

# A literature review of machine learning techniques for train delay prediction

D. White, N. Al-Moubayed

November 19, 2019

## Abstract

*Context:* train delay prediction (TDP)

Train delays impose a huge cost on both train operators and passengers. A preliminary analysis of historical delay attribution data released by Network Rail (NR) shows that 35 million minutes of delay were experienced by passengers in the 2018 - 2019 financial year. The prediction of train delays allows the rescheduling or re-routing of crews and rolling-stock, the reduction in the amount of delay, and improvement of information available to passengers, and subsequent better decision-making.

*Objective:* the goal of this work is to synthesise available research results to inform evidence-based selection of machine learning (ML) techniques for TPD.

*Method:* relevant studies about ML techniques were gathered via a systematic literature review. Supporting contextual studies were also gathered.

*Results:*  $x$  studies about data-driven TPDMs were selected. They use: Bayesian networks, support vector regression, random forests, neural networks, fuzzy Petri nets, and extreme learning machines. The data used, and results obtained, are compared.

*Conclusion:* to do.

# Contents

# 1 Introduction

## 1.1 Delays

A *delay* may be defined as "the difference between the minimum, or unopposed, travel time, and the actual travel time or [as] the difference between the scheduled running time and actual running time" [6].

A *delay* is a "positive deviation between the realized time and scheduled times of [an] activity" [3].

They are reduced by the provision of *slack* in railway timetables [3]. Slack is the amount of time a train can be delayed without the delay propagating.

Mention relationship between slack, capacity, and so on and so forth.

Although precise terminology differs, the literature agrees that there are two principal classes of delay [14]: primary and secondary.

A *primary* or *exogenous* delay is "caused by external stochastic disturbances" [15]. The causes of primary delays are varied and numerous [1], [12], [20], [3], and many different classifications exist. For the purposes of this dissertation, the following classification is proposed:

- *Weather*: severe heat, flooding, landslips, leaves, snow, and ice
- *Passenger*: prolonged alighting and boarding times, large flows
- *Maintenance*: construction work, repair work
- *Other*: accidents, vandalism, trespassing, fatalities, strikes

A large component of this dissertation is the effect of the inclusion of exogenous data on the predictive capability of various machine learning techniques: this classification broadly follows the sets of data that will be considered.

The relative importance of these factors is poorly studied. A limited study by [14] found that the punctuality of trains is correlated with the number of passengers, occupancy ratio, departure punctuality, and operational priority rules.

The number of passengers affects the *dwell time*, the time "devoted to the loading and unloading processes of the train" [24]. For this dissertation, this refers to the alighting and boarding of passengers. Dwell time is a key parameter of system performance, service reliability and quality [19]. Passenger volume is considered a key factor influencing both dwell time [24] and punctuality [14].

A *secondary* (*knock-on*, *consecutive*) delay is "generated by operations conflicts" [3], i.e. primary delays. Secondary delays often affect both the route on which the primary delay occurred and any connecting routes; delays 'cascade' as "trains, drivers, and crews aren't in the right place at the right time to run other services" [21], or o trains are held according to waiting policies between trains [1].

Although further classifications of secondary delay exist [5], the current level of detail will suffice for this dissertation.

Secondary delays cannot be exactly forecast [1], [12] because they are influenced by multiple interacting factors: the severity of the primary delay, the timetable of the train, the infrastructure, and even the behaviour of the driver, who may

drive faster than planned, or reduce dwell time at stations, in order to make up time.

The goal of this work is to synthesise available research results to inform evidence-based selection of ML techniques.

A systematic literature review (SLR) is "a means of evaluating and interpreting all available research relevant to a particular research question or topic area or phenomenon of interest" [28]. The specific objectives of this SLR are:

- To identify categories of ML techniques
- To summarise current research solutions for TDP
- To synthesis the current results from ML techniques for TDP
- To identify the research challenges and needs in the area of TDP

## 2 Overview of systematic literature review

### 2.1 Search strategy and terms

We identified key search terms from a preliminary exploration of relevant studies. Our main search term is "train delay prediction". Unfortunately, the verb "train" is a common word in scientific literature. Where possible, "train NOT training" was specified. Appropriate synonyms were also used for the other search terms: "delay" OR "delays"; "prediction" OR "predicting" OR "analysis" OR "analyzing" OR "analysing" OR "estimation" OR "estimating". The breadth of possible machine learning techniques eliminated the possibility of searching by technique.

