The Fifth International Conference on Railway Technology
RAILWAYS 2022
22–25 August 2022
Montpellier, France

Proceedings of the Fifth International Conference on
Railway Technology:
Research, Development and Maintenance
Edited by J. Pombo
Civil-Comp Conferences, Volume 1, Paper 10.11
Civil-Comp Press, Edinburgh, United Kingdom, 2022, doi: 10.4203/ccc.1.10.11

Predictive analysis of fatigue crack growth on railroad tracks using machine learning techniques

M. Leyli-abadi[1] and O. Vo Van[2]

[1]IRT Systemx, Palaiseau, France
[2]SNCF, France

Abstract

The railroad tracks are frequented by thousands of rolling stocks every day. Depending on the type of the rolling stocks (fret, passenger, etc.) and the corresponding conveyed weight, the different parts of railroad tracks are under a constant stress. In recent years, thanks to the technological advances, more data are collected using automatic inspections of railroad tracks and infrastructure and have been acquired by the French National Railroad Company (SNCF). In this article, the objective is to analyse the fatigue crack propagation on subsurface of rails (squat defects) with the aim to avoid the potential rail fractures. As a considerable amount of data is provided in this work, we propose the use of data-driven techniques for prediction of the evolution of crack lengths over time. These models have the advantage of considering a number of influent factors in the modelling unlike the mechanical models, e.g., infrastructure and traffic related factors, climatic variables, etc. However, the calibration of the hyperparameters of data-driven models is of utmost importance. We have conducted an analysis to study the effect of hyperparameters on the predictive capacity of models. Finally, a number of state-of-the-art machine learning techniques are evaluated for the prediction of fatigue crack length and their performances are compared. The neural network based models obtain the promising results and could be investigated in more depth in future works. We give also some insights of models which consider temporal dependency between observations.

Keywords: Fatigue crack propagation, Railroad fatigue, Machine Learning, Predictive analysis

1 Introduction

The railroad tracks are frequented by thousands of rolling stocks every day. Depending on the type of the rolling stocks (fret, passenger, etc.) and the corresponding conveyed weight, the different parts of railroad tracks are under a constant stress. This could initiate the apparition of cracks on the surface (headcheck defects) or subsurface (squat defects) of rails which evolve very fast over time. In the most extreme case, a fracture can happen which cause not only a heavy economic impact (delayed trains, maintenance cost), but also threaten the passenger safety (derailment). To avoid such situations, the French National Railroad Company (SNCF) performs a strict method with both corrective and preventive maintenance strategy which implies a significant cost for the company.

Many studies have been carried out to analyse the fatigue behaviour in mechanical fields [1] [2] [3]. These models being slow, their use is restricted to small-scale railroad networks. Furthermore, such models could not consider exogenous contextual factors which may have an impact on the evolution of crack propagation.

In recent years, thanks to the technological advances, more data concerning the crack propagation and contextual factors are obtained from automated inspections. This allows the use of data driven models to study the evolution of fatigue crack growth. A recent study has used the extreme Gradient Boosting algorithm to predict the broken rail rate [4]. The neural networks are also used for prediction of track structural defects [5].

In this article, we are interested in predictive maintenance strategy using data-driven techniques. The contribution of this paper can be summarized as below:

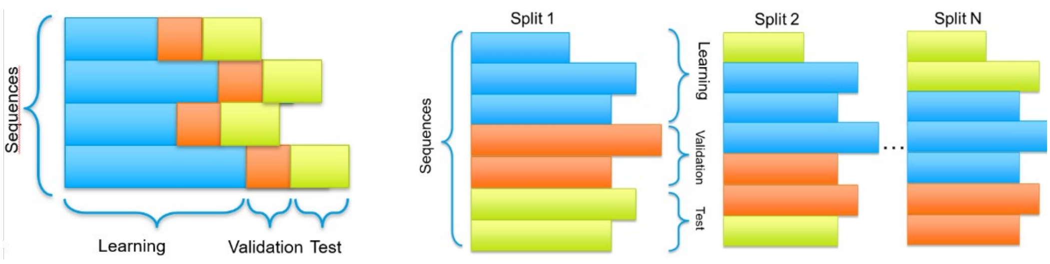
- Considering various available factors which may have an impact on the evolution of crack propagation, e.g. infrastructure-based variables (curvature, rail profile, etc.), climatic variables (temperature), traffic-related information (gross tonnage, number of passing cars). A crossing between the heterogeneous information is performed for each rail segment (using kilometre point);
- Extraction of relevant features such as the rail age, the duration after the first detected defect and cumulative temperature, which show a high correlation with fatigue propagation;
- Benchmarking the state-of-the-art statistical models (Random Forest, Gradient Boosting) and deep neural networks with the aim to select the best performed model for prediction of crack length over time;
- As the fatigue crack growth are represented by temporal series of crack lengths (in mm), we have also studied the models that consider the temporal dependency between consecutive observations, i.e., Markov models and Recurrent Neural Networks.

