Mérnöki Szimpózium a Bánkiban
(ESB 2020)
http://bgk.uni-obuda.hu/esb/2020

Challenges of the application of machine learning in the serial production.

[1]Balázs Szűcs, [2]Áron Ballagi

[1]*AUDI HUNGARIA Zrt., Győr, Hungary, balazs.szucs@audi.hu*
[2]*Széchenyi István University, Department of Automation. Győr, Hungary, ballagi@sze.hu*

Abstract

The industrial applications of the machine learning methods have promising results in the field of smart manufacturing and quality assurance; however, the development of these models have challenges. In this paper we present the classical steps of model creation and the challenges and questions of industrial applications in each step, especially for series production. Finally, we present a possible workflow for model development in the manufacturing.

Keywords: machine learning, industry 4.0, intelligent manufacturing, smart manufacturing, quality assurance

1 Introduction

The industrial applications of the machine learning methods have promising results in the field of smart manufacturing and quality assurance; however, the development of these models have challenges. The typical industrial use-cases of artificial intelligence are pattern recognition, classification, anomaly- or fault detection and fault prediction for predictive maintenance. The serial production plants are highly automated, and with the spread of industry 4.0 and the IoT equipment, enormous amount of data available to train models for the above-mentioned tasks. However, in these plants the processes are strict standardized and highly specialized, this can no longer said of the data acquisition, data analytics and data interpretation. Besides the heterogeneity of the data acquisition, the diversity of industrial and enterprise systems, the interconnectability of these systems and the lack of structured and standardized storage of data makes the creation of models even harder. At last, but not least, the data and the key feature selection are the key steps of successful model creation, this is where the process- or domain knowledge and the functional approach of data analytics plays very important role. In the following sections we present the classical workflow of model creation and their challenges in accordance with industrial application.

2 Methodology of data analytics and model creation

In this section we present a brief overview of the main steps in the methodology of data analytics and model creation with the short description and role of each steps in the procedure. In the followings we discuss the topics data preparation, data pre-processing, model selection, training, performance evaluation.

2.1 Data preparation

Data preparation is the very first step of data analytics and model development. The aim of data preparation is to get an overview of the process data and then select the most corresponding samples and features for modeling. The main tasks are data extraction from databases or systems, data structure analysis and data selection to build and train models. The sample and variable selection is based on the model to be built.

2.2 Data pre-processing

Data pre-processing is essential to be carried out to improve the quality of the data, and some appropriate data transformations may be needed to make the data modeling more efficient. This process boosts the model's performance and accuracy. In this step, the inconsistency in time needs to be eliminated, outliers and gross errors should be removed, the missing values need to be addressed, e.g. deletion of the sample, missing value estimation and the scale difference among process variables needs to be considered (the step of scaling or normalizing) to increase accuracy and decrease training time. In this step takes place the splitting of the dataset for training, testing and validation datasets, which are used for training and model validation and performance evaluation.

2.3 Model selection and training

After the data selection and pre-processing comes the selection of an appropriate machine learning algorithm for data model construction. Based on the detailed analyses of data characteristics, the architecture of the model and the complexity of the data model can be evaluated. After the architecture of the model has been selected, the model parameters can be determined by implementing and testing the machine learning algorithm on the training dataset. In this step, different data models (algorithms) can be formulated as different optimization problems. During the training we optimize the loss function of the model.

2.4 Performance evaluation

After the training, the performance of the model needs to be evaluated. The most used model validation and performance evaluation methods are cross-validation, model stability analysis, model robust analysis, parameter sensitivity analysis, etc. For model validation and performance evaluation, separate testing or validation dataset required.

3 Challenges of machine learning applications in serial production.

In this section we present the most common challenges of application of machine learning methods in industrial environment, with special attention to serial production. We inspect the challenges in accordance with the steps of the methodology of data analysis and model creation.

3.1 Challenges of data preparation

The data preparation and selection are essential steps of data analytics and model creation. In the manufacturing there are several systems for data acquisition, each with different purpose. A part tracking system tracks the manufactured parts and their statuses, an inline supervision system measures and stores process signals and values, the ERP system stores additional information like financial or supplier data; just to name a few possible systems. There are data, which are present only in one or all the systems. Which will be the lead system, which provide the 'ground truth'? The data representation can be different in every system. How do we know what data format to use? How do we convert the data between the systems? If a measurement system stores only the measured values and the timestamp of the measurement and the part ID is stored only in the part tracking system to prevent redundant data storage, how do we connect the data in the two systems? Which timestamp is used in which system, the timestamp of the measurement or the timestamp of the data acquisition? How do we synchronize them? If the identification or the name of the data source is different in the systems which must be interconnected, how do we connect the corresponding data? What technology to use to collect the data? What are the key features in the collected data? These are only the few of the possible challenges of data preparation in real-world industrial applications, which must be addressed. These questions cannot be answered without the exact knowledge of the system architecture, the data acquisition, infrastructure, and the process.

