A public transport departure time prediction algorithm based on operation strategies and real-time monitoring data

A A Agafonov[1]

[1]Samara National Research University, Moskovskoye shosse 34, Samara, Russia, 443086

Abstract. In this paper, we consider a public transport departure time prediction problem. The problem is considered in two notations: estimation of the mean expected departure time and estimation of the time interval required to ensure a predefined probability to depart on-time. We propose departure time estimation algorithms based on a real-time monitoring data, schedule and operation strategies for the public transport management. We consider a schedule-based holding strategy and a bus bunching prevention strategy. We also propose an algorithm for departure time interval estimation. Numerical tests based on real bus routes in Samara, Russia are used for a comparative study of different departure time prediction algorithms.

1. Introduction

It is a challenge for public transportation agencies to provide reliable service because public transport often operates under conditions of high uncertainty. Variation in travel time caused by different factors of uncertainty, such as real-time traffic situation and traffic congestion, weather conditions, change of passenger flow, driver behavior [1]. The unreliability of the transport schedule reduces the efficiency of the transport infrastructure because road users have to take into account risks of arriving late while planning the route. This is especially important for the passenger transport, where the discrepancy with the schedule can accumulate and lead to the bus bunching problem.

In this paper we consider a passenger transport departure time prediction problem. We compare algorithms based on the schedule, real-traffic data and operation strategies of the public transport management.

There are many papers focused on the similar arrival time prediction problem. Proposed methodologies include

- Regression models [2] that constructed as a regression function from the set of independent variables. Non-parametric regression (NPR) models is a relatively simple method for prediction without the need to estimate parameters. K nearest neighbor (k-NN) methods are one of the most popular NPR methods. Bus travel time prediction models using k-NN was developed in [3, 4].

- Kalman filtering models [5, 6] are an efficient recursive procedure that estimates the future states of dependent variables. In [7] authors developed a path-based model and a link-based model using Kalman filter to predict bus travel times.

- Machine learning models, including artificial neural networks (ANN) and support vector machine (SVM) models. ANN has been reported to be especially useful for finding solutions for complex non-linear problems. In [8] authors proposed two ANN-based models to predict bus arrival time: the link-based ANN and the stop-based ANN. The paper [6] proposed a dynamic algorithm that integrated the ANN model and a Kalman filter-based algorithm. In [9] authors used Bayesian inference theory to combine neural networks. Their results showed that the ANN model outperformed the historical data based model and the regression model in terms of prediction accuracy. SVM is a very specific type of learning algorithm characterized by the capacity control of the decision function, the use of the kernel functions, and the sparse solution. The SVM-based models to predict bus arrival time was developed in [10–12].

- Hybrid methods [13, 14] that combine two or more models to predict arrival time precisely.

However, the described models use the real-time and historical travel time information as the basis to predict arrival time of the public transport and cannot be used directly to predict departure time of the vehicle from the starting station.

Another problem that needs to be considered in the context of this paper is the schedule design and operational management of the public transport [15]. The use of operational control strategies makes it possible to reduce the discrepancy between the scheduled time and the real time of arrival at the control points and prevent the bus bunching on the route. One of the most popular control strategies is the schedule-based holding control strategy: if a bus arrives at a stop earlier than the scheduled time, it is held until the scheduled departure time is reached [16]. Although this strategy can improve schedule adherence, it delays the operation of early buses and causes impatient among passengers on board. Another strategy is a driver schedule recovery strategy that proposes to adjust speed over segments between two consecutive control points to arrive on-time [17, 18]. However, this strategy requires that drivers highly flexible response to the traffic situation and can be limited by traffic conditions.

In this paper, we consider the problem of the public transport departure time prediction from the starting station. The problem is considered in two formulations: estimation of the mean expected departure time and minimizing the travel time interval required to ensure a predefined probability of departure on-time.

The structure of this paper is organized as follows: the second section introduces the basic notations and the problem formulation, describes the departure time prediction algorithms based on real-time traffic data and schedule. In this section, we propose an algorithm based on the operation control strategies. The third section presents the problem formulation of the departure time interval estimation and describes a base algorithm for solving this problem. In the fourth section, we present an experimental study of the algorithms. Lastly, we present our conclusions and recommendations for future work.

2. Mean expected departure time estimation

2.1. Basic notation

Introduce the following notation. Let

V be the set of public transport vehicles;

R be the set of routes;

I_r be the bus run numbers on the route $r \in R$;

DT_i be the departure time from the starting station of the bus with the run number $i \in I_r$ on the route $r \in R$;

DT_i^{sch} be the scheduled departure time from the starting station of the bus $i \in I_r$;

AT_i be the arrival time at the ending terminal of the bus $i \in I_r$;

AT_i^{sch} be the scheduled arrival time at the ending terminal of the bus $i \in I_r$.

