ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.

In-Station Train Movements Prediction: from Shallow to Deep Multi Scale Models*

Gianluca Boleto[1], Luca Oneto[1], Matteo Cardellini[1], Marco Maratea[1], Mauro Vallati[2], Renzo Canepa[3], Davide Anguita[1]

[1]University of Genoa - Via Opera Pia 11a, 16145, Genova, Italy
[2]University of Huddersfield - Huddersfield, West Yorkshire, HD1 3DH, UK
[3]Rete Ferroviaria Italiana - Via Don Vincenzo Minetti 6/5, 16126, Genova, Italy

Abstract. Public railway transport systems play a crucial role in servicing the global society and are the transport backbone of a sustainable economy. While a significant effort has been devoted to predict inter-station trains movements to support stakeholders (i.e., infrastructure managers, train operators, and travellers) decisions, the problem of predicting in-station movements, while being crucial to improve train dispatching (i.e., empowering human or automatic dispatchers), has been far more less investigated. In fact, stations are the most critical points in a railway network: even small improvements in the estimation of the duration of trains movements can remarkably enhance the dispatching efficiency in coping with the increase in capacity demand and with delays. In this work we will first leverage on state of the art shallow models, fed by domain experts with domain specific features, to improve the current predictive systems. Then, we will leverage on a customised deep multi scale model able to automatically learn the representation and improve the accuracy of the shallow models. Results on real-world data coming from the Italian railway network will support our proposal.

1 Introduction

Rail transport is probably the most sustainable, whether in terms of CO_2 emissions, energy consumption, use of space, or noise levels. In Europe the increasing volume of people and freight transported on railways is congesting the network. The only viable solution to increase capacity, in the short/medium term, is then to improve the efficiency in order to be able to control a larger number of running trains without requiring massive public investments in new physical assets[1]. For this reason, in the last years, all the major players of the European railways grouped together[2] and started extensive modernisation programs that leverage on advanced ICT solutions, Artificial Intelligence (AI)-based especially, to improve system safety and service reliability, to enhance passenger experience, to provide higher transit capacity, and to reduce operational costs.

In this work, we will focus on the problem of analysing train movements. The study of train movements has a long history and, in the last 10 years, it has attracted the interest of both researchers and industry because of the ability of

*This work has been partially funded by Hitachi Rail STS through the RAIDLab (Railway Artificial Intelligence and Data Analysis Laboratory), a joint laboratory between Hitachi Rail STS and University of Genoa.

[1]`https://ec.europa.eu/transport/themes/infrastructure_en`

[2]`https://shift2rail.org/`

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.

the new generation of AI-based systems to positively impact on the efficiency of the railway. In particular, a significant effort has been spent in predicting inter-station trains movements to support railway stakeholders decisions [1, 2]: infrastructure managers and train operators can better plan the use and the exploitation of the railway infrastructure, while travellers can be informed timely of congestion and delays. Much less effort has been spent in studying the in-station movements because of the complexity in retrieving the related fine-grained data. In fact, the vast majority of trains are not equipped with a Global Positioning System (GPS) and usually their position is known in the so called points of measure [1, 3]. To analyse the in-station train movements it is required to retrieve fine-grained data from the interlocking system which is able to register the block section occupations [3]. Unfortunately, these data are seldomly available. Nevertheless, the ability to accurately predict the duration of trains movements is crucial for improving the in-station train dispatching [4–6]. Train dispatching (both human- or AI-based), in fact, requires accurately predicting the duration of trains movements to efficiently cope with the increase in capacity demand and with delays.

In this work we will adopt a two step approach to address the problem of predicting the duration of in-station train movements (Section 2). We will first leverage on state of the art shallow models based on Random Forests (RF) [7], fed by domain experts (i.e., operators of Rete Ferroviaria Italiana - RFI, the Italian infrastructure manager) with domain specific features, to improve the current predictive systems. Then, we will leverage on a custom deep multi scale models based Temporal Convolutional Network (TCN) [8] able to automatically learn a rich and expressive representation directly from the data and to improve the accuracy of the shallow models. Results on real-world data coming from a series of stations in the North West of the Italian railway network, provided by RFI, will support our proposal (Section 3). Section 4 will conclude the paper.

