Annals of Data Science (2023) 10(3):829–850
https://doi.org/10.1007/s40745-021-00362-9

A Comprehensive Survey of Anomaly Detection Algorithms

Durgesh Samariya[1,2] · **Amit Thakkar**[2]

Received: 28 December 2020 / Revised: 18 August 2021 / Accepted: 2 October 2021 /
Published online: 26 November 2021

Abstract
Anomaly or outlier detection is consider as one of the vital application of data mining, which deals with anomalies or outliers. Anomalies are considered as data points that are dramatically different from the rest of the data points. In this survey, we comprehensively present anomaly detection algorithms in an organized manner. We begin this survey with the definition of anomaly, then provide essential elements of anomaly detection, such as different types of anomaly, different application domains, and evaluation measures. Such anomaly detection algorithms are categorized in seven categories based on their working mechanisms, which includes total of 52 algorithms. The categories are anomaly detection algorithms based on statistics, density, distance, clustering, isolation, ensemble and subspace. For each category, we provide the time complexity of each algorithm and their general advantages and disadvantages. In the end, we compared all discussed anomaly detection algorithms in detail.

Keywords Anomaly · Anomaly detection · Outlier detection · Outlier analysis · Survey

1 Introduction

From early 20th century, Data Science, Machine Leaning, Deep Learning and Artificial Intelligence fields got so much attention from different industries. Recently IoT, decision making and artificial intelligence become key technologies of this century [1]. The advancement of computing power led to technologies like machine learning, deep learning, and similarly, the advancement in computer hardwares, the capacity

✉ Durgesh Samariya
samariya.durgesh@gmail.com

1 School of Engineering, Information Technology and Physical Sciences, Federation University, Churchill, VIC, Australia

2 Department of Computer Science and Engineering, Chandubhai S Patel Institute of Technology (CSPIT), Charotar University of Science and Technology (CHARUSAT), CHARUSAT Campus, Changa, Gujarat, India

of databases are growing significantly in terms of data entries [2]. Majority of such entries are normal, however, some of them are abnormal. Those abnormal data are rare and different then reminder of data points and called as *anomaly* or *outlier*.

Anomaly detection is considered as one of vital task of data mining with wide range of application domains [3]. Anomaly or outlier detection is the task of finding data points that do not conform with the rest of the data. Such data points are called as anomalies or outliers.[1] There is no complete definition of an anomaly; however, some of the classic definitions found in the literature are presented below:

1. "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs" [4].
2. "An outlier is an observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism" [5].
3. "An observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data" [6].
4. "Outliers are points that lie in the lower local density with respect to the density of its local neighbourhood" [7].
5. "Outliers are points that do not belong to clusters of a data set or as clusters that are significantly smaller than other clusters" [8].
6. "A point can be considered as an outlier if its own density is relatively lower than its nearby high density pattern cluster, or its own density is relatively higher than its nearby low density pattern regularity" [9].
7. "Anomalies are patterns in data that do not conform to a well defined notion of normal behaviour" [3].
8. "data records or instances which are deviated significantly and do not conform to normal data are called anomalies or outliers" [10].

These definitions somewhat differ from each other. However, they convey a similar concept of an anomaly. In literature, anomaly detection[2] is also termed as outlier detection, novelty detection, and abnormality detection.

The main aim of this survey is to provide better understanding of anomaly detection algorithms and their advantages and disadvantages to novice researcher with appropriate resources.

1.1 Type of Anomaly

Anomalies are classified as follows.

- *point anomaly* also known as *global anomaly*, is a data point that is significantly different than the rest of the data points.
- *group anomaly* is a collection of data points, which are anomalies compared to the rest of the data points.
- *local anomaly* is a data point which is an anomaly in terms of its neighbourhood.

[1] Anomaly and outlier are widely used terms. In this work, we will use both terms interchangeably.

[2] Anomaly detection and outlier detection are widely used terms. In this paper, we used both terms interchangeably.

Table 1 Different application domains of anomaly detection

Application domain	Explanation
Intrusion detection	In this application domain, anomalies refer to abnormal activity in a computer or network systems
Fault diagnosis(detection)	In this application domain, anomalies refer to faults in mechanical units
Healthcare	In this application domain, anomalies refer to unusual health condition of patients
Fraud detection	In this application domain, anomalies refer to fraudulent activity in credit card transactions or insurance claims
Detecting novelty from text	In this application domain, anomalies refer to novel text or news, or a collection of documents

– *collective anomaly* is collection of similar data points which is anomaly compared to entire dataset [11]. In this group, particular data instance might not be anomaly but consider as anomaly due to its presence in the anomalous region.

1.2 Different Applications of Anomaly Detection

Anomaly detection has a wide range of application domains; thus, it is beyond our limit to provide them in a single survey. As a result, we will summarise only interesting and recent application domains and how they define anomaly in that domain area which is summarised in Table 1. We refer readers to some surveys and chapter [3,12–14] for an extensive list of anomaly detection application domains.

1.3 Evaluation Measures

The purpose of the evaluation measure is to detect the performance of an algorithm. The three commonly used evaluation measures [18,19] are described in Table 2. The different mathematical symbols used in this paper are provided in Table 3.

1.4 Organization

The rest of the paper is organized as follows. In Sect. 2, we discussed about different types of anomaly detection algorithms. In subsequent sections, we discussed different anomaly detections based on statistics (Sect. 3), density (Sect. 4), distance (in Sect. 5), clustering (Sect. 6), isolation (Sect. 7), ensemble (Sect. 8), subspace (Sect. 9). In Sect. 10, we compared each anomaly detector. Finally, we conclude the paper in Sect. 11.

Table 2 Evaluation measures

Name	Equation	Explanation
Precision at n (P@n) [15]	$\text{P@n} = \frac{\{\lvert a \in \mathcal{A} \rvert rank(a) \leq n \rvert\}}{n}$	• Defined as the proportion of correct anomalies in the top n ranks • Where $\mathcal{A}$ is set of anomalies, n is the total number of anomalies
Average precision [16]	$\text{AP} = \frac{1}{n} \sum_{a \in \mathcal{A}} P@rank(a)$	• Assume that the total number of anomalies are known • Where $\mathcal{A}$ and n are defined as above
ROC AUC [17]	–	• A most popular evaluation measure • ROC curve is plot of true positive against false positive • ROC AUC value range in 0 to 1

Table 3 Symbols and notations

Symbol	Description
$\mathcal{R}^d$	A real-domain of d dimensions
$\mathbf{x}$	A data instance in $\mathcal{R}^d$
$\mathcal{X}$	A data set of $\mathbf{x}$, where $\lvert\mathcal{X}\rvert = n$
$\mathcal{A}$	A set of anomalies in data set, where $\mathcal{A} \subset \mathcal{X}$
a	An anomalous data point, where $a \in \mathcal{A}$
b	Number of bins in histogram
$\mathcal{D}$	A set of randomly selected subsamples, $\lvert\mathcal{D}\rvert = \psi$
ψ	Number of subsamples
t	Number of sets in ensemble
k	Size of nearest neighbour

2 Anomaly Detection

The anomaly detection algorithms can be categorized into seven categories which contains 52 anomaly detection algorithms, described in Table 4.

