Towards Data Driven Process Control in Manufacturing Car Body Parts

Bas van Stein✉, Matthijs van Leeuwen and Hao Wang
LIACS, Leiden University
Niels Bohrweg 1, Leiden
The Netherlands
Email: {b.van.stein, h.wang, m.van.leeuwen}@liacs.leidenuniv.nl

Stephan Purr, Sebastian Kreissl and Josef Meinhardt
BMW Plant Regensburg
93055 Regensburg, Germany
Email: {Stephan.Purr, Sebastian.Kreissl, Josef.Meinhardt}@bmw.de

Thomas Bäck
LIACS, Leiden University
Niels Bohrweg 1, Leiden
The Netherlands
Email: t.h.w.baeck@liacs.leidenuniv.nl

Abstract—**The manufacturing process of car body parts is a complex industrial process where many machine parameters and material measurements are involved in establishing the quality of the final product. Data driven models have shown great advantages in helping decision makers to optimize this kind of complex processes where good physical models are hard to build. In this paper a framework for on-line process monitoring and predictive modeling is proposed to optimize a car body part production process. Anomaly detection plays an important role in this framework as it can provide an early alert for operators on the production line using a complex set of machine parameters and material properties. In this paper an anomaly detection algorithm, GLOSS, that is successfully implemented as the first module in the process, is introduced. GLOSS finds local outliers in high dimensional mixed data-sets using a relative density measure that takes the global neighborhood into account while searching for outliers in subspaces of the data. An overview of the application and implementation of the algorithm in the car body part press shop is presented.**

Index Terms—**Anomaly Detection, Industry 4.0**

Short Paper for CSCI-ISBD

I. Introduction

In the car body parts manufacturing industry, data mining and on-line automated quality control are emerging and important topics [1], [2]. In an *Industry 4.0* factory, machines and products are interlinked with each other as one collaborative process, also known as the *Internet of Things* [3]. On-line quality control of the products and the prediction and avoidance of defects are among the key goals of applying data analytics and optimization to such processes. In the car body parts industry, blanks of sheet metal are cut from a coil and pressed into car body parts such as side frames, roofs as well as structural parts like B-pillars. For different parts, different materials are required and different machine settings are used. Due to the high variation as well as high dimensionality in both material properties and machine settings, the process is a very complex one with lots of parameters that influence the final product.

To estimate where defects might occur, data mining techniques have to be applied at the very beginning of the production process. Anomaly detection [4] plays an important role in this early stage, since most of the parameters are still unknown. However, applying anomaly detection in this real-world setting is a major challenge. The dimensionality of the problem is large and the data consists of heterogeneous coil types and suppliers used for many different body parts. Using anomaly detection techniques on material properties allows for the detection of anomalous sheet metal coils and more precisely, regions in the sheet metal that could later lead to problems in the production process. The results of anomaly detection algorithms can be presented to experts to gain additional knowledge about the process. In the near future, the results could directly be used in the press line, and depending on the anomaly scores, a "careful" flag could be set. Further steps in the optimization of the production process can be done by data-driven predictive models and model-based optimization algorithms [5].

In Section II the car body parts manufacturing process is explained into detail. Next, in Section III a data-driven framework is proposed for the manufacturing of car body parts together with one of the main modules, anomaly detection. A high-level overview of anomaly detection algorithms is given in Section IV. Finally, a novel anomaly detection algorithm designed specifically for the framework is introduced and early results are discussed in Section V and VI.

II. Manufacturing Car Body Parts

The manufacturing process, in the current study, consists of two main processes and a buffer period. First, the incoming steel coils are unrolled and cut into individual blanks. During the cutting process, the following properties are measured:

Impulse Magnetic Process On-line Controller (IMPOC) is an advanced measurement commonly used in steel manufacturing plants that measures the residual magnetic field strength of the material [6].

Oil Levels on the surface of the blanks are considered to be an important factor in the stamping process. The amount of lubricant affects the friction and thus plays an important role in the deep drawing process of sheet metals.

Roughness of the surface.

Thickness of the material.

