Exploiting Data Mining and Constraint Programming for Predictive Maintenance

TASC (LS2N-CNRS), IMT Atlantique

Supervisors: Samir Loudni & Charles Prud'Homme & Nicolas Beldiceanu

Contact:

- Charles Prud'Homme (charles.prudhomme@imt-atlantique.fr)
- Nicolas Beldiceanu (Nicolas.Beldiceanu@imt-atlantique.fr)
- Samir Loudni (samir.loudni@imt-atlantique.fr)

keywords: predictive maintenance, data mining, constraint programming, anomaly detection.

Context: Recently, with the emergence of Industry 4.0 (I4.0), predictive maintenance (PdM) based on data-driven methods has become the most effective solution to address smart manufacturing and industrial big data, especially for performing health perception (e.g. fault diagnosis and remaining useful life (RUL) estimation) [7, 18]. Here, maintenance corresponds to the process that deals with equipment or system components to ensure their normal operating under any circumstance. PdM relies on the continuous monitoring of the equipment or the machine to predict when maintenance actions are necessary; hence the maintenance can be scheduled [22]. Detecting and preventing failures is thus essential, and industries seek to minimise the number of operational failures, minimise their operational costs, and increase their productivity.

The general workflow of data-driven PdM system consists of the following three subprocesses:

1. Data acquisition and pre-processing. The first step is to collect a large set of sensor data representing **healthy** and **faulty operation**. These raw data are then preprocessed to bring it to a form from which **features** can be easily extracted.
2. Features extraction. The next step is to identify features that help distinguish healthy conditions from faulty, on the basis of historical sensor data.
3. Model training and predicting. The last step consists in using the extracted features to train a machine learning model that can (i) detect anomalies; (ii) classify different types of faults; or (iii) estimate the remaining useful life (RUL) of equipment.

The increased availability of large volumes of operational data, collected from various sensors over time, has paved the way for the development and deployment of the data-driven PdM, which utilises prediction tools to provide valuable information regarding the status of equipment [23]. According to [14, 15], the most common data collected from sensors are vibration, temperature and electrical signal. However, performing predictive maintenance in such continuously changing temporal data (aka **time series data**) has become a major challenge which received increasing attention from both the industry and the scientific community.

Failure Prediction is one of the critical components of PdM for which the main goal is to predict the approximate moment when some failure could occur. Recent works have addressed anomaly detection for PdM in order to predict incipient failures from historical data [10, 19]. In the context of condition monitoring, this is interesting because anomalies can tell us something about the "health state" of the monitored equipment [6]: Data generated when the equipment approaches failure, or a suboptimal operation, typically have a different distribution than data from "healthy" equipment. Machine learning techniques (ML) emerged as a promising tool to achieve this goal. However, the current use of ML is mainly focused on supervised learning, which means that data sets need to be labelled, that is the data must be annotated with the true machine health condition. Accordingly, **learning with unlabelled data**, namely unsupervised learning, is the focus of this internship.

In the last decade, new research have began connecting data mining to symbolic Artificial Intelligence (AI). Such fertilization leads to a number of algorithms that have been proposed within Constraints Programming (CP) and Satisfiability (SAT) for mining sequences [1, 16], frequent itemsets [11, 17, 20], association rules [2, 12], clustering [5, 8], classification [9, 21], etc. The main advantage of symbolic AI approaches for pattern mining is their declarativity and flexibility, which include the ability to incorporate new user-specified constraints without the need to modify the underlying system. Within CP, existing work on extracting quantitative pattern in the context of time series constraints was initialy done in [3]. The work was recently adapted to the context of sliding time-windows [4]. To the best of our knowledge, there exists no work that exploits CP for PdM.

Objective: The objective of this internship is to use constraint programming to apply symbolic data mining techniques on historical data to characterise the healthy behaviour of equipment. We will consider especially symbolic data mining techniques applicable to time series data where data are generated in streams. The extracted patterns will then serve to form a knowledge base to detect abnormal or uncertain behaviour in new data. To detect anomalies in newly monitoring data, concordance and discordance metrics will be defined and exploited to compute an anomaly score on the basis of the previously obtained knowledge.

The internship will address the two following principal tasks:

- Knowledge discovery process about normal behaviour;
- The anomaly detection in new data.

Candidate : We are looking for a motivated Engineering School or Master's degree candidate in Computer Science who is motivated by constraint programming and machine learning fields. Good programming abilities will be required.

References

[1] J. O. R. Aoga, T. Guns, and P. Schaus. An efficient algorithm for mining frequent sequence with constraint programming. In Paolo Frasconi, Niels Landwehr, Giuseppe Manco, and Jilles Vreeken, editors, *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II*, volume 9852 of *Lecture Notes in Computer Science*, pages 315–330. Springer, 2016.

[2] M-B. Belaid, C. Bessiere, and N. Lazaar. Constraint programming for association rules. In Tanya Y. Berger-Wolf and Nitesh V. Chawla, editors, *Proceedings of the 2019 SIAM International Conference on Data Mining, SDM 2019, Calgary, Alberta, Canada, May 2-4, 2019*, pages 127–135. SIAM, 2019.

