proceedings

Proceedings

Adaptive Real-Time Method for Anomaly Detection Using Machine Learning [†]

David Novoa-Paradela *, Óscar Fontenla-Romero and Bertha Guijarro-Berdiñas

CITIC Research Center, Universidade da Coruña, 15071 A Coruña, Spain; oscar.fontenla@udc.es (Ó.F.-R.); berta.guijarro@udc.es (B.G.-B.)

* Correspondence: david.novoa@udc.es

† Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Published: 22 August 2020

Abstract: Anomaly detection is a sub-area of machine learning that deals with the development of methods to distinguish among normal and anomalous data. Due to the frequent use of anomaly-detection systems in monitoring and the lack of methods capable of learning in real time, this research presents a new method that provides such online adaptability. The method bases its operation on the properties of scaled convex hulls. It begins building a convex hull, using a minimum set of data, that is adapted and subdivided along time to accurately fit the boundary of the normal class data. The model has online learning ability and its execution can be carried out in a distributed and parallel way, all of them interesting advantages when dealing with big datasets. The method has been compared to other state-of-the-art algorithms demonstrating its effectiveness.

Keywords: anomaly detection; convex hull; data streaming; big data

1. Introduction

Anomaly detection, also known as one-class classification, is the process of identifying unexpected events in data that differ from what is considered normal instances. It has two basic assumptions: anomalies only occur very rarely and their features differ from the normal data significantly. Because anomalous data occurs very sporadically, in most real-world problems only data of the normal class is available. Then, most machine learning approaches for two-class supervised classification are not applicable to develop automatic methods for anomaly detection. Therefore, it requires specific machine learning methods whose training phase is generally carried out using only normal data. These methods try to model the normal class boundaries, so that new data can be classified by checking whether they belong to the normal class or not. This type of problem is frequent in real-world scenarios, such as predictive maintenance of industrial machinery. In this of problems, the ability to learn in real time can be essential as there may not be a sufficient amount of data at the beginning of the learning process but over time. There are many use cases (medical, IT security, etc.) where this situation happens and the detection methods must be able to start making decisions as soon as possible with very little initial knowledge and adapt this knowledge as new data are available. This paper presents the adaptation of an anomaly-detection method [1,2] based on convex hulls and random projections. The main contributions are to allow the convex hulls to be dynamically changed in an online learning scenario and to represent non-convex regions as a union of several convex hulls. The limits of the normal class will adapt when new data is processed, without store all the data or retrain from scratch.

2. Base Methods

Calculating the CH (Convex Hull) in high-dimensional spaces is a computationally expensive task. Due to this, several anomaly-detection methods choose to project the data on 2-D spaces in which the calculation of CHs is simple [1]. This random projection technique is based on the idea that high-dimensional data spaces can be projected into a lower-dimensional space without significantly losing the data structure if multiple projections are used. Base methods consist of the following:

1. Learning phase: Given a data set (normal data), a number τ of random projections of the data e are made onto 2-D subspaces. First, τ random matrices are generated. Second, the training set is projected into the space generated by each projection matrix. Finally, the CH's vertices are calculated in each projection, this being the aim of training (see projections P_1, P_2, P_3 in Figure 1). These vertices are projections of the original data; therefore, the algorithms only need to store the vertices forming the CHs and the projection matrices, discarding the rest of the training data.
2. Classification phase: To predict the class of a new data point, it is first projected using the τ projections generated during the training phase. For each projection, and given the set of vertices of the CH of that 2-D space, it is possible to check if the point is inside the corresponding polygon. It will be classified as normal only if it is inside all CHs. This procedure is shown in Figure 1. The new point (green) will be classified as an anomaly since it falls out of CH in the P_1 projection.

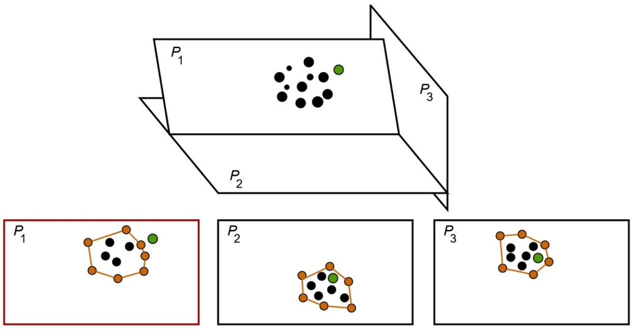

Figure 1. Projections of a 3-D data cloud (black), CHs of each 2-D space (orange) and a new point.