2 Problem and Proposal

The problem that we want to address in this paper is to predict the movements duration of a train on a block section inside a station. Since trains are not equipped with a GPS, the most fine-grained data are the one coming from the interlocking which allow to record the occupation and liberation of the block sections [3]. In fact, in order to effectively and efficiently plan the dispatching of the trains inside a station, it is required to predict the time needed to traverse each block section [5] (Figure 1). In particular, we need to be able to predict these movements duration at least 1 hour [4–6] in

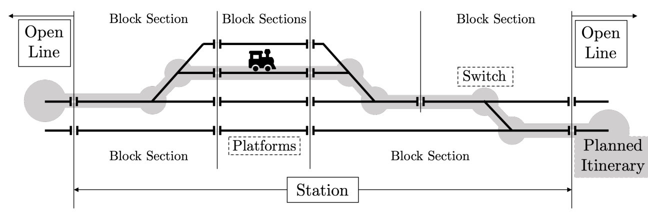


Fig. 1: A simple representation of a small station.

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

advance in order to being able to optimise the dispatching. For this purpose, at time t, to make a prediction at time t+1h, all the time series representing the behaviour of all trains in a sub portion of the network insisting on the station of interest are available from $t-\Delta$ to t. Moreover, additional (side) information is available like the planned itinerary of the trains through the station, if the train stops (and in case the scheduled stop time), its delay, day of the week, working day of holiday, type of train, time slot, weather conditions, etc. Note also that not all trains are circulating through the network every day so not all time series are present every day. This problem can be mapped in the class of the regression problems with structured input data [8]. The problem can be addressed with three approaches. The first one, which is currently used by RFI, is to rely on simple historical statistics able to make reasonable predictions with limited effort (i.e., lookup tables) [1]. The second one (Section 2.1) is to distil these structured input data into simple features (i.e., representation) via features engineering based on the experience of domain experts and then apply shallow models [1]. The last approach (Section 2.2) is to automatically learn a rich and expressive representation directly from the data, via deep architectures [8].

2.1 Shallow Models

Shallow models rely on a simple idea: map the raw input data into a series of features able, from one side, to well represent the input data and, from the other side, to remove redundant or negligible information. Then, based on these features, train a classical (shallow) machine learning model [9] able to learn from data the desired input/output relation. For this purpose, first an handcrafted careful feature engineering phase able to exploit the experience of domain experts has been performed. To facilitate the process we encoded the knowledge of the operators into an ontology (see Figure 2) and then, based on these ontology we generate the induced features set. In fact, thanks to a reasoner, the ontology allows to model and find easily new properties and relations. Then, computing classical statistical indexes (i.e., mean, median, variance, minimum, maximum, kurtosis, and skewness) of the subset of information defined by the ontology we easily get our final features. Exploiting these features, we tested different models (from Kernel Methods to Gaussian processes and Ensemble Methods [9]) selecting RF [7] as it generally outperformed all the others. RF also allows for computationally inexpensive

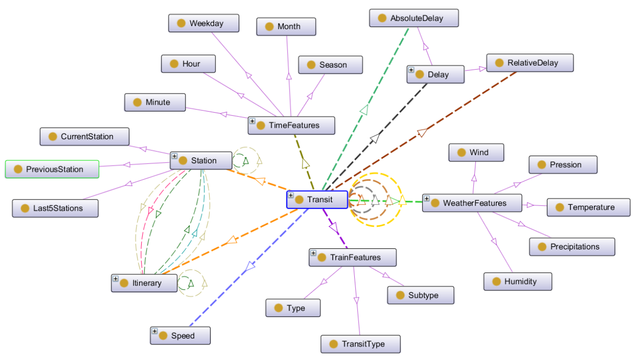


Fig. 2: Ontology supporting feature engineering

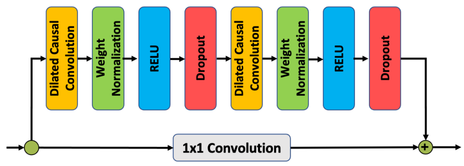
ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

feature ranking and then reduction [10]. In particular, using complete cross validation with 10 folds [11], we tuned the number of predictors to the randomly sampled in the construction of each tree in $\{1, 2, 4, 8, 16, 32, 64\}$, the minimum number of samples of each leaf in $\{1, 2, 4, 8, 16\}$, and the number of features to be kept during the feature reduction phase in $\{8, 16, 32, 64, 128\}$. The number of trees in the forest, instead, has been set to 1000 since increasing it has not shown to improve the accuracy.

