Wisdom of the contexts: active ensemble learning for contextual anomaly detection

Ece Calikus[1] and Slawomir Nowaczyk[2]

***Abstract*— In contextual anomaly detection, an object is only considered anomalous within a specific context. Most existing methods use a single context based on a set of user-specified contextual features. However, identifying the right context can be very challenging in practice, especially in datasets with a large number of attributes. Furthermore, in real-world systems, there might be multiple anomalies that occur in different contexts and, therefore, require a combination of several "useful" contexts to unveil them. In this work, we propose a novel approach, called the wisdom of the contexts (WisCon), to effectively detect complex contextual anomalies in situations where the true contextual and behavioural attributes are unknown. Our method constructs an ensemble of multiple contexts, with varying importance scores, based on the assumption that not all useful contexts are equally so. We estimate the importance of each context using an active learning approach with a novel query strategy. Experiments show that WisCon significantly outperforms existing baselines in different categories (i.e., active learning methods, unsupervised contextual and non-contextual anomaly detectors) on 18 datasets. Furthermore, the results support our initial hypothesis that there is no single perfect context that successfully uncovers all kinds of contextual anomalies, and leveraging the "wisdom" of multiple contexts is necessary. Full paper:**
https://doi.org/10.1007/s10618-022-00868-7

I. INTRODUCTION

Anomaly detection is the task of identifying patterns or data points that differ from the norm. To identify anomalies, most anomaly detection techniques treat all the features associated with a data point equally. In many real-world applications, while some attributes in a feature set provide direct information on the normality or abnormality of objects (i.e., behavioural attributes), others give clues on environmental factors affecting the system (i.e., contextual attributes). For instance, high energy usage for space heating in a household can be normal in winter, whereas identical behaviour would be abnormal in the summer. In this case, even though the "ambient temperature" is not the attribute that directly indicates the abnormality, it can be used to determine whether the energy consumption behaviour is as expected under a particular set of conditions.

The goal of contextual (or conditional) anomaly detection is to find objects that are anomalous within certain contexts, but are disguised as normal globally (i.e., in the complete feature space). Contextual attributes are used for defining contexts, while behavioural attributes help to determine whether an object x significantly deviates from other related objects, i.e., those sharing similar contextual information with x. The majority of existing techniques address this problem using pre-specified features—most commonly assumed to be either spatial or temporal ones—to define the context.

However, identifying the "true" contextual and behavioural attributes in complex systems is very difficult in practice and often requires extensive domain knowledge. A context can be specified in many different ways, especially in datasets with a large number of features. Considering that the actual roles of different attributes are unknown in most cases, an effective contextual anomaly detection should be able to make an automated decision on the "best" context among many possible combinations of different attributes.

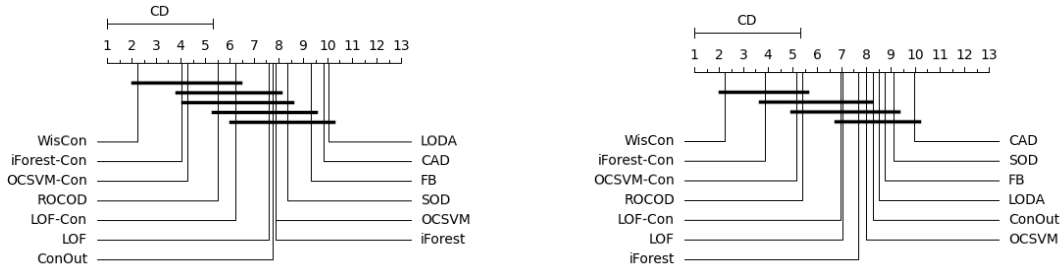
Anomaly detection is mostly formulated in an unsupervised fashion as ground truth information is often absent in practice, and acquiring labels can be prohibitively expensive. The lack of ground truth makes defining the "right" context for the problem more challenging, as we cannot verify whether the specified context actually reveals the contextual anomalies.

On the other hand, real-world systems often generate very diverse types of anomalies, and many of them may occur in different situations for different reasons. In such cases, a context that seems "right" for discovering a specific anomalous behaviour may be completely irrelevant when detecting other anomalies occurring in the same system. As in the previous example, a high level of heat consumption in summer is probably abnormal, and it can be detected when heat consumption is used as the behavioural attribute, and the ambient temperature is the context. On the other hand, broken valves cause the heating system to increase the hot water flow in the pipes unnecessarily. The high flow rate by itself does not indicate an anomalous behaviour if it is a result of high heat consumption, because a large flow is necessary to carry sufficient heat. However, a system that has a much higher flow rate than others *with similar heat consumption* is most likely faulty. In this case, the flow rate indicates the abnormality, and the heat consumption serves as the context. We cannot specify both the ambient temperature and heat consumption as a single context, as heat consumption was the behavioural attribute in the previous example. To be able to identify both anomalies, we need two different contexts. Evidently, there is no single "true" context to "rule" them all. Our main hypothesis is that anomalies in such complex systems can only be detected with a proper combination of multiple contexts.

