J Opt (April 2024) 53(2):756–774
https://doi.org/10.1007/s12596-023-01147-4

RESEARCH ARTICLE

A survey of anomaly detection techniques

Fatma M. Ghamry[1,2] · Ghada M. El-Banby[3] · Adel S. El-Fishawy[1] · Fathi E. Abd El-Samie[1,4] · Moawad I. Dessouky[1]

Received: 20 October 2022 / Accepted: 28 February 2023 / Published online: 16 February 2024

Abstract The phrase "anomaly detection" is often used to describe any technique that looks for samples that differ from expected patterns. Depending on availability of data labels, types of abnormalities and applications, many anomaly detection techniques have been developed. This study aims to give a well-organized and a thorough review of anomaly detection techniques. We think it will aid in a better understanding of the topic of anomaly detection. It also presents the different approaches introduced in the literature for anomaly detection from images as well as other patterns. Despite the common availability of categorical data in practice, anomaly detection from categorical data has received a relatively little attention as compared to that from quantitative data. We divide the anomaly detection research methodologies into distinct categories. We describe the fundamental anomaly detection techniques, as well as their modifications and importantance. In addition, we highlight the merits and demerits of each category. Finally, we discuss the research gaps and limitations encountered, when using anomaly detection techniques for categorical data to solve real-world problems.

Keywords Object detection · Anomaly detection · Image processing · Medical images

✉ Fatma M. Ghamry
f_ghamry@yahoo.com

Ghada M. El-Banby
ghadaelbanby75@gmail.com

Adel S. El-Fishawy
aelfishawy@hotmail.com

Fathi E. Abd El-Samie
feabdelhamid@pnuedu.sa; fathi_sayed@yahoo.com

Moawad I. Dessouky
dr_moawad@yahoo.com

[1] Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

[2] Communication and Electronics Department, Bilbeis Higher Institute for Engineering, BHIE, Bilbeis, Egypt

[3] Department of Industrial Electronics and Control Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

[4] Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

Introduction

Anomaly detection is a serious subject that has been widely researched across a wide range of research disciplines and application domains. First, this paper offers a review of research methodologies for anomaly detection. In addition, it introduces the analysis of the efficiency of anomaly detection techniques across a variety of application areas. Although the majority of strategies for identifying anomalies are similar, their importance and application areas may vary [1]. It is critical to consider the availability of data labels, while constructing an algorithm for classification or anomaly detection. We may separate anomaly detection techniques into three settings based on label availability [2], as shown in Fig. 1:

1) *Unsupervised anomaly detection.* We assume that only unlabeled data is available for training of the model in this context, which is perhaps the most prevalent scenario in anomaly detection [1, 3].
2) *Semi-supervised anomaly detection.* We assume that the training dataset is partially labeled and contains both labeled

Fig. 1 General types of anomaly detection techniques

and unlabeled samples in this case. Semi-supervised algorithms are useful in situations, when annotating the entire dataset is prohibitively expensive. This environment is especially common in the anomaly detection sector, since there is frequently both labeled and unlabeled data, yet labeling of data often needs specialist knowledge, and abnormalities can be costly in some circumstances, such as industrial and biomedical applications.

In supervised anomaly detection, we presume that the dataset is completely labeled. When abnormalities can be easily documented, supervised algorithms are preferable [4]–[7]. It is critical to distinguish between supervised anomaly detection and binary classification tasks at this stage. If normal and abnormal data are supplied during the training phase, the problem can be phrased as a supervised binary classification problem, and the job will no longer be an anomaly detection task. Different classifications will be illustrated in the following sections (Fig. 2).

Related work

This paper is concerned with the evaluation of the wide range of strategies that have been presented in the field of anomaly detection. We would like to classify the techniques, but we would like also to see if the analysis reveals any potentially general anomaly detection frameworks [8]. An anomaly detection assignment may face a variety of challenges depending on the type of data, such as high false positive rate, high computational cost and lack of a standard datasets for assessment.

Anomaly types

Anomalies are divided into three groups, depending on their nature [9, 10]:

i. *Point anomalies.* A point anomaly is a single anomalous sample that exhibits a pattern irregularity or deviation from the normal behavior [11], according to the authors

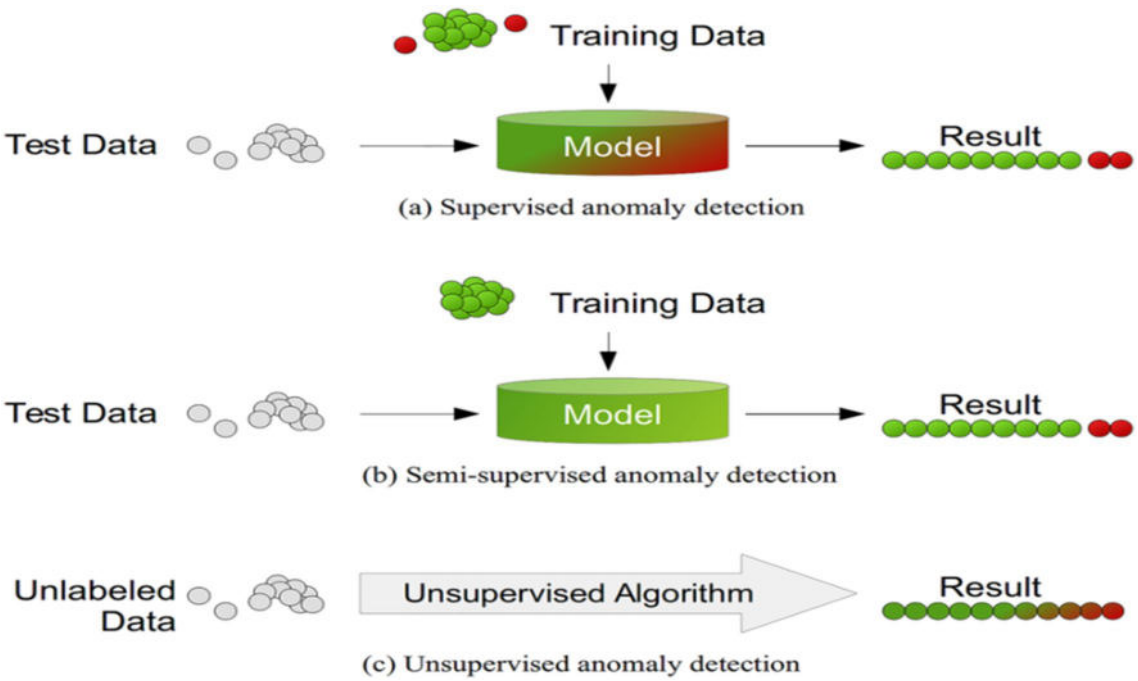

Fig. 2 Different anomaly detection classifications

Fig. 3 Anomaly types

of [12]. A credit card transaction with a large spend recorded at a Monaco restaurant, for example, seems to be a point anomaly as indicated in Fig. 3, since it deviates greatly from the rest of transactions [13].

ii. *Contextual Anomalies,* also identified as conditional anomalies, represent a data instance that may be regarded unusual under certain circumstances [14]. Both contextual and behavioral variables are used to identify contextual anomalies. The most common contextual factors are time and space. A pattern of spending money, the incidence of system log events, and any other feature used to characterize typical behavior are examples of behavioral characteristics.

iii. *Collective anomalies or group anomalies.* A subcategory of data points that are regarded as a group aberrant in comparison to the total dataset are referred to as collective or group anomaly. Each sample in the collective anomalies may or may not be an anomalous point in itself.

Anomaly detection techniques

Anomaly detection is an important problem that has been researched within diverse research areas and application domains. Many anomaly detection techniques have been specifically developed for certain application domains. This survey tries to provide a structured and comprehensive overview of the research on anomaly detection. We have grouped existing techniques into different categories based on the underlying approaches adopted by each technique. The majority of the common detection techniques can be categorized into classification-based, nearest-neighbor-based, clustering-based, and statistical techniques. Some techniques belong to research areas such as information theory, and spectral theory as shown in Table 1, revealing several different examples of anomaly detection techniques.

