Predictive Model of Train Delays in a Railway System

Weiwei Mou^a, Zhaolan Cheng^a, Chao Wen^{a,b,1}

 ^a National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu Sichuan 610031, China;
 ^b National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu Sichuan 610031

¹ E-mail: wenchao@swjtu.cn, Phone: +86 13882255078

Abstract

Delay prediction is an important issue associated with train timetabling and dispatching. Based on real-world operation records, accurate estimation of delays is of immense significance in train operation and decisions of dispatchers. In the study, we establish a model that illustrates the interaction between and the factors affecting the same via train operation records from a Dutch railway system. Based on the main factors that affect train delay and the time series trend, we identify the independent and dependent variables. A long short-term memory (LSTM) prediction model in which the actual delay time corresponded to the dependent variable is established via Python3. 6. Finally, the prediction accuracy of the random forest model and artificial neural network model is compared. The results indicate that the LSTM model outperforms other models.

Keywords

Railway, Real-world data, Delay prediction, LSTM model

1 Introduction

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 prediction based on actual operational data involves establishing the relationship between train delays and various characteristics of a railway system. This provides a basis for the operator's scheduling decision.

From a strategic and tactical viewpoint, the accurate prediction of train delays is of immense significance. At a strategic level, accurate train delay prediction is conducive to analyzing capacity of railway and effectiveness of its route planning. It is well known that operators tend to reduce train delays by investing in infrastructure. Accurate delay

prediction can also detect habitual delays in railway routes and potential conflicts in train operation in a timely manner. This enables operators to improve infrastructure for specific routes, and thereby promotes the overall transport efficiency of the railway system. With respect to tactical level, accurate delay prediction is tremendously significant in the establishment of a flexible and stable train diagram and aids in improving the stability of train operation plan. Timetables are tested for robustness via probability distributions of process durations that are derived from historical traffic realization data. Conclusions from the tests are subsequently used to improve timetable robustness (Medeossi et al. (2011))

2 Literature Review

Machine learning methods have been widely used in train delay prediction, which are roughly divided into two categories, namely traditional statistical machine methods (including correlation analysis, linear regression, Markov chain, Bayesian network, and random forest) and neural network machine learning (mainly including support vector, neural network, and deep learning).

Traditional statistical machine learning methods consider train operation performance as model-driven data to update algorithm structure and parameters in time such as delay probability updating in Bayesian network and pruning of a decision tree. Berger A. (2011) proposed a stochastic model of delay propagation to predict train arrival and departure delay events. The model is suitable for all public transportation systems and requires online prediction. The actual delay data of the train should be updated in real time. Based on the train operation data of the Netherlands railway network, extant studies established several models via traditional statistical machine learning methods including a train delay prediction model based on network graph (Huisman et al. (2002), Yuan and Hansen (2007)) and a train stop time and train operation performance prediction model based on distribution statistics(Meer et al. (2009),Goverde et al. (2013)). The results obtained by Olsson and Haugland (2004) indicate that passenger management is an important factor that affects train punctuality in congested areas while the management of train crossings is the key factor that affects train punctuality in non-congested areas. Flier H et al. (2009) combined linear regression and combination model to predict delay based on the on-line train delay monitoring data of the Swiss railway network. The model tested the regional corridor of Lucerne and achieved good prediction results without considering station capacity constraints. Gorman (2009) used statistical methods to forecast the average monthly train running time, and the average absolute percentage error corresponded to 4.6%. The train running process is typically considered as a Markov process. Train running delay is predicted (Barta et al. (2012), Şahin (2017),Kecman et al. (2015)) based on the deduction of train running state. The delay probability updating mechanism of the Bayesian network simulates the process of dispatcher updating the delay probability based on experience and train operation data. It is also used to establish a delay prediction model(Lessan et al. (2018), Francesco and Pavle (2018), Kecman and Goverde (2015b)) that utilizes robust linear regression, regression tree, and random forest models to predict the train running time and dwell time. Furthermore, robust linear regression was improved, and a local model was proposed for local routes and sections. The results indicated that the local model exhibited higher prediction accuracy.

