Linköpings universitet | Institutionen för datavetenskap
Examensarbete på avancerad nivå, 30hp | Datateknik
202018 | LIU-IDA/LITH-EX-A--2018/001--SE

Arrival Time Prediction On Buses in Linköping

–

Max Lund, Simeon Jackman, Per Lindström, Elias Alesand, Jakob Lundberg, Sebastian Callh

Handledare : Matthias Tiger
Examinator : Cyrille Berger

Linköpings universitet
SE–581 83 Linköping
013-28 10 00 , www.liu.se

Sammanfattning

`Abstract.tex`

Författarens tack

`Acknowledgments.tex`

Innehåll

1 Introduction

This report describes the work performed in the course *TDDE19 Advanced Project Course - AI and Machine Learning*. The task of the project was to make arrival time prediction for local buses in the Linköping area.

1.1 Problem

The problem at hand is to create machine learning models to do arrival time prediction for local buses. This is done using spatio-temporal data provided by the state transport company, Östgötatrafiken.

1.2 Terminology

Below is a list of the terminology used in this report:

- **Bus line:** All the data points from any bus driving a particular route, in either direction of the line.
- **Journey:** All the data points from a bus driving a between two defined stations (in one direction only).
- **Trajectory:** Several consecutive data points from the same bus driving on the same bus line.
- **Segment:** All data points between two adjacent bus stops in a journey.
- **Data point:** A single data point from any bus.

- **MAPE** The mean absolute percentage error.

1.3 Goals

- Literature study to find out about viable methods.

- Pre-process data into a suitable format for use in machine learning models.
- Predict bus arrival times using a simple baseline method.
- Predict bus arrival times using Gaussian processes.
- Predict bus arrival times using neural networks.
- Evaluate solutions.

1.4 Approach

In this report bus time arrival prediction will be performed using a simple statistical baseline method, Gaussian processes, and neural networks. The theory behind the solutions will be discussed in depth in chapter 2, and the methodology will be discussed in chapter 3. The models are evaluated by their MAPE on identical data sets.

1.5 Delimitations

Generalizing the created models to include any bus line is beyond the scope of this project. Therefore certain problems within the domain have been disregarded, and the problem limited to examining only a few bus lines. Spatial outliers and incomplete journeys within those routes have been removed to reduce the complexity of the models needed to make predictions.

2 Theory

This chapter aims to explain concepts and methods that are used is our implementations.

2.1 Gaussion Processes

A Gaussian Process (*GP*) is, in statistical terms, a stochastic process such that any set of the variables have a multivariate normal distribution. Such a distribution is defined by a mean vector μ and covariance matrix σ, where every random variable spans one dimension. Viewed through the lens of machine learning, the GP serves as a non-parametric regression model.

Gaussian Process Regression

GP regression (also known as kriging) is a non-parametric generative model that builds on the assumption that all data points are drawn from a multivariate normal distribution. A GP f is consequently defined as $f \sim \mathcal{N}(0, \Sigma)$. Zero mean can be assumed without loss of generality and is done for mathematical convenience, but in practice μ could be any suitable mean vector. The covariance matrix Σ is thus the only free parameter to be chosen. Unfortunately, a covariance matrix over a continuous function would exist only in an infinite dimensional space, which is indeed problematic. However, the spaces we are interested in have an inner product corresponding to some covariance function $c(x, x') = cov(f(x), f(x'))$ for the process f and data points x, x', and this function can be kernelised as $k(\theta, x, x') = cov(f(x), f(x'))$ for some valid kernel function k and its hyperparameters θ. It is consequently possible to compute Σ given a set of data points and a kernel function, circumventing the need to construct infinite dimensional spaces.

Thanks to the kernel trick the choice of parameters is consequently not a covariance matrix Σ but a kernel $k(\theta, x, x')$ and its hyperparameters θ. This choice represent our prior over f, and in particular it represents how smooth we believe f is by imposing certain covariance on $f(x)$ for nearby points. This is the key idea of a GP regression and what makes it very flexible: It allows us to specify a prior over f as a function of x, without even knowing f.