Contributions:

We describe the fundamental anomaly detection techniques in addition to (Fig. 4; Table 2)

- anomaly types
- common anomaly detection techniques
- some examples on general anomaly detection techniques
- illustration of the anomaly detection for categorical data
- some examples and applications of anomaly detection for categorical data
- anomaly detection based on classification

Classifiers include Support Vector Machines (SVMs) and neural networks to recognize normal and abnormal data in a particular feature space [37]. Anomaly detection strategies based on one-class classification presume that all training cases have just one class label. Such strategies use a one-class algorithm to develop a discriminative border around the normal examples. The available algorithms include one-class SVMs [38], and one-class Kernel Fisher Discriminants (KFDs) [39]. Anomaly is indicated for any check instance that does not fit inside the learned boundary.

In multi-class and one-class settings, neural networks have been used to identify anomalies. A rudimentary neural-network-based multi-class anomaly detection algorithm works in two phases. To learn the normal classes, a neural network is first trained on the normal training data. Second, each test case is sent into the neural network as an input. It is normal if the network accepts the test input, and it is an anomaly if the network discards the test input [40]. The following are some of the benefits of classification-based techniques:

(A) Classification-based techniques, particularly multi-class techniques, depend on sophisticated algorithms to discriminate between instances belonging to distinct classes.
(B) Because each test case must be checked according to a pre-computed model, the testing step of classification-based techniques will be quick.

The following are some drawbacks of classification-based techniques:

(A) Techniques based on multi-class classification depend on the obtainability of precise labels for distinct normal classes, which is frequently not the case.

Table 1 Examples of common anomaly detection techniques

Anomaly detection technique	Methodology	Benefits	Drawbacks
Nearest neighbor	Identifying anomalies by using neighborhood information Typical examples include kNN [15], kNNW [16], LOF [17], LoOP [18], ODIN [19], RBDA [20], etc	Independent of the data distributions Intuitively understood and easily interpreted Very easy to understand when there are few predictor variables Useful for building models that involve non-standard data types, such as text	Sensitive to parameters Relatively poor performance Have large storage requirements Sensitive to the choice of the similarity function that is used to compare instances Lack of a principled way to choose k, except through cross-validation Computationally-expensive technique
Subspace-based detection	Finding anomalies by sifting through different feature subsets Examples include SOD [21], Zhang et al. algorithm [22, 23], RODS [24], OR [25], Muller et al. algorithm [26], etc	High efficiency High effectiveness in some cases	Finding the relevant feature subspaces for outliers is non-trivial and difficult
Ensemble-based detection	Integrating various anomaly detection results to achieve a consensus FB [27], HiCS [28], Stein et al. [29], Zimek et al. [30], Passillas et al. [31] methods, and so on	High accuracy Less sensitive	Inefficient Choosing the right meta-detectors is difficult
Mixed-type detection	Making a unified model for different data types, or taking each data type, separately LOADED [32], ODMAD [33], Zhang et al. [34], Lu et al. [35], Do et al. [36] methods, and so on	Capable of handling the data with different types Relatively high accuracy	Obtaining the correlation structures of features is difficult High complexity
Neural Network	Training and testing sets	A neural network can perform tasks that a linear program cannot When an element of the neural network fails, it can continue operation, which makes it unique, especially for dimensionality reduction Problem with their parallel nature A neural network learns and does not need to be reprogrammed It can be implemented in any application	The neural network needs training to operate The architecture of a neural network is different from the architecture of microprocessors, and therefore it needs to be emulated Required high processing time for large neural networks
Decision Tree		Simple to understand and interpret Little data preparation Ability to handle both numerical and categorical data Uses a white box model Possibility to validate a model using statistical tests Robustness Perform well with large data in a short time	The problem of learning an optimal decision tree is known to be NP-complete under several aspects of optimality and even for simple concepts Decision-tree learners create over-complex trees that do not generalize the data well There are concepts that are hard to learn, because decision trees do not express them easily
Support Vector Machine	Finding the optimal separation hyper plane	Can deal with very high-dimensional data Some kernels have infinite Vapnik Chervonenkis dimension, which means that they can learn very elaborate concepts Usually work very well	Required positive and negative examples Need to select a good kernel function Required memory and CPU time There are some numerical stability problems in solving the constraint QP

Table 1 (continued)

Anomaly detection technique	Methodology	Benefits	Drawbacks
Self-organizing map	A topological clustering unsupervised algorithm that works with nonlinear dataset	Simple and easy-to-understand algorithm The excellent capability to visualize high dimensional data onto 1- or 2-dimensional space makes it unique, especially for dimensionality reduction	Time consuming algorithm
K-means		Low complexity	Necessity of specifying k Sensitive to noise and outlier data points Clusters are sensitive to initial assignment of centroids
Fuzzy C means		Allows a data point to be in multiple clusters A more natural representation of the behavior of genes	Need to define c, the clusters number Need to determine membership cutoff value Clusters are sensitive to initial assignment of centroids
Expectation Maximization		Ease to change the model to adapt to a different distribution of datasets Parameters number does not increase with the training data increase	Slown convergence in some cases

(B) Classification-based techniques provide a label to each test instance, which might be a disadvantage. Some strategies exist for obtaining a probabilistic classification, such as anomaly identification using the nearest-neighbor method.

These strategies are created on the premise that normal data examples exist in dense neighborhoods, whereas anomalies happen far away from their nearest neighbors. The distance to the kth nearest neighbor or relative density can be used for assessment.

The following are some of the benefits of using closest-neighbor techniques [40]:

(A) One of the biggest advantages of closest-neighbor-based algorithms is that they are unsupervised and do not create any expectations about the data propagative distribution. They are solely based on facts.
(B) Because the chance of an anomaly forming near a neighborhood in the training dataset is quite low, semi-supervised techniques outperform unsupervised ones in terms of missed anomalies.
(C) Adaptation of closest-neighbor-based techniques to a different data type is simple, requiring only the definition of an acceptable space measure for the supplied data.

The following are the drawbacks of closest-neighbor-based techniques:

(A) When using unsupervised techniques, whether the data contains normal examples with insufficient near neighbors or anomalies with sufficient near neighbors, the methodology is unsuccessful to appropriately identify them, resulting in missing anomalies [37].
(B) The false positive rate for semi-supervised techniques is significant if the normal examples in test data do not contain enough alike normal instances in the training data.
(C) The testing phase computational cost is also an insignificant problem, as it necessitates calculating the distance between every test instance and all instances belonging to either the test data or the training data in order to calculate the close neighbors.
(D) A distance measure created between two data instances, which may successfully discriminate between normal and abnormal instances, is crucial to the performance of a closest-neighbor-based technique. If the data is complicated, such as graphs or arrangements, defining distance metrics between instances might be difficult [40].

- Anomaly detection based on clustering.

Anomalies are defined as data instances that are distant from the centroid of their nearest cluster, whereas normal data instances are considered to belong to a cluster in the data. Clustering is a technique for grouping comparable data instances into clusters [41, 42]. Although semi-supervised clustering [43] has recently been investigated, clustering is generally an unsupervised approach. Some clustering-based anomaly detection techniques are advanced, despite the fact that clustering and anomaly detection seem to be basic. Anomaly detection techniques based on clustering may be divided into three groups. DBSCAN [44], ROCK [45], and SNN clustering [46] are examples of clustering algorithms that do not compel every data item to belong to a cluster.

The following are some of the benefits of clustering-based techniques:

i. Techniques based on clustering can be used in an unsupervised mode.
ii. Basically, a clustering-based technique that can deal with a particular data type may be typically extended to other complicated data types.
iii. Because the number of clusters is constant, the testing step for clustering-based techniques is quick.

The following items are some drawbacks of clustering-based techniques [37]:

i. The ability of clustering-based techniques to capture the cluster structure of normal instances is greatly dependent on the efficacy of the clustering algorithm.
ii. Anomalies are detected as a by-product of clustering in many cases. Therefore, clustering-based techniques may not be appropriate for anomaly detection.
iii. Quite a lot of clustering techniques require that each instance be allocated to one of several clusters. This might lead to anomalies being allocated to a big cluster, and so being misclassified as regular instances by techniques that assume anomalies not belonging to any cluster.
iv. Some clustering-based techniques are only useful, when anomalies do not form meaningful clusters.

- Anomaly detection based on statistical techniques.

The term "anomalies" refers to observations that are unlikely to have been created by the "background" stochastic model. As a result, anomalies appear in the background model as low-probability areas. The background models may be [37]:

A. parametric models such as Gaussian, Gaussian mixture and regression models.
B. non-parametric models such as kernel models.

The following basic assumption underpins statistical anomaly detection techniques.

