Revisiting Histogram Based Outlier Scores: Strengths and Weaknesses

Ignacio Aguilera-Martos(✉), Julián Luengo and Francisco Herrera

Department of Computer Science and Artificial Intelligence, Andalusian Institute of Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, Spain

nacheteam@ugr.es, {julianlm,herrera}@decsai.ugr.es

Abstract. Anomaly detection is a crucial task in various domains such as finance, cybersecurity or medical diagnosis. The demand for interpretability and explainability in model decisions has revived the use of traceable models, with Histogram Based Outlier Scores being a notable option due to its fast speed and commendable performance. Histogram Based Outlier Scores is a well-known and efficient unsupervised anomaly detection algorithm. Despite its popularity, it suffers from several limitations, including the inability to update its internal knowledge, model complex distributions, and consider feature relations. This work aims to provide a comprehensive analysis of the Histogram Based Outlier Scores algorithm status and its limitations. We conduct a comparative analysis of Histogram Based Outlier Scores with other state-of-the-art anomaly detection algorithms to identify its strengths and weaknesses. Our study shows that while Histogram Based Outlier Scores is efficient and computationally inexpensive, it may not be the best option in scenarios where the underlying data distribution is complex or where variable relations play a significant role. The presented alternatives and extensions to Histogram Based Outlier Scores provide valuable insights into the development of future anomaly detection methods.

Keywords: Anomaly detection · explainable AI · HBOS

1 Introduction

Anomaly detection is a fundamental task in data analysis and machine learning that aims to identify deviations from the expected behavior within a dataset. It involves detecting rare and unusual observations that differ significantly from the majority of normal data points. This process is crucial for uncovering novel patterns, outliers, and abnormal events, providing valuable insights across various domains such as network intrusion detection [10], fraud detection [17], system monitoring [16], quality control [7], and outlier identification in complex datasets. Anomaly detection techniques have gained widespread adoption due to the increasing volume of data and the need for automated decision-making systems.

P. García Bringas et al. (Eds.): HAIS 2023, LNAI 14001, pp. 39–48, 2023.
https://doi.org/10.1007/978-3-031-40725-3_4

Among the various methods developed, Histogram Based Outlier Scores (HBOS) [5] stands out for its simplicity, efficiency, and accuracy along with the interpretability that the model provides. HBOS creates histograms for each feature and computes the anomaly score based on the probability of each sample in the histograms. Despite its advantages, HBOS has several limitations, such as the inability to handle complex distributions, lack of update mechanisms, and the assumption of independence between variables.

In this contribution, we revisit the HBOS algorithm, providing a comprehensive analysis of its strengths and weaknesses, as well as comparing it to other state-of-the-art anomaly detection algorithms. We also present several extensions and alternatives to HBOS that address its limitations, such as Multi-step Histogram Based Outlier Scores (MHBOS) [2], Light Online Detector of Anomalies (LODA) [11] or Empirical-Cumulative-distribution-based Outlier Detection (ECOD) [8]. We analyze the performance of these methods using various datasets and metrics and provide insights into the suitability of each approach for different types of data and applications. Our work aims to provide practitioners with a better understanding of the HBOS algorithm and its extensions, and to guide the selection of the most appropriate method for a given anomaly detection task, where explainability is a key factor.

The rest of the text is structured as follows: in Sect. 2 the anomaly detection problem is presented. Section 3 introduces the HBOS algorithm as well as related algorithms. In Sect. 4 the experimental framework, results and analysis are presented. Section 5 presents the main strengths and weaknesses of HBOS. Finally Sect. 6 summarizes the lessons learned from this study.

2 Anomaly Detection

Anomaly detection [4] is a classical Machine Learning task consisting on identifying data points that deviate significantly from the norm or expected behavior of a given dataset. Unsupervised anomaly detection methods are particularly useful when labeled data is scarce or not available, as they do not require prior knowledge. This type of detection is only required to assume that the proportion of anomalies is low in relation to the number of normal data points.

Anomalies can be classified as contextual or collective [1]. Contextual anomalies occur in a specific situation which gives the anomaly the relevance, such as a high-priced item in a low-cost store. Collective anomalies are groups of data points that exhibit anomalous behavior as a whole piece, such as a group of hot days in winter.

