

SEMINAR TOPIC:

Train Delay Safety Measure Prediction using machine learning

ABSTRACT

Train Delay prediction is a process of estimating delay probability based on known data at a given checkpoint and is typically measured via arrival (departure) delay. The key to making delay predictions based on actual operational data involves establishing the relationship between train delays and various factors of a railway system. In this we have used Train delay dataset to predict Train delays.

The train delay some factors that affected mainly are:

- long-distance transport
- Whether attributes (temperature, Rain Level, wind speed, visibility) etc

We have a dataset of weather, train delay and train schedule records were collected and analysed in order to understand the patterns of train delays and to predict train delay time. We found that in severe weather train delays are determined mainly by the type of bad weather, while in ordinary weather the delays are determined mainly by the historical delay time and delay frequency of trains. Identifying the factors closely correlated with train delays, we developed a machine learning model to predict the delay time of each train at each station. This prediction model is useful not only for passengers wishing to plan their journeys more reliably, but also for railway operators developing more efficient train schedules and more reasonable pricing plans.

The arrival or departure delay of a train is measured as a difference between the predefined scheduled time of the event and its realized time-stamp. The prediction of future train delays can be communicated to passengers. Being informed about delays as early and as accurate as possible increases the service quality of railway operators for passengers significantly, even though delays will clearly not lead to passenger satisfaction. Another is an accurate predictions of train delay development are a crucial decision support information for traffic controllers who try to minimize the propagation of delay in railway networks.

The techniques that used:

Many different mathematical models and approaches have been used in literature for the purpose of train delay prediction. (Robust) linear regression, k-nearest neighbours (k-NN) random trees (RT), random forest (RF), timed event graph (TEG), time series analysis (TSA), Markov chains (MC), artificial neural networks (ANN), support vector machines (SVM) or bayesian networks (BN) are among them. Concerning input data for the prediction method, we want to distinguish between historic train movement (HTM), actual delays (AD), infrastructure indicators (II), timetable properties (TP) and external factors (EF).

Almost all approaches take into account observations of realized historic train movements which include basic information about the rolling stock (train ID, service ID, train category) and the general setting (weekday, time of the day). As soon as approaches try to predict delay development in real-time, actual train delays (AD) of arrival, departure or passing events are an indispensable input in prediction models.

Some of Research Questions are:

- Where do delays occur on the Railway network?
- What types of delays are most common or largest?
- How are delays connected on the network?
- Can we provide historical context to an emerging incident?
- Can we make predictions about emerging incidents?