GSM Based Bus Classification Using Machine Learning Algorithm

C. Nandhini and S. Sathees babu

Department of Computer Science and Engineering,
PSNA College of Engineering and Technology, Dindigul, India

Abstract: Effective bus arrival time prediction is more important in transit operation system. Accurate and real-time arrival time information of buses can help passengers' to plan accordingly and minimize long waiting time. A dynamic bus arrival time prediction is based on Kalman filtering and machine learning algorithm using support vector machine. In the proposed model, the Kalman filtering can adjust the bus arrival time with the latest bus information and estimate travel time, the well-trained support vector machine model predicts the baseline bus arrival time from the historical data. It also includes the mobile application that provides exact location and arrival time of bus according to the passengers' request. This system classifies the dynamic operational process according to the carter's base station.

Key words: Kalman filter · Location Tracking · Accuracy · Support vector machine

INTRODUCTION

With the development of public transportation system, the adoption of technologies such as bus tracking, location and communication has become more popular. These technologies make it available to predict bus arrival time at stops and help to provide reliable service in public transport. The passengers can enable efficient scheduling of their trips using the available accurate bus travel time prediction to avoid waiting times. The time series mainly rely on the similarity of the historical data and future data. The prediction result will apparently deviate with the variation of the historical data. Generally speaking, the passengers are interested in the predicted travel times of the next buses and the predicted arrival times at the bus stop. Passengers can choose different bus routes to reach their destinations. They would like to know when the next buses will arrive at the bus stop. But in previous studies specifies the bus generate the same road segments and bus stops to predict the bus arrival times. The range is more important because it contains geometric information needed for positioning and timing. A dynamic bus arrival time prediction model on road that are developed using real-world data, which can fill the gap that there is no dynamic model for bus arrival time prediction. Typically, automatic location systems are based on GPS measurements where the data

acquired in each vehicle were uploaded to a main server with a large periodicity (commonly daily), to a synchronous method. It is scheduled that if the predicted arrival times of the next bus could be known by the passengers' it would save passengers' waiting times and decrease anxieties. In this where the buses typically runs, real-time tracking can improve passengers' confidence in the transit service, allowing users to schedules closer connections with less built-in margin of error. The bus travel time of preceding buses of route is considered as one of the input variables in the proposed models.

The performances of dynamic operational models and the traditional models have been assessed and compared for forecasting bus arrival times of bus on road. The goal is to reduce complexity and cost of the services by creating an automatic system for transit tracking, arrival time prediction and mapping. The algorithm that basically combination of Kalman filtering and machine learning using support vector machine for classification of routes and specifies the arrival time of the bus according to the passengers' request and reply. The system provides dynamic operation of Kalman filtering and support vector machine which means Kalman filtering for adjusting the bus arrival time with the latest information and estimate the travel time and support vector machine predicts the baseline bus arrival time from the existing information. The combination of two algorithms leads to reduce cost and

Corresponding Author: C. Nandhini, Department of Computer Science and Engineering,
PSNA College of Engineering and Technology, Dindigul, India.

define accurate arrival of bus for the passengers' to plan for the next bus arrival. The mobile application provides exact location and arrival time of bus according to the passengers' request. Not much research focuses on more than one preventive action. One of the most promising research areas is learning from data bases.

Maintain schedule reliability using metrics such as on-time performance, movement constancy. Schedule coordination at terminals and hubs to facilitate transfer and possible to divide the process into a two-stage are filtering and tracking. According to increase in the number of smart phones users the transit agency can implement a sophisticated bus-tracking and arrival time prediction system simply and downloading the bus-tracking app to each phone. Manually collecting this type of information can be a time consuming and complex task for many transit agencies. Due to its automated nature, our system is able to balance the latest routes and schedules in response to latest road construction or predictable congestion events.

Theory and Backround

Historical Average Models: In the past decades, a variation of patterns and algorithms has been developed to predict the bus arrival times or bus travel times. The models widely used ones can be classified into the following categories: historical average models, regression models, machine learning algorithm models including artificial neural network (ANN) models and support vector machine (SVM) models, Kalman filtering-based models and dynamic models. Historical average models are based on the historical data and able to predicate the bus travel times or bus arrival times through previous bus trips. These models will be reasonable, convenient and reputable when the traffic flow is relatively small and stable. A historical model for estimating the link travel time between two bus stops, which are calculated as the average travel time between two bus stops reduced by the average dwell time at bus stops [1]. Historical average models could be valuable in the occurrence of prediction models but the reliability of the prediction was limited.