### 2.2 Databases

The following databases were searched:

- ACM Digital Library
- IEEE Xplore
- ScienceDirect
- SpringerLink

### 2.3 Study selection

Okay. Now I'll get all of my papers. Let's just get it DONE.

ScienceDirect: 701 results Title, abstract, keywords = (train NOT training) AND (delay OR delays) AND (predicting OR prediction OR estimating OR estimating OR analysis OR analysing OR analyzing)

Reduced to 156 by using -brain -neuroscience -neurobiology -behavioural -Lancet -Ultrasonics -Neurocomputing -medicine -Resuscitation as a filter.

Of those 42 were suitable by title. But we want by abstract...

SpringerLink: 125 results Search string = (delay OR delays OR prediction OR predicting OR estimating OR estimation OR analysis OR analysing) Discipline: Computer Science

If I say it isn't systematic, then we're in the clear. And I save myself a lot of time. It's unlikely, I think, that I've missed anything. Yeah: let's scrap out the systematic requirement. Search term: (delay OR delays OR prediction OR predicting OR estimating OR estimation OR analysis OR analysing)

Selection of studies for inclusion in the SLR is a three stage process:

1. Initial selection of studies based on title
2. Selection of studies based on abstract
3. Further selection of studies based on reading the paper

### 3 Timescales

There are several different timescales at which delays can be predicted, such as short-term (predicted using real-time operating data) and long-time (3 days to a week in advance), as in [26]; tactical (in which models are applied to both timetabling and resource planning), and operational (in which models are used for the real-time prediction of train delays), as in [11]. This dissertation is concerned with the short-term / operational level, henceforth referred to as the *real-time* level.

Models for real-time traffic have so far focused on overcoming the combinatorial complexity of train rescheduling, rolling stock and crew scheduling, and delay management [9]. Real-time train delay prediction (RTTDP) models are *online*, i.e. updated as data on train movements becomes periodically available. Many different models have been proposed; they will be discussed later.

### 4 Metrics

*Punctuality* is "a feature consisting in that a predefined vehicle arrives, departs, or passes at a predefined point at a predefined time" [23]. This definition has the interesting effect that trains that arrive *early* cannot be considered punctual. However, the use of punctuality hides a lot of information [25]; reliability and variability are better metrics [14].

*reliability* has several measures [22]:

- The probability that a train arrives  $x$  minutes late (punctuality)
- The probability of an early departure
- The mean difference between the expected arrival time and the scheduled arrival time
- The mean delay of an arrival given that one arrives late
- The mean delay of an arrival given that one arrives more than  $x$  minutes late
- The standard deviation of arrival times

*variability* is a "measurement of the uncertainty of trip journey times in transportation" [14]. It relates to the distribution of arrival times for a train [13]: a train that arrives the same amount of minutes behind schedule every day has low variability, but not would be considered punctual.

The Office of Road and Rail (ORR), the UK's railway regulator, uses Public Performance Measure (PPM) to assess punctuality, and Cancellations and Significant Lateness (CaSL) to assess reliability.

## 5 Existing models

### 5.1 Analytical

An *analytical model* is "primarily quantitative or computational in nature and represents the system in terms of a set of mathematical equations that specific parametric relationships and their associated parameter values as a function of time, space, and/or other system parameters" [8]. Current state-of-the-art TDPS use analytical models [15].

Simplistic early models, such as [7] made overly restrictive assumptions about railway operations by, for example, forbidding overtakes, assuming that departure times are uniformly distributed, and that the speed of each train is unique and constant.

Subsequent work in this area has largely relaxed these assumptions, by including factors such as overtakes, different speeds, priority systems, and uncertainties associated with train departure time [17], [4]. More complex models have also emerged, incorporating stochastic approximation [2] and the impact of dispatching strategies on train delays and passenger waiting time [16]. Although the state-of-the-art advances constantly, a good example of an recent *in-use* system is [1], which is currently used in the German rail network.