3 Anomaly Detection Algorithms Based on Statistical Model

The intuition behind this type of anomaly detection algorithms is, *anomalies resides in the low probability regions while inverse is true for the normal data instances*. This kind of anomaly detection algorithms are earliest work that have been proposed for anomaly detection. The process of this kind of anomaly detection algorithms is, first, a statistical model is fit to a given data set, and later some statistical test have been

Table 4 Anomaly detection algorithms

Category	Anomaly detection algorithm
Anomaly detection algorithms based on statistic model	Grabbs'test, Dixon test, Rosner's test, Student's t-test, Hostelling t^2-test, χ^2-statistics test, box plots, HBOS
Anomaly detection algorithms based on density	LOF, COF, LoOP, LOCI, RDF, INFLO, ROF, FastLOF, DWOF, SimplifiedLOF, LiNearN, GLOSH, SPAD, SPAD+
Anomaly detection algorithms based on distance	k-NN, kth-NN, RBRP, ABOD, GPA, LDOF, Sp, AntiHub
Anomaly detection algorithms based on clustering	OFP, FindOut, FindCBLOF, CBOD
Anomaly detection algorithms based on isolation	iForest, SciForest, HS-Tree, ReMass-iForest, iNNE, LeSiNN, LSHiForest, usfAD
Anomaly detection algorithms based on ensemble	LODA, DCSO, LSCP
Anomaly detection algorithms based on subspace	SOD, LSOF, HighDOD, COP, HiCS, CMI, Zero++

performed on the test data to identify whether test data belongs to the statistical model or not. If it belongs to the statistical model, then it is considered as normal instance or else anomaly.

Anomaly detection based on statistics models can be classified into two categories based on their assumption to the knowledge of underlying distribution [3].

- *parametric algorithms* assume that the underlying distribution of the given data is known (e.g. Gaussian distribution).
- *non-parametric algorithms* does not have any prior knowledge of data distribution of the given data.

3.1 Parametric Algorithms

The earliest work in this kind of statistical model assumes that the given data is generated from normal (Gaussian) distribution. Shewhart [20] is the earliest work which defines this kind of algorithms to detect anomaly in the process quality control application domain. Later, more advanced statistical techniques have been proposed to detect anomalies, are discussed in [4,6,21–27]. The commonly used statistical tests to detect anomaly are summarized in Table 5.

Other parametric algorithms employed are based on box plots [25,28–30]. Basically, box plot visually summarise the data.

3.2 Non-parametric Algorithms

The simplest and more intuitive non-parametric statistical algorithms are histogram-based methods. Histogram-based methods consist of two steps. In the first step, the

Table 5 Analysis of different statistic tests

Name	Function equation	Analysis
Grubbs's test [4]	$G = \frac{\max_{i=1,\ldots,n} \lvert x_i - \bar{\mathcal{X}} \rvert}{s}$	• Also known as maximum normalized residual test • G greater than critical value is considered as an anomaly • Easy to implement • For univariate data set only • One anomaly at a time • Where $\bar{\mathcal{X}}$ and s are sample mean and standard deviation, respectively
Dixon test [31]	$Q = \frac{x_n - x_{n-1}}{x_n - x_1}$	• Assume—only one outlier is present • Simple to implement • Applicable only on small datasets • First need to arrange data in ascending order and then compute the Dixon test on data
Rosner's test [21]	$R_{i+1} = \frac{\lvert x_i - \bar{\mathcal{X}} \rvert}{s}$	• where $\bar{\mathcal{X}}$ and s are as defined above • The data follow a normal distribution and the anomalies are employed from a different distribution • This test is good for larger data sets • Not simple as Dixon test • Must have knowledge of number of anomalies
Student's t-test [23,24]	$t = \frac{Z}{\sigma} = \frac{\bar{\mathcal{X}} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$	• Where $\bar{\mathcal{X}}$ is as defined above, μ and $\hat{\sigma}$ are mean and standard deviation of population, respectively and n is total number of samples • Normal samples are compared to test instance
Hostelling t^2-test [22]	$T^2 = n(\bar{\mathcal{X}} - \mu)' S^{-1} (\bar{\mathcal{X}} - \mu)$	• Where $\bar{\mathcal{X}}$, n and μ are as defined above. S is a covariance matrix
χ^2-statistics [26]	$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\mu}$	• μ is defined as above • *Assumption*—normal data has a multidimensional normal distribution • Data point with higher χ^2 value is considered as an anomaly

algorithm builds histogram in each dimension. In subsequent step, the algorithm compute the anomaly score as sum of the height of bins in each dimension in which point falls into. The histogram based classic techniques includes [32–36].

Other non-parametric statistical algorithms are based on kernel functions. Kernel based functions are used to estimate probability density and data instance having low probability density are considered as an anomaly [37–41].

We refer our readers to [42–44] for extensive list of this kind of anomaly detection algorithms.

Analysis.

- *Time Complexity* of statistical model are depended on the model they adopt and kernel based methods are quadratic to data size.
- *Advantages of anomaly detection algorithms based on statistical models are:*
  1. Histogram based methods compared to other anomaly detection algorithms are very simple and intuitive.
- *Disadvantages of anomaly detection algorithms based on statistical models are:*
  1. Most of the methods are applicable on univariate data only. Thus, it is computationally very expensive when dealing with multi-dimensional data set.
  2. Key weakness of histogram based methods is that they are unable to capture anomaly which is dependent on multiple features.
  3. Kernel based algorithms are computationally very expensive and sensitive to parameters.

4 Anomaly Detection Algorithms Based on Density

The intuition behind this type of anomaly detection algorithms is, *the density of the outlier object is significantly different from the normal instance*. The most used ones includes LOF [7], COF [45], LoOP [46] and LOCI [47]. LOF is the well-known and widely used anomaly detection algorithm, which is based on the relative density of data point with respect to its k-nearest neighbourhood. COF anomaly detection algorithm works same as LOF; however, COF uses a different mechanism to calculate density. LoOP employed a more robust local density estimator. LOCI proposed a measure called multi-granularity deviation factor, which is also a variant of LOF. The recent algorithms of this kind of anomaly detection includes RDF [48], INFLO [49] , ROF [50], FastLOF [51], DWOF [52], SimplifiedLOF [53], LiNearN [54], GLOSH [55], SPAD [56], SPAD+ [10].

