CLOUD-BASED MACHINE LEARNING FOR BUS ARRIVAL TIME PREDICTION

OLCZYK Adrian, GAŁUSZKA Adam

Silesian University of Technology, Institute of Automatic Control, Gliwice, Poland, EU

Abstract

The bus arrival time is one of the key elements in public transport information systems. The amount of Automated Vehicle Location (AVL) systems is growing, therefore in this paper we aim to provide a cloud-based machine learning solution of this problem. Using bus location data, we created models in Microsoft Azure Machine Learning Studio using different machine learning methods: Artificial Neural Network (ANN), Support Vector Machine (SVM) and Linear Regression (LR). We validated the methods using historical data and compared the results to naïve predictions that use either historical data with a delay or a vehicle speed.

Keywords: Public transport network, bus arrival time prediction, machine learning, artificial neural network, support vector machine, linear regression

1. INTRODUCTION

The bus arrival time is one of the key elements in public transport information systems. In this paper we aim to create a cloud-based machine learning solution for bus arrival time prediction problem. We will test the solution for an increasing distance between last known bus location and prediction location. Also we'll compare the models to naïve solution, where we'll get the delay information from the last known bus location and use it as a prediction on a subsequent location.

2. BACKGROUND

It the past years a number of methods have been developed for bus arrival time prediction. The basic, naïve approach uses involve bus delay measurement on preceding stop and transferring this value to the stop on which the prediction is being made. Fortunately, over the years more sophisticated methods have been introduced.

The bus arrival time prediction research begun by the end of 1900s with Lin and Zeng [1] work. They extracted bus operation information from vehicle monitoring systems and developed an algorithm that was based on historical data. They used bus location data and schedule data to calculate bus delay. Then they compared the difference between scheduled and actual arrival times and on this based they predicted the delay.

Chien et al. [2] proposed two models based on Artificial Neural Network (ANN). First was Stop Based Model, in which whey measured vehicle delay only on the stops. The difference between scheduled and actual arrival time was used to predict arrival time on succeeding stop. The second model (Link Based Model) were based on route division between stops (intersection were used as a point of reference). The results showed that although Link Based Model required more preparation (intermediate expected times had to be calculated upfront) it outperformed Stop Based Model even if the number of intersections was relatively small.

In recent years more machine learning methods have been developed, tested and compared [3,4]. It includes methods such as support vector machine (SVM), artificial neural network (ANN) [5], k-nearest neighbours algorithm (k-NN) and linear regression (LR). Depending on the study, quality of the data and implementation every method is able to outperform other ones given the right context [6].

Additionally, some research proposes some heuristic methods that contain two steps. First, a historical data trained SVM, and second, a real-time data Kalman Filter.

Most of the systems rely on the Global Positioning System (GPS) or some other vehicle location solution with some notable exceptions. For example, system could be developed to rely completely on collaborative passenger information. This approach is completely independent from bus operation companies and could be applied to almost every public transport system with enough passengers willing to participate. The prediction algorithm used are not different to other systems, only the vehicle location comes from other source.

3. THE MODEL

Bus arrival time prediction at bus stop can be described in the following way: given the bus route location, the goal is to predict arrival time at the bus stop at the desired bus stop. The **Figure 1** illustrates the prediction framework for this study.

When a bus of given bus route arrives at Location A the bus arrival time (T_a^A) is recorded. Then, given the scheduled arrival time (T_s^A) and actual arrival time (T_a^A) the delay (d^A) is calculated (**Equation 1**).

$$d^A = T_s^A - T_a^A \tag{1}$$

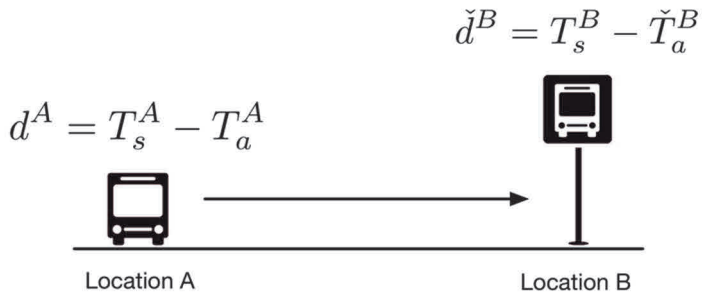

Figure 1 delay on location A and B

The predicted vehicle arrival delay $\check{d}^B$ on Location B can be calculated using **Equation 2**.

$$\check{d}^B = T_s^B - \check{T}_a^B \tag{2}$$

3.1. Support vector machine

SVM is a type of machine learning algorithms based on statistical learning. It can be used to map the input-output relationship for the non-linear system. The solution is always unique and globally optimal since training SVM is equivalent to solving a linearly constrainer quadratic programming problem.