## 6 Overview of studies

We identified 13 studies in the literature that focus on ML techniques for TPD. An preliminary analysis shows that all work occurred in the past decade (i.e. during, or after, 2010), with most studies (46%) published in 2015 and 2019.

| Year  | #  |
|-------|----|
| 2010  | 1  |
| 2013  | 2  |
| 2014  | 1  |
| 2015  | 3  |
| 2016  | 2  |
| 2018  | 1  |
| 2019  | 3  |
| Total | 13 |

11 distinct ML techniques were identified. Several papers compared and contrasted different techniques, or variations of the same technique (as in [10][15][12][11]); in this case, each distinct variation is counted separately. The most popular techniques are Bayesian networks (24%) and artificial neural networks (18%). Some studies use techniques that may be more accurately classified as 'statistics' rather than ML (i.e. simple variants of regression, as in [18][27]); however, as the line is blurred, and for completeness' sake, all are included here.

| ML technique                             | Acronym | #  |
|--|---------|----|
| Artificial neural fuzzy inference system | ANFIS   | 1  |
| Artificial neural network                | ANN     | 3  |
| Bayesian network                         | BN      | 4  |
| Extreme learning machine                 | ELM     | 1  |
| Fuzzi Petri net                          | FPN     | 1  |
| Genetic algorithm                        | GA      | 1  |
| Gradient-boosted regression trees        | GBRT    | 1  |
| Kernel method                            | KM      | 1  |
| $k$ -nearest neighbour                   | $k$ -NN | 1  |
| Random forests                           | RF      | 2  |
| Support vector regression                | SVG     | 1  |
| Total                                    |         | 17 |

I don't think that we should give a damn about where the papers come from. And I certainly don't want to have to calculate it. I'm not going to assess the quality of each paper. I want to expand more on data model inputs, however. Should I include results also? I don't think that we can generalise that easily

Objectives, method limitations, evaluation methodology, evaluation subjects (dataset), evaluation results.

