A Deep Learning Approach for Unsupervised Failure Detection in Smart Industry

(Discussion Paper)

Angelica **Liguori**[1], Giuseppe **Manco**[2], Ettore **Ritacco**[2], Massimilano **Ruffolo**[3] and Salvatore **Iiritano**[3]

[1]*Department of Computer Engineering, Modeling, Electronics and, Systems University of Calabria*
[2]*Institute for High Performance Computing and Networking of Italian National Research Council*
[3]*Revelis S.r.l.*

Abstract
We propose an unsupervised anomaly detection model that is able to identify abnormal behavior by analysing streaming data coming from IoT sensors installed on critical devices. The proposed model is based on a Siamese neural network which embeds time series windows in a latent space, thus generating distance-based clusters of normal behavior. We experiment the proposed model on a case study aimed at the predictive maintenance of elevators where specific sensors measure the oscillations of the lift during its daily use. The experiments show that the proposed model successfully isolates anomalous oscillations thus correlating them to prospective malfunctions and thus preventing possible faults.

Keywords
Anomaly detection, Failure detection, Fault detection, Time-series analysis, Embeddings, Siamese networks.

1. Introduction

Maintenance is one of the most important activities in support of all industrial production systems. It represents both an opportunity and a bottleneck at the same time. In fact, if on the one hand it makes it possible to avoid the occurrence of machinery breakdowns, service interruptions and possibly monetary penalties, on the other hand it risks slowing down industrial processes, since it requires starting activities not strictly correlated with the core business, committing the production machinery and numerous resources in terms of technicians, tools, time, possible costs.

Fault detection and prevention [1] is one of the most critical part of predictive maintenance, as it aims at identifying an anomalous behavior of the system and avoid sudden interruptions and catastrophic machine failures. Typically, the focus is on developing solutions that produce warnings when an anomalous behavior is detected and fault detection methods based on machine learning aim at analyzing historical data in order to devise models of failure or malfunction, based on supervised learning. In this respect, the predictive maintenance process has largely

benefited from the large-scale adoption of sensor devices. With the advent of the IoT technology, numerous sensors can be installed on production devices, which produce significant amounts of data based on their sampling frequency.

Despite their intrinsic predictive value, these enormous streams of data represent a challenge in several respects. First of all, sensor streams are very noisy temporal sequences, with excessive dimensionality and affected by specific issues (e.g. burst effects, seasonality). It is necessary to devise automatic methods capable of filtering out noise and produce effective representations that can be fruitfully exploited in predictive models. Moreover, often the presence of streaming data requires also the capability to meet real-time requirements. Maintenance solutions must be able to provide timely warnings to maintenance experts.

However, in a vast majority of industrial situations, there are no samples that represent the presence of failure, hence, the adoption of unsupervised methods seems better suited. The choice between supervised and unsupervised approaches does not depend only on the presence or absence of the fault indicator. There are numerous scenarios in which, despite the presence of such a flag, the classification cannot succeed in a quality prediction. Failures are rare events and are often not sufficient to define the prediction patterns that are necessary for any classification technique.

In this paper, we study a specific scenario where the above mentioned issues take place. The paper is focused on a study where the adoption of sensing technology, combined with machine learning, can effectively characterize the behavior of lifting systems and thus devise effective predictive maintenance strategies aimed at ensuring the stability and efficiency of the elevator, as well as preventing future breakdowns. The intuition is that, in their daily routine, elevators produce oscillations which can be registered and analyzed. A methodology for devising profiles of normal oscillations can hence help us detect any deviation from typical signatures and hence prospective efficiency issues.

