Improved Traffic Prediction Accuracy in Public Transport Using Trusted Information in Social Networks

Ahmad Faisal Abidin, Mario Kolberg and Amir Hussain

Department of Computing Science and Mathematics, University of Stirling, UK
{faa,mko,ahu}@cs.stir.ac.uk

Abstract. Bus arrival time prediction is a key service for improving public transport attractiveness. In this research, a model of bus arrival time prediction, which improves on the accuracy of such prediction approaches, is proposed. The bus arrival time will be predicted using a Kalman Filter (KF) model which is complemented with additional information from social network communication. Using social network communication in this context, allows to include road traffic information into the prediction, based on people's experience near the scene. This paper provides an introduction to Kalman Filters and their use for traffic prediction. The paper presents some initial results using a Kalman Filter model for prediction and details how information from social network communication will be added into the Kalman model. Verifying the trustworthiness of social network information is crucial to the success of the approach and the paper discusses some of the challenges.

Keywords: Kalman Filters, Social Networks, Traffic Arrival Prediction, Trust.

1 Introduction

Traffic flow on major urban roads is affected by several factors. It is influenced by a number of conditions and events, such as traffic lights, road conditions, number of vehicles on the road, time of travel, weather conditions, and driving style of vehicles. The provision of timely and accurate travel time information of public transport vehicles is valuable for both drivers and passengers. Recently, arrival time estimation approaches for public transport have attracted an increased interest by researchers applying various paradigms to tackle the problem.

When real-time travel time measurements are available, a dynamic calibration of travel time models is able to improve prediction performance [1]. Dynamic Kalman Filter (KF) models have been shown to not only be able to estimate traffic states on freeways [2], but also to improve the accuracy of urban travel time prediction due to their key property of updating their state continuously creating new observations[1,3].

There are many algorithms and statistical models that have been proposed for vehicle arrival time prediction. However, there is a gap amongst these algorithms. One particular issue is the question of how the algorithm can receive and incorporate live real-time traffic information. Without receiving such information, algorithms cannot produce an accurate result. If there is an accident 5 miles down the road with a traffic jam which takes 30 minutes to pass, then this information should be included in the arrival time estimation. Very few types of approaches can cope with dynamic information. Kalman Filters is one of them and exhibits the most accurate predictions.

This research proposes an approach which uses information from social networks, especially Twitter, and incorporates this with a Kalman Filter based arrival time prediction algorithm.

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

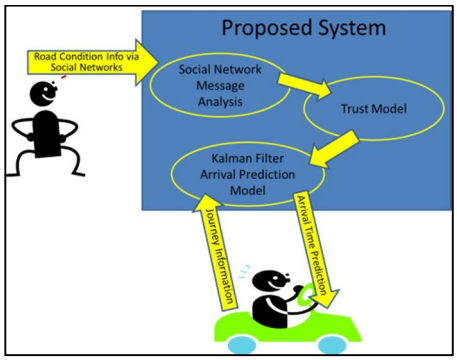

Figure 1: Overview of Arrival Time Prediction System.

Figure 1 depicts the proposed system with its major components. The system takes two types of input: information about the journey (location, speed etc) from the vehicle and road condition updates (accidents, road closures, traffic jams) via social network messages. The system will process the latter to extract the required information from the messages, establish if these messages are trustworthy and if so passes the information together with the journey information into a Kalman Filter model to estimate the arrival time. This is fed back to the vehicle and its passengers and other passenger announcement systems such as bus stop displays.

Using such information with public transport arrival time prediction approaches is novel, and has a major cost advantage over previous approaches employing road sensor data. In addition, this approach allows for the identification of unexpected traffic events, and the subsequent inclusion of this new, real-time information, as part of route calculations and updates. This provides updated information during journeys that may not have been available when travel was initially planned or started. In this situation, social networks can play a pivotal role as an input to the model. Trust is a major concern when using information from social networks. This research proposes to consider trust in the social media information used. We propose to use an approach to trust which takes into account a variety of aspects including location of sender, past behaviour, and reputation in other social networks.

2 Using Kalman Filters for Traffic Arrival Prediction

Recent literature surveys [4][5] have shown that arrival time prediction models are commonly based on historical arrival patterns and/or other explanatory variables correlated with arrival time. The explanatory variables used in these previous studies include historical arrival time (or travel time), schedule adherence, weather conditions, time-of-day, day-of-week, dwell time, number of stops, distance between stops and road–network condition [4][5].