Regression Models: Regression models use a multivariate statistical approach for examining the linear correlations between a set of independent variables and a single dependent variable [2]. A set of multiple linear regression models to estimate travel time from current bus stop to the target stop. Distance, bus schedule adherence and arrival time are at one specific bus stop having been chosen as

the independent variables in regression models [3]. Although different independent variables and different consolidation of these independent variables were set in different regression models, the results suggested that the prediction review of the regression models was good. In addition, multiple linear regression models have the ability to announce the degree of importance of each independent variable.

Artificial Neural Network Models: Artificial neural network (ANN) models are very popular in forecasting bus travel times and bus arrival times. ANN models had the ability to solve complex nonlinear relationships and they are very effective in bus travel time prediction. It demonstrates its superior performance as compared with the historical data based and multi linear regression models. The authors indicated that their model was powerful in modeling variations in bus-arrival times along the service route. These models have been proved to be effective for the provision of satisfactory bus arrival time information. Though the learning and testing process is inherently delicate and is slow to coverage to the optimal solution.

Support Vector Machine Models: In recent times, SVM has been proposed as a good technique for bus arrival time prediction and bus travel time prediction. There were many profitable attempts in bus travel time prediction. A number of different forecasting methods for travel time estimation including historic method, time series analysis, ANN and SVM are compared [3]. Comparison showed that the representation of both SVM and ANN models were comparable to each other and these two methods outperformed other methods. The results showed that the SVM model outperformed the historic mean prediction model, the autoregressive integrated moving average and the ANN Model.

SVM model proved to have better prediction performance than that of the ANN model. In general, SVM model outperformed other bus travel time prediction models in terms of prediction accuracy. SVM model can record the training data from the input space into a higher dimensional feature space. SVM model are trained and tested with the data sets which consist of some data instances. Each instance in the training set contains one target values and several attribute values. The goal of SVM is to produce a model which predicts target value of data occurrence in the testing data set which is given only the attributes. Then use the attributes data to train the SVM. The SVM can accurately predict the class label of

target values; consequently the system can recognize each target value correctly. The SVM algorithm is used to predict the arrival time based on the travel time of a current segment and the latest travel time of the next segment.

Algorithm 1

Candidate Support Vector (SV) = {closest pair from opposite classes}
While there are violating points do
Discover a violator
CandidateSV = candidateSV combines violator
If any $\alpha_\square$ <0 due to addition of S to D then
Candidate SV = candidate SV\p
Repeat till all such points are pruned
End if
End while

Based on the nearest neighborhood nodes and distance the bus available will be calculated. It considers arrival time and departure time as well as number of waiting passengers in all bus stops. Hence this research is easy to work with huge dimensional dataset and it aims at partition concept. Based on the location, location information, number of travelers and number of waiting passengers in diversity bus stops this scenario provides response as much as faster.

Kalman Filtering Models: Kalman filtering algorithm was introduced for travel time prediction because of its advantage in continuously updating the state variable as new observations. The adaptive Kalman filter technique based Algorithm was tested in a stretch of freeway. Compared with the probe based method and the double detector based method, the proposed algorithm outperformed under both recurrent and non recurrent traffic conditions. Although various approaches based on Kalman filter were explored to improve the prediction accuracy, this study lacks the comparison of other prediction models. Compared with a prediction method using space discretization, the proposed algorithm had better performance in prediction accuracy. There are many previous studies using Kalman filtering based dynamic algorithm in travel time prediction. All these studies showed that Kalman filtering-based models are feasible and have a strong theoretical foundation in travel time prediction. However, most of these Kalman filtering based models lack the performance comparison with other models and algorithms.