It is not necessary for neural network machine learning methods to be based on prior scheduling knowledge. They realize train delay prediction by learning useful features from massive data. Marković et al. (2015) determines the effect of the infrastructure on train delays by experts and then uses the support vector machine model to predict the arrival time of a train at a station. When compared with the ordinary artificial neural network model, this indicates that the support vector machine model exhibited better prediction effect. Based on the actual data of Wuhan-Guangzhou high-speed railway, Chen et al. (2015) proposed three models, namely least squares method, support vector machine and least squares support vector machine models, to determine train location and predict train delay. Specifically, ANN was used to establish the delay prediction model, and a data-driven model was constructed based on the train operation data in Iran and Germany. The model validation results indicated that the prediction accuracy of the model is high (Yaghini et al. (2013), Peters et al. (2005)).

Most recently, a shallow and deep extreme learning machine (DELM) was proposed in conjunction with the rapid development of big data technologies. Oneto et al. (Oneto et al. (2017b), Oneto et al. (2016)) presented a data-driven TDPS for a large-scale railway network to provide useful information on RTC processes by using state-of-the-art tools and techniques. The system extracted information from a large amount of historical train movement data using big data technologies, learning algorithms, and statistical tools. The described approach and prediction system were validated based on real historical data in six months. The results revealed that the DELM outperformed the current technique, and this was mainly based on the event graph proposed by Kecman and Goverde (2015a). Oneto et al. (2017a) developed a data-driven dynamic train delay prediction system based on the findings of Oneto et al. (2017b). This integrated heterogeneous data sources to deal with varying dynamic systems via DELM. The system exploited state-of-the-art tools and techniques, was completely data-driven, and did not require any prior information on the

railway network.

When compared with the traditional statistical machine learning model, deep learning uses deep neural network models for learning. The steps it learns corresponds to signal-feature-value. The first step involves not determining via learning the structure of the input data and not via random initialization. Therefore, the initial value is closer to the global optimum, and the model achieves better results. Overall, it corresponds to a layer-wise training mechanism. If the traditional neural network reaches more than seven layers, then the residual propagation to the foremost layer is extremely low, and gradient diffusion occurs, and this affects the accuracy of the model. When compared to traditional neural networks, deep learning reduces the number of neural network parameters and adds new structures (for e.g., LSTM and ResNet), a new activation function (ReLU), new weight initialization methods (for e.g., layer-by-layer initialization and XAVIEER), new loss functions, and new over-fitting methods (for e.g., Dropout and BN). It is characterized by a deep neural network selection that overcomes artificial choices. Currently, the prediction model of the train arrival delay is not refined. The research means and prediction accuracy are limited. Generally, from the time series perspective, it is common to consider multi-attributes to obtain the delay prediction. However, a few studies focus on the application of deep learning technology to predict train delays. In the study, the LSTM neural network model in deep learning is applied to prediction of train delays, and this is mainly because the propagation mechanism of train delays is complex and exhibits a non-linear relationship in time and space. The LSTM neural network exhibits a complex structure, and this can be used for non-linear fitting of data related to train delays to realize coding and decoding of time series data. The essential relationship between train delays and impact factors is better revealed via deep learning of large data samples and self-selection of features, and this improves the prediction accuracy of train delays.

Based on the actual running data of the Dutch railway Rotterdam Central to Dordrecht section, the study uses the LSTM model to predict the train arrival delay, and this lays a theoretical foundation for a dispatcher's decisions. The main structure of the study is as follows: Section 3 mainly describes the data of train delays. Section 4 introduces LSTM model for arrival delay prediction. Section 5 presents model forecast accuracy analysis and model evaluation. Section 6 discusses the main conclusions and applications.