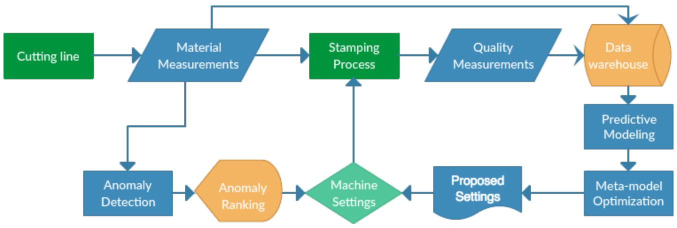

Fig. 1. A framework for optimizing and monitoring a car body parts manufacturing process.

Peak Count of the surface, representing the number of peaks per square meter.

The steel blanks are then stacked on top of each other and stored in the buffer. After a certain time in the buffer, the stack of blanks are moved to the press line. At the press line the blanks are pressed into a specific car body part. Depending on the body part produced, the press line consists of a number of operations, each of them controlled by a large variety of machine parameters. Anomaly detection is needed to detect deviating material properties to warn the press line controllers of risks as early as possible. When blank properties are deviating significantly or abruptly, it might affect the quality of the produced parts. When these sudden changes are identified before entering the press line, the machine parameters can be adjusted in advance, reducing the potential number of defects, resulting in higher utilization and therefore reduced cost.

III. A Framework for On-line Process Control

The optimization of such processes is far from trivial and receives increasing attention in research. In the field of *Optimal Operational Control* [7], [8] and *Real Time Optimization* (RTO) [9], [10] static mathematical models are constructed for the industrial process and used in the optimization procedure to search for good operational parameters for a specific process. These model-based control theories include both linear and nonlinear systems where set points of the controllers are assumed to be known. The downside of these mathematical models is that external disturbances and noise are usually not included and that the process of specifying these models can be very hard. Another downside is that the optimal configuration of operational parameters is almost never known and therefore these model based approaches will not be able to completely optimize the production process. Because of these limitations a data-driven approach for monitoring and optimizing production processes is proposed in this paper. Because the data-driven modeling requires minimal understanding of the mechanisms of the process and more importantly, the noises in measurements or machine settings can be reduced largely by incorporating a large amount of data, resulting in robust and reliable models. The proposed framework has to deal with high dimensional data coming in real-time. The framework needs to provide valuable feedback to the domain experts, decision makers and process controllers about the current and preferably also the future situation of the production process.

The proposed framework globally consists of the following steps: First data is collected per process unit (step) in a centralized data management system. The data being collected first for the car body part stamping process comes from the cutting line where material properties are measured (see Figure 1). Second, unsupervised algorithms such as anomaly detection are applied on the measurements to quickly detect deviations and interesting regions in the incoming blanks. Finally, predictive data driven models such as *Random Forests* [11] can be trained on the collected historical data containing material properties, machine parameters and quality indicators, to predict the final quality of the incoming steel blanks before stamping. Based on predictive models, the machine settings can be optimized to achieve a better predicted product quality using model-based optimization procedures [5].

IV. Anomaly Detection in Industry

In industry, anomaly detection algorithms are used in many areas to detect possible flaws in systems and processes [4]. Most applications of anomaly detection can be found in the security [12], [13], insurance and banking sectors [14], where anomaly detection algorithms are applied to detect possible intrusions and fraud cases, respectively. Anomaly detection in the car manufacturing industry is not applied on a wide scale as of yet, although several applications exist. For example, anomaly detection can be used in mixed-product assembly lines to detect abnormal logistic states (ALSs) [15]. These states seriously hinder efficient delivery of materials to the assembly line and it is therefore of major importance to detect them as early as possible. Anomaly detection can also be used in the automatic inspection of metal parts by using thermographic images [2]. Using these thermal images, cracks and other surface defects can be detected on the fly using existing anomaly detection procedures. However, a complex process application such as in this paper has not been reported yet.

V. Gloss: Anomaly Detection for Complex High Dimensional Data

Many anomaly detection algorithms already exist, from statistical anomaly detection algorithms [16] that detect global outliers to clustering based anomaly detection algorithms [17]. Most popular outlier methods used today are density based local outlier algorithms such as *Local Outlier Factor* (LOF) [18], *Local Outlier Probabilities* (LoOP) [19] and *Local Correlation Integral* (LOCI) [20]. These outlier detection algorithms compare the densities of data points with the relative densities of the direct neighboring data points. The main bottleneck of these local outlier algorithms is that they suffer from the curse of dimensionality when the dimensionality grows to several hundreds of parameters. That is why for the complex task of detecting anomalous steel coils in this high dimensional mixed data set, *Global Local Outliers in Sub Spaces* (Gloss) is introduced. Gloss is a local outlier detection algorithm based on the existing outlier detection algorithm LoOP [19] combined with a subspace search procedure from *High Contrast Subspaces* (HiCS) [21]. The algorithm searches for outliers in

