

Real-time forecasting of bus arrivals using GPS-data

October 2019

Background

Why does not the bus arrive when it is supposed to? That was the spark that ignited the idea for this project. The traffic in a city constitutes a very complex system, and there are numerous parameters that determine if a bus or car arrives at its destination on time. In Uppsala county the public transportation is provided by Upplands Lokaltrafik (UL). To provide real time estimates of bus arrivals to their passengers, they use a third-party tool built on top of a proprietary algorithm.

Preliminary works

During the spring of 2019 a team of students conducted their bachelor thesis at UL with the aim of predicting bus arrival times. To accomplish this they used a neural network approach with inputs: *time travelled since the start of journey*, *time of day*, *current bus stop* and *final bus stop*. The output of their model is the travel time to the specified stop. With this approach they managed to achieve a better prediction accuracy than the currently used third-party tool [1]. There has been other research in this field. Wei Fan and Zegeye Gurnu compared different models and found that *Artificial Neural Networks* (ANNs) can be a good model for this kind of problem [2].

Project description

This project is in collaboration with and under the supervision of UL, and intends to improve their buss arrival forecasts, providing more accurate predictions of the time when a bus will reach a particular bus stop. The project will not necessarily use a neural network as in [1]. For this to be achieved, real-time positional data will be used. The latter will be acquired in partnership with a project group from the course *Project in Embedded Systems*.

The partner group will design and produce a real-time position tracker and a web platform. This, combined with the forecasting provided by this project will make a proof-of-concept product that aims to give better predictions than currently available forecasting.

Work plan

1. Initial preparation and planning of the project.
2. Data preprocessing - Preprocess the GPS data into a useful state for analysis.
3. First stage of data analysis - Investigate baseline model performance and determine feasibility of a Gaussian process analysis versus a neural network approach.
4. Second stage of the data analysis - Implement the model of choice for real-time forecasting.
5. Validate the prediction accuracy of the model and compare to the forecasting method currently in use by UL.
6. Third stage of data analysis - Investigate if any external parameters (weather, traffic conditions, etc.) can be used to increase model performance.
7. Create a poster and report.

It is difficult to estimate how the time will be distributed between the above steps. But as an estimation we believe that most of the work will be spent on steps 2 and 4.

Tools

To perform the majority of the steps above the interpreted language Python version 3.7 will be used. The data analysis stage will make heavy use of the python libraries NumPy and PyTorch.

References

- [1] Johan Rideg, Max Markensten, 2019, 'Are we there yet?: Predicting bus arrival times with an artificial neural network', (Unpublished Bachelor Thesis), Uppsala Universitet, Uppsala, Sweden.
- [2] Wei Fan, Zegeye Gurmu, 2015, 'Dynamic Travel Time Prediction Models for Buses Using Only GPS Data', International Journal of Transportation Science and Technology, vol. 4, issue 4, pp. 353-366.