Improving the development and reusability of Industrial AI through Semantic Models

Giovanna Martínez-Arellano[1][0000-0003-3105-4151] and Svetan Ratchev[1][0000-0001-9955-2806]

[1] Institute for Advanced Manufacturing, University of Nottingham, UK

giovanna.martinezarellano@nottingham.ac.uk

Abstract. Despite some of the success of AI, particularly machine learning, in industrial applications such as condition monitoring, quality inspection and asset control, solutions are typically bespoke and not robust in the long term. There is a considerable amount of effort in developing these solutions to deliver accurate results within a very limited scenario. In addition, the operationalisation of these models in the factory floor is a challenge. Developing and maintaining these models requires of data science expert knowledge and the digital skills gap in the manufacturing industry is a major barrier. A step towards the development of AI skills can be facilitated in a Learning Factory environment, provided there is a way for operators to develop an understanding of how manufacturing problems can be addressed with different data science tools. To address this, this paper introduces a semantic based framework as one of the key elements to facilitate the development of Industrial AI solutions. By formalising the way data, manufacturing processes and AI models are described and linked before and after the creation of a solution, it is possible not only to automate model creation, but to enable reusability and management. A preliminary conceptualisation on the use of this framework through a process monitoring scenario is presented. By capturing the semantic relationships, it is possible to support a more automated machine learning pipeline, enable manufacturers to understand how solutions can be created and learn how they can then be reused in future similar scenarios.

Keywords: Machine learning, MLOps, semantic models, pipeline, scalability, Industry4.0

1 Introduction

Artificial Intelligence (AI) is expected to be one of the most disruptive digital technologies of this fourth industrial revolution (Industry 4.0) [1]. From shop floor applications at the asset level to smart supply chain management, over the last decade research developments have shown the potential of this technology at all levels of the value chain. However, most of these developments have low readiness levels. There is a lack of production ready AI solutions demonstrated in real industrial scenarios.

There are multiple reasons why AI, particularly Machine Learning (ML), industrial solutions are unscalable in their current form. One of the main challenges is access to good quality data. Data is typically scattered across multiple systems such as PLM and

ERP systems and the shopfloor equipment itself, which may be equipped with different sensors, making data inconsistent across multiple equipment. Lack of adequate data management is one the of main reasons for the failure of AI projects [2].

Another factor that hinders scalability is the expertise needed to develop such solutions. Most manufacturing companies do not have the inhouse expertise. Data scientists, which are not experts in the application domain, need to understand what the manufacturing challenge is, to then define what data, approach and deployment environment is suitable. Understanding the manufacturing problem and how it translates into a definition of the machine learning problem (i.e., classification or a regression challenge) is key to the success of the solution [3]. The same applies for curating the data, performing feature extraction and selection. Depending on the industrial application, data resolution needed may change and this affects how data preparation and feature engineering is performed [4]. This understanding and knowledge could be captured systematically to ease model development in future applications. Having an approach to capture and combine knowledge from both fields could support better the skills development of operators, as it can enable them to understand which data science tools, data and models are available to them for use and reuse.

Furthermore, there is a major limitation of machine learning approaches for dealing with changes in data distribution (i.e. concept drift). Two machines carrying out the same operation and using the same sensors can render completely different patterns, due to the nature of the machines themselves, requiring different trained models [5]. Concept drift might occur due to natural wear, requiring continuous model retraining. A robot may carry out different tasks as new end effectors are added. In this scenario, the model deployed needs to incrementally learn new patterns without forgetting previous ones. Continuous monitoring and understanding on how and when to retrain the model (transfer or continual learning) is needed. Once again, key context information on the process and on how models are developed is key to model reusability and robustness. Given these challenges, this paper presents preliminary work on the use of semantic modelling to facilitate the development and reuse of ML industrial solutions. This will be a step towards providing manufacturing operators with the tools to learn and easily develop these solutions without the need of a data scientist. The rest of the paper is organised as follows: Section 2 presents some of the existing work on facilitating the development and management of industrial AI solutions. Section 3 introduces the proposed approach, providing details on the different semantic models and how they are used to facilitate machine learning model development and reuse. Section 4 presents a scenario to demonstrate the use of the approach and finally conclusions and future work are presented in Section 5.

