Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 138–145

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)

# Bus arrival time prediction based on network model

Marko Čelan[a,*], Marjan Lep[a]

[a] *University of Maribor - Faculty of Civil Engineering, Transportation Engineering and Architecture, Smetanova ulica 17, 2000 Maribor*

**Abstract**

Providing accurate information on bus arrival and departure times at bus stops is one of the key parameters of high-quality public transport today. This paper proposes a model for the real-time prediction of arrival times at bus stops. The proposed model is based on information about the current location of the bus, the classification of runs into time periods with respect to the historical data, and the data model of the bus network. We discuss four types of data models: a data model defined by bus stops and crossings of the road network, a data model defined by bus stops, a data model which addresses the individual parts of the network in relation to the potential barriers that affect the travel speed of buses, and a data model with fixed-length links of the bus network. Travel times are classified according to the average travel speed into four time periods: morning, afternoon, early morning or late evening, and weekend. The results of the analysis showed that both the data model of the bus network and classifying runs into time periods affect the accuracy of predictions of bus arrival times at bus stops.



*Keywords:* public transportation; network data model; bus arrival prediction; tracking, real-time information system

## 1. Introduction

Information systems for real-time arrival times at bus stops (RTPI) are systems which provide passengers and potential passengers with real-time information about bus arrival times at bus stops. In order to deliver accurate and reliable information the system must be able to give a precise and probable estimate about the travel time predictions given all possible traffic conditions and circumstances[1]. The operation of these systems requires intelligent transport systems such as automatic vehicle location, automatic vehicle identification, systems for the validation of passengers entering and exiting the vehicles, and systems for the transmission and display of data. Real-time information about

* Corresponding author. Tel.: +386 2 22 94 366.
*E-mail address:* marko.celan@um.si


10.1016/j.procs.2017.08.331

bus arrival times at the bus stop has many positive effects on the users of public transport services and, consequently, (may) lead to an increase in the use of bus transport. Positive effects of RTPI systems – user perceived[2]:

- Decrease in perceived waiting time.
- Positive psychological factors, such as lower uncertainty.
- Increased ease of use and increased sense of security.
- Increased willingness to pay.
- Adapted travel behavior (i.e. better use of waiting time and more efficient travel).
- The impact on the choice of means of transport.
- Improved overall image of public transport.

The estimated bus arrival time at the target bus stop is calculated based on knowledge of current bus locations and estimated travel time that the bus needs from the current location to the target bus stop. Huge amounts of data on vehicle locations are nowadays obtained through the global positioning system (GPS)[3] that makes it possible to determine the current location and to obtain samples of past travel speeds.

Although numerous products and systems for predicting the operation of public transport in real time have been developed globally, there are still deviations between the predictions and the actual travel time[4]. Discrepancies between the prediction and the actual bus arrival time at the bus stop are particularly common in public bus transport systems that use the same road infrastructure as individual transport. Thus the bus travel times in different time periods, for example during peak hours, off-peak, or during weekends, can greatly vary.

## 2. Literature review

Many researchers, who used different paradigms in their models, have studied predictions of bus arrival times at the bus stop within public passenger transport. Altinkaya and Zontul[4] classified models that have been proposed thus far into four categories: models based on the historical data, statistical models, Kalman filtering models and machine learning models. Balasubramanian and Rao[5] proposed a bus arrival prediction model with cyclic variation explicitly incorporating information about seasonality in the data series (day of week, time of day). Deng et al.[6] have proposed a model for prediction of travel times based on a Bayesian network by comparing data on the traffic situation and travel times of buses from historical data. Chen et al.[7] used an automatic passenger counter as a key parameter in the model for prediction of the bus arrival time at a bus stop. Cats and Loutos[8] have developed three different schemes for the calculation of travel time:

- A method that calculates the arrival of the bus at the bus stop based on the current location of the bus and the pre-defined timetable.
- A method based on estimated travel times depending on the time period (day of week, time of day).
- A method which deals separately with travel times between bus stops and the time spent at the bus stop. To determine traffic conditions for travel times between bus stops it takes into account the speed of the preceding buses on the same route.

Numerous studies have been conducted to compare the accuracy of predictions using different methods and models. These methods and models can be classified as: Kalman filter models[9,10,11], artificial neuron network models[9,11,12], support vector methods[9,13,14], linear regressions[9] and nearest neighbour methods[9].

