BUS ARRIVAL TIME PREDICTION ALGORITHM BASED ON MARKOV CHAIN

Yuan Tian – PhD student
School of Transportation Science and Engineering, Harbin Institute of Technology, People's Republic of China
E-mail: ytian.phd@hotmail.com

Aleksandr Rakhmangulov – D.Sc (Tech.), Professor
Nosov Magnitogorsk State Technical University, Russia. Phone: +7-9000-285-000. E-mail: ran@magtu.ru

Dmitri Muravev – Doctoral Student
School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, People's Republic of China
Phone: +86-168-211-02-872. E-mail: Dmitri_Muravev@sjtu.edu.cn

Siqing Wang– PhD student
School of Transportation Science and Engineering, Harbin Institute of Technology, People's Republic of China

Abstract
Developing public transport is an effective way to solve traffic congestion and improve travel efficiency. Improving bus service quality can attract passengers to travel by public transport. In the past, as the bus arrival time is unknown, and the buses often arrive inaccurate, passengers feel anxious and the quality of public transport service declined. Even though some bus stations equipped with electronic bus stop boards, the predicted bus arrival time is often inaccurate. Therefore, in order to convenient for people travel by public transit, this paper puts forward the method of bus travel time prediction based on the Markov chain which considers the spatial-temporal characteristics of the bus travel time. The prediction method can improve the quality of the bus service, help the travelers to make travel planning and reduce the waiting time. The algorithm is verified by the actual operation data of No.114 bus line in Harbin. The results show that the prediction error is small, and the algorithm is easy to implement.

Keywords: bus arrival time, prediction, Markov chain, GPS.

1. Introduction*

In recent years, the problem of urban traffic is becoming more and more prominent. Traffic congestion, traffic accidents, environmental pollution, energy consumption and other issues have become the serious urban diseases. Public transport has the characteristics of large carrying capacity, high transport efficiency, low energy consumption, small relative pollution and low transportation cost, which is the best choice to solve urban traffic problems in the world.

Bus arrival time prediction is the core content of real-time bus travel information release, bus travel path guidance, bus operation dynamic scheduling management and so on. In recent years, many large cities in China have carried out research and practice on intelligent electronic bus stop board and public transport service information platform. It includes: (1) Entity electronic bus stop board. By installing the electronic bus stop board on bus stop to display the distance and time of the different bus lines in the form of electronic information, (2) The internet bus service information system. By developing computer and mobile phone client software, the users can inquire bus information in real time. It includes operation status, the bus stop numbers of the coming bus, distance and time of the vehicles. It provides real-time services for transit users.

According to the effect of the implementation, most cities provide the forecast of the number of bus stops or the distance to the bus stop, only a few cities try to predict the arrival time of the bus stop on specific lines. As there are many internal and external influence factors of the bus system and the influence factors are quite complex, the existing bus arrival time prediction systems are inaccurate, and some errors even reach more than 200%. Road traffic volume, passenger flow, interference in bus stops, emergencies and other uncertainties are the reasons causing the inaccurate.

The survey shows that the people who travel on the bus are very demanding on the punctuality and reliability. If the actual arrival time of the bus lag behind the forecast time seriously, it will lead to the prolongation of the waiting time for the passengers and cause the problems of travel anxiety and crowds. At the same time, the unreliable arrival time forecast will lead to the diversion of passengers to the other more reliable travel mode [1].

At present, the method of bus arrival time prediction mainly includes, historical averages and real-time GPS data weighted prediction method by Sun [2], neural network model [3, 4], Calman filter [5, 6], support vector machine (SVM) model [7] and so on.

2. Bus arrival time prediction algorithm based on Markov chain

2.1 Background of algorithm

The operation of the bus is related to temporal and spatial. The running state of buses is affected by various factors. For example, weather, road conditions, etc. In a short time, the running state of a running bus is very associated with the front bus stop's running state. For example, if the bus runs normally between two stations in the midway may also runs smoothly between the next two bus stops. And if the bus runs with delay in the previous stage, it may be accelerated in the subsequent operation. The travel time between bus stops can be used as the state of a specific bus route system. A series of bus stops internal travel time is the running state of the system at different stages. Therefore, the prediction method of discrete system can be used to estimate the travel time between bus stops [8]. By predicting internal travel time between bus stops and the arrival time of the bus, the arrival time of the bus to the next stop can be predicted.

The bus has a characteristic of state transition. The occurrence probability of the next stage is related to the current state, which is consistent with the basic idea of Markov chain. To apply the Markov chain algorithm, the following assumptions are made.

1. The travel time between bus stops in the same direction of the same bus route is conforming to Markov process. It is the bus stops internal travel time constitutes the Markov chain on the condition of same bus route and same direction;

2. In the same period, the same bus route and the same driving direction, the probability of state transition between two bus stops are same. The time difference characteristics of bus travel are fully taken into account. Besides, it is considered that the bus operation in the same direction at the same time has characteristic of homogeneous and it simplifies the complexity of the Markov state transfer matrix.