2 Related Work

2.1 Model development pipelines in industrial applications

There have been some recent efforts in the research community on the development of frameworks to facilitate shopfloor operators in developing and managing industrial models and data. From the model development point of view, there are some recent works on the development of ML pipelines that can speed up and guide the un-experienced user on the development of such models. For example, Zhou et al. present a framework for the development of Industrial ML solutions through four different pre-defined ML pipelines [6]. Using semantic models, the framework can identify the most appropriate pipeline to use and so it can automate the development through a predefined set of preprocessing steps which include statistical feature extraction, selection, and training two or three pre-defined models depending on the end application. One of the main limitations, in addition having a limited set of predefined models and architectures, is that it relies on the availability of an initial well bounded data set expected to be a good representation of all possible future scenarios. In practice this is almost never the case for dynamic environments. Even when data is coming from the same machine carrying out the same operation (*concept drift*), and so it is not possible to know a priori the bounds of such data set. In this context, it is not practical to assume data for that machine is available, instead it is important to know if there is another model from another machine that has been trained in a similar scenario and use that as a starting point or employ techniques that work better with limited data sets before the model can then be updated with a more robust one once more data is available.

2.2 Model lifecycle management in industrial environments

There have been several technologies that have been proposed to support the operationalisation of Industrial AI. With the success of continuous software engineering practices (DevOps), there has been an increasing interest in the rapid deployment of machine learning models, referred as MLOps [7]. General architectures based on MLOps have been proposed for industrial contexts. Zhao for example, proposes an MLOps architecture for industrial settings integrating technologies such as DVC for data and model versioning, AWS for data storage, MLFlow for pipeline implementation [8]. Raffin et al propose a cloud edge architecture for the operationalisation of machine learning models in manufacturing shopfloors [9]. The authors highlight the importance of the correct management of complex data as this enables successful improvement of machine learning models in the long term. Although the main architectural elements are proposed, it is not clear how data, models and processes can all be linked to ensure an effective reusability or further training of ML models. Elements such as meta data, code repository and feature store can really help to shorten the transition of models from prototype to production [10].

The Asset Administration Shell (AAS) has been recently proposed to lifecycle manage AI solutions. As proposed by Rauh et al, AI solutions can be described as an AI artifact or instance with three life cycle stages: before training to a data set, after training, and during runtime [2]. After training, the model instance is frozen. As a result, a simple hierarchical structure is used to capture all the relevant changes of the AI model. The paper, however, does not consider the monitoring of the model's performance after deployment, which is key for the continuous update and re-deployment of models.

2.3 Model reusability

Availability of large, good quality data sets is another of the great challenges in manufacturing. To deal with this, some of the techniques that are starting to get attention are transfer learning and generative AI. Wang and Gao explore the use of foundation models for transfer learning in machine condition monitoring. The authors stress the need to further understand the boundaries for effective model transferability, particularly when using pre-trained models with non-manufacturing-specific data [11]. Giannetti and Essien demonstrate the effectiveness of transfer learning for predicting operational parameters of can body maker machines [12]. The authors propose the F-score measure to select the model to use for transfer learning from a set of models from similar machines. Their main finding was that the effectiveness of the training strategy depends on the data, so contextual information of the process to understand the differences on the data is essential.

Across all these streams of research, there is a common limitation, and that is the lack of context and knowledge capture; (1) from the manufacturing process which determines the data characteristics that can support better model transfer learning, (2) from the operator that understands the process and provides an important guidance to the solution design, (3) from the decisions made during the model development process, and finally (4) from the model itself, capturing monitoring metrics to support ML model re-deployment. To tackle this major challenge, this paper integrates some of the already available semantic models to support the complete life cycle of the ML solution in a more automated and effective way. By using an ontology model, multiple elements that are involved in the development of ML solutions can be defined and standardised; from Asset and Process to capture *manufacturing knowledge*, to Data, Algorithm, and Model, to capture intrinsic knowledge related to *model development.* The ontology can then be used to support the operator to draw answers (i.e. inference) to questions such as which data, algorithm to use, or which model to reuse when a new model needs to be developed, and supporting this way the operator to develop its AI and data science skills. Semantic models are a scalable option when dealing with the complexity of the manufacturing environment and once defined, they can be used to instantiate any manufacturing environment and industrial solution.

