Paper Title: A Scalable Approach towards Predicting Travel Times of Commuter Buses, Trains & Ferries in Real Time

Authors: Steven Hayman[1], Vikash Kumar[2], Karim Cisse[1], Shiv Maheshwari[1], Edwin Ng[1]

Affiliations:
1 Connected Journey Solutions, New Zealand Transport Agency
2 Customer Experience & Behaviour, New Zealand Transport Agency

What's the strategic, technical or professional context for your paper?

*Context should outline your intent to develop knowledge in the profession and demonstrate the value of looking beyond your personal interest or that of your employer.

Context

Dynamics of transportation in urban environments is rapidly changing. This is triggered by a higher concentration of population resulting in increasingly complex mobility needs. Transportation systems need to adapt and innovate to keep up with the changes in a sustainable manner. Increasing the participation in public transport system is one way to do this. We want to benefit from recent advancements in AI to help improve the efficiency of the public transportation network across New Zealand. We have used data science techniques to forecast travel times of public buses, trains and ferries with acceptable accuracy. This will help in travel demand management, accurate real-time reporting, traffic planning and also increase public's trust in the efficiency of our networks.

What will attract people to read your paper and attend your presentation? What kinds of people or roles would benefit most? *This is your main selling point – for the people attending the conference	**Relevance** We will be presenting the lifecycle of modelling and deployment of a Machine Learning solution at scale. We will share the experience of our journey from drawing board to deployment and considerations we made to ensure the timelines and expectations stayed realistic. This will include a discussion on how we were able to quickly scale up to predicting more than fifteen thousand trips a day using cloud solutions.
What is the particular question, issue or idea you intend to address in this session? * Consider this an executive summary but be specific and relevant to your audience.	**Focus** We have developed a tool for predicting travel times of public buses, trains and ferries in real time and with acceptable accuracy. This tool uses live GPS data from on board units of these vehicles. It is presently being tested for forecasting arrival and departure times of buses, trains and ferries using data from Auckland region and for buses in Queenstown. We have consciously used techniques that allow us to scale this to do predictions for public transport in other cities if live GPS feed is available. We foresee several applications including use in real-time boards on stations, journey planning, demand management, etc.

Introduction

Accurate predictions of travel times of public transport vehicles in a transportation network are essential to delivering a high-quality service. Making these predictions is a non-trivial problem due to various factors such as weather, roadworks, and special events that can affect a vehicle's travel time.

Besides known factors, predictions can be affected by several unforeseen considerations such as a breakdown, crash, flash strike, etc.

This paper discusses a travel time prediction engine which was built for New Zealand transport Agency (NZTA) to provide real time information on the estimated arrival times of vehicles along routes in their network. An onboard GPS device provides the real time location of each vehicle. This information is used to dynamically update the estimates generated by the prediction engine.

A model was developed based on up to three months of historical data on 11 Auckland bus routes. The model utilizes Kalman filtering combined with historical means to generate predictions. It considers external factors such as variations arising due to the time of day or the day of week. By using the limited available data and accounting for the aforementioned external factors, promising results were achieved. The strength of this approach is the ability for the model to quickly adapt to uncertainties typically associated with urban road travel.

The model has since been scaled to predict travel times of all Buses, Trains and Ferries in Auckland region in real time using GPS readings from their on-board units. It is also performing predictions on live data from Queenstown buses. The system is currently being deployed by New Zealand Transport Agency (NZTA).

Structure of paper: In the next sections, we will describe the background work leading up to this solution including defining the problem statement. We then delve into some examples of how this problem was tackled in some other places. We then describe our solution in detail and provide evidence of its performance on different metrics. We then briefly discuss its ongoing implementation at NZTA. Finally, we conclude with a summary of how we are building upon this work to create more value.

Background

This project started as a partnership between Auckland Transport (AT) and New Zealand Transport Agency (NZTA).

Problem Statement

The problem we are trying to solve can be summarized under the following points:

1. Develop a technique that could reliably predict arrival and departure times of commuter buses, trains and ferries at upcoming stops along its route
2. The predictions are to be done in real time for all vehicles that are currently active on their routes

3. The solution should be scalable to different modes of public transport (e.g. trains, ferries)
4. The solution should also be general enough, so it is able to be deployed for travel time predictions in other cities

Figure 1: Real time board for buses

The end goal is to display expected arrival times of buses, trains and ferries on real time boards (Figure 1). It would also be used for journey planning by apps like Journey Planner[1].

