Towards Gaussian Processes for Automatic and Interpretable Anomaly Detection in Industry 4.0

Fabian Berns¹, Markus Lange-Hegermann² and Christian Beecks¹

¹Department of Computer Science, University of Münster, Germany
²Department of Electrical Engineering and Computer Science, OWL University of Applied Sciences and Arts, Lemgo, Germany

Keywords: Anomaly Detection, Gaussian Processes, Explainable Machine Learning, Industry 4.0.

Abstract: Discerning unexpected from expected data patterns is the key challenge of anomaly detection. Although a multitude of solutions has been applied to this modern Industry 4.0 problem, it remains an open research issue to identify the key characteristics subjacent to an anomaly, sc. generate hypothesis as to why they appear. In recent years, machine learning models have been regarded as universal solution for a wide range of problems. While most of them suffer from non-self-explanatory representations, Gaussian Processes (GPs) deliver interpretable and robust statistical data models, which are able to cope with unreliable, noisy, or partially missing data. Thus, we regard them as a suitable solution for detecting and appropriately representing anomalies and their respective characteristics. In this position paper, we discuss the problem of automatic and interpretable anomaly detection by means of GPs. That is, we elaborate on why GPs are well suited for anomaly detection and what the current challenges are when applying these probabilistic models to large-scale production data.

1 INTRODUCTION

Anomaly detection is an important data mining process to distinguish expected from unexpected data patterns. It enables researchers as well as practitioners to assess the condition and current state of a system of interest (Chandola et al., 2009). Applications are manifold, e.g. in medicine (Rajpurkar et al., 2017), credit card fraud (Sorournejad et al., 2016), or network intrusion detection (Ioannou et al., 2017). Especially with regards to industrial applications, monitoring sensor data from complex processes in order to detect outliers or low-performing production behavior caused by undesired patterns and trends, which we summarize as anomalies, is a challenging task (Beecks et al., 2019). Not only due to the massive amount of sensor data but also due to different types of anomalies, manual or automatic inspection systems are frequently supported by anomaly detection algorithms (Stojanovic et al., 2016). It is often stressed that anomaly detection should be part of understanding the production process as a whole (Niggemann and Lohweg, 2015), i.e. that it is not sufficient to detect anomalies but also to generate hypothesis as to why they appear.

While the last years have witnessed the development of different anomaly detection algorithms (cf. the work of Renaudie et al. (2018) for a recent performance evaluation in an industrial context) only less effort has been spent on the investigation of the inherent structure of an anomaly. Utilizing techniques of machine learning and Artificial Intelligence (AI) for anomaly detection seems like a natural choice to not only detect unexpected patterns, but to understand them. Still, achieving the level of trustworthiness required for sensitive industrial applications is a non-trivial task (cf. High-Level Expert Group on Artificial Intelligence, 2019; Kwon et al., 2019).

With regards to the key requirements for trustworthy AI defined by the High-Level Expert Group on Artificial Intelligence (2019) of the European Commission, we consider GPs (Rasmussen and Williams, 2006) as an appropriate machine learning model to fulfill those demands. GPs deliver robust and reliable statistical data models, which are able to cope with unreliable, noisy, or partially missing data (Boškoski et al., 2012). Since these models are composed from simple components defined by explainable hyperparameters, they are interpretable by nature and deliver the capabilities to trace their predictions back to those
components and the according training data (Lloyd et al., 2014; Duvenaud et al., 2013).

In this position paper, we discuss the problem of automatic and interpretable anomaly detection by means of GPs. That is, we elaborate on why GPs are well suited for anomaly detection and what the current challenges are when applying these probabilistic models to large-scale production data. To this end, our contributions are two-fold:

- We introduce GPs as an interpretable model for anomaly detection.
- We outline challenges that have to be addressed in order to scale GPs to large-scale production data.

This position paper is structured as follows. We outline related work in Section 2. We introduce GPs in Section 3, while the concept and rationale for anomaly detection by means of these statistical data models and the remaining challenges in that field are explained in detail in Section 4. We conclude our paper with an outlook on future work in Section 5.