2.2 Prediction algorithm

For a series of dependent random variables, the Markov chain model with the step length of one and the initial distribution are used to infer the absolute distribution of the future, that is, the basic travel time prediction algorithm based on Markov chain, and the concrete prediction algorithm steps are as follows.

1. Dividing the running time of the vehicle. Assuming that 1 hour is a period of time, then 5:00 to 21:00 can be divided to 16 periods. Dividing the operation hours of buses in one day.

2. Constructing the state transition matrix. The state transfer matrix depends on the GPS historical data of the bus running. For each operation period, statistics of bus stops internal running time between 3 adjacent bus stops a, b and c to get travel time t_{ab} and t_{bc}, as shown in **figure 1**.
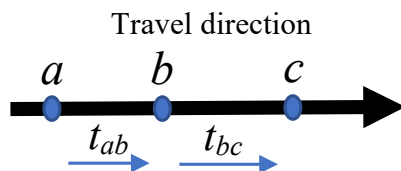
Travel direction

Fig. 1. Illustrations of bus stops internal travel time

(t_{ab}, t_{bc}) constructs a series of travel time transition pairs. For example, (255s, 350s) indicates that the travel time between bus stops a and b is 255 seconds and 350s between next two bus stops b and c. The probability of occurrence of state can be obtained by statistical historical data. The transition probability matrix of bus stops internal travel time can be obtained by a series of travel time transition pairs at the period T and bus route L

$$
\begin{array}{ccccc}
 & t_{bc\,0} & t_{bc\,1} & \cdots & t_{bc\,n-1} \\
t_{ab0} & P_{0\,0} & P_{0\,1} & \cdots & P_{0\,n-1} \\
t_{ab1} & P_{1\,0} & P_{1\,1} & \cdots & P_{1\,n-1} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
t_{abn-1} & P_{n-1\,0} & P_{n-1\,1} & \cdots & P_{n-1\,n-1}
\end{array} \quad (1)
$$

3. Bus arrival time deduction: When the vehicle arrives at the bus stop b, the travel time between stop a and stop b is known. By using the transition matrix, the travel time tbc between the stop b and stop c can be calculated.

$$t_{bc} = \sum_{j=0}^{n-1} P_{ij} t_{bcj}. \quad (2)$$

The arrival time of bus stop c $t_{arrive\,c}$ can be calculated by

$$t_{arrive\,c} = t_{arrive\,b} + t_{b\,c}, \quad (3)$$

where $t_{arrive\,b}$ is the arrival time of bus stop b.

3. Case study

3.1 Dataset

The dataset used for case study is the GPS data record of Harbin in December 2012. The bus route No.114 with direction from the Sugar industry institute to Chengde square is selected as case study verification route from more than 100 bus routes of the city. The GPS data from December 3rd to December 7th (Monday to Friday) are selected as training data, and the data of December 10th are used as test data. 9:00-10:00 is selected for the experiment period. Hexing 3rd Street bus stop, Hexing Road bus stop, Xidazhi Street are selected as experiment bus stops. **Fig**. **2** shows the data format that intercepted from the Oracle software. The fields are composed of bus line number (O_LINENO), bus name (O_BUSNAME), bus stop arrival time (O_ARRIVETIME), bus stop departure time (O_LEAVETIME), driving direction (O_UP), bus stop representative number (O_STATIONNO), etc. **figure 3** is the bus network map of GIS, which is used for GPS data matching and experiment.

After extracting the data from the database, replacing the result within a certain range with a numerical value to simplify the calculation and simplify the transfer matrix. 2'01-2'30 is replaced by 2 '15, 2'31-3'00 is replaced by 2' 45, and as so on. The constructed transfer matrix is shown in **table 1**.

Table 1

Transition matrix

Time	2'45	3'15	3'45	4'15	4'45	5'15	5'45	6'15	6'45	7'15	7'45
2'15	0.25	0.00	0.50	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2'45	0.09	0.00	0.00	0.18	0.18	0.27	0.18	0.00	0.00	0.00	0.09
3'15	0.00	0.00	0.14	0.00	0.00	0.43	0.29	0.00	0.00	0.14	0.00
3'45	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.33	0.00
4'15	0.00	0.00	0.00	0.20	0.20	0.00	0.20	0.20	0.00	0.00	0.20
4'45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5'15	0.00	0.00	0.00	0.00	0.00	0.25	0.50	0.00	0.00	0.00	0.25

The actual data of No.114 bus route at 9 am in December 10th are shown as **table 2**.