3 Proposed Framework

To address the challenges discussed in Section 2, we propose a semantic framework which becomes the backbone for the integration of knowledge across the lifecycle of an industrial AI solution. This allows to connect process to AI solution, which is important for the monitoring as well as reusability of the model as it will be explained later. The proposed framework is not relying on a predefined architecture in terms of edge and cloud technologies, it is in this sense independent to how it is implemented.

3.1 Semantic Models

Implemented in Protégé, this framework expands on the work from Järvenpää et al. on the Process and Resource Models [13], Sensors and Samples from Janowicz et al. [14], Algorithms and Application models from Braga et al. [15], and introduces the models Manufacturing Application, AI Pipeline and Data processing. A general view and interconnections of the eight models that comprise the framework are presented in Figure 1. Each of the models play a key role in the life cycle of a solution:

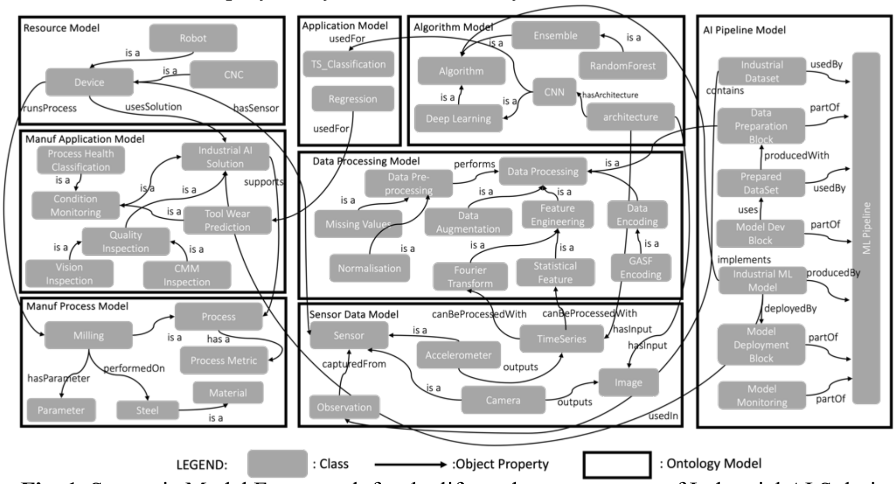

Fig. 1. Semantic Model Framework for the life cycle management of Industrial AI Solutions

- *Resource Model* – captures all data related to an asset (e.g. robot, CNC machine). The asset is linked to a process that is or was carried out and to the sensors on it. From this, the data related to an and process can be inferred. This helps identify candidate ML models for transfer learning or for its data to train other models. This is particularly useful when data is limited.
- *Manufacturing Process Model* – defines common processes that are needed to deliver a product, and which are linked to the capabilities of assets. A process is characterised by its parameters as well as by metrics that are used to monitor such process, which provide context when selecting models.
- *Manufacturing Application Model* – this captures the specific industrial ML solutions and how they are linked to the process. For example, a vision inspection solution may support/control the quality of a 3D printing process, or a condition monitoring solution may be related to a machining process. Capturing the industrial solution is essential for querying the ontology for similar ML models.
- *Application Model* – defines a solution or industrial problem from an ML perspective as either a classification or prediction problem. These classes help to link ML applications to industrial solutions as well as to the ML techniques that can be used to solve such application, supporting the design of ML pipelines.
- *Algorithm Model* – captures AI/ML approaches and their parameters. The same approach might be linked to multiple Manufacturing Applications the same way that a Manufacturing Application may be solved by different algorithms.

- *Data Processing Model* – defines different processing steps involved in a ML pipeline. Linking data types to techniques enable the automation of the pipeline.
- *Sensor Data Model*- captures characteristics of measuring devices on assets and enables semantic annotation.
- *AI Pipeline Model* – brings together all the elements that are used for the development of an industrial solution and to track its performance, supporting future model development by using complete/partial pipelines of existing models.

