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/.

Simultaneous failures classification in a predictive maintenance case

Antoine Hubermont[1,2],Elio Tuci[1] and Nicola De Quattro[2]

1- UNamur - Computer Science faculty
Rue Grandgagnage 21, 5000 Namur - Belgium

2- Telespazio Belgium
Rue devant les Hêtres 2, 6890 Libin - Belgium

Abstract. In industry 4.0, Machine Learning coupled with sensors monitoring leverages new ways to optimise maintenance strategies. In a predictive maintenance case, failure diagnoses are an excellent way to prevent any breakdowns. Up to now, failure diagnoses are focused on the classification of only one failure among many (multi-label classification), even if multiple failures can occur simultaneously. This study proposes an extension to classify simultaneous failures with the most popular classification methods such as random forests or artificial neural networks. Validated on a public predictive maintenance dataset, our methodology achieved classification with equal or best accuracy compared to multi-label classification.

1 Introduction

With the rise of Industry 4.0, sensors are widely used to monitor various industrial processes, in particular those based on production lines [1]. Scientific and technological developments allow to perform this monitoring process with the support of artificial intelligence based algorithms in order to develop adaptive and optimised maintenance strategies [2]. In the literature, there are three main types of maintenance: preventive, corrective, and predictive maintenance [3]. The first type (preventive) schedules maintenance operations before any failure; the second type (corrective) acts after a failure to repair the system's faulty components; the third type (predictive) predicts when and where a breakdown is likely to happen in order to carry out actions to prevent it. Predictive maintenance (hereafter, PdM) allows to reduce costs by avoiding both the replacement of healthy components and the occurrence of any service disruption by anticipating problems. Two main approaches have been identified in the literature to implement PdM: i) the diagnostic approach to identify the sources of the failure; and ii) the prognostic approach to forecast the remaining useful life of the system's components [4]. By taking a diagnostic approach, our intention is to provide operators and decision makers with more precise information (i.e., the source and the name) about the nature of the failure. In a systematic review of PdM applications [5], the authors state that, in the literature, the most frequently used AI techniques to implement PdM is Random Forest (RF) [3] used in 33% of the reviewed papers, followed by Artificial Neural Networks (ANN) used in 27% of the reviewed papers.

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/.

Our methodology is dedicated to diagnose failures with the most frequently used AI techniques. Up to now, these AI techniques are used in single-label classification in PdM context. Single-label classification is achieved by assigning a single failure among all the possible ones [6, 7]. Assuming that a maintenance process for a given monitored system could be carried out using a limited set of labels (e.g., no failure, failure A, failure B), single-label classification assigns one of these labels to any failure event. However, the approach currently employed in the PdM literature does not take into account the case where two types of failure occur on the same monitored system at the same time (simultaneous failure). This is the case in [6] where the authors applied multiple AI techniques to achieve single-label classification case on a public PdM dataset [8] containing simultaneous failure. As result, for every timestep affected by two failure, the label only refers to one failure while the other is ignored. In this paper, we have quantified the impact of this loss information for the datset in [8], and we found that up to 11% of the timesteps impacted by a failure are impacted by two failure. Our methodology aims to study the impact of multi-label classification to solve the present loss information problem in PdM case. In order to easily compare the performances of our approach with what already described in the literature, we replicate and extend the study described in [6] and we also contributing to the already made research on the public PdM dataset [8].

2 Methodology

A public dataset dedicated to PdM is made available by Microsoft for the Azure suite using synthetically-generated data [8]. This dataset is composed of 100 machines used in heavy industry, monitored from 2015-01-01 to 2016-01-01. The vibration, rotation, pressure and voltage are measured each hour with sensors placed on each machine. A machine is characterised by its service age (i.e., from 0 to 20 years) and is composed of 4 components. The nature of these components is not detailed, but it is assumed that each component can be replaced if faulty. The dataset includes the corrective and preventive actions operated to the components. Also, a troubleshoot record includes the type of error (5 errors types are listed) occurred at the machines and the failure/s occurred on the component/s. Failure implies a service interruption while error does not compromise the system functionality. Also, a component can be defined with 2 distinct states, faulty or normal (N). Therefore, failures are named according to which component is faulty (i.e., C1, C2, C3 and C4). In the dataset, each monitoring action, hereafter referred to as timestep, is timestamped. To minimise the noise on sensor readings performing the monitoring, the following features are added at each timestep: the time since last maintenance operation on each component (in hour), the number of errors (and their type) occurred in the last n hours, the machine service age, a long moving average and standard deviation of sensors' data computed over the last n hours, and a short moving average and standard deviation of sensors' data computed in the last 3 hours. We have varied the window length with $n \in [3h, 6h, 12h, 24h, 36h, 48h, 72h]$, where n cor-

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/.

responds to the lagging window size (LW). Due to lack of space, in this study we only show the results for $n = 24h$ and $n = 48h$.