The effect of congestion was treated in various ways in the approaches. For example, some have used traffic properties like volume and speed from simulation results [6], while others have clustered their data into different time periods [7]. Historical data based models were used in geographical areas where traffic congestion is less likely, as the models assumed a cyclical traffic pattern. On the other hand, Kalman Filter techniques and Artificial Neural Network (ANN) approaches were used mostly in urban areas [4]. Importantly, Kalman Filter approaches can be applied with updated input data while the journey is in progress. In fact, besides Kalman Filter approaches, only some statistical approaches [8] take dynamic route information into account.

2.1 Use of Kalman Filters

Kalman Filters are an estimation approach. That is they infer values of parameters from observations which may be noisy, inaccurate, and uncertain. Importantly, unlike many other approaches, Kalman Filters are recursive and hence can take new observations into account as they arrive. With this, Kalman Filters can be executed at *runtime* of the system under observation. The algorithm is referred to as a 'filter' as calculating the estimate from noisy input data essentially 'filters out' the noise.

Kalman filters estimate a process by estimating the process state at a given time and then obtaining feedback in the form of noisy measurements. Generally, the equations for the Kalman filter fall into two groups: time update equations and measurement update equations. Time update equations are responsible for predicting forward (in time), using the current state and error covariance estimates to obtain the a priori estimates for the next time step. The measurement update equations are responsible for the feedback, i.e. for incorporating a new measurement into the a priori estimate to obtain an improved a posteriori estimate. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. Indeed, the final estimation algorithm resembles that of a predictor-corrector algorithm for solving numerical problems.

A Kalman model implies that the state of a system at a time k+1 developed from the previous state at time k. This can be expressed by the following state equation:

$$x_{k+1} = \alpha x_k + \beta u_k + w_k \tag{1}$$

Here x_{k+1} is the state vector containing the terms of interest (e.g. position, velocity) at time $_{k+1}$. u_k is the vector with the control inputs (e.g. steering angle, braking force) at time $_{k.}$ α is the state transition matrix which applies the effect of system state parameters at time $_k$ to the state at time $_{k+1.}$ This basically means that the position and velocity at time $_k$ will affect the position at time $_{k+1}$. β is the control input matrix which applies the effects of the control input parameters (u_k) to the state vector. w_k is a vector containing process noise terms for each value in the state vector.

Measurements of the system are performed according to the formula:

$$y_k = \mu x_k + z_k \tag{2}$$

where y_k is the vector of measurements, μ is the transformation matrix which maps the state vector parameters to the measurements and z_k is the vector which contains the measurement noise for each element in the measurement vector. This measurement formula is also referred to as output equation.

If we want to use a standard Kalman filter to estimate a signal, the process that is being measured needs to be described by linear system equations.

2.2 Applying Kalman Filters to Travel Predication

Kalman filtering models have the potential to adequately accommodate traffic fluctuations with their time-dependent parameters [8-11]. They have been used extensively for predicting bus arrival time [8]. KFs use previous journey data and present observations and measurements as well as a dynamic model, in order to estimate how the 'state' evolves over time and to make future predictions [8,12]. Crucially, Kalman Filters can be applied during an active journey and hence observations of the current journey are included in the estimation. This also allows to include further information

about the journey to feed into the model. Specifically, in our approach we combine Kalman Filters with information obtained from social networks.

In order to determine the present 'state' of a vehicle, any available data for the position and speed (e.g from GPS) should be taken into account, however, it should be recognised that none of these observations are perfect. Every measurement has uncertainty, be it due to measurement inaccuracies or sampling issues. These observations would be combined with a 'model' of the travelling behaviour such as the waiting duration if the car stops at traffic lights, or the reduced speed of the car if it slow down due to being near a school. Importantly, the model and the observations are not perfect and the real-world travelling behaviour will not correspond exactly to the model. Thus the model includes dynamic parameters like traffic congestion etc.

Existing arrival time prediction approaches often detect unforeseen hold-ups only with considerable delay. For instance, traffic congestion somewhere on the route is often detected using travel speed information from other vehicles. Consequently, only after a rather large number of vehicles got stuck queuing to get past an accident, is the arrival time estimation updated.