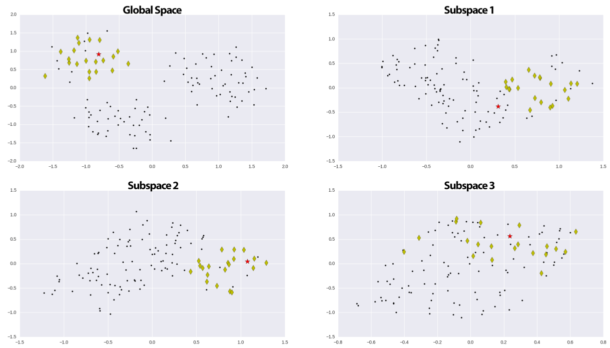

Fig. 2. Synthetic dataset with six dimensions, consisting of a mixture of samples from three distributions. Shown are the global $6D$ space projected onto $2D$ (top left), and three $2D$ subspaces (other). The implanted outlier (red star) can only be detected in Subspace 1 (top right) if local outlier detection uses its global neighbors (yellow diamonds) instead of its subspace neighbors.

subspaces of the data instead of the global feature space. In practice this means that each coil is split into 99 overlapping segments in the length of the coil. Each segment (with a length of 2% of the complete coil) is assigned an outlier probability using the relative density between this segment and the same position on globally similar coils. Because the data set consists of several different mixtures (steel grades and suppliers), it is important that the segments are related to the coils' global neighborhood instead of the local neighborhood. When only considering local neighborhoods, local outliers could "hide" behind different coil types (and thus not be detected).

Figure 2 illustrates the problem that we consider on a synthetic dataset. The data consists of three normally distributed clusters in six dimensions. When considering all the data points, the data point depicted by the red star is not a local outlier in any of the subspaces; neither in the global nor in any of the two-dimensional subspaces (only three shown). However, when only considering the data point's neighbors in the global space, depicted with yellow diamonds, we can observe that the red star is a clear outlier in the 2D subspace shown in the top right plot: *it is relatively far away from other data points belonging to this component of the mixture*. This special type of outliers we call *Local Subspace Outlier in Global Neighborhood* and the aim of the proposed algorithm GLOSS is to detect these outliers as well as more regular global and local outliers. Existing outlier detection algorithms are unable to accurately mark the above outliers, whereas our method can, especially in high-dimensional data.

On a high level, the algorithm uses the following procedure. First of all, the global k-neighborhood is computed for each data point. After that, for each data point a local outlier detection method is used to compute outlier scores for each considered subspace, *relative to its global neighborhood*. As mentioned, the instantiation in this paper uses LoOP because it computes (normalized) probabilities rather than hard-to-interpret scores. Finally, for purposes of ranking each data point is assigned the maximum probability assigned to one of the considered subspaces. In more detail, we combine and adapt a combination of LoOP and HiCS as follows. The *standard distance* of LoOP is altered to incorporate a feature subspace F and a *global* neighborhood relation G:

$$\sigma(p_F, G_p) = \sqrt{\frac{\sum_{s\in G_g} d(p_F, s_F)^2}{|G_p|}}, \tag{1}$$

where p_F and s_F are data points p and s projected onto subspace $F \in \mathcal{F}$ and G_p is the set of points in the global neighborhood of p. Then, based on the *probabilistic set distance* (pdist) as defined in LoOP [19], we define the *Probabilistic Global Local Outlier Factor* $PGLOF$ as:

$$PGLOF_{\lambda,G_p}(p_m) = \frac{pdist(\lambda, p_m, G_p)}{E_{s\in G_p}[pdist(\lambda, s, G_s)]} - 1 \tag{2}$$

Where λ is a constant that is set to 3 for a 98% confidence interval. Finally, subspace outlier probabilities are computed using $PGLOF$ as defined in Definition 1, i.e., with the *global* neighborhood projected onto the features in the *subspace*.