Let Δ_{min} be the minimum recovery time at the ending terminal (time between two consecutive runs).

Denote the deviation of the real observed arrival time of the bus with the run number $i \in I_r$ from the scheduled arrival time as Δ_i^{arr}:

$$\Delta_i^{arr} = AT_i - AT_i^{sch}.$$

The waiting time between runs denote as Δ_i^{wait}:

$$\Delta_i^{wait} = DT_j^{sch} - AT_i^{sch},$$

where $(i,j), i \in I_r, j \in I_r$ is the numbers of consecutive runs of the bus $v \in V$ on the route $r \in R$.

The main problem is to predict the departure time from the starting station DT_i, i.e., the following assessment construction:

$$\hat{DT_i}, \forall i \in I_r, \forall r \in R. \tag{1}$$

In the next subsections, we describe algorithms for estimation the departure time $\hat{DT_i}$ based on the real-time monitoring information, schedule, and operation control strategies: a holding control strategy and a bus bunching prevention strategy.

2.2. Monitoring data based algorithm

The monitoring data based departure time prediction algorithm uses the real-time monitoring data taking into account the planned schedule. The main assumption of the algorithm is that the deviation of the arrival time at the ending terminal from the scheduled arrival time will be maintained when the vehicle departs from the starting station. The departure time estimation based on the monitoring data can be described as follows

$$\hat{DT}_j^M = DT_j^{sch} + \Delta_i^{arr}. \tag{2}$$

Obviously, this estimation is quite simple and does not consider any control strategies to decrease the deviation from the scheduled departure time. In the next subsections, we present departure time prediction algorithms based on schedule and operation strategies.

2.3. Schedule based algorithm

The schedule-based holding control strategy is one of the most popular strategies in the bus route schedule design problem. In this strategy, if a bus arrives early at the timing point, it will be held until the scheduled departure time is reached. If the bus is already late, it will depart from the timing point immediately after passenger pickup. We use the similar approach to adapt schedule-based holding strategy to estimate the departure time from the starting terminal.

The schedule-based algorithm with the holding strategy considers the deviation of the arrival time Δ_i^{arr} and the waiting (recovery) time Δ_i^{wait} to control the departure time in the following way:

$$\hat{DT}_j^S = \begin{cases} AT_i + \Delta_{min}, & \Delta_{wait} < 0, \\ DT_j^{sch} + \alpha\Delta_i^{arr}, & \Delta_i^{arr} < 0, \\ DT_j^{sch} + \max\left(0, \beta\Delta_i^{arr} - \gamma\Delta_i^{wait}\right), & \text{otherwise,} \end{cases} \tag{3}$$

where $\alpha \in [0, 1], \beta \in [0, 1], \gamma \in [0, 1]$ are the control strategy coefficients.

The proposed control strategy based algorithm consists of the following steps:

- if the vehicle arrived at the terminal stop later than the scheduled departure time of the next run, then shorten the waiting time to a minimum;
- if the vehicle arrived earlier than the scheduled arrival time, then reduce the deviation from the scheduled departure time by the recovery time increasing;
- otherwise, reduce the deviation from the scheduled departure time by reducing the recovery time.

2.4. Operation strategy based algorithm

The operation strategy based algorithm considers a minimum delay between departure times of the different buses on the same route to prevent bus bunching.

We assume that the minimum delay time between two consecutive runs can be different for different routes and is determined based on the route schedule.

Denote the minimum delay between the departure times as $\Delta_r^{route}, r \in R$.

Then the departure time estimation for the run number $j \in R$ of the vehicle $v \in V$ on the route $r \in R$ can be expressed as follows:

$$\hat{DT}_j^{OS} = \max(\hat{DT}_j^S, DT_k + \Delta_r^{route}), \tag{4}$$

where k is the last run number, $\hat{DT}_j^S$ is the departure time estimation obtained using the schedule-based algorithm.

Experimental study of the described departure time estimation algorithms is presented in section 4.

3. Departure time interval estimation

In the previous section, we describe the mean expected departure time estimation problem. Sometimes this formulation cannot be suitable for passengers. Often it is desirable to know not a time instant of the departure, but a departure time interval required to ensure a predefined probability of departing on-time.

Let τ be the required departure time instant, p be the required probability to depart on-time, r be the selected route.

The departure time interval estimation problem can be expressed in the following form:

$$P(\tau - \Delta^{dep} < DT_i < \tau) > p, \tag{5}$$

where DT_i is the departure time,
Δ^{dep} is the estimated departure time interval.

The base departure time Δ^{dep} estimation algorithm uses traffic statistics. Let $\Pi_r^d = \{DT_i, i \in I_r\}$ be the dataset with the departure time of the vehicles $v \in V$ on the routes $r \in R$ in the selected day d, $\Pi_r = \{\Pi_r^d\}_{d=1}^N$ be the statistics for N days.

