Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Procedia CIRP 00 (2021) 000–000

www.elsevier.com/locate/procedia

54th CIRP Conference on Manufacturing Systems

Predictive analytics in quality assurance for assembly processes: lessons learned from a case study at an industry 4.0 demonstration cell

Peter Burggräf[a], Johannes Wagner[a], Benjamin Koke[a], Fabian Steinberg[a], Alejandro R. Pérez M.[a], Lennart Schmallenbach[a], Jochen Garcke[b,c], Daniela Steffes-lai[b], Moritz Wolter[b,d]

[a] *Chair of International Production Engineering and Management (IPEM), Universität Siegen, Paul-Bonatz-Straße 9-11, Siegen - 57076, Germany*
[b] *Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven 1, Sankt Augustin- 53757, Germany*
[c] *Institut for Numerical Simulation, Universität Bonn, Endenicher Allee 19b, 53115 Bonn*
[d] *Institut for Computer Science, Universität Bonn, Endenicher Allee 19a, 53115 Bonn*

* Corresponding author. Tel.: +49-271-740-4509; fax: +49-271-740-2630. *E-mail address:* alejandro.perez@uni-siegen.de

Abstract

Quality assurance (QA) is an important task in manufacturing to assess whether products meet their specifications. However, QA might be expensive, time-consuming, incomplete, or delayed. This paper presents a solution for predictive analytics in QA based on machine sensor values during production while employing machine-learning models based on logistic regression in a controlled environment. Furthermore, we present lessons learned while implementing this model, which helps to reduce complexity in further industrial applications. The paper's outcome proves that the developed model was able to predict product quality, as well as to identify the correlation between machine-status and faulty product occurrence.

Keywords: Machine-Learning; Predictive Quality; Production; Quality Assurance; Logistic Regression

1. Introduction

(AP:

- **Products' quality must always be guaranteed.**
- **In certain cases (especially in series assembly production) there is a delay between quality inspection and end of the production process high potential for producing scrap**
- **By doing quality inspection based on samples, there is still a risk of sending NOK parts to customers**
- **The cost of quality checks might be expensive**
- **Quality assurance is normally done at the end of production (for assembly processes)**
- **Early prediction of the production's quality would help to decrease the amount of scrap and the risk of sending NOK parts to customers**
- **Variable predictions by using machine learning models is currently widely used in diverse fields due to its accurate predictions with the data collected from the machines, prediction models can be implemented for specific predicted values**

)

2. Related Works

2.1. Quality check in production

The quality control of production processes and products is one fundamental part of the general quality management. It contains the measures and activities needed to fulfill specific quality requirements [6]. According to Illés, there are three established tools and methods of quality control [6]: The Statistical Process Control (SPC), Auditing and Total Quality Control (TQC). The SPC is normally used to monitor the production process using recorded data values for the quality characteristics, as well as to indicate any significant changes in the quality characteristics in the production process [13]. This method can take place right before and/or after a manufacturing process and be implemented as an "off-line" or "in-line" SPC. **(AP: @Lennart, last sentence is still quite confising to me)**

The biggest advantage of working with SPC is that large batches can be checked without major time impact on the QC process, however, this would be only possible if the variance of the process product parameters is small enough, to ensure process capability [7].

(AP: 100% method is not introduced)

Product and process auditing is the independent examination of product and process quality to provide information and it is considered the most fundamental quality control technique. This method is executed by an independent auditor, who based on sampling or checking records limits the finding on reporting incorrect processes procedures or products documentation. In a further step, the findings would be later corrected by other than the auditor [5].

Both statistical methods and hundred percent controls can be applied as an audit tool.

Statistical methods such as acceptance sampling or quality control charts represent a compromise between zero and hundred percent control [13, 10].

