Gaussian Processes for Anomaly Description in Production Environments

Christian Beecks
University of Münster and Fraunhofer Institute for Applied Information Technology FIT, Germany
christian.beecks@uni-muenster.de

Kjeld Willy Schmidt
University of Münster, Germany
kjeld.schmidt@uni-muenster.de

Fabian Berns
University of Münster, Germany
fabian.berns@uni-muenster.de

Alexander Grass
Fraunhofer Institute for Applied Information Technology FIT, Germany
alexander.grass@fit.fraunhofer.de

ABSTRACT
Concomitant with the rapid spread of cyber-physical systems and the advancement of technologies from the Internet of Things (IoT), many modern production environments are characterized by vast amounts of sensor data which are generated throughout different stages of production processes. This paper, we propose a novel method for discovering the inherent structures of anomalies arising in IoT sensor data. Our idea consists in modeling and describing anomalies by means of kernel expressions, which are combinations of well-known kernels. The results of our empirical analysis show that our proposal is suitable for modeling differently structured anomalies. Moreover, the results indicate that Gaussian processes provide a powerful tool for future algorithmic investigations of IoT sensor data.

1 INTRODUCTION
Concomitant with the rapid spread of cyber-physical systems and the advancement of technologies from the Internet of Things (IoT), many modern production environments are characterized by vast amounts of sensor data which are generated throughout different stages of production processes. These sensor data streams are often considered as valuable information sources with a high economic potential and are characterized by high volume, velocity and variety. Their data-driven value is indisputable for optimizing and fine-tuning industrial production processes.

Monitoring sensor data from complex production processes in order to detect outliers or low-performing production behavior caused by undesired shifts and trends, which we summarize as **anomalies**, is a challenging task. Not only due to the massive amount of sensor data but also due to different types of anomalies, which are potentially unknown in advance, manual or automatic inspection systems are frequently supported by anomaly detection algorithms. While the last years have witnessed the development of different anomaly detection algorithms, cf. the work of Renaudie et al. [21] for a recent performance evaluation in an industrial context, only less effort has been spent to the investigation of the inherent structure of an anomaly.

In this paper, we thus propose a novel method to discover the inherent structure of an anomaly. Our idea consists in modeling and describing anomalies by means of kernel expressions, which are combinations of well-known kernels. By fitting kernel expressions to the corresponding sensor data, we are able to decompose the inherent structure of an anomaly and to describe its individual behavior such as linearity and periodicity by natural language. For this purpose, we make use of Gaussian processes [20] and the Compositional Kernel Search model [11].

We carry out our analysis on the recently proposed IoT dataset [5], a real-world industry 4.0 dataset, which has been collected within the EU project MONSOON\(^1\). To sum up, we make the following contributions:

- We propose a machine-learning-based method in order to model anomalies and to describe their inherent components.
- We enrich the MONSOON IoT dataset with a novel ground truth derived from domain experts in order to further stimulate research of anomaly detection algorithms on this real-world dataset.

The paper is structured as follows. In Section 2, we outline related work. In Section 3, we briefly introduce Gaussian processes and their application to adapt kernel expressions to sensor data. The preliminary results of our proposed method are reported and discussed in Section 4, before we conclude our paper with an outlook on future research directions in Section 5.

2 RELATED WORK
Strongly related to our approach are anomaly detection algorithms. There is a plethora of these algorithms including Z-Score [10], Mahalanobis Distance-Based, Empirical Covariance Estimation [18] [9], Mahalanobis Distance-Based, Robust Covariance Estimation [22] [9], Subspace-based PCA Anomaly Detector [9], One-Class SVM [23] [18] [9] [12], Isolation Forest (I-Forest) [16] [18], Gaussian Mixture Model [18] [9] [19], Deep Auto-Encoder [8], Local Outlier Factor [7] [18] [9] [1], Least Squares Anomaly Detector [24], GADPL [14] and k-nearest Neighbour [13] [1] [12].

While these algorithms are all possible options for anomaly detection, as shown in different surveys such as [13], [19] and [9], they are not directly suited for describing the inherent structure of anomalies, which is the major focus of this paper. We choose the means of Gaussian processes for anomaly description due to their capability to not only gather statistical indicators, but deliver the very characteristics of specific anomalous behavior from the data [20].

For describing these characteristics, Lloyd et al. [17] have proposed the Automatic Bayesian Covariance Discovery System that adapts the Compositional Kernel Search Algorithm [11] by

---

\(^1\)www.spire2030.eu/monsoon
adding intuitive natural language descriptions of the function classes described by their models. In [15], these models are expanded to discover kernel structures which are able to explain multiple time series at once.

In this work, we make use of these algorithms in order to describe the inherent structures of anomalies, as shown in the following section.

3 GAUSSIAN PROCESSES

In this section, we describe the analysis of anomalies in sensor data via Gaussian processes. To this end, we assume the sensor data to be univariate and an anomaly \( A \) to be a finite subsequence of timestamp-value pairs \( A = \{(t_i, v_i)\}^{n}_{i=1} \) with timestamps \( t_i \in \mathbb{T} \) and values \( v_i \in \mathbb{R} \).

