Handling positional uncertainty in a real-time bus tracking system

Bashir Shalaik and Adam Winstanley
Computer Science Department
National University of Ireland
Maynooth, Kildare
{bsalaik, adam.winstanley}@cs.nuim.ie

Abstract— **Automatic Vehicle Location (AVL) systems are increasingly being used by transit agencies for the real time monitoring of their vehicles. AVL systems can be used to improve the service given to passengers by using information on the current position of buses to maintain headways or increase reliability by improved operational control and provide an estimate of as to the arrival time of the next bus at the stop. In real-time bus tracking systems, some positional uncertainty is usually associated with the location of buses in service that are tracked using a locational device such as a Global Positioning System (GPS) receiver. Rather than raw coordinates, the location is usually better understood in terms of the landmarks along the route, particularly the named stop. Three prediction models have been implemented to estimate the location of vehicles on bus routes. Analysis indicates that one based on mining historical data for patterns gives more accurate results than regression and Kalman filter models when travel is disrupted by one-off events.**

Keywords: AVL; GPS; APC; GPRS; Real Time \Bus Tracking

I. INTRODUCTION

A main component of transport information services is predicting the arrival times of the next bus and providing this information to the public via various media. Due to the different operational conditions, the technologies used for arrival time estimation in transit modes such as rail systems can not directly be transferred to fixed route bus services. Rail transit is a well-controlled mode; vehicles have few external factors interfering with their progress and their locations are tracked as a matter of course for safety, signaling and other reasons. Real-time bus locations can not be determined or estimated as accurately without external sensors, so bus location or arrival time information often has a degree of uncertainty. This uncertainty is likely to be higher due to the presence of a number of factors such as traffic congestion, intersection delays, and weather conditions. These can force buses to deviate from their schedules and headways and can result in poor services and increases in the length of passenger waiting-times at stops.

Providing accurate real-time bus arrival information enhances the credibility of public transit systems and makes them more competitive against other transportation modes. GPS devices have become the central technology enabling this function and allowing travel and arrival time data to be comprehensively collected. Additionally, recent advances in cheap mobile communication have made it possible to transmit GPS data via GSM/GPRS technology to a central server inexpensively using public networks. This data can be displayed and visualized through different media such as PDAs, mobile phones or web browsers in a real-time environment. One major use of this data is to provide vehicle arrival time predictions for passengers along the route.

Work sponsored by Science Foundation Ireland, Libyan Ministry of Education and Blackpool Transport Ltd.

II. REAL TIME BUS TRACKING SYSTEMS

Real time bus tracking systems are a combination of the installation of a location-aware transmitter on a bus plus purpose-designed computer software to collect the data and enable its storage, display and analysis. Modern bus tracking systems commonly use GPS technology for locating the bus, but other types of location technology can also be used. Bus real time information can be viewed on electronic maps and diagrams via a web browser or using specialized software. Urban public transit authorities are an increasingly common user of bustracking systems, particularly in large cities. By using real time bus tracking systems, bus services provide on-the-fly information to their users, including the current locations of buses and the predicted arrival times at bus stops. Urban transit agencies use the technology for a number of other purposes, including monitoring schedule adherence of buses in service. In order to improve services, as well as providing real-time information, real time bus tracking systems can build up an archive of data that can be analysed and mined for information to show the behaviour of the transport system over time, indicating common problems such as vehicle bunching and delays due to congestion.

In a joint project between NUI Maynooth and Blackpool Transport (Great Britain), a real-time bus tracking system has been built allowing the operator to assess and improve the quality of their public service. Figure 1 shows the real-time user interface of the bus tracking system.

In our system, an on-board GPS receiver is interfaced with a GSM modem which uses a standard mobile phone network. The GPS receiver records point locations in latitude-longitude pairs, speed, date, time and other vehicle data (battery voltage, engine temperature) at pre-specified intervals (typically 30 or 45 seconds). The data is transmitted via GPRS to a standard web-server that stores it in a database. A web browser can be used to display the data for various purposes (tracking, arrival time prediction, quality-of-service metrics) in various formats (maps, diagrams, text) using server-side processing.

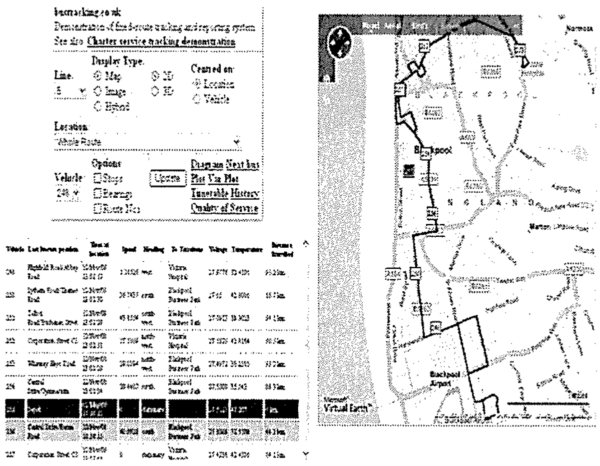

Figure1. User interface of real-time bus tracking system.