Related Work

There have been several attempts to come up with an efficient technique for predicting travel times of buses. Most of these fall in one of the following categories (Zegeye Kebede Gurmu, 2014; Mehmet Altinkaya, 2013):

Historical Averages

Travel times of buses are estimated based on previous journey of the buses along the same route. There is often no consideration to current traffic conditions and therefore the model works best in situations where traffic conditions are stable. While this simple technique may work well in some rural areas, it will provide erroneous estimates in Auckland due to the variability in traffic conditions.

[1] https://www.journeys.nzta.govt.nz/

Regression Techniques

Regression techniques aim to predict a dependent variable (arrival times in our case) based on their relationship (a mathematical function) with a set of independent variables. Independent variables like distance between stops, dwell times at stops, passenger count and weather have been used in the past to predict the arrival times of buses. Regression is able to take into account unstable traffic conditions and has been shown to give decent results when compared to historical averages. Using regression, one can also establish which features are more and which are less relevant to predict travel times. However, due to the non-deterministic nature of this problem, it is hard to encode "unforeseen" events that may affect the travel times. Regression techniques are not good at handling such randomness.

Kalman Filters

Kalman filtering approach is used to provide estimates of unknown variables by calculating a joint probability distribution over past observations of noisy data. It predicts future states of a system based on its current readings by finding parameters of interest from inaccurate and uncertain observations. This approach has been tried in various hybrid forms and shown to outperform historical averages and regression techniques (Yanguo Huang, 2013; Jinjin Zhang, 2015).

Artificial Neural Networks

Artificial Neural Networks (ANNs) can be used in modelling complex non-linear relationships between predictors (input variables) and output data. This feature of ANNs can be used to make travel time predictions which themselves are a result of complex relationships between a number of external factors. Different types of ANNs have been tried in research with good results for travel time prediction of buses (Zegeye Kebede Gurmu, 2014; Yuhan Jia, 2017; Jiang, 2017; JoharAmita, 2016). Evidence has been provided of how they outperform other methods of prediction in such problems.

Our Solution

Our aim was to develop a technique that could make the best use of information that is readily available from public buses, trains and ferries in New Zealand.

Data Description

Real time GPS data has been collected from vehicles, with a reading recorded approximately every 10 seconds. Up to 3 months of historical data from several Auckland buses were made available to develop and test the model. Data from up to 17 representative bus routes in Auckland were received and after cleansing, we were able to use 11 of these routes for training and testing. The dataset contained the following fields:

- Service date – Date for which the data was recorded
- Route number – Unique number identifying the bus route
- Route name – Name of the route (e.g. Te Atatu Peninsula to Henderson)
- Route destination – Final stop on the route
- Service start time – Start time of the bus on this trip from the first stop
- Stop sequence number – Incremental sequence of numbers from first till last stop on the route
- Stop name – Name of the corresponding stop (e.g. 55 Matipo Rd)
- Timetable arrival time – Scheduled arrival time of the bus at the corresponding stop
- Sighting time entry – Actual arrival time of the bus at that stop
- Timetable departure time – Scheduled departure time of the bus from the corresponding stop
- Sighting time exit – Actual departure time of the bus from that stop
- VehicleID – Unique ID for each vehicle

The Hybrid Kalman Filter Solution

Despite the promise of ANNs, we started to build our algorithm by applying Kalman Filter (KF) technique. The decision was based on some practical considerations:

1. It was unclear if there would be access to sufficient data to train a neural network upfront – KF can work with less data
2. KF is simple to implement and would allow us to quickly test the viability of our approach
3. KFs are good at handling random variations in processes. They can constantly update as variations get too large. This was appealing as we wanted a system that adapts quickly to uncertain behavior that bus journeys naturally exhibit.

Departure times derived from GPS data were used to build our model and enhanced using historical data on departure times. KF works on the principle of minimizing the loss function and works on a "predict-update" cycle. A model and a sequence of approximate measurements are provided to the Kalman filter. As the measurements are read into the filter, it can correct its estimate of the state of the system. If the model is sufficiently good, the filter can make accurate predictions of states further along the sequence. The Kalman filter's ability to correct itself during the process provides an advantage over other techniques.

