Reliability prediction of further transit service based on support vector machine

Xiaoning Gu¹, Chao Chen¹, Yunong Yang², Xingzhi Miao¹ and Baozhen Yao¹

Abstract

The requirement for transit reliability grows with the increase of pace of life since unstable bus arrivals can raise the anxiety of waiting passengers. This paper proposes a reliability assessment method to evaluate the reliability of each bus stop on the route and the reliability of bus routes. In reliability prediction, the prediction target is locked by rolling horizon to reduce the interference of other information. In addition, a prediction method of the reliability of further transit service using the accurate online support vector machine is proposed. This prediction can provide more accurate and stable data for the arrival of buses and reduce unnecessary waiting of passengers. Finally, the reliability prediction method proposed is tested with the real data of a bus route in Dalian, China. The results show that the accurate online support vector machine with reasonable parameters can predict the reliability of transit service accurately.

Keywords

Reliability prediction, support vector machine, public transport

Date received: 31 August 2017; accepted: 26 May 2019

Introduction

Traffic congestion has been a common problem in large cities of the world. In order to effectively alleviate traffic congestion, the urban public transit system has been pushed into a very important position. However, in many large and medium-sized cities of China, the development of public transit is not effective because of the pressure and difficulty from the large numbers of citizens. The specific drawbacks are the poor punctuality of buses, uneven headway between buses, and the appearance of bus bunching and large spare space. With the increase of pace of life, travelers are willing to arrive at their destinations quickly and promptly. In other words, reaching the destination within the expected time (i.e. the reliability of buses) is very important. Moreover, in many large cities, it is common that there might be no timetable of buses at each bus stop because traffic conditions usually change quickly, which makes it hard for buses to follow timetables. However, although timetable of buses might not exist, buses still have schedule at the beginning bus stops. Considering the actual condition of these timetable-lacking buses, their reliability seems to be more important owing to the anxiety of waiting passengers without an approximate waiting time. Therefore, concentrating on the reliability of public transit is helpful to evaluate and further improve public transit service.

The research on the reliability theory on road traffic networks has developed rapidly since the 1980s, and the main aspects include network connectivity reliability, travel time reliability, capacity reliability, waiting time reliability, and so on. Abkowitz investigated the San Francisco Bay Area to study travel time reliability, and the result showed that the travel time reliability has a significant impact on the departure-time choice of commuters. Asakura and Kashiwadami used a simulation-based method to examine the impact of variability in O-D demand levels. Strathman et al. used the headway ratio, run time ratio, and average waiting time to evaluate the bus reliability. From the passenger’s perspective, Bruinsma et al. regarded travel time reliability as an index for the evaluation of

¹School of Automotive Engineering, Dalian University of Technology, Dalian, P.R. China
²Transportation Management College, Dalian Maritime University, Dalian, P.R. China

Corresponding author:
Baozhen Yao, School of Automotive Engineering, Dalian University of Technology, Dalian 116024, P.R. China.
Email: yaobaozhen@hotmail.com
transit service reliability. Taking into account the interactions between network performance and passengers’ route choice behavior, Yin et al.\textsuperscript{20} used the Monte Carlo simulation approach with stochastic user equilibrium transit assignment model to quantify the reliability of transit running time, schedule reliability, and stop waiting time reliability. Liang et al.\textsuperscript{21} proposed unblocked reliability by defining that travelers can travel under a specified level of service in transit network.

As can be seen in the literature, studies mainly focus on the assessment of current transport reliability, rather than the prediction of transit reliability. Xuan et al.\textsuperscript{22} recognized the importance of public transportation stability and conducted a reliability analysis of the bus scheduling problem to reduce bus congestion and improve the possibility of on-time arrival schedule. The prediction, in the field of urban transport, mainly involves two aspects: traffic volume prediction and bus arrival time prediction. The related prediction methods could be roughly divided into two classes: traditional mathematics prediction methods and data-driven approach. The traditional mathematics prediction methods include historical average model, time series model, Kalman filter model, and so on, while the data-driven approach includes neural networks, nonparametric regression, $k$-nearest neighbor algorithm ($k$-NN), and so on. Kim and Hobeika\textsuperscript{23} applied the autoregressive integrated moving average model to the freeway volume forecasting. Chen et al.\textsuperscript{24} analyzed the Kalman filter theory and then established a short-term traffic volume forecast model based on the Kalman filter theory. Chien and Kuchipudi\textsuperscript{25} developed a path-based model and a link-based model using a Kalman filter to predict bus travel times. Chien et al.\textsuperscript{26} proposed two artificial neural network (ANN)-based models, the link-based ANN model, and the stop-based ANN model, to predict bus arrival time. Taking the city of Jinan, China, as an example, and based on the historical GPS data and AFC system data, Lin et al.\textsuperscript{27} proposed two ANN models to predict the real-time bus arrivals. Yu et al.\textsuperscript{28} adopted several methods, including support vector machine (SVM), ANN, $k$-NN, and linear regression (LR), for the bus arrival time prediction and found that the SVM model performs the best among the four proposed models. In spite of numerous studies about traffic flows and running time prediction, the research on transit reliability prediction is rare, which is mainly discussed in this paper.

