Adaptive Anomaly Detection and Root Cause Analysis by Fusing Semantics and Machine Learning

Chapter · August 2018

DOI: 10.1007/978-3-319-98192-5_46

1 author:

Bram Steenwinckel
Ghent University
38 PUBLICATIONS **356** CITATIONS

SEE PROFILE

Adaptive anomaly detection and root cause analysis by fusing semantics and machine learning

Bram Steenwinckel[0000−0002−3488−2334]

Ghent University - imec, IDLab, Ghent, Belgium
Bram.Steenwinckel@ugent.be

Abstract. Anomaly detection (AD) systems are either manually built by experts setting thresholds on data or constructed automatically by learning from the available data through machine learning (ML). The first requires profound prior knowledge and are non-adaptive to changing environments but can perform root cause analysis (RCA) to give an understanding of the detected anomaly. The second has a huge need for data, is unable to perform RCA and is often only trained once and deployed in various contexts, leading to a lot of false positives. Fusing the prior knowledge with ML techniques could resolve the generation of these alarms and should define the causes. The primary challenges to create such a detection system are: 1) Augmenting the current ML techniques with prior knowledge to enhance the detection rate. 2) Incorporate knowledge to interpret the cause of a detected anomaly automatically. 3) Reduce of human-involvement by automating the design of detection patterns.

Keywords: Anomaly Detection, Root Cause Analysis, Machine learning, Expert knowledge, Semantic Web, Knowledge Graphs

1 Introduction

In recent years, there is an increasing interest in Internet-connected devices and sensors, called the Internet of Things (IoT). These IoT devices continuously generate data that describe their state and their context or environment. Sensor monitoring systems have found their way into almost all industries and a variety of research fields and applications such as transportation [5] and healthcare [19]. Such systems can yield valuable insights into a company's physical assets and the interaction between these assets. However, awareness is growing across industries that strategically placed sensors have small added value without data analysis. Companies that invest in and successfully derive value from their data hold a distinct advantage over their competitors [29]. Both Anomaly detection (AD) and root cause analysis (RCA) are methods to investigate irregularities in the data. They are becoming more accessible as more relevant data is generated and tools for data analysis becoming widely available.
AD is the identification process of events or observations, which do not correspond to an expected pattern or other items inside a dataset [19]. RCA helps to

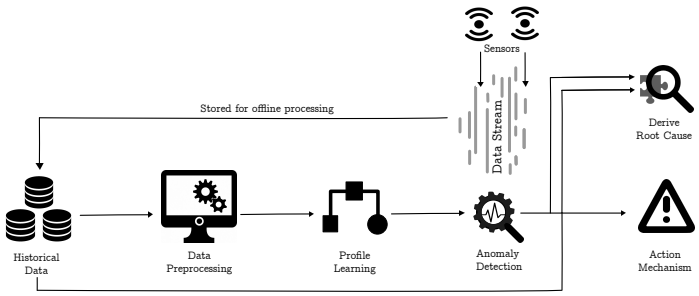

Fig. 1. Overview of the workflow of current anomaly detection systems

guide the problem solver understand the real causes of detected anomalies [17]. The detection process visualised in Figure 1 represents the usual workflow. First, historical data from different sources are used in a préprocessing step to make sure that all the available records are uniform. Based on this cleaned data, algorithms will learn the regular patterns and will detect the most relevant characteristics. This pattern can now be used to identify different anomalies in newly, unseen data. When the learned patterns diverge from this new data, an action mechanism will be able to alert this anomalous behaviour, or the cause of this unusual event can be investigated to resolve it. More concrete, suppose for example a fully automated ventilation system available in modern houses today. Historical sensor data will be used to determine the average levels of CO_2. New sensor data will be used to determine which room is currently underventilated and the system acts by adapting the fan speed to resolve the high level of CO_2. Modern techniques monitor the household's behaviour to react to the possible causes of the detected anomalies. In our example, self-learning techniques optimise the performance of the fan to get the house fully ventilated when people arrive after work.

However, the AD and RCA tools of today have difficulties to adapt to changing behaviours. If in our ventilation example, a person works at home, the ventilation system will detect abnormal behaviour and will be unable to react to this new situation or is even unable to derive the cause if it is not explicitly programmed or taught. Human-involved tuning is, therefore, frequently needed to adapt these systems to multiple environments. Sensors can produce new information at a fast pace, while enhanced analysis is needed to investigate whether current observations are anomalous, resulting in many false alerts and undetected events [6].