3.2 Challenges of data pre-processing

Data pre-processing is essential to be carried out to improve the quality of the data. We need to deal with the inconsistency, the missing values, the outliers or delete corrupted samples. But are these samples really corrupted or they are only exotic values in the process? How do we replace the missing values? Is the usage of the average value of the samples are good enough? If there are labels, are the labeling valid? Are the data valid and describes the process well? To answer these questions, the knowledge of the physical process and the manufacturing technology is essential.

3.3 Challenges of model selection and training

To choose the best performing model first we need to determine the desired behavior of the model. The architecture, the algorithm, the loss, and optimization function can be very different for the different tasks. It is essential to frame the problem. Is it a regression or a classification problem? Is there any seasonality in the data or not? Is it a timeseries or discrete values or features? Do we need a standalone model, a stacked or a multi model architecture with different architectures which corresponds to the relevant features? The industrial processes are well standardized and controlled, thus the quality of the product is highly unbalanced and the data also: the percentage of waste and rejected products are relatively small. How to train a classification model if the input data highly unbalanced? Which training strategy to choose? Are there similar problems and solutions? What is the best batch size for the training? How many epochs do we have to train the model? These questions are the addressed to the data scientist or to the machine learning engineer.

3.4 Challenges of performance evaluation

Before deploying the model in production, to evaluate and tests are needed. But if the training and the input data also highly unbalanced, the evaluation can take long time. In the production the appearance of a rare event can vary between days and years. If the model misclassifies or cannot detect these rare events, was the result of the poor training data or was the results of the behavior of the model, which thought to be the good? Do we have to detect events which are occurs in every few years or not? That depends on if we manufacture candy bars or operate a nuclear power plant. Often there is no time to make expensive and well-prepared manufacturing or process tests. We need to decide which features and events are worth the time and cost. These are both the questions addressed to the model's creator and for the process engineers and the process hosts.

4 Conclusion

As it seems form section 3, the industrial application of machine learning not just machine learning problem, rather a complex, multidisciplinary engineering task. Not only the domain knowledge, but the knowledge of the industrial processes, the products and their characteristics, the quality characteristics, the data acquisitions methods, the automation technologies, and the IT architecture are essential for the successful industrial application of ML methods. In addition, the application of these methods and the tasks of these models are based on rational technologically and financial decisions.

The aim of ML in the manufacturing is to support decision making and the efficient productions of high-quality products, therefore the applications of the methods should be placed on the shopfloor, near the production, where the real value is produced and where knowledge for application are present.

4.1 Workflow for industrial applications of ML

Based on the above-mentioned experiences, we propose one possible workflow for successful industrial application of machine learning consist the following steps:

1. Framing the problem with technological approach, domain knowledge: we determine what event or cases has the model detect or classify, then we choose the relevant technological parameters and features based on the product and the manufacturing process. (In contrast with classical BigData methods, where we collect every data, and we try to find pattern in the data.) Consultation with product-, process-, automation- and quality assurance engineers.

2. Identifying source systems and data format, data map: we identify which data in which system(s) are present, identification of data and creation of data dictionary. Defining data acquisition technology to collect data from systems. If possible, we use uniform naming convention for naming the data in the different systems. Consultation with automation-, IT-engineers and applications hosts.

3. Model creation, training, and evaluation: based on the needs of the production, we create the model for the predefined task, then we test and evaluate the model on the real process data. Consultation with data scientist-, product-, process- and automation engineers.

4. Deploying the model in production: after successful evaluation of the model deployment of the model to the production. Full life-cycle management needed for further development and fine-tuning if needed. Consultation with product-, process-, automation- and quality assurance engineers and process host.

4.2 Further research

Based on experience, further research on the topic and the proposal of a framework for the industrial application of machine learning with standardized, homogenized data collection technology, uniform naming conventions and a modular toolbox are required.

5 References

[1] Z. Ge, Z. Song, S. X. Ding and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," in IEEE Access, vol. 5, pp. 20590-20616, 2017, doi: 10.1109/ACCESS.2017.2756872.

[2] Thorsten Wuest, Daniel Weimer, Christopher Irgens & Klaus-Dieter Thoben (2016) Machine learning in manufacturing: advantages, challenges, and applications, Production & Manufacturing Research, 4:1, 23-45, DOI: 10.1080/21693277.2016.1192517

[3] Wickham, Hadley. (2014). Tidy data. The American Statistician. 14. 10.18637/jss.v059.i10.

[4] P. Kadlec, B. Gabrys, and S. Strandt, ''Data-driven soft sensors in the process industry,'' Comput. Chem. Eng., vol. 33, no. 4, pp. 795–814, Apr. 2009.