列 | 数据 | 约束条件 | 授权 | 统计信息 | 触发器 | 闪回 | 相关性 | 详细资料 | 分区 | 索引 | SQL

排序... | 过滤器:

	O_LINE...	O_BUSNAME	O_ARRIVEDATE	O_ARRIVETIME	O_LEAVETIME	O_UP	O_STATIONNO
1	114	3284	2012-12-01	05:24:13	05:24:32	1	33
2	114	3284	2012-12-01	05:24:47	05:24:58	1	34
3	114	3293	2012-12-01	05:37:49	05:37:58	1	33
4	114	3293	2012-12-01	05:38:15	05:38:32	1	34
5	114	3288	2012-12-01	05:51:16	05:51:26	1	34
6	114	3341	2012-12-01	05:51:24	05:51:45	1	2
7	114	3341	2012-12-01	05:52:14	05:52:28	1	3
8	114	3299	2012-12-01	05:50:54	05:50:57	1	34
9	114	3288	2012-12-01	05:50:56	05:51:06	1	33
10	114	3341	2012-12-01	05:55:23	05:55:45	1	5
11	114	3341	2012-12-01	05:52:52	05:53:19	1	4
12	114	3282	2012-12-01	05:53:29	05:53:39	1	33
13	114	3282	2012-12-01	05:53:50	05:54:00	1	34
14	114	3341	2012-12-01	05:59:01	05:59:33	1	8
15	114	3341	2012-12-01	05:56:22	05:56:52	1	6
16	114	3341	2012-12-01	05:57:36	05:58:02	1	7
17	114	3293	2012-12-01	05:40:09	05:59:45	0	1

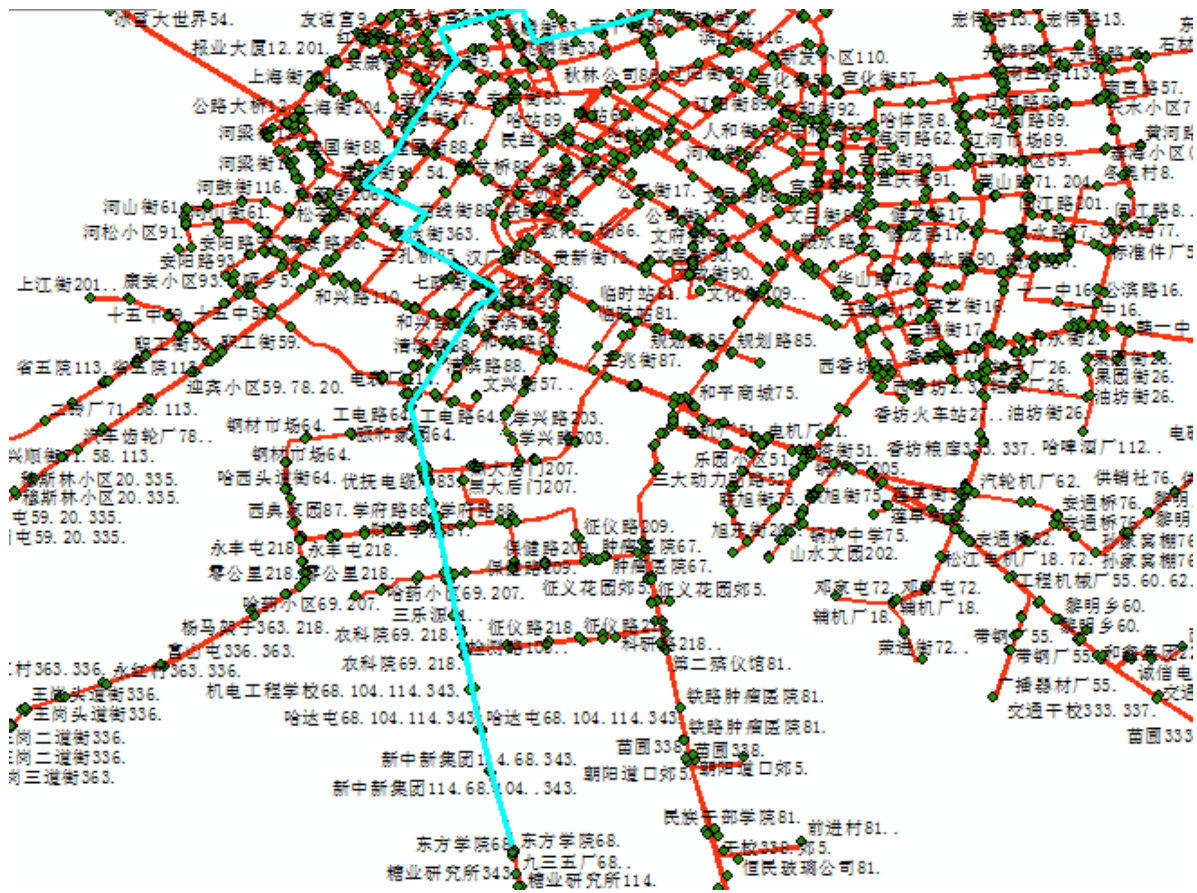Fig. 2. Data format of GPS data

Fig. 3. Bus network map and No.114 bus route in GIS

Table 2

The actual data of No.114 bus route

Hexing 3rd Street bus stop- Hexing road bus stop	Hexing road bus stop - Xidazhi Street bus stop
2'15	3'45
2'45	4'15
2'45	5'15
2'45	5'15
3'15	5'45
3'45	5'15
4'15	5'45