Alejandro et al.[15] used the average speed along the line with free traffic flow for prediction, and additionally their algorithm accounted time loss for each signalised intersection, roundabout, acceleration and deceleration at the bus stops, and the opening and closing of the bus door and passengers boarding.

Numerous authors have developed various proposals for predicting bus arrival times at the bus stop, but due to the complexity and nature of the data individual models or algorithms are not suitable for all environments[4].

Previous research in most cases did not focus on the data model of the bus network. Chen et al.[16] proposed and compared a model for prediction of the bus arrival time at stops, which is based on a section-based model of the bus network and a link-based model of the bus network. Zegeye et al.[1] proposed a model with bus stops and reference points that are defined by the distance from the selected bus stop.

## 3. Proposed model

This study presents a model that calculates the predicted travel time to the selected bus stop based on the current location of the bus and knowledge of a predetermined bus route trajectory. The proposed model is based on a pre-defined data model of the bus network and the classification of individual runs in time periods according to the average bus travel times in the past.

Knowledge of the current location of the bus is the basis for operation of the RTPI system. In general, there is an absolute or geographical location, which is expressed in latitude and longitude, and logical location – georeference – which is expressed in relation to the topological elements of the network data model (in the section, in the intersection, at the stop). Information on the location of the bus can be obtained in different ways. The most commonly used system for locating is the global positioning system (GPS). Nowadays this system can be used free of charge. This method of positioning the vehicle is not 100% accurate, therefore, for the purpose of locating buses on the bus network it is necessary to perform map-matching – positioning the coordinate of the point obtained from the tracking device on the bus network. The proposed model considers only logical track points extracted from the dynamic properties (velocity, acceleration, deceleration) of bus motion.

Due to stochasticity of the transport system, travel speeds in individual sections of the network vary, and travel times also vary between individual runs in the same sections. Travel speeds of buses are affected by numerous parameters such as density of traffic flow, administrative limitations, number of passengers, weather situation, possible delays of preceding buses, drivers, etc.

In the proposed model, the time periods are structured according to the time series obtained from runs on specific routes. The model considers travel times at different network sections from the past, which indirectly determines traffic situation and the dwell time at bus stops. In general the proposed model does not considered unpredictable incidents as road accidents and bus breakdowns.

### 3.1. Bus network data model

Besides identifying the current location, accurately predicting the arrival time of the bus at the stop requires knowledge of the bus route trajectory and the estimated time the bus will need to get from the current location to the target bus stop. Travel times between the current location and the point of prediction can vary significantly, especially when public transport is integrated with other motorized traffic. Travel times are calculated for specific sections of the bus network. Therefore, when devising an algorithm for the prediction of bus arrival times at bus stops it is necessary to define a data model of the bus network. Primarily, the data model of the bus network is defined by bus stops and links between adjacent bus stops[17]. In this study, the data model or bus route is defined by the nodes and links between the adjacent nodes.

The study compares the prediction of bus arrival times at bus stops according to different data models of the bus network, which are:

- Potential speed barrier data model (DM1).
- Bus-stop-based data model (DM2).
- Potential speed barier adaptive data model (DM3).
- Fixed link distance data model (DM4).

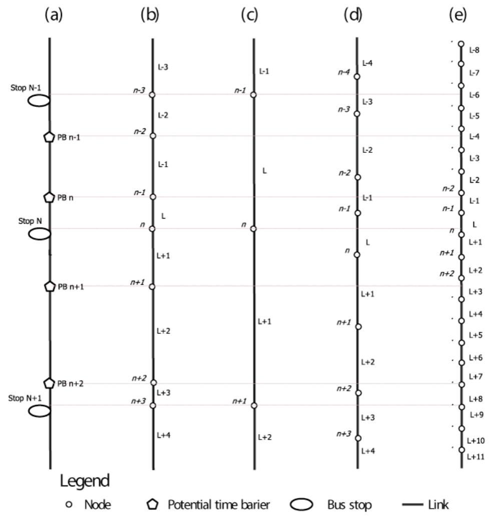


Fig. 1. Bus network data models: (a) bus route with bus stops and potential barriers; (b) potential speed barrier data model (DM1); (c) Bus-stop-based data model (DM2); (d) Potential speed barrier adaptive data model (DM3); (e) Fixed link distance data model (DM4)