Assumption: Anomalies arise in the little-probability areas of a stochastic model, whereas normal data examples occur in the high-probability regions [40].

The following are some of the benefits of statistical techniques:

i. Statistical techniques give statistically-defensible solutions for anomaly identification if the assumptions about the underlying data distribution are valid.
ii. A statistical technique anomaly score is accompanied with a confidence interval, which may be utilised as a supplementary data, when creating a conclusion about every test occurrence.
iii. Statistical techniques can work in an unsupervised situation devoid of the necessity for labeled training data if the distribution approximation phase is resilient to data anomalies.

The following are some of the drawbacks of statistical techniques:

i. Assuming that data is derived from a certain distribution is a major drawback of statistical techniques. This assumption is frequently incorrect, mainly for high-dimensional real-world datasets.
ii. For the statistical assumption to be reasonable, there are a variety of hypothesis test statistics that may be used to discover anomalies; selecting the optimum statistics is typically difficult [47].
iii. While histogram-based algorithms are straightforward to construct, they lack the ability to capture interactions between distinct features, which is a major drawback for multivariate data. An anomaly may include attribute values that are quite common separately, but extremely unusual when combined, yet an attribute-wise histogram-based approach may fail in detecting such anomalies [37].

- Anomaly detection based on spectral techniques.

Principal Component Analysis (PCA) and its extensions are the most important tools in this case. The basic premise is that an anomaly has deviant coordinates in comparison to typical PCA coordinates. Spectral techniques attempt to approximate the data using a set of qualities that represent the majority of the data variability. The scalability of PCA is

good. Nonlinear techniques can reduce temporal complexity to a constant sum of dimensions [40]. The dimensionality of techniques that conduct Singular Value Decomposition (SVD) on the data is often quadratic.

The following are some of the benefits of spectral anomaly detection techniques:

(A) Spectral techniques provide dimensionality reduction automatically, making them suited for high-dimensional datasets. Moreover, they may be utilized in preprocessing stages before applying any other anomaly detection technique to the modified space.
(B) In an unsupervised context, spectral techniques can be applied.

The following are some of the drawbacks of spectral anomaly detection techniques:

(A) Spectral techniques are only helpful if the normal and abnormal examples can be distinguished in the data low-dimensional embedding.
(B) Spectral techniques are notoriously difficult to implement.

- Detection of anomalies based on information theory.

Some techniques depend on information-theoretic measurements, such as entropy, and relative entropy to examine the information content of collected data. Entropy, relative entropy, and other terms are used to describe the complexity of a system. The fundamental information-theoretic anomaly detection technique has exponential time complexity, while approximation solutions with linear time complexity have been presented.

The following are some of the advantages of information-theoretic techniques:

(A) They can work in an unattended environment.
(B) They do not make any assumptions regarding the statistical distribution of the data.

The following are some of the drawbacks of information-theoretic techniques:

(A) The choice of the information-theoretic metric has a significant impact on the performance of such strategies.
(C) When it comes to arrangements and geographical datasets, information-theoretic techniques rely on the size of the substructure, which is sometimes hard to get.
(D) With an information-theoretic technique, it is challenging to link an anomaly score to a test case.

Classification of anomaly detection techniques

A wide range of techniques for detecting anomalies have been developed in recent years [58]-[64]. This paper is concerned with anomaly detection techniques, which are covered in some survey publications (e.g., [65] - [69]). Some authors classify techniques based on application, technical strategy, and/or data type.

The techniques can be grouped based on application domain as follows:

- *Network intrusion detection.* The works of [70–73], and [74] surveyed the techniques used for network intrusion detection. The work of [75] surveyed the distance and similarity metrics used in intrusion detection systems.

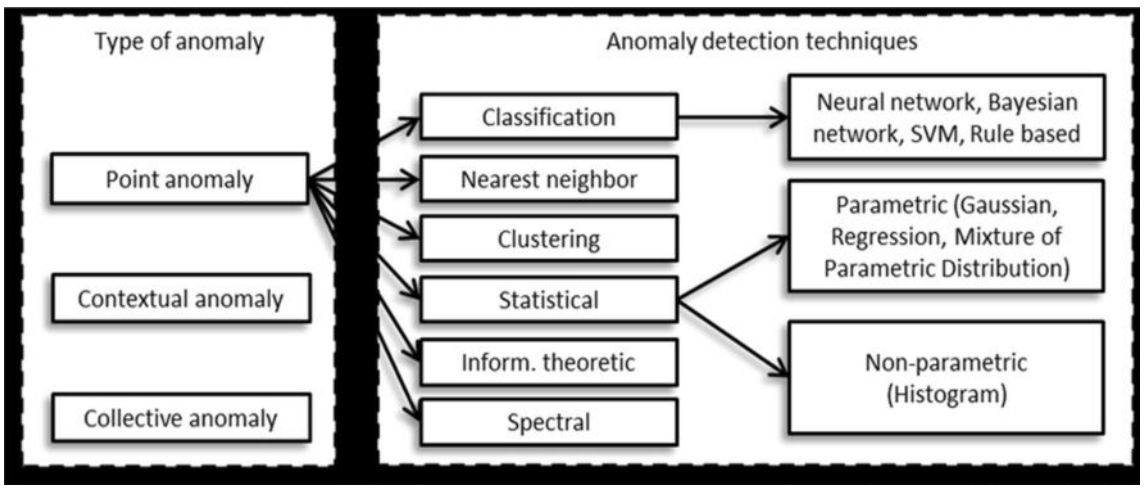

Fig. 4 Common anomaly types with detection techniques

Table 2 Examples of general anomaly detection techniques

Detection type	Reference	year	methodology
Supervised	49	2018	Uncertainty measure based on gradient of negative log likelihood is used as a measure of confidence
Supervised	50	2018	Confidence scores based on Mahalanobis distance from different layers are combined using weighted averaging
Supervised	51	2018	Invariance of classifier performance under various transformations of input image is used as a measure of confidence
Supervised	52	2017	Binary detector trained on intermediate feature representations is proposed to detect adversarial examples
Semi-supervised	53	2019	Likelihood ratio-based method is used to differentiate between in-distribution and OOD examples
Semi-supervised	54	2019	A two-head CNN consisting of a common feature extractor and two classifiers with different decision boundaries is trained to detect OOD examples
Unsupervised	55	2016	Predicted softmax probability is used to detect OOD examples
Unsupervised	56	2018	Temperature scaling by adding small perturbations to the input is used to better separate the softmax score for OOD detection
Unsupervised	57	2019	Resampling uncertainty estimation approach is proposed as an approximation to the bootstrap
Unsupervised	58	2020	Sensitivity of adversarial examples under compression-based transformations is used as a measure of confidence

- *Wireless Sensor Network* (*WSN*). The series of publications [76–85] focused on anomaly detection methods for wireless sensor network data. In [86] and [87], characteristics of anomaly detection techniques for wireless sensor networks in non-stationary and challenging situations were surveyed. The detection of anomalies in automated surveillance and smart homes has been discussed in [88] and [59].
- *Data streams.* The characteristics and categorization of anomaly detection in data streams are the subject of a few review publications (see, e.g., [89, 90], and [91]). The approaches for detecting anomalies in dynamic streaming data have been discussed in [92]. Focus was also given to anomalies in continuous-time variation data streams in [93].
- *Fraud detection.* There are numerous fields in which fraud and abuse detection techniques are effective. Techniques for detecting healthcare fraud have been reviewed in [94] and [95]. The authors of [96] reviewed fraud prevention strategies for chemical and biological data. In [97], fault detection techniques in industrial processes were examined.
- *Financial, business, and recommender systems* In [98–101], and [102], financial and credit card fraud detection techniques have been examined. The authors of [103] studied outlier profile assaults on recommender systems.

Another set of survey articles focused on the techniques or the methodologies that are used to identify outliers. These techniques include

• Data mining technique

Numerous review articles examined anomaly detection techniques based on data mining (see, e.g., [64, 94, 104, 105], and [106]). Specific data mining patterns have been covered in other review papers. Neural-network-based novelty detection techniques have been covered in [107]. Common pattern-based anomaly detection techniques have been evaluated in [108]. The methods for clustering-based anomaly identification have been examined in [91]. In [106], anomaly detection techniques for distributed data were examined. The authors of [109] discussed common pattern-based anomaly detection techniques and related score metrics.

• Machine learning techniques

There are survey publications that discussed anomaly detection techniques based on machine learning (see, e.g., [97, 110], and [58]).

