Dissertation project plan

Dominic White

March 14, 2020

Description

Although considerable work has been done for real-time train delay prediction, the time-frame considered by this project (medium-term; up to 5 days' in advance) is poorly, if at all, explored. This project will therefore be based on real-time work and suitably adapted. A machine learning system will be developed using weather data, historic train delay data, and various other data for this purpose. An API will be developed to expose the model to forecast data and unseen schedules, with the ultimate objective of a simple application.

Preliminary preparation

- Sourcing suitable datasets
- Merging said datasets

Objectives

Minimum

- Implementation of a machine learning system

Intermediate

- Hosted API to expose model to unseen scheules and weather forecast data

Advanced

- User interface to access predictions via API

Project plan

Although the preliminary preparations are of great importance (as for any machine learning, the data is vital), they cannot in of themselves constitute a deliverable. Considerable time will therefore be dedicated to these preparations. To ensure maximum scope for exploration, as little pre-processing as possible will be performed prior to the start of the minimum objective. Most of the time spent on this project will be on the minimum objective.

The completion of the intermediate objective should be fairly easy – the dataset will be formatted to match both schedule data and forecast data. However, the amalgamation of these various APIs, as well as hosting an API exposing the trained model, will likely be time-consuming. It is likely in the course of this objective the work of the previous objective will have to be modified slightly.

The completion of the advanced objective is likely unrestrained. For the system to be useful, consumers must be able to access delay predictions through a convenient interface. This will likely take the form of either a hybrid application or website, which allows the user to look up train routes (itself a complex problem) and then predicts the likelihood of delays for those routes. For actual value, the system would also have to incorporate a booking system, or delegate this to an API, but this is beyond the scope of the project at this stage.