RDF (stands for Relative Density Factor) method uses P-trees to detect anomalies from the data set. Data points with higher RDF values are considered as an anomaly. INFLO is another variant of LOF, which uses the symmetric neighbourhood relationship to identify anomalies. ROF (stands for Resolution-based Outlier Factor) detects anomalies using the growing window technique. Instead of using a growing window like ROF, DWOF uses a dynamic window-based outlier factor, which overcomes the shortcomings of ROF. SimplifiedLOF replaces local reachability distance to k-NN distance in LOF anomaly score measure. LiNearN is the new nearest neighbour-

Table 6 Time complexity of anomaly detection algorithms based on density

Methods	Time complexity	
	Training stage	Testing stage
LOF	–	$O(n^2d)$
COF	–	$O(n^2d)$
LoOP	–	$O(n^2d)$
LOCI	–	$O(n^3)$
RDF	–	$O(n^2d)$
INFLO	–	$O(n^2d)$
ROF	–	$O(n^2d)$
FastLOF	–	$O(n^2d)$
SimplifiedLOF	–	$O(n^2d)$
LiNearN	$O(t(\psi + \Psi)\psi d))$	$O(nt\psi d)$
SPAD	–	$O(nd)$
SPAD+	$O(nth + t\psi d)$	$O(t(h + d))$

based density estimator, which has linear time complexity. SPAD is a histogram-based method and uses probability density-based measure. SPAD+ is a variant of SPAD, which overcomes shortcomings of SPAD using principal component analysis [57]. HBOS [35] mentioned in anomaly detection algorithm based on statistics model also falls in this category.

For more detailed information on anomaly detection algorithms based on density, we refer our readers to [18,58,59].

Analysis.

- *Time complexity*[3] of anomaly detection algorithms based on density is presented in Table 6.
- *Advantages of anomaly detection algorithms based on density are:*
  1. This kind of anomaly detection algorithm is very intuitive; thus, they are widely used.
  2. Perform better than anomaly detection algorithm based on statistics and distance.
- *Disadvantages of anomaly detection algorithms based on density are:*
  1. A major disadvantage of this kind of anomaly detection algorithms is that they are computationally expensive, as they require to compute a pair-wise distance.
  2. Not suitable for large and high dimensional datasets.
  3. Sensitive to the parameter, such as the size of nearest neighbours (k).

[3] The time complexity of this kind of algorithms can be reduced to $O(n \log(n))$ by using good indexing structure, but they are not feasible in high dimensional space. Thus we mention time complexities without such index throughout the paper.

Table 7 Time complexity of anomaly detection algorithms based on distance. #m is total number of cells

Methods	Time complexity
kNN	$O(n^2d)$
kth-NN	$O(n^2d)$
RBRP	$O(n^2d)$
ABOD	$O(n^3d)$
FastABOD	$O(n^2 + nk^2)$
GPA	$O(kn^2 + m)$
LDOF	$O(n^2d)$
Sp	$O(nd\psi)$
AntiHub	$O(n^2d)$

5 Anomaly Detection Algorithms Based on Distance

The intuition behind this type of anomaly detection algorithms is, *anomalies are far away from their nearest neighbours*. In distance-based methods, anomaly score of the data point is calculated as a sum of the distance between a data point and it's k-nearest neighbours. The classic algorithms of anomaly detection based on distance includes k-NN [60,61], kth-NN [62], RBRP [63], ABOD [64], GPA [65], LDOF [66], Sp [67] and AntiHub [68].

The core idea behind k-NN is, anomalous data point is far from its k-nearest neighbour, anomaly score computed as distance between data instance and its k-NN. Later, kth-NN was introduced, which compute anomaly score as distance between data point to it's kth-nearest neighbour. kth-NN overcome the shortcomings of k-NN. Instead of nearest neighbour, approximate nearest neighbour is used in RBRP; which enables RBRP to run faster than k-NN and kth-NN. The core process of computing ABOD (angle-based outlier detection) anomaly score is, for each point, the weighted variance of angles between all other data points are considered as anomaly factor. FastABOD is first variant of ABOD, which is faster version of ABOD with time complexity of $O(n^2 + nk^2)$. GPA use a grid based pruning technique to detect anomaly in uncertain data. Rather than searching k-nearest neighbour in complete data set, Sp search nearest neighbour in randomly selected small sub-samples. Anomaly score is computed as the distance between data point to its nearest neighbour in sub-sample. LDOF compute anomaly score based on how data instance deviates from its neighbourhood. AntiHub employs reverse nearest neighbour counts based anomaly score.

For more information about anomaly detection algorithms based on distance, we refer our readers to [3,12,58,69].

Analysis.

- *Time complexity* of anomaly detection algorithms based on distance is presented in Table 7.
- *Advantages of anomaly detection algorithms based on distance are:*

  1. Easy to implement and straightforward.
  2. Independent to data distribution.

Table 8 Time complexity of anomaly detection algorithms based on clustering

	OFP	FindOut	FindCBLOF	CBOD
Time complexity	$O(n^2d)$	$O(Tdn\log_2(n))$	$O(n^2d)$	$O(n^2d)$

3. Using indexing structure(such as R*-Tree [70], R-Tree [71], R+-tree [72], kd-tree [73] and X-tree [70]) time complexity of this kind of algorithms can be improved to $O(n\log(n))$.

– *Disadvantages of anomaly detection algorithms based on distance are:*

1. In general, distance based methods have high time complexity, which accounts in $O(n^2)$. Using appropriate indexing scheme it can be reduced to $O(n\log(n))$. However, one common disadvantage of such indexing scheme is that, they breakdown in high dimensional data set.

6 Anomaly Detection Algorithms Based on Clustering

The intuition behind this type of anomaly detection algorithms is, *each point of the data set is either a part of a cluster or an anomaly*. The classic works that belongs to this category are OFP [8], FindOut [74], FindCBLOF [75] and CBOD [76]. OFP (Outlier Finding Process) is two-phase clustering process to detect outliers. In first phase, modified k-means clustering algorithm is used. Then, in subsequent phase, minimum spanning tree (MST) is constructed. The tree with less nodes are considered as anomalies. FindOut anomaly detection techniques use wave cluster algorithm, FindOut, first detects clusters and remove them from the data set. remaining points are consider as anomalies. FindCBLOF compute anomaly score based on cluster size in which data point falls into and distance between data point to cluster centroid, this scoring measure is called as CBLOF. CBOD use one pass clustering algorithm to detect clusters in first stage and in next stage, it computes outlier factor of each cluster.