The number of nodes and their locations and thus also the number of links along the bus route varies depending on the defined data model, namely:

- Nodes in DM1 are potential barriers (signalised intersections of the road network, road and rail traffic level junctions, roundabouts, yield junctions), which may affect the travel speed of the bus and the points just before the bus stops.
- Nodes in DM2 are located just before the bus stops.
- Nodes in DM3 are located midway between two points – the potential barriers (yield junction, signalised intersection, roundabout, railway crossing), which may affect the travel times of buses and the points just before the bus stops.
- Links in DM4 have a fixed length.

### *3.2. Prediction algorithm*

Predicted bus arrival time at the bus stop is equal to the time (T) that a bus needs to the end of link where it is currently located and the sum of the predicted travel times on other links of the network, from the next link to the link where the bus stop is located. Predicted information on travel times is determined by the average bus travel times in each time period. Predicted bus travel time from the current location to the target bus stop is:

$$T^{p} = t_{Lc}^{p} \cdot (1 - n_{Lc}) + \sum_{x=L_{c+1}}^{L_{s-1}} t_{x}^{p} \tag{1}$$

The terms in equation 1 have the following meanings:

$T^{p}$ - predicted bus arrival time in the time period p from the current location to the location of the target bus stop

$t_{Lc}^{p}$ - travel time that the bus in period p uses for the link of bus network on which it is currently located

$1 - n_{Lc}$ - percentage of the travel time for the link of bus network that remains for the bus to cover until the end of the link on which it is currently located

$\sum_{x=L_{c+1}}^{L_{s-1}} t_{x}^{p}$ - predicted bus travel time for all the links of bus network from the next link to the target bus stop

The data that can vary in the proposed model is the data model of the bus network and the time period for which the travel times are calculated.

## 4. Experiment

The study analysed bus travel times for individual links of the bus network in different time periods for bus line 1, Tezenska Dobrava, Maribor, Slovenia. The total length of line 1 is 7.495 km. Along the line are 15 bus stops. The bus route runs almost entirely along other motorized traffic. Part of the line between bus stops 1 and 2 runs separately on a dedicated bus lane, while bus service between bus stops 2 and 3 runs entirely separate from motorized traffic. Along the entire route buses pass through 11 signalised intersections and 4 roundabouts, while the bus route trajectory in the unsignalised intersections runs on a priority road. The line crosses a railway crossing with gates, which is rarely closed. The first part of the route with a length of 3 km runs in the centre with a speed limit of 50 km/h, while second part of the route runs predominantly on a four-lane

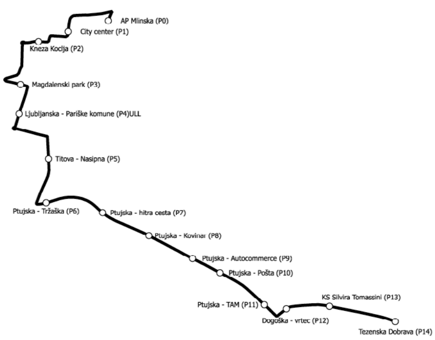


Fig. 2. Scheme of bus line

road with a speed limit of 70 km/h.

The proposed model has been evaluated on the basis of real data in the case study of city bus line 1 in the direction from the centre to the outskirts of the city. Line 1 is the most frequented line in Maribor and makes stops at 15 bus stops. The model is evaluated on the basis of collected information about the locations of buses over a 2-week period from 30 May 2016 to 12 June 2016. Data collected during the first week was used to determine the time periods and the calculation of travel times on each link of the bus network, and the data in the second week was used for the evaluation of the proposed model.

### *4.1. Determination of time periods*

Time periods are based on 270 runs, which were made in the first week of observation. Periods are based on the average travel time along the entire route. The obtained data show that the average travel velocities vary significantly between individual runs (fig. 3). Buses have the greatest average velocities late at night, early in the morning, and during most runs on weekends.