• Statistical techniques

Before data mining and machine learning, the anomaly detection problem has been considered with statistical techniques. In contrast to complicated structures (such as categorical, graphical, and/or spatial data), statistical anomaly detection techniques focus on simple data types, such as numerical and quantitative data. Methods for detecting statistical anomalies have been discussed in [111–115].

A third group of review articles concentrated on the types of data, where anomaly detection techniques are proposed to identify outliers. The articles in this area can be classified according to data types including

• Social networks and graph data:

In certain survey studies, anomaly detection techniques in social media were reviewed [105, 116, 117], and [118]. Some studies were concerned with broader coverage and graph data (see, e.g., [119, 120], and [121].

• Time series and spatial data:

Several techniques for anomaly detection for time series and spatial data have been proposed [89, 122, 123], and [93].

• Big and complicated data:

Big and complicated data are important topics of study, particularly in the literature of computer science. Recently, a number of techniques for the detection of abnormalities in high-dimensional, complicated, and massive data have been presented [124] and [125]. In [126] and [127], anomaly detection for high-dimensional data was provided. Instead of analyzing all dimensions, other strategies concentrated on finding anomalies in subspaces [128].

• Experimental studies:

Experimental comparison studies of a particular class of anomaly detection techniques, such as various statistical and distance-based anomaly detection techniques, are the topic of some papers in empirical comparative reviews [60] and [129]. Additionally, empirical analysis of unsupervised anomaly detection techniques has been considered in [130] (Fig. 5).

Categorical data

Nominal and ordinal categorical data are the two types available. Gender, nationality, and network protocol type are examples of the first type. The latter type includes things like a course letter grade (such as A, B, C, D, or F), the amount of network traffic (such as low, medium, or high), and a Likert scale variable (such as 1 for strongly disagreeing, 2 for disagreeing, 3 for neutral, 4 for agreeing, and 5 for strongly agreeing). Because the categories in nominal variables lack

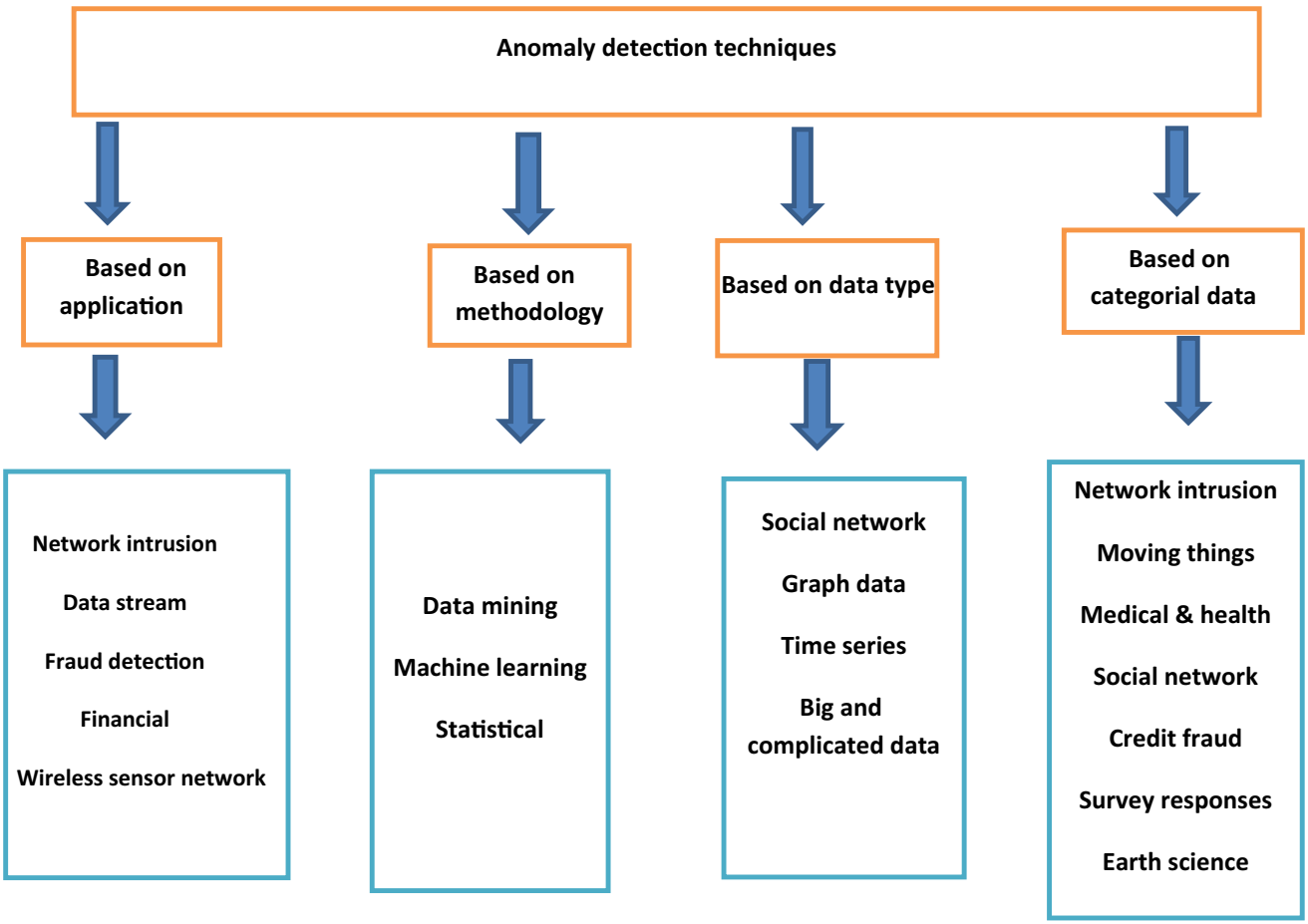

Fig. 5 Types of anomaly detection techniques

a natural ordering, finding anomalies in ordinal variables is easier than finding abnormalities in nominal data [131].

Applications of anomaly detection in categorical data are numerous. We mention here just few examples:

- *Network intrusions.* An intriguing use case for anomaly detection in cloud computing is the identification of users, who make unusual access [132]. Profiles and actions of users are typically coded as categorical variables in this context.
- *Moving things.* Anomaly detection can be applied to moving object data analysis to discover objects that move unexpectedly and to categorise the different sorts of road segments, where moving objects behave abnormally [113] and [133].
- *Medical and health data.* Anomaly detection techniques can help decision-makers by identifying instances of unreasonably-high medical expenses and/or negligence. The management of healthcare can be improved with the help of this information.
- *Social networks.* Anomaly detection is used in numerous applications in social networks, such as identifying unusual users in social groupings, who have varying interests, beliefs, and opinions [135] and [136].
- *Credit fraud.* When an unexpected activity is noticed, it may be possible to identify unauthorised credit card use. Behavioral data is frequently represented by categorical attributes [137].
- *Survey responses.* Survey responses are typically translated into category or ordinal properties. The authors of [138] emphasized the significance of anomaly detection in surveying data.
- *Earth science.* Finding anomalies in spatiotemporal data, e.g., weather patterns or climate changes in various geographical areas gives an explanation for interesting spatiotemporal patterns [139]
- *Law enforcement.* Examples of anomaly detection applications for law enforcement are discovering anomalies in trading activities and insurance claims [140].

Challenges facing anomaly detection in categorical data

The identification of anomalies in categorical data faces some challenges. These challenges include:

- Anomaly detection techniques concentrate on determining the statistical distribution patterns (patterns suggested by the majority of data), and then classify observations that deviate from the presumed patterns as anomalies [141]. The literature reveals a few distance functions to calculate the separation between categorical observations [142] and [143]. In categorical data, it might be challenging to detect patterns and calculate distances. As a result, techniques for anomaly identification are more frequently used with quantitative data than with categorical data.
- The literature contains a number of additional but distinct definitions of anomalies in categorical data [144]. Depending on the definition used, anomaly detection techniques can classify various sets of observations as being anomalous.
- There are very few benchmark datasets that may be used to evaluate how well anomaly detection techniques for categorical data perform in terms of computation time, detection rate, etc. Additionally, due to the lack of techniques that produce such data, it is difficult to develop synthetic categorical data with recognized abnormalities [145].
- Because most real applications have enormous datasets in terms of number of observations, number of categorical variables, and number of categories in each, computational complexity is a difficult problem in the identification of anomalies, especially in categorical data. As a result, temporal complexity is a major problem, when using anomaly detection techniques on categorical data.