2.2 Artificial Neural Networks

Artificial neural networks have shown to to be useful when predicting travel times due to their ability to model nonlinear relationships between features [**brazilANN**][**malaysiaANN**]. This section introduces the parameters that have been considered when creating artificial neural networks for this report.

Activation Functions

The purpose of the activation function is to compute the hidden layer values [**Goodfellow-et-al-2016**]. Two of the most popular activation functions are the sigmoid function:

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.1}$$

and the rectifier function:

$$f(x) = max\{0, x\} \tag{2.2}$$

Using the rectifier function leads to less computationally complex learning than when using the sigmoid function although problems can occur where the backpropagation is blocked by a "dead" neuron due to the hard saturation at 0 [**pmlr-v15-glorot11a**].

Loss Functions

For an artificial neural network to be able to update its weights it needs a loss function that should be minimized in the case of gradient descent [**Goodfellow-et-al-2016**]. Some examples of common loss functions are **mean squared error (MSE)**, **mean absolute error (MAE)** and **mean absolute percentage error (MAPE)**. A problem with MAPE is that when the true value is 0 the function is undefined [**MAPE**]. MSE gives a larger weight to outliers since the metric has an exponential relation to the error.

Hidden Layers and Neuron Count

The more hidden layers and neurons a network has the more computationally complex the learning becomes. Therefore you should not use a model that is more complex than the problem in question requires. Problems that involve large amount of input features like the *ImageNet* contest has been shown to be a case where deep neural networks perform well [**ImageNet**]. However, it has been shown that virtually any function can be approximated with two hidden layers given there are enough neurons [**Demuth**].

2.3 Kalman Filters

Kalman filtering is an iterative algorithm used to increase measurment accuracy by considering multiple consecutive data points observed over time. It can be used in real time as new measurements are made. Kalman filtering has been shown to be useful when predicting travel times [**kalmanPrediction**]. It has also been used together with artificial neural networks with promising results [**kalmanANN**].

3 Method

3.1 Pre-processing

The data for this project was delivered by Östgötatrafiken. The whole dataset is over 300GB in size. The data contains, among other things, GPS coordinates and different types of events representing actions performed by a bus. To be able to feed the data into our models, preprocessing had to be done. The initial data exploration was performed by using jupyter notebooks and the python pandas package. Data operations such filtering and plotting the statistical distributions of the raw data provided an initial overview of the data set. By creating a simple finite state-machine, the data covering a complete journey from a specific bus line could be extracted.

Data structure

The provided data consists of 90 files, where each file represents one day of data. Each file has an approximate size of 5GB. Within the data there are over 20 event-types that represent the state of a bus during a day. For this project four types of events were used, while the others were discarded. The four events used are:

ObservedPositionEvent: Gets triggered every second, contains the GPS data of a given bus.

EnteredEvent: Gets triggered when the bus is within a certain distance to a bus-station. Is used to split the journey into segments.

JourneyStartedEvent: Gets triggered when the bus is assigned a new journey. Is used to determine which line a bus is currently serving.

JourneyCompletedEvent: Gets triggered when the bus has completed a journey. Is used as a flag to determine when a journey has ended.

TODO (this section) For this project XXXX bus lines were selected: a subset of the bus line three and bus line number eleven. Line three has been used as a basis, since we all know that bus and we could assure that there were no irregularities on that line. Line eleven had been chosen because it captures a lot of problems that need to be taken into consideration within our predictions. Such problems are GPS variance due to high buildings, high red light density, and sharp corners.

Detecting Segments

Bus journeys are split into several segments. The *EnteredEvent* is used to detect when a bus is approaching a station, and is also used to split the raw data into segments. The first segment of a journey starts when a *JourneyStartedEvent* triggers. When all the journeys of interest have been collected from the raw data, the collection is investigated in order to find and remove severe anomalies (such as drivers taking a wrong turn). These faulty journeys are discarded since it is not feasible to create a model which takes such complexity into consideration, given the amount of training data available.

