Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 2 (2012) 293 – 299

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Power and Energy Systems

Research of the Data Mining Based Public Traffic Transfer Algorithm

Yin Kui-ying* , Jin Lin ,Chang Wen-sheng

Nanjing Research Institute of Electronics Technology Nanjing 210039, China

Abstract

The route relationship matrix M is employed in this paper to indicate relationship between routes. The travel solutions for bus stations are obtained by matrix operation. Through the data mining to the history record of transportation information, the road transportation status at specific time interval can be predicted. Based on the predication, the best travel solutions can be found. The algorithm fully uses history transportation information, optimizes the travel solutions according to the time consumed. It can satisfy the time requirement of traveler and is very useful.

Key words: Digital City; Public Traffic Transfer; Data Mining

1. Introduction

Digital city emerge with the development of internet and computer technology. The major functionality of digital city is providing information service for city residents. One of the valuable services in digital city is providing the city public travel route information. Normally, a medium city has about one hundred public travel routes and one thousand stations. How to select proper travel route and save travel time is the problem that city residents have to consider.

For convenience of travel, city public travel route system is developed. According to the start and target station, travel route system will inquire city pubic travel information database and find proper public travel route solution.

* Corresponding author. Tel.: +86-25 51824342; fax: +86-25 51824307.
E-mail address: yinkuiying@sina.com

 doi:10.1016/j.aasri.2012.09.049

Traffic status in city changes with time. But, there's regularity in traffic status change. For example, during rush hour of working days, traffic flow on city main road is very big and traffic jam is easy to happen. In order to provide convenient and time-saving travel route solution for traveler, while designing travel route system, the regularity of traffic status change should be considered.

2. Main idea of public traffic transfer algorithm design

The data source of public traffic transfer system is the bus route and station information in city public traffic system. According to start and target station, the public traffic transfer algorithm will output all possible transfer solutions between two stations. Then, these transfer solutions are sorted according to time or distance.

Therefore, to find the optimal public traffic transfer solution between two stations, the transfer algorithm includes following two major steps.

First, according to the start and target station, find all possible public traffic transfer solution, including direct route, one time transfer, two times transfer and more than two times transfer.

Second, according input parameter, provide optimal transfer solutions. According to investigation, when selecting public traffic solution, the major considerations of people are times of transfer, total spent time and distance.

Actual public traffic route is very easy to be affected by the traffic status of city road. When choosing travel route, people always want to avoid the section of road which is in bad traffic status. Therefore, the public traffic transfer algorithm need to predict the traffic status of the route of bus transfer solution at specified time. In this paper, the traffic status prediction is achieved through data mining of traffic status history statistics data.

3. Design of public traffic transfer mathematical model

3.1. Information of public traffic routes

Each bus route in city has a unique number. The different directions of same route have different numbers. The reason is as follows. First, some streets in city are one-way street. Second, the stations on different directions of same bus route are not fully identical. Third, start and end time of different directions of same route are different. This number is the unique identifier of bus route.

Each bus station in a city has a unique number. Because the different bus stations in city may have same name, this number is used as the unique identifier of bus stations.

The set of all bus routes pass station s is represented as CR(s)。

All stations in a bus route are ascending numbered from start station. The number of start station is 1. Number of following stations is increased by one according to the order in route. The number of one station in one bus route is represented as $K(R, r) = xxx$, where R is the number of bus route, r is number of station. If station r is not contained in route R, then $K(R, r) = 0$。

Suppose the number of two bus routes are A and B respectively, the common station set of route A and B is represented as CS(A, B)。 This set contains the intersection points of two routes and these points are the candidate stations for bus route transfer. Two routes may have multiple common stations, therefore, multiple elements may contained in CS(A, B)。If there's no common stations in two bus routes, the $CS(A, B) = \Phi$。

3.2. Bus route relation matrix

The relation of two bus routes can be divided into two categories. One is the two bus routes can be transferred from one to another directly, another is the two bus routes can't be transferred directly[1]。

Bus route relation matrix is used to represent the relation of bus routes. Suppose the number of two bus routes is i and j respectively, if CS(i, j) is not null, it means that traveler can transfer from route i to j directly, and in route relation matrix, $M_{i,j} = 1$。If can't transfer from route i to j directly, $M_{i,j} = 0$。

In the bus route relation matrix, $M_{i,i}$ is defined as zero. If i and j represent the two directions of same bus route, then both $M_{i,j}$ and $M_{j,i}$ are defined as zero.

