Anomaly Detection and XAI Concepts in Swarm Intelligence

Mathias ANNEKEN, Manjunatha VEERAPPA, Nadia BURKART
Fraunhofer IOSB
Karlsruhe
GERMANY

{mathias.anneken, manjunatha.veerappa, nadia.burkart}@iosb.fraunhofer.de

ABSTRACT

For human operators in swarm intelligence, decision support in critical situations is crucial. The large amount of data shared by the autonomous systems, can easily make the human decision makers too overwhelmed by it and hence there is a need for support in analysing the data in an intelligent way. For this purpose, automatic systems for assessing situations and indicating suspicious behaviour or statistical outliers is employed. This strengthen their situation awareness as well as decrease the work load. Therefore, in this work, we emphasize that the data fusion services developed for detecting anomalies in surveillance tasks, e.g. in the maritime domain, can be adapted to support operators in swarm intelligence. Furthermore, in order to make the behaviour of the swarm and the results of the data fusion services understandable to the human operator, explainable artificial intelligence (XAI) concepts are introduced. This makes the autonomous system's behaviour more intelligible and understandable to humans by providing explanations for certain decisions.

1.0 INTRODUCTION

As drones become on the one hand more affordable and on the other hand more capable, they are used in an increasing amount of applications. In general, the necessary technologies for controlling one drone are rather mature, the same cannot always be said for a fleet or swarm of drones. For utilizing a swarm efficiently, decision making is necessary and important in two ways: Firstly, the swarm and its elements have to decide on a course of action, and secondly the operator has to identify his next actions based on the available information and either interact with the swarm or taking other next steps. The paper covers the latter part of decision making: the human in the loop and how to improve it.

One possible application for swarms is surveillance, e.g. in the maritime domain. A classical approach for surveillance is to install the necessary sensors stationary and fuse the available data in a common picture. While this approach might work quite well in many scenarios, in others a more dynamic approach is necessary. This involves non-stationary assets for collecting all necessary information. So far, this was done by only using a limited number of assets, making a larger 24/7 area coverage quite challenging. In these kinds of applications swarms can support in generating a holistic overview of the situation at hand. This would not only make the surveillance more cost efficient by utilizing cheaper assets, it will also decrease the needed personnel. But the latter is only true for the operation of the swarm itself, not for actually assessing the actual situation. This paper will give an overview on how modern artificial intelligence based methods can be used in order to support operators in gaining better situational awareness and how explainable artificial intelligence (XAI) can be used to make these results explainable and more interpretable for the operator.

The paper is structured as follows: After this brief introduction, anomaly detection algorithms for decision support applications are explained. Afterwards, the importance of explainable and interpretable results of artificial intelligence based algorithms are discussed. At the end we will summarize the findings.

2.0 ANOMALY DETECTION FOR DECISION SUPPORT IN SWARM INTELLIGENCE

While a swarm can be used to cover a larger area in surveillance tasks or in order to get a better understanding of the surroundings, it will also increase the amount of data to be processed by human operators and decision makers. In order to strengthen their situation awareness, automatic systems for assessing situations and indicating suspicious behaviour or statistical outliers is important. This will help to decrease the workload for the operator. For these automatic methods, one can look into the either already available or currently in development algorithms for supporting surveillance tasks, e.g. in the maritime domain. These approaches assume, that there is a huge amount of data available through a more classical setup of sensors. Here, these assets are complemented by intelligent swarms of drones. Let’s say, we have a swarm in support of maritime vessels, these vessels will collect not only the data available by their own sensor systems, but by all assets. The information gathered by all sources needs to be fused into one coherent picture. This should not be limited to the first level of JDL data fusion, but should include higher level data fusion processes in order to elicit the available information about all objects in the vicinity. This will include methods as described by Riveiro et al. [5] for detecting anomalies based on data-driven or signature-based approaches or a combination of both. These methods can not only detect positional or kinematic anomalies, but also anomalies considering the context (e.g. season, time of day, etc.) or complex moving patterns and interactions of multiple objects.