SVM can solve machine learning problems in classification and induction with good generalization ability. SVM has global optimal characteristics and the advantage of not falling into local optimal,\textsuperscript{29,30} which makes it not requiring to perform a complex nonlinear optimization. SVM maps data from nonlinear low-dimensional spaces to higher dimensional linear spaces and searches for the optimal LR hyperplane algorithm to solve the convex programming problem with convex constraints, so as to obtain the global optimal solution.\textsuperscript{31,32} In the aspect of prediction methods, SVM performs well at the prediction of financial time series, short-term wind speed, traffic flow, and so on. In order to make SVM better solve some practical problems, researchers modified the objective function of quadratic programming to construct a number of SVMs with new properties, for example, $v$-SVM, BSVM, One-Class SVM, RSVM, Ls-SVM, WSVM, and so on.

In addition, since Vapnik proposed SVM, the studies have proved that the generalization performance of SVM did not depend on all the training data, but support vectors. The reason is that the number of support vectors is smaller compared with the number of entire training data sets, which provides the possibility of incremental learning. Syed et al.\textsuperscript{42} first proposed an incremental algorithm based on SVM. Mitra et al.\textsuperscript{43} put forward to achieve incremental learning through the error-drive technology. Similar to Syed, they abandoned non-support vectors at each training, but the difference is that the training set consists of a support vector set and a misclassified sample set, and the SVM model was retrained. Poggio and Cauwenberghs\textsuperscript{44} developed a method to incrementally obtain exact solutions for global optimization problems, increasing or decreasing the impact of a support sample on the Lagrange coefficient and support vector. Considering the nuclear localized features of radial basis functions (RBFs), Ralaivola and d’Alché-Buc\textsuperscript{45} applied incremental learning based on the characteristics of data locality and put the data adjacent to new data into a training set to correct its original classification criteria. Ma et al.\textsuperscript{46} proposed an accurate online support vector machine (AOSVM) algorithm, which could learn online and update model. This method was proved to be suitable for solving issues like large changes of time series data property, frequent model updates, and so on. Thus, it is well applied for online analysis and prediction.

Research on transit reliability can provide travelers with more accurate travel information. Thus, unexpected delays can be avoided if travel time is reasonably arranged. For operating companies, the operational level can be improved effectively, and the available resources can be adequately distributed at the maximum degree. In addition, the reliability is also a decision indicator for transit network planning and optimization, which ensure that public transport has a higher service level and stronger marketplace position in the future.

The purpose of this work is to enrich the study on the prediction of transit reliability, especially for nonschedule bus routes and bus stops. Therefore, a specific measure of reliability is put forward for the prediction of future reliability. The machine learning method is adopted in this paper, that is, the idea of incremental learning is adopted, and the accurate online support vector machine (AOSVM) is used to predict the reliability of public transit (buses).
The remainder of this paper is organized as follows. Section “Model development” shows the description of the problem and develops a model to predict the transit reliability of buses using AOSVM. In section “Accurate online support vector machine,” we present the principle of AOSVM, as well as the process of algorithm. A case study of line 10 in Dalian, China, is presented in section “Case study.” And finally, the conclusions are provided in section “Conclusion.”

Model development

Problem description

Getting to the destination within the expected time (i.e. the reliability of the bus) is becoming increasingly important. In order to improve bus reliability, this paper first establishes a bus reliability evaluation model, which includes the reliability assessment of bus stops and bus routes. Then, the rolling horizon is used to select the target and the prediction function is established to predict the reliability of transit.

Transit reliability assessment

Stop reliability assessment. Reliability of bus stops can be seen as the property of fairness of each bus stop in this paper. That is, the time interval between two successive bus stops is used to measure stop reliability and fair arrival (same arrival interval) at a stop leads to a high stop reliability. The level of the reliability of a bus stop should regard the actual time interval between two successive vehicles arriving at the same stop as the basis of reliability evaluation. Theoretically, the service is good if the actual time interval between two successive vehicles arriving at the same stop is equal to the departure interval. However, affected by weather, traffic accidents, and other random factors (e.g. short-term large traffic volume, road pavement maintenance), the time intervals between two successive vehicles arriving at the same stop are actually fluctuating. And the higher the volatility of bus arrival time, the worse the bus reliability. Thus, by evaluating the volatility of bus arrival at each stop, the bus operation reliability can be described. In this paper, $F_j$ is used to define the fluctuation degree of bus stops during the operation process in a certain period of time

$$F_j = \frac{\sqrt{\sigma_j^2}}{\bar{h}_j}$$  \hspace{1cm} (1)$$

$$\sigma_j^2 = \frac{\sum_{i \in \Omega_k} (h_{i,j} - \bar{h}_j)^2}{k}$$  \hspace{1cm} (2)$$

$$h_{i,j} = t_{i,j} - t_{i-1}$$  \hspace{1cm} (3)$$

$$\bar{h}_j = \frac{\sum_{i \in \Omega_k} h_{i,j}}{k}$$  \hspace{1cm} (4)$$

where $h_{i,j}$ is the actual time interval between bus $i$ and bus $i-1$ arriving at stop $j$, and $k$ is the total trip number in a certain period of time, which is called rolling horizon in this paper. That is, only the information of the $k$ buses is used to assess the arrival interval volatility of stops, while the information beyond the rolling horizon is skipped. An illustration of rolling horizon of buses can be seen in Figure 1.

