Research on Time-of-station Prediction of Tram Based on Support Vector Machine

Kang Yang, Cui Zhao
Wuhan Business School, School of Information Engineering, Wuhan 430056, China

Abstract. At present, the traffic of the T1 line of the trams in Hanyang District of Wuhan City is based on the GPS positioning. The arrival time prediction based on GPS positioning will be biased due to the sudden situation. Based on the GPS data of the tram, this paper proposes a tram-to-station time prediction model based on SVM (Support Vector Machine). The SVM support vector machine model after a large amount of historical data training is used as the time reference and the arrival time. Make dynamic adjustments. Taking the T1 line of the vehicle-based tram in Wuhan Hanyang District as an example, the prediction result of the arrival time prediction model is compared with the prediction result obtained by single GPS positioning. The results show that the model is applied to the prediction of the arrival time of the tram. Good applicability and higher prediction accuracy.

1. Introduction
Tram vehicles have special driving tracks, large capacity, convenient and fast, and their running time is less affected by social vehicles. Because of its special operation mode, the tram has the characteristics of high stability and good data availability. At the same time, the tram is equipped with a car GPS system, which provides conditions and data basis for predicting the arrival time of the tram. Accurately predicting the running time of each section of trams is not only an important part of building an efficient dynamic dispatching system for rail transit operations, but also a key technology for realizing intelligent transportation modern information services, and improving the overall operational efficiency of intelligent transportation systems. Service levels are important. Support Vector Machine (SVM) is a new tool to solve machine learning problems using optimization methods. Has a complete theoretical foundation of statistical learning and good learning ability. Even if the data sample size is small and the dimension is high, the prediction method can be applied well in this case.

Vanajakshi [1] compared the prediction accuracy of the four models of historical average model, time series model, ANN and SVM. The experimental comparison shows that the prediction accuracy of SVM and ANN methods is higher. Yu [2] proposed a SVM-based bus travel time prediction model, and conducted experimental tests based on the operational data of Dalian 23 bus. The experiment shows that the SVM model is more predictive than the historical average model, ARMA model, and ANN model. The data is more accurate. In addition, Yu[3] is compared with SVM, ANN, nearest neighbor algorithm and regression model. The results show that SVM model is the most accurate model among the four models. Thissen [4] and Wu [5] also proposed the use of SVM model to analyze the bus travel time prediction results. The SVM model generally has better prediction accuracy than the ANN model in the bus travel time prediction field. Generally speaking, the SVM model is also more accurate than other single bus travel time prediction models [3].
2. Overall design of arrival time prediction model

The prediction model is composed of a road condition information acquisition system, an in-vehicle terminal system, a network transmission system, a terminal display system, a database system, and a prediction algorithm system. [7] The road condition information collection system is composed of various information collection tools set in the tram route. For example, a radio frequency identification information exchange module set by a station, intersection monitoring, and the like.

The key to the arrival prediction model is the in-vehicle terminal system. The on-vehicle terminal system of the tram provides GPS real-time information based on the GPS, and realizes even communication between the tram and the dispatch center. On the basis of obtaining the running position and speed data of the tram, it provides reliable operation information for the tram-to-station prediction model.

Through the network transmission system, the train, the stop platform and the dispatch center exchange information. The terminal display system displays the arrival time prediction result through the electronic display screen of each station platform.

The database system provides the basic data needed to predict the model. There are two parts of information in the database, one is real-time information, and the other is historical information. The real-time information is pre-processed and provided to the model, and the historical information is collected through long-term collection. These two kinds of data information are this model. The basis and premise of the realization.

The prediction algorithm system is the most critical module of the prediction model. It determines the accuracy and accuracy of the prediction time. The prediction algorithm system generates the prediction result by calculating and analyzing the information in the database system, and then transmits it to the terminal display system to display the train arrival time. Figure 1 shows the structure of the arrival time prediction model.

![Figure 1. Arrival time prediction model structure](image)

3. Static SVM Support Vector Machine Predictive Model Establishment

Support vector machines have been successfully used in classification, regression and other fields. SVM has been applied to the prediction of bus travel time. In the bus forecasting model, it is more complicated than the application to the tram arrival time prediction because it takes into account the important factors of traffic congestion caused by bus operation. Some factors in the public transport forecasting model can be ignored, not included in the input variables, only need to consider the number of traffic lights in the road section, weather and other important factors affecting train travel time. Therefore, the static SVM support vector machine model suitable for the tram is adjusted and optimized.
3.1. Static SVM support vector machine model

The main idea of the support vector machine is to establish a classification hyperplane as the decision surface, so that the isolation edge between the positive and negative examples is maximized, which is an approximate implementation of structural risk minimization [8].