2 Methods

The main objective of this article is to benchmark and compare various state-of-the-art data-driven models with the aim to select the best performed technique for prediction of fatigue crack propagation. We are in supervised machine learning setting where each method requires a set of (X, y) pair of data for learning. Where X represents a vector of available features and y represents corresponding ground truth crack propagation value. In this study, the data gathering procedure is performed from 2013 to 2018 with variable frequency.

The set of features can be categorized in following groups: infrastructural features which are mainly static over time (nominal velocity of trains, curvature, rail material and profile, etc.), the traffic data which is represented by the tonnage of rolling stocks for each segment of the railroad network and is also static, climatic data including the temperature from 2013 to 2018 with daily and monthly frequency, and finally each considered railroad segment is supervised with variable frequency and the crack propagation measure (crack length in mm) is reported with the corresponding dates which is a time series data.

It should be mentioned that a huge work of data preparation is also performed to be able to conduct this analysis. As each above-mentioned information is recorded and reported separately, they have been crossed at the rail track segment level. The anomalies such as inconsistencies in crack propagation are identified and removed with the aim to avoid to introduce some bias into the models. Furthermore, some important features are also extracted to improve the learning convergence and to obtain the better predictions, which are the age of the rail segments and the duration after the first detected defect (this helps to consider the irregular time interval into models).



(a) Classic strategy (b) Leave-k-out sequence cross validation

Figure 1 Experimental settings. Each horizontal rectangle presents a sequence of observations (crack length over time) and each color represents a specific partition of data.

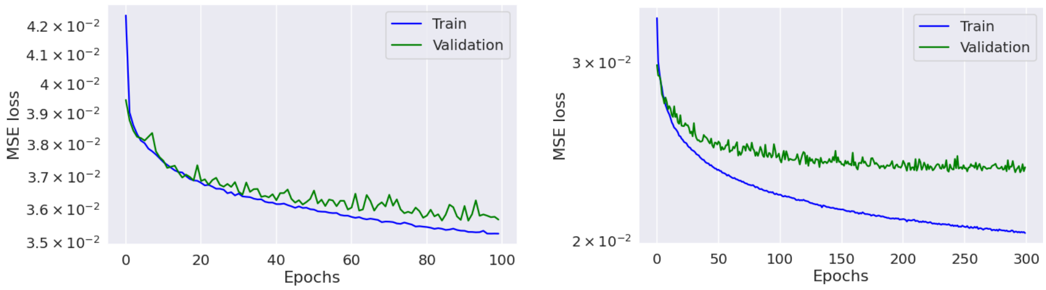
We consider also two different ways to evaluate the performance of the previously mentioned approaches. The first approach (see figure 1a) consists in dividing each sequence into three partitions, where 60% of each sequence is dedicated for training, 20% for validation and 20% for testing. We have also experimented a shuffled version of this first approach where the temporal nature of data is neglected. The second approach (see figure 1b) consists in partitioning by sequence. To do this, we have considered 60% of sequences for training, 20% for validation and 20% for testing. It should be mentioned that the validation dataset is used for fine-tuning and the test dataset has never been observed during the training phase.

.

3 Results

We have considered two different features set for evaluation of previously mentioned approaches. A first set includes only the features coming directly from the available datasets, and the second set adds the extracted features ε to the first set comprising 15 features. In these preliminary results, we have considered the first experimental configuration that has been mentioned in previous section which does not consider the temporal nature of observed data. All the methods are fine-tuned such as to select the set of best hyper-parameters.