We refer our readers to [58] for more detailed information on anomaly detection algorithms based on clustering.

Analysis.

– *Time complexity* of anomaly detection algorithms based on clustering is presented in Table 8.

– *Advantages of anomaly detection algorithms based on clustering are:*

1. This kind of methods are easily operated in unsupervised setting.
2. By simply replacing clustering algorithms, this type of methods work with complex and different data types.

– *Disadvantages of anomaly detection algorithms based on clustering are:*

1. Performance heavily rely on clustering algorithm.
2. High time complexity.
3. Sensitive to parameters.

Table 9 Time complexity of different anomaly detection algorithms based on isolation

Methods	Time complexity Training stage	Testing stage
iForest	$O(t\psi \log(\psi))$	$O(nt \log(\psi))$
SCiForest	$O((t\tau\psi(q\psi + \log(\psi) + \psi))$	$O(qnt\psi)$
HS-Tree	–	$O(t(h + \psi))$
ReMass-iForest	$O(t\psi \log(\psi))$	$O(nt \log(\psi))$
iNNE	$O(t\psi^2 d)$	$O(ntd\psi)$
LeSiNN	$O(\psi td)$	$O(n\psi td)$
LSHiForest	$O(t\psi \log(\psi)d)$	$O(nt \log(\psi)d)$
usfAD	$O(nth + t\psi d)$	$O(t(h + d))$

4. This kind of anomaly detection algorithms use binary score, thus it can not differentiate strong and weak anomaly, while it is possible in other kind of anomaly detection algorithms (such as distance, density and isolation).

7 Anomaly Detection Algorithms Based on Isolation

The intuition behind this type of anomaly detection algorithms is, *anomalies are few and thus they are easy to isolate compared to normal instances*. The earliest algorithm in this line of work, to best to our knowledge, is isolation forest [77,78], and later many algorithms have been proposed such as SCiForest [79], HS-Tree [80], Re-Mass iForest [81], iNNE [82,83], LeSiNN [84], LSHiForest [85], and usfAD [86,87]. iForest, in training stage, constructs isolation trees on randomly selected subsamples(ψ) and isolate each instances using random axis-parallel splits. In test stage, anomalies are detected using average path length of isolation tree. SCiForest is variant of iForest, which is able to detect global and local clustered anomalies[4] efficiently. HS-Tree is fast one-class anomaly detection algorithm for streaming data, which employs mass [88] as anomaly ranking measure. ReMass-iForest is variant of iForest, which use relative score instead of global score as used in iForest. Due to relative score, ReMass-iForest is able to detect local anomalies, which is shortcoming of iForest. Instead of axis-parallel isolation, iNNE employs nearest neighbour based isolation. LSHiForest is isolation based anomaly detection which uses local sensitive hashing forest. usfAD constructs ensemble of unsupervised stochastic forest(USF) [89] and compute anomaly score as average of USF tree.

We refer our reader to [90] for an explanation on why nearest neighbour based methods with small subsamples work.

Analysis.

– *Time complexity* of different anomaly detection algorithms based on isolation is presented in Table 9.
– *Advantages of anomaly detection algorithms based on isolation are:*

[4] Clustered anomalies are anomalies, which form cluster of few points outside of the normal cluster.

Table 10 Time complexity of anomaly detection algorithms based on ensemble

Methods	LODA	DSCO	LSCP
Time complexity	$O(nkd^{-\frac{1}{2}})$	$O(nd + n\log(n))$	$O(nd + n\log(n))$

1. In general, this type of anomaly detection algorithms have relatively low run time (time complexity) and high outlier detection accuracy.
2. High scalability.
3. Suitable for data with global and local anomalies.
4. SCiForest have better accuracy than other methods such as iForest, LOF in detecting local and global clustered anomalies.

– *Disadvantages of anomaly detection algorithms based on isolation are:*

5. Tree-based methods (such as iForest and usfAD) are not suitable for data with local anomalies.
6. HS-Tree is only applicable on streaming data.
7. More shortcomings of iForest can be found in [91].

8 Anomaly Detection Algorithms Based on Ensemble Technique

The intuition behind this kind of anomaly detection algorithms is, *combination of different anomaly detector is beneficial, if they do not have same error.* In many machine learning tasks, ensemble based methods are well known for their better performance compared to other methods. Recently, researchers have been interested in ensembles for anomaly detection. The classic algorithms include LODA [92], DCSO [93] and LSCP [94].

We refer our readers to [95–97] for more detailed and broader information about anomaly detection algorithms based on ensemble.

Analysis.

– *Time complexity* of anomaly detection algorithms based on ensemble is presented in Table 10.
– *Advantages of anomaly detection algorithms based on ensemble are:*

1. In general this type of methods are highly stable and perform comparatively better.
2. Useful for outlier analysis.

– *Disadvantages of anomaly detection algorithms based on ensemble are:*

1. Compare to other data mining tasks, only very less methods have been developed.
2. Comparatively high time complexity compared to isolation based algorithms.

Table 11 Time complexity of anomaly detection algorithms based on subspace

Methods	Time complexity
SOD	$O(n^3d)$
LSOF	$O(n^2d)$
HighDOD	$O((x+n)Ndim(S))$
COP	$O(n^2d^3)$
HiCS	$O(n^2d)$
CMI	$O(n^2d)$
Zero++	$O(ntq+dtq)$

9 Anomaly Detection Algorithms Based on Subspace

The intuition behind this type of anomaly detection algorithms is, *anomalies are hidden in subset of features*. The basic idea of this kind of algorithms is they will select subspace[5] and then compute anomaly score in that subspace. The algorithms which falls in this categories includes SOD [98], LSOF [99], HighDOD [100], COP [101], HiCS [102], CMI [103], and Zero++ [104].

SOD (Subspace Outlier Degree) explore axis-parallel subspace and consider point as anomaly if it deviates from its neighbours in that subspace. LSOF (Local Subspace based Outlier Factor), first detects interesting features using variance and later enhance variant of LOF used to detect outlier in those features. HighDOD uses distance based measure which is dimensionally unbias. COP search for outlier in arbitrarily oriented subspaces and benefit of COP is that, it not only detects anomaly but also provides explanation of how anomaly is different than others. HiCS search for high contrast subspace and then it computes outlier score in high contrast subspace only. CMI is another contrast measure employs on cumulative entropy of subspace. Zero++ employs the number of zero appearances as a score to detect anomalies in categorical dataset. Assumption behind Zero++ is, anomalies have high number of zero appearances compared to normal instance.

We refer our readers to [105,106] for comprehensive list of anomaly detection algorithms based on subspace.

Analysis.

