ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Automated green machine learning for condition-based maintenance

Afonso Lourenço*, Carolina Ferraz*, Jorge Meira*, Goreti Marreiros*, Verónica Bolón-Canedo[ⱡ] and Amparo Alonso-Betanzos[ⱡ]

GECAD* - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Polytechnic of Porto, Portugal

CITIC[ⱡ] - University of Coruña, A Coruña 15071, Spain

Abstract. Within the big data paradigm, there is an increasing demand for machine learning with automatic configuration of hyperparameters. Although several algorithms have been proposed for automatically learning time-changing concepts, they generally do not scale well to very large databases. In this context, this paper presents an automated green machine learning approach applied to condition-based maintenance with automatic data fusion and density-based anomaly detection based on locality sensitivity hashing. Experiments on numerical simulations of train-track dynamic interactions demonstrate the utility of the approach to detect railway wheel out-of-roundness. This unlocks the full potential of scalable machine learning, paving the way for environment-friendly systems and automated decision-making.

1. Introduction

The railway sector offers a great illustration of how green automated machine learning can be used in real-world applications, due to its huge network of long rail lines and sensors dispersed over great distances. Ensuring reliability is crucial for the efficient transportation of passengers and freight in the railway sector [1]. In particular, the small wheel-rail contact patch has received a lot of attention in the literature, due to its increasing hazard function and associated risk of train derailment [2]. The high stick and sliding stresses in rolling contact leads to wheel out-of-roundness (OOR) in the form of flats and polygonization [3]. However, it is extremely challenging to distinguish between damage and environmental or operational variations (EOVs) [4]. To overcome these issues in the form of large amounts of unlabeled, noisy, and non-stationary time series signals, several contributions were made, mostly with data fusion techniques from multiple sensors [5]. However, few works achieved automated fault detection by

* The present work has been developed under project FERROVIA 4.0 (POCI-01-0247-FEDER-046111) and WAY4SafeRail (NORTE-01-0247-FEDER-069595). It has been supported by Portuguese Foundation for Science and Technology under project UIDB-00760-2020 and Spain´s Ministerium of Science and Innovation MCIN-AEI-10.13039-501100011033 under grant PID2019-109238GB-C22.

a user-friendly solution [6]. To bridge this gap, the added value of this paper lies in building a green automated configuration for OOR detection, with two failure modes: wheel flats and wheel tread polygonization. The developed strategy was evaluated with numerical train-track dynamic interaction simulations [7].

2. Fault detection methodology

The proposed method encompasses a four-stage process with: (1) windowing by hidden Markov model (HMM); (2) extraction of localized features with short-time Fourier transform (STFT); (3) data fusion with Principal Component Analysis (PCA); and (4) anomaly detection based on locality sensitive hashing technique (LSHAD).
The HMM was deployed to extract the wheel-rail dynamics from the vibration signals, associating sequence of observations with different hidden internal states [8]. The attainment of segments that adapt to the shape of the signal allowed to automatically index each segment to the corresponding wheel and enable the direct comparison of individual wheel passages. A total of 15 time-domain features were extracted for each segment, including mean, max, min, power, and margin. Envelope metrics, such as peak-to-peak amplitude, root mean square, variance, standard deviation, skewness, kurtosis, crest factor, form factor, and pulse indicator, were also selected to capture the overall shape of the data. Secondly, to address the challenges imposed by non-linear factors in the dynamic railway environment, the STFT was adopted to generate frequency-domain features [9]. This approach works by computing the Fourier transform for overlapping short-time windows of a segmented signal, considered as locally stationary. Windows were generated by splitting the signal into the HMM-based sequences and creating windows corresponding to overlaps of adjacent segments by 50%, which helped to provide better time and frequency resolution for different parts of the vibration signal. The resulting number of windows is equal to twice the number of segments minus one. From the magnitude spectra of each time segment, a power spectrum of the signal was computed, and 7 features were extracted to detect magnitude variation over frequency, i.e., the mean, peak, sum, variance, skewness, kurtosis, and relative margin. To reduce the dimensionality and mitigate the effects of EOVs, a latent variable approach of Principal Component Analysis was adopted [10]. The number of components to extract was determined with scree plots and eigenvalue thresholds.
In this reduced dimension space, LSHAD was used for fault detection. The algorithm, previously proposed by the authors [11], is highly parallelizable and incorporates automatic hyperparameter tuning so that users do not implement costly manual tuning. In addition, it has the added advantage of being an unsupervised learning process, that does not need labelled data. The tuning mechanism can adjust regardless of the input data domain. Nonetheless, in groups of two axles there is a partial confoundment in the vibration signal of consecutive wheels with one another due to the closely connected wagons. This leads to a different data structure representation, tackled with two separate