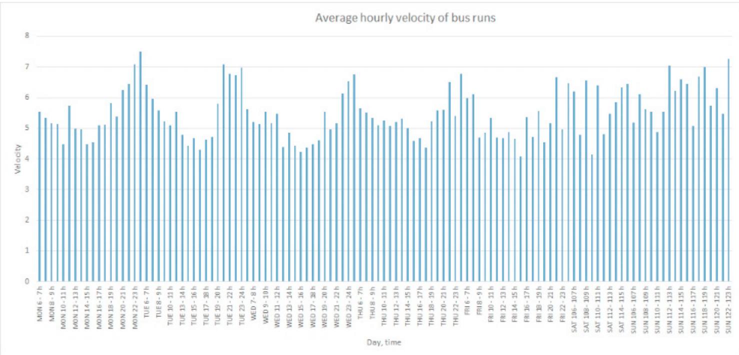


Fig. 3. Average velocity of bus runs

Table 1. Average velocity of bus runs by time period

| Time period | Number of runs | Average velocity (m/s) |
|---|---|---|
| Afternoon | 93 | 4.79 |
| Morning | 86 | 5.30 |
| Early morning, late evening | 57 | 6.16 |
| Weekend | 34 | 5.9 |
| Cumulative | 270 | 5.38 |

According to the measured data on the travel speed along the entire route, four time periods have been formulated: Morning period (Mon-Fri, 6 am – 12.00 pm), Afternoon period (Mon-Fri, 12 pm – 7 pm), Early morning, late evening period (Mon-Fri, before 6 am, after 7 pm) and Weekend period (Sat, Sun).

The obtained data showed that the average travel speed along the entire route is the lowest on weekdays during the afternoon. There is no morning spike which is expected, since the analyzed data is for the direction away from the city centre. The average travel speed in the morning and during the weekend is very similar.

### *4.2. Bus line data model*

As mentioned above, four models were proposed for data bus line 1. DM1 consists of 30 links. Link beginnings and ends were determined by the potential barriers and bus stops. The beginnings and ends of links in DM2 are determined by bus stops. In DM3, the points of the beginning and end of each link were determined midway between two potential barriers that affect the bus travel speed. There are 22 links along line 1. Links in DM4 are 100 meters long. Therefore this type of model has 75 links along the line.

### *4.3. Validation of the proposed model*

The validation of the proposed model was carried out on 340 runs (116 in the afternoon period, 106 in the morning period, 34 during the weekend, and 84 rides in the early morning/late evening period) from 6 June 2016 to 12 June 2016 for the four different types of data networks. Data is also disaggregated by time periods, namely for the morning hours from Monday to Friday, the afternoon hours from Monday to Friday, weekends, and early morning and late evening runs.

There is also model validation for just one time period. In total, 192,084 predictions of bus arrival times at bus stops have been calculated. For each piece of information on the current location of the bus, a predicted bus arrival time for all bus stops along the line was calculated.

The results of the model are based on the difference between the observed and calculated predicted bus arrival times at bus stops evaluated by three different approaches. The accuracy of the model was measured by the mean absolute error (MAE) between the observed and calculated travel time and given as a percentage of deviation between the observed and calculated travel time (MRA – mean relative error). Evaluation was carried out in percentage of the user perceived corresponding predictions (UPCP), namely a maximum of 1 minute prediction deviation for travel times from current location till target stop up to 5 minutes, a maximum of 2 minutes prediction deviation for the travel times up to 10 minutes, and a maximum of 3 minutes prediction deviation for the travel times greater than 10 minutes. A comparison of predictions of bus arrival times at the bus stop and the actual arrival time was performed for all 14 bus stops along the line. The fewest predictions were compared for bus stop P1 (2,871) and the most for the last bus stop on the line, P14 (21,793).

