Real-time Delay Prediction In Railway Transportation

Veerle Hennebel[1] Bart Roets[2,3] Léon Sobrie[3]
Marijn Verschelde[4,3]

[1]University of Leuven [2]Infrabel

[3]On Track Lab, Ghent University [4]IÉSEG School of Management

INFORMS Annual Meeting 2020

m.verschelde@ieseg.fr

www.ontracklab.com

Outline

Introduction

- Trains operate in a dense network leading to a complex environment with high pressure on traffic managers and traffic controllers.
- Accurate information concerning train arrival times is useful for both passengers and railway operators.
- Railways are an important means for freight transport which should be tracked precisely to the benefit of the customer.

Introduction

Existing literature

Predominantly focused on predicting delays for a given railway line (e.g. Wen et al. (2020), Huang et al. (2020)). Exceptions are Oneto et al. (2018) and Yaghini et al. (2013).

Contributions

1. Real-time analysis on the entire railway network.
2. Inclusion of spatial features. The directly connected train stations (neighbors) are together with their importance included via a spatial matrix.
3. We benchmark our advocated approach with both a spatial regression and a rules-based approach for the entire network

Real-time data structure

- The close-to-real-time data structure is developed by INFRABEL, the Belgium railway infrastructure company, and comprises the entire train network which is one of the most dense in the world.
- From this data, 3 feature types are constructed to predict train delays in real-time: temporal, spatial and operational features.
- With this input data, we predict the delay for the 5 upcoming stops for a specific train.
- We focus on September 2020, resulting in a data structure of over 2 million observations of trains passing a signal

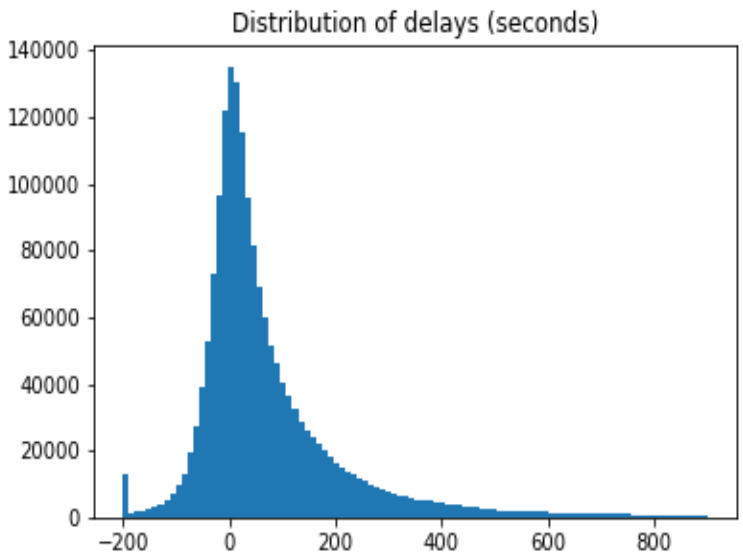
Real-time data structure

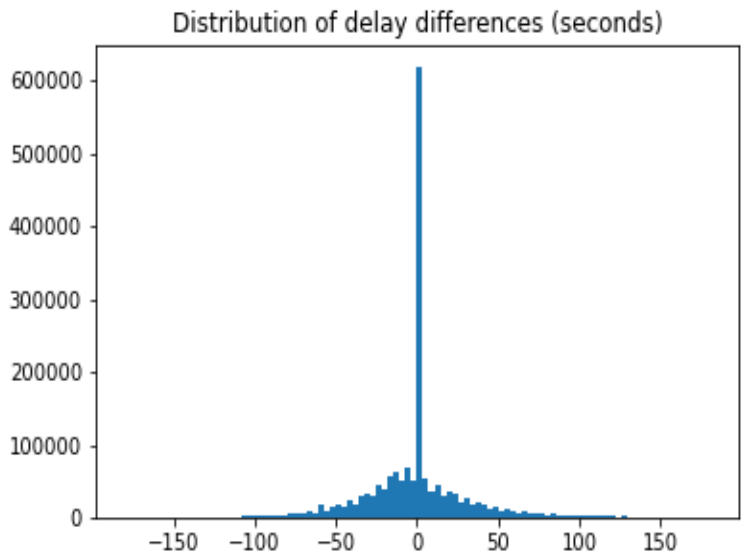
- In total, we have 130 input features that belong to one of the following features categories.