For detecting positional and kinematic anomalies, several approaches are discussed in the literature: Based on a statistical interpretation, outlier are in comparison to the normal behaviour observed in a monitored area are e.g. introduced in [12, 13, 17]. In [12], the conformal prediction framework is used to detect reliable and well calibrated predictions. In [13], an Ornstein–Uhlenbeck process is used to model vessel behaviour, resulting in better estimations for the position compared to similar approach in the literature. B-Splines are used in [17] to model recorded trajectories, which in turn are used to classify deviations as anomalies. A cluster analysis is e.g. used in [14, 15, 16] in order to extract the normal behaviour of vessels. In [14], the clustering is based on dividing the recorded trajectories into segments based on movement and stop parts. Furthermore, the movement parts are divided into linear segments. These segments are then clustered by using the OPTICS algorithm. The TREAD algorithm is introduced in [15] and is used to cluster the normal vessel behaviour based on a DBSCAN approach. The clustering in [16] is based on Fuzzy-k-Nearest-Neighbour and Fuzzy-c-Means. These algorithms are used to extract correlations between the trajectories, while fuzzy logic is used to model the uncertainties in the trajectories. In [18], positional and kinematic anomalies are extracted by utilizing a deep learning approach based on a recurrent neural network.

While these approaches will enable a decision support system to identify some anomalies, based solely on the kinematics of a vessel, this approach is still quite limited in more complex scenarios. Thus, anomaly detection algorithms, which include the context around the vessel, are introduced. The TREAD algorithm as introduced in [15] is extended in [19] by explicitly modelling IMO regulations for maritime vessel traffic. In [21], a recurrent neural network is given not only the positional information of a vessel, but also contextual information about harbours in order to make a better estimation regarding service time of vessels in said harbour. A multi-agent approach is used in [22] in order to predict vessel behaviour an calculate an anomaly score for each vessel. The approach is based on a game-theoretic foundation. The type of a vessel is derived by its moving patterns using a convolutional neural network with positional information together with geographical features in [27].

Furthermore, in order to identify complex anomalies, especially rule-based, probabilistic and multi-agent models are introduced. These models not only integrate contextual information but try to capture the important information in order to assess complex situations. In [28], a multi-agent system is introduced, which defines different roles as agents (e.g. for smuggling). These are used to estimate the probabilities for situations. Other approaches define probabilistic graphical models, e.g. dynamic Bayesian networks in [29], in order to model situations of interest based on expert knowledge. Compared to the multi-agent approach, these models give directly the probability for certain situations.

All these above-mentioned methods can be used in order to strengthen the situation awareness of an operator: The automatic extraction of information from incoming data streams is crucial whenever the amount of data can overwhelm a human operator, especially in critical situations. Some of these algorithms are black-box models, thus quite complicated for an operator to be completely understood and furthermore make its usage in critical situations challenging. A possible way to counter this problem is XAI. Possible ways to circumvent this challenge will be discussed in the next section.

3.0 IMPORTANCE OF EXPLAINABLE ARTIFICIAL INTELLIGENCE

A further way to strengthen the situation awareness is by adding the concept of eXplainable Artificial Intelligence (XAI). The application of XAI makes the autonomous system's (swarm robots, UxVs) behaviour more intelligible and understandable to humans by providing explanations for certain decisions. It should be possible for the XAI system to clarify its decision-making to explain what it has done, what it is doing now, and what will happen next. This approach is very relevant to situation awareness which involves the ability to perceive, interpret and predict future events. It intends to provide ethics, privacy, confidence, trust, and security [2].

The ability to explain why a decision was made is a crucial aspect of system intelligence. This is because the explanation of one's decision might be a prerequisite to establish a trust relationship between the autonomous system and the decision maker. There exists already much work on categorizing and developing XAI models from a cognitive science perspective. For instance, works such as [6, 7, 8] concern on how to unbox the artificial intelligence (AI) models that are considered as black-box models. Another work [26] presented the type of explanations needed for the Human-Robotic Interaction (HRI). These XAI models are useful for producing explanations that can be easily understood by the user for a given application.