2. Related Work

In the maintenance field, traditional unsupervised approaches use one-class classifier, e.g. One-class Support Vector Machines [2, 3] or distance-metrics, such as e.g. Isolation Forest [4, 5]. Autoencoders [6] are the core of Unsupervised Anomaly Detection methods based on Deep Learning and most recent literature on unsupervised failure detection is based on them, such as [7, 8, 9] in which the reconstruction error is used as anomaly score and, their variants, such as [10, 11] that propose an unsupervised fault detection based on a stacked autoencoder and on a sparse autoencoder, respectively. Autoencoders can also be adapted to cope with streaming data. Lindemann et al. [12] propose an anomaly detection system by combining an autoencoder with Long short-term memory (LSTM) [13]. By contrast, Jiang et al. [14] propose an unsupervised fault detection based on denoising autoencoder (DAE). Xiang et al. [15] propose a framework based on Variational Autoencoder (VAE) [16] in which the Gated Recurrent Network (GRU) [17] network unit is introduced into VAE network to replace the traditional neural unit in encoder and decoder. Alfeo et al. [18] combines an autoencoder with a heuristic-based discriminator in order to improve the interpretability of the detection. Jian and Zhiyan [19] propose an unsupervised fault detection method based on adversarial auto-encoders in which

a discriminator is able to refine the reconstruction errors and hence. In [20] a one-class fault detection based on unsupervised training of Generative Adversarial Networks [21] is proposed, where the problem of distinguish real from fake data is converted in distinguish normal from anomalous data, by forcing the generator in producing normal-like generated data.

Most of the unsupervised failure detection systems in the literature exploit the reconstruction error to discriminate between normal data and anomalies. Usually, when the reconstruction error is used as measure of outlierness, a threshold is defined such that the data whose reconstruction error is above the threshold is marked as outlier. Defining a threshold is very hard, especially when there is no knowledge background.

Unlike these works, our proposal exploits the philosophy of the Siamese networks [22] to map the data into a latent space so that data that belong to the same category are located in the same area with respect that data that belong to different category. This idea mitigates the overfitting problem of the cited approaches: sequences are analyzed according to a collective approach instead of comparing any rebuilt sequence with only its original version.

3. Setting

Let $\mathscr{D}$ be a set of devices characterized by functionality, structure, purpose and/or working environment. Each device $d \in \mathscr{D}$ is equipped with a set sensors $\mathscr{M}_d$ emitting temporal sequences of events $S_d = \{e_d^{(1)}, e_d^{(2)}, e_d^{(3)}, ...\}$, where the superscript indicates the time step. Each event $e_d^{(t)}$ is a real vector of size $|\mathscr{M}_d|$ containing the values observed by the sensors. We further assume that S_d is labeled by a specific category which characterizes the process being monitored. Categories can also refer to the underlying device (i.e., each device can be seen as a distinct category); however, they can also describe the situations that the sensors are measuring, such as an elevator moving up with a load of two persons.

The objective of our research is to detect anomalous situations within a sequence S_d associated with a device. The basic approach consists in building a machine learning model which is capable of characterize the profile of each sequence, and mark anomalies whenever a given sequence does not fit that profile. However, in order to apply such a methodology, we need to face two main problems. First of all, there are several different profiles that can characterize each sequence. The categories encode different situations and we can expect that events marked with a specific category are different from events labeled with a different one. We can hence assume that the expected number of categories is high. Furthermore, we assume that $|\mathscr{M}_d|$ is large but the overall number of sequences is small compared to the number of categories. That is, we can expect that each sequence is labeled with almost a different category. This clearly poses a problem in the learning stage, since it is not possible to build specific profiles due to the lack of sufficient training data for each category.

The solution we propose is a methodology divided into two parts. The first part consists in a data transformation approach that allows the definition of a classification problem. The second part consists in the exploitation of a modular Siamese network able to map the input (sequence fragments) into data points lying on a latent space, by solving the aforementioned classification problem.

The latent space has a geometric interpretation: points that are close in the space correspond
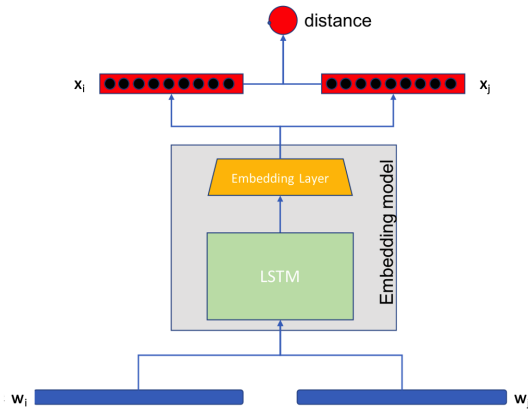
Figure 1: Network architecture.

to devices that exhibit, in some time interval of their working process, similar behavior. The core concept is that clusters of these latent data points represent the different working modes of the target devices. Any element that is on the edge of a cluster or out of all of them can be highlighted to maintenance experts for further investigation. This allows to approach situation, typical in many industrial processes, where critical anomalies (e.g. failures) are extremely rare.