III. POSITIONAL DETERMINATION IN REAL-TIME BUS TRACKING SYSTEMS

Knowledge of the vehicle location (and sometimes the azimuth and speed as well) enables fleets of vehicles to be tracked and managed. Various real time bus tracking systems use different techniques to determine a vehicle's location, usually in terms of the distance to a particular bus stop on the route. In (Ramakrishna, Lakshmanan, 2006), the timestamp of GPS data points at or near a bus stop are considered as arrival times at the bus stops. (Kidwell,2001) suggests creating artificial "sign-posts" inside the computer software. The software watches the GPS data and sets a flag when the bus has moved close enough to an identified location (such as a bus stop). In (Predic, Rancic,Stojanovic,Milosavlejevic 2007), the arrival of the vehicle at the bus stop is assumed if a single GPS positional update for that vehicle is detected inside a predefined circular proximity zone and the detected speed is lower than a predefined upper threshold. The reason for this condition is due to the nature of streamed GPS positional data. A more precise condition would be if there are two consecutively detected GPS positional updates within a defined proximity to the bus stop with a speed of 0km/h.

In the system implemented for Blackpool, an efficient method has been implemented for measuring the identity and distance of the nearest bus stop to the moving buses as well as different prediction models to provide real-time information on bus arrival time. A spatial indexing and search strategy efficiently locates the nearest stop within the database of stops for the route. This uses the maximum latitude and longitude interval between any two stops on a route to eliminate most stops before an accurate spatial look-up is performed to determine the closest stop.

A. Bus Arrival Time Prediction Models

Accurate real-time prediction of bus arrival times have been the subject of research for at least two decades using different types of data collected from different sources, such as probe vehicle data, magnetic ticket data, Automatic Passenger Counting (APC), Automatic Vehicle Location (AVL) and other systems. Different methods are used for predicting bus arrival time, some systems using simple statistical/mathematical models, e.g. prediction according to deviation from the schedule. Kalman filtering and more complicated artificial intelligence and machine learning algorithms are also in use. In our system, three bus arrival time models were used, namely a historical-data model, a multiple linear regression model and a one-dimensional Kalman filter model (based on Shalaby, Ontario and Farhan, 2003). Bus arrival times at stops in urban areas are difficult to estimate because travel times on routes (links), dwell time at stops, delays at intersections and traffic congestion are changing spatially and temporally. These changes impact the bus schedule and headway adherence as the vehicle moves farther from the starting terminal.

Developing a Credible travel time prediction model could accurately estimate vehicle arrival times and improve transit vehicle performance. However developing such a model that takes into account the effects of time, space, traffic conditions, dwell-time, rider-ship and weather conditions is a difficult task. Our three models were implemented and tested using data collected from bus line 5 in Blackpool. Different variables act as inputs to each model. The output variable is the travel time from the current bus stop to the target bus stop in minutes.

IV. HISTORICAL DATA AVERGING MODEL

The historical data model predicts travel time using the average travel time for the same journey over similar conditions obtained from the data archive. These models assume that traffic conditions are cyclical and that the usual travel time on a specific route at the same time-of-day and day-of-week is a good predictor of the journey time today. This model requires sufficient observations (samples) to produce meaningful results. The drawback of is that the accuracy of the results is dependent on the similarity of travelling patterns.

An algorithm has been implemented to estimate bus arrival times using our data archive. This algorithm involves using of historical operational data on the route and averaging bus travelling time (between bus stops) for each hour of each day of the week with the underlying assumption that buses will be running in the same operational conditions at the same time each week. This averaged data is used to predict how long it will take an individual bus to get from where it is to a particular stop:

$$\eta = \frac{1}{N}\sum_{i=1}^{N} traveltime_i \qquad (1)$$

where: η = mean travel time.

V. REGRESSION MODELS

Regression models are conventional approaches for predicting travel time. These models predict a dependent variable with a mathematical function formed by a set of independent variables. This requirement for independent variables limits the applicability of the regression model in transportation because variables in transportation systems tend to be highly inter-correlated. However, an arrival time estimation model can be developed as a function of several significant variables (for example average speed, distance, number of intermediate stops and time of the day).