3 Data Description

The actual data of train operation in the study ranges from Rotterdam Central to Dordrecht

section of the Dutch railway system, and this contains seven stations, namely Rotterdam Central (Rtd), Rotterdam Blaak (Rtb), Rotterdam Lombardijen (Rlb), Barendrecht (Brd), Zwijndrecht (Zwd), and Dordrecht (Ddr). The data includes delays of all trains in seven stations and six sections. The time span corresponds to 66 working days ranging from September 4, 2017 to December 8, 2017. The data records include the date, train number, train characteristic, location, train activity, planned time, realization, delay jump, and delay cause. A few examples of the data are shown in Table 1.

	Table 1.1 art of the original data table										
Traffic		Trainn	Train	Location	Activity	Planned	Dealication				
	Date	umber	Characteristic	Location	Activity	Time	Keansation				
	2017/9/4	5195	SPR	Zwd	K_A	1:14:18	1:14:43				
	2017/9/21	2274	IC	Brd	K_A	22:59:00	22:59:10				
	2017/9/29	5131	SPR	Ddr	А	9:20:00	9:20:21				
	2017/11/13	5025	SPR	Rtd	А	7:39:00	7:39:15				

Table 1:Part of the original data table

4 Train Arrival Delay Prediction Model

4.1 Selection of characteristic variables

Delay prediction is a process of estimating the probability of train delays at subsequent recording points based on train operation history data, and this is typically determined by arrival delays. It is assumed that a train is currently located at station s_n , the former station and the subsequent station to arrive are donated by s_{n-1} and s_{n+1} respectively. s_{n+1} . Based on the train delays at s_n and s_{n-1} stations and scheduled running time of trains at sections (s_{n-1}, s_n) , (s_n, s_{n+1}) , the study predicts the arrival delays of trains at s_{n+1} stations. As shown in Fig. 1, the train arrives at the station s_n at time t_n^A on schedule and starts at the same station at time t_n^D . However, in the actual operation process, given various interference factors, the train can deviate from the timetable to generate the actual arrival time \hat{t}_n^A and actual departure time \hat{t}_n^D . Figure 1 shows successive stations (s_{n-1}, s_n) , and s_{n+1}) with the parameters in parentheses indicating the scheduled time and actual time of the event. The train delay can be typically divided into arrival delay and departure delay. The difference between the actual and scheduled times $(\hat{t}_n^A - t_n^A)$ and $(\hat{t}_n^D - t_n^D)$ indicate the arrival and departure delays, respectively, of the train at station s_n .

The train can be delayed due to various disturbances in the operation process. Six

parameters are selected after the analysis of the train arrival delays at the station to constitute the feature space (F). The study assumes that the parameters affect the future delay of the train, and thus the future arrival of the train is predicted based on the selected parameters.

The feature variables included in the feature space (F) are as follows:

1. Train Characteristic(X_1)

There are three main characteristics of trains running in Rotterdam Central to Dordrecht section of the Netherlands railway system, namely regional train stopping at station (SPR), intercity train stopping at large station (IC), and empty train (LM).

2. Departure delay time of the train at the current station(X₂)

The actual departure delay time of the train at the current station s_n indicates the difference between the actual departure time of the train at station s_n and the planned departure time. The equation corresponds to $\hat{t}_n^D - t_n^D$, which is accurate to seconds.

3. Arrival delay time of the train at the current station(X_3)

The actual arrival delay time of the train at the current station s_n indicates the difference between the actual arrival time of the train at s_n station and planned arrival time. The equation corresponds to $\hat{t}_n^A - t_n^A$, which is accurate to seconds.

4. Departure delay time of the train at the last station(X_4)

The actual departure delay time of the train at the last station indicates the difference between the actual departure time of the train at s_{n-1} station and planned departure time. The equation corresponds to $\hat{t}_{n-1}^D - t_{n-1}^D$, which is accurate to seconds.

5. Planned running time of the train in the last section(X_5)

The calculation equation for the planned running time between the last station s_{n-1} and current station s_n corresponds to $t_n^A - t_{n-1}^D$, which is accurate to seconds.

6. Planned running time of the train in the next section (X_6)

The calculation equation for the planned running time between the current station s_n and next station s_{n+1} corresponds to $t_{n+1}^A - t_n^D$, which is accurate to seconds.