There is a need for adaptive, but still, accurate AD and RCA system that can take directly into account the prior knowledge to optimise the detection of anomalies and identify their causes. In our example, work schedules and agendas could be used as prior knowledge to optimise the ventilation process.

In summary, the challenges this research will tackle are: 1) Reducing the number of falsely generated anomalies by incorporating available prior knowledge 2) Automatic determination of the most plausible cause of the detected anomalies, to give additional interpretability to human operators. 3) Adaptively change

the detection behaviour to a various number of contexts to reduce the human-involvement during the design of such detection techniques.

2 State of the Art

Different research domains make use of the available prior knowledge to improve AD and RCA. This section gives an overview of 1) learning models which uses the available knowledge as input. 2) techniques which work directly with the available information. 3) rule-based detection mechanisms.

2.1 Knowledge incorporation in data-driven AD and RCA

Approaches for detecting anomalies in a dataset which do not require any predefined rules, models, or prior knowledge limit the efforts needed for systems designed by experts [8]. Most of these approaches are based on machine learning (ML) techniques and can process vast amounts of data. ML models can be supervised or unsupervised, based on the availability of labelled data. Labelling a significant amount of data for a domain-specific problem requires much human involvement. Therefore, most AD problems belong to the category of unsupervised learning due to unpredictable aspects of the data. In critical domains, where faults have a significant impact, the primary goal of AD is not the speed of the detection, but the accuracy or the reduction of the false negative and positive rates [23].
The goal of AD detection technique is, therefore, to model the normal behaviour. Statistical tests can be devised to determine if this behavioural model explains the data samples, uncovering both temporal and spatial anomalies when it does not succeed [20, 28]. The detected anomalies are mostly hard to interpret [22]. RCA techniques are therefore based on detection models using tree structures and logics [2, 30]
Knowledge nowadays is usually represented as a mesh of information, linked up in such a way that it should be interpretable by machines. Such a mesh of information is more generally known as the semantic web [3]. Many of the semantic concepts inside various domains are described in so-called ontologies, providing structured relations and the ability to reason on these concepts. Data annotated by these ontologies is stored using node-edge triples in a knowledge graph, relating prior information over multiple domains [14]. ML methods, in general, are currently not able to take advantage of these graphical knowledge representations. Therefore, techniques to transform graphs into a vectorial representation are becoming more popular, resulting in embedding techniques [11]. Knowledge graph embeddings usually map entities and relations to a vector space and predict unknown triples by scoring the candidate triples [21]. Embeddings are mostly designed to perform a single statistical relational learning task, like predicting missing edges or predicting properties of nodes [12]. Recently designed embedding techniques transform the graph triples (subject, object and relation pairs) directly into vectors which can be reused for various tasks [15]. These more general embeddings are particularly attractive for cross-domain knowledge

graphs, which can be used in a variety of scenarios and applications. Constructing embeddings for dynamic knowledge graphs is, however, still problematic. High variable behaviour, such as in sensor data streams, usually involve recalculations due to the changed graphical representation.
Embedding techniques used in combination with the traditional ML techniques usually have a low level of interpretability because decisions are based on the vectors themselves, not on the interpretable initial graphical data. Techniques to resolve this loss in interpretability are usually expensive and do not scale for large graphs [25]. Song et al. [18] gave a broad overview of how to use the existing general-purpose knowledge to enhance the ML processes, by enriching the features or reducing the labelling work using prior knowledge. No efforts within this research domain are taken to use such an approach for developing AD systems to our knowledge, probably due to the unsupervised nature of the original detection problems.

2.2 Knowledge-based machine learning

While embeddings translate the information into a manageable form for which we already have many methods available, techniques exist to learn directly over the knowledge graphs without any loss of information due to embedding transformations. One such technique is the Relational Graph Convolutional Network (RGCN), a method similar to neural networks but operating on graphs, developed specifically to deal with the highly multi-relational data characteristic of realistic knowledge graphs [10]. Another research area focuses on the development of predicate descriptions, using the available data and the existing prior knowledge. This Inductive Logic Programming (ILP) techniques are based on sound principles from both Logic and Statistic. Other combinations of ML techniques and prior knowledge models exist [1, 5], but none of them is currently adapted to work with sensor streams or highly variable data because the evaluation of the prediction also requires the additional prior knowledge.