## References

- [1] Annabell Berger et al. “Stochastic Delay Prediction in Large Train Networks”. In: *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*. Ed. by Alberto Caprara and Spyros Kontogiannis. Vol. 20. OpenAccess Series in Informatics (OASICS). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2011, pp. 100–111. ISBN: 978-3-939897-33-0. DOI: 10.4230/OASICS.ATMOS.2011.100. URL: <http://drops.dagstuhl.de/opus/volltexte/2011/3270>.
- [2] Malachy Carey and Andrzej Kwieciński. “Stochastic approximation to the effects of headways on knock-on delays of trains”. In: *Transportation Research Part B: Methodological* 28.4 (1994), pp. 251–267. DOI: 10.1016/0191-2615(94)90001-9.
- [3] Fabrizio Cerreto et al. “Causal Analysis of Railway Running Delays”. In: *11th World Congress on Railway Research (WCRR 2016)*. 2016.
- [4] Bintong Chen and Patrick T. Harker. “Two Moments Estimation of the Delay on Single-Track Rail Lines with Scheduled Traffic”. In: *Transportation Science* 24.4 (1990), pp. 261–275. DOI: 10.1287/trsc.24.4.261.
- [5] Winnie Daamen, Rob M. P. Goverde, and Ingo A. Hansen. “Non-Discriminatory Automatic Registration of Knock-On Train Delays”. In: *Networks and Spatial Economics* 9.1 (2008), pp. 47–61. DOI: 10.1007/s11067-008-9087-2.
- [6] Mark Dingler et al. “Determining the causes of train delay”. In: *AREMA Annual Conference Proceedings*. 2010.
- [7] Ove Frank. “Two-Way Traffic on a Single Line of Railway”. In: *Operations Research* 14.5 (1966), pp. 801–811. DOI: 10.1287/opre.14.5.801.
- [8] Sanford Friedenthal, Alan Moore, and Rick Steiner. *A Practical Guide to SysML*. 2nd ed. Morgan Kaufmann Publishers, 2012.
- [9] Pavle Kecman, Francesco Corman, and Lingyun Meng. “Train delay evolution as a stochastic process”. In: *Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis: RailTokyo2015*. 2015. URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-119746>.
- [10] Javad Lessan, Liping Fu, and Chao Wen. “A hybrid Bayesian network model for predicting delays in train operations”. In: *Computers & Industrial Engineering* 127 (2019), pp. 1214–1222. DOI: 10.1016/j.cie.2018.03.017.
- [11] Nikola Marković et al. “Analyzing passenger train arrival delays with support vector regression”. In: *Transportation Research Part C: Emerging Technologies* 56 (2015), pp. 251–262. DOI: 10.1016/j.trc.2015.04.004.
- [12] Sanjin Milinković et al. “A fuzzy Petri net model to estimate train delays”. In: *Simulation Modelling Practice and Theory* 33 (2013), pp. 144–157. DOI: 10.1016/j.simpat.2012.12.005.
- [13] Robert B. Noland and John W. Polak. “Travel time variability: A review of theoretical and empirical issues”. In: *Transport Reviews* 22.1 (2002), pp. 39–54. DOI: 10.1080/01441640010022456.
- [14] Nils O.E. Olsson and Hans Haugland. “Influencing factors on train punctuality—results from some Norwegian studies”. In: *Transport Policy* 11.4 (2004), pp. 387–397. DOI: 10.1016/j.tranpol.2004.07.001.
- [15] Luca Oneto et al. “Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data”. In: *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (2016). DOI: 10.1109/dsaa.2016.57.
- [16] Süleyman Özekici and Selim Şengör. “On a Rail Transportation Model with Scheduled Services”. In: *Transportation Science* 28.3 (1994), pp. 246–255. DOI: 10.1287/trsc.28.3.246.
- [17] E. R. Petersen. “Over-the-Road Transit Time for a Single Track Railway”. In: *Transportation Science* 8.1 (1974), pp. 65–74. DOI: 10.1287/trsc.8.1.65.
- [18] Suporn Pongnumkul et al. “Improving arrival time prediction of Thailand’s passenger trains using historical travel times”. In: *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (2014). DOI: 10.1109/jcsse.2014.6841886.
- [19] Andre Puong. “Dwell Time Model and Analysis for the MBTA Red Line”. In: (Mar. 2000).
- [20] Network Rail. *Delays explained*. 2019. URL: <https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/> (visited on 11/19/2019).
- [21] Network Rail. *Knock-on delays*. 2019. URL: <https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/knock-on-delays/> (visited on 11/19/2019).
- [22] P Rietveld, F.R Bruinsma, and D.J van Vuuren. “Coping with unreliability in public transport chains: A case study for Netherlands”. In: *Transportation Research Part A: Policy and Practice* 35.6 (2001), pp. 539–559. DOI: 10.1016/S0965-8564(00)00006-9.
- [23] Andrzej Rudnicki. “Measures of Regularity and Punctuality in Public Transport Operation”. In: *IFAC Proceedings Volumes* 30.8 (1997), pp. 661–666. DOI: 10.1016/S1474-6670(17)43896-1.

- [24] Hor Peay San and Mohd Idrus Mohd Masirin. “Train Dwell Time Models for Rail Passenger Service”. In: *MATEC Web of Conferences* 47 (2016), p. 03005. DOI: 10.1051/mateconf/20164703005.
- [25] R Skagestad. “Kritiske prestasjonsindikatorer i jernbanedrift”. PhD thesis. Norwegian University of Science and Technology, 2004.
- [26] Pu Wang and Qing-peng Zhang. “Train delay analysis and prediction based on big data fusion”. In: *Transportation Safety and Environment* 1.1 (2019), pp. 79–88. DOI: 10.1093/tse/tdy001.
- [27] Ren Wang and Daniel B. Work. “Data Driven Approaches for Passenger Train Delay Estimation”. In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (2015). DOI: 10.1109/itsc.2015.94.
- [28] C.C. Williams and J.K. Hollingsworth. “Automatic mining of source code repositories to improve bug finding techniques”. In: *IEEE Transactions on Software Engineering* 31.6 (2005), pp. 466–480. DOI: 10.1109/tse.2005.63.