3.2 Building an initial Knowledge Graph and matching models to applications

To use the semantic models proposed, an initial knowledge graph is constructed. Instances for each asset on the shopfloor as well as sensors can be created, and data sets can be semantically annotated. To generate instances of ML pipelines, assuming no model has been created, the proposed starting point is to take advantage of the plethora of research literature to build *template* pipeline instances. There is no general model that works for every industrial problem, so research literature can be a good way to build pipeline templates that can then be fine-tuned through AutoML tools.

Once instances are created, the semantic model can be queried to infer which existing solutions can be reused if they exist, or which pipelines can be used to create a model. To facilitate the matching of existing solutions/data, inference through SPARQL queries and semantic distance are implemented to find similar cases. Depending on the results, if a model exists, then this can be used for transfer learning. If there is no such model but a *similar* one (e.g. from a different industrial application, or different asset type) exists, then the ontology can be used to identify the model's pipeline. If no case is found, then an existing ML pipeline template can be used.

3.3 Semantic framework as a tool for Learning Factories

There are some emerging works on the use of Learning Factories to support the acceptance and scaling of AI solutions in the shopfloor [16]. However, these are focused on the acceptance and learning of *existing* solutions. The semantic framework can complement existing learning environments, as a tool that trainees can use to understand how different AI techniques can be used for different problems, what works best, and understand why they work well. This ultimately not only empowers the employee with data science skills, but also build trust and confidence on the technology they are using.

4 A machine tool monitoring use case scenario

Using Bosch's process monitoring use case [5], instances for asset, process, sensors, data and models were created (some instances shown in Fig. 2). Three machines perform 15 operations with different tools and cutting parameters with a tri-axial accelerometer mounted on the spindle that samples at 2KHz. Data is collected from 2018 to 2021 in periods of 6 months and labelled "OK" and "NO OK" after product quality inspection. An ML model for process monitoring of Machine01 is developed and captured through the semantic model. Using the semantic relationships, the ontology can be queried to develop models for Machine02 in the following ways:

(a) Searching for an existing model of OP00 for process Monitoring – assuming an existing model for Machine02 exists, a sparql query is used to look for an instance that is linked to a solution type *processMonitoring* and to a process of type *stepDrilling* with speed=250 and feed=100. If a model exists but not for the given process parameters, it can be further trained to expand the model's capabilities.

(b) Searching a model from a similar machine for OP00 – the ontology is queried to find assets of type *cncMachine* that have a model for *stepDrilling* and the needed parameters. The ontology matches it with the instance from Machine01 OP01.

(c) Look for a similar machine for OP10 – as Machine02 has no model for OP10, inference is used to retrieve similar assets. As M01 has no model for OP10 either, other *MaterialRemoval* processes from M01 that may be running with the same parameters can be queried. The ontology matches this with OP02. Although the process uses a different tool, parameters match, which is useful as it is known that frequency is a feature relevant to the problem and is correlated to the speed [5].

(d) No model from similar machine/process – in this case, a template pipeline can be retrieved and model development can be supported by AutoML techniques.