Anomaly detection of categorical data based on indicator variables

The representation of categorical data by numerical values is a method for the detection of abnormalities in categorical data. Then, one can apply anomaly detection for quantitative data. Indicator Variables (IVs) [146] and a method based on Multiple Correspondence Analysis (MCA) [146] are two examples of methodologies that use this approach. The IV approach substitutes indicator variables for each classification variable [146]. Due to the computation of SVD as well as the quadratic cost of distance estimation, the MCA-based anomaly detection technique is computationally expensive.

Frequency-based techniques

Instead of using distances, frequency-based techniques employ frequency (i.e., the number of times a category occurs). Three different frequencies can be utilized to spot irregularities in categorical data: marginal frequency, itemset frequency, and diversified frequency.

Marginal frequency: Observations with low marginal frequencies are known as anomalies in categorical data. The Attribute Value Frequency (AVF) [147] and [148], Square of the Complement Frequency (SCF) [149], Weighted Attribute Value Frequency (WAVF) [150], Weighted Density-based

Outlier Detection (WDOD) [151], Cloud Model-based Outlier Detection (CMBOD) [152], and Bouguessa's technique [153] are some techniques that use this definition.

For each observation, X_i, a frequency score is computed as the basis for the AVF.

$$\mathrm{AVF}(X_i) = \frac{1}{q}\sum_{j=1}^{q} f(X_{ij}) \tag{1}$$

where the marginal frequency of X_{ij} within the variable X_j is denoted by $f(X_{ij})$. When M is a parameter selected by the user, the AVF identifies the M objects with the lowest AVF scores as outliers. Instead, the authors of [154] proposed estimating the number of anomalies M by assuming that the frequency scores exhibit a normal distribution.

For each observation, X_i, in the dataset, the Square of the Complement Frequency (SCF) gives an outlier score as

$$\mathrm{SCF}(x_i) = \sum_{j=1}^{q} \frac{\left[1 - p(X_{ij})\right]^2}{c_j} \tag{2}$$

where $p(X_{ij})$ is the marginal relative frequency of X_{ij} and c_j is the number of categories of the variable X_j (the number of occurrences within X_j divided by number of observations). The M observations with the highest outlying scores are then classified as outliers. The difference between frequent and infrequent categories is increased by SCF using the square of the complement frequency. Additionally, SCF takes the number of categories c_j into account to give variables with low c_j greater weights in the scoring function.

The sparseness of the frequencies in categorical variables is not taken into consideration by the AVF and SCF techniques. As a result, by providing weight to the frequencies, these techniques can better capture the sparsity of categorical data. The WAVF [150] and WDOD [151] are two instances of this trend.

A weighting formula is used in the WAVF to show the variable sparsity in calculating the outlier scores. The outlier score is more significantly impacted by the variable higher level of sparsity. As a result, observations with more sparse categories are more likely to be outliers. By using statistical functions for sparsity, such as the range or standard deviation of the marginal frequencies, one can determine the degree of sparsity of a categorical variable. The AVF can be extended using the WAVF as

$$\mathrm{WAVF}(X_i) = \frac{1}{q}\sum_{j=1}^{q} f(X_{ij}) \cdot R_j \tag{3}$$

where R_j is the jth categorical variable frequency range. The WDOD technique for categorical data is another way to weight the frequencies [151]. It is based on calculating the weighted density for the entire dataset after estimating the density of each categorical variable. To be more precise, WDOD begins by calculating a weight, $W(X_j)$, for each variable X_j as

$$W(X_j) = \frac{1 - E^c(X_j)}{\sum_{i=1}^{q} 1 - E^c(X_i)} \tag{4}$$

where,

$$E^c(X_j) = \sum_{i=1}^{c_j} p(X_{ij}) \cdot (1 - p(X_{ij})) \tag{5}$$

$p(X_{ij})$ is the category relative frequency, and c_j is the number of categories in X_j. The complement entropy, $E^c(X_j)$, is used in [155] to assess the information gain or uncertainty to assign various weights to categorical variables in order to highlight significance, according to [151]. In contrary to the Shannon entropy logarithmic tendency, the complement entropy gauges both fuzziness and uncertainty. The weighted relative frequencies are then added to determine an object WDOD score, which is calculated as follows:

$$\mathrm{WDOD}(X_i) = \sum_{j=1}^{q} p(X_{ij}) \cdot W(X_j) \tag{6}$$

As a result, the uncertainty in each variable density is taken into account by WDOD(X_i). An item is more likely to be a density-based outlier if its weighted density WDOD(X_i) is low. Therefore, the WDOD declares an observation, X_i, as an outlier.

The time complexity of the AVF, SCF, WAVF, and WDOD is $\approx O(nq)$, which linearly increases with the number of observations and the number of categorical variables.

The AVF, SCF, WAVF, WDOD, and CMBOD are fast and scalable techniques. They can efficiently deal with large-scale categorical datasets, but they only consider the marginal frequency and ignore any dependency among the categorical variables. Moreover, they require specifying the number of anomalies M in advance, which is impractical in real applications. In addition to the difficulty of defining the suitable values of these parameters, the results are very sensitive to those values.

Itemset frequency. The techniques in this group take into account the frequency of itemsets, which are combinations of categories with a maximum predetermined size. The term "anomalies" refers to observations with low itemset frequencies. These techniques begin by creating a collection of often-occurring itemsets (itemsets with a predetermined minimum frequency), and then classify them as outlier observations with fewer frequent item sets. Frequent Pattern Outlier Factors (FPOFs) [156], Link-based Outlier and Anomaly Detection in Evolving Datasets (LOADEDs) [157] and [158], Outlier Detection for Mixed Attribute Datasets

(ODMADs) [148], and Frequent Non-Derivable Itemsets-Outlier Detection (FNDI-OD) [148] are a few examples of these techniques.

Anomalies are defined by the FPOF, the categorical portion of LOADED, and the categorical component of ODMAD, as observations with uncommon patterns in their itemsets. These techniques depend on the minimum frequency of frequent itemsets and the maximum length to define the list of frequent itemsets (S).

A significant problem with the itemset-based techniques is that the number of frequent itemsets tends to be huge. For each observation, the set of frequent items is scanned to identify frequent items belonging to a certain observation. Thus, they take a very long time in processing. The time complexity of itemset frequency-based techniques can be attributed to two tasks; finding frequent itemsets and calculating outlying scores.

Diversified frequency Under this category, we discuss a method for identifying anomalies in categorical data, that is, Couple Biased Random Walk (CBRW) [159]. The CBRW method takes into account intra-feature coupling (distribution) among categories for the same categorical variable as well as inter-feature coupling (interactions) among categories for different categorical variables. First, to begin with, let m_j be the modal (the most frequent) category for the categorical variable X_j and $p(m_j)$ be the relative frequency of this modal category. Then, for each category X_{ij} in X_j, we compute an intra-feature coupling deviation score as

$$\delta\left(X_{ij}\right) = \frac{\left[\text{dev}\left(X_{ij}\right) + \text{base}\left(m_j\right)\right]}{2} \tag{7}$$

where

$$\text{dev}\left(X_{ij}\right) = \frac{\left[p\left(m_j\right) - p\left(X_{ij}\right)\right]}{p\left(m_j\right)} \tag{8}$$

and

$$\text{base}\left(m_j\right) = 1 - p\left(m_j\right) \tag{9}$$

Thus, the intra-feature coupling considers both the marginal frequencies of categories as well as the whole frequency distribution within each variable.