2.3 Defining prior knowledge into rule-based systems

Rule-based detection systems utilising expert information have the advantage of being explainable and can determine the cause of the problem. They are however language dependent, do not scale with the increasing amount of data, and the development is time-consuming because much human involvement is needed [1, 17]. Detection systems use techniques which track unwanted patterns in a data stream to provide more scalable solutions to the highly variable data of today. Complex Event Processing (CEP) can be used to identify these abnormal events using pattern matching techniques such as rule-based, model-based or parametric statistical approaches [19]. These approaches, however, lack of expressiveness and flexibility to cope with complex events in different situations or different contexts. Therefore, semantic complex event processing (SCEP) proposes the semantic enrichment of the event streams, in which derived events are added in addition to the already observed pattern [19]. SCEP has been used in diverse applications comprising a variety of complex events including security and threat

detection events [7, 13], sensor networks [4, 26] and eHealth or ambient assisted living [16, 27]. SCEP systems make self-constrained decisions using a rule base, making them ideal candidates to perform RCA in data streams. Despite the benefits of SCEP, most patterns are static, and the anomalies must be defined upfront to work correctly. It requires some human involvement to adapt and update these patterns inside the multiple processing units [19].

3 Problem Statement

By analysing the state of the art methods, the following open problems can be identified:

P1 Current AD techniques only use the data itself to determine the occurrence of the unwanted behaviour.
P2 The frequently used accuracy metric misleads the functioning of the models due to the high impact of the falsely generated alarms.
P3 AD and RCA techniques are optimised for offline purposes, making them inappropriate to work with variable data, such as streaming environments.
P4 AD en RCA models are usually trained once and are therefore hard to adapt to new contexts, sensors or environments.
P5 The design of RCA models for a specific domain requires much human involvement.
P6 Most AD methods do not make interpretable decisions, reducing the ability to perform RCA.

From this, the following hypotheses can be deducted:

H1 Incorporating prior knowledge in learning and reasoning algorithms will outperform the detection rate of original AD ML techniques by at least 1% in real-life cases.
H2 The F1-score, which relates the number of false negatives and false positives, will be increased at least by 3% by incorporating the prior knowledge.
H3 Techniques which are adaptable to changing environments will reduce the human involvement by more than 50%.
H4 Designed techniques must be applicable in streaming contexts, without causing any data-driven congestions.

The following research question will be resolved to deliver the hypotheses proves:

Q1 Can prior knowledge, in the form of knowledge graphs or linked datasets, be incorporated as simple input features in the currently existing AD ML models to improve the detection of both false positives and false negatives?
Q2 Can current AD outcomes be transformed to enable RCA-based reasoning for finding the cause with the highest probability of an anomalous observation or be representative for the decision they make?
Q3 Is it possible to reduce the human involvement by deriving explainable rules from existing AD models inside a data stream and detect newly derived types of events without retraining or increasing the computational costs?

4 Research Methodology and Approach

A system which fuses both ML and semantics will be designed to improve the detection of anomalies together with the ability to determine their causes inside a stream of data accurately. An overview of such a system is given in Figure 2. Prior knowledge will be used to derive rule patterns directly from the data stream and improve both the AD and RCA to address the research questions defined in Section 3. How this prior knowledge is incorporated in each of these three parts is discussed in the following sections.