In general, XAI systems can be categorized according to various criteria. Based on the scope of the explanation, XAI systems deliver two kinds of explanations: local and global explanations. Local explanations, in particular, explain a single prediction result over the entire model, i.e., it explains the conditional interaction between dependent and independent variables with respect to the single prediction. Global explanations explain the behaviour of the entire model, e.g. in the form of rule lists.

Further, XAI models can be classified as model-specific and model-agnostic interpretation. While model-specific techniques are limited to certain mathematical models such as sparse linear models, model-agnostic techniques can be applied to any type of model. In this work, we focus on model-agnostic approaches.

As specified in section 2.0, higher-level data fusion services are required to strengthen the situation awareness in swarm intelligence. That means, two or more modalities are combined to develop an AI agent which is then used for prediction or anomaly detection. Deep learning is one of the architectures widely used in order to achieve this objective. However, it has a few limitations. One of these is a fact that their decision-making is opaque and it is very hard for the user to understand the reason behind their prediction. Therefore, in order to make the results of such AI agents understandable and interpretable, we imply the following well-known XAI approaches for the surveillance tasks [9, 10].

3.1 Feature attributions

Feature attributions illustrate the contribution of each feature towards the model's prediction for a given instance. This shows the relationship between a feature and the prediction. As a result, users will be able to understand which features their network relies on. The two most prominent feature attribution methods are Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME).

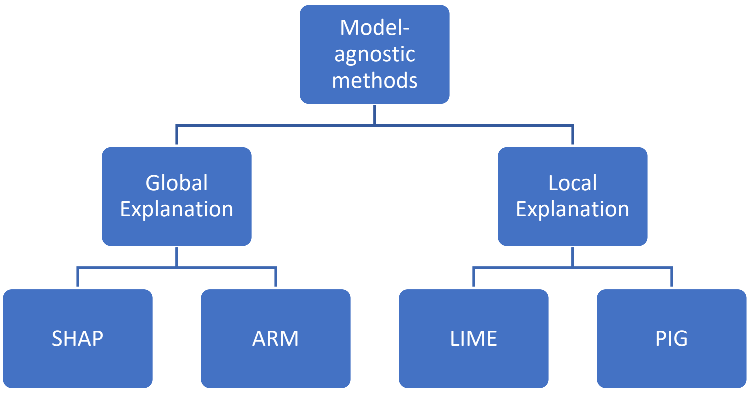

Figure 1: Model-agnostic methods categorization

SHAP [8] is a model-agnostic global approach, which evaluates the contribution of a feature to the overall prediction. It is primarily based on Shapley value from the concept of "cooperative game theory", which indicates what reward players can expect depending on a coalition function. To extend this approach to the explainability of AI agents, players are considered as features and reward as the outcome of the model. SHAP provides global interpretability (behaviour of the entire model) as well as local interpretability (behaviour of a single prediction). On the other hand, LIME [7] is a model-agnostic local approach suitable only for local explanations. This method perturbs the input in the neighbourhood of an instance and examines the output of the model. Thus, indicates the input features the model considers when making a prediction.

3.2 Path attributions

Path attribution methods explain the output of the model that is based on gradients. That means, the contribution of each feature is computed by aggregating the gradients from baseline values to the current input along the path. One such method is Path Integrated Gradients (PIG) [11], which determines the features contributing the most towards the model prediction. This assists users in comprehending the reasoning for the decision.

3.3 Association Rule Mining

Association Rule Mining (ARM) is one of the many ways to discover patterns in data, which finds correlations and co-occurrences between features in a large dataset. They are considered as most interpretable prediction models with their simple if-then rules. A rule is essentially an if-then statement with two components: an antecedent and a consequent. The input feature with a condition is an antecedent and a prediction is its consequent. The popular techniques to extract the rules from a large dataset are Scalable Bayesian Rule Lists (SBRL) [20], Gini Regularization (GiniReg) [23] and Rule Regularization (RuleReg) [24]. All three techniques are suitable for the classifiers in surveillance tasks as presented in [25].