The output variable of the delay prediction in the study denotes the arrival delay time (Y) of the train at the next station. The delay prediction data based on the aforementioned characteristic variables are shown in Table 2. The expression is detailed as follows:



Figure 1: General scheme of train movements at three successive stations

$$Y = \varphi(X_1, X_2, X_3, X_4, X_5, X_6).$$
(1)

Where Y denotes the train arrival delay (output variable), X_1, X_2, X_3, X_4, X_5 , and X_6 denote the train delay influence factors (input variables), and φ denotes the machine learning algorithm model.

Date	Train number	The Last Station	The Current Station	The Next Station	X1	X2	X3	X4	X5	X6	Y
2017/ 9/4	5139	Brd	Zwd	Ddr	SPR	113	17	-7	180	300	140
2017/ 10/2	2216	Sdm	Rtd	Rtb	IC	106	124	138	132	300	124
2017/ 10/26	2214	Rtd	Rtb	Rtz	IC	819	809	781	120	132	812
2017/ 11/7	2218	Rlb	Brd	Zwd	IC	215	215	190	60	240	179
2017/ 12/8	5027	Dtz	Sdm	Rtd	SPR	128	94	116	378	300	80

Table 2: Modeling data table

4.2 LSTM model

The LSTM model was proposed by Hochreiter et al. to improve the model based on RNN. In a conventional RNN, the hidden layer generally corresponds to an extremely simple node such as Tanhwhile the LSTM improves the simple node of the hidden layer into a storage unit. The basic structure of the storage unit is shown in Figure 2. The storage unit is composed of an input gate i, an output gate o, a forgetting gate f, and a memory cell c. In forward propagation, the input gate determines when to activate the incoming storage unit while the output gate determines when to activate the outgoing storage unit. In reverse propagation, the output gate determines when to allow errors to flow into the storage unit, and the input gate determines when to let it flow out of the storage unit. The input gate, output gate, and forgetting gate constitute keys to control information flow. The operation principle of the storage unit is expressed in terms of equations (2)–(6) (Bengio et al. (2002),Greff et al. (2016),Gers et al. (2002)) as follows:

$$i_{t} = \delta(\mathbf{W}_{i}x_{t} + \mathbf{U}_{i}h_{t-1} + \mathbf{V}_{i}C_{t-1} + b_{i}).$$
(2)

$$f_t = \delta \Big(W_f x_t + U_f h_{t-1} + V_f C_{t-1} + b_f \Big).$$
(3)

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tan \, h(\mathbf{W}_c x_t + \mathbf{U}_c h_{t-1} + b_c). \tag{4}$$

$$o_t = \delta(W_o x_t + U_o h_{t-1} + V_o C_t + b_o).$$
(5)

$$h_t = o_t \cdot tan \, h(c_t). \tag{6}$$

Where ct denotes the calculation method of memory cells at time t; ht denotes all outputs of LSTM units at time t; W, U, V, and b denote the matrix of coefficients and vector of offset; δ denotes the activation function sigmoid; \cdot denotes a point multiplication operation; and it, ft, and ot denote the calculation methods of the input gate, forgetting gate, and output gate at time t, respectively. As shown in Figure 2, the outputs of the three gates of the input gate, forgetting gate, and output gate at connected to a multiplier element to control the input and output of information flow and the status of cell units respectively.



Figure 2: Basic structure of the LSTM storage unit

In the actual operation of trains, given the mutual restriction between trains, the delay of the forward train can affect the backward train and result in the lateral propagation of the delay. The LSTM model assumes time series format data as input, and its results at any *t*-time are based on the results at the previous time and input data at the current time. This mechanism enables the preservation and reuse of time series information in the model for a long period such that it learns the knowledge of time series correlation in time series data.