Algorithm 2

Repeat {every ΔT_p}
Read location information from Bdb
Extract all the measurements from Bdb within interval [t_k-ΔT_{db},t_k]
For b=1 to B do {bus index}
If there are HF-badge events for b then
Set estimation position of b to the location of the HF badge reader according to recent detection event
Else
Select measurements related to bus b
Select reference location information according to the selected measurements
If there are measurements of b then
Estimate the mobile position using KF
Else
Position estimation is not available and do not do any estimation
End if
End if
Display the estimated position on the map
Update the estimated position to the Bdb
End for
Pause if ΔT_p is not fully consumed
Until stop

The Kalman filter smoothes out and tracks the estimation errors by adopting linear prediction from the previous estimation data while the bus is dynamically moving. The formulation of scheme, a feasible accuracy can be acquired for location tracking, including position, distance and time. However, the accuracy is significantly affected by the precision of the external location estimator.

Dynamic Models: Different researchers have different opinions on dynamic models and as a result different algorithms are proposed in the dynamic models. A dynamic algorithm [4] integrating the ANN model and Kalman filter-based algorithm, because the history data based models had difficulty in dealing with dynamic traffic conditions. Results showed that this dynamic model was powerful in predicting bus arrival times along the service route. A hybrid model that was based on SVM and Kalman filtering technique to predict bus arrival times, which performed better than the ANN-based methods. The Kalman filter algorithm was applied to adjust the bus travel times predicted by SVM. The prediction results of the proposed model outperformed the Kalman filter model,

but it lacked the results comparison with that of SVM model. By integrating bus travel times of different bus routes on the same road segments, the estimation accuracy of traffic conditions could be improved. The previous researches only used the information of the same bus route to predict the bus travel/arrival times, but the integration of bus information of multiple bus routes was not included in these studies.

Over View of Structre: The dynamic model consists of two particular components: the first component is the support vector machines (SVMs) model estimating the baseline bus travel times on road with bus routes; the second component is the Kalman filtering-based dynamic algorithm, so the prediction results of the first component can be adjusted based on the latest travel time information. Due to the limited capacity of bus stops, the buses might queue up at the bus stops and therefore the bus travel time becomes much longer and unreliable. Thus, to find a feasible and applicable predicting method for bus arrival times on road with bus routes is very meaningful and important.

SVM model can map the training data from the input sphere into a higher dimensional feature sphere. In this higher dimensional feature sphere, a separating wired plane is constructed which can make the maximum margin in the feature sphere. Points on the edge are called the support vectors. Although the performances of SVM or ANN models outperform other models in terms of prediction accuracy, the SVM or ANN models still cannot allocate the prediction results dynamically. The SVM or ANN model is based on the historical data and no importance how they are trained and tested them can only estimate the bus travel times based on historical data but not the real-time information. So the Kalman filtering-based dynamic algorithm is proposed in this dynamic model so as to take full value of the current bus travel time data.

The Architecture diagram consists of two steps, namely, the offline prediction step and the dynamic adjustment step. The primary step is the offline prediction, which uses the historical bus travel time data and the well-trained SVM or ANN models. The output of the primary step is the baseline bus travel time, which serves as the input of the next step. The secondary step is the dynamic adjustment. In the second step, the Kalman filtering-based algorithm can adjust the baseline bus travel time with the latest travel time data. This dynamic model is SVM-Kalman model. The Kalman filter is integrated according to the navigation in maps the navigation solution is computed and combines to define the particular location.
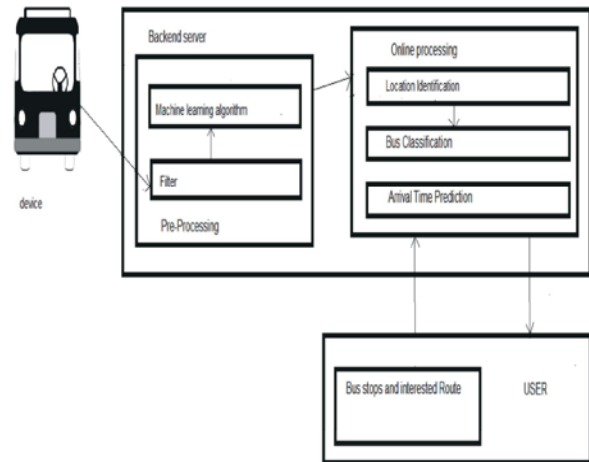

Fig. 1: Architecture diagram

Performance Evaluation: We have presented an algorithm that is efficient, intuitive and fast. Thus the algorithm significantly outperforms the early process. Due to the complex nature of the data, generally a single method or algorithm has not achieved feasible results. However there is increasing trend to utilize hybrid algorithms to improve the prediction accuracy. There is no single method could produce dynamic prediction due to nature of the methods. While kalman filters and machine learning using support vector can separately predict the accuracy of bus. Combination of these two algorithms can produce better prediction result to the passengers in case of limited information of data. In case of insufficient information about the conditions, some positive data could be used most representative support vector data. In such bounded fragments, regression or time serious further application of Kalman filters, would isolate, any other dynamic factors, while improving accuracy. Due to complex nature of the data, generally a single method or algorithm has not achieved any robust and feasible results. However there is an increasing trend to utilize hybrid algorithms to improve the prediction accuracy.