In ARM, a global surrogate model is trained to approximate the black-box model's predictions. Then the conclusions are drawn by interpreting the surrogate model. SBRL trains an interpretable surrogate model using Bayesian rule lists algorithm, while GiniReg, a regularization technique that allows gradient-based optimization in order to train a surrogate model. The RuleReg method fits the global surrogate model using a regularization term that acts as a degree of explainability.

With the help of the explanations from the above-mentioned approaches, the decision makers can understand how autonomous systems such as swarm robots, UxVs work, and can better control, communicate with and use them in challenging environments.

4.0 CONCLUSION

As swarm robots work without any supervision, it is important for them to be able to explain their decisions in a user understandable way. Therefore, this work has discussed some procedures that can be considered in order to improve the user understanding and the trust towards swarm intelligence. For instance, a swarm of UxVs will provide capabilities to monitor larger areas with better coverage. While this will already be a huge advantage for human operators, they might get overwhelmed by the sheer amount of additional information. Therefore, automatic systems, like higher level data fusion services for anomaly detection are needed. In order to make the behaviour of the swarm and the results of the data fusion services understandable to the human operator, XAI methods are needed to gain better inside into the reasoning of the systems. Explainable AI has made it possible to examine AI models using visuals, text, and rules in the form of local and global explanations. However, an approach to assess the quality of explanations is needed, which would be our future work.

REFERENCES

[1] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human factors, 37(1), 32-64.

[2] Kiesenberg, P., Weippl E. W., Holzinger, A. (2016). Trust for the doctor-in-the -loop. European Research Consortium for Informatics and Mathematics (ERCIM) News: Tackling Big Data in the Life Sciences. 104(1), pp. 32–33

[3] Iba, Hitoshi. (2019). AI and SWARM Evolutionary Approach to Emergent Intelligence.

[4] Bouvry, P., Chaumette, S., Danoy, G., Guerrini, G., Jurquet, G., Kuwertz, A., ... & Sander, J. (2016, September). Using heterogeneous multilevel swarms of UAVs and high-level data fusion to support situation management in surveillance scenarios. In 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (pp. 424-429). IEEE.

[5] Riveiro, M.; Pallotta, G. & Vespe, M.: Maritime anomaly detection: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Wiley, 2018, 8

[6] Burkart, N., Huber, M.F., 2021. A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research70, 245–317.

[7] Ribeiro, Marco & Singh, Sameer & Guestrin, Carlos. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 1135-1144. 10.1145/2939672.2939778.

[8] Lundberg, Scott & Lee, Su-In. (2017). A Unified Approach to Interpreting Model Predictions.

[9] Burkart, Nadia & Huber, Marco & Anneken, Mathias. (2021). Supported Decision-Making by Explainable Predictions of Ship Trajectories. 10.1007/978-3-030-57802-2_5.

[10] Veerappa, Manjunatha & Anneken, Mathias & Burkart, Nadia. (2021). Evaluation of Interpretable Association Rule Mining Methods on Time-Series in the Maritime Domain. 10.1007/978-3-030-68796-0_15.

[11] Sundararajan, Mukund & Taly, Ankur & Yan, Qiqi. (2017). Axiomatic Attribution for Deep Networks.