While the SPC mostly optimizes the capability of the process, for example the acceptance plan prevents the nonconforming product to be pass and then delivered to the next process or the customer [10]. In the case of a hundred percent control, this is associated with higher costs and a higher requirement of time, compared to a quality control based on a machine learning model. Statistical methods are faster as **(AP: as "or" "than")** hundred percent controls, but have the same problem like the SPC: Because they are based on assumptions, this methods are prone to errors, if the variance of the product quality is to high. The concept of TQC is to extend the scope of quality control to the whole company and product life cycle, by involving all departments of the organization [6]. The quality characteristics of a product can be classified as variables or attributes. Variables are measurable characteristics, which are shown as numbers. On the contrary, attributes are not measureable and can only take the value go or no-go [11]. The advantage of variable measurements is that in the event of a defective intermediate or end product, precise measured values are available, which enable precise adjustment of the process parameters. This allows the desired quality to be restored. With that in mind, a predictive quality control is a promising solution.

(AP:

- **How quality check is done in production**
- **Concepts and sampling plan**
- **Attributive vs Variable measurements [VM has higher measurement accuracy, reason why it's better to use regression model]**

)

2.2. *Predictive analytics in production*

According to Delen and Damirkan, predictive analytics (PA) consist on unravel the inherent relationships, if any, between data input and output by using data and mathematical techniques [4].

PA offers a wide range of applications, and it can be implemented as long as there is enough data available. By conosidering industrial and production environment, PA would be a great asset due to the high flexibility and high amound of data available comming from processes, single machines or cells, producs, and others. Some commmon industrial applications of PA at the industry are focused products (design and optimization), machine (predictive maintenance) and production (scheduling) [8].

Applications that combines PA with QC might be a great improvement in the quality control process because it is an improvement compared with the earlier mentioned methods 2.1, in terms of speed, accuracy and time delay.

The outcome of an explorative literature review on this matter, shows that the findings that combines PA and QC are very scarce in literature. This could be in great part, due to the novelty of the area [1, 12]. Related work, e.g from Krauß focused on automated machine learning for predictive quality in production [9] and a product quality prediction in a process chain [8].

Based on an explorative literature review on this matter, the short outcome is that the findings that combines PA and QC are very scarce in literature. This could be in great part, due to computer power limitations in early researchs and the recent increase of ML applications on this area [1, 12]. Related work, e.g from Krauß focused on automated machine learning for predictive quality in production [9] and a product quality prediction in a process chain [8].

(AP: - Introduce research question - why do we do it? - what do we expect to find or prove - the motivation is clear)

The few existing approaches are extended by this paper in that a demonstrator executes an industry-unspecific production/assembly process, the collection and processing of the data from the process parameters in continous and a few different ML algorithms are used for prediction, in order to determine, which algorithms are most advantageous for predicting the quality of an outcoming product. In addition, it is described in detail how the collected data was collected, prepared and processed.

(AP: Last paragraph goes in intro)

3. The case industry 4.0 demonstration cell

In our work we applied a case-based research approach [14] to answer the research question. One central motivation for a case-based research approach is to gain insights for real needs, rather than to develop theories without practical relevance [3]. In addition case-based research approaches have already been successfully applied in the area of quality prediction based on process parameters (see, for example, ...). **(AP: - How does the previous sentences apply to the democell? - This suits better in the methods section)**

