Restoration Time Prediction in Large Scale Railway Networks: Big Data and Interpretability

Luca Oneto[1], Irene Buselli[1], Paolo Sanetti[1],
Renzo Canepa[2], Simone Petralli[2], Davide Anguita[1]

[1]DIBRIS - University of Genoa, Via Opera Pia 13, Genova, 16145, Italy
{luca.oneto,irene.buselli,paolo.sanetti,davide.anguita}@unige.it
[2]Rete Ferroviaria Italiana S.p.A., Via Don Vincenzo Minetti 6/5, Genova, 16126, Italy
{r.canepa,s.petralli}@rfi.it

Abstract. Every time an asset of a large scale railway network is affected by a failure or maintained, it will impact not only the single asset functional behaviour but also the normal execution of the railway operations and trains circulation. In this framework, the restoration time, namely the time needed to restore the asset functionality, is a crucial information for handling and reducing this impact. In this work we deal with the problem of building an interpretable and reliable restoration time prediction system which leverages on the large amount of data generated by the network, on other freely available exogenous data such as the weather information, and the experience of the operators. Results on real world data coming from the Italian railway network will show the effectiveness and potentiality of our proposal.

Keywords: Railway Network, Restoration Time Prediction, Big Data, Data Driven Models, Interpretable Models, Experience Based Models.

1 Introduction

The functional behavior of railway infrastructure assets degrades for many different reasons [8]: age, extreme weather conditions, heavy loads, and the like. For example, the influence of snow on switches is critical, in particular when switch heating is not functioning properly. Even worse is the case of wind in combination with snowfall, when the assets belonging to a specific area can be significantly affected. Additionally, problems can be introduced unknowingly by performing maintenance actions, for example by a simple human error or as a reaction of the system to changes made on an object [1]. For instance, some maintenance activities (e.g. tamping or ballast dumping) performed close to a switch can change the track geometry, then other parts of this asset must be adjusted to the new situation. One of the most crucial pieces of information needed to reduce the impact on train circulation of assets failures and maintenance is the restoration time, namely the time needed to restore the complete functionality of the asset [2].

For this reason in this work we will investigate the problem of predicting the time to restoration for different assets and different failures and malfunctions. In other words, the objective of this analysis is to estimate the time to restoration for planned (anticipating faults) and corrective maintenance (rectifying faults) by looking at the past maintenance reports, correlated to the different assets and different types of malfunctions. The predictive model needs to take into account the knowledge enclosed into maintenance reports, exogenous information such as the weather conditions (e.g. weather condition) and the experience of the operators in order to predict the time needed to complete a maintenance action over an asset and to restore its functional status. Moreover, the model should be interpretable enough to give insight to the operators in what the main factors influencing the restoration time are in order, for example, to better plan the maintenance activities. This information will help the Traffic Management System to assess the availability of the network, for example by estimating the time at which a section block including a malfunctioning asset will become available again, and properly reschedule the train circulation.

For this purpose we will build a rule-based model which is able to exploit real maintenance historical data provided by Rete Ferroviaria Italiana (RFI), the Italian infrastructure manager that controls all the traffic of the Italian railway network, the historical data about weather conditions and forecasts, which is publicly available from the Italian weather services, and the experience-based model currently exploited by the train operators for predicting the restoration time of planned maintenance. Results on these real world data will show the effectiveness and potentiality of our proposal.

2 Proposed Approach

The Restoration Time prediction problem can be easily mapped into a classical regression [7] problem where we have an input space $\mathcal{X}$ and an output space $\mathcal{Y} \subseteq \mathbb{R}$ and the purpose is to find the unknown relation μ between $\mathcal{X}$ and $\mathcal{Y}$. In our case $\mathcal{X}$ is the description of the maintenance action or the failure, plus some exogenous information such as the weather conditions, plus the experience of the operator that, thanks to this last information, is sometimes able to provide, for example in case of planned maintenance, an estimation of the restoration time. $\mathcal{Y}$, instead, is the actual restoration time that we want to predict. In this framework, we would like to find a model $h : \mathcal{X} \rightarrow \mathcal{Y}$ which approximates μ just based on a finite set of observations of μ, called dataset $\mathcal{D}_n = \{(X_1, Y_1), \cdots, (X_n, Y_n)\}$, composed of n tuples where $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. The model h should be a good approximation of μ, but it should also be easy to understand or interpret by a human operator who wants to get insights from the model h, and not just a prediction of the restoration time. For this purpose, in order to measure the quality of h in approximating μ we have to define one or more accuracy measures. Since we will exploit $\mathcal{D}_n$ for building h, first we need to exploit an additional fresh dataset of cardinality m, called test set $\mathcal{T}_m = \{(X'_1, Y'_1), \cdots, (X'_m, Y'_m)\}$ in order to be sure that the accuracy of h on $\mathcal{T}_m$ is an unbiased estimator of the true accuracy [6]. Then we will use as measures of accuracy the following quantities:

- the Mean Absolute Error: $\text{MAE} = {}^1\!/_m \sum_{i=1}^{m} |h(X'_i) - Y'_i|$;
- the Mean Absolute Percentage Error $\text{MAPE} = {}^{100}\!/_m \sum_{i=1}^{m} {}^{|h(X'_i) - Y'_i|}\!/_{|Y'_i|}$;
- the Pearson Product Moment Correlation Coefficient, (or bivariate correlation) $\text{PPMCC} = {}^{\sum_{i=1}^{m} (h(X'_i) - \bar{h})(Y'_i - \bar{Y}')}\!/_{\sqrt{\sum_{i=1}^{m} (h(X'_i) - \bar{h})^2} \sqrt{\sum_{i=1}^{m} (Y'_i - \bar{Y}')^2}}$ where $\bar{h} = {}^1\!/_m \sum_{i=1}^{m} h(X'_i)$ and $\bar{Y} = {}^1\!/_m \sum_{i=1}^{m} Y'_i$.

In order to ensure the interpretability of h, instead, we have to perform two different steps. First of all, we need a functional form of the model that is interpretable by construction, and hence the most reasonable choice is to use a rule based model, namely a decision tree. A decision tree can be efficiently and effectively learned from the data even when millions of samples are available [4, 5]. The second step is to find how to map $\mathcal{X}$ in such a way that the model, learned on this mapping, still remains interpretable. More formally we have to find a function $\phi : \mathcal{X} \to \Phi$ where Φ must be, from one side, a rich representation of $\mathcal{X}$ and, from another side, it must contain features that are easy to understand, grounded, and physically connected with the nature of the problem. In this way the final functional form of the model $h \circ \phi : \mathcal{X} \to \mathbb{R}$ will be an interpretable model based on a rich and interpretable feature set. Note that Φ may contain both numerical and categorical features since decision trees can efficiently and effectively handle naively both types of feature spaces. In order to create this rich and interpretable feature space we operated a two step approach. First we have enriched the original space $\mathcal{X}$ of new features designed together with the RFI experts in order to include their knowledge inside the ϕ. Note that the experience of the operator is also included in $\mathcal{X}$ as a feature which estimates the restoration time for the planned maintenance based on a model developed by the RFI experts during the years[1]. As a second step we estimated the most important and easy to interpret statistical descriptors of the features designed by the experts (mean, variance, skewness, kurtosis, etc.) Finally, we learned over $\Phi \times \mathcal{Y}$ a decision tree, pruning the tree based on the 10-fold cross validation principle [6]. The pruning procedure has been performed optimizing the number of points per leaf [4]. In order to handle the size of the problem, as in our case n can count even millions of samples, we implemented everything in Scala using Spark with the Decision Tree library included in MLlib [5]. As a final remark we would like to underline that, given the nature of the problem which evolves in time, in $\mathcal{D}_n$ we have the data until December 2016 while in $\mathcal{T}_m$ we have the data from January 2017 on; in this way we are actually simulating to apply the model in the future based on data about the past.

3 Experimental Evaluation

In order to test our proposal, we first have to describe the available data.

- RFI provided data from January 2015 until December 2018 about the Italian Railway Infrastructure[2]. They contain data about the location of the maintenance or fault, the stations and track involved, the estimated restoration time based on the RFI model for just the maintenance (RFI has not a model for the faults), and the actual duration of the maintenance. In the data there are also notes of the operators.
- From the Italian weather services [3] we retrieved data about the actual and predicted weather information regarding the same years of the RFI's data.

[1] The details about this model cannot be disclosed because of confidentiality issues.

[2] In order to give an idea of the number of maintenance and faults, just for a small region of the north of Italy we have more that 100.000 records; we cannot disclose more details because of confidentiality issues.

From these services it is possible to retrieve the hourly information about wind, temperature, rain, snow, solar radiation, and fog.

From these data, together with the RFI experts we extracted the feature set Φ described in Table 1. Then we implemented the Decision Tree using the MLlib [5]

Table 1: Feature set Φ extracted with the RFI experts.

