Comparison and Development Trend of Bus Arrival Time Prediction Model based on Intelligent Algorithm

Zeyuan (Uriah) Wu

YK Pao School, China

Keywords: Bus arrival time prediction, historic average method, ANN model, Kalman filter (KF) model, Support Vector Machine (SVM) model, Prediction Model Based on Probability.

Abstract: The bus transportation, as one of the major forms of the public transportation system, has played important role in the modern society. As the bus transportation system, and the intelligent transport system as a whole, develops, the precise bus arrival time prediction has become in great need, for which benefits the endusers by provide better prediction service thus saving their waiting time, rendering increased satisfaction rate. To accurately predict the arrival time, an assortment of measures was taken by different entities. This essay would briefly introduce some of them and illustrate how they functions. And then the different approaches are compared.

1. Introduction

As in the modern society the sustainable development has gradually became the main trend of the 21-century urban development, the public transportation was valued the vital composition of the accomplishment of the sustainable development goals. Take China for instance. As the progress of urbanization went further, the economic development of the megacities got faster, and the population boosts rapidly, traffic congestion and environmental degradation have become more catastrophic. Admittedly, the bus transport has advantages in carrying capacity, transport efficiency, and the energy consumption. Hence, the development of the bus transport system, with the Advanced Public Transportation System (APTS) as a whole, became the best solution to the problem.

As the core concept of the ground bus transport system, the prediction of the bus arrival time attributes to the issue of the bus transport information, the implement of the route guidance, and the dynamic dispatching of the bus traffic control. As the intended customers of the bus transport system are majorly students and workers, the delay rendered by the inaccurate arrival time prediction could prompt dissatisfaction and apprehension. Additionally, the crowds could thus convert to the more reliable approaches of transportation and thus put more pressure on them.

2. Research Status

The study of the bus arrival time prediction started in 1970s, when the research environment is extremely limited. Only static prediction model based on routes and departure times provided by the dispatching system are formed then and thus being lack of the capacity to track the dynamic operating state of the buses. Therefore, the latency system could not predict the congestions and accidents and thus could not fulfill the needs of the consumers. As the computer technology became more sophisticated, the accuracy of the prediction raised rapidly, and thus various kinds of modeling measures could perform the prediction of bud arrival time. Among all of them they could be divided into three types: the statistical-based models, which means the models consider the correlation between the historical arrival time and current arrival time to make the prediction; the analysis-based models, which refers to those models that analyze the internal contributory factors of the arrival time and build causal relationships; and the machine learning models.

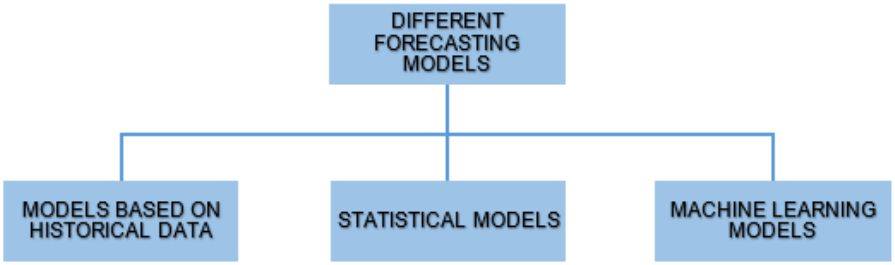

Fig 1. Different forecasting models

As in the current stage of study the way to obtain data and the prediction measures are still limited, the accuracy and reliability of the forecast are still not satiating. The major problems are:

Firstly, the current prediction models are based on certain factors, which don't incorporate the integration research of the uncertain factors, the randomness, and the dynamics.

Secondly, the current result only takes consider to the normal conditions, ignoring the extreme conditions like the conditions in peak-hours, the situation during road construction, and the situations with accidental events, which, in most cases, are ironically the first concern of the customers.

3. Main Methods of Bus Arrival Prediction and Their Comparison

3.1 Main Methods

Several models have been issued and implemented to more precisely predict the bus arrival time. The majority of the measures and the intercorrelation between them could be briefly shown in the following graph:

And here some of the main-stream prediction model would be introduced with respect to the usage of various models, including historic average method, ANN model, Kalman filter (KF) model, Support Vector Machine (SVM) model, Prediction Model Based on Probability.

3.1.1 Historic Average Method

The operation of the bus transport system has very valid periodicity, moving according to a certain time scheme and along fixed sets of stations. Hence, the historical data derived from the preceding bus could serve as valid evidence when predicting those of the following buses. The historic average method is using the average time to deduce the arrival time of the current bus. The method could be briefly classified as three major types: unweighted average method, moving average method, and index average method. Lin had applied the unweighted average method to predict the average latency of the bus arrivals, which equals to the expected value minus the actual arrival time, and then the arrival times predicted based on the dispatching time, GPS location, and average arrival latencies. Hence, the arrival time could be predicted with the average latency, the GPS location, and the dispatching time. Sun had put forward a method combining weighted historic data method and GPS location data predict the arrival times, in which the weighted average of the actual velocity and output of the GPS as the computation speed. Hence, the arrival time reached with the speed and distance. Within the method, weights are dependent on the length of the pathway of the buses: the longer the path, the higher the weight of the historical data.