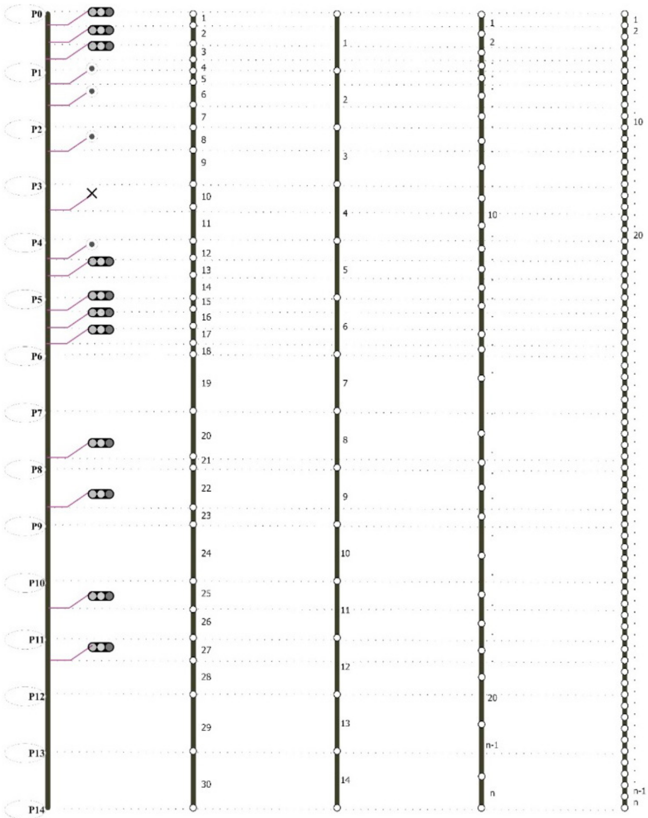

Fig. 4. Bus network data model: (a) bus route; (b) DM 1; (c) DM 2; (d) DM 3; (e) DM 4

Table 2: Number of predictions per bus stop

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Target stop (P1-P8) | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 |
| Number of predictions | 2,871 | 4,526 | 7,332 | 7,974 | 10,397 | 12,408 | 16,826 | 15,516 |
| Target stop (P9-P14) | P9 | P10 | P11 | P12 | P13 | P14 | Total | |
| Number of predictions | 16,676 | 17,969 | 19,352 | 20,415 | 21,029 | 21,793 | 192,084 | |

MAE is calculated by comparing the difference between the projected and observed bus travel times according to the proposed bus network data model, the time period with regard to bus travel times for the complete observed period, and for the observed period within a particular time period.

$$MAE = \frac{1}{n} \cdot \sum_{x=1}^{n} \left| t_{p_x} - t_{o_x} \right| \tag{2}$$

The results (table 3) showed that predicting using time periods improves the prediction of bus arrival times at bus stops, irrespective of the network data model. Data model selection primarily affects prediction accuracy for closer bus stops, while the difference in prediction accuracy for farther bus stops is very small. The best predictions were given for DM3 and DM4, while the absolute error is much larger for DM2, 61 seconds on average. Among all the combinations, the least deviation between the observed and predicted travel time was at DM3 in the morning period, namely a little under 45 seconds. MAE increases with distance and is 25–35 seconds when the distance between the current location and the target bus stop is approximately 1 km, at a distance of 2 km it is approximately 40 seconds, and at 6 km between 80 and 100 seconds, depending on the selected data model (fig. 5).

MRE is calculated based on the absolute difference between predicted and observed travel time given the actual travel time from the current location to the target bus stop. he results (table 4) showed that the average deviation between the observed and the predicted travel time given the data model is between 13.3% and 16.5%.

Table 3: MAE comparison according to the data model and with regard to time periods

| Data model | Time conside-ration | E. mor/ l. even. period (s) | Morning period (s) | After-noon period (s) | Weekend period (s) | Average (s) |
|---|---|---|---|---|---|---|
| DM1 | Period | 57.5 | 45.7 | 60.5 | 52.7 | 54.4 |
| | Together | 59.2 | 45.9 | 60.8 | 53.2 | 55.0 |
| DM2 | Period | 70.4 | 51.1 | 60.8 | 68.9 | 61.0 |
| | Together | 71.7 | 51.4 | 61.1 | 69.3 | 61.5 |
| DM3 | Period | 54.7 | 44.9 | 55.1 | 49.8 | 51.3 |
| | Together | 56.1 | 45.2 | 55.4 | 50.1 | 51.9 |
| DM4 | Period | 57.2 | 45.1 | 59.4 | 53.5 | 53.8 |
| | Together | 58.5 | 45.4 | 59.7 | 53.8 | 54.3 |

$$MRE = \frac{1}{n} \cdot \sum_{x=1}^{n} \frac{\left| t_{p_x} - t_{o_x} \right|}{t_{o_x}} \tag{3}$$

MAE increases as the distance from the target bus stop increases, while the MRE decreases as the distance increases (fig. 6). With bus arrival times, predictions at the bus stop within 500 m MRE is approximately 25%, while when the distance is more than 3 km the error is approximately 10%. UPCP is calculated based on the time-distance of the bus from the target bus stop according to the observed travel time. Correct predictions are those that differ from observed values up to 1 minute for observed travel times under 5 minutes, up to 2 minutes for observed travel times between 5 and 10 minutes, and up to 3 minutes for observed bus travel times over 10 minutes.