The classification of the different types of failure is performed using two different approaches. The first approach, referred to as Random Forest single failure (hereafter, RF-s), replicates the methodology originally described in [6], where the authors base their analysis on the assumption that only one failure can occur per timestep. In this approach, the classification method Random Forest is used to assign a label to each timestep. As in [6], RF-s is validated via a grid searching algorithm with a number of estimators sets to 70. Given that in the considered dataset, 11% of the total timesteps impacted by a failure are characterised by two distinct types of failure, this approach suffers from an information loss problem. Thus, in this study we have developed a second alternative method based on the classification of multiple failures per timestep. The classification of multiple failures per timestep is accomplished with the following three different approaches: i) Multi Layers Perceptron (hereafter, MLP) with hidden layers size of 500 and optimised with *adam*, ii) Logistic Regression (hereafter, LR) with *newton cg* solver, and iii) Random Forest multi failure (hereafter, RF-m) with a number of estimators sets to 70. With LR and MLP, each timestep is labelled with a 5 bits vector to represent all the possible states of the components. The first bit indicates with a 0 if there is a failure and with a 1 if the timestep is normal to avoid confusion between a normal timestep and a classification error where no label has been assigned. The following bits signal the state for each component, with 0 indicating no failure, and 1 indicating a failure for each respective component. In this way, For example, if at timestep t there is no failure the $label_t$ is $[1, 0, 0, 0, 0]$. If instead there are failures for component 1 and component 3, $label_t$ is $[0, 1, 0, 1, 0]$. With RF-m, for each timestep of each component the method outputs 1 if there is a failure, 0 otherwise. Finally, for every training case, each data related to sensors have been $z-$normalized following this equation:

$$Z = \frac{x - \mu}{\sigma}; \tag{1}$$

with μ indicating the mean, σ the standard deviation and x the value before normalisation. The labels distribution follows the same behaviour as the lagging features, and each label is mirrored in the n previous timesteps. The label distribution is detailed in Table 1.

	N	C1	C2	C3	C4
Distribution	93.85%	2.31%	1.62%	1.29%	0.94%

Table 1: Label distribution in the dataset

The training set includes the first 8 months of records. The classifiers have been validated with the following 2 months. Finally, the best classifier has been tested during the last 2 months. Regarding the evaluation metrics, the classifiers have been evaluated to determine the number of true positive (TP), false negative (FN) and false positive (FP). The main evaluation metrics are $F1$

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/.

score and WS_{label}, computed in the following;

$$\begin{aligned} F1 &= 2 \times \frac{P \times R}{(P+R)}; \quad WS_{label} = F1 * \frac{O_{label}}{100}; \\ P &= \frac{TP}{(TP+FP)}; \quad R = \frac{TP}{(TP+FN)}; \end{aligned} \tag{2}$$

with P indicating precision, and R the recall. O refers to the labels percentage of occurrence in the dataset which is described in Table 1. $F1$ refers to the F1 score giving an weighted accuracy score with respect to precision and recall. WS is the weighted score added to evaluate the classification accuracy for a label according to its weight inside the dataset.

3 Results

	LW	N	C1	C2	C3	C4	WS
RF-s	24h	0.9983	0.8909	0.9306	0.9056	0.8985	0.9927
	48h	0.9985	0.9126	0.9187	**0.9113**	0.9235	0.9935
MLP	24h	0.9971	0.814	0.8741	0.8231	0.839	0.9872
	48h	0.9964	0.7744	0.8318	0.7581	0.7703	0.9835
LR	24h	0.9938	0.7812	0.8301	0.8489	0.8369	0.983
	48h	0.9886	0.5737	0.6933	0.7296	0.718	0.9684
RF-m	24h	**0.9987**	**0.9235**	**0.9354**	0.9003	**0.9574**	**0.9944**
	48h	0.9985	0.9179	0.9327	0.8904	0.9458	0.9938