We introduce an indicator variable of the following form:

$$p(d, \tau, \Delta) = \begin{cases} 1, & \exists DT_i \in \Pi_r^d : \tau - \Delta < DT_i < \tau; \\ 0, & \text{othewise}, \end{cases} \tag{6}$$

Then the departure time interval estimation algorithm can be expressed in the following form:

Algorithm 1: Departure time interval estimation

$\hat{p} = 0$;
$\Delta = 0$;
while $\hat{p} < p$ **do**
    $\Delta = \Delta + step$; // $step$ - increment step
    $\hat{p} = \sum_{d=1}^{N} p(d, \tau, \Delta)/N$;
end
$\Delta^{dep} = \Delta$;

---

4. Simulation setup and results

An experimental study of the algorithms was carried out on the public transport data in Samara. To estimate the departure time, we chose the 1070 runs of different bus routes, for which the departure / arrival times to the control points according to the schedule are known.

Firstly, we estimate the deviation of the real observed arrival time to the terminal stop from the scheduled arrival time. Figure 1 shows the histogram of the arrival time deviation. The positive deviation indicates the delay in the arrival, and the negative deviation means that the vehicle arrives ahead of schedule. As can be seen from the histogram, vehicles often arrive at the ending terminal later than the scheduled time.
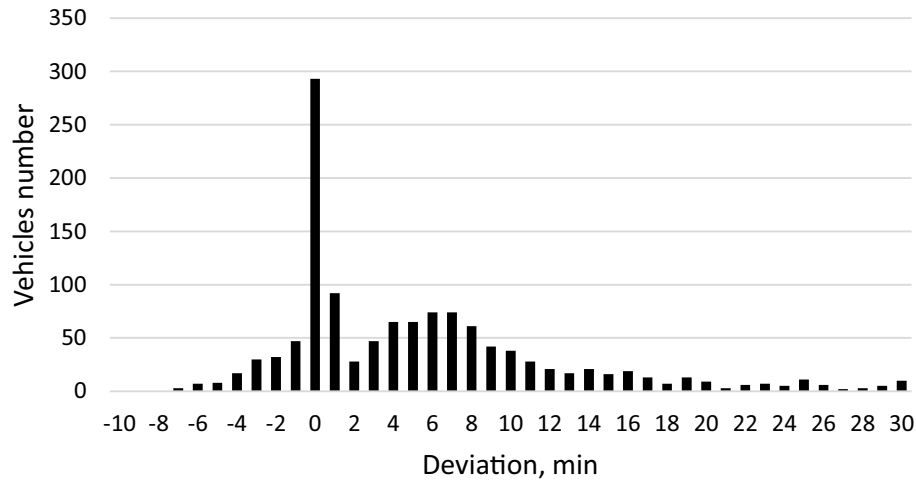

Figure 1. Arrival time deviation

Next, we compare the proposed algorithms based on monitoring data, schedule and operation strategies. Table 1 provides the mean absolute error and standard deviation of the proposed algorithms.

Table 1. Algorithms comparison

	Mean absolute error	Standard deviation
Monitoring data based algorithm	601.0	1147.35
Schedule based algorithm	249.95	385.22
Operation strategies based algorithm	260.85	377.78

The schedule-based algorithm with the holding strategy provides the best results by the selected criteria. Using bus bunching prevention control strategy do not improve the departure time estimation precision.

At the final stage of the experimental study, we estimate the deviation of the real observed departure time from the estimated departure time. Figure 2 depicts the histogram of the departure time deviation. A positive deviation means that the vehicle departed after the estimated departure time.
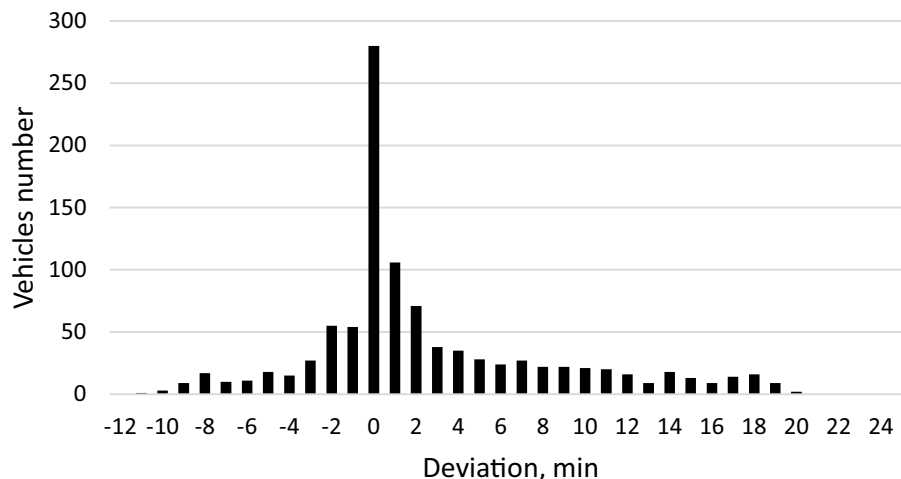

Figure 2. Estimated departure time deviation

As can be seen from the histogram, the schedule-based algorithm estimates the departure time without deviation from the observed time.