According to above definition, the n-th power of route relation matrix M represents the relation of bus routes after n times transfer. M^n has following characteristic.

Theorem 1: the n-th power of matrix M represents the number of n times transfer solution between two different bus routes. If value of $M_{i,j}^{n}$ is bigger than zero, it means traveler can change from route i to j via n times transfer. The value of $M_{i,j}^{n}$ is the number of transfer solutions via different intermediate routes for changing from route i to route j.

The route relation matrix only represents the relation of two bus routes. It divides the relation of bus routes into two categories: intersections existing and intersections not existing. The n-th power of matrix represents whether two routes can be reached via n times transfer, also records the number of transfer solutions between two routes via different intermediate routes. In the real bus routes, two routes may have multiple intersections. Therefore, the number of transfer solutions between two real routes may not equal to value of $M_{i,j}^{n}$.

Suppose three bus routes A, B and C. There's multiple intersections between A and B. Also multiple intersections exist between C and B. In all intersections of A and B, there's a station which has smallest number in route B. In all intersections of B and C, there's a station which has the biggest number in route B. If the smallest number on route B of all intersections of A and B is smaller than the biggest number on route B of all intersections of B and C, then it can insure there's a route which can transfer from A to C via B. As shown in Fig 1.

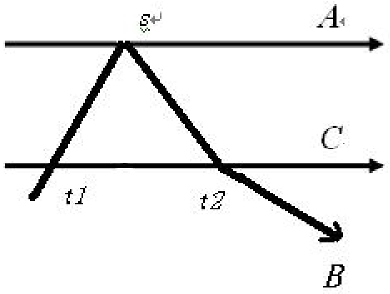

Fig. 1. bus route with multiple intersections

In Fig 1, there's an intersection between A and B, represented by s. There are two intersections between route B and C, represented by t1 and t2. The order of these three stations are represented as follows: K(B,t1)< K(B,s)< K(B,t2)。Because K(B,s)< K(B,t2), traveler can transfer from route A to B at station s, then transfer to route C at station t2, which is an intersection of B and C. Because K(B,t1)< K(B,s)，traveler can transfer from route C to B at station t1, then transfer to route A at station s.

The necessary condition of changing from route A to C via route B is represented as follow formula:

$$\underset{s \in CS(A,B)}{Min} K(B,s) < \underset{t \in CS(B,C)}{Max} K(B,t) \tag{1}$$

Formula 1 defines the necessary condition of changing from route A to B via B. But it's not sufficient condition. Whether the actual transfer solution is feasible or not is also determined by following conditions: position of start station, position of route intersection and position of target station. Even a feasible transfer solution existing, since the route has direction, the solution is only feasible for partial stations on route.

4. Algorithm of bus transfer between two stations

Suppose start station is s, target station is t. Following algorithm will use bus route information and bus route relation matrix to get the transfer solution from s to t.

(1) If $CR(s)\bigcap CR(t)\neq\phi$, it means there's route which is from s to t directly. For any bus route $R\in CR(s)\bigcap CR(t)$, if $K(R,s)<K(R,t)$, then traveler can go from s to t via route R.

(2) For route $i\in CR(s)$ and route $j\in CR(t)$, if the corresponding element of route relation matrix $M_{i,j}\neq 0$, then there's one time transfer solution between station s and station t.

Suppose there's station $m\in CS(i,j)$, which meet follow condition: $K(i,s)<K(i,m)$, $K(j,m)<K(j,t)$, then transfer solution is: go to station m by route i, change from route i to j at station m, go to station t by route j.

(3) For route $i\in CR(s)$ and route $j\in CR(t)$, if the corresponding element of second power of route relation matrix $M_{i,j}^{2}\neq 0$, then there's two time transfer solution between station s and station t. The intermediate stations can be found through formula 1 and the bus route relation matrix. Then all the two time transfer solution between two stations can be found.

(4) Similar method can be used to find the solutions which have more than two time transfers. For route $i\in CR(s)$ and route $j\in CR(t)$, if the corresponding element of n-th power of route relation matrix $M_{i,j}^{n}\neq 0$, then there's n time transfer solution between station s and station t. The intermediate stations can be found through formula 1 and route relation matrix.