Definition 1 (Global Local Outlier in Subspaces): The probability of a point p being a global local outlier in subspaces is defined as:

$$GLOSS_S(p) = \max\left\{0, \operatorname{erf}\left(\frac{PGLOF_{\lambda,S}(p)}{nPGLOF\cdot\sqrt{2}}\right)\right\}$$

where $nPGLOF = \lambda \cdot Stddev(PGLOF)$ the standard deviation of PGLOF values, assuming a mean of 0, and erf is the standard *Gauss error function*.

VI. Application to Industry

The GLOSS algorithm is applied on an industrial proprietary dataset made available to us by the *BMW Group* at plant Regensburg, Germany. This dataset is the original motivation of GLOSS since it is high dimensional and consists of a highly mixed set of steel coils from different suppliers and steel grades. The steel coil dataset consists of 2204 coils (data points) from the time period December 2014 to December 2015. Each coil is represented by 1188 features, grouped into 99 12-dimensional subspaces. Each subspace represents 2% of the coils length and consists of 3 tracks in width. Each subspace consists of 3 averaged IMPOC measurements and 9 averaged Oil level values (3 for each track). GLOSS and LoOP are compared using all global features. Other algorithms are not included in the evaluation because of the high dimensionality of the data; run times would be unreasonably long. Two sample coils of the results are shown in Figures 3 and 4.

It can be noticed from Figures 3 and 4 that GLOSS is capable of detecting a complete outlier region in the coil, while LoOP is only capable of capturing the sudden changes at the start and end of these anomalous regions. For example, in coil $\#1$ there is a region near the end of the coil with a sudden drop of oil levels and IMPOC. LoOP is capable of detecting the sudden drop at the beginning of this region but reports that everything is fine once the oil levels are being stable again. GLOSS on the other hand detects this complete region of low IMPOC and oil levels as being unexpected behavior for this coil. Process experts confirm that the results given by GLOSS are more informative and leading to better outlier rankings.

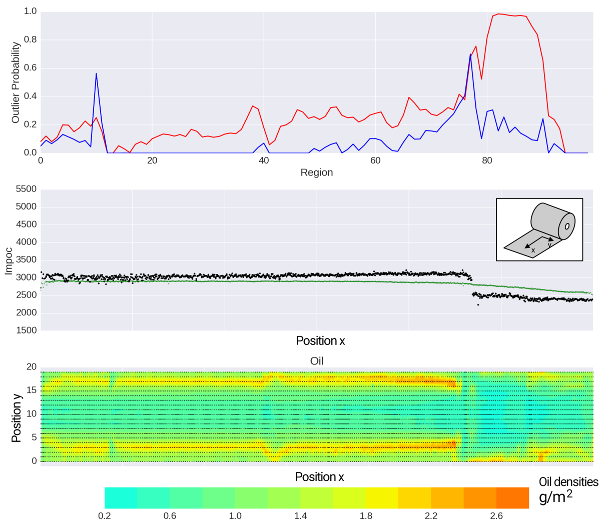

Fig. 3. Results for coil #1 Top: GLOSS (red line) and LoOP (blue line) outlier probabilities for each of 99 consecutive coil segments. Middle: IMPOC measurements over the whole length of the coil, both for this particular coil (black) and averaged over its 20 global neighbors (green). Bottom: Oil level measurements visualized in 2D, representing the entire surface of the coil.

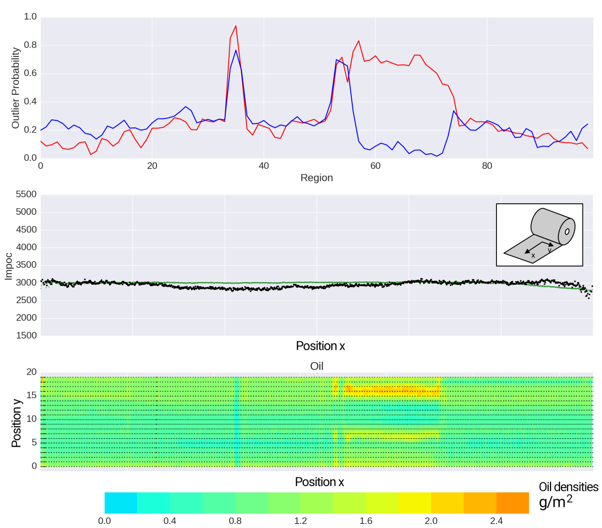

Fig. 4. Results for coil #2. Details identical to that of Figure 3.