[12] Laxhammar, R. (2014). Conformal Anomaly Detection - Detecting abnormal trajectories in surveillance applications. University of Skövde, University of Skövde, 2014

[13] Millefiori, L. M.; Braca, P.; Bryan, K. & Willett, P. (2016). Long-term vessel kinematics prediction exploiting mean-reverting processes. 2016 19th International Conference on Information Fusion (FUSION), 2016, 232-239

[14] Guillarme, N. L. & Lerouvreur, X. (2013). Unsupervised Extraction of Knowledge from S-AIS Data for Maritime Situational Awareness. Information Fusion (FUSION), 2013 16th International Conference on, 2013, 2025-2032

[15] Pallotta, G.; Vespe, M. & Bryan, K. (2013). Traffic knowledge discovery from AIS data. Proceedings of the 16th International Conference on Information Fusion, 2013, 1996-2003

[16] Shao, H.; Japkowicz, N.; Abielmona, R. & Falcon, R. (2014). Vessel Track Correlation and Association using Fuzzy Logic and Echo State Networks. Evolutionary Computation (CEC) 2014, IEEE Conference on, 2014

[17] Anneken, M.; Fischer, Y. & Beyerer, J. (2016). Anomaly Detection using B-Spline Control Points as Feature Space in Annotated Trajectory Data from the Maritime Domain. Proceedings of the 8th International Conference on Agents and Artificial Intelligence, 2016, 2, 250-257

[18] Nguyen, D.; Vadaine, R.; Hajduch, G.; Garello, R. & Fablet, R. (2019). GeoTrackNet-A Maritime Anomaly Detector using Probabilistic Neural Network Representation of AIS Tracks and A Contrario Detection. arXiv preprint arXiv:1912.00682, 2019

[19] Liu, B.; de Souza, E. N.; Matwin, S. & Sydow, M. (2014). Knowledge-based clustering of ship trajectories using density-based approach. 2014 IEEE International Conference on Big Data (Big Data), 2014, 603-608

[20] Yang, Hongyu & Rudin, Cynthia & Seltzer, Margo. (2016). Scalable Bayesian Rule Lists.

[21] Abualhaol, I.; Falcon, R.; Abielmona, R. & Petriu, E. (2018). Data-Driven Vessel Service Time Forecasting using Long Short-Term Memory Recurrent Neural Networks. 2018 IEEE International Conference on Big Data (Big Data), 2018, 2580-2590

[22] Anneken, M.; Fischer, Y. & Beyerer, J. (2017). A Multi-agent Approach to Model and Analyze the Behavior of Vessels in the Maritime Domain. Proceedings of the 9th International Conference on Agents and Artificial Intelligence, ICAART 2017, Volume 1, Porto, Portugal, February 24-26, 2017., SCITEPRESS - Science and Technology Publications, 2017, 200-207

[23] Burkart, Nadia & Faller, Philipp & Peinsipp-Byma, Elisabeth & Huber, Marco. (2020). Batch-wise Regularization of Deep Neural Networks for Interpretability. 10.1109/MFI49285.2020.9235209.

[24] Burkart, Nadia & Huber, Marco & Faller, Phillip. (2019). Forcing Interpretability for Deep Neural Networks through Rule-Based Regularization. 700-705. 10.1109/ICMLA.2019.00126.

[25] Veerappa, Manjunatha & Anneken, Mathias & Burkart, Nadia. (2021). Evaluation of Interpretable Association Rule Mining Methods on Time-Series in the Maritime Domain. 10.1007/978-3-030-68796-0_15.

[26] Sheh, R. (2017). “Different XAI for Different HRI.” AAAI Fall Symposia (2017).

[27] Anneken, M.; Strenger, M.; Robert, S. & Beyerer, J. (2020). Classification of Maritime Vessels using Convolutional Neural Networks. Artificial Intelligence: Research Impact on Key Industries; the Upper-Rhine Artificial Intelligence Symposium (UR-AI 2020), 2020, 103-114

[28] Brax, C. (2011). Anomaly detection in the surveillance domain. University of Skövde, University of Skövde, 2011

[29] Anneken, M.; de Rosa, F.; Kröker, A.; Jousselme, A.-L.; Robert, S. & Beyerer, J. (2019). Detecting illegal diving and other suspicious activities in the North Sea: Tale of a successful trial. 2018 20th International Radar Symposium (IRS), 2019