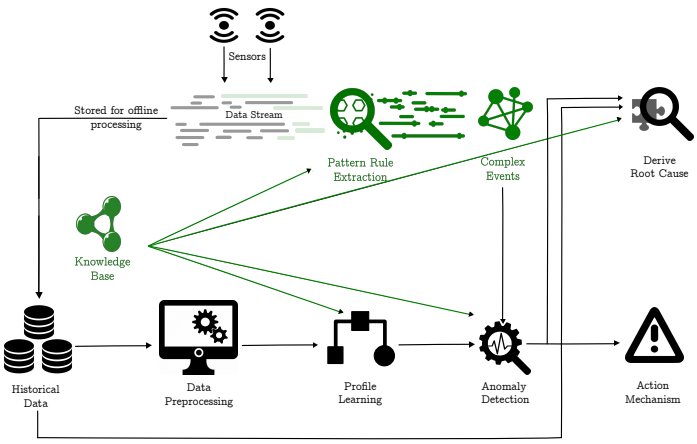

Fig. 2. Overview of the enhanced AD and RCA system

4.1 Improved feature selection for enhanced AD

Embeddings can be used to incorporate prior knowledge into ML models as discussed in Section 2.1. These embeddings can be used as features for anomaly-based ML systems but will operate in a pipeline of discrete steps. More concretely, the detection models will no further improve once the feature vectors are extracted because the error signal from the performed detection task can no longer be used to fine-tune the extraction step further. Instead of using the embedded representation of the knowledge graph directly, a new technique will transform the knowledge graph directly into a matrix formation. Rows will represent subject-object pairs, while the columns represent the relation types. Analysing this matrix can require some computational effort because the computations scale linearly with the number of cells or thus the number of links within the knowledge graph. There is a high probability that important information is scattered all over the matrix. The proposed technique will extract information from this matrix by adaptively selecting a sequence of regions of interest. These regions of interest represent the cells within the matrix with the most valuable information concerning the anomalous links. A (bandit) reinforcement learning

(RL) agent is, therefore, an excellent candidate to control the choice of the region of interest, as it can work with partially available information. The agent will select actions related to the number of regions and the location in the matrix. The feedback on the correctly detected anomalies will improve the region selection process of the RL agent. Figure 3 gives an overview of this process. The detection rate can be improved because the extracted information only focusses on the informative links within the available prior knowledge.

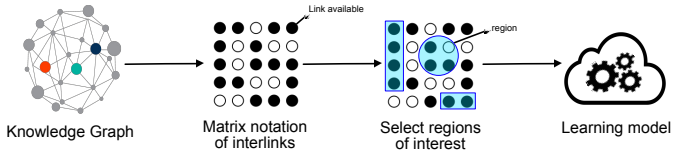

Fig. 3. Overview of the matrix knowledge learning process.

4.2 Interpretable knowledge for RCA

ML-based detections reveal the relations between the selected features and the provided outcome. Most of these feature vectors have, however, a low level of interpretability, reducing the capability of determining the underlying cause of the detected anomaly. The generated vectors, both from Section 4.1 and the previously mentioned embedding techniques, are called black-box features due to the reduced interpretation of the generated values. Research is needed to reproduce valuable information, which was available in the original knowledge graph, from these embedded vectors. One method based on Generative Adversarial Networks (GAN) can be used to transform the vectors back into an interpretable graph, giving back the power to determine the cause of an anomaly. In such a GAN network, the generator network constructs a graphical representation from a generated vector and inputs these graph structures to the discriminative network. The discriminative system is supposed to detect whether the structure generated by the generative network resembles a part of the original knowledge graph. Both networks update their performance until a low number of faults are generated, and the discriminative system has difficulties in finding differences between fake subgraphs from the original subgraphs. An overview of such a GAN process is given in Figure 4. Further analysis to determine the cause of the detected anomalies is possible with these embedding interpretations.

4.3 Adaptive detection and analysis for streaming data

Techniques to incorporate prior knowledge directly into data streams resulted in the design of SCEP systems. Problems arise when many different anomalous events need to be tracked. Rules are human maintained and can contradict. The correct functioning of the system can, therefore, not be guaranteed. Rules should be able to directly derivable from the ML-learning techniques used for the detection of anomalies in Section 4.1. White-box models can be translated

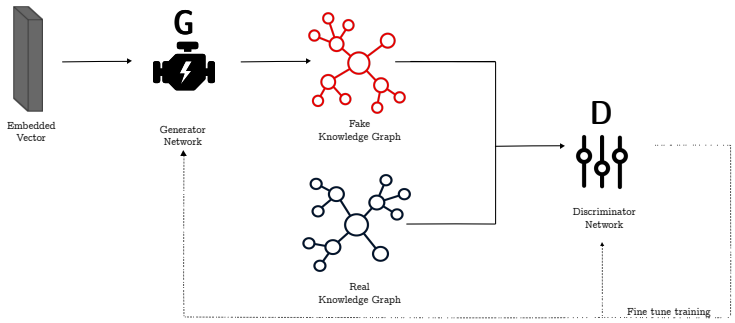

Fig. 4. Overview of knowledge graph embedding using a Generative Adversarial Network.