Name	Meaning	Input/Output
Prov	Province of the intervention	Input
BegStat, EndStat	Geolocation of the beginning and end stations involved	Input
Track	What track are involved	Input
Type	Maintenance or Failure	Input
Intervention	Type of Intervention	Input
Day, Month, Hour	Information about the time of the beginning of the intervention	Input
Rain, Temp, Sun, Wind, Snow, Fog	Information about the weather at the beginning of the intervention	Input
PredictedTime	RFI estimated restoration time (just for the maintenance and not available for the failures)	Input
ActualTime	Actual restoration time (in minutes)	Output

in Spark and we deployed an infrastructure of four machines equipped with 32GB of RAM 500GB of SSD disk space and 4 CPU on the Google Compute Engine[3]. We implemented the 10-fold cross validation for optimizing the number of points per leaf n_l by searching $n_l \in \{5, 10, 20, 50, 100, 200, 500\}$. All experiments have been performed 30 times in order to ensure the statistical robustness of the results. Table 1 reports

- the error of the RFI model measured with the MAE, MAPE, and PPMCC on the maintenance since no model for the failures is available to RFI on $\mathcal{T}_m$;
- the error of our model measured with the MAE, MAPE, and PPMCC on all the intervention, on the maintenance, and on the failures on $\mathcal{T}_m$.

Figure 2 reports the first three levels of the model derived from the data[4].

Fig. 1: Quality of the models.

Int.	MAE	MAPE	PPMCC
RFI			
Maint.	30.5	31.5	0.75
Our Proposal			
All	11.3±1.1	10.7±0.9	0.93±0.03
Maint.	8.1±1.0	7.8±0.7	0.97±0.03
Fail.	15.2±1.3	14.3±1.1	0.88±0.04

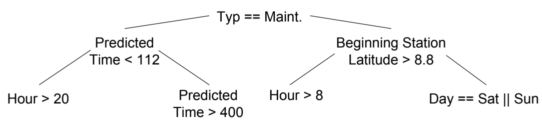

Table 2: First three levels of the model.

From Table 1 and Figure 2 it is possible to note that the quality of our model is remarkably higher than the one of the RFI model. In particular for

[3] `https://cloud.google.com`

[4] The full model which has more levels is not reported for confidentiality issues.

the maintenance we reach a MAPE lower than 10% which is more than 3× better than the accuracy of the RFI model. Even for the failures the accuracy of the restoration time model is remarkable if compared with the model of RFI. Note also that, by looking at Figure 2, the model is very easy to interpret and reasonable (e.g. the type of intervention is on top, together with the location of the intervention and the weather information). Note also that the model of RFI is taken considerably into account in case of maintenance as expected since the RFI model has already a quite high predictive power.

4 Discussion

In this work we dealt with the problem of predicting the restoration time of a part of the railway network from an intervention on an asset based on data coming from the railway information system, exogenous variables like the weather information, and the experience of the operators. Moreover, given the particular application which is very human oriented, it was required to build a model as interpretable as possible in order to help the operators in taking decisions not just based on a prediction but also based on the functional form of the model. For these reasons, we proposed an approach which produces easy-to-interpret models, is computationally efficient and effective, and is able to handle a huge amount of historical interventions. Results on data provided by RFI coming from Italian Railway Network support our proposal both in terms of quality and interpretability of the derived models.

Acknowledgments This research has been supported by the European Union through the projects IN2DREAMS (European Union's Horizon 2020 research and innovation programme under grant agreement 777596).

References

1. Ahren, T., Parida, A.: Maintenance performance indicators (MPIs) for benchmarking the railway infrastructure: a case study. Benchmarking: An International Journal 16(2), 247–258 (2009)
2. Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J.: An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological 63, 15–37 (2014)
3. Italian Weather Services: Weather Data. `http://www.cartografiarl.regione.liguria.it/SiraQualMeteo/script/PubAccessoDatiMeteo.asp` (2018), online; accessed 16 October 2018
4. James, G., Witten, D., Hastie, T., Tibshirani, R.: An introduction to statistical learning. Springer (2013)
5. Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D.B., Amde, M., Owen, S.: Mllib: Machine learning in apache spark. The Journal of Machine Learning Research 17(1), 1235–1241 (2016)
6. Oneto, L.: Model selection and error estimation without the agonizing pain. WIREs Data Mining and Knowledge Discovery 8(4), e1252 (2018)
7. Shalev-Shwartz, S., Ben-David, S.: Understanding machine learning: From theory to algorithms. Cambridge university press (2014)
8. Zoeteman, A.: Asset maintenance management: state of the art in the european railways. International journal of critical infrastructures 2(2-3), 171–186 (2006)