3.1.2 Artificial Neural Network (ANN) Model

Since the bus transport system is a complicated system without valid linear correlation, the artificial neural network could be applied with its capacity of surmounting complex non-linear system. A typical neural network model consists of three parts: the input layer, implicit strata, and the output layer. In the input layer great amount of non-linear information could be input; the number of neurons could be determined by the dimension of the sample data. The output layer output the data that is calculated and analyzed by the neurons. The implicit strata are the layer lays between the input and output layer, which could be classified as AML, BP, and RBF according to different types of neural network. The setting of the model is on top of the conclusion that the bus arrival times are

correlated to the arrival time of the proceeding bus, and influenced by multiple factors like time period and traffic condition. Chien proposed the self-adaptation back propaganda neural network, predicting with the operating time based on the path and bus stops, which could adjust the parameters of the model according to the error of real-time prediction. Lin proposed a two-layer neural network based on the consideration of different operating features of different operating period and the influence of the traffic lights. Nonetheless, the model demands great amount of training data and thus being less time-efficient.

3.1.3 Kalman Filter (KF) Model

The buses, during operation time periods, could be influenced by other vehicles, traffic light signals, and pedestrians, results rendered not that accurate. The elimination of the noice is crucial in order to enhance the accuracy of the prediction. Kalman filter model constructs the state space of signal and noice, applies optimal estimation recursion theorem, utilize the discrepancy between the expected value and the actual value to construct the optimal estimation recursion modeling index, and thus entitled the capability to cope with multi-dimensional and non-stationary random processes. Shalaby divides the process of the bus system into traveling on the roads and spotting on the stations, with the data from the AVL and APC system, presuming the function of the bus transport system is circulating on a daily basis. Vanajakshi proposed KF model of the mixed traffic situations, dividing the prediction object into several subsidiary road units, constructing the relationship between the intended units and units from the upper course, and adjusting the parameters accordingly. Since the model doesn't require large amount of historical data and could carry out real-time prediction, it could suit the prediction in mixed traffic conditions.

Kalman filter model was first issued by Kalman in the year of 1960, adapting the state space model of the linear stochastic system constructed with the combination of state equation and observation equation.

3.1.4 Support Vector Machine (SVM) Model

The SVM don't need to form any functions, a well-trained SVM could capture the complicated relationship between the input variable and output variable in a nonlinear or time varying system. Hence, the SVM system is especially useful when coping with some intricate system without valid correlations that could be easily described by mathematical formulas. Additionally, one of its major advantages of the system is that the system could get the maximum or minimum value instead of the extremes values of certain domain.

There could be various kinds of factors that could potentially influence the arrival time of the buses, which could thus be incorporated to the system as input variables, such as the weather, time period in a day, or the length of the roads. Before the forecasting, the type of the kernel function and its parameters need to be set in order to ensure the system works. Then, the SVM system need to be trained with database; as the database changed, the system should be trained again accordingly. Then, the system could be implemented to forecast the arrival time of the buses.

Nevertheless, the SVM system based all its prediction on the historical data. Therefore, the trained system might not be able to deal with some accidental occasions and thus needs to collaborate with other models to solve the actual situations.

Yu had proposed prediction model based on the SVM system considering the five factors, including time period, weather, road, the travel time of current section, and the travel time of downstream section. Within the factors, the travel time in the downstream sections are estimated with the time spent by the previous car. The prediction results manifested to be better than those derived with historical average method and ANN model.

3.1.5 Prediction Model based on Probability

As the arrival time could still be significantly influenced by the random factors, the Prediction Model Based on Probability is issued. There are majorly two approaches to reach the model: with the modeling measure and with the analysis of historical statistics. Meng reckons that the prediction of the bus arrival made through traditional approaches not accurate, and he thus proposed a

regeneration probability method. His focus is majorly the interrelationship of the passengers, buses, and other vehicles that could influence the system and the calculation of the range of the arrival times with the influences carried. The model could, to some extent, reflect the uncertainty of the bus arrival times.

3.2 Comparison of Various Forecasting Methods

Based on the analysis of the forecasting models mentioned above, the advantages and disadvantages of bus arrival time forecasting models are compared in terms of forecasting accuracy, reliability, complexity and real-time, etc. See Table 1.