On the other hand, information on road incidents can appear on social networks very rapidly. Thus this paper proposes to integrate such information with Kalman Filter prediction approaches in order to improve the accuracy of the prediction. We propose to compare prediction of conventional GPS based systems with social media information supported systems. As this paper discussed previously, any delay information could initially be 'linearly' added to determine total KF based arrival estimation times. For instance, if a KF model estimated arrival time without external delay information is 10 minutes, and the delay info is 2 minutes, then the updated/total arrival time = 12 minutes.

2.3 Experimentation with a Kalman Filter Model

In order to demonstrate the use of Kalman Filters, this subsection presents a Kalman model and initial experimentation employing it to estimate the position of a car during a journey. In the simulation, it is assumed that the journey takes around 55 minutes, and that the car travels at about 40 km/h for some distance, then slows down to less than 10 km/h and then continues to travel at about 40 km/h until just before it arrives at its destination. Process noise during the journey (delays caused by traffic lights, roadwork etc.) is set at 26 minutes and the noise around 10 minutes. The measurements are updated and fed into the model once every minute.

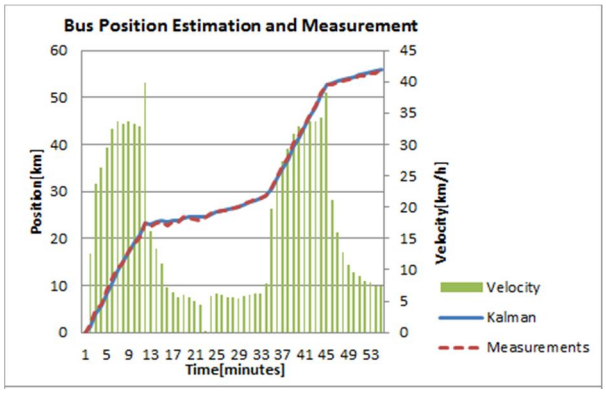

Figure 2: Kalman Filter model results.

Figure 2 shows the results from the Kalman Filter model. The solid line represents the estimated position by the Kalman Filter whereas the dashed line indicates the measured vehicle position. The vertical lines indicate the velocity of the vehicle. When compared with the measured vehicle location, the curve for the Kalman estimated position is smoothed due to the noise removal. The good accuracy of the model when compared to the measured values is one of the features of the Kalman Filter. In the period between the 13th and 20th minutes, there is a gradual change of velocity from 40 km/h to 10 km/h, possibly due to an accident on the road. The delay is estimated to be around 15 minutes, and then (after 35 minutes) the vehicle continues to move at about 40km/h until the 45th minute when it slows down again before arriving at its destination. In this experiment, the noise/delay is added linearly to system. It is expected that the information pertaining to the delay is received from social network sources (e.g. Twitter).

3 Using Kalman Filters with information from social network

Social networks such as Facebook or Twitter allow users to exchange messages. Especially short messages posted on Twitter reflect events in real time as they happen. Hence, such content is particularly useful for real-time identification of events. Twitter provides hashtags, which allows users to relate their messages to particular events or topics. Specifically, road conditions and incidents are frequently discussed on Twitter. An example of this real time information is shown in Figure 3.

Figure 3: Information on Social Networks relating to road traffic.

It is important to realise that our approach does not make use of the social network of an individual, but rather gains information from looking at a larger scope, e.g. all messages sent from a certain geographic location such as a city or messages containing particular keywords. Filtering useful information on Twitter in real time is a challenging problem, due to the use of natural language and immense volume of data [13]. In fact extracting all of the information from any Twitter message is an AI-complete problem. Twitter users post messages with a variety of content types, including personal updates and fragments of related information [14]. However, a number of tools exist which analyse and extract information from Twitter messages [15]. Figure 4 shows Topsy as an example. The use of key words or the use of a specific format for Twitter messages including traffic information can help the analysis. Indeed Topsy allows to filter messages according to keywords, hashtags etc. The use of such approaches requires further study.

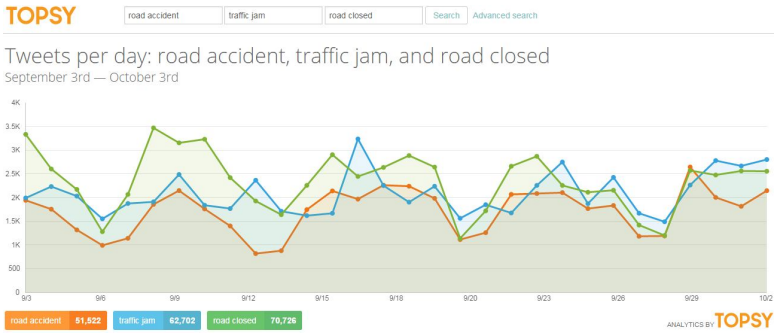

Figure 4: Keyword analysis of Twitter messages using Topsy.

