Exploring Methods to Apply Gaussian Processes in Industrial Anomaly Detection

Andreas Besginow
University of Applied Sciences and Arts
Department of Electrical Engineering and Computer Science
Lemgo, Germany
andreas.besginow@th-owl.de

Jan David Hüwel
Computer Science Department
Westfälische Wilhelms-Universität
Münster, Germany
jan.huewel@uni-muenster.de

Markus Lange-Hegermann
Unversity of Applied Sciences and Arts
Department of Electrical Engineering and Computer Science
Lemgo, Germany
markus.lange-hegermann@th-owl.de

Christian Beecks
Computer Science Department
Westfälische Wilhelms-Universität
Münster, Germany
christian.beecks@uni-muenster.de

***Index Terms*—Gaussian Process, Differential equations, Streaming methods, Interpretability, Explainable AI, Anomaly Detection**

I. Introduction

Machine learning describes the field of computers solving problems without explicit solutions. There are many different methods in machine learning that can be used, the best known being Neural Networks which, despite its popularity gives little to no insight into its inner workings. Contrarily, models like Gaussian Processes (GPs) [15] provide an innate interpretability. A GP can be understood as a probability distribution over the space of functions $f \sim \mathcal{GP}(\mu, k)$, completely determined by its mean function $\mu(x)$ and covariance function $k(x, x')$ (also called kernel). In practice, the covariance function is built from kernel functions that represent different patterns in the data. For example, the Linear (LIN) kernel describes linear trends while the Squared Exponential (SE) kernel describes smoothness of the data via its lengthscale hyperparameter. The interpretability of GP models mainly comes from knowing the kernel composition of its covariance function and understanding the corresponding hyperparameters. For instance, if the SE kernel's lengthscale is large the expected behaviour is a very slow change in value. This interpretability can be used in GP-based anomaly detection [5, 11, 8, 4, 2] to not only find an anomaly but also explain how it differs from standard system behaviour. To extend the work on GPs in anomaly detection we explore implementations of GPs which can be applied to datastreams (i.e. real-time capability) and the development of differentiable kernels. To enable the application of GPs in actual industrial use-cases the speed of GPs has to be increased significantly, despite current developments in sparse GPs [12, 14, 7, 6]. As for the differentiable kernels, we explore the cases where underlying physical processes in the form of differential equations dictate the system behaviour and through differentiable kernels in combination with multi-output GPs we can learn these behaviours.

Overall we aim to develop methods that advance the field of GPs in multiple different applications by providing new methods for common forms of data. In the following, we briefly present our current research.

II. Hybrid, physical-data driven GP models

In most production systems there is an underlying physical process that determines the system's behaviour, which can be written as differential equations. We plan to make use of this by exploring the application of differentiable covariance functions, based on theorethical works in this area [9, 10], on data from these production systems. By doing this we enable the GP models to learn the system's behaviour and make predictions about the next system state. But in contrast to applying the differential equations directly, the innate ability of GPs to handle noise makes the predictions more stable and better suited for real world applications, which behave different from the ideal scenarios required for differential equations. In later stages of the work we will further explore systems where the differential equations are not known and try to find a suitable representation for the system based on our model.

III. Stream-capable GPs

With the increase in online data generation and the construction of smart factories, real-time datastream processing rises in importance. Applying standard GPs to the problem is non-trivial since their evaluation cost lies in $O(n^3)$ and their required space is in $O(n^2)$ for n datapoints [15]. There are multiple works on sparse GPs [13, 12, 14, 7, 6], and while they are successful in accelerating calculations, their approximating nature can cause them to miss local trends and ignore small anomalies. On the other hand, window-based GPs [3, 11] can be unfit to model large-scale trends and their dependency on the window size combined with a high frequency of incoming data can cause the calculations to slow down significantly.

Although there has been research on methods to solve these problems [1], there is currently no way to use previous results to accelerate a kernel search on new data. In our research, we want to propose an algorithm that can adjust a kernel expression based on new information. The adjustment only happens when a drift in the data is detected. This guarantees fast processing while also delivering accurate and interpretable results. Our method in development uses the likelihood of a kernel expression to verify its continued validity and performs a window-based adjusting kernel search in the case of a drift.