Define a set of data \((x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_l, y_l)\) \((x_i \in X \in R^n, y_i \in Y \in R)\), where \(X\) is the input vector space, \(Y\) is the output variable space, and \(l\) is the spatial size of the data training set. The expression of its function is:

\[
f(x) = \omega \emptyset(x) + b
\]  

In the formula (1), \(\omega \in R^n, b \in R\). Under the constraint condition, the minimum regularized risk index function \(Q\) is defined, and the optimal regression function of the kernel function is as follows:

\[
Q = \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^{l} L_e(y_i, f(x_i))
\]  

Where \(\omega\) is the first standard vector, (1) is the minimum cost function, and the function can improve the generalization ability; (2) is the empirical risk generalization function, where \(C\) is a constant and greater than zero. \(L_e(y_i, f(x_i))\) is an insensitive loss function, and its specific form is:

\[
L_e(y_i, f(x_i)) = \max(|y_i - f(x_i)| - \epsilon, 0)
\]  

At this point, the function (2) is transformed into a convex quadratic programming to find the optimal solution problem. The Lagrange multipliers \(a_i\) and \(a_i^*\) are introduced, and the vector \(\omega\) is expressed as follows:

\[
\omega = \sum_{i=1}^{l} (a_i - a_i^*) x_i
\]  

Substituting this formula into (2) gives:

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

The kernel function is \(K(x_i, y_j)\), which has the advantage that the operation in the low-dimensional space can be realized without the need to find a specific transformation form of the function, and the step of mapping to the feature space of the high dimension is eliminated. Substituting the kernel function into equation (5) gives:

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

Where \(K(x_i, y_j)\) is the inner product of the vectors \(x_i\) and \(x_j\) in the feature spaces \(\emptyset(x_i)\) and \(\emptyset(x_j)\). In this model, the kernel function with smaller parameters is selected to reduce the difficulty of selecting the optimal parameters. Therefore, the radial basis function \(K\) is selected \(K(x_i, y_j) = \exp(-\gamma ||x_i - y_j||^2)\), using the grid search for the radial basis function to find the optimal parameters \(C, \epsilon, \gamma\), the experimental optimization analysis, the optimization parameters of this model are \(\gamma=8, C=256, \epsilon=1\).

The parameter determination method often used here is the Cross Validation method. The specific steps are as follows: The GPS measurement data of the tram is converted into a matrix, provided that the format required by the support vector machine is satisfied. The data includes travel times and parking times for different trains. All matrix data is divided into \(n\) equal subsets of data and then input as a training set of support vector machines. Finally, traverse and validate each subset and complete all training sets. In this case, the correct percentage of the verification subset as a percentage of the total data is the accuracy of the cross-validation. After the cross-validation is completed, the highest precision corresponding values of \(\epsilon, C, \gamma\) are the optimal values of the model parameters. The input variables of the model include: 1. the station number of the train start station; 2. the length of the road section; 3. the number of traffic lights on the road section; 4. whether the weather is rain or snow, if it is 1, then 0; otherwise 0; 5. Visibility; 6. Average temperature; 7. Monday to Sunday; 8 departure time.
3.2. Specific steps for SVM-based arrival time prediction

Combined with the support vector regression theory, the arrival time prediction model of the tram is proposed. The structural diagram of the prediction model is shown in Figure 2.

Figure 2. Forecast model structure

The specific steps using this prediction model are as follows [9]:

1. According to the principle of the support vector machine and the possible conditions affecting the train travel time, each station along the tram is divided into n sub-sections, and the section between the two adjacent stations is a sub-section.

2. Select the type, kernel function and loss function of the support vector machine. The prediction model uses the support vector machine as the basic algorithm, and the radial basis function \( K(x_i, y_j) = \exp \left( -\gamma ||x_i - y_j||^2 \right) \) as the kernel function, with \( L_{\epsilon}(y_i, f(x_i)) \) as an insensitive loss function as a loss function.

3. Divide the data into two matrices. One is to use the training matrix to train the support vector regression machine in the prediction model, and the other is to use the prediction matrix to predict and test the results.

4. After calculating the optimal parameters of the support vector regression machine, the training matrix is used to train the support vector regression machine according to the optimal parameters, and the trained support vector regression machine is used to calculate the running time of the predicted train in each section.

4. Experimental test and results analysis

The GPS data collected by the trolley car GPS equipment is used as the data base of different road segments. The prediction results are compared with the current arrival time algorithm based on single GPS positioning. The prediction model based on support vector machine theory proposed in this paper can be obtained. The conclusion of the arrival time of the car is predicted more accurately.

4.1. Experimental evaluation indicators

In order to compare with the current single GPS positioning time prediction results, the prediction results are evaluated by three evaluation index factors: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). [9] Its calculation formula is as follows:
MAE = \frac{\Sigma|P-O|}{N} \quad (7)

MAPE = \frac{1}{N} \Sigma \frac{|O-P|}{O} \quad (8)

RMSE = \sqrt{\frac{\Sigma(O-P)^2}{N}} \quad (9)

Where O is the observed value; P is the predicted value; N in the formula is the logarithm of the predicted value and the observed value. In the selection of experimental verification, the data analysis of the first three sections of Wuhan Business School towards the wheel square was selected in the example route.