In this work we used multiple linear regression methods to construct a formula that will predict travel time values from two variables (current time at a particular bus stop and number of intermediate stops). Multiple linear regression is a form of regression analysis in which the relationship between one or more independent variables and another variable, called the dependent variable, is modelled by a least squares function, called a linear regression equation (Patnaik, Chien, Bladikas, 2004).

The current time (time stamp) of a data point on the nearest bus-stop (in the travelling direction) and the number of intermediate stops were selected as independent variables. Some other variables such as distance between the current position of the bus and nearest terminus, the bus average speed from a bus trip origin point to the current location were tested but they have a low correlation with the dependent variable. The relationship between these two variables and the dependent variable was evaluated as a linear relationship. The least square method is used to estimate regression parameters and these were applied to calculate a predicted arrival time.

VI. Kalman Filtering Models

The purpose of a Kalman filter is to estimate the state of a system from measurements which contain random errors. The Kalman filter is a multi-dimensional model based on the numbers of state variables to be estimated. In this work we implemented a one-dimensional Kalman filter model to estimate bus arrival time. We used one parameter (last observed bus running time). One of the main characteristics of a Kalman filter is its ability to update itself based on new data that reflects the changing characteristics of the vehicle operating environment. Bus travel time between stops was modelled to predict the arrival time. This model makes use of the historical data (from several previous days) of bus running times for the instant ($k+1$) as well as the bus running time for the previous bus on the current day at the instant *(k)* to predict the bus running time at $k+1$. (Shalaby, Farhan, 2004).

VII. Prediction Models Performance Evaluation

In order to evaluate the performance of the three models, Mean Absolute Percentage Error (MAPE) was used as a measure of closeness between predicted and observed values. (Chung, Shalaby, 2007), (MAPE 2009), (Shalaby, Farhan, 2004).

$$MAPE = \frac{1}{n}\sum_{i}^{n}\left|\frac{y_p - y_{ob}}{y_p}\right| \times 100\% \qquad (2)$$

where:

y_p = predicted travel time.

y_{ob} =observed travel time.

n= number of validation data points.

MAPE represents the average percentage difference between the observed value and the predicted value. Table 1 shows MAPE values of the three prediction models.

TABLE I. MAPE values of bus arrival times prediction models

Model	MAPE
Historical Data Model	13
MLR Model	29
Kalman Filter Model	20

VIII. Conclusion

The preliminary results obtained show that the historical averaging model performs better than the Kalman filter and multiple regression models. This is to be expected due to the cyclic similarity of bus travelling patterns. However, analysis of results shows that the historic model can sometimes perform much worse than the others. This is in congested traffic and extremely slow vehicles conditions caused by unpredictable rare incidents such as road traffic accidents. Therefore a combination of methods may be more optimal. Future work will consider how to combine the models to take this into account.

It is also true that all arrival time prediction models can perform well with more continuous data points from GPS devices. The use of more accurate GPS devices may well improve the prediction performance of these models.

Acknowledgments

Thanks are due to Blackpool Transport Ltd for facilitating this project and particularly to Oliver Howarth for his comments and feedback. This project was partly supported by the StratAG project funded under the SFI Strategic Research Cluster Program (07/SRC/I1168). One author is supported by a PhD studentship from the Libyan Ministry of Education.

References

Lakshmanan V. (2006). Bus travel time prediction using GPS data. *Proceeding of Map India.*

Kidwell, B. (2001) *Predicting transit vehicle arrival times.* Geographic Laboratory, Bridgewater State College, Bridgewater, MA, USA August http://www.glump.net/content/bus_predict/predicting_transit_vehicle_arria.htm

Predic, B. Rancic, D. Stojanovic, D and Molosavijevic, A. (2007). Automatic vehicle location in public bus transportation system. *Proceedings of the 11th WSEAS International Conference on Computers*, Agios Nikolaos , Crete Island, Greece

Patnaik, J. Chien, S. and Bladikas, A. (2004). Estimation of bus arrival times using APC data. *Journal of Public Transportation.* 7(1)

Shalaby, A., Ontario, T., and Farhan, A. (2003). *Bus travel time prediction model for dynamic operations control and passenger information systems.* TRB 82nd Annual Meeting, Washington D.C.

Shalaby, A and Farhan, A. (2004). Prediction model of bus arrival and departure times using AVL and APC data" *Journal of Public Transportation.* 7(1)

Chung, E., Shalaby, A. (2007). Expected time of arrival model for school bus transit using real-time Global Positioning System-based automatic vehicle location data. *Journal of Intelligent Transportation Systems*, 11(4), 157-167.

MAPE (2009): http://en.wikipedia.org/wiki/MAPE