

GITSS2015

## Bus Dwell Time Prediction Based on KNN

Jianxia Xin<sup>a,b,\*</sup>, Shuyan Chen<sup>a,b</sup>

<sup>a</sup>*Jiangsu Key Laboratory of Urban ITS, Southeast University, Si Pai Lou #2, Nanjing, 210096, China*

<sup>b</sup>*Jiangsu Province Collaboration Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou #2, Nanjing, 210096, China*

---

### Abstract

The objective of this research is to develop a dynamic model to predict bus dwell time at downstream stops. The research also intends to test the proposed model using real-world data. This model is based on k-Nearest Neighbour (KNN) algorithm using history and current data collected by GPS (Global Position System) fixed on buses. In the research, the data of buses of No.B1 line of Changzhou in China is used. In the test with real-world data, the proposed bus dwell time prediction model performed effectively both on accuracy and calculating speed.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Department of Transportation Engineering, Beijing Institute of Technology

*Keywords:* bus dwell time; prediction; k-nearest neighbor algorithm

---

### 1. Introduction

In order to promote competitiveness with other modes of travel such as the passenger car, bus transit agencies need to improve the quality of service to attract more people. Providing the information about bus arrival time makes public transit more attractive [1]. Bus travel time between two points consists of two parts: bus travel time on segments between every two bus stations, and bus dwell time at bus stations. This article mainly focuses on prediction method of bus dwell time.

Bus dwell time is composed by three parts: bus door opening, passengers up and down, bus door closing [2]. Bus dwell time is influenced by many factors. In the TCQSM (Transit Capacity and Quality of Service Manual), factors that affect the dwell time consist of passenger boarding and alighting volumes, fare payment method, in-vehicle circulation, and stop spacing [3]. Prediction models can be set up between dwell time and these influence factors.

---

\* Corresponding author. Tel.:15805151683.

E-mail address: [nicholas.xjx@163.com](mailto:nicholas.xjx@163.com)

Martin Milkovits [4] established bus dwell time model using data from the automatic passenger counting, automatic fare counting, and automatic vehicle location systems installed on Chicago Transit Authority buses. Peng Qingyan [5] added coefficients related to weather, time (flat or peak period) and station location, to adjust the primary models. Because of the size of BRT station, Influence of passengers walking time should be considered [6]. Beside of these mathematical regression models, Machine learning algorithms also have been used to establish the relationship between bus dwell time and influence factors [7, 8].

However, all of these models are poor in Generalization ability and practicality. Because, different bus stop with different situation need different prediction model, and, how many passengers will get on the bus and how many get off are both needed to be predicted. With the advent of big data era, relationship between the bus dwell time and influence factors can be ignored. What we need to study is its changing rule. Time series based methods has been used in bus dwell time prediction using history data of bus dwell time [9]. Compared with previous models, this kind of model can be used in practice without need of adjustment according to bus style, station form and also without need of prediction the number of passengers will on and down.

This article mainly focus on analysis of bus dwell time changing rule and propose a dynamic model based on KNN to predict bus dwell time at bus station.

## 2. Methodology

### 2.1. Model exploration

Bus dwell time at each stop is determined by passenger demand and traffic condition .In a certain passenger demand, and the traffic condition is so stable that the headways of each bus are equal to the departure interval. Obviously every bus will experience similar stop time in each station. However, in fact, the passenger demand and traffic condition are in changing. First, higher passenger demand makes dwell time at stops longer, and vice versa. So compared with other periods, in peak, more passengers will wait at stations after a certain interval. Second, the change of the traffic condition on segments may change the headway of buses which influence the number of passengers will get on the bus. For example, if one segment becomes crowded, then, the travel time of the segment become longer, so the headway of the buses run through it become larger, and there will generate more passengers at downstream stops. In short, bus dwell time at each stop is determined by passenger demand and traffic condition, and both of them are in changing. Because passenger demand and traffic condition are both in approximate periodic change [10], so there are similar passenger demand and traffic condition, and similar characters of stopping at stations at the same time every day.