IV. Conclusion

The GAIA project aims to make Gaussian Processes more applicable for anomaly detection in new areas like datastreams and production systems with underlying physical processes. With our current research into kernels derived from differential equations we will provide more accurate kernel expressions for many industrial systems, while our developments in stream-capable GPs will present a fast and interpretable algorithm for model selection. In future works we will expand and combine these methods and apply them in other areas of interest.

Acknowledgment

This research was supported by the research training group “Dataninja” (Trustworthy AI for Seamless Problem Solving: Next Generation Intelligence Joins Robust Data Analysis) funded by the German federal state of North Rhine-Westphalia.

References

[1] Fabian Berns and Christian Beecks. “Complexity-Adaptive Gaussian Process Model Inference for Large-Scale Data”. In: *Proceedings of the 2021 SIAM International Conference on Data Mining, SDM 2021, Virtual Event, April 29 - May 1, 2021*. SIAM, 2021, pp. 360–368. DOI: 10.1137/1.9781611976700.41.

[2] Fabian Berns, Markus Lange-Hegermann, and Christian Beecks. “Towards Gaussian Processes for Automatic and Interpretable Anomaly Detection in Industry 4.0”. In: *IN4PL 2020, Budapest, Hungary, November 2-4, 2020*. 2020. DOI: 10.5220/0010130300870092.

[3] Fabian Berns et al. “3CS Algorithm for Efficient Gaussian Process Model Retrieval”. In: *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*. IEEE, 2020, pp. 1773–1780. DOI: 10 . 1109 / ICPR48806 . 2021 . 9412805.

[4] Kai-Wen Cheng, Yie-Tarng Chen, and Wen-Hsien Fang. “Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 2909–2917.

[5] Minghao Gu, Jingjing Fei, and Shiliang Sun. “Online anomaly detection with sparse Gaussian processes”. In: *Neurocomputing* 403 (2020), pp. 383–399.

[6] James Hensman, Nicolo Fusi, and Neil D Lawrence. “Gaussian processes for big data”. In: *arXiv preprint arXiv:1309.6835* (2013).

[7] James Hensman, Alexander Matthews, and Zoubin Ghahramani. “Scalable variational Gaussian process classification”. In: *Artificial Intelligence and Statistics*. PMLR. 2015, pp. 351–360.

[8] Kira Kowalska and Leto Peel. “Maritime anomaly detection using Gaussian process active learning”. In: *2012 15th International Conference on Information Fusion*. IEEE. 2012, pp. 1164–1171.

[9] Markus Lange-Hegermann. “Algorithmic Linearly Constrained Gaussian Processes”. In: *NeurIPS*. 2018, pp. 2141–2152.

[10] Markus Lange-Hegermann. “Linearly Constrained Gaussian Processes with Boundary Conditions”. In: *CoRR* abs/2002.00818 (2020).

[11] Jingyue Pang et al. “Anomaly detection based on data stream monitoring and prediction with improved Gaussian process regression algorithm”. In: *2014 International Conference on Prognostics and Health Management*. IEEE. 2014, pp. 1–7.

[12] Joaquin Quinonero-Candela and Carl Edward Rasmussen. “A unifying view of sparse approximate Gaussian process regression”. In: *The Journal of Machine Learning Research* 6 (2005), pp. 1939–1959.

[13] Edward Lloyd Snelson and Zoubin Ghahramani. “Local and global sparse Gaussian process approximations”. In: *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, AISTATS 2007, San Juan, Puerto Rico, March 21-24, 2007*. Ed. by Marina Meila and Xiaotong Shen. Vol. 2. JMLR Proceedings. JMLR.org, 2007, pp. 524–531.

[14] Michalis Titsias. “Variational learning of inducing variables in sparse Gaussian processes”. In: *Artificial intelligence and statistics*. PMLR. 2009, pp. 567–574.

[15] Christopher K Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006.