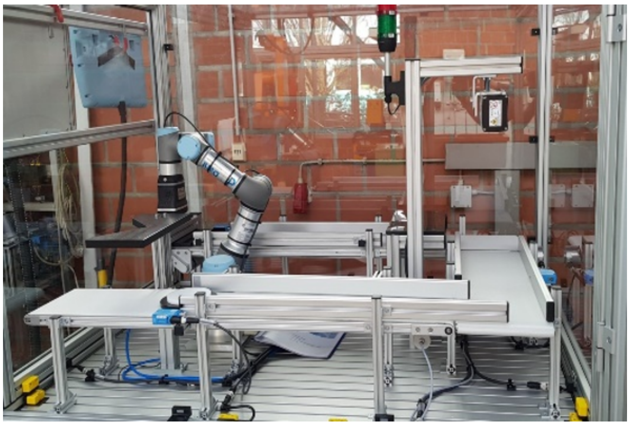

Fig. 1. The industry 4.0 demonstration cell at the University of Siegen

Our case of study is an abstraction of an assembly process at the industry. The industry 4.0 demonstration cell is composed by three independent conveyor belts, a robotic assembly arm, a laser scanner used for quality control as well as a wide range of sensors, all orchestrated by a SIEMENS PLC S7-1200. The assembly process done at this demostration cell consist on stacking two disk of different size on the top of each other. The quality of the pseudo products is later evaluated by the laser scanner by evaluating the concentricity of both disk. If the disk's concentricity is within a tolerance of 1.5mm, the piece is classified as OK, if not, it shall be classified as NOK. By considering that malfunctions can occurs during real production processes at industries, our demonstration cell is able to simulate failures in diverse areas, such as:

- Assemble errors due to robotic arm
- Bearing damage on the conveyor belt
- Resistance on the conveyor belt
- Leakage in the compressed air system
- Productivity limit reached
- Missing material

Those simulated failures have a real impact in the data collected by the sensors. Anomalies can be observed at the vibration sensors, temperature rise, fluctuations in the air pressure system, noticiable shifting on pieces' position as well as the stop of the process.

The continous data collection and clear correlation between the machine parameter is the ideal play-ground to implement a sandbox for machine learching implementations. With this in mind, our goal is to predict the quality of the product (concentricity of the assambled parts), prior to the actual quality check done by the laser scanner. The findings gained with the demonstration cell can be transferred to various real assembly processes and use cases, due to its manageable size and high degree of scalability. This would be an step forwards for future implementations at the industry, where the quality predictions might determine early enough, the next steps for the assambled pieces.

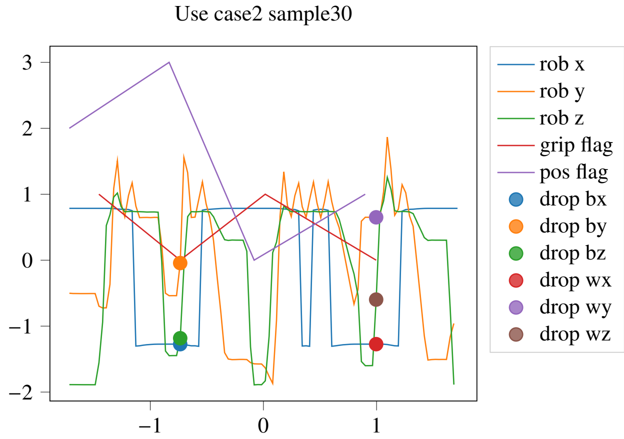

Fig. 2. Robot arm movement patterns. We show the movement of the cell's robot arm, its grippers, and the encoding of the overall position in the system. The changes in the grip flag are triggered, when the gripper opens or closes. Zero indicates an open, while one indicates a closed gripper. The position flags marks the state of the cell's arm. 0 means the black disk is in store, one that the black disk is at the gate, two that the white disk is in store and three that the white disk is located at the gate.

4. Machine Learning Methods

We compare Support Vector Machines (SVM), Multilinear Perceptrons [2] as well as Random forests in the task of quality prediction.

5. Experiments

5.1. Recording the data

(DSt: - could we include collection via OPCUA server? or manually?)

5.2. Classifier optimization and testing

The robot's position at disc drop is marked in Figure 2. The quality prediction problem is framed as a classification task. The input vectors which we feed into our classifiers consists of the arm positions at the disc drops for both discs in three dimensions, as well as the largest recorded belt speed of all three belts. The interpretation of the quality measurement is used as the training target. It can be good or bad which we encode as zero and one.

We work with a total of 132 measurements. Each containing arm and belt data logs of individual cell runs. 20 samples are set aside at random for testing purposes, leaving 112 training samples. The random number generator seed is set to one in order to ensure the train and test set splits are identical for all experiments.