Besides, bus dwell time at a stop relates to its dwell time at upstream stops. Assuming that, the passenger generation rate at one bus stop is “ $G$ ”, and headway between two buses is “ $T$ ”. So the number of passengers who will get up at the station is  $N_{up}$ , and the proportion of passengers who will get down at anther station in these passengers is “ $P$ ”. For exempla, at station A, every unit of time generates  $G_a$  Passengers, and  $P_{ab}$  percent of them will get down at station B,  $P_{ac}$  percent of them will get down at station C, and so on. The headway when the bus running into station A with the previous bus is “ $T_a$ ”. Then:

(1)How many passengers will get on can be obtained by

$$N_{up} = G * T \quad (1)$$

Where  $G$  is passenger generation rate,  $T$  is headway when a bus running into a station,  $N_{up}$  is the number of passengers will get on the bus.

(2)How many passengers will get down can be obtained by

$$\begin{aligned}
 N_{down a} &= 0 \\
 N_{down b} &= N_{up a} * P_{ab} \\
 N_{down c} &= N_{up a} * P_{ac} + N_{up b} * P_{bc} \\
 N_{down d} &= N_{up a} * P_{ad} + N_{up b} * P_{bd} + N_{up c} * P_{cd}
 \end{aligned}
 \tag{2}$$

The subscript *a, b, c, d...*denote station name.

(3)The total number of passengers on the bus can be obtained by

$$\begin{aligned}
 N_{Ta} &= N_{up a} - N_{down a} \\
 N_{Tb} &= (N_{up a} + N_{up b}) - (N_{down a} + N_{down b}) \\
 N_{Tc} &= (N_{up a} + N_{up b} + N_{up c}) - (N_{down a} + N_{down b} + N_{down c})
 \end{aligned}
 \tag{3}$$

The subscript *a, b, c, d...*denote station name.

(4)Dwell time is:

$$T_d = \max \left( a * t_{up} * \frac{N_T}{C} * N_{up}, b * t_{down} * \frac{N_T}{C} * N_{down} \right)
 \tag{4}$$

$t_{up}$  is the time of one person getting on the bus,  $t_{down}$  is time of one person getting down the bus,  $\frac{N_T}{C}$  is the vehicle loading rate,  $\alpha, \beta$  are coefficients.

So, bus dwell time at a stop relates to its dwell time at upstream stops, and in this research, bus dwell time at upstream stops will be used as input state vector. And, database is divided in to several parts according to time type (morning peak of weekdays, flat period of weekdays, evening peak of weekdays, weekends).Only the data of same type are used for prediction.

### 2.2. Prediction model based on KNN

K-nearest neighbour (KNN) method is one of the simplest methods of pattern recognition. First, it measures the distance between the test data point and all training points. Then it chooses *k* data points closest to the test data point. The output for the test data point is the weighted mean of the outputs for all those *k* training data points.

In the dynamical bus dwell time prediction model, bus dwell time at upstream stops will be used as input state vector. In order to improve predict accuracy and calculating speed, history data is divided into four databases: morning peak of weekdays, flat period of weekdays, evening peak of weekdays, weekends. When searching neighbours, the search scope is just the database of same type.

In k-nearest neighbour algorithm, the principle of choosing neighbours is calculating there Euclidean distance. For example, bus dwell time at stop *n* of trip *m* is to be predicted, and history data of bus dwell time are known, as shown in figure 1. Then, the Euclidean distance can be calculated using the equation 5.

$$ED_i = \sqrt{\sum_{j=1}^{n-1} (T_j^i - T_j^m)^2}
 \tag{5}$$

Superscript “*i*” is trip number ( $i=1,2,3...m$ ), subscript “*j*” is stop number ( $j=1,2,3...n$ ). “*n*” is the objective station number, “*m*” is the objective bus trip number.  $T_j^i$  is the bus’ dwell time of trip *i* at *j* station.