4. Methodology

Since sequences may have different sizes, our methodology applies a sliding window extraction procedure to generate, for each sequence, a set of fixed-size observation windows. Notice that if the window shift is lower than the window size, contiguous windows partially overlap. Each window is a time frame that partially describes the behavior of the target devices during the interval of observation. For the rest of the paper we will name $\mathcal{W} = \{w_1, w_2, ...\}$ the set of all the time windows and we will use the function $cat(w_i)$ for denoting the category relative to the device that generated the i-th time window.

The set $\mathcal{W}$ and the function cat enable the definition of the following classification problem: given two arbitrary time windows w_i and w_j, the goal is to predict if they belong to the same category, i.e. whether $cat(w_i) = cat(w_j)$.

The solution we propose to address this problem is a Siamese neural network, shown in Figure 1, composed by two modules. The first one is the *Embedding model* that maps a time window into the high-dimensional latent space. The Embedding subnet is a sequential model composed by a recurrent neural network (we exploited a LSTM layer [13]), for catching the time dependencies within each window, and a feed-forward neural network (a dense embedding layer), that generates the data points in the latent space. The second module is the *Distance subnet* that outputs the euclidean distance between two embeddings.

The working flow of the whole architecture is described as follows. The input of the network is a pair of time windows, w_i and w_j, that are randomly sampled from $\mathcal{W}$. The set of all the built-up pairs, called $\mathcal{W}_s$, suitably represent each category, providing a sufficient number of
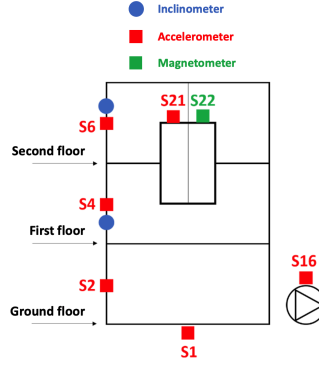
Figure 2: Graphical representation of the elevator. All sensors are set with a sampling frequency of 200Hz except S16 (4000Hz) and S22 (100Hz)

positive (windows belonging to the same category) and negative (windows belonging to different categories) comparisons. Both w_i and w_j pass through the Embedding subnet, the Siamese part of the network, that will produce two embeddings, respectively x_i and x_j, that will feed the Distance subnet that will computes and returns their euclidean distance. The loss function we chose is the following:

$$loss = \frac{1}{|\mathscr{W}_s|} \sum_{w_i, w_j \in \mathscr{W}_s} y_{i,j} \cdot f(w_i, w_j) - (1 - y_{i,j}) \cdot \log\left(1 - e^{-f(w_i w_j)}\right) \tag{1}$$

where $y_{i,j}$ is equal to 1 if $cat(w_i) = cat(w_j)$, 0 otherwise, and $f(w_i, w_j)$ is a function that computes the euclidean distance between the embeddings x_i and x_j of the time windows w_i and w_j, respectively.

The proposed loss function encourages the network to generate pairs of embeddings that are close in the latent space if $y_{i,j} = 1$; on the other hand, if the two categories are different, the network will produce pairs of embeddings that are distant.

5. Predicting Elevator Anomalous Oscillations

We applied our methodology to a real case study whose objective is to monitor the health status and the working process of an elevator in an office building. The sensor system placed in the elevator (see Figure 2 for details) records the movements in the x, y and z axes of the oscillations of the elevator guides and cabin, the inclination of the cabin and the magnetic field intensity in a relational data structure. Each record is a time-series collecting sensor emissions in a interval of time. Moreover, records provide further information about the operation status of the elevator, highlighting, for each time step its position, movement and door status.