Table 4: Comparison of MRE according to the data model and time period

| Data model | E. mor./ l. even. period | Morning period | Afternoon period | Weekend period | Total |
|---|---|---|---|---|---|
| DM1 | 16.7% | 12.2% | 13.1% | 15.8% | 14.0% |
| DM2 | 20.5% | 14.5% | 14.4% | 20.4% | 16.5% |
| DM3 | 16.1% | 11.9% | 12.1% | 15.1% | 13.3% |
| DM4 | 17.3% | 12.5% | 13.1% | 16.4% | 14.3% |

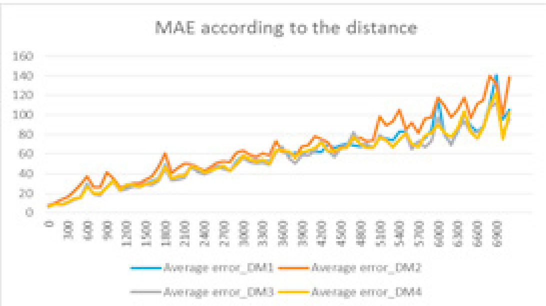


Fig. 5. MAE according to the distance between current location and target stop

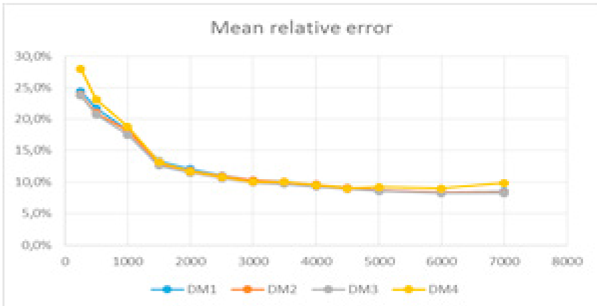


Fig. 6. MRE according to the distance between current location and target stop

UPCP is equal to the ratio of the sum of correct predictions for observed times up to 5, up to 10 and over 10 minutes, and the sum of all the predictions (equation 4).

$$UPCP = \frac{n_p^{5-} + n_p^{5-10} + n_p^{10+}}{n_o^{5-} + n_o^{5-10} + n_o^{10+}} \tag{4}$$

where $n_p^{5-}$, $n_p^{5-10}$, $n_p^{10+}$ are numbers of predictions according to the time-distance of the bus from the target bus stop and $n_o^{5-}$, $n_o^{5-10}$, $n_o^{10+}$ are corresponding predictions.

DM3 and DM4, while DM2 gives the worst results (table 5). The proposed user-perceived model using DM3 gives 95% accuracy of predicted information.

Table 5: Comparison of user perceived corresponding prediction according to the data model and time period

| Data model | E. mor./ l. even. period | Mor-ning period | After-noon period | Week-end period | Total |
|---|---|---|---|---|---|
| DM1 | 94.2% | 95.9% | 92.1% | 96.0% | 94.2% |
| DM2 | 88.2% | 93.7% | 91.6% | 88.3% | 91.1% |
| DM3 | 95.1% | 96.0% | 93.7% | 96.4% | 95.1% |
| DM4 | 94.7% | 96.3% | 92.3% | 95.8% | 94.5% |

## 5. Conclusion

This paper proposes an approach for predicting travel times based on historical data and according to the data model of the bus network. Four different approaches of bus network division were compared. On the basis of 192,000 bus location data points, the results of observed and predicted travel times for 4 different data models, divided into 4 different time periods, were compared. The results showed that the choice of a data network affects the accuracy of the prediction of bus arrival times at the stop in real time. The division of travel times to time periods also affects prediction accuracy. The analysis showed that the predictions are most accurate in the case of DM3, when segments of the network are determined in the middle of two potential time barriers. In the case of DM3, MAE is 51.3 seconds and MRE is 13.3%. Compared to the other specified data models, MAE for DM3 is reduced by 2 to 10 seconds, and MRE by 1% to 3.2%. The share of appropriate predictions for users of public bus services for all DMs exceeds 90%, and the best result is obtained for DM3 with an accuracy of just over 95% for all predictions.