We compare a total of three different classifier architectures on the data. A Support Vector Machine (SVM), a multilinear perceptron and a Random Forest structure. All three methods

Support Vector Machine	95.%
Multilinear Perceptron	95.%
Random Forest	100.%

Table 1. Method comparison for our annomaly detection task using the industry demo data test set.

are trained using standard hyperparameters. [1] Results are shown in table 1. For 55 of the total 132 samples quality measurements indicate a problem. This sets the baseline over the entire data set, that would be obtained by simply labeling all samples as good. Over the entire data set, we require to classify more than 58.3% of the data correctly. The test set contains eight incorrect samples. We would therefore expect any naive classifier to produce a 60% accuracy, which we have to beat. In Table 1, we observe that this is indeed the case for all three approaches evaluated here. The random forest performed best in this case.

6. Conclusion

6.1. Lesson Learned

(AP: explain experiences by doing the research - Connection problem between devices ¡¡¡¡¡¡¡ HEAD - Network connection - Device restriction: SIM4000, SIEMENS. - Server Timestamp syncr. - Interpretation of data: preprocessing ======= - Missing tags - Interpretation of data: preprocessing ¿¿¿¿¿¿¿ 5272fc99b92062cad1141c5f767ee0d54ba8d267 - What data is relevant? - Missing data tags: modifying PLC and robot code to have tags -¿ key point - Without label we cannot use supervised ML methods - Dataset size - ML implementation @Moritz)

7. Conclusion

Acknowledgments

This project was partly supported by the german and european IKT.NRW project ”‘ManuBrain”’ **(DSt: do we need to include ManuBrain reference here?)**

References

[1] S. Aumi, B. Corbett, and M. P. Model prdictice quality control of batch processes. *American Control Conference*, (5646-5651), 2012.
[2] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.
[3] D. Cutcheon and J. Meredith. Conducting case study research in operations management. *Journal of Operations Management*, 11(3):239–256, 1993.
[4] D. Delen and H. Demirkan. Data, information and analytics as services. *Decision Support Systems*, 55(359-363), 2013.
[5] M. Fox. *Quality Assurance Management*. Springer, 1993.
[6] B. Illés, P. Tamás, P. Dobos, and R. Skapinyecz. New challenges for quality assurance of manufacturing processes in industry 4.0. *Solid State Phenomena*, 261(481-481), 2017.
[7] W. Kahle and E. Liebscher. *Zuverlässigkeitsanalyse und Qualitätssicherung*. Oldenbourg Verlag München, 2013.
[8] J. Krauß, J. Dorißen, H. Mende, M. Frye, and R. Schmitt. Machine learning and artificial intelligence in production: Application areas and publicly available data sets. *Production at the leading edge of technology*, 9(493-501), 2019.
[9] J. Krauß, B. M. Pacheco, H. M. Zang, and R. Schmitt. Automated machine learning for predictive quality in production. *Procedia CIRP*, 93(443-448), 2020.
[10] N. kurniati, R.-H. Yeh, and J.-J. Lin. Quality inspection and maintenance: the framework of interaction. *Procedia Manufacturing*, 4(244-251), 2015.
[11] A. Mitra. *Fundamentals of Quality Control and Improvement*. wiley, 2016.
[12] C. Ritter and F. Schweitzer. Neue wegene der prozeßmodellierung für die rechnergestützte qualitätsvorhersage und -sicherung in der spanplattenfertigung. *Holz als Roh- und Werkstoff*, 50(62-66), 1992.
[13] D. Selvamuthu and D. Das. *Introduction to Statistical Mehods, Design of Experiments and Statistical Quality Control*. Springer Nature Singapore Pte Ltd., 2018.
[14] R. Yin. *Case study research and applications: Design and methods*. Sage publications, 2017.

[1] In order to allow exact reproduction of these results source code is available at `https://github.com/manubrain/Demo-Cell-Classification`