As can be seen from formula (1), a smaller fluctuation coefficient means the bus line operation at the stop is more stable and the reliability is better. Particularly, the fluctuation coefficient $F_j = 0$ indicates that during this time period, the actual arrival interval of buses at the stop is strictly equal to each other, and buses are operated in a stable status. In other words, for passengers, they will feel fairer to a certain extent resulting from the same bus arrival interval. Considering the short headway and unstable operation of buses in reality, the range of fluctuation coefficient usually belongs to $[1, +\infty]$. However, the reliability should be a bounded value in the range of $[0, 1]$. Therefore, a method is needed to convert the value of $F_j$ to the range $[0, 1]$. Finally, a reliability index $\alpha_j$ is defined in this paper and shown as follows

$$\alpha_j = \frac{1}{F_j}$$  \hspace{1cm} (5)$$

The reliability of transit service of bus route. By summing the reliability of each stop on a bus route according to a certain weight, the reliability of the bus route is finally obtained. Essentially, the weights measure the
importance of a single stop on the route. In a certain period of time, the bus stop plays an important role in the bus route if the passenger flow is large. Thus, the proportion of passenger flow is taken as the weight of each bus stop. Noting that, for the sake of simplification, the entering and exiting times of passengers are set as the same, the reliability of the bus route $\beta$ could be calculated as follows

$$\beta = \sum_{j=1}^{Q} \frac{q_j}{Q} \times \alpha_j$$  \hspace{1cm} (6)

where $q_j$ denotes the number of passengers at the stop $j$ and $Q$ denotes the total number of passengers at all the stops of the bus route.

**Prediction of transit reliability**

With the help of the GPS technology, the bus arrival time at each stop can be easily obtained. Thus, the reliability of the bus route is available according to the formulas in section “Problem description.” However, the future reliability of the bus route is of great significance for both bus operating and passenger travel choice, especially in short term. Therefore, it is necessary to predict future reliability of bus route. To predict the reliability of the further transit service, the potential relation between the current and further transit services should be deduced. In this paper, the SVM method is adopted to formulate the reliability of the future transit service based on the reliability of the current and recent transit services

$$\hat{\beta}_{i+w} = f(\beta_i, \beta_{i-1}, \beta_{i-2}, ..., \beta_{i-k+1})$$  \hspace{1cm} (7)

where $\hat{\beta}_{i+w}$ denotes the prediction value of the reliability of the bus stop when bus $i+w$ arrives at the stop in the future, $\beta_i, \beta_{i-1}, \beta_{i-2}, ..., \beta_{i-k+1}$ refer to the observed (calculated) reliability value of the former $k$ buses. $f(\cdot)$ denotes the reliability prediction function. Figure 2 shows an example of the prediction of future reliability of transit service. $\beta_{i+w}$ is predicted when bus $i$ arrives at the predicted bus stop, and $\beta_{i+1}, ..., \beta_{i+w}$ are the future reliability of transit service.

**Accurate online support vector machine**

**The principle of AOSVM**

SVM is a kind of novel machine learning method based on statistical learning theory proposed by Vapnik. SVM can well solve the small sample, nonlinear, and other practical problems. SVM has the advantages of finding global solutions and higher generalization capability. SVM is developed from the optimal classification hyperplane when the training data is linear separable. Its essence is to find that the support vector can be used to construct the optimal classification hyperplane from the training data.

The main principle of SVM is to choose a nonlinear mapping $\Phi(x)$ at first, which maps $n$-dimensional vector samples to a higher dimensional feature space from the input space, and then construct linear decision function $f(x) = w \cdot \Phi(x) + b$ in the high-dimensional feature space using the maximum interval method.

Supposing the training set $T = \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ is known, where $x_i \in \mathcal{X} \subseteq \mathbb{R}^n$, $y_i \in \mathcal{Y} \subseteq \mathbb{R}$, $i = 1, 2, ..., l$, a regression function can be obtained using the SVM method

$$f(x) = w \cdot \Phi(x) + b$$  \hspace{1cm} (8)

where $w \in \mathbb{R}^n, b \in \mathbb{R}, \Phi(x)$ is a mapping function that maps the input vector $x$ into a feature space vector, and $w$ and $b$ are the weight vector and bias, respectively, which can be determined by solving the following optimization problem.

As shown in Figure 3, circles and squares, respectively, represent samples of two different categories, $\mathcal{H}$ represents the hyperplane that correctly separates the two types of samples, and its direction is represented by hyperplane normal vector. $H1$ and $H2$ represent the plane that is parallel to the hyperplane. $H1$ and $H2$ pass through each type of samples and is closest to hyperplane $H$. The distance between $H1$ and $H2$ is called classification interval. The optimal classification hyperplane refers to the fact that the obtained hyperplane can not only correctly separate the two types of samples and minimize the model training error but also maximize the classification interval between the two types.

Make $|f(x)| > 1$ with normalizing, while the sample nearest to the plane of classification $|f(x)| = 1$. The classification interval is $(2/||w||)$. If the required classification interval is the largest, then $||w||$ or $||w||^2$ is required to be the smallest. And if the classification is required to correctly classify all samples, then the following requirements are met

$$\min R(w, b) = \frac{1}{2} w^T w$$  \hspace{1cm} s.t. $y_i - w \Phi(x_i) + b \leq \epsilon$  \hspace{1cm} $w \Phi(x_i) + b - y_i \leq \epsilon$  \hspace{1cm} $i = 1, 2, ..., l$

The relaxation factors $\xi_i$ and $\xi_i^*$ are introduced, which represent the deviation from $\epsilon$ tube, $C$ is the penalty
When training samples of the traditional SVM reach a certain level, in which the computational memory exceeds the limited memory size, the training results cannot be obtained. Therefore, the practical and effective online SVM, which can meet the requirement of large-scale problems and online applications, is recommended.