## References


1. Zegeye, G, Thomas, T, Haarsman, G & Vonk, W, 2009, 'Development of a Travel Time prediction algorithm for InfoRio', InfoRio project report.
2. Dziekan, K, L 2008, 'Ease-of-Use in Public Transportation – A User Perspective on Information and Orientation Aspects', Doctoral Thesis in Traffic and Transport Planning, Royal Institute of Technology Stockholm, Sweden.
3. Elragal and Raslan, 2014, Analysis of Trajectory Data in Support of Traffic Management: A Data Mining Approach, Industrial Conference on Data Mining, Advances in Data Mining, Applications and Theoretical Aspects, pp. 174-188.
4. Altinkaya, M & Zontul, M 2013, ' Urban Bus Arrival Time Prediction: A Review of Computational Models', International Journal of Recent Technology and Engineering', vol. 2, pp. 164 – 169.
5. Balasubramanian, P & Rao, K, R, 2015, 'An Adaptive Long-Term Bus Arrival Time Prediction Model with Cyclic Variations', Journal of Public Transportation, vol.18, pp. 1-18.
6. Deng, L, He, Z & Zhong, R, 2013, 'The Bus travel Time Prediction Based On Bayesian Networks', Intrenational Conference on Information Technology and Application, pp. 282 – 285.
7. Chen, M, Liu, X, Xia, J & Chien, S, L 2004, 'A dynamic bus-arrival time prediction model based on APC data', Computer-Aided Civil and Infrastructure Engineering, vol. 19, pp. 364 – 376.
8. Cats, O & Loutos, G, 2016, 'Evaluating the added-value of online bus arrival prediction schemes', Transportation research Part A, vol. 86, pp. 35 – 55.
9. Yu, B, Lam, W, H, K, & Mei Lam Tam, M, L, L 2011, 'Bus arrival time prediction at bus stop with multiple routes', Science direct, Transportation Research Part C: Emerging Technologies, vol. 19, pp. 1157 – 1170.
10. Chen, X, Gong, H & Wang, J, 2012, 'BRT Vehicle Travel Time Prediction Based on SVM and Kalman Filter', Journal of Transportation Systems Engineering and Information Technology, vol. 12, pp. 29 – 34.
11. Zaki, M, Ashour, I, Zorkany, M, Hesham, B, L 2013, 'Online Bus Arrival Time Prediction Using Hybrid Neural Network and Kalman filter Techniques', International Journal of Modern Engineering Research., vol. 3, pp. 2035 – 2041.
12. Chien, S, I, Ding, Y; Wei , G, 2002, 'Dynamic Bus Arrival Prediction with Artificial Neural Networks', Journal of Transportation Engineering, vol. 128, st. 429 – 438.
13. Maiti, S, Pal, A, Pal, A, Chattopadhyay, T, Mukherjee, A, 2014, 'Historical Data based Real Time Prediction of Vehicle Arrival Time', IEEE 17th International Conference on Intelligent Transportation Systems (ITSC).
14. Yang, M, Chen, C, Wang, L, Yan, X, & Zhou, L, 2015, Bus Arrival Time Prediction Using Support Vector Machine with Genetic Algorithm, Neural Network World 3, pp. 205–217.
15. Alejandro, T, L 2006, 'Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services', Science direct, Transportation Research Part C: Emerging Technologies, vol. 30, pp. 239 – 256.
16. Chen, G, Yang, X, An, J & Zhang, D, 2012, 'Bus-Arrival-Time Prediction Models: Link-Based and Section-Based', Journal of Public Transportation, vol.138, pp. 1-18.
17. Xiang, L & Hui, L 2006, 'A trajectory-oriented carriageway-based road network data model, part 1: Background', Geo-spatial Information Science, vol 1, pp. 65 – 70.