5. Conclusion

In this paper, we considered the problem of public transport departure time estimation. We compared the algorithm based on the real-time monitoring, the schedule-based algorithm with the holding operation strategy, and the operation strategy based algorithm with bus bunching prevention strategy. The schedule-based algorithm with the holding control strategy showed the best results by mean absolute error criteria. Using bus bunching prevention strategy did not decrease the error of the estimation.

The main disadvantage of the proposed algorithm is the fact that the traffic statistics are not used to estimate the departure time of the public transport.

In this paper, we considered a public transport departure time prediction problem. The problem is considered in two notations: estimation of the mean expected departure time and estimation of the time interval required to ensure a predefined probability to depart on-time. We compare algorithms based on a real-time traffic data and public transport schedule. The paper proposes an original prediction algorithm based on the operation strategies of the public transport management. Numerical tests based on real bus routes in Samara, Russia are used for a comparative study of different departure time prediction algorithms.

In addition, we consider the problem of the time interval estimation required to ensure a predefined probability to depart on-time. The base algorithm for solving this problem is proposed. The development of more complex algorithms will be conducted as our future research.

6. References

[1] Chen X, Yu L, Zhang Y and Guo J 2009 Analyzing urban bus service reliability at the stop, route, and network levels *Transportation Research Part A: Policy and Practice* **43** 722-734

[2] Agafonov A, Sergeev A and Chernov A 2012 Forecasting of the motion parameters of city transport by satellite monitoring data *Computer Optics* **36** 453-458

[3] Chang H, Park D, Lee S, Lee H and Baek S 2010 Dynamic multi-interval bus travel time prediction using bus transit data *Transportmetrica* **6** 19-38

[4] Smith B, Williams B and Keith Oswald R 2002 Comparison of parametric and nonparametric models for traffic flow forecasting *Transportation Research Part C: Emerging Technologies* **10** 303-321

[5] Vanajakshi L, Subramanian S and Sivanandan R 2009 Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses *IET Intelligent Transport Systems* **3** 1-9

[6] Chen M, Liu X, Xia J and Chien S 2004 A dynamic bus-arrival time prediction model based on apc data *Computer-Aided Civil and Infrastructure Engineering* **19** 364-376

[7] Chien S J and Kuchipudi C 2003 Dynamic travel time prediction with real-time and historic data *Journal of Transportation Engineering* **129** 608-616

[8] Chien S J, Ding Y and Wei C 2002 Dynamic bus arrival time prediction with artificial neural networks *Journal of Transportation Engineering* **128** 429-438

[9] van Hinsbergen C, van Lint J and van Zuylen H 2009 Bayesian committee of neural networks to predict travel times with confidence intervals *Transportation Research Part C: Emerging Technologies* **17** 498-509

[10] Yu B, Lam W and Tam M 2011 Bus arrival time prediction at bus stop with multiple routes *Transportation Research Part C: Emerging Technologies* **19** 1157-1170

[11] Bin Y, Zhongzhen Y and Baozhen Y 2006 Bus arrival time prediction using support vector machines *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* **10** 151-158

[12] Yu B, Yang Z and Wang J 2010 Bus travel-time prediction based on bus speed *Proceedings of the Institution of Civil Engineers: Transport* **163** 3-7

[13] Park T and Lee S 2004 A bayesian approach for estimating link travel time on urban arterial road network *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **3043** 1017-1025

[14] Agafonov A and Myasnikov V 2014 An algorithm for city transport arrival time estimation using adaptive elementary predictions composition *Computer Optics* **38**(2) 356-368

[15] Wu Y, Tang J and Luo X 2015 Comparative analysis of operation strategies in schedule design for a fixed bus route *International Transactions in Operational Research* **22** 545-562

[16] Cats O, Larijani A, Ólafsdóttir Á, Burghout W, Andréasson I and Koutsopoulos H 2012 Bus-holding control strategies *Transportation Research Record* 100-108

[17] Kalaputapu R and Demetsky M J 1995 Modeling schedule deviations of buses using automatic vehicle-location data and artificial neural networks *Transportation Research Record* 44-52

[18] Yan Y, Meng Q, Wang S and Guo X 2012 Robust optimization model of schedule design for a fixed bus route *Transportation Research Part C: Emerging Technologies* **25** 113-121