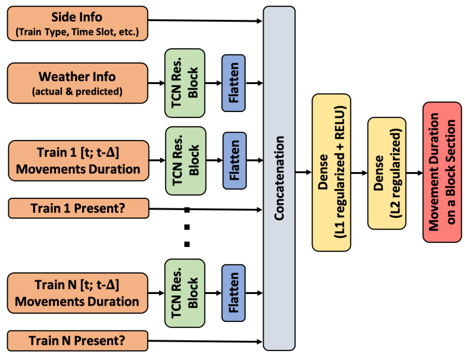
2.2 Deep Models

Note that shallow models have two main limitations. The first one is to rely on a handcrafted and experience-based feature engineering phase. The second, and most important one, is that these features cannot effectively model and extract all the multiscale information from the input time series.

In order to overcome these limitations we first rely on a state-of-the-art approach based on Long Short Term Memory (LSTM) networks [12]. In particular, we defined a standard architecture where the different raw time series are fed to an LSTM layer which is in charge to extract the representation (of length equal to 10) that is then fed directly to a dense L1 regularised layer and then to dense L2 regularised layer, together with a Boolean variable which indicated if the time series is present or not, and the side information. Network has been trained with the ADAM optimiser [12]. This architecture has a series of hyperparameters to be tuned via complete cross validation with 10 folds [11] and the space of hyperparameters have been explored via random search [13]. In particular we searched the cyclical learning rate [14] in $\{0.0001, 0.0005, 0.001\}$ and the L2 and L1 regularisation [12] in the last two dense layers in $10^{\{-4.0,-3.2,\cdots,4\}}$.



(a) TCN Residual Block

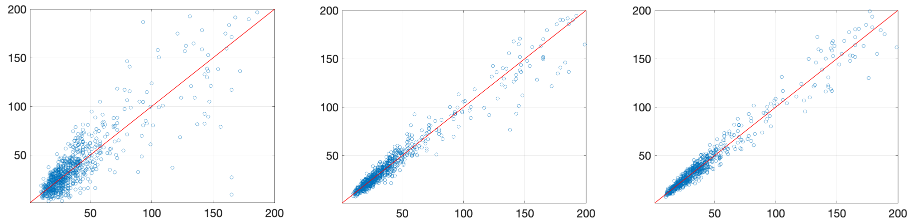


(b) High level architecture

Fig. 3: Proposed Deep Model

Unfortunately, this architecture was not able to outperform the shallow model as LSTM is only able to handle two temporal scales (a long and a short temporal scale). To overcome this limitation we decided to rely on TCN residual blocks [8], (replacing the LSTM) which is capable of learning different temporal scales for each raw input time series (Figure 3(a)). The peculiarities of the proposed Deep Multi Scale Models architecture based on TCN are mainly three: first the convolutions in the architecture are causal, namely there is not information leakage from future to past, second the architecture can handle different

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.



(a) Actual (MAE: 19.3±5.1) (b) Shallow (MAE: 7.8±2.5) (c) Deep (MAE: 6.9±2.1)

Fig. 4: Comparison between the models (everything is in seconds)

sequence lengths and map it to an output sequence of the same length as the LSTM, and finally is able to handle long effective history. For what concerns the first characteristic, the TCN uses causal convolutions. For what concerns the second characteristic, it is due to the use of 1D fully-convolutional network model where each hidden layer is the same length as the input layer. As for the last point we employ dilated convolution [15] that enables a large receptive field without employing too deep TCN residual blocks. The network has been trained with the ADAM optimiser. In this case, we explored the cyclical learning rate in $\{0.0001, 0.0005, 0.001\}$, the dropout after the ReLU activation functions in TCN residual blocks in $\{0.1, 0.15, \cdots, 0.5\}$ and the number of TCN residual blocks for each input data. Moreover, for each TCN residual block, we search the number of convolutional filters in $\{32, 64, 128\}$ and the kernel size of each convolutional filter in $\{3, 5, 7, 9, 11\}$. The high level proposed architecture is reported in Figure 3(b).