In this case study, we applied our methodology on a limited set of similar elevators in an office building. Thus, their sequences were split into fixed size time windows that were randomly paired up to be processed by the Siamese network. The category labels we used were the operational modalities: (*i*) Stationary; (*ii*) Moving up; (*iii*) Moving down; (*iv*) Opening doors; (*v*)
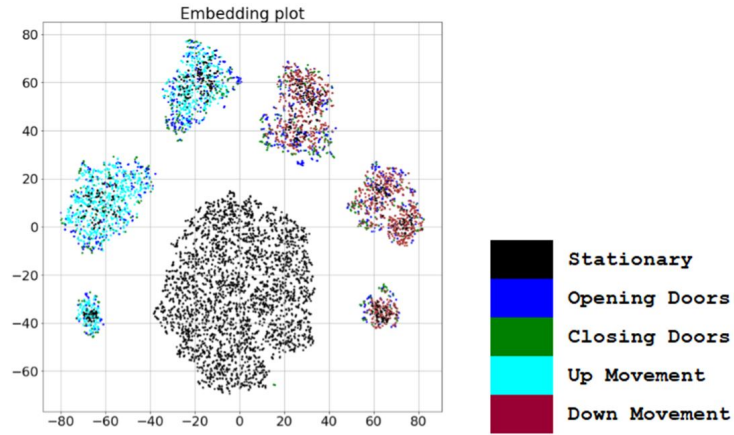
(a) Normal embeddings
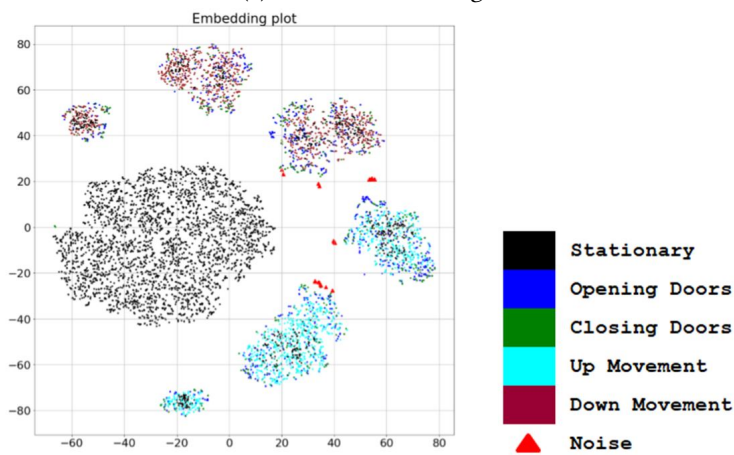
(b) Normal and anomalous embeddings

Figure 3: Embedding space

Closing doors. Each of these modalities is further labeled by contour conditions representing the load (number of people) in the lift.

The result of the training phase was that, for each elevator, the Network understood the different normal behavior models, which map into clusters of the latent space. To allow a friendly visualization of the embeddings, we used the t-distributed stochastic neighbor embedding (t-SNE) library [23]. A 2D t-SNE plot of the learnt behavior clusters is shown in Figure 3a, where each point is related to the embedding of a time window, while colors indicate the categories the windows belong to. As can be seen, the network found out 7 different behavioural clusters, in which there is a dominance of a color. The partial color overlapping is due to two factors. On one hand, category labels were noisy since they were produced by humans with external chronometers, thus, making, for each category session, the initial and final time windows imprecise. On the other hand, the sensors we used where not able to find appreciable differences in vibration when doors were opening or closing.

In order to observe the capability of the model to isolate anomalies we performed new experiments in which passengers stopped and restarted several times the elevator movements or produced (weak) unexpected vibrations. As shown in the Figure 3b, the 2d t-SNE transformation of the embeddings, provided by the network fed with these anomalous sequences, generated points, labeled as *Noise*, that are outside the clusters.

6. Conclusions

We proposed a new methodology for addressing the early detection problem of faults in critical devices that equip sensors that generate time sequences of observations. In particular, the methodology is designed to effectively work in settings where explicit information about previous failures is missing, overcoming the hindrance of the exploitation of supervised detection approaches. Assuming that failures are rare events during the life time of a device, the proposed methodology supports a maintenance expert in easily identifying them as anomalous elements that are distant from all the clusters of normal behavior.

