Anomaly Explanation

Véronne Yepmo
University of Rennes 1, IRISA - UMR 6704, F-22305 Lannion, France
veronne.yepmo-tchaghe@irisa.fr

Abstract

With the surge of deep learning and laws aiming at regulating the use of artificial intelligence, providing explanations to algorithms outputs has been a hot topic in the recent years. Most works are devoted to the explanation of classifiers outputs. The explanation of unsupervised machine learning algorithms, like anomaly detection, has received less attention from the XAI community. But this little interest is not imputable to the irrelevance of the topic. In this paper, we demonstrate the importance of anomaly explanation, the areas still needing investigation based upon our previous contributions to the field, and the future directions that will be explored.

1 Introduction

According to Hawkins [Hawkins, 1980], an outlier is *an instance that deviates so much from others as to arouse suspicion it was generated by a different mechanism*. Those deviating instances are also called anomalies, and the machine learning task aiming at identifying them is called anomaly detection.

Assuming we have a dataset $D = \{x_1, ..., x_n\}$ containing n instances each having m attributes/features $f_1, ..., f_m$, *anomaly detection* consists in dividing the dataset into two subsets R and A such that $D = R \cup A$. The features can be categorical or numerical. R is the subset of regular instances or inliers, and A is the subset of anomalies or outliers. Anomaly detection has many applications ranging from spam detection to cancer detection, including fraud detection and fault detection. It can be handled as a supervised problem. In this case, anomaly detection is a binary classification problem (inlier VS outlier) with a strong imbalance between the classes, since anomalies are few in the dataset. This setting is used for example in credit card fraud detection or spam detection, where having access to labels is relatively easy (the user provides the labels). However, in most applications, labelling is a tedious task. The unsupervised setting, where no labels are required, is therefore the most adequate one. Methods like Local Outlier Factor [Breunig *et al.*, 2000], One-Class Support Vector Machines [Amer *et al.*, 2013] and Isolation Forest [Liu *et al.*, 2012] are unsupervised.

Anomaly explanation consists in providing explanations to the output of a particular anomaly detection system. It helps to answer the question:

Why did the system find this instance anomalous?

Some explanation systems are built for a particular anomaly detector. These are called model specific methods. It is the case of DIFFI [Carletti *et al.*, 2020] which was built for the Isolation Forest. Other methods, called model agnostic, are designed to be used by any anomaly detector. LookOut [Gupta *et al.*, 2018] is an example of model agnostic explanation method.

2 Motivation

Explaining anomalies helps first of all the final user to determine if the detected outlier is really an anomaly. In fact, sometimes deviating instances are not true anomalies when the context is taken into consideration. For example, a temperature of 35 degrees Celsius can be deemed excessive. But, if this temperature was recorded during summer in France, it is perfectly normal. An explanation (e.g: *The instance is anomalous because the temperature is too high*) allows to discard quickly that outlier, as a high temperature during that season in this location is normal. Without explanations, the user would have needed to take a closer look at the instance, which is time consuming, especially in real time situations.

When facing a classification problem, binary classification without limiting the generality, we already know what to expect from a new unknown instance: it is part of the first class or the second class. When dealing with anomaly detection, things can be more tricky. Anomalies are by definition deviating instances, and an instance can deviate in more ways than it can converge towards a class. Considering a dataset which contains cars and bikes, a new instance in a classification problem will be either a car or a bike. In anomaly detection, the instance can be a car or a bike (regular instances), but also a truck, a scooter, or any other objects which are neither cars nor bikes. This deviation also propagates in the explanation: if the instance is a scooter, why is it an anomaly? Because it has an engine like a car but not four wheels? Because it has two wheels like a bike but it has an engine, which a bike does not have? The space of possible explanations grows bigger, similarly to the space of possible deviations. For this reason, anomaly explanation deserves a particular attention.

3 Contribution

Our main contribution up to now is a thorough analysis of the requirements of anomaly explanations, and an insightful review of the literature related to the topic. In [Yepmo *et al.*, 2022], we realized a state of the art of the anomaly explanation field. We proposed a taxonomy of anomaly explanation methods. This taxonomy had four categories/levels of explanations:

- explanation by feature importance,
- explanation by feature values,
- explanation by data points comparisons,
- explanation by structure analysis.