Distinguishing between noise and anomalies is essential to ensure the accuracy and reliability of the results. Noise refers to random or irrelevant variations with no meaningful information. The main sources where noise comes from are measurement errors, data acquisition artifacts or environmental factors. In contrast, anomalies are observations that significantly deviate from the normal behavior of the data and contain useful information from unusual events. Removing noise and anomalies from the data is crucial for accurate data analysis and the enhancement of the applied techniques afterwards.

3 Histogram Based Outlier Scores: Analysis

This section is dedicated to discussing HBOS and the extensions that have been proposed to enhance the algorithm. Section 3.1 will provide an overview of the original HBOS algorithm, while Sect. 3.2 will introduce the MHBOS algorithm, which aims to overcome the limitation of updating histograms. In addition, Sect. 3.3 will explain the LODA algorithm as a means of addressing the issue of feature interaction within the model. Finally, Sect. 3.4 will provide an explanation of the ECOD algorithm, which to addresses the problems associated with the use of histograms in the HBOS algorithm.

3.1 Histogram Based Outlier Scores

Histogram are a graphical representation that divide the domain of a variable into fixed-sized intervals or bins and counts the number of values that fall within each bin. This structure provides the practitioner with the frequency or probability, if scaled, of a certain value range.

The HBOS [5] algorithm leverages histograms to determine if a value is rare or anomalous. To this end, the algorithm generates one histogram for each feature, scales the frequencies to reach the maximum value of one, and combines the information using the formula $HBOS(p) = \sum_{i=1}^{d} \log_2(\frac{1}{histogram(p_i)})$.

The logarithmic function, when applied to the inverse of the histogram frequency of the corresponding bin, is an increasing monotonic function. The logarithmic function reaches its minimum when the frequency is one, resulting in a value of zero. In contrast, if the frequency is less than one, the fraction becomes greater than one, resulting in a positive number. Hence, a higher value is obtained for samples with lower frequency, which corresponds to more anomalous samples. Finally, the information from all histograms is aggregated.

3.2 Updating the Histograms: MHBOS

MHBOS [2] is a novel algorithm that addresses the inability of HBOS to handle data streams. This algorithm introduces several update mechanisms for both static and dynamic histograms, which maintain the performance of HBOS while enhancing the algorithm's flexibility.

MHBOS defines an initial histogram for each feature with the available data. By means of one of the update mechanisms the histogram values (bin edges and frequencies) are updated with the incoming data slice. Thefore the algorithm is able to train iteratively on the available data and therefore face streams of data.

3.3 Modeling Relationships Across Features: LODA

LODA [11] leverages weighted and random one-dimensional projections of the features in the dataset. By doing so, it transforms the variables into a single feature that contains mixed information from all of them. Histograms are then constructed over these projections to represent joint distributions, and probabilities are computed in the same manner as with HBOS. This process is repeated multiple times to obtain several histograms and represent as many feature interactions as possible.

This algorithm tackles one of the key weaknesses of HBOS. The original algorithm assumes that the features are independent and therefore histogram modelling each feature is enough to detect anomalies. This is not the case in all datasets and therefore LODA extends the behavior to solve this problem.

3.4 Histograms Aside: ECOD

ECOD [8] is a probabilistic detector that uses the empirical cumulative distribution function to score the samples based on the assumption that rare events occur in the tail of the distribution. The algorithm checks if the distribution of the data is right or left skewed and analysis the corresponding cumulative distribution function to evaluate the probability of each sample. The bigger the cumulative probability the more anomalous the value is as more data points are present to the right or left.

While ECOD has its own weaknesses the method gets rid of histograms. HBOS suffers from modelling certain types of distributions such as distributions with holes and heavy tails. When using ECOD the holes in the distribution do not matter as the important value is the cumulative probability of the sample.

4 Experimental Analysis

In this section, we present the experimental framework used to analyze the impact of the modifications implemented in the alternative methods compared with HBOS, as well as recent Deep Learning approach. First, we describe the experimental setup used to evaluate the performance of the algorithms (Sect. 4.1. Then, we present the experimental results and study the performance of HBOS with state-of-the-art anomaly detection algorithms (Sect. 4.2) checking whether HBOS is still competitive against their alternatives or Deep Learning approaches or not.

4.1 Experimental Framework

In this study, a set of state-of-the-art algorithms for anomaly detection are opposed to HBOS. For this purpose MHBOS, LODA and ECOD are included as they were presented as solutions to the limitations of HBOS. In order to perform an updated comparison, a Deep Learning approach based on an Autoencoder is included. Autoencoder models [6] have been applied for anomaly detection as well as many other tasks. These neural networks learn to encode and reconstruct normal data. This process of dimensionality reduction learning enable a good reconstruction of normal samples yielding high error in anomalous instances.