Second, a directed graph, G, is constructed to represent the interactions between category values in addition to inter-feature value coupling. Each category in this network is represented by a node, and edge connecting nodes u and v allow inter-feature coupling to pass between them.

$$A(u,v) = p(u|v) = \frac{p(u,v)}{p(v)} \tag{10}$$

The inter-feature coupling determines if one outlying category for a categorical variable is associated with another. $A(u,v)$ gauges how strongly u and v are coupled. Finally, using $A(u,v)$, CBRW constructs a biased random walk matrix W_b.

$$W_b(u,v) = \frac{\delta(v)A(u,v)}{\sum_{u\in V}\delta(v)A(u,v)} \tag{11}$$

where V is the collection of G nodes (all categories in the whole dataset). The transition from node u to node v is represented by $W_b(u,v)$, with a probability proportional to $\delta\,(v)A\,(u,v)$. To determine an outlying score for each category, CBRW first constructs W_b before generating the probability distribution of biased random walk column vector π_0. Then CBRW initializes π_0 values by a uniform distribution. Thereafter, it computes π_{t+1} at time $t+1$ on the basis of π_t as

$$\pi_{t+1} = (1-\alpha)\frac{1}{|V|}1 + \alpha W_b^t \pi_t \tag{12}$$

where α is a damping factor to guarantee convergence. The outlying v score for a certain category is computed as

$$\text{CBRW}(v) = \pi^*(v) \tag{13}$$

CBRW scores determine the importance of a certain categorical variable as

$$\text{Rel}(X_j) = \sum_{i=1}^{c_j} \text{CBRW}\left(X_{ij}\right) \tag{14}$$

where X_{ij} is the i^{th} category in the j^{th} variable and c_j is number of categories in the j^{th} variable. CBRW scores can identify outliers by computing an outlying score for each observation X_i as:

$$\text{CBRW} - \text{OD}\left(X_i\right) = \sum_{j=1}^{q} W_j \cdot \text{CBRW}\left(X_{ij}\right) \tag{15}$$

where

$$W_j = \frac{\text{Rel}(X_j)}{\sum_{j=1}^{q}\text{Rel}(X_j)} \tag{16}$$

Accordingly, CBRW labels observations having the highest M CBRW-OD scores as outliers. One advantage of the CBRW is that it takes into consideration correlation among the categorical variables, since it captures frequency distribution of categorical variables as well as inter-feature coupling. However, it is computationally expensive as it has a quadratic complexity with respect to the number of

categories. Moreover, it requires two parameters: the number of outliers M and the damping factor α.

Bayesian/conditional frequency-based techniques

These techniques define categories of anomalies, differently. They look for observations that have low joint frequencies and high marginal frequencies. In other words, they define anomalies as observations with uncommon category combinations, when the categories themselves are common. The abnormalities found by conditional anomaly detection techniques differ significantly from those found by conventional anomaly detection techniques. Conditional anomaly detection techniques search for observations that have common categories that are infrequently observed together. Anomaly Pattern Detection (APD), Conditional Algorithm (CA) [160], Attribute Association (AA) algorithm [161], and the method suggested in [162] (denoted below by RHH, the first letters of the authors' names) are a few examples of techniques approaches that use the conditional anomaly detection strategy.

Outliers are defined with CA as observations having uncommon combinations of common categories, [49]. The CA determines the ratio $r(X_{ij}, X_{ik})$ between the categories X_{ij} and X_{ik} in the categorical variables X_j and X_k as a measure of rarity. The definition of the $r(X_{ij}, X_{ik})$ ratio is

$$r(X_{ij}, X_{ik}) = \frac{p(X_{ij}, X_{ik})}{p(X_{ij})p(X_{ik})} \tag{17}$$

where the marginal and joint relative frequencies, respectively, are denoted by $p(X_{ij})$ and $p(X_{ij}, X_{ik})$. A low r-ratio indicates a higher likelihood of anomalous co-occurrence, since the marginal probabilities multiply considerably more frequently than their combined probabilities.

The Anomaly Pattern Detection (APD) method was proposed to detect anomalous patterns, groups of related observations having outliers percentage higher than expected [163]. The APD method depends on the CA algorithm [160] as a first step to identify individual outliers. Then, a rule-based technique is used to study the behavior of outliers in each pattern. Similar to the conditional probability method, the RHH method [162] searches for observations with frequent attribute values but infrequent joint co-occurrence. It can find anomalies in categorical datasets as well as mixed datasets. In the mixed datasets, it transforms quantitative variables into categorical variables by discretizing quantitative variables into fixed-length intervals. In the training phase, it builds a Bayesian network (see, e.g., [164] and [165]) to capture the dependency among attributes.

The AA algorithm [161] defines conditional anomalies as observations that contain frequent categories, but their itemsets are rarely observed together. It starts with deriving a set of association rules with high confidence from the data. Then, it computes an outlying score called outlier degree.

Conditional anomaly detection techniques have time complexity problems, since they require a combinatorial time for building itemsets and a long time for searching in the high conditional probability space. Moreover, they require many parameters.

Density-based techniques

The goal of the density-based anomaly or local anomaly detection strategy is to locate observations that deviate from the norm in their immediate surroundings [167]. Local anomalies differ from global anomalies, which do not only include observations made locally but also those made globally [166–168], and [169]. Global anomalies are incongruous with the pattern given by the majority of all other observations. The Hyperedge-based Outlier Test (HOT), the k-Local anomalies Factor (k-LOF), and the WATCH method are three local anomaly identification methods for categorical data. The k-LOF computes another type of similarity named the accumulated similarity of k-walk between two observations X_i and X_j, as

$$s^k(X_i, X_j) = \sum_{i=1}^{k} s^i(X_i, X_j) \tag{18}$$

Then, the above-mentioned two similarity measures are combined in an outlying score for X_i, which is given by

$$k - \mathrm{LOF}(X_i) = \frac{1}{s^k(X_i, X_j) \times n} \sum_{j=0}^{n} s^k\left(X_i, X_j\right) \tag{19}$$

The k-LOF labels an observation, X_i, as an outlier if k-LOF$(X_i) > \theta$, where θ is a predefined parameter. The k-LOF takes into consideration the direct relationships between an observation and its direct neighbors as well as the indirect relationships among the neighbors and the neighbors' neighbors, where $k > 1$. Thus, it requires the parameters θ and the maximum length of indirect relationships k.

The time complexity of k-LOF is attributed to two tasks: building a similarity graph G and computing similarity and outlying scores.

Clustering-based techniques

Categorical datasets are clustered using clustering-based anomaly detection techniques, which subsequently classify observations in sparse regions as outliers. This category of techniques includes Rough-ROAD [170] and [171] and Ranking-based Outliers Analysis and Detection (ROAD) [172] and [173].

Frequency-based and clustering-based outliers are the two categories of outliers that are defined by the ROAD. Observations with rare categories are considered frequency-based outliers. However, observations with uncommon combinations of common categories are considered clustering-based outliers.

A ranking system is created by ROAD for each sort of outlier. The average marginal frequencies are first calculated for each observation as density scores, where

$$\text{den}(X_i) = \frac{1}{q}\sum_{j=1}^{q} f\left(X_{ij}\right) \tag{20}$$

which is equivalent to AVF(X_i) in Eq. (1). Observations are sorted by ROAD according to their densities. Then, ROAD assigns those observations with low density scores higher likelihoods of being frequency-based outliers. Second, ROAD partitions the given categorical dataset into k clusters using the $k-$mode algorithm [174]. Then, it defines the set of big clusters, BC, as the clusters that contain at least $\alpha\%$ of observations, where α is the big cluster threshold.

The time complexity of ROAD is attributed to three tasks. Firstly, building the frequency-based ranking scheme requires $O(nqc_{max})$, where c_{max} is maximum number of categories per categorical variable. Secondly, the the cluster-based ranking scheme needs $O\left(nqk^2 + nqkt\right)$ computations, where k is the number of clusters and t is the number of iterations required for convergence in the $k-$mode algorithm. Thirdly, the complexity of ranking observation is $O(n\log n)$. Therefore, the overall complexity of ROAD is $O(nqc_{max} + nqk^2 + nqkt + n\log n)$. The RoughROAD is as complex as that of ROAD. Therefore, the clustering-based techniques are computationally expensive.

Distance-based techniques

The concept of distance-based anomalies for quantitative data is expanded by the distance-based anomaly detection techniques for categorical data ([175–177] and [178]). Anomalies have numerous definitions in distance-based techniques. These consist of

- Anomalies are the M observations whose average distances to the k nearest neighbors are the greatest [178] and [179].
- Anomalies are the M observations whose distances to the k-th nearest neighbor are the greatest [177].
- Anomalies are the observations that have fewer than p observations within a certain distance d [175] and [176].
- Anomalies are observations that have the highest z-scores of the average distances to the k-nearest neighbors. That is, first we compute the average distance of each observation to its k-nearest neighbors. Then, we standardize these average distances and obtain their z-scores. The observations with z-scores greater than a threshold θ (e.g., 3) are declared as outliers. This method does not require the parameter M in advance, but assumes that the z-scores follow a standard normal distribution to help in choosing the value of θ.