However, leveraging multiple contexts effectively is not a trivial task, at least not without knowing which con-

[1]KTH Royal Institute of Technology, Stockholm, Sweden, calikus@kth.se

[2]Center for Applied Intelligent Systems Research, Halmstad University, Sweden, slawomir.nowaczyk@hh.se



(a) CD diagram (AUC-PR)

(b) CD diagram (AUC-ROC)

Fig. 1: The distribution of the performances (AUC-PR) in different contexts for all datasets.

texts are useful. Solutions that treat all available contexts equally would incorporate incorrect decisions when many uninformative contexts are present. Therefore, we need an approach that carefully decides on important contexts that reveal different kinds of contextual anomalies in the system while eliminating the impact of the irrelevant ones.

In this work, we focus on two major challenges: (1) effectively incorporating multiple contexts, given that the usefulness of a context is unknown, and (2) effectively estimating whether a context is useful or not, without ground truth information concerning which attributes are contextual vs behavioural. The paper introduces the "Wisdom of the Contexts" (WisCon) approach to address the problem of contextual anomaly detection, in which the "true" roles of attributes are unknown a priori. WisCon leverages the synergy of two worlds, active learning and ensembles. Active learning is concerned with quantifying how useful or irrelevant a context is in unveiling contextual anomalies with a low labelling cost. Ensemble learning is used to combine decisions over multiple useful contexts with varying importance, instead of relying on a single pre-specified one.

Various active learning methods have been previously developed from different perspectives to decide which samples in a dataset are more informative than the others. However, the concept of informativeness is inherently subjective and depends on the problem at hand, the dataset, and the machine learning model. Our work is different from all prior work in one key aspect——the purpose of our active learning is to query instances that help distinguish between useful and irrelevant contexts accurately. To the best of our knowledge, there is no existing sampling strategy designed based on this objective. It is inherently different from the use of active learning in, for example, supervised machine learning, where classification uncertainty has been used successfully. In anomaly detection, one of the classes is intrinsically more interesting than the other, which presents unique challenges. Moreover, the fact that WisCon uses queries to evaluate *contexts* adds an extra layer of complexity.

To fill this gap, we propose a novel query strategy that aims to select samples enabling the most accurate estimation of the "usefulness" of different contexts. It is a committee-based approach, in which each committee member is a different base anomaly detector built based on a particular context. Our strategy mainly ensures that anomalous samples that cannot be detected under the majority of the contexts and, therefore, cannot gain the confidence of the committee, are queried sufficiently often. We claim that a context successfully revealing these anomalies has a higher probability of being "useful." Using this approach, we can easily discard many irrelevant contexts from the decision-making by using a limited number of queries. In summary, the contributions in this paper are as follows:

- **WisCon approach:** We propose a novel approach for contextual anomaly detection, namely Wisdom of the Contexts (WisCon), that automatically creates contexts, where true contextual and behavioural attributes are not known beforehand, and constructs an ensemble of multiple contexts with an active learning model,
- **Sampling strategy:** We propose a new committee-based query strategy, low confidence anomaly (LCA) sampling, designed to select anomalous samples that cannot be detected under the majority of the contexts. This strategy allows us to actively query for the anomalies that provide more information about which contexts are "relevant" or "irrelevant" so that importance of contexts can be estimated within a small budget.
- **Context Ensembles:** We design an ensemble over the weighted combination of different contexts, in which the results from different contexts are aggregated using their importance scores estimated with active learning and a pruning strategy that eliminates irrelevant contexts.
- **An empirical study:** We conduct comprehensive experiments, including statistical comparisons with baselines in different categories; performance comparisons and budget analysis among the different state-of-the-art query strategies and our novel strategy; and a study on showing individual benefits of two core concepts, i.e., active learning and multi-context ensembles, in this problem.

Reproducibility: The datasets used in the experiments are publicly available, and the implementation of WisCon is open-sourced at `https://github.com/caisr-hh`.

II. CONCLUSIONS

Our WisCon approach can significantly boost detection performance by effectively building ensembles using active learning with a proper query strategy.