easily into rules and benefit from the ease of interpretability. In contrast, black-box models do not have this interpretability, but methods exist to convert these models to a set of rules [24]. To cope with the adaptive character of adding and removing these generated model-based rules, a RL agent will decide which rules to activate. The pattern rule extractor in Figure 2 will still test a subset of rules (actions) using this approach, while adaptations and further improvements are ensured. Feedback based on the number of rules or the complexity of the rule tests guarantees the efficiency of the RL agent. An overview of the learning agent is given in Figure 5. The designed technique will be able to operate in a changing environment where high variable data, such as in data streams, need to be analysed. Automatic derivation of simple rules reduces the human involvement and can be explainable, making RCA possible after the detections took place.

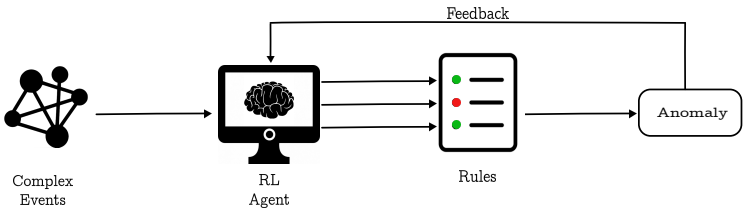

Fig. 5. RL agent for rule selection.

5 Evaluation Plan

Each phase of Section 4 can be evaluated separately. The proposed technique of Section 4.1 will be compared with existing embedding techniques, to show the advantages. A comparison using the prior knowledge directly and the technique proposed in Section 4.2 will reveal the benefits of fusing semantics with ML for RCA. At last, evaluation of models using the original rule sets and the technique proposed in Section 4.3 will be made to determine their adaptiveness and scalability. All this will be done using standardised benchmark RDF datasets

used for classification purposes in [15], but adapted here to detect the minority classes as anomalies. The functioning of the full system will be tested using two different proof of concepts:

- **Pervasive Healthcare:** In the eHealth domain, the available data from intelligent devices and sensors correlate to prior knowledge. Profile information of the patients can improve the detection of anomalies. The imec SWEET study[1] is such a case were stress analyses with sensor data can be improved by incorporating additional context. Data will be used during this research containing both raw sensor data available from wearables and context parameters based on the person's habits.
- **Transport and maintenance:** In the transport and ventilation sector, many devices are equipped with different types of sensors, investigating the onboard electronics and engines [9]. Televic[2] is such a partner in the railway domain, utilising a high number of sensors on trains which produce floats of data for further analysis. Renson[3] controls the airflow of many households, based on sensor observations per room. Several contextual parameters, such as the weather, influence the measures. To reduce these unwanted alerts, this prior knowledge must be fused with the available sensor data.

6 Conclusions

In this proposed research, techniques will be developed for improved AD and RCA in a sensor stream environment. While current techniques only focus on the data themselves, the proposed methods will incorporate prior knowledge to reduce the number of false positive and negatives. Analysing the cause of an anomaly will be possible with the design of an interpretable embedding technique. At last, adaptiveness with less human involvement can be achieved in data streams by automatic derivation of learning rules from already existing models. The full system will be evaluated with two use cases for two different domains. I would like to thank my promotors prof. dr. ir. Filip De Turck and dr. Femke Ongeae for their support and valuable input in the realisation of this work. I would also like to thank Televic, Renson and imec for participating in this research.

References

1. Abele, Lisa, e.a.: Combining Knowledge Modeling and Machine Learning for Alarm Root Cause Analysis. IFAC Proceedings **46**(9) (2010) 1843–1848
2. B. A. Smith, e.a.: Fault diagnosis using first order logic tools. In: Proceedings of the 32nd Midwest Symposium on Circuits and Systems,. (Aug 1989) 299–302 vol.1
3. BERNERS-LEE, T., HENDLER, J., LASSILA, O.: The semantic web. Scientific American **284**(5) (2001) 34–43

[1] http://sweet-study.be/
[2] https://www.televic-rail.com
[3] https://www.renson.eu