Prediction value which is calculated by using the data from first column and the transition matrix is shown in **table 3**. For the convenience of comparison, the result is converted into a decimal format, and the visual comparison results are shown in **figure 4**.

Table 3

Prediction results

Prediction value	Real value	Error (absolute value)
3.625	3.75	0.125
5.068182	4.25	0.818
5.464286	5.25	0.214
5.75	5.25	0.500
5.75	5.75	0.000
0	5.25	5.250
6.125	5.75	0.375

3.2 Analysis of prediction results

As shown in **figure 4**, the predicted value and the real value are completely different in the sixth vehicle. Analyzing the reasons, as the limited amount of data, there is no historical data from Hexing 3rd Street bus stop to Hexing road bus stop at 4'45, the constructed transition matrix is not accurate. But this problem can be solved by adding a large amount of historical data. The errors between predicted values and real values of other vehicles are small. Finally, the bus arrival time prediction is completed by calculating arrival time and predicted bus stops internal travel time.

Conclusion

Based on the operation characteristics of bus vehicles, a Markov chain based algorithm is proposed to predict internal travel time between bus stops and bus arrive time. A case study by using Harbin No.114 bus route verified the algorithm. The prediction results of the model are

relatively good except for the imperfection of transition matrix.
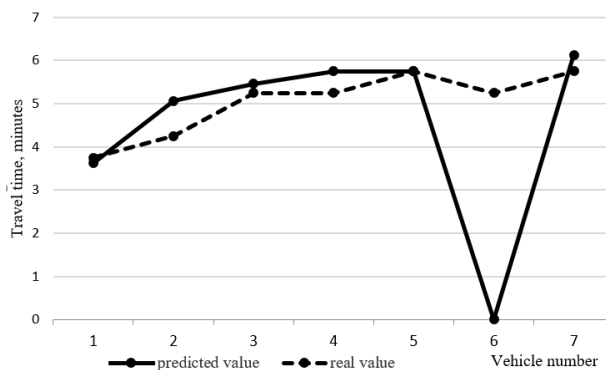

Fig. 4. The comparison of predicted value and real value

As the dataset is limited, the verification in this paper is inadequate. The prediction model can be improved by increasing the amount of data and improving the algorithm in the future.

References

1. XIANG Hongyan, PENG Xuewen Current Study and Development Trend of Bus Arrival Time Prediction // Journal of Transport Information and Safety. 2014, no. 4, pp. 57–61. (In Chinese).
2. Sun D., Luo H., Fu L., Liu W., Liao X., Zhao M. Predicting Bus Arrival Time on the Basis of Global Positioning System Data // Transportation Research Record: Journal of the Transportation Research Board. 2007, vol. 2034, no. 1, pp. 62–72. doi: 10.3141/2034-08.
3. Chien S. I.-J., Ding Y., Wei C. Dynamic Bus Arrival Time Prediction with Artificial Neural Networks // Journal of Transportation Engineering. 2002, vol. 128, no. 5, pp. 429–438. doi: 10.1061/(ASCE)0733-947X(2002)128:5(429).
4. Lin Y., Yang X., Zou N., Jia L. Real-Time Bus Arrival Time Prediction: Case Study for Jinan, China // Journal of Transportation Engineering. 2013, vol. 139, no. 11, pp. 1133–1140. doi: 10.1061/(ASCE)TE.1943-5436.0000589.
5. Shalaby A., Farhan A. Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data // Journal of Public Transportation. 2004, vol. 7, no. 1, pp. 41–61. doi: 10.5038/2375-0901.7.1.3.
6. Vanajakshi L., Subramanian S. C., Sivanandan R. Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses // IET Intelligent Transport Systems. 2009, vol. 3, no. 1, p. 1. doi: 10.1049/iet-its:20080013.
7. Bin Y., Zhongzhen Y., Baozhen Y. Bus Arrival Time Prediction Using Support Vector Machines // Journal of Intelligent Transportation Systems. 2006, vol. 10, no. 4, pp. 151–158. doi: 10.1080/15472450600981009.
8. Jihua Hu, Guoyuan Li, Zhifeng Cheng Algorithm for predicting bus travel time between stops based on Markov chain // Journal of Transport Information and Safety. 2014, vol. 32, no. 2, pp. 17–22. (In Chinese)