The core of the incremental learning in online SVM is that, when the regression data set sample is updated, the new sample will be added to the training set, and the algorithm dynamically updates the trained SVM model. The basic idea is that when the sample set changes, update three coefficients \( \theta_S, \theta_E, \) and \( \theta_R \) in each set of training samples in equation (12) directly.

Based on the original training results, AOSVM updates the model online according to sample set change. There are five conditions in equation (12), but they can be identified into three categories.\(^{46,51}\) The three categories are separated according to formula (14)

\[
S = \{i | ((\theta_i \in [0, +C]) \land h(x_i) = -\varepsilon) \lor ((\theta_i \in [-C, 0]) \land h(x_i) = +\varepsilon) \}
\]

\[
E = \{i | ((\theta_i = -C) \land h(x_i) \geq +\varepsilon) \lor ((\theta_i = +C) \land h(x_i) \leq -\varepsilon) \}
\]

\[
R = \{i | (\theta_i = 0) \land |h(x_i)| \leq \varepsilon\}
\]

where \( S \) is the support vector set, \( E \) is the error set, and \( R \) is the reserved set. Different from the offline SVM, the AOSVM directly increases or removes one sample data instead of restarting training when the data are updated and then adjusts SVM structure and data characteristics of the corresponding training data set online dynamically, making SVM corresponding to KKT conditions.

In the incremental algorithm, a new sample \( x_c \) is hoped to enter one of the three sets of formula (13) while being able to satisfy the KKT condition, when adding \( x_c \) into sample.\(^{52}\) The Lagrange multiplier of the new added data is set as \( \theta = 0 \). Then, equation (13) can be obtained as

\[
h(x_c) = f(x_c) - y_c = \sum_{i=1}^{N} \theta_i K(x_i, x_c) + b - y_c
\]

The value of Lagrange multiplier of the new added data \( \theta_i \) will be gradually updated to satisfy the KKT condition (so as to enter one of the three sets). When adding a new sample, the values of \( \theta_i \) and \( h(x_i) \) in equation (13) may change according to the adjustment of Lagrange parameter of the new added data.\(^{51}\) Therefore, the sample sets \( S, E, \) and \( R \) previously learned may also change, and the progress is illustrated in Figure 4.

The farthest historical sample points in the data samples of \( S, E, \) and \( R \) sets are selected to “forget” in the decremental algorithm.\(^{53}\) When removing an existing sample from the training set, the decremental algorithm is used to eliminate the old sample. That is, a trained coefficient, \( \varepsilon \) is for nonsensitive coefficient. The above formula can be expressed as

\[
\min R(w, b) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)
\]

\[\text{s.t.} y_i - w^T \Phi(x_i) + b \leq \varepsilon + \xi_i \]

\[w^T \Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \]

\[\xi_i \geq 0, \xi_i^* \geq 0, \quad i = 1, 2, ..., l\]

where \( \xi_i \) and \( \xi_i^* \) are slack variables whose specific values are determined by the following formula

\[\xi_i^{(*)} = \begin{cases} 0, & f(x_i) - y_i < \varepsilon \\ |f(x_i) - y_i| - \varepsilon, & f(x_i) - y_i \geq \varepsilon \end{cases}\]

Formula (9) can be written as formula (11) by introducing Lagrange multipliers

\[f(x) = \sum_{i=1}^{l} (a_i - a_i^*) K(x_i, x) + b\]

where \( K(x_i, x) = \Phi(x_i)^T \Phi(x) \) is the kernel function,\(^{50}\) and \( a_i, a_i^* \) are the corresponding multipliers. According to the Karush–Kuhn–Tucker (KKT) theorem, the following can be obtained

\[
\begin{cases}
  h(x_i) \geq \varepsilon & \theta_i = -C \\
  h(x_i) = \varepsilon & -C < \theta_i = 0 \\
  -\varepsilon \leq h(x_i) \leq \varepsilon & \theta_i = 0 \\
  h(x_i) = -\varepsilon & 0 < \theta_i = C \\
  h(x_i) \geq -\varepsilon & \theta_i = C
\end{cases}
\]

where \( \theta_i = a_i - a_i^* \) and

\[h(x_i) = f(x_i) - y_i = \sum_{i=1}^{l} \theta_i K(x_i, x) + b - y_i\]
sample in $S$ set which does not contribute to the model will be removed from the training set. The decremental algorithm follows the incremental algorithm with a few small adjustments.

Detailed information on the matrix deduction in the incremental and decremental algorithms is available in the literature\textsuperscript{46,52} (Ma et al. 2003; Wang et al. 2009). Through the update rules, $\theta_i$ and $h(x_i)$ can be adjusted via $\Delta \theta_i$.\textsuperscript{46}

The process of algorithm

In the incremental algorithm, when a new sample $(x_c, y_c)$ is introduced, the basic learning procedure is to change the $\theta_c$ corresponding to the new sample data in an infinite number of discrete steps, until $x_c$ meets the KKT conditions.\textsuperscript{46,48} In addition, ensure that the existing samples in the training set obey the KKT conditions at each step. Figure 5 shows the flowchart of the algorithm. The general steps of the incremental algorithm include:

Step (1) Set $\theta_c = 0$.
Step (2) $\Delta \theta_c$ is determined according to the algorithm described in Kivinen et al.\textsuperscript{54}
Step (3) Update $\theta_i$ for all training data sets, and simultaneously update the three subsets $E$, $S$, and $R$.
Step (4) If $(x_c, y_c) \notin E$, then go back to step (2).