As we do not know in advance the number of values and the distances between individual timestamps, we can also thought of an anomaly \( A \) as a mathematical function \( A : \mathbb{T} \rightarrow \mathbb{R} \), which assigns every timestamp \( t \in \mathbb{T} \) a real-valued value \( v(t) \in \mathbb{R} \). By considering the individual values \( v(t) \) to be random variables following a Gaussian distribution, we can formalize the Gaussian process as

\[
v(t) \sim GP(m(t), k(t, t')),
\]

where \( m(t) = \mathbb{E}[v(t)] \) is the mean function and \( k(t, t') = \mathbb{E}[(v(t) - m(t)) \cdot (v(t') - m(t'))] \) is the covariance function \( k: \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R} \). In other words, a Gaussian process is a stochastic process over random variables, where every subset of random variables from the Gaussian process follows a normal distribution. The distribution of the Gaussian process is the joint distribution of all of these random variables and it is thus a probability distribution over (the space of) functions in \( \mathbb{R}^\mathbb{T} \).

While the covariance function \( k \) defined above is a general way to model the behavior of data, we aim to describe each anomaly \( A \) by its own covariance function \( k_A \). That is, we aim to learn a covariance function \( k_A \), which is then also denoted as kernel expression in the domain of machine learning, by fitting combinations of well-known kernels, such as

- the constant kernel \( k_C(t, t') = \lambda \in \mathbb{R} \),
- the linear kernel \( k_{\text{LIN}}(t, t') = (t - l) \cdot (t' - l) \),
- the squared exponential kernel \( k_{\text{SE}}(t, t') = \exp\left(-\frac{|t-t'|^2}{2\sigma^2}\right) \),
- or the periodic kernel \( k_{\text{PER}}(t, t') = \exp\left(\frac{2\sin^2\frac{|t-t'|}{\tau}}{\tau^2}\right) \).

In order to individually fit a kernel expression to each anomaly based on the aforementioned kernels, we use the compositional kernel model, as utilized for instance in [17]. This allows us to decompose an anomaly into individual components, which can be ranked by their contribution towards explaining the data. As an example, an anomaly \( A \) with a highly weighted linear kernel \( k_{\text{LIN}} \) indicates a hidden linearity component while a highly weighted periodic kernel \( k_{\text{PER}} \) indicates an inherent periodicity in the anomaly.

The resulting kernel expressions are reported and discussed in the next section.

4 PRELIMINARY RESULTS

In this section, we report and discuss the results of our preliminary performance evaluation. For this purpose, we use the recently introduced MONSOON IoT dataset [5] which comprises 357,383 data records in total. This dataset is based on a real production line of coffee capsules and the attribute under observation is the plastification time, that is the time which is needed to melt

![Figure 1: An example of the MONSOON IoT dataset with three anomalies.](image-url)
(plastify) the plastic melt for the actual injection molding cycle. More information about this process can be found in [3].

An overview of this attribute value, i.e. the pastification time, as a function of the cycle number is shown in Figure 1. As can be seen in the figure, while the normal pastification time is at approximately 4.2 seconds, it drops down to less than 3 seconds in case of an anomaly. Supported by domain experts, we figured out 28 anomalies in total in this dataset, of which three are shown in the above figure.

In the first series of experiments, we computed the best fitting kernel expressions by means of the ABCD algorithm. The results are shown in Table 1 for each anomaly. Together with the kernel expression of the corresponding anomaly, we also show the Bayesian Information Criterion (BIC) value which models the trade-off between model accuracy and size. As can be seen in the table, all anomalies are well described by their corresponding kernel expression (lower BIC values indicate better fit and vice versa). Surprisingly many kernel expressions do not show a linear component $k_{\text{LIN}}$, although some anomalies clearly show this linear tendency. We figure out that this is due to overfitting of the kernel expression in the ABCD algorithm. We aim to address this issue in future research.

In the second series of experiments, we evaluated how suitable a kernel expression of a certain anomaly fits to other anomalies. The results in form of the corresponding BIC values are summarized in Table 2. As can be seen in this table, kernel expressions of a certain anomaly do in general not fit to other anomalies. One reason for this behavior is the high degree of idiosyncrasy of the anomalies. Another reason might be the overfitting issue mentioned above.

To sum up, we have investigated the potential of describing anomalies in IoT sensor data by means of kernel expressions. Our preliminary results indicate that our proposal is well suited for this purpose. As one major challenge, we figure out that the problem of overfitting needs to be addressed in future research.

5 CONCLUSIONS AND FUTURE WORK
In this paper, we have addressed the problem of discovering the inherent structures of anomalies arising in IoT sensor data. To this end, we have proposed to model and describe anomalies by means of kernel expressions, which are combinations of well-known kernels. The results of our empirical analysis show that our proposal is suitable for modeling differently structured anomalies. Moreover, the results indicate that Gaussian processes provide a powerful tool for future algorithmic investigations of IoT sensor data.

In future work, we aim to address the problem of overfitting by modifying the grammar used within the ABCD algorithm for computing the kernel expressions. In addition, we aim to further develop our proposal in order to not only describe anomalies but also detect anomalies (which is not the focus of the current paper). For this purpose, we aim to measure similarity in IoT sensor data by incorporating Gaussian processes into adaptive distance-based similarity models, such as the Signature Matching Distance [6], and query processing algorithms [2, 4].

ACKNOWLEDGMENTS
The project underlying this paper has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 723650 (MONSOON). This paper reflects only the authors’ views and the commission is not

<table>
<thead>
<tr>
<th>Kernel Function</th>
<th>BIC Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>4.23</td>
</tr>
<tr>
<td>Quadratic</td>
<td>3.97</td>
</tr>
<tr>
<td>Cubic</td>
<td>3.85</td>
</tr>
</tbody>
</table>

Table 2: Evaluation of the BIC for every kernel expression against every anomaly.
REFERENCES