3.2. Artificial neural network

ANN is a mathematical model that has been inspired by the human's brain neural structure. ANN processes processes information by interaction between neurons with differently weighted weights. ANN has the ability to model complex input-output relationships.

3.3. Linear regression

Linear regression is a mathematical approach for modelling the scalar relationship between variables. It's the first type of regression analysis. For vehicle arrival time prediction linear regression model is one of the simplest methods and is used extensively.

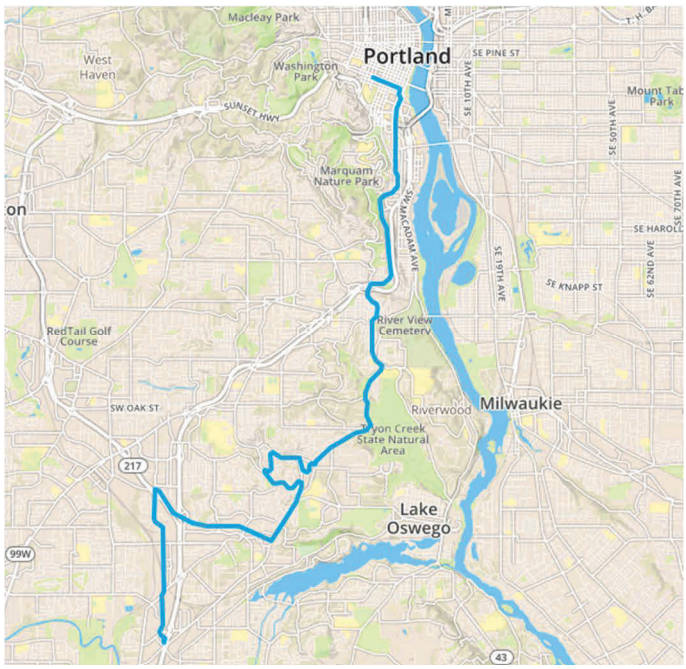

Figure 2 The visualisation of the route 38

4. CASE STUDY

We collected data for bus and rail services in the Portland, Oregon. The acquired data spans over one month, starting on June 1st 2016. General network information and schedules were obtained in GTFS format, and real-time data with vehicle location were recorded and stored in the separate database.

For the analysis we selected a bus route 38, which includes 59 stops. The route is visualised on **Figure 2.**

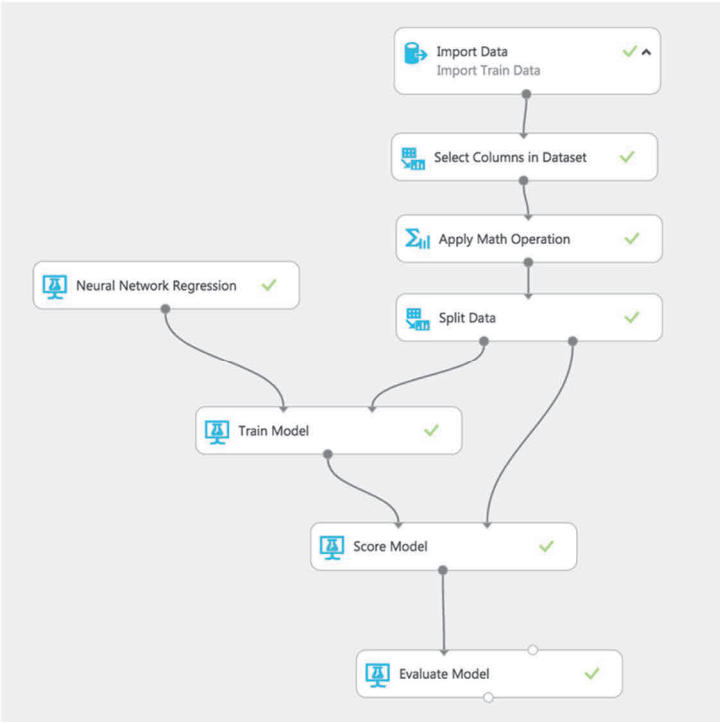

Figure 1 System design on Microsoft Azure Machine Studio

For data analysis and bus arrival time prediction we've used Microsoft Azure Machine Learning Studio. The software allowed computations in the cloud and convenient machine learning model switching.