Examples of distance-based techniques for categorical data are (a) Orca (name of software) [180], (b) a method called iOrca [181], (c) the Common Neighbor Based distance (CNB) [182], and (d) the Recursive Binning and Re-Projection (RBRP) [183]. Distance-based anomaly detection techniques are very sensitive to the number of nearest neighbors.

Compression-based techniques

Compression algorithms are usually used in the fields of communication and storage rather than in data mining. Recently, few compression-based anomaly detection techniques have been proposed based on the fact that anomalies do not comply with the model suggested by the other points in the data. Accordingly, observations that could not compress well are considered as outliers. These techniques look for the best compression model that suits the non-outlying data points. Objects that deviate (have bad compression measures) are highlighted as anomalies [184]. Examples of compression-based anomaly detection techniques for categorical data are (a) KRIMP [185] and (b) Comprex [186].

KRIMP is based on the idea of code tables. A code table consists of two columns. The first column contains the itemsets and the second column contains their codes. Itemsets are sorted in descending order according to their lengths then their frequencies. Higher order (that is, longer and more frequent) itemsets take shorter codes. Each observation is represented by a set of non-overlapping itemsets that completely cover all values of that observation. An anomaly can be seen as an observation, which contains infrequent itemsets, and hence, its code is longer than those of other observations.

Although compression-based anomaly detection techniques for categorical data outperform other relevant techniques in terms of error and detection rates, they are computationally very expensive, especially for datasets that contain large numbers of variables. Similar to other anomaly detection techniques for categorical data, compression-based techniques require decision parameters to decide whether an observation is or is not an outlier

Finally, existing anomaly detection techniques for categorical data face many problems. Important problems are discussed below:

- Computational complexity
- Human intervention:

Human intervention (input parameters) is a problem for anomaly detection techniques for categorical data. Existing techniques require one or more input parameters. In real applications, defining the suitable values of these parameters is a hard and critical task. In addition, the results are sensitive to these input parameters.

Existing anomaly detection techniques for categorical data compute a score for each observation. To identify whether an observation is an outlier, they require one of the following parameters:

(1) M: The number of assumed anomalies in the dataset. Most of existing methods (AVF, SCF, CMBOD, FPOF, FNDI-OD, CBRW-OD, WATCH, ROAD, Rough-ROAD, CNB, Orca, CNB, LSA, GA, ITB-SP, EEB-SP, ITB-SS, EEB-SS and CompreX) require M in advance.

(2) Itemset-based parameters: The minimum frequency and the maximum length are used in defining frequent itemsets. They are required by FPOF, LOADED, ODMAD, FNDI-OD, HOT, AA, CA and KRIMP.

(3) k: The number of nearest neighbors in distance-based anomaly detection techniques for categorical data is required, especially for Bouguessa's method, Orca, CNB, and k-LOF.

Conclusions

In this review article, we have covered a variety of research techniques for anomaly detection and their use in a number of different fields as discussed previously in [187]. Anomaly detection techniques can be classified based on application, methodology, data type and finally categorical data. We discussed various techniques for detecting anomalies, which are very important and beneficial in many actual applications, such as identifying computer network intrusions and fraud detection. We discussed the strengths and weaknesses of most of these techniques, and discussed various application domains. Finally, this paper surveyed some available techniques for the identification of anomalies from categorical data in the statistics as well as machine learning and computer science literatures. There is no overall agreement on a distinct definition of an anomaly in categorical data. We reviewed several techniques for anomaly detection from categorical data. We identified the strengthes and weaknesses of each technique. In addition, we have discussed the common challenges in the detection of anomalies from categorical data. There are several directions for further research in anomaly detection from categorical data. Extending categorical data anomaly detection to novel fields of study, such recommender systems [188], categorical data streams [189], moving objects [190], and information network may be studied in future research. Another area for future research is to define automatic critical values rather than predetermining the number of abnormalities.

References

1. L. Ruff, J. Kauffmann, R. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. Dietterich, K. M¨uller, *A Unifying Review of Deep and Shallow Anomaly Detection* (IEEE, 2021)
2. H. Hojjati, T. Ho, N. Armanfard, *Self-Supervised Anomaly Detection: A Survey and Outlook*, (IEEE, 2022)
3. V. Hodge, J. Austin, A survey of outlier detection methodologies. Arti. Int. Rev. **22**(10), 85–126 (2004)
4. R. Feinman, R. Curtin, S. Shintre, A. Gardner, Detecting adversarial samples from artifacts. arX., 0410 (2017)
5. K. Lee, K. Lee, H. Lee, J. Shin, A simple unified framework for detecting out-of-distribution samples and adversarial attacks. Adv. Neur. Info. Proc. Syst. 31 (2018)
6. V. Jumutc, J.A. Suykens, Multi-class supervised novelty detection. IEEE Trans. Pattern. Anal. Mach. Intell. **36**(12), 2510–2523 (2014)
7. S. Kim, Y. Choi, M. Lee, Deep learning with support vector data description. Neur. Comput. **165**, 111–117 (2015)
8. T. Ehret, A. Davy, J. M. Morel, M. Delbracio, Image anomalies: a review and synthesis of detection methods. J. Math. Imaging Vis. (2018)
9. V. Chandola, A. Banerjee, V. Kumar, Outlier detection: a survey, ACM Comp. Surv. **14**(15) (2009)
10. G. Pang, C. Shen, L. Cao, A. V. D. Hengel, Deep learning for anomaly detection: a review, ACM Comp. Surv. **54**(2) (2021)
11. R. Chalapathy, S. Chawla, Deep learning for anomaly detection: a survey (2019)
12. R. Chalapathy, S. Chawla, Deep learning for anomaly detection: a survey, Res. Gate (2019)
13. H. Peng, R. Marculescu, Multi-scale compositionality: identifying the compositional structures of social dynamics using deep learning. PLoS ONE **10**(4), e0118309 (2015)
14. X. Song, M. Wu, C. Jermaine, S. Ranka, Conditional anomaly detection. IEEE Trans. Knowl. Data Eng. **19**, 631–645 (2007)
15. S. Ramaswamy, R. Rastogi, K. Shim, Efcient algorithms for mining outliers from large datasets, ACM. SIGM. ICMD, pp. 427–438 (2000)
16. F. Angiulli, C. Pizzuti, Fast outlier detection in high dimensional spaces. ECP. DMKD, pp. 15–26 (2002)
17. M. Breunig, H. Kriegel, R. Ng, J. Sander, LOF: identifying density-based local outliers, ACM. SIGM. Rec., 29(93–104) (2000)
18. H. Kriegel, P. Kroger, E. Schubert, A. Zimek, LoOP: local outlier probabilities. ACM. CIKM. **09**, 1649–1652 (2009)
19. H. Ville, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph. IEEE, ICPR. **3**, 330–433 (2004)
20. H. Huang, K. Mehrotra, C. Mohan, Rank-based outlier detection. J. Stat. Comput. Simlut. **83**, 518–531 (2013)
21. H. P. Kriegel, P. Kroger, E. Schubert, A. Zimek, Outlier detection in axis-parallel subspaces of high dimensional data, Asia Conf. AKDDM., pp. 831–838 (2009)
22. J. Zhang, Y. Jiang, K.H. Chang, S. Zhang, J. Cai, L. Hu, A concept lattice based outlier mining method in low dimensional subspaces. Patt. Reco. Lett. **15**, 1434–1439 (2009)