Experiments on a real case study showed the capability of the proposal in effectively isolating anomalous time frames, suggesting that its application fields can span in many different and more complex scenarios.

Acknowledgments

References

[1] N. Amruthnath, T. Gupta, A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance, in: 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), 2018, pp. 355–361.

[2] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, J. C. Platt, et al., Support vector method for novelty detection., in: NIPS, volume 12, Citeseer, 1999, pp. 582–588.

[3] M. Amer, M. Goldstein, S. Abdennadher, Enhancing one-class support vector machines for unsupervised anomaly detection, in: Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, ODD '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 8–15.

[4] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation forest, in: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, ICDM '08, IEEE Computer Society, USA, 2008, p. 413–422.

[5] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation-based anomaly detection, ACM Trans. Knowl. Discov. Data 6 (2012).

[6] D. Bank, N. Koenigstein, R. Giryes, Autoencoders, 2020. `arXiv:2003.05991`.

[7] D. F. Oliveira, L. F. Vismari, J. R. de Almeida, P. S. Cugnasca, J. B. Camargo, E. Marreto, D. R. Doimo, L. P. F. de Almeida, R. Gripp, M. M. Neves, Evaluating unsupervised anomaly detection models to detect faults in heavy haul railway operations, in: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019, pp. 1016–1022.

[8] D. F. N. Oliveira, L. F. Vismari, A. M. Nascimento, J. R. de Almeida Jr au2, P. S. Cugnasca, J. B. C. J. au2, L. Almeida, R. Gripp, M. Neves, A new interpretable unsupervised anomaly detection method based on residual explanation, 2021. `arXiv:2103.07953`.

[9] E. Principi, D. Rossetti, S. Squartini, F. Piazza, Unsupervised electric motor fault detection by using deep autoencoders, IEEE/CAA Journal of Automatica Sinica 6 (2019) 441–451.

[10] K. H. Park, E. Park, H. K. Kim, Unsupervised fault detection on unmanned aerial vehicles: Encoding and thresholding approach, Sensors 21 (2021).

[11] X. Liang, F. Duan, I. Bennett, D. Mba, A sparse autoencoder-based unsupervised scheme for pump fault detection and isolation, Applied Sciences 10 (2020).

[12] B. Lindemann, F. Fesenmayr, N. Jazdi, M. Weyrich, Anomaly detection in discrete manufacturing using self-learning approaches, Procedia CIRP 79 (2019) 313–318.

[13] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.

[14] G. Jiang, P. Xie, H. He, J. Yan, Wind turbine fault detection using a denoising autoencoder with temporal information, IEEE/ASME Transactions on Mechatronics 23 (2018) 89–100.

[15] G. Xiang, R. Tao, Y. Peng, K. Tian, C. Qu, Unsupervised deep learning for fault detection on spacecraft using improved variational autoencoder, in: 2020 Chinese Automation Congress (CAC), 2020, pp. 5527–5531.

[16] D. P. Kingma, M. Welling, Auto-encoding variational bayes, 2014. `arXiv:1312.6114`.

[17] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014. `arXiv:1412.3555`.

[18] A. L. Alfeo, M. G. Cimino, G. Manco, E. Ritacco, G. Vaglini, Using an autoencoder in the design of an anomaly detector for smart manufacturing, Pattern Recognition Letters 136 (2020) 272–278.

[19] W. Jian, H. Zhiyan, A novel fault detection method based on adversarial auto-encoder, in: 2020 39th Chinese Control Conference (CCC), 2020, pp. 4166–4170.

[20] P. Spyridon, Y. S. Boutalis, Generative adversarial networks for unsupervised fault detection, in: 2018 European Control Conference (ECC), 2018, pp. 691–696.

[21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, MIT Press, Cambridge, MA, USA, 2014, p. 2672–2680.

[22] G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, 2015.

[23] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, Journal of Machine Learning Research 9 (2008) 2579–2605.