As an input data we used a bus delay on a four preceding locations **(Equation 1**). The system was designed as on **Figure 3**. We divided data into equal training and validation sets. After a training we validated the model using the validation set and prediction a bus delay on the one of subsequent stops (**Figure 4**).

We used three machine learning models: Artificial Neural Network (ANN), Linear Regression (LR) and Support Vector Machine (SVM) and compared them to naïve delay transfer (COPY) from the last known bus location. The models were validated for a different number of stops between last known bus location and a bus stop on which we were making a prediction.

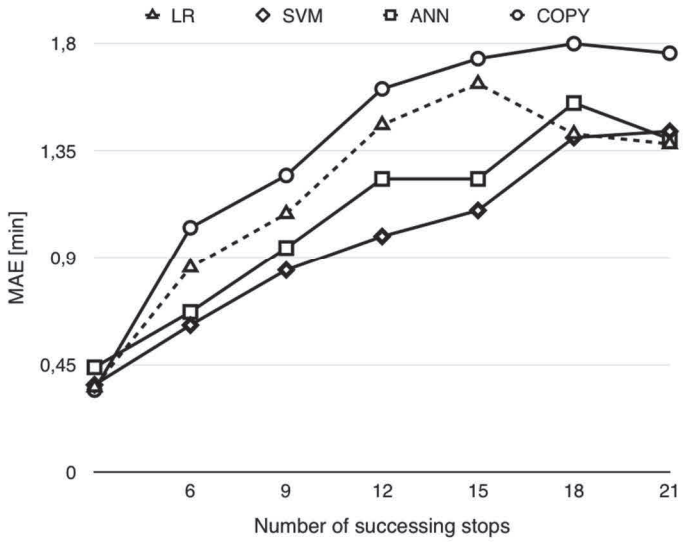

Figure 4 machine learning models (ANN, SVM, LR) compared to route delay copy (COPY)

The results show that with an increase of a number of stops between last known bus location and the prediction location the mean average error (MAE) also increases. We noticed that as overall MAE value doesn't exceed 2 minutes the error for large number of subsequent stops is not growing. The reason for this might be that a average delay for a location is not exceeding certain value, thus prediction should trend to the same value.

From thee machine learning models the SVM outperformed other methods, although ANN also yielded similar results. Naïve method of transferring delay from the last known bus location to the predicted location resulted in a worst prediction, but not completely useless.

5. CONCLUSION

In this paper we case study where we collected a real-time bus location data in Portland, Oregon. Using stored historical values, we created three machine learning models in Microsoft Azure Machine Learning Studio. We used created models for bus arrival time prediction. Results showed that SVM model outperformed ANN, LR and naïve methods.

REFERENCES

[1] LIN W. ZENG J., Experimental Study on Real-Time Bus Arrival Time Prediction with GPS Data, Transportation Research Record: Journal of the Transportation Research Board, No. 1666, TRB, National Research Council, Washington, D.C. 1999, pp. 101-109.

[2] CHIEN S.I.J., DING Y., WEI C., Dynamic bus arrival time prediction with artificial neural networks. Journal of Transportation Engineering 128 (5), 2002, pp. 429-440

[3] YU B., LAM W. H. K. TAM M. L., Bus arrival time prediction at bus stop with multiple routes. Transportation Research Part C: Emerging Technologies 19, 2011, pp. 1157-1170

[4] CHAN, K.S., LAM, W.H.K., TAM, M.L., Real-time estimation of arterial travel times with spatial travel time covariance relationships. Transportation Research Record 2121, 2009, pp. 102-109.

[5] JEONG R., RILETT L.R., Bus Arrival Time Prediction Using Artificial Neural Network Model. In: 7th International IEEE Conference on Intelligent Transportation Systems: (ITSC 2004), 2004, Washington DC.

[6] SHALABY A., FAHRAN A., Prediction models of bus arrival and departure times using AVL and APC data. Journal of Public Transportation 7 (1) 2004, 41-61.