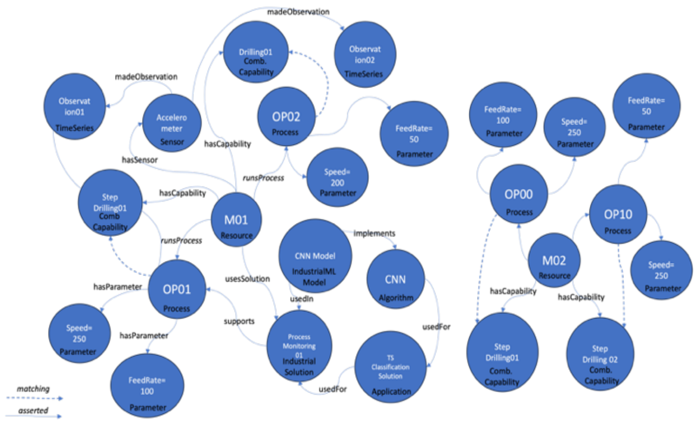

Fig. 2. Some instances from Bosch Process Monitoring use case

5 Conclusions and Future Work

In this paper a framework for unifying industrial a machine learning context information to support ML solution development and reuse is presented. By semantically capturing context, it is not only possible to manage the lifecycle of solutions but also find similar solutions that are useful for transfer learning. In a way the ontology acts as a meta-learning approach that is enhanced as more models are developed. This is particularly useful when not enough data is available for a given machine, taking advantage of pretrained models in a very similar context. By capturing this knowledge and semiautomating model development, it is possible to support the non-experts on how to develop such solutions. Future steps will involve fully implementing the pipelines so not only query to the ontology can be automated but the execution of the ML pipeline for building and retraining models effectively.

References

1. Maier, J., Made Smarter Review, https://www.gov.uk/government/publications/made-smarter-review (2017), last accessed 23/Oct/2023.
2. Rauh, L., Gärtner, S., Brandt, D., Oberle, M., Stock, D., and Bauernhansl, T. (2022). Towards AI Lifecycle Management in Manufacturing Using the Asset Administration Shell (AAS), Procedia CIRP, 107.
3. Huyen, C. (2022). Designing Machine Learning Systems: An iterative process for production-ready applications. O-Reilly.
4. Ompusunggu, A.P., Carcel, C. R. (2023). Low-cost vibration sensor with low frequency resonance for condition monitoring of low speed bearings: a feasibility study. In Low-Cost Digital Solutions for Industrial Automation (LoDiSA 2023) 2023, 158-164. IET.
5. Tnani, M., Feil, M., and Diepold. K. (2022). Smart data collection system for brownfield CNC milling machines: A new benchmark dataset for data-driven machine monitoring, Procedia CIRP, vol. 107, 131-136.
6. Zhou, B., Zhou, D., Chen, J., Svetashova, Y., Cheng, G., and Kharlamov, E. (2022). Scaling Usability of ML Analytics with Knowledge Graphs: Exemplified with A Bosch Welding Case. In Proceedings of the 10th International Joint Conference on Knowledge Graphs (IJCKG '21). Association for Computing Machinery, New York, NY, USA, 54–63.
7. Mäkinen, S., Skogström, H., Laaksonen. E., and Mikkonen, T. (2021). Who needs mlops: What data scientists seek to accomplish and how can MLOps help? In 2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN'21). IEEE, 109–112.
8. Zhao, Y. (2021). MLOPs Scaling ML in an Industrial Setting. Master Thesis
9. Raffin, T., Reichenstein, T., Werner, J., Kühl, A., and Franke, J. (2022). A reference architecture for the operationalization of machine learning models in manufacturing, Procedia CIRP, 115, 130-135.
10. John, M. M., Olsson, H. H., and Bosch, J. (2021). Towards MLOps: A Framework and Maturity Model 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Palermo, Italy, 1-8.
11. Wang, P. and Gao, R. X. (2020). Transfer learning for enhanced machine fault diagnosis in manufacturing, CIRP Annals, Volume 69(1), 413-416.
12. Giannetti, C., and Essien, A. (2022). Towards scalable and reusable predictive models for cyber twins in manufacturing systems. J Intell Manuf 33, 441–455.
13. Järvenpää, E., Siltala, N., Hylli, O. and Lanz, M. (2019). The development of an ontology for describing the capabilities of manufacturing resources. J Intell Manuf 30, 959–978.
14. Janowicz, K., Haller, A., J.D. Cox, S., Le Phuoc, D., Lefrançois, M. (2019). SOSA: A lightweight ontology for sensors, observations, samples, and actuator. Journal of Web Semantics 56, 1-10.
15. Braga, J., Dias, J., and Regateiro, F. (2020). A machine learning ontology. Frenxiv Pap., preprint.
16. Rueckert, P., Papenberg, B., Sievers, T.S. and Tracht, K. (2021). On-site learning factories for implementing artificial intelligence systems in production. In Proceedings of the Conference on Learning Factories (CLF).