While travel time data can be obtained through various sources, such as loop detectors, microwave detectors, radar, etc., it is unrealistic to hope that the whole roadway network is completely covered by such data collection devices.

Using Twitter messages such as "M8 Delays of 15min Near J25/J24 westbound" can be leveraged. The information about 15 minutes delay can be input linearly into Kalman Filter models.

Thus besides extracting the actual piece of information from social network messages, a major concern is if the sender and thus the message can be trusted.

4 Using Trust in the System

When receiving messages relating to road traffic in social networks, trusting the sender of messages is crucial. Malicious users may insert bogus messages and distribute false information. Thus a number of criteria can be considered to establish a level of trust in the information. For instance:

- Time of the message (messages sent very soon after the particular event are more valuable, very old messages are out-of-date and counter-productive).
- Location of sender of message (a sender close by the incident is more reliable)
- Sender of the message (messages from a trusted sender are more valuable).

Clearly, the reputation of the sender is important. We propose to use user reputation built up in online social networks as a base to verify the credibility of senders and their messages. Trust within social networks has received a lot of attention recently. An overview can be found in [22]. In our approach, the concepts of Social Trust [16] and degrees of separation [17] are employed to identify the level of trust in messages sent by the sender. To find a trustworthy opinion, relationship and experience are two major features that have to be considered. Thus trust can be determined by the degree of relationship between two people, a closer relationship implies more trustworthy information and opinions. Personal information and experience from previous interactions could be seen as another trust metric [18]. Furthermore, trust can be founded in the roles of users. For example, information from a government agency or the police are seen as trustworthy. In terms of degree of relationship, the number of mutual friends, their behaviour, and relationship history can be used to define a level of trust.

Building trust in social networks, based on popularity and engagement [19], is another way to identify trusted information. Popularity trust refers to the acceptance and approval of a member by others in the community, while engagement trust captures

the involvement of someone in the community. Popularity trust can be seen to reflect the trustworthiness of a member in the community, while engagement trust reflects how much a member trusts others in the community.

The analysis of social media data allows to establish whether particular messages can be used to verify the credibility of messages. Such an analysis can be improved by joining information from different social networks. For instance a Twitter user may have an extensive Facebook or Linked-in network which when combined allows building up a more comprehensive picture. Below we highlight trust measures offered by a number of example social network and e-commerce systems.

Amazon uses a ranking system to rank the sellers based on the customer feedback and the number of products sold by the seller on Amazon. Amazon's seller rating is based on an equation, seller rating = total points/total orders.
eBay uses a reputation system to establish trust. eBay collects information on the past behaviour of sellers and buyers. Negative feedback can only be left by buyers. Buyers can leave detailed feedback for sellers in 4 specific categories [20].
Linkedln is a social networking website for people in professional occupations. One purpose of the site is to allow registered users to maintain a list of contact details of people with whom they have some level of relationship. LinkedIn offers a higher degree of linking of people in certain settings. For instance, LinkedIn profiles usually have a larger number of work colleagues as part of a person's profile than other social media networks. As users often know people in their network personally, this can be used to indicate a higher level of trust.
Facebook provides a social network platform for users to connect and to exchange messages. It provides tools to establish the degree of separation between two users. On average, the distance between any two members is 4 degrees [21]. A lower degree of separation indicates a closer acquaintance and thus a higher level of trust.

Finally, there may be some privacy concerns with such a system as users will need to share their current location and information on journeys. Privacy is a complex issue will require further study. However, the system will support appropriate message encryption when sending location and journey information to/from the system. Furthermore, the system will not store such information after the processing is complete. Information shared by users via social networks (road condition information) is regarded to be provided with the knowledge that it can be accessed by other users. Information on group membership and connections/friends on social networks will not be shared with other users.

5 Conclusion

This paper has presented a general framework for a novel approach complementing Kalman Filter models with information from social networks. Kalman Filter models are well established models to predict bus arrival times. The strength of KF models is their ability to predict or estimate the state of a dynamic system from a series of noisy measurements as well as their ability to be executed during a journey. Based on their strength, additional credible and trusted information from social networks can be used to increase the prediction accuracy of KF models. Crucially, as malicious messages can easily be inserted in social networks, only trusted messages from a trusted source must be used. Integrating the information from social networks with the Kalman model, defining a robust approach to trust and ensuring the users' privacy are key elements of future work.