3. Online and Sub-Divisible Distributed Scaled Convex Hull

The main objective of OSHULL (**O**nline and **S**ub-divisible Distributed Scaled Convex **Hull**), is to carry out online one-class learning, so that starting from a minimum set of initial data points, it allows learning and adjusting the model with the arrival of new normal data. Also, because base methods are not suitable for dealing with datasets with non-convex shapes, OSHULL implements this capability. The procedures described below are applied independently to the CH at each of the projections.

3.1. Convex Hull Adjustment

To provide the ability to learn in real time, the limits of CHs must be readjusted as new data is available to the system. To do this, and at the same time guarantee some system stability, these limits will be expanded whenever a considerable number of data falls systematically and concentrated around the same area between the limits of the CH and a margin. In this case, the behavior of the algorithm will be to extend the limits of the CH to cover, as far as possible, that area where normal data did not appear when building the initial CH.

3.2. Region Subdivision

Because representing datasets using a single CH produces poor results in datasets with non-convex geometric shapes, the OSHULL method implements an iterative process that subdivides CHs for a better fit to the shape of the data. In this way, starting from an initial CH, it can be recursively subdivided into as many convex hulls as necessary to properly approximate the shape of normal data. Therefore, the method may represent non-convex regions as the union of various convex regions. At each projection several convex hulls can coexist that will continue to be readjusted individually.

3.3. Freezing Process

In the same way that a CH must be subdivided, a CH with all its edges at low distances from the data, and therefore well adjusted, must be maintained until the end of training and prevent it from being unnecessarily subdivided. This process is called freezing. A CH will freeze if all its edges are at normal distances and it has not been subdivided for a given number of iterations.

3.4. Pruning Process

After several subdivisions, some convex hulls can cover empty regions of normal data. Also, small convex hulls can be found that overlap the margins of other adjacent convex hulls. To get rid of them, a periodic pruning process is performed that eliminates those convex hulls that have not received data inside during a period, as it is assumed that they do not represent normal data.

4. Results

To evaluate the method, 8 datasets were employed (5 artificial and 3 real ones). Artificial datasets were used to evaluate the ability of the method to adapt to certain 3-D shapes. Real datasets were used to evaluate it in higher-dimensional datasets (6, 19 and 30 features). The results obtained by our method in an anomaly-detection scenario were compared to state-of-the-art algorithms working in batch mode. As they do not have the ability to adapt in real time, they were trained with the full training dataset. Conversely, the training process of OSHULL was carried out in real time: first creating the convex hulls with a small dataset and then iteratively readjusting them with the remaining data. Table 1 contains the average results for the test sets. Similarity was used as metric because is a balance between accuracy and recall. Although our method had the fourth position, its average similarity is close to the top, despite the disadvantage of being trained in online mode compared to batch mode.

Table 1. Average similarity (%) ± standard deviations for the different algorithms.

Algorithm	LOF	O-SVM	RC	OSHULL	IF	O-DSCH
Avg. similarity	**91.6± 5.9**	90.6± 6.6	89.4± 8.8	89.3± 7.8	85.1± 6.3	71.7± 19.1

5. Conclusions

OSHULL is an anomaly-detection method that offers online learning without significant loss of performance compared to classic batch methods, and its subdivision capacity favors it for treating non-convex problems. It is a light model as it is not necessary to store all the data for the creation or adaptation of CHs and it just needs to store their vertices. As an additional feature, the convex closures of each projection could be executed in parallel and distributed to achieve greater efficiency.

Author Contributions: These authors contributed equally to this work.

Funding: This work has been supported by Spanish Government's Secretaría de Estado de Investigación (Grant TIN2015-65069-C2-1-R), Xunta de Galicia (Grants ED431C 2018/34 and ED431G/01) and EU FEDER funds.

References

1. Casale, P.; Pujol, O.; Radeva, P. Approximate polytope ensemble for one-class classification. *Pattern Recognit.* **2014**, *47*, 854 –864.
2. Fernández-Francos, D.; Fontenla-Romero, O.; Alonso-Betanzos, A. One-class convex hull-based algorithm for classification in distributed environments. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 386–396.