In the decremental algorithm, when a trained sample $(x_o, y_o)$ is removed from the training set, a decremental algorithm is employed to adjust its coefficient to zero, while ensuring all the other samples in the training set continue to satisfy the KKT conditions. The general steps of the decremental algorithm include:

Step (1) If $(x_o, y_o) \in R$, directly remove it and return.
Step (2) If $(x_o, y_o) \in E$ or $S$, remove it and start a calculation circle in a reverse direction, with a decremental algorithm that ends until the coefficient reduces to zero and the KKT conditions are still satisfied in the remaining samples.

Case study

In this paper, the proposed model and algorithm are tested based on the real-world data in Dalian, China (i.e. GPS data of the bus route 10). The raw data were collected from public transport providers, including bus Id, time, current location (latitude and longitude), heading, average speed, number of boarding/alighting passengers at bus stops, and so on. And GPS data include bus Id, time, current location (latitude and longitude), heading, and average speed. GPS is always accurate in general, but in the process of data acquisition, it may be affected by a variety of factors and produce some data with large errors or wrong data. In order to improve the accuracy of the research results, it is necessary to preprocess GPS data and delete the inaccurate or incorrect data, including incorrect time format, data field missing, latitude and longitude range crossing, data redundancy, and so on.

Bus route 10 is from Baihe village to Shahekou station with a total length of 14.5 km and 28 stations (see Figure 6). There are 11,729 groups of valid data from...
6:00 to 20:00 during 1 September to 10 September 2015. The input variable of SVM is the recent processed data, and the objective/output is the future reliability. Thus, the time interval (headway) arriving at each stop between two successive buses can be calculated and measured in bus reliability. In order to avoid overfitting of the model in the training process, 80% of the data were randomly selected as the training data. Original sample data are divided into two subsets: training sample set and test sample set, where the test samples are set as 20% (about 2.4 thousand groups) and the rest are set as the training samples.

Headway of buses is the essential factor that can directly represent the level of transit service (reliability). Figure 7 shows a boxplot of headway with minimum, first-quartile, median, third-quartile, and maximum value of headway at different bus stops. Figure 7 shows that the average headway at all stops maintains 125 s, and the headway has a trend that increases with the sequence of bus stops. In other words, bus stops near the terminal usually have higher headway variation than that of bus stops near the originating station. This higher variation results from the accumulation and progressive increase of abnormal running, such as traffic congestion and so on. Figure 8 gives an example of headways of five buses on line 10 during the peak hours.

From Figure 8, it can be seen that there is a bunching between buses 4 and 5 since headway of bus 5 is small (close to 0). The operational statuses of some buses are strange, such as buses 3 and 4. By observing the small headways between buses 2 and 3, it can be concluded that bus 2 is always following bus 3 from stop 8 to 16. It shows a high possibility of passenger accumulation after buses 2 and 3, and leaving the “accumulated waiting passengers” to the next following bus (i.e. bus 4). Therefore, we can observe the large headways of bus 4 from stop 8 to stop 20, resulting from larger passenger flows and propagation of this “negative” effect.

To find an AOSVM with good features-learning capacity, two main steps need to be preprocessed: (1) selection of the kernel function and (2) selection of parameters in SVM. The kernel parameter, C, and epsilon were determined via grid search algorithm. The kernel RBF shows advantages in bus running time prediction. Considering that SVM prediction in this paper is related to bus running, RBF is selected as the kernel function. In the literature, there are many methods to calibrate the parameters (C, e, o) of the RBF, in which the most common and reliable one is the grid search. Using this approach, parameters in this paper are set as: (2^{-3}, 2^{-6}, 0, 4).

In order to verify the stability of the algorithm, the proposed AOSVM is used to predict for 10 times. The performance of the model is evaluated with two metrics: mean absolute percent error (MAPE) and root mean square error (RMSE). The formulas of MAPE and RMSE can be seen in formulas (16) and (17), and the prediction results are shown in Figure 9.

\[
MAPE(\%) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \hat{Y}_i}{Y_i} \right)^2 \times 100\%
\] (16)
As we can see in Figure 9, the MAPE of 10 times calculation is lower than 15%, and RMSE has a mean value of about 0.08. The results illustrate that the AOSVM has a good stability and is feasible to predict the transit reliability.

In order to verify the accuracy of AOSVM, we use AOSVM, classic SVM (Cl-SVM), and GA-SVM, respectively, to predict the reliability of bus route 10. The performances (if data updates) of different methods with limited running time are listed in Table 1. And in order to verify the influence of different methods of searching kernel parameters on the results of AOSVM, the kernel parameters of AOSVM are searched via both GA and grid search algorithm. And GA-SVM (SVM of which the kernel parameters searched by GA) is also compared with GS-SVM (SVM of which the kernel parameters are searched by grid search). Actually, the grid search is a kind of traversal algorithm and gets through all the parameters. As the result, the results obtained by the grid search method are better than those of the GA. However, the convergence speed of GA is much faster than that of the grid search. And the performances of different search algorithms with limited running time are summarized in Table 1. In this running-time-limit procedure, we substituted 5% (about 400 samples) of the training data set to inspect samples for 10 times with each time 0.5%. The running time is limited to 5 min with a dual-core processor (3.2 GHz) and 4G memory, and the procedure will stop if time-out, no matter where the progress bar (no. of the substitution, second or ninth substitution) is.