Features		
Temporal	Spatial	Operational
Day Hour Minute	Spatial delay Distance BC/AN GPS coords station Delay influencers Av. delay (diff) station	Current delay (diff) Arr-Dep-Pass Buffer times Route planning char.

Real-time data structure

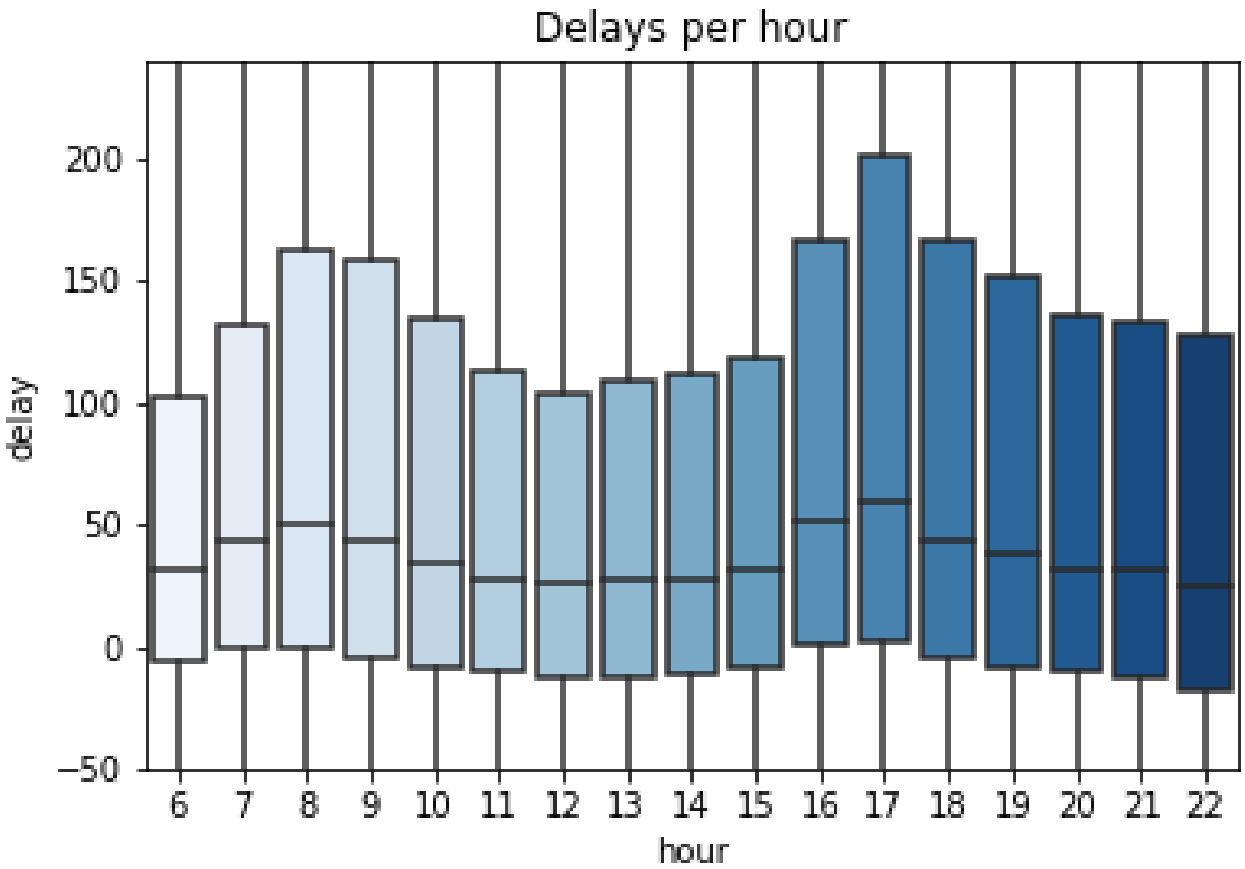
Descriptive analysis: Distribution of delays