In Fig. 2 comparing the two different algorithms in early result and evaluated the efficiency of two algorithms. The separate features of the kalman filter and support vector can provide the prediction accuracy of the information from the database. The combined process of kalman filtering and support vector can provide the efficient result and reduce the error while the evaluating and transferring of information to the passengers' through mobile application. The analysis of the years are specified and evaluated. Efficiency of the algorithms are defined and compared according to the recent years of information collection.
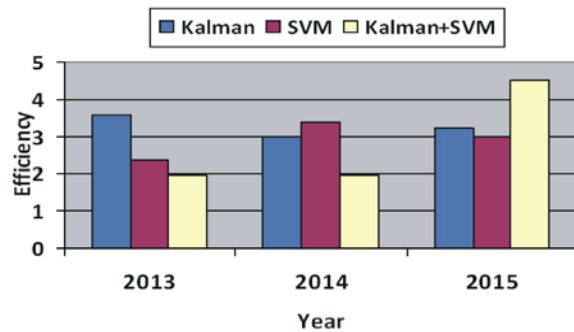
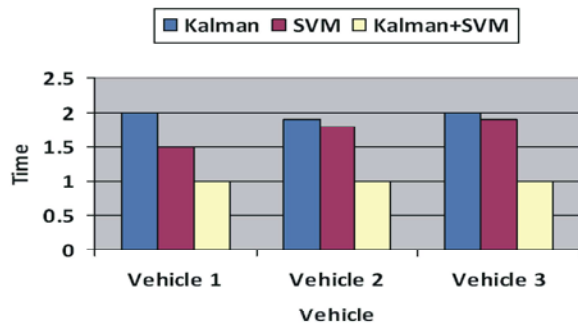Fig. 2: Efficiency of combined algorithm

Fig. 3: Server response time

In Fig. 3 represents the server side and time required for detecting bus. The variations are defined by the Time of seconds for each bus traveling and accuracy. Creation of application are used for easy identification of passengers' were the bus traveling and when it reaches the position of passengers'. The chart represents the vehicles that how many seconds it takes to show the result for Kalman filter, Support vector machine and the combine algorithm. According to the number of bus and arrival time information are defined accuracy and fast detection of information.

CONCLUSION

In this paper, the dynamic travel time prediction models for buses on road with multiple bus routes. The weighted standard bus travel time of preceding buses of any route was introduced as one of the input variables. The detailed approaches of the support vector machine and Kalman filtering-based dynamic algorithm were presented in this paper, together with the formation of the dynamic bus travel time prediction models on road with multiple bus routes. Furthermore, the comparison sequence showed that in general the SVM-Kalman model was the most accurate one among all the models. The SVM Kalman model was a little advance than the ANN-Kalman model in terms of prediction accuracy, but it

outperformed the pure SVM, ANN and Kalman models. This prediction system also helps to attract more passengers and increase the demand of buses. This system is able to handle and track a large number of buses, identifying there service routes and the forecasting model predicts the arrival time with great efficiency.

REFERENCES

1. Jeong, R. and L.R. Rilett, 2004. Bus arrival time prediction using artificial neural network model. in Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, pp: 988-993. IEEE.

2. Marx, M.L. and R.J. Larsen, 2006. Introduction to Mathematical Statistics and Its Applications. Pearson/Prentice Hall.

3. Vanajakshi, L. and L. Rilett, 2007. Support vector machine technique for the short term prediction of travel time. in Proceedings of the IEEE Intelligent Vehicles Symposium, pp: 600-605.

4. Chen, M., X. Liu, J. Xia and S.I. Chien, 2004. A dynamic bus-arrival time prediction model based on APC data. Computer-Aided Civil and Infrastructure Engineering, 19(5): 364-376.