VII. Summary and Outlook

A global framework for data driven control of complex production processes is proposed that aims to predict and optimize the quality of the final products. One of the main modules contributing to this framework is a novel anomaly detection algorithm for detecting anomalies in high dimensional mixed data sets; GLOSS. The anomaly detection module is applied successfully in a car body parts stamping process where detecting outliers as early as possible can be very helpful/beneficial. Several alternative state-of-the-art anomaly detection algorithms are assessed and local outlier detection is compared with GLOSS on a high dimensional dataset of steel blanks. Process experts from BMW confirm that results obtained with GLOSS seem to possess a higher quality and is easier to interpret.

The global framework needs many additional modules, such as the predictive data driven models and meta-model optimization modules to realize and complete the on-line modeling and model-based optimization for the industrial production process.

References

[1] S. Purr, J. Meinhardt, A. Lipp, A. Werner, M. Ostermair, and B. Glück, “Stamping plant 4.0–basics for the application of data mining methods in manufacturing car body parts,” in *Key Engineering Materials*, vol. 639. Trans Tech Publ, 2015, pp. 21–30.

[2] M. Benmoussat, M. Guillaume, Y. Caulier, and K. Spinnler, “Automatic metal parts inspection: Use of thermographic images and anomaly detection algorithms,” *Infrared Physics & Technology*, vol. 61, pp. 68–80, 2013.

[3] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong, “Data mining for the internet of things: literature review and challenges,” *International Journal of Distributed Sensor Networks*, vol. 2015, pp. 12–26, 2015.

[4] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–15, 2009.

[5] R. R. Barton and M. Meckesheimer, “Metamodel-based simulation optimization,” *Handbooks in operations research and management science*, vol. 13, pp. 535–574, 2006.

[6] K. Herrmann and M. Irle, “24 impoc©: An online material,” *Flat-Rolled Steel Processes: Advanced Technologies*, pp. 265–271, 2009.

[7] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1, no. 2.

[8] T. Chai, “Optimal operational control for complex industrial processes,” *IFAC Proceedings Volumes*, vol. 45, no. 15, pp. 722–731, 2012.

[9] K. B. Ariyur and M. Krstic, *Real-time optimization by extremum-seeking control*. John Wiley & Sons, 2003.

[10] S. Engell, “Feedback control for optimal process operation,” *Journal of process control*, vol. 17, no. 3, pp. 203–219, 2007.

[11] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[12] J. Hong, C.-C. Liu, and M. Govindarasu, “Integrated anomaly detection for cyber security of the substations,” *Smart Grid, IEEE Transactions on*, vol. 5, no. 4, pp. 1643–1653, 2014.

[13] A. Sari, “A review of anomaly detection systems in cloud networks and survey of cloud security measures in cloud storage applications,” *Journal of Information Security*, vol. 6, no. 02, pp. 142–154, 2015.

[14] W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen, “Effective detection of sophisticated online banking fraud on extremely imbalanced data,” *World Wide Web*, vol. 16, no. 4, pp. 449–475, 2013.

[15] X. Cao, Q. Li, and O. Miller, “A rfid-based anomaly detection approach for material supply of mixed-product assembly,” *International Journal of RF Technologies*, vol. 5, no. 3-4, pp. 183–201, 2013.

[16] V. Bamnett and T. Lewis, “Outliers in statistical data,” *Journal of the Royal Statistical Society*, vol. 141, no. 4, 1978.

[17] A. Loureiro, L. Torgo, and C. Soares, “Outlier detection using clustering methods: a data cleaning application,” in *Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector. Bonn, Germany*, 2004.

[18] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying Density-Based Local Outliers,” *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, jun 2000. [Online]. Available: http://portal.acm.org/citation.cfm?doid=335191.335388

[19] H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, “LoOP: local outlier probabilities,” *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 1649–1652, 2009. [Online]. Available: http://doi.acm.org/10.1145/1645953.1646195

[20] S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos, “Loci: Fast outlier detection using the local correlation integral,” *Data Engineering, 2003. Proceedings. 19th International Conference on*, pp. 315–326, 2003.

[21] F. Keller, E. Muller, and K. Bohm, “Hics: high contrast subspaces for density-based outlier ranking,” in *2012 IEEE 28th International Conference on Data Engineering*. IEEE, 2012, pp. 1037–1048.