The current model considers variations observed from historical data. So, the effect of e.g. traffic lights on the travel time of the bus is considered implicitly. The external forces are calculated by considering the average travel times between stops from previous trips on the same route starting at the same time of day as the trip that predictions are being made for. This calculation is being done by looking at up to 3 months of historical data, meaning that differences between days of the week are also being accounted for.

The current implementation still does not consider external factors like one off events (concerts, sports, parade, …), weather, school holidays, etc. (Yuhan Jia, 2017).

Evaluation

In this section we will describe the performance of the model in terms of accuracy and its ability to handle uncertainties.

Accuracy

Visual evidence of the accuracy of the algorithm on Auckland routes 955 and 142 can be observed from Figures 2 and 3 respectively. The forecast departure times of the buses (blue lines) when predicted from around midpoint in their journey all the way to the end which is roughly 15 stops away, closely follow the actual departure times (green lines) of the buses.

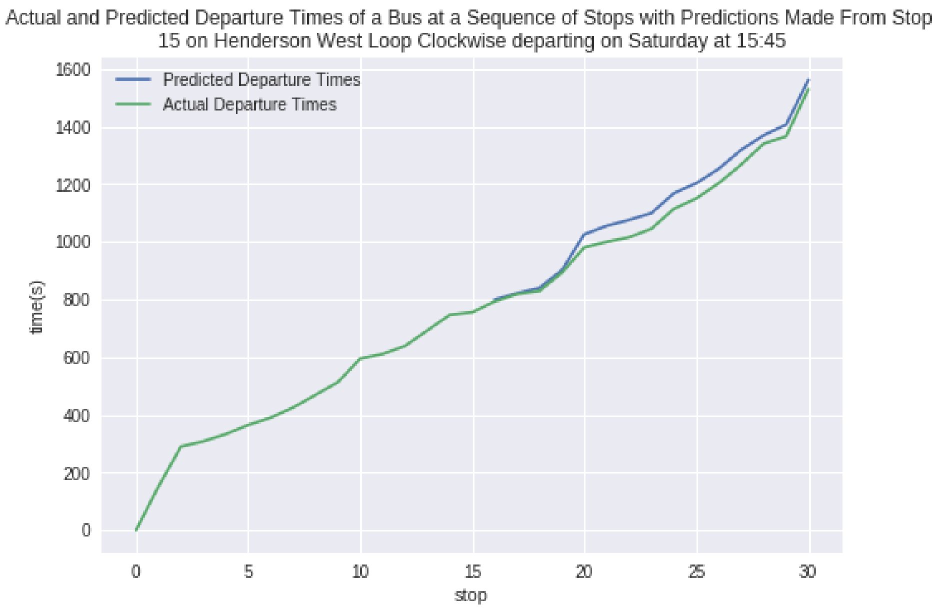

Figure 2: Predictions from stop 15 (roughly midpoint) on route 955

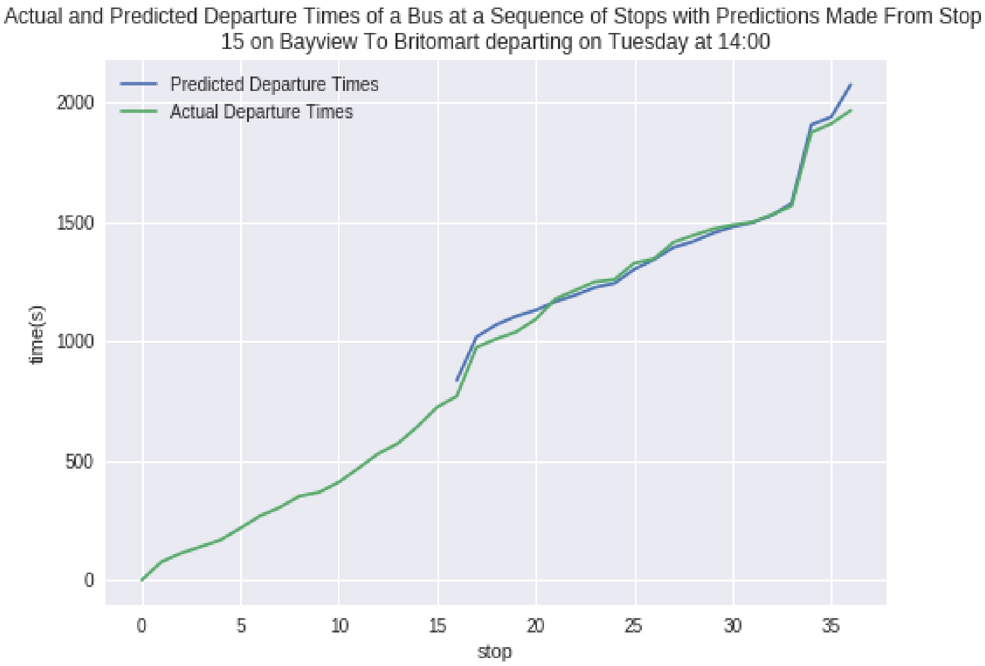

Figure 3: Predictions from stop 15 (roughly midpoint) on route 142

It is observed that when predicting bus departure times 29 stops away, 75% of all the predictions are within 145 seconds of actual arrival times of the buses. When predicting 10 stops away, the accuracy improves to being within 70 seconds - 75% of the time (Figure 4).

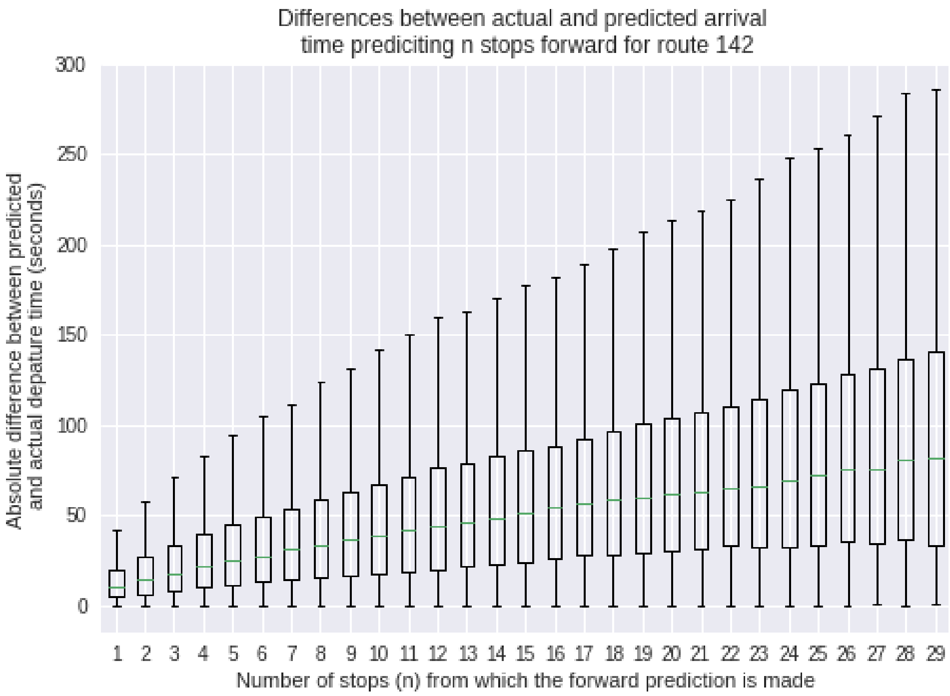

Figure 4: Accuracy based on number of forward stops predicted

Dealing with Uncertainties

The developed model is quick to adapt to unexpected delays along the route. As seen in Figure 5, the bus takes around 500 seconds more than the usually observed time to travel between stops 200 and 300 (red dot). Travel times between subsequent stops (300-400 and 400-500) follow usual pattern. As expected, future predictions when made from stop 1 are inaccurate due to the unexpected delay between stops 2 and 3 (red line in Figure 6). However, this is quickly rectified when information of that delay is available in next stops as shown by purple and yellow lines in Figure 6.