Table 1. Comparison among different prediction models

Model Name	Advantages	Disadvantages	Applicability
Historical Data Method	The model is simple and easy to understand. Fast calculation speed.	Poor real-time performance. Low prediction accuracy of complex traffic state.	Suitable for non-peak forecast
Neural network	Complex Nonlinear Functions can be worked out. Discovery of dependencies between different inputs through training.	Difficult to determine the network structure. Easy to produce problems of past learning and under-learning.	Prediction suitable for peak Period and hump period
Kalman filtering	High-dimensional problems and non-stationary processes. It has good real-time performance due to recursive algorithm.	When doing multi-step prediction, the accuracy has declined markedly.	Suitable for multiple line prediction.
Support Vector Machine	It can deal with complex non-linear problems and has strong learning ability. Suitable for large-scale data	It is difficult to determine the kernel function and its corresponding parameters. The calculation process is time-consuming and unsuitable for real-time prediction	Suitable for single line prediction
Prediction Model Based on Probability	Considering the influence of uncertainties. Mean and variance of the predicted object are obtained at the same time. Improve the reliability of prediction.	It needs a lot of data to determine the probability distribution model, and it has poor portability.	Suitable for uncertainty prediction

4. Trends and Prospects

Considering the aforementioned problems, the main trend of the arrival time analysis is mainly manifested in the following aspects:

Improving the prediction accuracy with the multi-source data fusion technology. Nowadays, the technologies like GPS and GIS are greatly boosted since most vehicles implemented the Automatic Vehicle Location (AVL), the Automatic Passenger Counter (APC), and the Automatic Fare Collection (AFC), and most cities have built the information platform of public transportation. As the hardware and software of the bus transport system continues to update, the diversification and complicate of the data structures are inevitable trend. Hence, how to effectively utilize the multi-source data fusion technology became crucial.

Based on the large, complicated public transport network, study the time correlation of multi-car, multi-point system. As the need of the bus transport became more and more urgent, the density of the bus network became excessively large in some megacities. Therefore, building the space-time prediction model is the inexorable trend of future development.

Consider the impact exerted by the random factors, as the urban traffic system is, in its nature, a dynamic random system.

The research focus on the optimization of the bus arrival time prediction model in different operating state, more detailedly: peak-hour/ non-peak-hour, with/ without bus transit lane, unfavorable weather influences, accidental event, and route construction situations.

References

[1] Lin W H, Zeng J. Experimental study on real-time bus arrival time prediction with GPS data[C]// Transportation Research Board 78th annual meeting, Washington, D. C., Transportation Research Record 1999: 101-109.

[2] Dihua Sun, Hong Luo, et.sl, Predicting bus arrival time on the basis of global positioning system data[C]//Transportation Research Board86thannual meeting, Washington, D. C., Transportation Research Record,2007:62-72.

[3] Chien S I J, Ding Y, Wei C. Dynamic bus arrival time prediction with artificial neural networks[J]. Journal of Transportation Engineering,2002,128(5):429-438.

[4] 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.

[5] Shalaby A, Farhan A. Prediction model of bus arrival and departure times using AVL and APC data [J]. Journal of Public Transportation,2004, 7(1)41-61.

[6] Vanajakshi L, Subramanian S C, Sivananda R. Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses[J]. IET Intelligent Transport Systems, 2009, 3 (1):1-9.

[7] Yu B, Yang Z, LinJ. Bus arrival time prediction using support vector machine [J]. Systems Engineering -Theory & Practice,2006,10(4):151-158.

[8] CHEN X, GONG H, Wang J. BRT Vehicle Travel Time Prediction Based on SVM and Kalman Filter [J]. Journal of Transportation Systems Engineering and Information Technology, 2012, 12 (4): 29 -34.

[9] Meng Q, Qu X. Bus dwell time estimation at bus bay: A probabilistic approach [J]. Transportation Research: PartC.2013, (36):61-71.

[10] Guojun C, Xiaoguang Y, Haode L, etc. Bus arrival time prediction model based on dynamic percentile travel time[J]. Highway Traffic Science and Technology, 2009,126(S1):101-106.

[11] Chen Guojun, Yang Xiaoguang, Liu Haode, et. al. Prediction model of bus arrival time based on dynamic prediction link travel time[J], Journal of Highway and Transportation Research and Development, 2009, 126(S1):101-106.

[12] Jianrong L, Wei D, Bing Z. Study on Lost Time of Buses at Bus Stops[J]. Traffic Information and Safety, 2011,29 (4):44-47. Liu Jianrong, Deng Wei, Zhang Bing, Research on bus dwell time at bus stop [J]. Journal of Transport information and safety,2011,29(4):44-47.

[13] Changshu P, Xuemei Z, Daozhi Z, etc. Analysis of Influencing Factors of Bus Service Quality Based on Passenger Perception[J]. Traffic Information and Safety,2013,31(4):40-43.

[14] Peng Changshu, Zhou Xuemei, Zhou Daozhi, et al. Influence factor analysis of bus transit service based on rider's perception[J]. Journal of Transport Information and Safety,2013,31(4):40-43.

[15] Watkins K E, Ferris B, et. Where is my bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders[J]. Transportation Research: Part A.2011(45):839848.

[16] Yetiskul E, Senbil M. Public bus transit travel time variability in Ankara: Turkey[J]. Transport Policy,2012(23):50-59.