- *Time complexity*[6] of anomaly detection algorithms based on subspace is presented in Table 11.
- *Advantages of anomaly detection algorithms based on subspace are:*
  1. This kind of anomaly detection algorithms are good at detecting hidden anomalies.
- *Disadvantages of anomaly detection algorithms based on subspace are:*

[5] Some algorithms choose subspace based on statistical test (e.g. HiCS, CMI) and some choose randomly(e.g. Zero++).

[6] Anomaly detection algorithms based on subspace are required to search for the subspace, which requires additional time, which depends on a search method. We only provide scoring time in a subspace.

Table 12 Comparison of different anomaly detection algorithms

Category	Methods	Equation	Time Complexity	Scalability	Reference
Based on density	LOF	$\mathrm{LOF}(x) = \frac{\sum_{y \in N^k(x)} lrd(y)}{\lvert N^k(x)\rvert \times lrd(x)}$	High	×	[7]
	COF	$\mathrm{COF}(x) = \frac{\lvert N_k(x)\rvert \cdot dist_{N_k(x)}(x)}{\sum_{y \in N_k(x)} \cdot dist_{N_k(y)}(y)}$	High	×	[45]
	LoOP	$\mathrm{LoOP}(x) = \max\left\{0, \mathrm{erf}\left(\frac{PLOF_{\lambda,S}(x)}{nPLOF \cdot \sqrt{2}}\right)\right\}$	High	×	[46]
	LOCI	$\mathrm{MDEF}(x_i, r, \alpha) = 1 - \frac{n(x_i, \alpha r)}{\hat{n}(x_i, r, \alpha)}$	High	×	[47]
	RDF	$\mathrm{RDF}(x, r) = \frac{DF_{nbr}(N_k(x), r)}{DF(N_k(x), r)}$	High	×	[48]
	INFLO	$\mathrm{INFLO}_k(x) = \frac{\sum_{y \in IS_k(x)} den(y)}{\lvert IS_k(x)\rvert \cdot den(x)}$	High	×	[49]
	ROF	$\mathrm{ROF}(x) = \sum_{i=1}^{R} \frac{\mathrm{ClusterSize}(x, r_{i-1}) - 1}{\mathrm{ClusterSize}(x, r_i)}$	High	×	[50]
	LiNearN	–	Low	✓	[54]
	SPAD	$\mathrm{SPAD}(x) = \sum_{i=1}^{d} \log \frac{\lvert H_i(x)\rvert + 1}{n + b}$	Low	✓	[56]
	SPAD+	$\mathrm{SPAD+}(x) = \sum_{i=1}^{d} \log \frac{\lvert H_i(x)\rvert + 1}{n + b} + \sum_{j=1}^{d} \log \frac{\lvert H_j(x')\rvert + 1}{n + b}$	Low	✓	[10]
Based on distance	kNN	$k\text{-NN} = \sum_{y \in kNN(x)} dist(x, y)$	High	×	[60,61]
	kth-NN	$k\text{th-NN}(x) := dist_k(x; X)$	High	×	[62]
	RBRP	–	High	×	[63]
	ABOD	–	High	×	[64]
	FastABOD	–	High	×	[64]
	GPA	–	High	×	[65]

Table 12 continued

Category	Methods	Equation	Time Complexity	Scalability	Reference
	LDOF	$\mathrm{LDOF}(x) = \frac{\bar{d}_x}{D_x}$	High	×	[66]
	Sp	$\mathrm{Sp}(x) = \min_{y \in \mathcal{S}} dist(x, y)$	Low	✓	[67]
	AntiHub	–	High	×	[68]
Based on clustering	OFP	–	High	×	[8]
	FindOut	–	Low	✓	[74]
	FindCBLOF	–	High	×	[75]
	CBOD	–	High	×	[76]
Based on isolation	iForest	$\mathrm{iForest}(x) = \frac{1}{t} \sum_{i=1}^{t} l_i(x)$	Low	✓	[77,78]
	SCiForest	$\mathrm{SCiForest}(x) = \frac{1}{t} \sum_{i=1}^{t} l_i(x)$	Low	✓	[79]
	HS-Tree	$\mathrm{HS\text{-}Tree}(x) = \frac{1}{t} \sum_{i=1}^{t} m_i(x)$	Low	✓	[80]
	ReMass-iForest	$\mathrm{ReMass\text{-}iForest}(x) = \frac{1}{t} \sum_{i=1}^{t} s_i(x)$	Low	✓	[81]
	iNNE	$\mathrm{iNNE}(x) = \frac{1}{t} \sum_{i=1}^{t} I_i(x)$	Low	✓	[82,83]
	LeSiNN	$\mathrm{LeSiNN}(x) = \frac{1}{t} \sum_{i=1}^{t} \min_{y \in \mathcal{S}} dist(x, y)$	Low	✓	[84]
	LSHiForest	–	Low	✓	[85]
	usfAD	$\mathrm{usfAD}(x) = \frac{1}{t} \sum_{i=1}^{t} l_i(x)$	Low	✓	[86,87]

Table 12 continued

Category	Methods	Equation	Time Complexity	Scalability	Reference
Based on ensemble	LODA	–	Low	✓	[92]
	DSCO	–	Low	✓	[93]
	LSCP	–	Low	✓	[94]
Based on subspace	SOD	$\mathrm{SOD}_{R(x)}(x) = \frac{dist(y, \mathcal{H}(R(x)))}{\|v^{R(x)}\|_1}$	High	×	[98]
	LSOF	$\mathrm{LSOF}(x) = \frac{1}{\|N_{minPts}(x)\|} \sum_{y \in N_{minPts}(x)} \frac{lsrd(y)}{lsrd(x)}$	High	×	[99]
	COP	$\mathrm{COP}(x, \psi) = norm(1 - \cos(x), \psi)$	High	×	[101]
	CMI	$\mathrm{CMI}(x) = -\sum_{i=1}^{n-1} \left(x_{i+1} - x_i\right) \frac{i}{n} \log \frac{i}{n}$	High	×	[103]
	Zero++	$\mathrm{Zero}(x, \mathcal{D}\|S) = \sum_{i=1}^{\psi} \sum_{S \in \mathcal{S}} I(P_S(x\|\mathcal{D}) = 0)$	Low	✓	[104]

1. In general this type of anomaly detection algorithms have high time complexity (approximately similar to distance and density based methods).
2. Zero ++ [104] is available only for categorical data set.
3. Sensitive to irrelevant attributes.

10 Comparison of Different Anomaly Detection Algorithms

Table 12 presents comparison of all discussed anomaly detection algorithms based on the time complexity and scalability. We also provides mathematical equation to compute outlyingness of data point x for each methods where it is available in respective paper. It is clear from given Table 12 isolation based methods are comparatively faster than other anomaly detection methods. Apart from that, recent subsample based methods such as Sp and LinearN are also faster than tradition distance and density based methods.