2 RELATED WORK

In the era of Industry 4.0, the field of anomaly detection has become crucially important. As a result, there is a plethora of classical anomaly detection algorithms that have been proposed in recent years such as Z-Score (Domingues et al., 2016), Mahalanobis Distance-Based, Empirical Covariance Estimation (Pedregosa et al., 2011; Chandola et al., 2009), Robust Covariance Estimation (Rousseeuw, 1984; Chandola et al., 2009), Subspace-based PCA Anomaly Detector (Chandola et al., 2009), One-Class SVM (Schölkopf et al., 2001; Pedregosa et al., 2011; Chandola et al., 2009; Eskini et al., 2002), Isolation Forest (I-Forest) (Liu et al., 2008; Pedregosa et al., 2011), Gaussian Mixture Model (Pedregosa et al., 2011; Chandola et al., 2009; Phua et al., 2010), Deep Auto-Encoder (Candell et al., 2018; Gong et al., 2019), Local Outlier Factor (Breunig et al., 2000; Pedregosa et al., 2011; Chandola et al., 2009; Auslander et al., 2011), Self-Organizing Maps (Von Birgelen et al., 2018), Least Squares Anomaly Detector (Tavallaee et al., 2010), GADPL (Graß et al., 2019), Automata (Vodencarević et al., 2011), and k-Nearest Neighbor (Goldstein and Uchida, 2016; Phua et al., 2010; Chandola et al., 2009), they are not directly suited for describing the inherent structure of anomalies, which is the major focus of this position paper. We choose GPs (Rasmussen and Williams, 2006) 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.

For automatically describing the underlying data characteristics, Lloyd et al. (2014) have proposed the Automatic Bayesian Covariance Discovery System that adapts the Compositional Kernel Search Algorithm (Duvenaud et al., 2013) by adding intuitive natural language descriptions of the function classes described by their models. Hwang et al. (2016) further expand on those concepts by expanding these models to discover kernel structures which are able to explain multiple time series at once. Recently, the 3CS (Berns et al., 2020) and LARGe algorithms (Berns and Beecks, 2020) have shown to outperform the aforementioned approaches in terms of efficiency. As these methods are all based on GPs, we give a short introduction of GPs in the following section.

3 GAUSSIAN PROCESSES

A Gaussian Process (GP) (Rasmussen and Williams, 2006) is a stochastic process over random variables \( \{ f(x) | x \in X \} \), indexed by a set \( X \), where every finite subset of random variables follows a multivariate normal distribution. The distribution of a GP is the joint distribution of all of these random variables and it is thus a probability distribution over the space of functions \( f : X \to \mathbb{R} \). A GP is formalized as

\[ f(\cdot) \sim GP(m(\cdot), k(\cdot, \cdot)), \]

where the mean function \( m : X \to \mathbb{R} \) and the covariance function \( k : X \times X \to \mathbb{R} \) are defined \( \forall x, x' \in X \) via

\[ m(x) = \mathbb{E}[f(x)] \]

(2)

\[ k(x, x') = \mathbb{E}[(f(x) - m(x)) \cdot (f(x') - m(x'))] \]

(3)
Given a finite dataset $D = \{X,Y\}$ with input $X = \{x_i \mid x_i \in X \land 1 \leq i \leq n\}$ representing the underlying index values, such as timestamps or locations, and target $Y = \{y_i \mid y_i = f(x_i), x_i \in X\}$ representing the actual data values, such as sensor values or other complex measurements, a GP can be used to statistically represent the dataset $D$ by optimizing the hyperparameters $\theta$ of both mean and covariance function. This optimization is frequently carried out by maximizing the log marginal likelihood $\mathcal{L}$ (Rasmussen and Williams, 2006; Kim and Teh, 2018) of the GP:

$$
\mathcal{L}(m,k,\theta \mid D) = -\frac{1}{2} \left[ (y - \mu)^T \Sigma^{-1} (y - \mu) + \log |\Sigma| + n \log(2\pi) \right]
$$

As can be seen in Equation 4, the marginalization of a GP for a given dataset $D$ of $n$ records results in a mean vector $\mu \in \mathbb{R}^n$, and a covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ which are defined as $y[i] = f(x_i)$, $\mu[i] = m(x_i)$, and $\Sigma[i,j] = k(x_i, x_j)$ for $1 \leq i, j \leq n$, respectively.