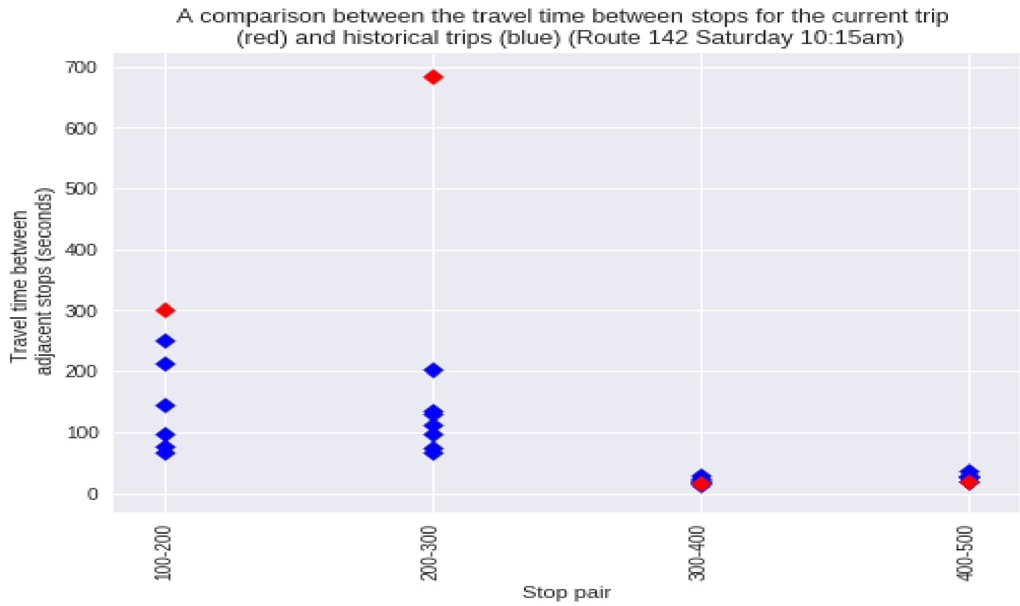

Figure 5: Travelling from stop 200 to 300 takes unusually longer in this trip

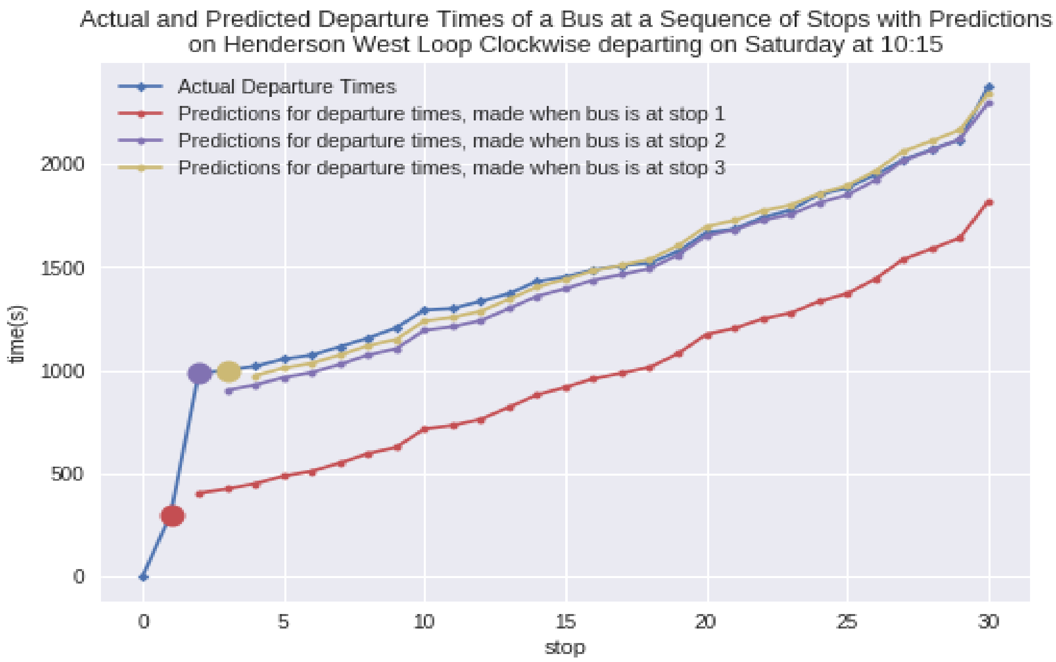

Figure 6: Adapting to unexpected delay along the route

Implementation

Most of the components for this project have been deployed on the Google Cloud Platform (GCP)[2]. Live GPS feed of vehicles is received from Auckland Transport via a REST connection which is then processed to detect scheduled stops. The Prediction Engine is triggered whenever a stop is detected or when any on demand prediction is requested. The forecast arrival and departure times for all subsequent stops along the route is computed and sent back to Auckland Transport or fed into NZTA's journey planner app.