3.2 Artificial neural networks

The neural network models were made using *Keras* on top of *Tensorflow*.

Baseline ANN model

A simple model was created to use as a baseline model that can be used to compare other neural network models to. This model predicts the time it will take to travel to the next bus stop. As input this model use time of day normalized to a value in the range [0,1] and the segment for which the observation has been made. The segment input is one-hot encoded meaning that there is an input for each segment in the journey which all have a value of 0 except for the segment of the observation which has the value 1. The network has one fully connected hidden layer with 13 nodes and an output layer with one node. The network uses the *relu* activation function. This model predicts the time in seconds it will take to travel the whole segment. To get a prediction of the time to the next bus stop you need to subtract the actual known time travelled since the previous bus stop from the output of the neural network.

4 Results

This chapter presents the results. Note that the results are presented factually, striving for objectivity as far as possible. The results shall not be analyzed, discussed or evaluated. This is left for the discussion chapter.

In case the method chapter has been divided into subheadings such as pre-study, implementation and evaluation, the result chapter should have the same sub-headings. This gives a clear structure and makes the chapter easier to write.

In case results are presented from a process (e.g. an implementation process), the main decisions made during the process must be clearly presented and justified. Normally, alternative attempts, etc, have already been described in the theory chapter, making it possible to refer to it as part of the justification.

5 Discussion

This chapter contains the following sub-headings.

5.1 Results

Are there anything in the results that stand out and need be analyzed and commented on? How do the results relate to the material covered in the theory chapter? What does the theory imply about the meaning of the results? For example, what does it mean that a certain system got a certain numeric value in a usability evaluation; how good or bad is it? Is there something in the results that is unexpected based on the literature review, or is everything as one would theoretically expect?

5.2 Method

This is where the applied method is discussed and criticized. Taking a self-critical stance to the method used is an important part of the scientific approach.

A study is rarely perfect. There are almost always things one could have done differently if the study could be repeated or with extra resources. Go through the most important limitations with your method and discuss potential consequences for the results. Connect back to the method theory presented in the theory chapter. Refer explicitly to relevant sources.

The discussion shall also demonstrate an awareness of methodological concepts such as replicability, reliability, and validity. The concept of replicability has already been discussed in the Method chapter (3). Reliability is a term for whether one can expect to get the same results if a study is repeated with the same method. A study with a high degree of reliability has a large probability of leading to similar results if repeated. The concept of validity is, somewhat simplified, concerned with whether a performed measurement actually measures what one thinks is being measured. A study with a high degree of validity thus has a high level of credibility. A discussion of these concepts must be transferred to the actual context of the study.

The method discussion shall also contain a paragraph of source criticism. This is where the authors' point of view on the use and selection of sources is described.

In certain contexts it may be the case that the most relevant information for the study is not to be found in scientific literature but rather with individual software developers and open source projects. It must then be clearly stated that efforts have been made to gain access to this information, e.g. by direct communication with developers and/or through discussion forums, etc. Efforts must also be made to indicate the lack of relevant research literature. The precise manner of such investigations must be clearly specified in a method section. The paragraph on source criticism must critically discuss these approaches.

Usually however, there are always relevant related research. If not about the actual research questions, there is certainly important information about the domain under study.

5.3 The work in a wider context

There must be a section discussing ethical and societal aspects related to the work. This is important for the authors to demonstrate a professional maturity and also for achieving the education goals. If the work, for some reason, completely lacks a connection to ethical or societal aspects this must be explicitly stated and justified in the section Delimitations in the introduction chapter.

In the discussion chapter, one must explicitly refer to sources relevant to the discussion.

6 Conclusion

This chapter contains a summarization of the purpose and the research questions. To what extent has the aim been achieved, and what are the answers to the research questions?

The consequences for the target audience (and possibly for researchers and practitioners) must also be described. There should be a section on future work where ideas for continued work are described. If the conclusion chapter contains such a section, the ideas described therein must be concrete and well thought through.