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

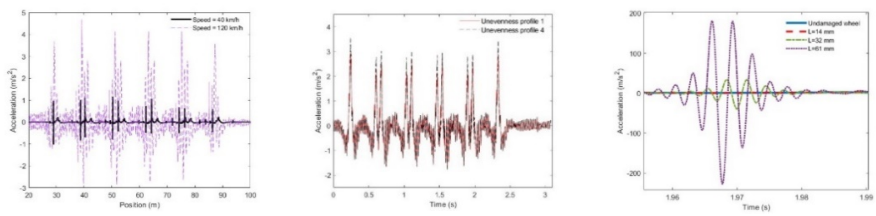
pipelines of PCA and LSHAD, allowing to better capture these differences. To evaluate the model, two criteria were used: the area under the receiver operating characteristic (AUC) and the relative increase in modelling time with larger datasets. For comparison, three state-of-the-art models were considered: Isolation Forest (iForest), Support Vector Machine (SVM), and Local Outlier Factor (LOF).

3. Numerical simulation

The robustness of the proposed methodology was validated with the virtual simulation of undamaged and damaged wheel scenarios under varying operational effects, i.e., rail irregularities, train speed and train loading. To accomplish this, the passage at speeds ranging from 80 to 120 km/h for a Laagrss-type freight train composed of five wagons were simulated with a train-track interaction model, using Hertzian theory. Regarding the track, it was modelled with a multiple-layer scheme, in which the ballast, sleepers and rails are linked through elastic elements. This study considers ten unevenness rail profiles corresponding to wavelengths D1 and D2 of the European Standard EN 13848-3 [12]. The resultant dynamic responses were evaluated with 24 accelerometers and 24 strain gauges at a sampling and cut-off frequency of 10 kHz and 500 Hz, respectively. Additionally, an artificial noise with 5% amplitude was added to represent possible environmental effects. With these settings, datasets of increasing sizes (1x to 25x) were generated as depicted in Table 1. For damaged scenarios, three different polygonal wheel profiles were considered, with wavelengths comprising 1-3, 6-8 and 18-19 harmonics, as well as two different wheel profiles with flat lengths drawn from uniform distributions between 25-50 and 50-100 mm. Note that the perimeter of the wheel for a Laagrss-type rail vehicle is 2.7m. The generated defects were placed randomly in one of the ten axles. Finally, a low-pass Chebyshev type II digital filter was applied. Fig. 1 illustrates the effects of varying speeds (a), different rail irregularity (b), and wheel flat lengths (c). Full description of train, track and interaction can be found elsewhere [7].

	1x	2x	5x	10x	25x	% Poly	% Flat
1 axle *(n)*	7440	14880	37200	74400	18600	9.68	6.45
2 axles *(m)*	16080	32160	80400	160800	40200	0	4.48

Table 1: Number of wheel passages in various datasets



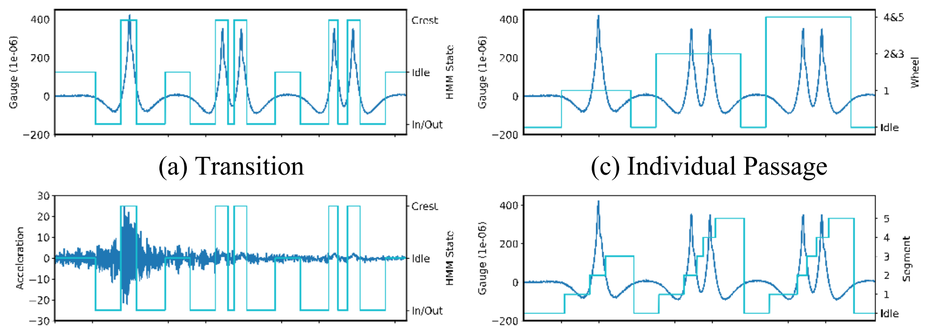
(a) Speed (b) Unevenness (c) Flat length