All feasible transfer solutions between two stations can be found through above algorithm. In the real city public traffic system, usually from one station can reach any other station via two times transfer[3]. Most travelers also do not like the solutions which have many time transfers, since they have purpose for travelling. Therefore, the 2 time power of route relation matrix can meet the normal requirement.

5. Sorting the transfer solutions

5.1. Analysis of public traffic traveler

All feasible transfer solutions between two stations can be obtained by the algorithm mentioned in above section. Maybe there are many feasible transfer solutions between two stations. Therefore, the optimal solution should be selected according some criterion.

In order to recommend the optimal transfer solution, we need to know the consideration of traveler when selecting the travel route. According to the survey of top considerations of traveler while selecting public traffic route in HeFei city, 2002[4], 38.24% of travelers select minimum times of transfer as the top consideration, 32.53% of travelers select minimum travel time, 20.51% of travelers select shortest distance, 8.71% of travelers select minimum travel expense.

According the above survey, while selecting travel route, the major considerations of traveler are number of transfer and travel time. The route transfer algorithm between two stations in previous section can insure the solutions with less times of transfer will be obtained earlier than the solutions with more times of transfer. Therefore, the algorithm can insure the optimal solution of transfer times.

5.2. Minimum time transfer solution

The second consideration of traveler for transfer solution is time. The travel time of a transfer solution is very easy to be affected by the road traffic status. Therefore, the best method of saving travel time is selecting the transfer route which excludes the section of road in bad traffic status.

In this paper, firstly, the history data of city traffic status is reorganized and analyzed. Then, through data mining of the traffic history data, the prediction to traffic status of some areas in city is present. The traffic status prediction is used as reference when selecting bus transfer solution.

5.3. Design of traffic status data

The information of traffic status includes traffic flow, vehicle speed, etc. These data need to be categorized according to time and position.

The traffic status information of some road section can be recorded in a table. Following fields are contained in this table: name of road section, ID, date, time, traffic flow, vehicle speed, etc. The date field is used to record the detail date when traffic status monitoring is performed, including year, month and day. The time field is used to record detail time of monitoring. The information of traffic flow and vehicle speed is used to estimate the road traffic status.

All road sections which are monitored in a city can be managed by a road section table. This table will record the geographical position of road sections, the bus routes which go through this road section and the information of nearby bus stations. According to the information of bus routes and stations in transfer solution, by inquiring the road section table, the road sections which are passed by the bus transfer solution can be obtained.

The date management table is used to manage the date in traffic information table. It’s used to record the map of date and days of the week.

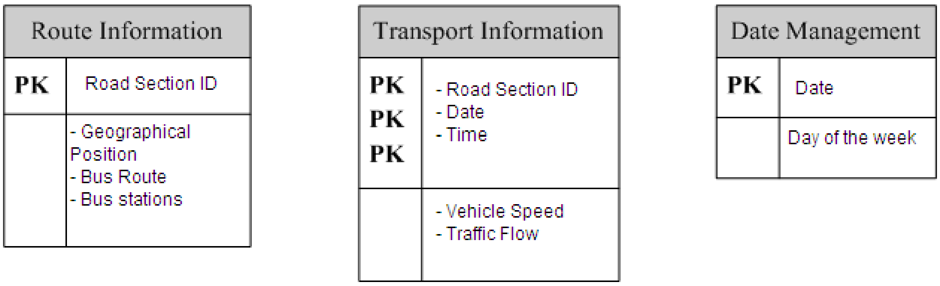

Fig. 2. Traffic information table

5.4. The traffic data cube

Data cube is used to modeling and inspecting the data in database in multi dimensional way. When evaluating the transfer solutions, the most useful traffic information data is the traffic status of some road sections at particular moment. Therefore, the data cube is built according to following rules.

(1) Road section ID, time, days of the week and vehicle speed are selected to build 4-dimensional data cube.

(2) Discretization is performed on time and vehicle speed.

(3) Following traffic status information of specified road section is inspected by data cube： the frequency of vehicle speed falling into some range in particular time period of some day.