At the time of writing, this engine is detecting up to 300,000 stops across up to 15,000 full trips per day. Data from 19th March 2019 is shown in Figure 7.

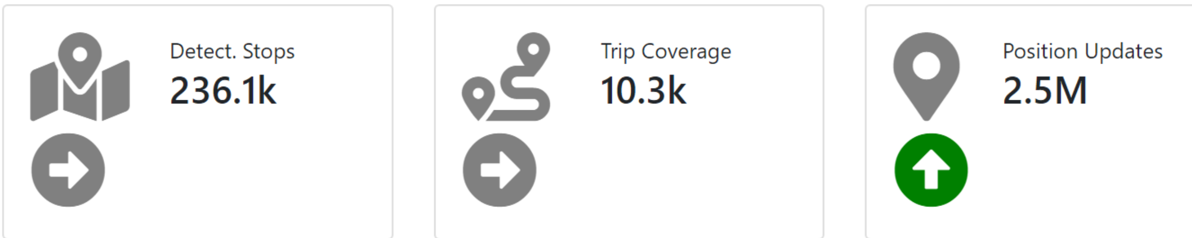

Figure 7: Predictions Statistics on 19th March 2019

Conclusions and Future Directions

In this paper, we have discussed our four-month journey from drawing board to deployment of a scalable algorithm for travel time prediction of public transport vehicles. We described some common techniques that have been tried to solve similar problems in the past. We elaborated on why we chose a Kalman Filter based approach and showed the results we achieved. We also described how we are deploying this on Google Cloud Platform.

While we have achieved satisfactory results and are in the process of deploying it live, there are a few aspects we are looking to improve this technique with.

Artificial Neural Networks

ANNs have been shown to outperform Kalman filters in literature (Jiang, 2017; JoharAmita, 2016). They are also capable of modelling complex relationships with external factors better than KFs. We have developed a prototype using ANNs with encouraging results and are in the process of adding additional features to see if it outperforms our original model.

---

[2] https://cloud.google.com/

Additional Features

The current model does not cover effects of special events, weather conditions, school holidays, etc. We have implemented a prototype using ANNs that includes the effects of rainfall in predicting travel times. School holiday dates is another feature we are looking to use to improve the accuracy of predictions, especially during the school holiday period.

Big Picture

This algorithm has so far been developed in isolation with a goal of displaying expected arrival times in real time boards. We are exploring possibilities of linking this with other projects like NZTA's Journey Planner to allow users to plan their travel based on these predictions instead of actual schedules that often go off during peak hours.

Finally, we would like to thank Auckland Transport for their support with data without which this work would not be possible.

References

Jiang, F. (2017). *Bus Transit Time Prediction using GPS Data with Artificial Neural Networks.*

Jinjin Zhang, Y. W. (2015). Travel Time Prediction Method Based on Improved Kalman Filter., (pp. 630-640).

JoharAmita, S. P. (2016). Prediction of Bus Travel Time Using ANN: A Case Study in Delhi. *Transportation Research Procedia*, 263-272.

Mehmet Altinkaya, M. Z. (2013). Urban Bus Arrival Time Prediction: A Review of Computational Models. *International Journal of Recent Technology and Engineering*.

Yanguo Huang, L. X. (2013). Urban Expressway Travel Time Prediction Method Based on Fuzzy Adaptive Kalman Filter. *An International Journal of Applied Mathematics & Information Sciences*, 625-630.

Yuhan Jia, J. W. (2017). Traffic Flow Prediction with Rainfall Impact Using a Deep Learning Method. *Journal of Advanced Transportation*.

Zegeye Kebede Gurmu, W. F. (2014). Artificial Neural Network Travel Time Prediction Model for Buses Using Only GPS Data. *Journal of Public Transportation*.