Note that, In Table 12 time complexity are given as high and low. We consider quadratic as high, linear or constant as low. We consider methods with low time complexity as scalable methods.

11 Conclusion

We start this survey paper with the basic definitions of anomaly detection, then list existing and recent real-world application of anomaly detection, evaluation measures and finally divide this survey in seven parts: (i) anomaly detection based on statistical models, (ii) anomaly detection based on density, (iii) anomaly detection based on distance, (iv) anomaly detection based on clustering, (v) anomaly detection based on isolation, (vi) anomaly detection based on ensemble technique and (vii) anomaly detection based on subspace. The main aim of this survey paper is to introduce the basic idea and comparison of existing anomaly detection algorithms and analyse strengths and weakness of each one. We provide detailed comparison of each presented methods in Table 12. It is impossible to present each anomaly detection algorithms in one survey due to wide range of application areas of anomaly detection and the diverse research communities.

Acknowledgements The authors would also like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

Author Contributions DS conducted the systematic literature review and examined various outlier detection techniques. DS wrote the first draft of the manuscript. DS made significant contributions to design and structure of review. AT review the work and edit the manuscript. All authors read and approved the final manuscript.

Funding No funding recieved.

Data Availability Not applicable.

Declarations

Conflict of interest Not applicable.

Ethical approval This article does not contain any studies with human participants by any of the authors.

References

1. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. Ann Data Sci 4(2):149–178
2. Ahmed M, Najmul Islam AKM (2020) Deep learning: hope or hype. Ann Data Sci 7(3):427–432
3. Chandola V, Banerjee A, Kumar V (2007) Outlier detection: a survey. ACM Comput Surv 14:15
4. Grubbs FE (1969) Procedures for detecting outlying observations in samples. Technometrics 11(1):1–21
5. Hawkins DM (1980) Identification of outliers, vol 11. Springer, Berlin
6. Barnett V, Lewis T (1984) Outliers in statistical data, 3rd edn. Wiley, New York
7. Breunig MM, Kriegel HP, Ng RT, Sander J (2000) Lof: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data, SIGMOD '00, Association for Computing Machinery, New York, NY, USA, pp 93–104
8. Jiang MF, Tseng SS, Su CM (2001) Two-phase clustering process for outliers detection. Pattern Recogn Lett 22(6):691–700
9. Hu T, Sung SY (2003) Detecting pattern-based outliers. Pattern Recogn Lett 24(16):3059–3068
10. Aryal S, Baniya AA, Santosh KC (2019) Improved histogram-based anomaly detector with the extended principal component features. arXiv preprint arXiv: 1909.12702
11. Ahmed M (2018) Collective anomaly detection techniques for network traffic analysis. Ann Data Sci 5(4):497–512
12. Hodge V, Austin J (2004) A survey of outlier detection methodologies. Artif Intell Rev 22(2):85–126
13. Aggarwal CC (2017) An introduction to outlier analysis. Springer, Cham, pp 1–34
14. Olson DL, Shi Y, Shi Y (2007) Introduction to business data mining, vol 10. McGraw-Hill/Irwin, New York
15. Nick C (2009) Precision at n. Springer, Boston, pp 2127–2128
16. Zhang E, Zhang Y (2009) Average precision. Springer, Boston, pp 192–193
17. Hand DJ, Till RJ (2001) A simple generalisation of the area under the roc curve for multiple class classification problems. Mach Learn 45(2):171–186
18. Goldstein M, Uchida S (2016) A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS ONE 11(4):1–31, 04
19. Campos GO, Zimek A, Sander J, Campello RJGB, Micenková B, Schubert E, Assent I, Houle ME (2016) On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. Data Min Knowl Disc 30(4):891–927
20. Shewhart WA (1930) Economic quality control of manufactured product1. Bell Syst Tech J 9(2):364–389
21. Rosner B (1983) Percentage points for a generalized ESD many-outlier procedure. Technometrics 25(2):165–172
22. Liu J-P, Weng C-S (1991) Detection of outlying data in bioavailability/bioequivalence studies. Stat Med 10(9):1375–1389
23. Surace C, Worden K, Tomlinson G (1997) A novelty detection approach to diagnose damage in a cracked beam. In: Proceedings-SPIE the international society for optical engineering, Citeseer, pp 947–953
24. Surace C, Orden K et al (1998) A novelty detection method to diagnose damage in structures: an application to an offshore platform. In: The eighth international offshore and polar engineering conference, International Society of Offshore and Polar Engineers
25. Laurikkala J, Juhola M, Kentala E (2000) Informal identification of outliers in medical data. In: Fifth international workshop on intelligent data analysis in medicine and pharmacology, vol 1, pp 20–24
26. Ye N, Chen Q (2001) An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. Qual Reliab Eng Int 17(2):105–112
27. Rousseeuw PJ, Leroy AM (2005) Robust regression and outlier detection, vol 589. Wiley, New York