When building the data cube, firstly, the time and vehicle speed need to be discretized. Half an hour is the unit to divide time and 5 kilometer per hour is the range unit to divide vehicle speed.

The traffic information of same road section in two days of one week is possible to be big different. For example, in some main traffic roads, at rush hours of work days, the traffic flow is very big and easy to lead to traffic jam. But at weekends, the traffic flow of same roads is relatively small. Therefore, the days of the week field in date management table of database is selected as one dimension of data cube for inspecting traffic status.

In one road section, the schematic diagram of data cube is shown in Fig.3.

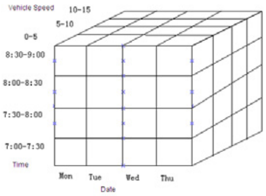

Fig. 3. Traffic information data cube

According to above definition of data cube, the road section traffic status information of given time period in some day of a week can be acquired. For example, in data cube, the history data of vehicle speed information in 8:00-8:30 time period of Monday can be analyzed. If 80% of all history vehicle speed records are lower than 5 kilometers per hour, then, it can be predicted, traffic jam is very likely to happen in this time period.

According to experience, rollup operation can be performed on data cube. Through dimension conflation to aggregate on the data cube[6]. For example, according to common knowledge, there's big difference between the traffic status of work day and weekend. But the difference of traffic status between different workdays is small. Therefore, conflation can be performed on the date dimension. For statistics, date is divided into two categories work day and weekend.

5.5. Sorting of transfer solutions

According to the route transfer algorithm in above section, all feasible transfer solutions between two stations can be obtained. Then, for each transfer solution, through inquiring the road section table, get all measurable road sections contained in this transfer solution. The prediction of road section traffic status in specified time is acquired by analyzing the history data of traffic status in road section data cube according to the date and time field.

By predicting the traffic status of road section contained tin transfer solution, the road section which is easy to be jammed can be avoided while selecting travel route. This is helpful for guiding traveler to select proper bus travel route. For example, there are two transfer solutions A and B acquired through the route transfer algorithm. By analyzing the road sections contained in solution A through data cube, 70% of vehicle speed record is lower than 10 kilometers per hour. For solution B, 75% of all vehicle speed history record is between 10 kilometer per hour and 20 kilometer per hour. It means that the traffic status of road sections in B is better than A at this time, therefore, recommend B solution to traveler.

6. Conclusion

In this paper, we design the mathematics model of relation between bus route and stations. Bus route relation matrix M is used to describe the relation between routes in public traffic system. By matrix calculation, the transfer relation between bus routes can be acquired. The bus transfer algorithm between two stations present in this paper can be used to get all feasible transfer solutions between two stations.

According to the investigation of traveler considerations, through mining the traffic information history data, the prediction to road traffic status at specified date and time can be acquired. All feasible transfer solutions between two stations are sorted and recommended according to the prediction of road traffic status.

In this algorithm, data mining is used to predict road traffic status. Therefore, the history data of road section traffic status is fully utilized and the correctness of algorithm is improved. All feasible transfer solutions are sorted according to the time cost, which meet the requirement of traveler for short travel time. Compared with some present bus transfer algorithm which aiming at the shortest travel distance, this algorithm has better practicability.

References

[1]H Y Chen，ZH M Jia, The Hierarchy and Incremental Bus Route Guide Algorithm Based on WebGIS，Computer ERA，2003，12: 26-27.

[2] Chao-Lin Liu, Tun-Wen Pai, Chun-Tien Chang, and Chang-Ming Hsieh.
Path-planning Algorithms for Public Transportation Systems. 2001 Intelligent Transportation Systems Conference Proceedings, 2001, 1061-1066.

[3] Bin Li，Chao Yang，Pei-kun Yang，The Shortest Bus Route Algorithm and Network Accessibility Index Calculation，Journal of Tongji University，1997，25(6): 651-655.

[4] Jun-jun Tang，Dai-sheng Zhang，Chao-yang Chen， Research of Travel Route Selection Model in Bus Travel Query System， Journal of Hefei University of Technology，2004，27(10): 1227-1230.

[5] Jiawei Han Micheline Kamber，Data mining：concepts and technologys， China MachinePress，2001.8

[6] Berman F. From Tera Grid to knowledge Grid[J].Communications of the ACM, 2001.44(11):27-28..