The LSTM model for delay prediction is constructed as follows:

(1) Seven stations in Rotterdam Central to Dordrecht are selected to extract the arrival delay time Y and corresponding feature space (F). All train delays and their extraction attributes are sorted based on the actual train operation sequence, and training data sets and test data sets are divided. As shown in Fig. 3, the first row in the figure indicates the

train arrival delay time (Y), and the second row indicates the characteristic space (F) of the influence factors of the delay time. Specifically, *i* denotes the train number; s_n denotes the station number; and the sliding window length *l* denotes the number of trains that are predicted to be entered each time. Hence, the effect of the previous *l* trains is considered on the current train delay.

(2) Determination of parameter l: The delay time and influencing factors of each train are treated as time series. The model considers the interaction relationship between different train numbers by inputting multiple trains each time. After repeated verification, when l=1 the best predictions can be obtained. Thus, only the effect of the previous train delay on the arrival of the current train is considered. This is mainly due to the long arrival time interval between different trains in Rotterdam Central to Dordrecht section of the Netherlands and tweak interaction between trains.

(3) After determining the optimal number of input trains, the model structure and parameters (for e.g., hidden layer number, neuron number, learning rate, optimizer, and dropout rate) are optimized to obtain the optimal parameters and structure of the model and predict the arrival delay of the train at the station. Finally, the LSTM model with time series input form is shown in Fig. 4. The arrival delay time($Y_i^{s_n}$) of the current train is predicted based on the feature space ($F_i^{s_n}$) of the current train and the effect of only the previous train ($F_{i-1}^{s_n}$). The aforementioned step is repeated to finally realize the prediction for all stations from Rotterdam Central to Dordrecht section.



Figure 3: LSTM input data format



Figure 4: LSTM prediction model

5 Precision and Evaluation of the Model Prediction

5.1 Model prediction accuracy analysis

In order to evaluate the prediction effect of the model, the following analysis is initially performed. As shown in Fig. 5, the actual and predicted arrival delays of trains at stations are compared. Second, as shown in Figure 6, the scatter plots of the observed and predicted arrival delays of trains are illustrated. The results indicate that the predicted values of train arrival delays exhibit a good match with the observed values. Specifically, in the interquartile range, the whiskers and right tail closely match in the figures for each station. Furthermore, as shown in Fig. 6, the majority of predictions are close to the depicted diagonal lines for arrival events ,which implies that the predicted value is extremely close to the observed value.

Figure 7 shows the distribution of predicted residuals for train arrival delays at different stations. The figure assumes the train actual arrival delay time as abscissa and the residuals as ordinate for visualization purposes. As shown in the figure, in the seven stations of Rotterdam Central to Dordrecht section of the Netherlands railway system, all stations (with the exception of the Rtd station) exhibit good prediction results. Increases in the prediction error of the Rtd station can be due to the increasingly significant influence of the outliers. Figure 8 shows the prediction accuracy histogram of LSTM model for the seven stations. As shown in the figure, the model accuracy corresponds to 87.6% with an allowable error within 30 s, and thus the model exhibits a good prediction effect.

Prediction vs Observation



Figure 5: Comparison of predicted and observed arrival delay distribution for different stations



Figure 6: Scatter plots of actual vis-á-vis predicted arrival delays.



Figure 7: Distribution of the residuals of train arrival delays at different stations



Figure 8: LSTM model prediction accuracy

5.2 Model evaluation

(1) Benchmark model

In order to better evaluate the prediction effect of the model, two benchmark models are selected and compared with the LSTM model, namely the random forest model and artificial neural network model. They are detailed as follows:

Random forest: The random forest is a joint prediction model that is composed of multiple decision trees (Cutler et al. (2004),Loh (2011)), and this can be used as a fast and effective classification and prediction model. Each decision tree in RF consists of several forks and nodes. Each decision tree is regressed and predicted. Finally, the predictive effect of random forest is determined via the predictive effect of multiple decision trees. The random forest corresponds to an ensemble learning algorithm, which belongs to the Bagging type. The final result is voted or averaged by combining multiple weak classifiers, and thus the overall model results exhibit higher accuracy and generalization performance. Thus, the model obtains good results, and this is mainly due to the "random" and "forest" elements, which make it resistant to overfitting and increase the precision.