According to the results of distance, K nearest neighbours can be found. Then, the bus dwell time at station “*n*”

can be calculated by equation 6:

$$T_n^m = \sum_{k=1}^K \frac{\frac{1}{ED_k}}{\sum_{l=1}^K \left(\frac{1}{ED_l}\right)} * T_n^k \quad (k=1,2,3,4,\dots,K) \tag{6}$$

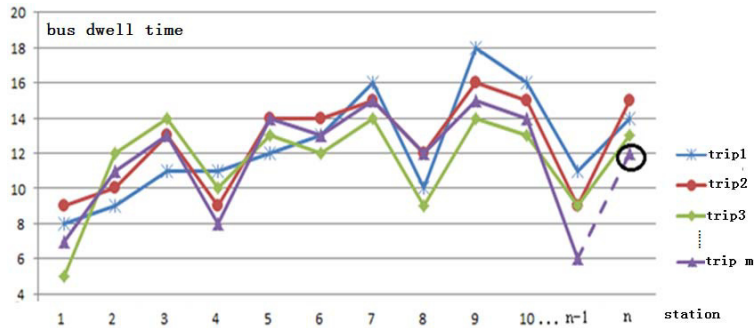


Fig. 1. bus dwell time on each station of destination trip and history trips.

### 3. Case study

#### 3.1. Data collection

Bus dwell time at every station as input of the proposed model can be obtained by GPS data. But, GPS only record bus travel speed, location and time. So, firstly, GPS data should be processed. The method of computing bus dwell time is as following:

- Step 1: find all the bus data that it is close to a bus stop and its speed is zero (in the study, the scope of 50 meters from a bus station is considered as stop area);
- Step 2: find the previous data and the next one, as shown in figure 2;
- Step 3: calculate bus dwell time by equation 6:

$$T = T_2 - T_1 - \left( \frac{V_1}{a_1} + \frac{V_2}{a_2} \right) \tag{7}$$

$$a_1 = a_2 = 2m / s^2$$

$D_1$  is the distance between the location of the previous data with the station, and  $D_2$  is the distance between the location of the next data with station.  $\alpha_1, \alpha_2$  is bus acceleration and deceleration respective.

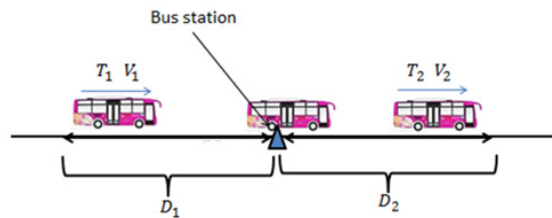


Fig. 2. bus stop at a bus station and its adjacent moment GPS collected.

The GPS data of bus route of line B1 in Changzhou is studied in the research. There are 30 stations along the bus line. The data is collected for a week, as shown in table 1. In the research, only data on morning peak period of weekdays is studied. 71 groups of bus dwell time have been extracted. And they are divided into two parts: 1/4 of the data are tested, the other is used as training data.

Table 1. Bus dwell time.

trip no.	Dir	type	bus dwell time(s)						
			stop1	stop2	stop3	stop4	stop5	.....	stop30
1	1	1	5	16	18	20	18	.....	25
2	1	3	9	17	20	19	21	.....	26
3	1	2	11	16	19	20	19	.....	22

Dir: 1-drive from north to south; 0- drive from north to south; Dir of 1 is tested in the research;

Type1: morning peak of weekdays, type2: flat period of weekdays, type3: evening peak of weekdays, type4: weekends,

Type1 is tested in the research

### 3.2. Result and analysis

The proposed model has been compared with other two methods: one is predicted by average dwell time, the other one is predicted by the method based on KNN using the total data rather than just using same type data. In the study, parameter “MAE”(Mean absolute error) is used to measure the prediction accuracy.

Due to the limitation of the predict method based on KNN, the dwell time at the first station can't be predicted, and it performs bad on prediction bus dwell time at the second station. So the test will skip the first two stations.

“MAE1” is Mean absolute error of predict method based on average dwell time; “MAE2” is Mean absolute error of predict method based on KNN using the total data; “MAE3” is Mean absolute error of predict method based on KNN using the same type data.

Because the purpose of the study is contrast the accuracy of these methods, so the test use the same K value in the two KNN based methods. Several tests show that, when K=7, the predict effect is best. The results are shown in table 2.

Table 2. MAE comparison table using different method at different station.