Explanation by feature importance consists in identifying the features that led to the detection of the anomaly. This type of explanations is the most explored in the literature. But, as we showed in [Yepmo *et al.*, 2022], feature importance is generally unable to capture all the subtleties of an anomaly explanation. In explanation by feature values, the values of the features which helped to flag the instance as an anomaly are returned. Explanation by data points comparisons consists in identifying an anomaly prototype or a regularity prototype, and comparing that prototype to the instance to explain. Methods belonging to the last category identify the properties that the instance shares with the groups of regular data points found in the dataset, and how that instance differs from the data points in these groups. The structure of the dataset is examined first, to detect if there are any groups of data points, hence the name explanation by structure analysis. This category provides the most detailed explanations, but has been the least explored in the literature. The explanations generated are local to a group of data, and not relative to all the dataset like in the three first categories. It starts at the detection level with the identification of local anomalies (anomalies relatively to a group of inliers), and their explanation afterwards. But there is currently no end-to-end method doing that. All the methods providing explanations by structure analysis are pipelines (anomaly detection - clustering - explanations).

This preliminary contribution allowed us to conclude that, as we imagined, anomaly explanation has not yet been intensively explored. The lack of end-to-end methods for explanation by structure analysis makes it even clearer. It is therefore our main objective.

4 Future Work

The principal direction of work is to build an end-to-end anomaly explanation method by structure analysis. To this end, we focus on one anomaly detection method: the Isolation Forest. It is efficient, fast, has very few hyper-parameters and leads to an interpretable structure (Isolation Trees). From a modified version of the Isolation Forest, we will infer the structure of the dataset in order to detect local and global anomalies, and provide explanations consequently. To do that, instead of completely random attribute splits, we perform the splits pseudo-randomly by avoiding to break clusters. The splits frame the dense regions, and by combining them we can identify the subspaces in which subgroups of data are located. The Isolation Forest will therefore be used to identify and describe the clusters, as well as the local anomalies. The first results we got on artificial and real world datasets are really encouraging.

Another topic we are currently working on regarding this research is the evaluation of anomaly explanations. The evaluation of explanation methods has been investigated in the XAI field, but there is still work to do, especially with anomaly explanation. The evaluation of anomaly explanations is currently done using ground truths. These ground truths, just like labels for the detection task, can be unavailable. We are therefore working on the definition of some objective metrics to evaluate anomaly explanations. Anomalies have several candidate explanations, and each candidate explanation is evaluated on the basis of each metric (grouped according to a taxonomy). The final user or the domain expert then chooses the most important metrics and the most adequate explanations are returned. This work is inspired by the one done in the data quality field.

Acknowledgments

This research is part of the SEA DEFENDER project funded by the French DGA (Directorate General of Armaments). It is supervised by Grégory SMITS (http://people.irisa.fr/Gregory.Smits/) and Olivier PIVERT (http://people.irisa.fr/Olivier.Pivert/).

References

[Amer *et al.*, 2013] Mennatallah Amer, Markus Goldstein, and Slim Abdennadher. Enhancing one-class support vector machines for unsupervised anomaly detection. In *Proceedings of the ACM SIGKDD workshop on outlier detection and description*, pages 8–15, 2013.

[Breunig *et al.*, 2000] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.

[Carletti *et al.*, 2020] Mattia Carletti, Matteo Terzi, and Gian Antonio Susto. Interpretable anomaly detection with diffi: Depth-based feature importance for the isolation forest. *arXiv preprint arXiv:2007.11117*, 2020.

[Gupta *et al.*, 2018] Nikhil Gupta, Dhivya Eswaran, Neil Shah, Leman Akoglu, and Christos Faloutsos. Beyond outlier detection: Lookout for pictorial explanation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 122–138. Springer, 2018.

[Hawkins, 1980] Douglas M Hawkins. *Identification of outliers*, volume 11. Springer, 1980.

[Liu *et al.*, 2012] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1):1–39, 2012.

[Yepmo *et al.*, 2022] Véronne Yepmo, Grégory Smits, and Olivier Pivert. Anomaly explanation: A review. *Data & Knowledge Engineering*, 137:101946, 2022.