To provide a comprehensive context for HBOS within the classical algorithm landscape, we have incorporated a range of algorithms for comparison. Our selection encompasses Principal Component Analysis [15] (PCA), One-class Support Vector Machine [13] (OCSVM), Local Outlier Factor [3] (LOF), K-Nearest Neighbors [12] (KNN), and Isolation Forest [9] (IForest). These algorithms represent a diverse set of anomaly detectors, encompassing distance-based, density-based, and decision tree-based approaches.

The benchmark datasets are obtained from ODDS Library [14], that collects classic labeled datasets for anomaly detection using distance-based methods. For our experimentation, we employed 6 datasets from the ODDS Library, which encompassed a broad range of instances and features, enabling us to obtain more generalized conclusions. The chosen datasets are: Arrhythmia (Arr) with 452 instances and 274 features, Breastw (Br) with 683 instances and 9 features, Glass (Gl) with 214 instances and 9 features, Letter (Lt) with 1,600 instances and 32 features, Thyroid (Th) with 3,772 instances and 6 features and finally Vertebral (V) with 240 instances and 6 features.

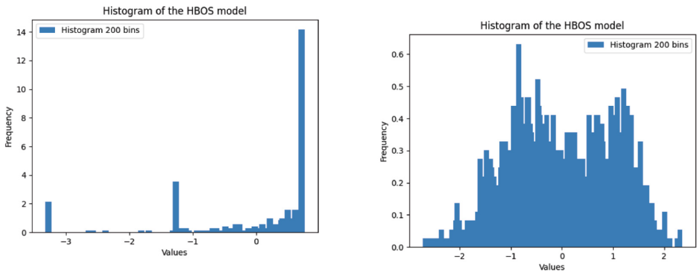
4.2 Experimental Results

The analysis of Table 1 reveals that HBOS algorithm outperforms other algorithms on most datasets, with MHBOS, OCSVM and LOF as the only algorithm that come close. HBOS demonstrates consistent performance across all datasets, achieving the highest F1 score on the Breastw dataset and the highest AUC on the Thyroid and Arrhythmia datasets. However, HBOS does not perform well on three out of the six datasets, namely Glass, Vertebral, and Letter.

Table 1. F1 and AUC for each algorithm. Datasets abbreviations in Sect. 4.1.

Datasets	Gl		Br		Lt		Th		Arr		V	
	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC
HBOS	0.111	0.705	**0.949**	0.990	0.12	0.623	**0.860**	**0.995**	**0.530**	**0.814**	0.066	0.328
AutoEncoder	0.111	0.601	0.933	0.973	0.18	0.724	0.387	0.961	0.424	0.775	0.1	0.524
ECOD	0.111	0.620	0.928	0.991	0.09	0.572	0.548	0.977	0.484	0.805	0.133	0.42
LODA	0	0.715	0.937	0.988	0.15	0.628	0.290	0.951	0.393	0.779	0.066	0.352
MHBOS	**0.444**	0.727	**0.949**	**0.993**	0.18	0.637	0.806	0.991	0.515	0.801	0.1	0.548
PCA	0.111	0.635	0.934	0.959	0.1	0.496	0.526	0.978	0.424	0.775	0.133	0.569
OCSVM	0.111	0.824	0.907	0.957	0.08	0.710	0.376	0.955	0.424	0.770	**0.233**	**0.694**
LOF	0.285	0.784	0.213	0.486	**0.538**	**0.912**	0.092	0.713	0.365	0.747	0.072	0.522
KNN	0.125	**0.839**	0.924	0.980	0.395	0.910	0.346	0.965	0.409	0.780	0.037	0.378
IForest	0.111	0.665	0.933	0.991	0.13	0.661	0.591	0.980	0.484	0.797	0	0.248

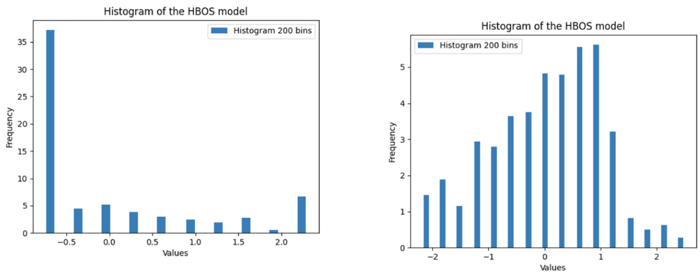
In order to gain further insight on why HBOS is having problems in those datasets, we analyzed the histograms of HBOS in the datasets where its performance was poor. We selected four histograms to summarize the internal state of the model, which are presented in Figs. 1a, 1b, 2a, and 2b. Our findings indicate that the performance of HBOS decreases when it encounters spaces in a distribution that has already been filled, resulting in a poor representation of the underlying domain. In contrast, we can observe an informative histogram in Fig. 1b as a comparison.