4.2. Introduction to the example road section
Taking the T1 line of the car-duty tram in Hanyang District of Wuhan as an example, the GPS data was collected. The data covers data from December 28, 2018 to January 19, 2019, for three consecutive weeks. Finally, the driving data of the processed GPS data matrix is extracted as a prediction matrix, and the remaining sample matrices are used as training matrix training support vector machines. A total of 22 stations along the tram line, about 16.8 kilometers, the experiment here only used the nine segments from the Wuhan Business School Station to the terminal wheel square station as the experimental data basis, define the nth station and n+1 station The sub-section number is n, and the T1 line of the tram in Hanyang District is shown in Figure 3.

4.3. Experimental results and analysis
Taking the Wuhan Commercial College Station to the terminal wheel square station as an experimental case, the experimental results are compared with the actual arrival station mean value, and the SVM support vector machine prediction time is compared with the single GPS prediction time. Figure 4 is a comparison chart.
Figure 4. Comparison of travel time comparison chart

Through the preliminary comparative analysis of the above figure, the current time prediction based on single GPS data is larger than the time travel time, that is, the phenomenon that the train is often late. The prediction model of SVM support vector machine theory used in this experiment is very close to the actual travel time, which is more accurate than the GPS prediction travel time method, which verifies the applicability of the prediction model. In order to make a further comparative analysis of the travel time prediction results, the arrival time of each road segment is recorded in the relative error record table, and the table shows that the GPS and SVM are compared with the actual time to obtain a relative error. The relative error of the predicted travel time is shown in Figure 5.

Figure 5. Relative error of the predicted travel time

It can be seen from Fig. 5 that the GPS prediction results are generally fluctuating and the errors are large. The SVM prediction method model has good stability in predicting the arrival time of the tram, and the fluctuations are small and the error is obviously smaller than GPS. The predicted result of the positioning. The GPS error fluctuation is large, and the volatility in the road segment 2 is the largest, while the SVM has an error in the road segment 2, but it is significantly smaller than the GPS prediction result. Overall, the SVM prediction effect is significantly better than the GPS prediction result.

Using the above evaluation indicators to compare the performance of the first three sections, the results of the evaluation indicators are shown in Figure 6.
As shown in Fig. 6, although the MAE, MAPE and RMSE values of the three sections differ greatly, there is one thing in common: the prediction error of the SVM prediction model proposed in this paper is much smaller than the prediction error based on GPS data, and the SVM prediction model accuracy is higher than the GPS prediction model.

5. Summary

At present, the use of GPS data to predict the train arrival time model in the tram has a large error, which will give the planning problem for the passengers in the line. In order to improve the intelligent level of rail transit and provide travellers with timely and reliable forecast information of arrival time of trams, this paper proposes a SVM-based predictive model of tram arrival time, which is obtained in SVM model. Based on a large amount of historical data, the model is established, and the T1 line of the tramway in Hanyang District of Wuhan City is selected as the case study. The results show that the proposed arrival time prediction model based on support vector machine can effectively solve the tram. The arrival time is predicted, and the prediction time is more accurate than the single GPS model. In the latter study, organic algorithms can be combined with other algorithms to make the predicted arrival time more accurate.

Acknowledgments

Supported by the project of Research on Energy-saving Coverage of Wireless Sensor Networks Based on Particle Swarm Optimization, the NO. of project is 2018KY030.

References

[1] Vanajakshi L,Rilett L. Support vector machine technique for the short term prediction of travel time[C]/Intelligent Vehicles Symposium,2007: 600-605.
[2] Yu B,Yang Z Z,Wang J. Bus travel-time prediction based on bus speed[J].Proceedings of the ICE-Transport,2010, 163(1): 3-7.
[3] Yu B, Lam W H K,Tam M L. Bus arrival time prediction at bus stop with multiple routes[J].Transportation Research Part C: Emerging Technologies, 2011, 19(6): 1157-1170.
[4] Thissen U,van Brakel R,de Weijer A P,et al. Using support vector machines for time series prediction[J]. Chemometrics and Intelligent Laboratory Systems,2003, 69(1): 35-49.
[5] Wu C H, Ho J M, Lee D T. Travel-time prediction with support vector regression[J]. IEEE Transactions on Intelligent Transportation Systems,2004, 5 (4): 276-281.
[6] Bai Cong, Peng Zhongren. Prediction of bus travel time based on dynamic model[J]. Computer Engineering and Applications, 2016, 52(3): 103-107
[7] Yan Junjia. Research on GPS-based vehicle arrival time prediction model [D]. Nanchang: Nanchang University, 2015.
[8] Fan Sizhen, Zhou Qicai, Xiong Xiaolei, et al. Tunnel deformation prediction model based on particle swarm optimization and support vector machine[J]. Computer Engineering and Applications, 2014, 50(5): 67-71.
[9] Chen Xumei, Gong Huibo, Wang Jingnan. Study on BRT Travel Time Prediction Model Based on SVM and Kalman Filter [J]. Transportation Systems Engineering and Information, 2012, 12(4):29-34.