4. Calvier, Francois Élies, e.a.: Ontology driven complex event pattern definition. In: Lecture Notes in Computer Science. Volume 10033., Springer (oct 2016) 522–530
5. Camossi, Elena, e.a.: Semantic-based Anomalous Pattern Discovery in Moving Object Trajectories. CoRR **abs/1305.1** (2013) 1–20
6. Ehsani-Besheli, Fatemeh, e.a.: Context-aware anomaly detection in embedded systems. In: Advances in Intelligent. Volume 582., Springer (2018) 151–165
7. Hammar, K.: Modular Semantic CEP for Threat Detection. Operations Research and Data Mining ORADM 2012 workshop proceedings (2012) 978–607
8. Huang, Hao, e.a.: Streaming Anomaly Detection Using Randomized Matrix Sketching. Proc. VLDB Endow. **9**(3) (2015) 192–203
9. Kdouh, H., e.a.: Wireless sensor network on board vessels. In: 2012 19th International Conference on Telecommunications, ICT 2012, IEEE (apr 2012) 1–6
10. Michael Sejr Schlichtkrull, e.a.: Modeling relational data with graph convolutional networks. CoRR **abs/1703.06103** (2017)
11. Nguyen, D.Q.: An overview of embedding models of entities and relationships for knowledge base completion. arXiv preprint arXiv **1703.08098** (2017)
12. Nickel, Maximilian, e.a.: A Review of Relational Machine Learning for Knowledge Graph. Proceedings of the IEEE **104**(28) (2015) 1–23
13. Patri, O., e.a.: Sensors to Events: Semantic Modeling and Recognition of Events from Data Streams. International Journal of Semantic Computing **10** (2016)
14. Paulheim, Heiko, e.a.: Exploiting Linked Open Data as Background Knowledge in Data Mining. International Workshop on Linked Data (2013) 1–10
15. Ristoski, Petar, e.a.: RDF2Vec: RDF Graph Embeddings and Their Applications. IOS Press **0** (2016)
16. Sandha, Sandeep Singh, e.a.: Complex Event Processing of Health Data in Real-time to Predict Heart Failure Risk and Stress. (2017)
17. Solé, Marc, e.a.: Survey on Models and Techniques for Root-Cause Analysis. Clinical Orthopaedics and Related Research (CoRR) (2017) 1–18
18. Song, Yangqiu, e.a.: Machine Learning with World Knowledge: The Position and Survey. arXiv preprint arXiv **1705.02908** (2017) 1–20
19. Souiden, Imen, e.a.: A survey on outlier detection in the context of stream mining. In: Advances in Intelligent Systems. Volume 557. Springer (2017) 372–383
20. Subutai Ahmad, e.a.: Unsupervised real-time anomaly detection for streaming data. Neurocomputing **262** (2017) 134 – 147
21. Takuma Ebisu, e.a.: Toruse: Knowledge graph embedding on a lie group. CoRR **abs/1711.05435** (2017)
22. T.T. Ademujimi, e.a.: A review of current machine learning techniques used in manufacturing diagnosis, Cham, Springer International Publishing (2017) 407–415
23. Ukil, Arijit, e.a.: Iot healthcare analytics: The importance of anomaly detection. Conference on Advanced Information Networking and Applications (2016) 994–997
24. Uzun, Yusuf, e.a.: Rule extraction from training artificial neural network. Multidisciplinary Engineering Science and Technology **3**(8) (2016) 2458–9403
25. Wang, Quan, e.a.: Knowledge Base Completion via Coupled Path Ranking. Acl (2014) 1308–1318
26. Xiao, Fuyuan, e.a.: New parallel processing strategies in complex event processing systems with data streams. Distributed Sensor Networks **13**(8) (2017) 1–15
27. Xu, Yongchun, e.a.: Semantic-based Complex Event Processing in the AAL Domain. 9th International Semantic Web Conference (ISWC2010) (2010)
28. Yan He, e.a.: Mechanism-indepedent outlier detection method for online experimentation. IEEE International Conference on Data Science (2017) 640–647
29. YE: Big data: Changing the way businesses compete and operate (2014)
30. Zheng, Alice X., e.a.: Failure diagnosis using decision trees. In: Proceedings of the First International Conference on Autonomic Computing. (2004)