(a) Heavy tailed histogram (Glass).

(b) Normal histogram (Breastw).

Fig. 1. Borderline situations on histograms



(a) Heavy tailed and spaced (Vertebral).

(b) Spaced histogram (Letter).

Fig. 2. Borderline situations on histograms

Table 2 presents the execution times of the algorithms. Based on the results, it can be inferred that AutoEncoder require a significant amount of time to execute, owing to its high complexity. On the other hand, MHBOS exhibit moderate execution times across most datasets. Notably, HBOS demonstrate remarkably low execution times for most datasets, being the fastest PCA, LOF or OVSVM depending on the dataset. Consequently, HBOS can be regarded as a well-balanced algorithm, exhibiting good performance with low execution times.

Table 2. Time consumed for each algorithm in execution (time in seconds).

Datasets	Gl	Br	Lt	Th	Arr	V
HBOS	1.733	2.197	4.262	1.450	7.075	2.128
AutoEncoder	53.991	102.025	470.736	596.141	240.764	66.344
ECOD	1.694	2.144	4.261	1.636	32.205	1.418
LODA	5.097	4.297	5.578	7.265	4.039	5.466
MHBOS	2.531	4.782	9.735	7.823	32.126	5.789
PCA	1.238	1.344	**1.714**	**1.247**	3.95	1.835
OCSVM	**1.172**	1.359	3.515	15.400	1.929	**1.128**
LOF	1.123	**1.226**	2.665	4.596	**1.428**	1.186
KNN	3.836	11.051	138.535	233.417	11.501	4.103
IForest	28.887	90.344	92.455	87.549	87.792	34.203

4.3 Algorithm Analysis

The performance and limitations of the HBOS algorithm are presented in this study. The algorithm exhibits fast processing time and consistently produces high-quality results, with computational efficiency measured at $O(B \cdot F)$, where B represents the number of histogram bins and F denotes the number of features.

However, limitations of the algorithm were identified in three datasets due to gaps that occur in the histograms. These gaps contribute to an increased number of false positives in the output, particularly impacting the F1 metric. False positives in anomaly detection may imply a very sensitive algorithm which raises too many alarms.

While neural networks are considered the state-of-the-art for many problems, it is not always the case that they outperform other methods for a given dataset or problem. In our study, we found that the Autoencoder model was not the best performer, with HBOS proving to be superior. This result may be triggered by the number of instances and features present in datasets, as deep neural networks require larger volumes of data to train properly and therefore lighter algorithms could be more suitable. Although neural networks offer greater flexibility and the ability to update their knowledge as new data arrives, they can be difficult to interpret due to their black-box nature. On the other hand, simpler methods like HBOS can achieve comparable results to neural networks in certain scenarios and offer a more interpretable output that can be easily understood by practitioners.

5 Considerations About Strengths and Weaknesses

HBOS has established itself as a prominent algorithm for anomaly detection, utilizing histograms as a simple yet effective tool. The algorithm's main strengths lie in its straightforward design, high performance, and low execution times as discussed in the previous section. Furthermore, due to its reliance on histograms, HBOS is highly interpretable, providing insights into the specific features that determine whether a sample is anomalous and how they are related to the underlying distribution.

Notwithstanding the aforementioned advantages, HBOS suffers from significant limitations. The algorithm assumes that the variables are independent and do not have any correlation. As a result, the histograms generated by the algorithm only capture information about individual variables not considering any relationships between them. Consequently, the algorithm is unable to detect collective anomalies, identifying only contextual anomalies. This is clearly visible in the Letter data set, in which the relationships between variables are fundamental to understanding the letter to be recognized.