Real-time data structure

Descriptive analysis: Temporal variation



Descriptive analysis: Summary statistics

	count	mean	std	min	25%	50%	75%	max
delay	1802033.0	99.15	235.38	-200.0	-6.00	33.00	117.00	1800.0
delaydif	1758674.0	1.06	48.30	-2000.0	-13.00	0.00	7.00	2000.0
delayPTCARpass	1155305.0	102.19	198.66	-200.0	4.00	47.00	130.00	1800.0
delayspatialpass	1802033.0	49.59	124.18	-200.0	0.00	1.00	55.87	1800.0
delayinfluencerpass1	1783298.0	83.27	185.82	-200.0	8.36	32.87	80.20	1800.0
delayinfluencerpass2	1784413.0	105.50	107.21	-200.0	34.77	76.44	145.85	1800.0
delayinfluencerpass3	1125140.0	132.27	245.45	-200.0	11.00	62.00	149.50	1800.0
delayinfluencerpass4	1644804.0	83.16	138.14	-200.0	15.33	42.14	98.33	1708.0
delayinfluencerpass5	1780419.0	66.37	117.11	-200.0	3.50	33.75	87.50	1800.0
delayinfluencerpass6	1748435.0	122.49	158.03	-200.0	25.00	73.25	168.33	1800.0
delayinfluencerpass7	1725345.0	96.73	173.40	-200.0	2.50	46.33	128.00	1800.0
delayinfluencerpass8	1684687.0	114.88	158.17	-200.0	22.33	74.50	156.33	1800.0
delayinfluencerpass9	1743145.0	123.74	174.30	-200.0	32.70	73.75	149.25	1800.0
delayinfluencerpass10	1393256.0	161.31	180.29	-200.0	58.50	122.00	210.50	1800.0
delayinfluencerpass11	1790879.0	95.60	92.22	-200.0	31.50	71.18	135.29	1800.0
delayinfluencerpass12	1788730.0	118.00	101.38	-200.0	47.10	92.78	160.84	1800.0

Descriptive analysis: Correlogram

	Delay	Lagged Delay	Spatial delay	Delay Infl. 12	Peak
Delay	1.00	0.89	0.18	0.08	0.07
Lagged Delay	0.89	1.00	0.19	0.08	0.07
Spatial delay	0.18	0.19	1.00	0.10	0.08
Delay Influencer 12	0.08	0.08	0.10	1.00	0.27
Peak	0.07	0.07	0.08	0.27	1.00