Fig. 1: Numerical simulation

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

4. Results and discussion

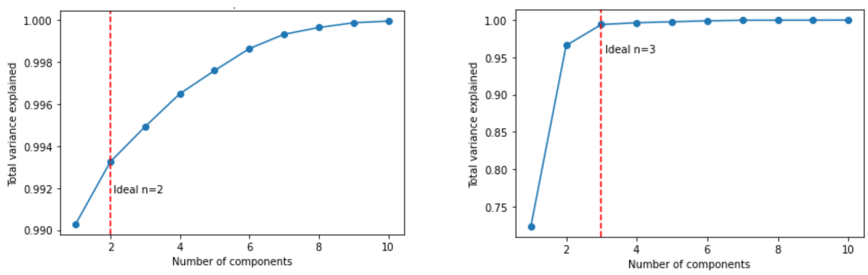
Firstly, the HMM was automatically fitted to the shape of the gauge signal with the hmmlearn python package. An analysis of the effect of parameters confirmed the Expectation-Maximization approach achieved convergence in less than 100 iterations with 3 gaussian states. Leveraging on the obtained indexing, automatic segmentation could be performed for the attainment of 3 and 5 subsequences in the groups of one and two axles, respectively. To offer a concise discussion of the results, Fig. 2 shows the gauge indexing (a), corresponding acceleration (b) and segmentation both based on a passage (c) and subsequences (d) for an exemplary train passage.



(a) Transition
(c) Individual Passage
(b) Indexing
(d) Sequences for STFT

Fig. 2: Data segmentation

For each subsequence, a total of 22 features were extracted obtaining matrices of n-by-66 and m-by-110, where n and m are number of passages with groups of one and two axles, respectively. To reduce the dimensionality and disregard the noise effect of EOVs, PCA was applied to both datasets. The scree plots for groups of one and two axles are presented in Fig. 3, showing that the selected principal components explained well over 90% of the data variance. The obtained matrices were n-by-2 and m-by-3.



(a) One axle
(b) Two axles

Fig. 3: Data fusion

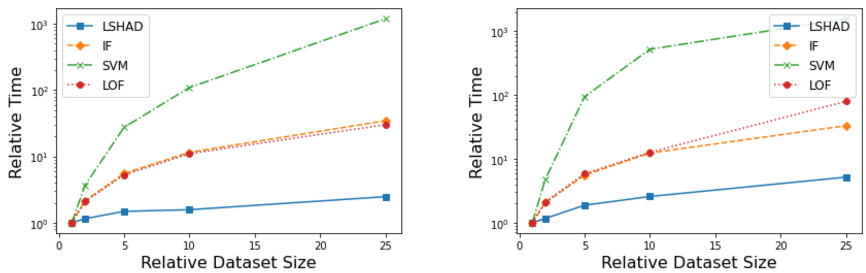
ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

For anomaly detection, LSHAD was performed using the Apache Spark implementation, with two folds applied, while the remaining algorithms were implemented in scikit-learn python package. To allow for a fair performance benchmark of the LSHAD, the other algorithms were also submitted to hyperparameter tuning. In iForest, a grid search was performed varying the number of trees from 150 to 450 and samples proportion required to split from 0.01 to 0.81. In SVM, the gamma parameter that determines the influence of each training example on the model's decision boundary was varied between 0.1 and 0.9. In LOF, it was studied a range between 100 and 500 of nearest neighbours to consider when calculating the local density of each data point. Table 2 presents the AUC with the best hyperparameter set.

	LSHAD		iForest		SVM		LOF	
	One Axle	Two Axles	One Axle	Two Axles	One Axle	Two Axles	One Axle	Two Axles
1x	0.759	0.885	0.835	0.873	0.513	0.316	0.861	0.873
2x	0.754	0.882	0.833	0.874	0.512	0.255	0.854	0.863
5x	0.764	0.912	0.864	0.880	0.513	0.503	0.779	0.721
10x	0.726	0.909	0.851	0.875	0.513	0.504	0.683	0.585
25x	0.769	0.909	0.869	0.877	0.513	0.504	0.555	0.526

Table 2: AUC for anomaly detection

In terms of time, the hyperparameter tuning process was excluded from the analysis. Fig. 4 shows the duration of fitting and predicting operations. Experiments were performed in a laptop with 8GB of RAM and a 1.30GHz Intel Quad-Core i7 processor.