3 Numerical Results

The experiments have been conducted exploiting the real-world data coming from the Italian railway network and provided by RFI. In particular, we had access to data about trains movements (and the related side information), of 6 months (at the end of 2019 since, because of the COVID pandemic, circulation has been strongly reduced in 2020 and 2021) of trains movements of a series of stations in the North West of Italy[3]. We also exploit, as exogenous information, the weather conditions from the Italian open data weather stations service (e.g., solar radiation, rain, and wind).

We trained and validated [1] the different models (the one actually in use by RFI, the shallow one of Section 2.1, and the deep one[4] Section 2.2) on the first 5 months of data and we report in Figure 4 the Mean Average Error (MAE) and the scatter plot (thousand random actual against predicted movements duration in seconds) on the 6th month of data.

Figure 4 shows how the big leap in performance improvements is due to the

[3]All RFI-related data have been anonymised through the paper (e.g., name of the stations, id of the trains, and id of the block sections) because of confidentiality issues.

[4]The one based on LSTM has not been reported since it underperformed (MAE: 10.9±3.3) the shallow one (MAE: 7.8±2.5).

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.

use of data driven models with respect to the simple statistics actually exploited in RFI. Deep models are able to outperform the shallow ones showing that it is actually able to extract automatically an informative representation from data.

4 Discussion

In this work we focused on predicting in-station train movements which is a fundamental building block for developing and improving the current train dispatching. We took as baseline the current RFI system, which uses simple statistics, and we improve it using a shallow model on top of a carefully crafted feature engineering phase which exploits the domain knowledge of the experts in RFI. Then we further improve the shallow model with a deep architecture based on TCNs able to automatically learn expressive features from the data. Results on six months of data coming from a series of stations in the North West of Italy support the proposal.

This work is a first step towards improving the prediction of in-station train movements, and more tests (on more stations and for a longer period of time) on a larger period of time and space (once trains will restart to circulate after the COVID pandemic) need to be performed to extensively asses the potentiality of this solution. Moreover, research needs to be conducted to assess the potentiality of this tool to improve the state-of-the-art train dispatching systems [5, 6].

References

[1] L. Oneto, I. Buselli, A. Lulli, R. Canepa, S. Petralli, and D. Anguita. A dynamic, interpretable, and robust hybrid data analytics system for train movements in large-scale railway networks. *Int. Journal of Data Science and Analytics*, 9(1):95–111, 2019.

[2] F. Corman and P. Kecman. Stochastic prediction of train delays in real-time using bayesian networks. *Transportation Research Part C*, 95:599–615, 2018.

[3] G. Theeg and S. Vlasenko. *Railway signalling & interlocking*. Eurail-press, 2009.

[4] C. Wen, P. Huang, Z. Li, J. Lessan, L. Fu, C. Jiang, and X. Xu. Train dispatching management with data-driven approaches: a comprehensive review and appraisal. *IEEE Access*, 7:114547–114571, 2019.

[5] M. Cardellini, M. Maratea, M. Vallati, G. Boleto, and L. Oneto. In-station train dispatching: a PDDL+ planning approach. In *ICAPS*, 2021.

[6] D. Abels, J. Jordi, M. Ostrowski, T. Schaub, A. Toletti, and P. Wanko. Train scheduling with hybrid answer set programming. *Th. and Prac. of Logic Program.*, (4):1–31, 2020.

[7] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[8] S. Bai, J. Z. Kolter, and V. Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[9] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[10] Y. Saeys, T. Abeel, and Y. Van de Peer. Robust feature selection using ensemble feature selection techniques. In *ECML PKDD*, 2008.

[11] L. Oneto. *Model Selection and Error Estimation in a Nutshell*. Springer, 2019.

[12] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[13] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

[14] L. N. Smith. Cyclical learning rates for training neural networks. In *IEEE winter conference on applications of computer vision*, 2017.

[15] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.