Method	Mean absolute error(s)									
	stop3	stop4	stop5	stop 6	stop 7	stop 8	stop 9	stop 10	stop 11	stop 12
MAE1	2.37	2.01	1.75	2.68	1.57	1.89	3.16	2.16	2.54	3.14
MAE2(k=7)	2.64	1.92	2.13	0.19	2.05	1.03	3.38	2.06	3.33	3.78
MAE3(k=7)	3.07	2.25	1.74	0.11	1.31	1.72	3.14	2.10	2.31	3.00
Method	Mean absolute error(s)									
	stop 13	stop 14	stop 15	stop 16	stop 17	stop 18	stop 19	stop 20	stop 21	stop 22
MAE1	3.45	3.37	3.81	1.97	0.76	2.20	2.14	2.99	1.06	2.37
MAE2(k=7)	3.65	3.49	4.21	1.95	0.77	2.47	2.55	3.48	1.24	2.44
MAE3(k=7)	3.14	3.22	3.48	1.95	0.72	2.45	2.04	2.69	0.98	2.20
Method	Mean absolute error(s)									
	stop 23	stop 24	stop 25	stop 26	stop 27	stop 28	stop 29	stop 30	AVERAGE	
MAE1	1.23	2.16	2.22	1.09	1.50	1.55	0.88	6.24	2.30	
MAE2(k=7)	1.34	2.70	1.98	0.94	1.65	1.58	1.81	7.49	2.44	
MAE3(k=7)	1.18	2.10	1.31	0.93	1.40	1.40	0.81	8.50	2.19	

The test result shows that method 3 that is the proposed method performs best. The prediction error is high at the end stop, because it often park around the end station for a long time. At station 3、4、18、30, MAE3 is not the minimum, the possible reason is the processing error of bus dwell time.

#### 4. Conclusions

The proposed prediction model based on KNN can predict bus dwell time at downstream stations using history bus GPS data of same time. The proposed model can be used in practice without need of adjustment according to bus style, stop form and also without need of prediction the number of passengers will on and down.

The test results shown that compared with the predict method based on average dwell time and the method based on KNN using the total data rather than just using same type data, the proposed method performs better. The processing error of bus dwell time may influence the prediction accuracy. For example some bus stations are close to crossroads, so the time of waiting for traffic light may considered as dwell time. If sensors can be fixed at bus station in future to collect bus dwell time instead of GPS, then the proposed model will perform better.

#### Acknowledgements

This work was supported by Natural Science Foundation of China under Grant No.61374195. The authors would like to thank the anonymous referees for their valuable criticism and fruitful comments.

#### References

- [1] Rahman, M. M., Wirasinghe, S.C. and Kattan, L., 2013. Users' views on current and future real-time bus information systems. *Journal of advanced transportation*, 47, 336–354
- [2] Wu, Y., 2007. The analysis of berthing time character at bus stop. *Transportation*. 2,78-80
- [3] Chen, G.J., 2013. Effect of bunching at stops on bus dwell time, *Intelligent Transportation Systems, Proceedings, ITSC*, pp. 1499-1502
- [4] Milkovits, M. N., 2008. Modeling the Factors Affecting Bus Stop Dwell Time: Use of Automatic Passenger Counting, Automatic Fare Counting, and Automatic Vehicle Location Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2072, 125–130
- [5] Peng, Q.Y..2002. Mathematical model of the dwell time at the bus stop. *Journal of Chang'an University (Natural Science Edition)*, 22, 60-62.
- [6] Sumeet, J., Jonathan, B. and Luis F., 2010. Influence of Platform Walking on BRT Station Bus Dwell Time Estimation. *J. Transp. Eng.* 136, 1173-1179
- [7] Wang, J., Deng, W., 2012. Bus dwell time prediction based on Bayesian network combined model. *Urban Transport of China*, 10.
- [8] Ding, J., Yang, M., 2014. Dwell time prediction of bus rapid transit using Arima-svm hybrid model. *Applied Mechanics and Materials*, 587-589, 1993-1997
- [9] Soroush Rashidi ,Prakash Ranjitkar .2015.Estimation of bus dwell time using univariate time series models. *Journal of Advanced Transportation*, 49(1), 139-152
- [10] Hyunho, C., Dongjoo, P., Seungjae, L., Hosang, L., Seungkirl, B., 2010. Dynamic multi-interval bus travel time prediction using bus transit data. *Transportmetrica* , 6(1), 19–38