Concerning the Multi-Layer Perceptron neural network, we have demonstrated the convergence curves for two different set of hyperparameters in Figure 2. A first set chosen randomly (3 hidden layers with 20, 40 and 20 neurons, batch size=128 and number of epochs=100) and the second obtained after fine-tuning (3 hidden layers with 50, 100 and 50 neurons, batch size=32 and number of epochs=300). It can be seen that the MSE loss (logarithmic scale) has been reduced significantly using the fine-tuned set of hyperparameters.



(a) Random set of hyperparameters (b) Fine-tuned hyperparameters

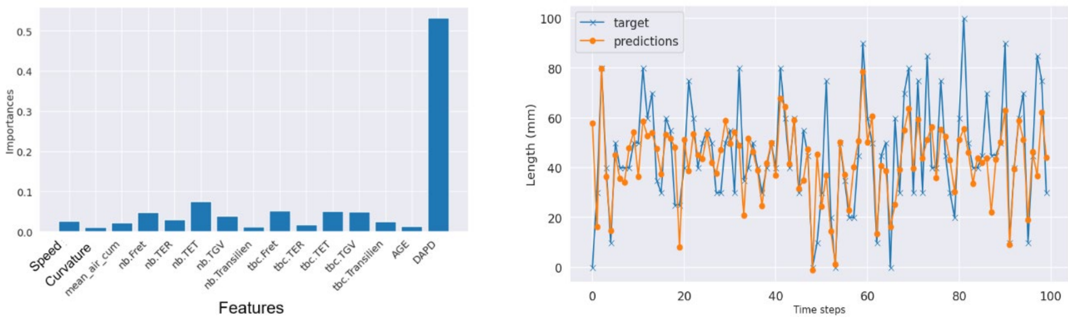
Figure 2 Convergence with respect to hyperparameters

The evaluation results are shown in Table 1. In this table, the best performances are highlighted in bold for each evaluation criteria. Two statements could be concluded from this table. First, the combination of original features with the extracted ones improve significantly the performances of all the methods. Second, the neural network based approach outperforms the traditional approaches for the regression task regarding all the evaluation criteria. An example of predictions for 100 different observations is also shown in Figure 3b. This encourages the study of more complicated architectures which may improve still the performance by considering the temporal dependencies among observations.

Table 1 Comparison table. The evaluated methods are: Random Forest (RF), Gradient Boosting (GB), Multi-layer perceptron (MLP). The evaluation criteria are: Mean absolute error (MAE), Mean Squared error (MSE), Root Mean Squared Error (RMSE) and Normalized Root Mean Squared Error (NRMSE).

Approaches	Features	Evaluation criteria			
		MAE	MSE	RMSE	NRMSE
RF	X	14.42	342.85	18.52	0.21
	X + ε	12.01	249.75	15.80	0.20
GB	X	13.17	290.88	17.06	0.17
	X + ε	11.73	235.27	15.34	0.16
MLP	X	14.61	351.26	18.74	0.25
	X + ε	**11.43**	**234.21**	**15.30**	**0.11**

Although, the neural network based algorithms show outstanding performances recently in many domains, their black box nature does not allow to analyse the impact of the input features on the response variable. It is of the utmost importance for rail network operators to identify this impact to be able to select the right action or still to improve the materials used with respect to specific regions. In Figure 3a, we can see the impact of different variables on the fatigue crack propagation. It can be seen that the extracted feature, i.e., duration after the first detected defect (DAPD) is the most important feature as the cracks propagates very fast, once they are initiated. The frequency and tonnage of the rolling stocks show also some impacts.



(a) Feature importance (b) Predictions using MLP

Figure 3 Feature importance obtained using Random Forests in (a) and predictions vs. ground truth crack lengths in (b)

4 Conclusions and Contributions

In this article, we have compared various data-driven models for the prediction of fatigue crack propagation. Different strategies have been considered for the evaluation of methods to verify their generalization capacity on unseen data. It has been shown that the use of extracted features helps to improve significantly the performance of the evaluated approaches. On the basis of Random Forest analysis, the duration after the first detected defect (DAPD) was the most important variable and the tonnage of rolling stocks show important impact on the crack propagation.

The analysis highlighted the superiority of neural network based method over traditional machine learning algorithms.