As we can see from Table 1, the MAPE or RMSE of GA-SVM is smaller than the Cl-SVM, and the performance of the classic SVM is the worst. It is because that in the GA-SVM technology, the parameters of SVM are optimized to reduce the margin of error. Considering the performance of AOSVM, we will find that the MAPE of AOSVM is obviously smaller than that of both Cl-SVM and GA-SVM. Owing to the strategy of adding new sample data to provide better support vector and updating the model online, the prediction accuracy of AOSVM is greatly improved.

The results of AOSVM are also compared with other three models: k-NN (k-nearest neighbors), ANN, and LR. The main reasons of selecting these three models for comparison are that LR is a classic method for regression with mathematical formulation, ANN is a well-known method in aspect of machine learning, and k-NN is an effective non-parameter method in prediction. In order to ensure fairness and scientificity when comparing the models, the same input data and output variable are used in the training and testing of different models. The parameter k in the k-NN model\textsuperscript{56} in this paper varies from 1 to 5 and is finally set as 4 according to a comparison. When k = 1, 2, 3, or 5, the accuracy of k-NN is worse than that when k = 4. Therefore, we adopt the best parameter (k = 4) for the further

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \hat{Y}_i}{Y_i} \right)^2}
\]  

(17)

![Figure 9. Prediction results using AOSVM for 10 times.](image)

Table 1. Comparison of performances among AOSVM, GA-SVM, GS-SVM, and Cl-SVM with limited running time.

| Predict time       | Algorithm | Search algorithm | Evaluation index |
|-------------------|-----------|------------------|-----------------|
|                   |           |                  | MAPE | RMSE  |
| 8 September 2015  | Cl-SVM    | Empirical selection | 9.99% | 0.086 |
|                   | GA-SVM    | GA search        | 9.88% | 0.074 |
|                   | GS-SVM    | Grid search      | 8.78% | 0.070 |
|                   | AOSVM     | GA search        | 7.32% | 0.062 |
|                   |           | Grid search      | 6.29% | 0.057 |
| 9 September 2015  | Cl-SVM    | Empirical selection | 8.78% | 0.068 |
|                   | GA-SVM    | GA search        | 8.66% | 0.066 |
|                   | GS-SVM    | Grid search      | 7.54% | 0.063 |
|                   | AOSVM     | GA search        | 6.33% | 0.059 |
|                   |           | Grid search      | 5.24% | 0.052 |
| 10 September 2015 | Cl-SVM    | Empirical selection | 10.87% | 0.094 |
|                   | GA-SVM    | GA search        | 10.12%| 0.092 |
|                   | GS-SVM    | Grid search      | 9.42% | 0.090 |
|                   | AOSVM     | GA search        | 9.46% | 0.084 |
|                   |           | Grid search      | 8.94% | 0.069 |

AOSVM: accurate online support vector machine; GA-SVM: genetic algorithm support vector machine; GS-SVM: grid search support vector machine; Cl-SVM: classic support vector machine; MAPE: mean absolute percent error; RMSE: root mean square error.
comparison with other models. And a standard three-
layer ANN\(^5\) is used to predict the bus stop reliability.
The number of hidden units are selected according to
experience, generally set as 75\% of the number of input
layer nodes. The number of hidden units in ANN
model in this paper varies from four to six and is finally
set as five according to a comparison. As seen in Figure
10, AOSVM has the highest accuracy of prediction, fol-
lowed by ANN, while LR performs the worst. Using
AOSVM to predict the reliability of bus line 10 during
a week, the results are shown in Figure 11.

As we can see in Figure 11, the level of the reliability
of public transport is around 0.8 in workdays. However,
the reliability at the weekend is lower. This is
because at the weekend, the passenger flows (Figure 12
shows passenger flows at a bus stop on various days)
are relatively larger than workdays, which brings more
uncertainty to public transit service. In peak times,
the reliability of transit also shows the same trend and
the reliability is lower than that in off-peak time. Thus,
we can conclude that there is a great potential to improve
the reliability of the public transit.

Concerning on the reliability of bus stops, we use
the AOSVM method to predict the reliability of stops
in morning peak (6:30–8:00) and evening peak (4:30–
6:30). Noting that the directions of the upstream and
downstream are handled separately. The prediction
results are shown in four different scenarios: morning
and upstream (MU), morning and downstream (MD),
evening and upstream (EU), and evening and down-
downstream (ED). The results are shown in Figure 12.

As Figure 13 shows, the reliability reduces in the
direction of upstream, while it increases in the opposite
direction. And it can be seen that the transit reliabilities
during morning peak and evening peak are similar.
Furthermore, in both directions, the reliability during
morning peak is lower than that during evening peak.
Because of the bottleneck of traffic, the reliability of
bus stop 6 is obviously low in both MU and EU. In
contrast, the road line 18–19 is straight and smooth,
making the reliability of bus stop 19 increase from 0.45
to 0.5.