28. Horn PS, Feng L, Li Y, Pesce AJ (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. Clin Chem 47(12):2137–2142
29. Solberg HE, Lahti A (2005) Detection of outliers in reference distributions: performance of Horn's algorithm. Clin Chem 51(2):2326–2332, 12
30. Dovoedo YH, Chakraborti S (2015) Boxplot-based outlier detection for the location-scale family. Commun Stat Simul Comput 44(6):1492–1513
31. Gibbons RD (1994) Statistical methods for groundwater monitoring. Wiley, New York
32. Javitz HS, Valdes A (1991) The SRI ides statistical anomaly detector. In: Proceedings of 1991 IEEE computer society symposium on research in security and privacy, pp 316–326
33. Gebski M, Wong RK (2007) An efficient histogram method for outlier detection. In: Ramamohanarao KP, Krishna R, Mohania M, Nantajeewarawat E (eds) Advances in databases: concepts, systems and applications. Springer, Berlin, pp 176–187
34. Jiang X-B, Li G-Y, Lian S (2011) Outlier detection algorithm based on variable-width histogram for wireless sensor network. J Comput Appl 31(3):694–697
35. Goldstein M, Dengel A (2012) Histogram-based outlier score (hbos): a fast unsupervised anomaly detection algorithm. In: KI-2012: poster and demo track, pp 59–63
36. Xie M, Hu J, Tian B (2012) Histogram-based online anomaly detection in hierarchical wireless sensor networks. In: 2012 IEEE 11th international conference on trust, security and privacy in computing and communications, pp 751–759
37. Latecki LJ, Lazarevic A, Pokrajac D (2007) Outlier detection with kernel density functions. In: Perner P (ed) Machine learning and data mining in pattern recognition. Springer, Berlin, pp 61–75
38. Oh JH, Gao J (2009) A kernel-based approach for detecting outliers of high-dimensional biological data. In: BMC bioinformatics, vol 10, Springer, p S7
39. Gao J, Hu W, Zhang Z, Zhang X, Wu O (2011) Rkof: robust kernel-based local outlier detection. In: Huang JZ, Cao L, Srivastava J (eds) Advances in knowledge discovery and data mining. Springer, Berlin, pp 270–283
40. Askari A, Yang F, Ghaoui LE (2018) Kernel-based outlier detection using the inverse christoffel function
41. Liu F, Yanwei Yu, Song P, Fan Y, Tong X (2020) Scalable KDE-based top-n local outlier detection over large-scale data streams. Knowl Based Syst 204:106186
42. Siegel AF, Morgan CJ (1988) Statistics and data analysis: an introduction, 2nd edn. Wiley, New York
43. Zhang Y, Hamm NAS, Meratnia N, Stein A, van de Voort M, Havinga PJM (2012) Statistics-based outlier detection for wireless sensor networks. Int J Geogr Inf Sci 26(8):1373–1392
44. Zimek A, Filzmoser P (2018) There and back again: outlier detection between statistical reasoning and data mining algorithms. WIREs Data Min Knowl Discov 8(6):e1280
45. Tang J, Chen Z, Fu AWC, Cheung DW (2002) Enhancing effectiveness of outlier detections for low density patterns. In: Chen MS, Yu PS, Liu B (eds) Advances in knowledge discovery and data mining. Springer, Berlin, pp 535–548
46. Kriegel H-P, Kröger P, Schubert E, Zimek A (2009) Loop: local outlier probabilities. In: Proceedings of the 18th ACM conference on information and knowledge management, CIKM '09, Association for Computing Machinery, New York, NY, USA, pp 1649–1652
47. Papadimitriou S, Kitagawa H, Gibbons PB, Faloutsos C (2003) Loci: fast outlier detection using the local correlation integral. In: Proceedings 19th international conference on data engineering (Cat. No. 03CH37405), pp 315–326
48. Ren D, Wang B, Perrizo W (2004) Rdf: a density-based outlier detection method using vertical data representation. In: extitFourth IEEE international conference on data mining (ICDM'04), pp 503–506
49. Jin W, Tung Anthony KH, Han J, Wang W (2006) Ranking outliers using symmetric neighborhood relationship. In: Proceedings of the 10th Pacific-Asia conference on advances in knowledge discovery and data mining, PAKDD'06, Springer, Berlin, pp 577–593
50. Fan H, Zaïane OR, Foss A, Wu J (2009) Resolution-based outlier factor: detecting the top-n most outlying data points in engineering data. Knowl Inf Syst 19(1):31–51
51. Goldstein M (2012) Fastlof: an expectation-maximization based local outlier detection algorithm. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), pp 2282–2285
52. Momtaz R, Nesma M, Gowayyed MA (2013) Dwof: a robust density-based outlier detection approach. In: Sanches JM, Micó L, Cardoso JS (eds) Pattern recognition and image analysis. Springer, Berlin, pp 517–525

53. Schubert E, Zimek A, Kriegel H-P (2014) Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection. Data Min Knowl Disc 28(1):190–237
54. Wells JR, Ting KM, Washio T (2014) Linearn: a new approach to nearest neighbour density estimator. Pattern Recogn 47(8):2702–2720
55. Campello Ricardo JGB, Moulavi D, Zimek A, Sander J (2015) Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Trans Knowl Discov Data 10(1):1–51
56. Aryal S, Ting KM, Haffari G (2016) Revisiting attribute independence assumption in probabilistic unsupervised anomaly detection. In: Michael C, Alan Wang G, Hsinchun C (eds) Intelligence and security informatics. Springer, Cham, pp 73–86
57. Abdi H, Williams LJ (2010) Principal component analysis. WIREs Comput Stat 2(4):433–459
58. Aggarwal CC (2017) Proximity-based outlier detection. Springer, Cham, pp 111–147
59. Domingues R, Filippone M, Michiardi P, Zouaoui J (2018) A comparative evaluation of outlier detection algorithms: experiments and analyses. Pattern Recogn 74:406–421
60. Knorr EM, Ng RT (1998) Algorithms for mining distance-based outliers in large datasets. In: Proceedings of the 24rd international conference on very large data bases, VLDB '98, Kaufmann Publishers Inc, San Francisco, CA, USA, Morgan, pp 392–403
61. Knorr EM, Ng RT, Tucakov V (2000) Distance-based outliers: algorithms and applications. VLDB J 8(3):237–253
62. Ramaswamy S, Rastogi R, Shim K (2000) Efficient algorithms for mining outliers from large data sets. SIGMOD Rec 29(2):427–438
63. Ghoting A, Parthasarathy S, Otey ME (2008) Fast mining of distance-based outliers in high-dimensional datasets. Data Min Knowl Disc 16(3):349–364
64. Kriegel HP, Schubert M, Zimek A (2008) Angle-based outlier detection in high-dimensional data. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '08, Association for Computing Machinery, New York, pp 444–452
65. Wang B, Xiao G, Yu H, Yang X (2009) Distance-based outlier detection on uncertain data. In: 2009 Ninth IEEE international conference on computer and information technology, vol 1, pp 293–298
66. Zhang K, Hutter M, Jin H (2009) A new local distance-based outlier detection approach for scattered real-world data. In: Theeramunkong T, Kijsirikul B, Cercone N, Ho T-B (eds) Advances in knowledge discovery and data mining. Springer, Berlin, pp 813–822
67. Sugiyama M, Borgwardt K (2013) Rapid distance-based outlier detection via sampling. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ (eds) Advances in neural information processing systems, vol 26, Curran Associates Inc, pp 467–475
68. Radovanović M, Nanopoulos A, Ivanović M (2015) Reverse nearest neighbors in unsupervised distance-based outlier detection. IEEE Trans Knowl Data Eng 27(5):1369–1382
69. Wang H, Bah MJ, Hammad M (2019) Progress in outlier detection techniques: a survey. IEEE Access 7:107964–108000
70. Berchtold S, Keim DA, Kriegel H-P (1996) The x-tree: an index structure for high-dimensional data. In: Proceedings of the 22th international conference on very large data bases, VLDB '96, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, pp 28–39
71. Guttman A (1984) R-trees: a dynamic index structure for spatial searching. SIGMOD Rec 14(2):47–57
72. Sellis TK, Roussopoulos N, Faloutsos C (1987) The r+-tree: a dynamic index for multi-dimensional objects. In: Proceedings of the 13th international conference on very large data bases, VLDB '87, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, pp 507–518
73. Bentley JL (1975) Multidimensional binary search trees used for associative searching. Commun ACM 18(9):509–517
74. Dantong Yu, Sheikholeslami G, Zhang A (2002) Findout: finding outliers in very large datasets. Knowl Inf Syst 4(4):387–412
75. He Z, Xiaofei X, Deng S (2003) Discovering cluster-based local outliers. Pattern Recogn Lett 24(9–10):1641–1650
76. Jiang S, An Q (2008) Clustering-based outlier detection method. In: 2008 Fifth international conference on fuzzy systems and knowledge discovery, vol 2, pp 429–433
77. Liu FT, Ting KM, Zhou Z (2008) Isolation forest. In: 2008 Eighth IEEE international conference on data mining, pp 413–422
78. Liu FT, Ting KM, Zhou Z-H (2012) Isolation-based anomaly detection. ACM Trans Knowl Discov Data 6(1):1–39