References

1. Liu, Hao, et al. "Predicting urban arterial travel time with state-space neural networks and Kalman filters." Transportation Research Record (2006): 99-108.
2. Wang, Yibing, Markos Papageorgiou. "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach."Transportation Research (2005): 141-167.
3. Li, Li, Xiqun Chen, and Lei Zhang. "Multimodel ensemble for freeway traffic state estimations." IEEE Transactions on Intelligent Transportation Systems 15.3 (2014): 1323-1336..
4. Zhu, Tongyu, Jian Dong, Jian Huang, Songsong Pang, and Bowen Du. "The bus arrival time service based on dynamic traffic information." 6th Intl. Conf. Application of Information and Communication Technologies (AICT), 2012, pp. 1-6. IEEE, 2012.
5. Williams, Billy M., and Lester A. Hoel. "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results." Journal of transportation engineering 129, no. 6 (2003): 664-672.
6. Li, Shu-Bin, Bai-Bai Fu, and Wen-Xiu Dang. "Dynamic Analyses of Urban Expressway Network with Microscopic Traffic Flow Model Integrated Variable Speed Limits." Journal of Nonlinear Dynamics 2014 (2014).
7. Nath, Rudra Pratap Deb, et al. "Modified K-means clustering for travel time prediction based on historical traffic data." Knowledge-Based and Intelligent Information and Engineering Systems. Springer Berlin Heidelberg, 2010. 511-521.
8. Altinkaya, Mehmet, and Metin Zontul. "Urban Bus Arrival Time Prediction: A Review of Computational Models." Intl. J. Recent Technology and Engineering (IJRTE)(2013):2.4
9. Simon, Dan. "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms." IET Control Theory & Applications 4.8 (2010): 1303-1318.
10. Yang, Jiann-Shiou. "A study of travel time modeling via time series analysis." Control Applications, 2005. CCA 2005. Proceedings of 2005 IEEE Conference on. IEEE, 2005.
11. Chien, Steven I-Jy, Yuqing Ding, and Chienhung Wei. "Dynamic bus arrival time prediction with artificial neural networks." J. Transportation Engineering128.5 (2002): 429-438.
12. Simon, Dan. "Using nonlinear Kalman filtering to estimate signals." Embedded Systems Design 19.7 (2006): 38.
13. Agrawal, Divyakant, et al. "Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks." *Databases in Networked Information Systems*. Springer International Publishing, 2014. 1-16.
14. Jussila, Jari, et al. "Visual network analysis of Twitter data for co-organizing conferences: case CMAD 2013." *47th Hawaii Int. Conf. System Sciences (HICSS),* IEEE, 2014.
15. Menéndez, Héctor D., Carlos Delgado-Calle, and David Camacho. "TweetSemMiner: A Meta-Topic Identification Model for Twitter Using Semantic Analysis." Intelligent Data Engineering and Automated Learning–IDEAL 2014. Springer, 2014. 69-76.
16. Golbeck, Jennifer. "Introduction to computing with social trust." Computing with Social Trust. Springer London, 2009. 1-5.16.
17. Abidin, Ahmad Faisal Amri and Kolberg, Mario. "Establishing Trust in VANETs Using Information From Social Networks." Postgraduate Network Symp. (PGNet) (2013):20-25.
18. Golbeck, Jennifer, and James Hendler. "Inferring binary trust relationships in web-based social networks." ACM Transactions on Internet Technology (TOIT) 6.4 (2006): 497-529.
19. Nepal, Surya, Sanat Kumar Bista, and Cecile Paris. "An association based approach to propagate social trust in social networks." UMAP Workshops. 2012.
20. Jøsang, Audun, Roslan Ismail, and Colin Boyd. "A survey of trust and reputation systems for online service provision." Decision support systems 43.2 (2007): 618-644.
21. Backstrom, Lars, et al. "Four degrees of separation." Proceedings of the 3rd Annual ACM Web Science Conference. ACM, 2012.
22. Sherchan, W., Nepal, S., and Paris, C. A survey of trust in social networks. ACM Comput. Surv. 45, 4, Article 47 (August 2013), http://doi.acm.org/10.1145/2501654.2501661.