**Conclusion**

In this paper, we proposed a reliability assessment
method to obtain the reliability of bus stops and bus
route service level. And then a reliability prediction
method for future transit service using AOSVM was
proposed. Studying the reliability of public transit can
support the improvement of the transit service and
shorten the actual travel time of passengers. Taking
bus route 10 of Dalian as an example, the reliability

---

**Figure 10.** Performance of AOSVM, k-NN, ANN, and LR on predicting bus stop reliability.

**Figure 11.** Reliability prediction of transit during a week using AOSVM.

**Figure 12.** Passenger flows at bus stop 2 in different days (upstream and downstream).

**Figure 13.** Reliability prediction of bus stops during peak using AOSVM.
prediction method proposed in this paper is verified. The prediction of bus reliability can improve the bus service. Better bus reliability prediction can improve the prediction accuracy of bus arrival time and provide better services for passengers.

Based on the above research, herein, we further studied transit service reliability. However, there are still some areas that have not been studied. In the future work, further research and analysis can be made, including the impact and treatment measures under the condition of insufficient reliability.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Natural Science Foundation of China 51578112 and The State Key Laboratory of Structural Analysis for Industrial Equipment S18307.

ORCID iD
Baozhen Yao https://orcid.org/0000-0002-8508-8632

References
1. Mine H and Kawai H. Mathematics for reliability analysis. Tokyo, Japan: Asakura-Shoten, 1982.
2. Bin Y, Wenzuan S, Zhen G, et al. A heuristic algorithm based on leaf photosynthate transport. Simulation 2018; 94(7): 593–607.
3. Asakura Y and Kashiwadani M. Road network reliability caused by daily fluctuation of traffic flow. In: Proceedings of the 19th PTRC summer annual meeting, University of Sussex, 1991, https://trid.trb.org/view/1173049
4. Bell MGH, Cassir C, Iida Y, et al. A sensitivity based approach to network reliability assessment. In: Proceedings of the 14th international symposium on transportation and traffic theory, Jerusalem, Israel, 20–23 July 1999.
5. Yao BZ, Chen C, Cao Q, et al. Short-term traffic speed prediction for an urban corridor. Computer-Aided Civ Inf 2017; 32(2): 154–169.
6. Yao BZ, Chen C, Zhang L, et al. Allocation method for transit lines considering the user equilibrium for operators. Transport Res C-Emer 2019; 105: 666–682.
7. Yao BZ, Yu B, Hu P, et al. An improved particle swarm optimization for carton heterogeneous vehicle routing problem with a collection depot. Ann Oper Res 2016; 242(2): 303–320.
8. Yu B, Kong L, Sun Y, et al. A bi-level programming for bus lane network design. Transport Res C-Emer 2015; 55: 310–327.
9. Yu B, Song X, Guan F, et al. k-nearest neighbor model for multiple-time-step prediction of short-term traffic condition. J Transp Eng-ASCE 2016; 142(6): 04016018.
10. Yu B, Wang H, Shan W, et al. Prediction of bus travel time using random forests based on near neighbors. Computer-Aided Civ Inf 2018; 33(4): 333–350.
11. Sun D, Ni X and Zhang L. A discriminated release strategy for parking variable message sign display problem using agent-based simulation. IEEE T Intell Transp 2016; 17(1): 38–47.
12. Chen A, Yang H, Lo HK, et al. A capacity related reliability for transportation networks. J Adv Transport 1999; 33(2): 183–200.
13. Lo H and Tung YK. A chance constrained network capacity model. In: Bell M and Cassir C (eds) Reliability of transport networks. Baldock: Research Studies Press, 2000, pp. 159–172.
14. Bowman LA and Turnquist MA. Service frequency, schedule reliability and passenger wait times at transit stops. Transport Res A 1981; 15(6): 465–471.
15. Peng Z, Shan W, Jia P, et al. Stable ride-sharing matching for the commuters with payment design. Transportation. Epub ahead of print 3 December 2018. DOI: 10.1007/s11116-018-9960-x.
16. Yu B, Guo Z, Peng Z, et al. Agent-based simulation optimization model for road surface maintenance scheme. J Transp Eng B 2019; 145(1): 04018065.
17. Abkowitz MD. Understanding the effect of transit service reliability on work-travel behavior. Transp Res Record 1981; 794: 33–41.
18. Strathman J, Duerer K, Kimpel T, et al. Automated bus dispatching, operations control, and service reliability: baseline analysis. Transp Res Record 1999; 1666: 28–36.
19. Bruinsma FR, Rietveld P and Van Vuuren DJ. Unreliability in public transport chains, 1999, https://research.sussex.ac.uk/en/publications/coping-with-unreliability-in-public-transport-chains
20. Yin Y, Lam WH and Miller MA. A simulation-based reliability assessment approach for congested transit network. J Adv Transport 2004; 38(1): 27–44.
21. Liang Y, Chen Y and Ren F. Reliability analysis of urban road network mobility. J High Transport Res Develop 2005; 22(12): 105–108.
22. Xuan Y, Argote J and Daganzo CF. Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis. Transport Res B 2011; 45(10): 1831–1845.
23. Kim C and Hobeika AG. A short-term demand forecasting model from real-time traffic data. In: Proceedings of the infrastructure: planning and management, 1993, pp. 540–550. ASCE, https://cedb.asce.org/CEDBSearch/record.jsp?dockey=0082432
24. Chen F, Jia Y, An W, et al. Research of short-term traffic flow forecast method based on the kalman filter. In: Proceedings of the 11th international conference of Chinese transportation professionals (ICCTP), 2011, https://ascelibrary.org/doi/10.1061/41186%28421%29995
25. Chien SIJ and Kuchipudi CM. Dynamic travel time prediction with real-time and historic data. J Transp Eng 2003; 129(6): 608–616.
26. Chien SIJ, Ding Y and Wei C. Dynamic bus arrival time prediction with artificial neural networks. J Transp Eng 2002; 128(5): 429–438.
27. Lin Y, Yang X, Zou N, et al. Real-time bus arrival time prediction: case study for Jinan, China. J Transp Eng 2013; 139(11): 1133–1140.
28. Yu B, Lam WH and Tam ML. Bus arrival time prediction at bus stop with multiple routes. *Transport Res C-Emer* 2011; 19(6): 1157–1170.