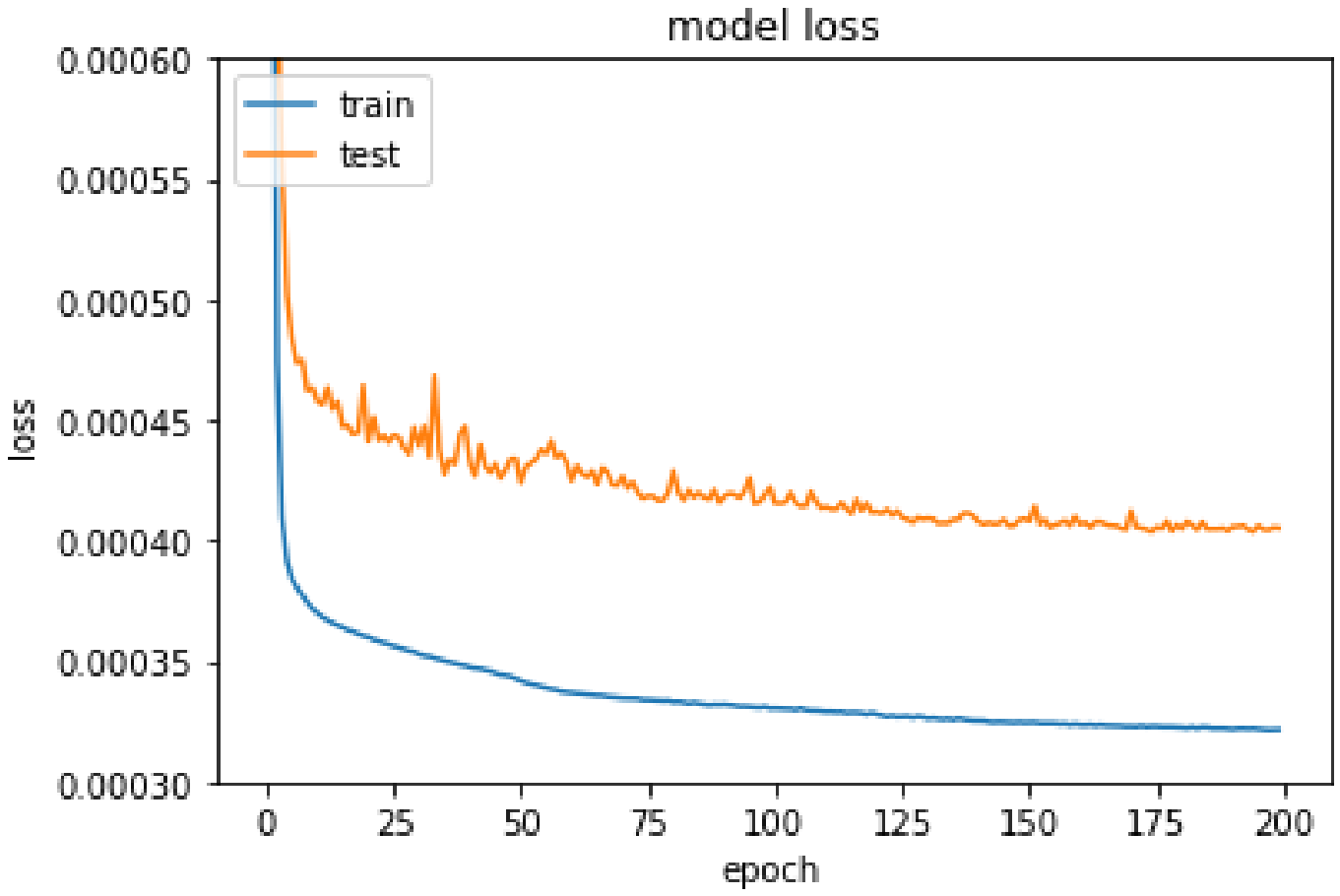
Model set-up

Recurrent Neural Network

- A sequence-to-sequence Recurrent Neural Network with Long Short-Term Memory (Hochreiter and Schmidhuber, 1997).
- We train on 2 weeks, validate on 1 week and test on one day.
- The RNN is benchmarked against the rules-based system and the spatial regression by the use of MAE and RMSE.