Artificial neural networks: An artificial neural network is one of the most commonly used train delay prediction model (Peters et al. (2005), Yaghini et al. (2013) , Malavasi (2001)). It mainly models the relationship between a set of input signals and set of output signals. The model is derived from the reaction of the human brain to stimuli from a sensory input. In a manner similar to how the brain uses a network of interconnected cells

of a neuron to create a large parallel processor, artificial neural networks use artificial neurons or a network of nodes to solve learning problems. There are three main characteristics of artificial neural networks as follows: ① Activation function that converts the net input signal of a neuron into a single output signal for further propagation in the network; ②network topology that describes the number of neurons in the model, number of layers, and the manner in which the layers are connected; and ③training algorithm that specifies the setting of the connection weight to suppress or increase the proportion of neurons in the input signal. This model is suitable for situations involving simple input and output data albeit an extremely complex input-to-output process.

(2) Model evaluation index

With respect to model evaluation, the study mainly selects MAE and RMSE as evaluation indexes. The equation to calculate the index is as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |p_i - y_i|.$$
⁽⁷⁾

$$RMSE = \sqrt{\frac{1}{N} \sum_{l=1}^{N} (p_i - y_i)^2}.$$
(8)

where p_i and y_i denote the predicted and observed delay values for *ith* arrival events, respectively, and *n* denotes the total number of observations. The measures quantify the average deviation of the predictions from the observed values. The model's performance level improves when the measures are closer to zero.

Figure 9 and Figure 10 show a comparison of MAE and RMSE values for LSTM, RF, and ANN models of different stations. As shown in Fig. 9, from the MAE perspective, the prediction effects of the LSTM model and the random forest model do not significantly differ and both are superior to the artificial neural network model. As shown in Fig. 10, from the RMSE perspective, the prediction effect of LSTM significantly exceeds that of random forest and artificial neural network models. In summary, the LSTM model exhibits a good predictive effect.



Figure 9: Comparison of MAE values at different stations



Figure 10: Comparison of RMSE values for different stations

6 Conclusions

The study presents a machine learning model to analyze the relationship between train

arrival delays and various characteristics of a railway system, which is important for planning changes and investments to reduce delays. In the study, the LSTM model is used to construct a prediction model of train arrival delay, and the model is trained and tested based on the historical data of train operation. The results show that the LSTM model exhibits a better predictive effect than random forest and artificial neural network models. The performance of the LSTM model is superior as indicated by the data validation results. Specifically, the LSTM model exhibits better MAE and MSE values, and its prediction accuracy reaches 87% within 30 s.

The LSTM model is a good measure of the lateral and vertical propagation of train delays. This feature ensures that the model exhibits good generality and can be extended to other high-speed railway routes. Additionally, the model exhibits two main advantages as follows: (a) The simplicity of the model makes it more explanatory and efficient. (b) It includes interrelationships between various delay factors and superposition of arrival delays.

The model in the study can be applied to other stations although similar data must be collected. With respect to the expansion direction of the model, the current model does not consider an excessive number of infrastructure factors. With respect to further model expansion, it is possible to consider additional train delay influence factors and extract increasingly accurate feature variables to obtain better prediction results.

Acknowledgment

This work was supported by the National Nature Science Foundation of China [grant number 71871188] and the Science & Technology Department of Sichuan Province [grant number 2018JY0567]. We are grateful for the contributions made by our project partners.

References

Barta, J., Rizzoli, A.E., Salani, M., Gambardella, L.M.,2012.Statistical modelling of delays in a rail freight transportation network. *Wint Simul C Proc.* 2012.

Bengio, Y., Simard, P., Frasconi, P., 2002.Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 5(2):157-166.

Berger A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M.,2011.Stochastic delay prediction in large train networks. In: OASIcs-OpenAccess Series in Informatics (Vol. 20). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Chen, D., Wang, L.,Li, L.,2015.Position computation models for high-speed train based on support vector machine approach. *Applied Soft Computing* 30(C):758-766.

Cutler, A., Cutler, D.R., Stevens, J.R., 2004. Random Forests. *Machine Learning* 45(1):157-176.