How GPs are utilized in the context of anomaly detection and how the resulting challenges can be solved are described in the following section.

4 GPs FOR ANOMALY DETECTION

GP models are robust Bayesian models that intrinsically prevent overfitting. This sturdiness is a clear advantage of GP models in dynamic and noisy production environments with always changing circumstances like evolving processes or sensors modifications. GP models yield smooth models even when only a few data points are given, which is particularly important in production, where the batch size is steadily decreasing.

In addition to recognizing an anomaly, it is important to understand and interpret its underlying structure. Only this makes it possible to rectify the underlying problem. Being interpretable is immanent in GPs, due to their rigid mathematical structure. Their covariance functions can induce combinations of various physically motivated behaviors like trends, periodicity, differential equations (Lange-Hegermann, 2018, 2020), and change points, and each of these covariance functions comes with hyperparameters, which are often interpretable physical constants like periods. The model prediction can even be disassembled into parts stemming from the individual interpretable parts of the covariance function. All this physically interpretable structure can be learned from data, and additionally domain knowledge can be pre-encoded into the models by experts.

What makes the application of GPs difficult to large-scale, streaming production data is the super-cubic computation time complexity of optimizing GPs. Though efficient solutions exist (Snellson and Ghahramani, 2007), they do not provide a fast solution for automatic and interpretable anomaly detection. For this reason, we identified the following major challenges.

4.1 Efficient Algorithms for Large-scale GP Models

The first challenge is concerned with the development of efficient retrieval algorithms for GP models. This includes the central question of how to efficiently search and determine suitable covariance functions for anomaly detection. Instead of applying a greedy search through the space of possible covariance functions, as done by state-of-the-art algorithms (Duvenaud et al., 2013; Lloyd et al., 2014; Steinruecken et al., 2019), one could learn the impact of individual hyperparameters in order to specifically derive new covariance functions. In addition to the development of such intelligent covariance search heuristics, another challenge is to process and infer GP models for multivariate, event-based sensor data in (near) real time. This demands for resource-efficient streaming GP retrieval algorithms that scale to state of the art big data processing frameworks such as Apache Hadoop, Spark, and Storm.

4.2 Model Selection for GPs in Anomaly Detection

Besides the development of efficient algorithms, the second major challenge lies in the development of suitable GP model selection approaches. While prominent model quality estimators, such as the likelihood function, tend to prefer complex models containing many hyperparameters, Laplace approximations enable to capture the model evidence of GP models more properly. The particular challenge is thus to couple Laplace approximations with unsupervised GP Latent Variable Models (Damianou et al., 2016) in order to enable conclusions on the underlying physical processes and individual sensor dimensions.

To sum up, the development of real-time GP model retrieval and selection algorithms for multivariate data streams that are open-source and are based on open industry standards is necessary for detecting, analyzing, and understanding anomalies in an domain-agnostic, automatic, and interpretable manner.
5 CONCLUSION

In this paper, we have argued for the application of GPs for automatic and interpretable anomaly detection. GPs are robust bayesian machine learning tools, which enable inference for noisy as well as unreliable data. GP models yield smooth models even when only a few data points are given, which is particular useful in industrial scenarios, where creating those records is potentially expensive.

Along with the advantages, several challenges have to be met. In particular, concepts and algorithms are needed to facilitate efficient GP model retrieval and selection on large-scale streaming data. We aim to address these challenges in our future work and to elaborate our developments in various application domains.

REFERENCES


anomaly detection. In 2016 International Joint Conference on Neural Networks (IJCNN), pages 1015–1022. IEEE.