(a) One axle (b) Two axles

Fig. 4: Relative processing time

Across all dataset sizes, LSHAD has demonstrated comparable AUC results. Additionally, LSHAD has significantly outperformed the other algorithms in terms of processing times, with detection times ranging from 30 seconds to 3 minutes, faster than other algorithms with detection times up to 14 minutes for the largest dataset. These results demonstrate the superior performance of LSHAD in detecting anomalies both efficiently and sustainably, making it more suitable for real-world applications.

5. Conclusions

In this paper, a scalable and sustainable methodology was developed for condition-based maintenance of wheel out-of-roundness. To achieve this, an anomaly detection model based on locality sensitive hashing processed large time series datasets generated by sensors in the railway network. The approach employs an Apache Spark framework for parallelized implementation, enabling efficient data processing across multiple clusters. Moreover, the model automates the hyperparameter tuning process, saving time, and reducing the potential for errors, while making the algorithm accessible to non-expert users in the machine learning field. As a preprocessing step, hidden Markov models and short-time Fourier transform were used for window-based feature extraction in the statistical and frequency domain from multiple redundant signals.
This performance study demonstrates the effectiveness and efficiency of the fault detection method, regardless of different wheel profiles, track irregularities, train speeds, position of the sensors and artificial noise, posed by non-linear environmental and operational factors that can distort the signal. Future efforts rely on the implementation of an online version of this methodology for dealing with data streams. With successful validation of this extension, the method has the potential to be applied in a field trial based on on-site measurements.

References

[1] S. Hidirov and H. Guler, ‘Reliability, availability and maintainability analyses for railway infrastructure management’, Structure and Infrastructure Engineering, vol. 15, no. 9, pp. 1221–1233, 2019.

[2] R. D. Fröhling and G. Hettasch, ‘Wheel-rail interface management: a rolling stock perspective’, Proc Inst Mech Eng F J Rail Rapid Transit, vol. 224, no. 5, pp. 491–497, 2010.

[3] S. Y. Chong, J.-R. Lee, and H.-J. Shin, ‘A review of health and operation monitoring technologies for trains’, Smart Struct Syst, vol. 6, no. 9, pp. 1079–1105, 2010.

[4] A. Lourenço et al., ‘Adaptive time series representation for out-of-round railway wheels fault diagnosis in wayside monitoring’, Eng Fail Anal, 2023.

[5] A. Lourenço, J. Meira, and G. Marreiros, ‘Online Adaptive Learning for Out-of-Round Railway Wheels Detection’, in Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, in SAC ’23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 418–421.

[6] M. Chenariyan Nakhaee et al., ‘The recent applications of machine learning in rail track maintenance: A survey’, in Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification: Third International Conference, RSSRail ‘19, Lille, France, 2019, Springer, 2019, pp. 91–105.

[7] P. A. Montenegro et al., ‘Wheel–rail contact formulation for analyzing the lateral train–structure dynamic interaction’, Comput Struct, vol. 152, pp. 200–214, May 2015.

[8] L. R. Rabiner, ‘A tutorial on hidden Markov models and selected applications in speech recognition’, Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.

[9] D. Griffin and J. Lim, ‘Signal estimation from modified short-time Fourier transform’, IEEE Trans Acoust, vol. 32, no. 2, pp. 236–243, 1984.

[10] I. Jolliffe, ‘Principal component analysis’, Encyclopedia of statistics in behavioral science, 2005.

[11] J. Meira, et al., ‘Fast anomaly detection with locality-sensitive hashing and hyperparameter autotuning’, Inf Sci (N Y), vol. 607, pp. 1245–1264, 2022.

[12] E. C. for Standardization, EN 13848-3 Railway Applications: Track. Track Geometry Quality. Part 3: Measuring Systems. Track Construction and Maintenance Machines. CEN, 2009.