29. Ahmadi MA, Galedarzadeh M and Shadizadeh SR. Low parameter model to monitor bottom hole pressure in vertical multiphase flow in oil production wells. *Petroleum* 2015; 2(3): 258–266.

30. Fazeli H, Soleimani R, Ahmadi MA, et al. Experimental study and modeling of ultrafiltration of refinery effluents using a hybrid intelligent approach. *Energy Fuels* 2013; 27(6): 3523–3537.

31. Gan XS, Duanmu J and Cong W. Fatalness assessment using a hybrid intelligent approach. *J Saf Sci Technol* 2015; 2(3): 258–266.

32. Yu B, Lam WH and Tam ML. Bus arrival time prediction. *J Saf Sci Technol* 2010; 6(3): 206–210.

33. Guan Z, Song T and Qi L. An application of support vector machine in foundation settlement prediction. *Trans Shenyang Ligong Univ* 2008; 2: 024.

34. Chang CC and Lin CB. Training v-support vector classifiers: theory and algorithms. *Neural Comput* 2001; 13(9): 2119–2147.

35. Frie TT, Cristianini N and Campbell C. The Kernel-Adatron algorithm: a fast and simple learning procedure for support vector machines. In: *Proceedings of the 15th international conference on machine learning (ICML '98)*, San Francisco, CA, 24–27 July 1998, pp. 188–196. New York: ACM.

36. Mangasarian OL and Musicant DR. Successive relaxation for support vector machines. *IEEE T Neural Networ* 1999; 10(5): 1032–1037.

37. Schölkopf B, Platt JC, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution. *Neural Comput* 2001; 13(7): 1443–1471.

38. Schölkopf B, Platt JC, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution. *Neural Comput* 2001; 13(7): 1443–1471.

39. Scaglia F, Manganelli G and Musicant DR. Successive relaxation for support vector machines. *Neural Comput* 2000; 12(5): 1207–1245.

40. Chang CC and Lin CB. Training v-support vector classifiers: theory and algorithms. *Neural Comput* 2001; 13(9): 2119–2147.

41. Frie TT, Cristianini N and Campbell C. The Kernel-Adatron algorithm: a fast and simple learning procedure for support vector machines. In: *Proceedings of the 15th international conference on machine learning (ICML '98)*, San Francisco, CA, 24–27 July 1998, pp. 188–196. New York: ACM.

42. Yu B, Lam WH and Tam ML. Bus arrival time prediction. *J Saf Sci Technol* 2010; 6(3): 206–210.

43. Guan Z, Song T and Qi L. An application of support vector machine in foundation settlement prediction. *Trans Shenyang Ligong Univ* 2008; 2: 024.

44. Schölkopf B, Smola AJ, Williamson RC, et al. New support vector algorithms. *Neural Comput* 2000; 12(5): 1207–1245.

45. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning. In: *Proceedings of the 14th European conference on machine learning (ECML '03)*, Cavtat-Dubrovnik, Croatia, 22–26 September 2003.

46. Schölkopf B, Platt JC, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution. *Neural Comput* 2001; 13(7): 1443–1471.

47. Vapnik V. *The nature of statistical learning theory*. Berlin: Springer, 1995.

48. Iplikci S. Online trained support vector machines-based generalized predictive control of non-linear systems. *Int J Adapt Control Signal Process* 2006; 20(10): 599–621.

49. Smola AJ and Schölkopf B. A tutorial on support vector regression. *Stat Comput* 2004; 14(3): 199–222.

50. Smola AJ and Schölkopf B. *A tutorial on support vector regression*. NeuroCOLT technical report no. TR-98-030, 30 September 1998. London: Royal Holloway College, University of London.

51. Ucak K and Günel GÖ. An adaptive support vector regressor controller for nonlinear systems. *Soft Comput* 2015; 20(7): 2531–2556.

52. Wang X, Du Z, Chen J, et al. Dynamic modeling of bio-technical process based on online support vector machine. *J Comput* 2009; 4(3): 251–258.

53. Cauwenberghs G and Poggio T. Incremental and decremental support vector machine learning. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

54. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

55. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

56. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

57. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

58. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

59. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

60. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

61. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.

62. Ralaivola L and d’Alché-Buc F. Incremental support vector machine learning: a local approach. In: *Proceedings of the international conference on artificial neural networks (ICANN)*, Vienna, 21–25 August 2001, pp. 322–330. Berlin: Springer.