79. Liu FT, Ting KM, Zhou ZH (2010) On detecting clustered anomalies using sciforest. In: Balcázar JL, Bonchi F, Gionis A, Sebag M (eds) Machine learning and knowledge discovery in databases. Springer, Berlin, pp 274–290
80. Tan SC, Ting KM, Liu TF (2011) Fast anomaly detection for streaming data. In: Proceedings of the twenty-second international joint conference on artificial intelligence, vol 2, IJCAI'11, AAAI Press, pp 1511–1516
81. Aryal S, Ting KM, Wells JR, Washio T (2014) Improving iforest with relative mass. In: Tseng VS, Ho TB, Zhou ZH, Chen ALP, Kao HY (eds) Advances in knowledge discovery and data mining. Springer, Cham, pp 510–521
82. Bandaragoda TR, Ting KM, Albrecht D, Liu FT, Wells JR (2014) Efficient anomaly detection by isolation using nearest neighbour ensemble. In: 2014 IEEE International conference on data mining workshop, pp 698–705
83. Bandaragoda TR, Ting KM, Albrecht D, Liu FT, Zhu Y, Wells JR (2018) Isolation-based anomaly detection using nearest-neighbor ensembles. Comput Intell 34(4):968–998
84. Pang G, Ting KM, Albrecht D (2015) Lesinn: detecting anomalies by identifying least similar nearest neighbours. In: 2015 IEEE international conference on data mining workshop (ICDMW), pp 623–630
85. Zhang X, Dou W, He Q, Zhou R, Leckie C, Kotagiri R, Salcic Z (2017) Lshiforest: a generic framework for fast tree isolation based ensemble anomaly analysis. In: 2017 IEEE 33rd international conference on data engineering (ICDE), pp 983–994
86. Aryal S (2018) Anomaly detection technique robust to units and scales of measurement. In: Phung D, Tseng VS, Webb GI, Ho B, Ganji M, Rashidi L (eds) Advances in knowledge discovery and data mining. Springer, Cham, pp 589–601
87. Aryal S, Santosh KC, Dazeley R (2020) usfad: a robust anomaly detector based on unsupervised stochastic forest. Int J Mach Learn Cybern 12:1–14
88. Ting KM, Zhou G-T, Liu FT, Tan JSC (2010) Mass estimation and its applications. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '10, Association for Computing Machinery, New York, NY, USA, pp 989–998
89. Fernando TL, Webb GI (2017) Simusf: an efficient and effective similarity measure that is invariant to violations of the interval scale assumption. Data Min Knowl Disc 31(1):264–286
90. Ting KM, Washio T, Wells JR, Aryal S (2017) Defying the gravity of learning curve: a characteristic of nearest neighbour anomaly detectors. Mach Learn 106(1):55–91
91. Bandaragoda TR (2015) Isolation based anomaly detection: a re-examination. PhD thesis, Monash University
92. Pevnỳ T (2016) Loda: lightweight on-line detector of anomalies. Mach Learn 102(2):275–304
93. Zhao Y, Hryniewicki MK (2018) DCSO: dynamic combination of detector scores for outlier ensembles. In: ACM SIGKDD ODD workshop, London, UK
94. Zhao Y, Nasrullah Z, Hryniewicki MK, Li Z (2019) LSCP: locally selective combination in parallel outlier ensembles. In: Proceedings of the 2019 SIAM international conference on data mining, SDM 2019, Calgary, Canada, pp 585–593
95. Aggarwal CC (2013) Outlier ensembles: position paper. SIGKDD Explor Newsl 14(2):49–58
96. Aggarwal CC (2017) Outlier ensembles. Springer, Cham, pp 185–218
97. Zimek A, Campello RJGB, Sander J (2014) Ensembles for unsupervised outlier detection: challenges and research questions a position paper. SIGKDD Explor Newsl 15(2):11–22
98. Kriegel H-P, Kröger P, Schubert E, Zimek A (2009) Outlier detection in axis-parallel subspaces of high dimensional data. In: Theeramunkong T, Kijsirikul B, Cercone N, Ho TB (eds) Advances in knowledge discovery and data mining. Springer, Berlin, pp 831–838
99. Agrawal A (2009) Local subspace based outlier detection. In: Ranka S, Aluru S, Buyya R, Chung Y-C, Dua S, Grama A, Gupta SKS, Kumar R, Phoha VV (eds) Contemporary computing. Springer, Heidelberg, pp 149–157
100. Nguyen HV, Gopalkrishnan V, Assent I (2011) An unbiased distance-based outlier detection approach for high-dimensional data. In: Jeffrey XY, Myoung HK, Rainer U (eds) Database systems for advanced applications. Springer, Berlin, pp 138–152
101. Kriegel H, Kröger P, Schubert E, Zimek A (2012) Outlier detection in arbitrarily oriented subspaces. In: 2012 IEEE 12th international conference on data mining, pp 379–388
102. Keller F, Muller E, Bohm K (2012) Hics: high contrast subspaces for density-based outlier ranking. In: 2012 IEEE 28th international conference on data engineering, pp 1037–1048

103. Nguyen HV, Müller E, Vreeken J, Keller F, Böhm, K (2013) Cmi: an information-theoretic contrast measure for enhancing subspace cluster and outlier detection. In: Proceedings of the 2013 SIAM international conference on data mining, SIAM, pp 198–206
104. Pang G, Ting KM, Albrecht D, Jin H (2016) Zero++: harnessing the power of zero appearances to detect anomalies in large-scale data sets. J Artif Intell Res 57:593–620
105. Aggarwal CC (2017) High-dimensional outlier detection: the subspace method, Springer International Publishing, Cham, pp 149–184
106. Zimek A, Schubert E, Kriegel H-P (2012) A survey on unsupervised outlier detection in high-dimensional numerical data. Stat Anal Data Min ASA Data Sci J 5(5):363–387