The primary drawback of using histograms as a distribution modeling tool in HBOS relates to distributions with non-covered spaces by bins and/or heavy tails, resulting in suboptimal algorithm performance. Due to the fixed bin size, some empty bins may be present between non-empty bins, leading to poor density estimation of the distribution. Consequently, the probability assigned to a sample may be non-conclusive, and an anomaly may be placed in a normal bin. Similarly, dynamic histograms with a high number of repeated values may result in distorted bins, further reducing the algorithm's accuracy. Vertebral is a good example of this, as we have been able to observe and corroborate by means of the histograms shown in Figs. 1a, 1b, 2a and 2b. Lastly, the lack of a histogram update mechanism renders HBOS inflexible and unresponsive to new data, making it unsuitable for dynamic datasets.

The importance of explainability and interpretability in machine learning has been increasingly recognized in recent years. Being able to understand and explain the reasoning behind a model's predictions is crucial for building trust and ensuring accountability. Regarding HBOS this is a valuable characteristic, as it enables users to understand how the model works and identify potential weaknesses or biases. In the future, it will likely become even more important as more complex models are developed, and regulatory requirements around explainability continue to evolve.

6 Conclusions

In this contribution, we have conducted a thorough analysis of the HBOS algorithm and identified its key strengths and limitations. Additionally, we have compiled recent proposals such as MHBOS, LODA, and ECOD, which address some of the exposed limitations. The study has pursued to prove whether these alternatives are effective, which has been accomplished specially with MHBOS proving to be more performant than the rest of the extensions.

Our results have led us to conclude that more complex Deep Learning models do not necessarily perform better than classical alternatives in simpler datasets. Furthermore, while deep neural network models lack explainability, the HBOS model offers a straightforward interpretation of results.

We conclude that HBOS is a reliable and fast tool that provides easy-to-interpret results. While the extensions discussed in this work address some issues with the algorithm, they do not solve all of them simultaneously. LODA is unable to update histograms dynamically, MHBOS maintains problems associated with histogram usage, and ECOD assumes independence between variables. We suggest that the development of probabilistic detectors has ample room for improvement by addressing these issues in combination or by integrating new tools beyond or along with the use of histograms.

Acknowledgment. This work has been supported by the Ministry of Science and Technology of Spain under project PID2020-119478GB-I00 and the project TED2021-132702B-C21 from the Ministry of Science and Innovation of Spain. I. Aguilera-Martos was supported by the Spanish Ministry of Science under the FPI programme PRE2021-100169.

References

1. Aggarwal, C.C.: An introduction to outlier analysis. In: Outlier Analysis, pp. 1–34. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47578-3_1
2. Aguilera-Martos, I., et al.: Multi-step histogram based outlier scores for unsupervised anomaly detection: arcelormittal engineering dataset case of study. Neurocomputing **544**, 126228 (2023)
3. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. SIGMOD Rec. **29**(2), 93–104 (2000)
4. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. (CSUR) **41**(3), 1–58 (2009)
5. Goldstein, M., Dengel, A.: Histogram-based outlier score (HBOS): a fast unsupervised anomaly detection algorithm
6. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313**(5786), 504–507 (2006)
7. Leigh, C., et al.: A framework for automated anomaly detection in high frequency water-quality data from in situ sensors. Sci. Total Environ. **664**, 885–898 (2019)
8. Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., Chen, G.: ECOD: unsupervised outlier detection using empirical cumulative distribution functions. IEEE Trans. Knowl. Data Eng. (2022)
9. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422 (2008)
10. Moustafa, N., Hu, J., Slay, J.: A holistic review of network anomaly detection systems: a comprehensive survey. J. Netw. Comput. Appl. **128**, 33–55 (2019)
11. Pevný, T.: LODA: lightweight on-line detector of anomalies. Mach. Learn. **102**(2), 275–304 (2016)
12. Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets, pp. 427–438 (2000)
13. Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection. In: Proceedings of the 12th International Conference on Neural Information Processing Systems, pp. 582–588 (1999)
14. Shebuti, R.: ODDS Library (2016). http://odds.cs.stonybrook.edu
15. Shyu, M.L., Chen, S.C., Sarinnapakorn, K., Chang, L.: A novel anomaly detection scheme based on principal component classifier. In: Proceedings of International Conference on Data Mining (2003)

16. Xue, Q., Li, G., Zhang, Y., Shen, S., Chen, Z., Liu, Y.: Fault diagnosis and abnormality detection of lithium-ion battery packs based on statistical distribution. J. Power Sources **482**, 228964 (2021)
17. Zhang, G., et al.: eFraudCom: an e-commerce fraud detection system via competitive graph neural networks. ACM Trans. Inf. Syst. (TOIS) **40**(3), 1–29 (2022)