We are currently working to add more complex techniques in the experimentation which allow to consider the temporal dependency of crack propagation series in modeling. One approach from probabilistic domain that we would like to experiment is based on Markov chains [6] and the other one is based on recurrent neural networks. The Markov chains are shown in Figure 4 for multiple sequences **z** and input vectors ***u***, and is based on the hypothesis that the current state (time *T*) depends only on the last observed state (time *T-1*).

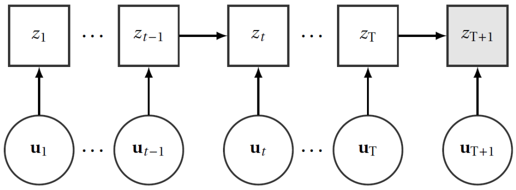


Figure 4 Joint hidden Markov models

Concerning the recurrent neural networks (see Figure 5), multiple variants exist, e.g., LSTM [7], GRU [8], etc. As an example, the LSTM model will allow to consider long term dependencies through a memory cell. In our case, the sequence lengths and the number of measurements per sequence differs from one sequence to another depending on inspection frequency per rail segment. Furthermore, the exogenous factors could be classified into two groups: temporal (temperature) and static inputs variables (traffic and infrastructure). These models should be adapted to consider the variable length sequences and also the two groups of exogenous factors.

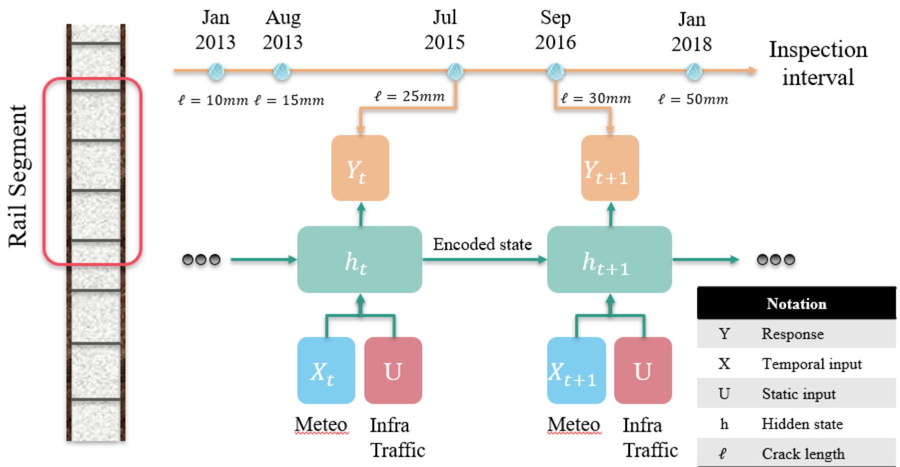


Figure 5 Modelling the fatigue crack propagation using Recurrent Neural Network

As a perspective for future investigations, we aim to use recent explicability tools, e.g., Lime or SHAP, to analyse the influence of input features on crack propagation when using neural networks. Furthermore, the data-driven approaches could be combined with mechanical models to build a "hybrid model" that takes advantage of both. More precisely, as the hybridation result, the fatigue crack propagation could be classified as critical for a given time stamp, and a maintenance operation like rail grinding or rail replacement should be performed.

References

[1] T. Bonniot, V. Doquet and S. H. Mai, "Fatigue crack growth under non-proportional mixed-mode loading in rail steel," *PhD Thesis, Ecole Polytechnique,* p. 134, 2020.

[2] M. Nguyen-Tajan and C. Funfschilling, "A numerical modeling strategy for the rolling contact fatigue analysis of rails," in *WCRR*, 2011.

[3] M. C. Burstow, "Whole Life Rail Model application and development. Continued development of an RCF damage parameter," 2004.

[4] Z. Zhang, K. Zhou and X. Liu, "Broken rail prediction with machine learning-based approach," in *ASME/IEEE Joint Rail Conference*, 2020.

[5] J. Sadeghi and H. Askarinejad, "Application of neural networks in evaluation of railway track quality condition," *Journal of mechanical science and technology,* vol. 26, pp. 113--122, 2012.

[6] K. Fujinaga, M. Nakai and H. Shimodaira, "Multiple-regression hidden Markov model," in *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, 2001.

[7] H. Sak, A. Senior and F. Beaufays, "Long short-memory recurrent neural network architectures for large scale acoustic modeling," 2014.

[8] J. Chung,, C. Gulcehre, K. Cho and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555,* 2014.