Table 2: Classification F1 score across classifiers and LW

The classification F1 scores for every classifier for $LW = 24h$ and $LW = 48h$ are detailed in Table 2. Regarding the classification accuracy by labels, all the classifiers detect nearly perfectly the timesteps without failure. This score is expected since failures rarely occurred in the dataset. Regarding the failure classification, MLP and LR classify the failures with less accuracy while RF-s and RF-m achieve the best classification results for all the types of failures, confirming the trend observed in [9]. LR clearly distinguishes the difference between a failure and a normal timestep but failed to classify the correct failure name. The LR scores are particularly low to classify failures due to an higher sensibility to the label imbalanced. Regarding the LW, $n = 48h$ turns out to be the best length to classify one failure per timestep. Regardless of the classifier, $n = 24h$ is the best LW to classify simultaneous failures. It is interesting to note that classifying the failures earlier with a LW set to $n = 48h$ does not significantly lower the classification scores and this for all the classifiers. However, the scores with a LW for $n = 24h$ are slightly higher than those with a LW for $n = 48h$ since the degradation of the machine is higher. This degradation is highlighted in the features since the failure probability increase. Also, the classification scores for

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/.

RF-m also demonstrates that it is slightly better to classify simultaneous failures than single failure since the WS for RF-m is higher than the score achieved by RF-s.

Label	P	R	F1
N	**0.9982**	**0.9992**	**0.9987**
C1	0.9437	0.9042	0.9235
C2	0.9674	0.9056	0.9354
C3	0.9211	0.8805	0.9003
C4	0.9783	0.9375	0.9574

Table 3: Validation scores of RF-m with LW sets to 24h

The classification scores for the best classifier are detailed in Table 3. RF-m generally achieved very high P and R scores for every failures. This demonstrates the robustness and the reliability of the classification with a low false positive and negative rates. Also, the RF-m classifies, with a very high score, timesteps without failure from timestep with failures. This result is especially interesting in PdM context where the false negatives can have a lot of impact in leading to service interruption.

Finally, the RF-m classification has been tested on the testing dataset containing only timesteps with multiple failures. The aim is to precisely evaluate its ability to detect simultaneous failures and to weight the number of FP and TP in the special case of simultaneous failures. As shown in Table 4, RF-m can classify correctly all failures with a very high level of precision. This results is even more important in PdM case where FP leads to unnecessary components replacement. However, RF-m is prone to generate FN even if the R score is still adequate with a lowest value of 0.88.

Label	P	R	F1
C1	1	0.937	0.968
C2	1	0.877	0.935
C3	1	0.875	0.933
C4	1	0.88	0.9361

Table 4: Validation of RF-m with the testing set

4 Conclusions

In conclusion, our approach demonstrates that is possible to detect all the possible simulatneous failures that could happen inside a same measurement interval without losing any information and without any classification performance loss. Also, the failures can be classified up to 48h before their occurrence, leaving time for maintenance operators and decision makers to plan the necessary operations to maintain their system at his best level.

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/.

5 Acknowledgments

The authors acknowledge the SPW-recherche (Walloon Region) and Telespazio Belgium throughout the research funding WIN4DOC. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

References

[1] Robert King and Kevin Curran. Predictive Maintenance for Vibration-Related failures in the Semi-Conductor Industry. *Journal of Computer Engineering & Information Technology,*, 8(1):1, April 2019.

[2] Ameeth Kanawaday and Aditya Sane. Machine learning for predictive maintenance of industrial machines using IoT sensor data. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 87–90, November 2017. ISSN: 2327-0594.

[3] Leo Breiman. Machine Learning, Volume 45, Number 1 - SpringerLink. *Machine Learning*, 45:5–32, October 2001.

[4] Yuxin Wen, Md Rahman, Honglun Xu, and Bill Tseng. Recent Advances and Trends of Predictive Maintenance from Data-driven Machine Prognostics Perspective. *Measurement*, 187:110276, October 2021.

[5] Thyago Carvalho, Fabrizzio Soares, Roberto Vita, Roberto Francisco, João Basto, and Symone G. Soares Alcalá. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137:106024, September 2019.

[6] Diogo Cardoso and Luis Ferreira. Application of Predictive Maintenance Concepts Using Artificial Intelligence Tools. *Applied Sciences*, 11:18, December 2020.

[7] Juan C. Quiroz, Norman Mariun, Mohammad Rezazadeh Mehrjou, Mahdi Izadi, Norhisam Misron, and Mohd Amran Mohd Radzi. Fault detection of broken rotor bar in LS-PMSM using random forests. *Measurement*, 116:273–280, February 2018.

[8] Predictive maintenance modelling guide data sets. https://gallery.azure.ai/Experiment/Predictive-Maintenance-Modelling-Guide-Data-Sets-1. [Online; accessed 14-April-2023].

[9] Abdelfetah Ouadah, Leila Zemmouchi-Ghomari, and Nedjma Salhi. Selecting an appropriate supervised machine learning algorithm for predictive maintenance. *The International Journal of Advanced Manufacturing Technology*, 119, April 2022.