Francesco, C., Pavle, K., 2018. Stochastic prediction of train delays in real-time using Bayesian networks. *Transportation Research Part C: Emerging Technologies*.

Flier, H., Graffagnino, T., Nunkesser, M.,2009. Finding Robust Train Paths in Dense Corridors. IAROR RailZurich 2009.

Gers, F.A., Schraudolph, N.N., Schmidhuber, J., 2002. Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research* 3(1):115-143.

Gorman, M.F.,2009.Statistical estimation of railroad congestion delay. *Transportation Research Part E Logistics & Transportation Review* 45(3):446-456.

Goverde, R.M.P., Corman, F.,D ' Ariano, A.,2013.Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. *Journal of Rail Transport Planning & Management* 3(3):78-94.

Greff, K., et al., 2016.LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks & Learning Systems* 28(10):2222-2232.

Huisman, T., Boucherieb, R.J.,Dijka, N.M.V.,2002.A solvable queueing network model for railway networks and its validation and applications for the Netherlands. *European Journal of Operational Research* 142(1):30-51.

Kecman, P., Corman, F., Meng, L., 2015. Train delay evolution as a stochastic process. International Conference on Railway Operations Modelling & Analysis-railtokyo, 2015.

Kecman, P., Goverde, R.M.P., 2015a. Online data-driven adaptive prediction of train event times. *IEEE Transactions on Intelligent Transportation Systems* 16(1):465-474.

Kecman, P.,Goverde, R.M.P.,2015b.Predictive modelling of running and dwell times in railway traffic. *Public Transport* 7(3):1-25.

Lessan, J., Fu, L., Wen, C., 2018. A Hybrid Bayesian Network Model for Predicting Delays in Train Operations. *Computers & Industrial Engineering*.

Loh, W.Y.,2011.Classification and regression trees. *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery* 1(1):14-23.

Malavasi, G.,2001.Simulation of stochastic elements in railway systems using self-learning processes. *European Journal of Operational Research* 131(2):262-272.

Marković, N., Milinković, S., Tikhonov, K.S., Schonfeld, P.,2015. Analyzing passenger train arrival delays with support vector regression. *Transportation Research Part C: Emerging Technologies* 56:251-262.

Medeossi, G., Longo, G., Fabris, S.D., 2011. A method for using stochastic blocking times to improve timetable planning. *Journal of Rail Transport Planning & Management* 1(1):1-13.

Meer Van der, D.J., Goverde, R.M.P., Hansen, I.A., 2009. Prediction of train running times using historical track occupation data. *Delft University of Technology*, 2009..

Meester, L.E., Muns, S., 2007. Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research Part B Methodological* 41(2):218-230.

Olsson, N.O.E., Haugland, H., 2004. Influencing factors on train punctuality—results from some Norwegian studies. *Transport Policy* 11(4):387-397.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D.,2016. Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data. *IEEE International Conference on Data Science & Advanced Analytics*, 2016. *IEEE, Montreal, Canada*, pp. 458-467.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D.,2017a.Dynamic Delay Predictions for Large-Scale Railway Networks: Deep and Shallow Extreme Learning Machines Tuned via Thresholdout. *IEEE Transactions on Systems Man & Cybernetics Systems* PP(99):1-14.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D.,2017b.Train Delay Prediction Systems: A Big Data Analytics Perspective. *Big Data Research*11, 54-64

Peters, J., Emig, B., Jung, M., Schmidt, S.,2005.Prediction of Delays in Public Transportation using Neural Networks. *Computational Intelligence for Modelling, Control and Automation*, 2005 and *International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, 2005, pp. 92-97.

Şahin, İ.,2017.Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances. *Journal of Rail Transport Planning & Management* 7(3).

Yaghini, M., Khoshraftar, M.M., Seyedabadi, M., 2013. Railway passenger train delay prediction via neural network model. *Journal of Advanced Transportation* 47(3):355-368.

Yuan, J., Hansen, I.A., 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transportation Research Part B Methodological* 41(2):202-217.