[3] Nicolas Beldiceanu, Mats Carlsson, Rémi Douence, and Helmut Simonis. Using finite transducers for describing and synthesising structural time-series constraints. *Constraints An Int. J.*, 21(1):22–40, 2016.

[4] Nicolas Beldiceanu, Mats Carlsson, Claude-Guy Quimper, and Maria-Isabel Restrepo-Ruiz. Classifying pattern and feature properties to get a $\Theta(n)$ checker and reformulation for sliding time-series constraints. *CoRR*, abs/1912.01532, 2019.

[5] Abdelhamid Boudane, Saïd Jabbour, Lakhdar Sais, and Yakoub Salhi. Clustering complex data represented as propositional formulas. In Jinho Kim, Kyuseok Shim, Longbing Cao, Jae-Gil Lee, Xuemin Lin, and Yang-Sae Moon, editors, *Advances in Knowledge Discovery and Data Mining - 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part II*, volume 10235 of *Lecture Notes in Computer Science*, pages 441–452, 2017.

[6] Ece Calikus, Sławomir Nowaczyk, Anita Sant'Anna, and Onur Dikmen. No free lunch but a cheaper supper: A general framework for streaming anomaly detection. *Expert Systems with Applications*, 155:113453, 2020.

[7] Xuewu Dai and Zhiwei Gao. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, 9(4):2226–2238, 2013.

[8] Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Constrained clustering by constraint programming. *Artif. Intell.*, 244:70–94, 2017.

[9] Emir Demirovic and Peter J. Stuckey. Optimal decision trees for nonlinear metrics. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 3733–3741. AAAI Press, 2021.

[10] Kilian Hendrickx, Wannes Meert, Yves Mollet, Johan Gyselinck, Bram Cornelis, Konstantinos Gryllias, and Jesse Davis. A general anomaly detection framework for fleet-based condition monitoring of machines. *Mechanical Systems and Signal Processing*, 139:106585, 2020.

[11] A. Hien, S. Loudni, N. Aribi, Y. Lebbah, M. El Amine Laghzaoui, A. Ouali, and A. Zimmermann. A relaxation-based approach for mining diverse closed patterns. In Frank Hutter, Kristian Kersting, Jefrey Lijffijt, and Isabel Valera, editors, *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part I*, volume 12457 of *Lecture Notes in Computer Science*, pages 36–54. Springer, 2020.

[12] Yacine Izza, Saïd Jabbour, Badran Raddaoui, and Abdelhamid Boudane. On the enumeration of association rules: A decomposition-based approach. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1265–1271. ijcai.org, 2020.

[13] Mikolás Janota and António Morgado. Sat-based encodings for optimal decision trees with explicit paths. In Luca Pulina and Martina Seidl, editors, *Theory and Applications of Satisfiability Testing - SAT 2020 - 23rd International Conference, Alghero, Italy, July 3-10, 2020, Proceedings*, volume 12178 of *Lecture Notes in Computer Science*, pages 501–518. Springer, 2020.

[14] Bai Jie, Liqun Hou, and Ma Yongguang. Machine fault diagnosis using industrial wireless sensor networks and support vector machine. In *2015 12th IEEE International Conference on Electronic Measurement Instruments (ICEMI)*, volume 01, pages 153–158, 2015.

[15] Chao Jin, Wenyu Zhao, Zongchang Liu, Jay Lee, and Xiao He. A vibration-based approach for diesel engine fault diagnosis. In *2014 International Conference on Prognostics and Health Management*, pages 1–9, 2014.

[16] A. Kemmar, Y. Lebbah, S. Loudni, P. Boizumault, and T. Charnois. Prefix-projection global constraint and top-k approach for sequential pattern mining. *Constraints An Int. J.*, 22(2):265–306, 2017.

[17] N. Lazaar, Y. Lebbah, S. Loudni, M. Maamar, V. Lemière, C. Bessiere, and P. Boizumault. A global constraint for closed frequent pattern mining. In *Proceedings of the 22nd CP*, pages 333–349, 2016.

[18] Jay Lee, Edzel Lapira, Behrad Bagheri, and Hung an Kao. Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, 1(1):38–41, 2013.

[19] Afrooz Purarjomandlangrudi, Amir Hossein Ghapanchi, and Mohammad Esmalifalak. A data mining approach for fault diagnosis: An application of anomaly detection algorithm. *Measurement*, 55:343–352, 2014.

[20] P. Schaus, J. O. R. Aoga, and T. Guns. Coversize: A global constraint for frequency-based itemset mining. In *Proceedings of the 23rd CP 2017*, pages 529–546, 2017.

[21] Hélène Verhaeghe, Siegfried Nijssen, Gilles Pesant, Claude-Guy Quimper, and Pierre Schaus. Learning optimal decision trees using constraint programming. *Constraints An Int. J.*, 25(3-4):226–250, 2020.

[22] Weiting Zhang, Dong Yang, and Hongchao Wang. Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal*, 13(3):2213–2227, 2019.

[23] Zeki Murat Çınar, Abubakar Abdussalam Nuhu, Qasim Zeeshan, Orhan Korhan, Mohammed Asmael, and Babak Safaei. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*, 12(19), 2020.