 \longrightarrow v_5 \longrightarrow v_7$

The application of the model in this article is limited to solving the shortest path problem between distribution centers. But in reality, for avoiding unexpected situations, items need to be shipped to the demand point as soon as possible regardless of the transportation cost. In the ideal situation, average speed is used as the weight to calculate the delivery route.

4. The Application of the Improved Shortest Path Algorithm in Logistics Distribution

4.1 Dijkstra Algorithm Overview

The basic idea of Dijkstra's algorithm is that assigning each vertex label of the empowerment graph to a temporary label T and a fixed label P, and then change the T label at a point to a P label by calculation. The T label indicates that there is not a shortest path from the starting point to this point. The P label indicates that a shortest path from the starting point to this point has been found. The calculation stops when the end point gets the P label. The shortest path from the starting point to some point can be found after a maximum of N-1 steps (N is G's vertex' number).

4.2 Improved Shortest Path Algorithm Instance Analysis

As can be seen in the Figure 3.1 and the Figure 3.2, Dijkstra's algorithm cannot effectively solve the specific decision-making problem of logistics distribution routes. For example, when the weight of the shortest route changes or the path cannot pass, the logistics distribution need to be decided again to recalculate or even change the route.

To solve this problem, this article designs an improved shortest path algorithm. Firstly, the logistics route matrix is set as R:

The matrix R is a weighted adjacency matrix with $r_{ij}=0$ when i = j and $r_{ij}=\infty$ when there is no path between vi and v_j. In the current case, the specific setting is: starting from v_1 to v_2, v_3 has direct Channel, and the distance is 1, 4, respectively, so set r1,1=0，r1,2=1，r1,3=4,3 = 4 in the matrix, and the rest $r_{0, i} = \infty$, and so on, we can get all the points .

Let M be the set of shortest path vertex and DT (i) be the total length of the various routes from the starting point to the point i. Starting with M={v0},DT(i)=R(0,i)={0,1,4,∞,∞,∞,∞},DT (2) = 1 and M={v_0,v_2} are obtained from DT(j)=min(DT(i)|Ti∈ T-M), where DT (i) becomes {0,1,3,5,8,6,∞}, and continue to calculate DT (3) = 2 according to the above algorithm, and modify M={v_0,v_1,v_2}, DT (i) becomes {0,1,2,5,8,3,∞}.

Here, the intermediate matrix Y is introduced under the following conditions: if 0<R<∞, then Y (i,j)=R (i,j)+DT(i), does not change if not satisfied, that is Y (i,j) = R (i, j). So we get Y:

$$Y=\begin{bmatrix} 0 & 1 & 4 & \infty & \infty & \infty & \infty \\ 2 & 0 & 3 & 5 & 8 & 6 & \infty \\ 6 & 4 & 0 & \infty & \infty & 3 & \infty \\ \infty & 6 & \infty & 0 & 4 & \infty & \infty \\ \infty & 10 & \infty & 5 & 0 & 6 & 5 \\ \infty & 6 & 2 & \infty & 4 & 0 & 7 \\ \infty & \infty & \infty & \infty & 4 & 8 & 0 \end{bmatrix} \quad R=\begin{bmatrix} 0 & 1 & 4 & \infty & \infty & \infty & \infty \\ 1 & 0 & 2 & 4 & 7 & 5 & \infty \\ 4 & 2 & 0 & \infty & \infty & 1 & \infty \\ \infty & 4 & \infty & 0 & 2 & \infty & \infty \\ \infty & 7 & \infty & 2 & 0 & 3 & 2 \\ \infty & 5 & 1 & \infty & 3 & 0 & 6 \\ \infty & \infty & \infty & \infty & 2 & 6 & 0 \end{bmatrix}$$

Next, the set of vertex M is modified to M, repeated on the shortest path length and the accumulation of the same time to modify the set of vertex to get:

$M(v_1)=\{v_2\}$, $M(v_2)=\{v_4,v_3\}$, $M(v_4)=\{v_5\}$, $M(v_5)=\{v_7\}$, $M(v_3)=\{v_6\}$, $M(v_6)=\{v_5\}$,

Comprehensive can get all the shortest routes for：

$v_1 \longrightarrow v_2 \longrightarrow v_4 \longrightarrow v_5 \longrightarrow v_7$ and: $v_1 \longrightarrow v_2 \longrightarrow v_3 \longrightarrow v_6 \longrightarrow v_5 \longrightarrow v_7$

5. Conclusion

Based on the data mining, the article studies the vehicle scheduling problem. Firstly, the current application status of data mining is analyzed to indicate the development of data mining. Then, the application of data mining in vehicle scheduling is explained and suggestions for current logistics distribution are proposed, which make the vehicle scheduling can both meet the delivery tasks and minimize the total distance effectively. Next, the article proposes the shortest path concept model and the logistics distribution path optimization problem model after analyzing Chinese status. Finally, Dijkstra algorithm is improved for the shortest path model to optimize allocation of resources.

The current data mining has been valued and applied by various industries, and its achievements are obvious to all. With the development of data mining and the needs of the logistics industry, the time is ripe for data mining to apply to logistics information, which makes it possible for enterprises to respond quickly to the market demand. The prospect is promising.

References

Cheung, D., Han, J., Ng, V., & Wong, C. Y. (1996). *Maintenance of discovered association rules in large databases: An incremental updating technique.* Stanley Y. W. Su. Proceedings of the twelfth International Conference on Data Engineering(ICDE'96), New Orleans, Louisiana, IEEE Computer Society, 106-114. https://doi.org/10.1109/ICDE.1996.492094

Ramaswamy, R., Chakravarti, N., & Chakravarti, N. (2005). Sensitivity analysis for shortest path problems and maximum capacity path problems in undirected graphs. *Mathematical Programming,* *102*(2), 355-369. https://doi.org/10.2139/ssrn.489804

Zhang, G. Y., Yin, J. D., & Gong, L. Y. (2010). Data mining in the logistics business management applications based on SQL. *Enterprise Economics, 2010*(11), 45-47.