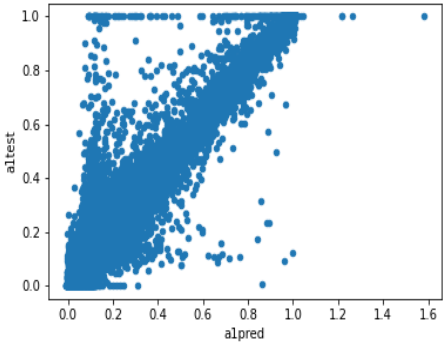
Benchmarking delay predictions

RNN: Training

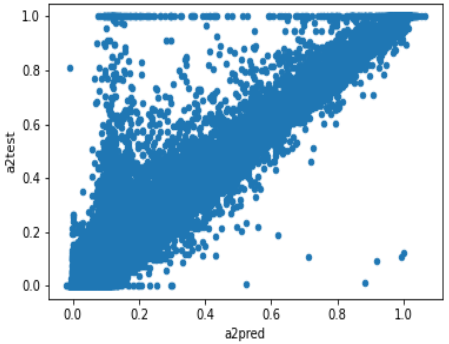


Benchmarking delay predictions

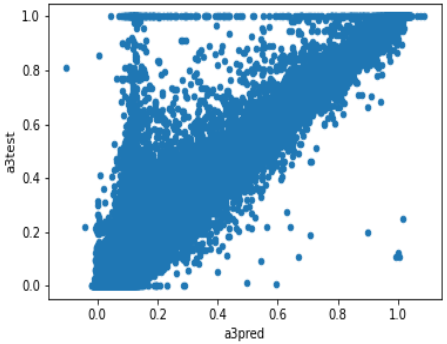
RNN: Predicted vs. actual delay per output sequence



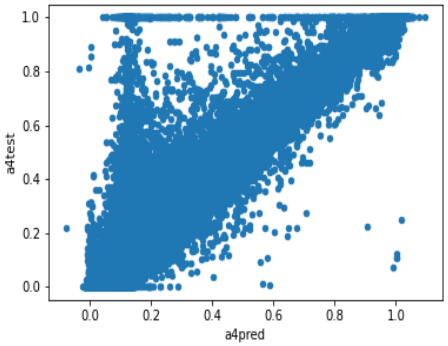
(a) Sequence 1



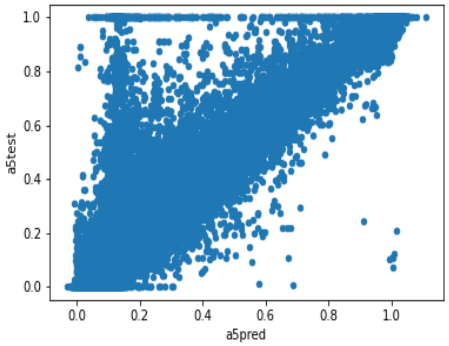
(b) Sequence 2



(c) Sequence 3



(d) Sequence 4



(e) Sequence 5

A first benchmarking exercise

	Spatial reg.	RNN
	All sequences	
RMSE	0.042	0.040
MAE	0.024	0.022
	Sequence 1	
RMSE	0.027	0.025
MAE	0.013	0.013
	Sequence 2	
RMSE	0.035	0.033
MAE	0.019	0.018
	Sequence 1	
RMSE	0.042	0.040
MAE	0.024	0.0230
	Sequence 4	
RMSE	0.049	0.046
MAE	0.029	0.027
	Sequence 5	
RMSE	0.054	0.052
MAE	0.032	0.030

Conclusion

- The Belgian railway transportation network is one of the most dense in the world and is an ideal testing ground for transportation delay prediction in a spatially interdependent environment
- We propose a sequence-to-sequence neural network framework to predict train delays while including spatial features
- Our preliminary benchmarking exercise highlights the usefulness of our approach
- Further research is warranted on the importance of the different feature groups for delay prediction

References I

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Huang, P., Wen, C., Fu, L., Lessan, J., Jiang, C., Peng, Q., and Xu, X. (2020). Modeling train operation as sequences: A study of delay prediction with operation and weather data. *Transportation Research Part E: Logistics and Transportation Review*, 141:102022.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., and Anguita, D. (2018). Train delay prediction systems: a big data analytics perspective. *Big data research*, 11:54–64.

References II

Wen, C., Mou, W., Huang, P., and Li, Z. (2020). A predictive model of train delays on a railway line. *Journal of Forecasting*, 39(3):470–488.

Yaghini, M., Khoshraftar, M. M., and Seyedabadi, M. (2013). Railway passenger train delay prediction via neural network model. *Journal of advanced transportation*, 47(3):355–368.