23. J. Zhang, X. Yu, Y. Li, S. Zhang, Y. Xun, X. Qin, A relevant subspace based contextual outlier mining algorithm. Knowl. Based. Syst. **99**(72), 1–9 (2016)
24. J. Dutta, B. Banerjee, C. Reddy, RODS: rarity based outlier detection in a sparse coding framework. IEEE, TKDE. **28**(2), 483–495 (2016)
25. E. Muller, I. Assent, U. Steinhausen, T. Seidl, OutRank: ranking outliers in high dimensional data. IEEE 24th ICDE., pp. 600–603, (2008)
26. E. Muller, M. Schifer, T. Seidl, Adaptive outlierness for ¨ subspace outlier ranking, in 19th Int. Conf. Info. CIKM, vol. 10, pp. 1629–1632 (2010)
27. A. Lazarevic, V. Kumar, Feature bagging for outlier detection, KDD, in 11th ACM. SIGK, pp. 157–166 (2005)
28. F. Keller, E. Muller, K. ohm, HiCS: High contrast ¨ subspaces for density-based outlier ranking, IEEE 28th, ICDE, pp. 1037–1048 (2012)
29. B. Stein, M. Leeuwen, T. Back, Local subspacebased outlier detection using global neighbourhoods, in 4th IEEE, ICBD, pp. 1136–1142 (2016)
30. A. Zimek, M. Gaudet, R. Campello, J. Sander, Subsampling for efcient and efective unsupervised outlier detection ensembles, in 19th ACM, KDD, pp. 428–436 (2013)
31. J. Pasillas-Diaz, S. Ratte, Bagged subspaces for unsupervised outlier detection. IJCI. **33**(3), 507–523 (2017)
32. A. Ghoting, M. Otey, S. Parthasarathy, "LOADED: Linkbased outlier and anomaly detection in evolving data sets, Fourth IEEE, ICDM, pp. 387–390 (2004)
33. A. Koufakou, M. Georgiopoulos, A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes. Data Mining Knowl Discov **20**(2), 259–289 (2010)
34. K. Zhang, H. Jin, An efective pattern based outlier detection approach for mixed attribute data. AI, LNCS. **6464**, 122–131 (2010)
35. Y. Lu, F. Chen, Y. Wang, C. Lu, Discovering anomalies on mixed-type data using a generalized student-t based approach. Exp. Syst. Appl. **28**(10), 1–10 (2016)
36. K. Do, T. Tran, D. Phung, S. Venkatesh, "Outlier detection on mixed-type data: an energy-based approach", ADMA., SIP.,111–125, (2016)
37. T. Ehret, A. Davy, J. Morel, M. Delbracio, " Image Anomalies: a Review and Synthesis of Detection Methods", Math. Img. and Vis., (2018)
38. L. Manevitz, M. Yousef, " One-Class SVMs for Document Classification", Jour. of Mach. Lear. Res., 139–154, (2001)
39. V. Roth, " Outlier Detection with One-class Kernel Fisher Discriminants", CANIPS., 17, (2004)
40. V. chandola, A. Banerjee, V. kumar," Anomaly Detection: A Survey", ACM Comp. Sur., 1–72, (2009)
41. P. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining", Add., Wesl., (2005)
42. A. Jain, R. Dubes, *Algorithms for Clustering Data* (Hall Inc, Pren., 1988)
43. S. Basu, M. Bilenko, R. Mooney, *A Probabilistic Framework for Semi-Supervised Clustering*, tenth ACM SIGKDD. ACM Press, pp. 59–68 (2004)
44. M. Ester, H. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, ICKDDM, 226–231, (1996)
45. S. Guha, R. Rastogi, K. Shim, ROCK: A robust clustering algorithm for categorical attributes, Inf. Sys., 25(5), 345–366, (2000)
46. L. Ert¨oz, M. Steinbach, V. Kumar, Finding topics in collections of documents: a shared nearest neighbor approach, CIR, pp. 83–104, (2003)
47. H. Motulsky, *Intuitive Biostatistics: Choosing a statistical test*, Oxford University Press, Oxford (1995)
48. P. Oberdiek, M. Rottmann, H. Gottschalk, 'Classification uncertainty of deep neural networks based on gradient information, CoRR, 1805–08440 (2018)
49. K. Lee, K. Lee, H. Lee, J. Shin, A simple unified framework for detecting out-of-distribution samples and adversarial attacks, Adv. Neu. Inf. Proc. Syst., 7167–7177 (2018)
50. Y. Bahat, G. Shakhnarovich, Confidence from invariance to image transformations, arXiv 1804-00657 (2018)
51. J. Metzen, T. Genewein, V. Fischer, B. Bischoff, "On detecting adversarial perturbations, arXiv 1702-04267 (2017)
52. J. Ren, P. J. Liu, E. Fertig, J. Snoek, R. Poplin, M. DePristo, J. Dillon, B. Lakshminarayanan, Likelihood ratios for out-of distribution detection, arXiv 1906-02845 (2019)
53. Q. Yu, K. Aizawa, " Unsupervised out-of-distribution detection by maximum classifier discrepancy", IEEE Int. Conf. Comp. Vis., 9518–9526, (2019)
54. D. Hendrycks, K. Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv 1610-02136, (2016)
55. S. Liang, Y. Li, R. Srikant, Enhancing the reliability of outofdistribution image detection in neural networks, Int. Conf. Learn. Repr. (2018)
56. P. Schulam, S. Saria, Can you trust this prediction? Auditing pointwise reliability after learning. Mach. Learn. Res. **89**, 1022–1031 (2019)
57. Y. Kantaros, T. Carpenter, S. Park, R. Ivanov, S. Jang, I. Lee, J. Weimer, Vision Guard: runtime detection of adversarial inputs to perception systems. arXiv 09792 (2020)
58. T. Bailetti, M. Gad, A. Shah, Intrusion learning: an overview of an emergent discipline. Tech. Inn. Man. Rev. **6**(2), 15–20 (2016)
59. U. Bakar, H.Ghayvat, S. Hasanm, S. Mukhopadhyay, Activity and anomaly detection in smart home: a survey, Mukh., Nex. Gen. Sens. and Sys., Spr., pp. 191–220 (2016)
60. Z. Bakar, R. Mohemad, A. Ahmad, M. Deris, A comparative study for outlier detection techniques in data mining, IEEE ICCIS, pp. 1–6 (2006)
61. V. Barnett, T. Lewis, *Outliers in Statistical Data, 3rd* (Wiley, New York, 1994)
62. S. Bay, M. Schwabacher, Mining distance-based outliers in near linear time with randomization and a simple pruning rule, ACM, SIGKDD, pp. 29–38 (2003)
63. E. Beh, Simple correspondence analysis of nominal-ordinal contingency tables. App. Math. Dec. Sci. **228**, 1–17 (2008)
64. A. Beldar, V. Wadne, The detail survey of anomaly/outlier detection methods in data mining. Int. Mult. Cur. Res. **3**, 462–472 (2015)
65. K. Singh, S. Upadhyaya, Outlier detection: applications and techniques. Comput. Sci. Issues. **9**(1), 307–323 (2012)
66. Ghosh, D., Vogt, A.: Outliers: an evaluation of methodologies. Join. Stat. Meet., pp. 3455–3460 (2012)
67. X. Ding, Y. Li, A. Belatreche, L. Maguire, An experimental evaluation of novelty detection methods. Neur. Comput. **135**, 313–327 (2014)
68. K. Malik, H. Sadawarti, G. Kalra, Comparative analysis of outlier detection techniques. Comput. Appl. **97**(8), 12–21 (2014)
69. K. Divya, N. Kumaran, Survey on outlier detection techniques using categorical data. Int. Eng. Technol. **3**, 899–904 (2016)
70. P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, E. Vázquez, Anomaly-based network intrusion detection: Techniques, systems and challenges. Comput. Secur. **28**(1), 18–28 (2009)

71. P. Gogoi, D. Bhattacharyya, B. Borah, J. Kalita, A survey of outlier detection methods in network anomaly identification. Comput. J. **54**(4), 570–588 (2011)
72. G. Golub, v. Loan, Matrix computations, 3rd edit. (2012)
73. T. Bailetti, M. Gad, A. Shah, Intrusion learning: an overview of an emergent discipline. Tech. Innov. Man. Rev. **6**(2), 15–20 (2016)
74. M. Ahmed, A. Mahmood, J. Hu, A survey of network anomaly detection techniques. Netw. Comput. Appl. **60**, 19–31 (2016)
75. W. Fahy, B.J. Borghetti, A. Sodemann, A survey of distance and similarity measures used within network intrusion anomaly detection. IEEE Commun. Surv. Tutor. **17**(1), 70–91 (2015)
76. Y. Zhang, N.H. Meratnia, Outlier detection techniques for wireless sensor networks: a survey. IEEE Commun. Surv. Tutor. **12**(2), 159–170 (2010)
77. M. Rassam, M.A. Maarof, M. Zainal, A survey of intrusion detection schemes in wireless sensor networks. Appl. Sci. **9**(10), 1636–1652 (2012)
78. J. Daniel, V. Joshna, S. Manjula, A survey of various intrusion detection techniques in wireless sensor networks. Comput. Sci. Mob. Comput. **2**(9), 235–246 (2013)
79. A. Mahapatro, A. Khilar, Fault diagnosis in wireless sensor networks: a survey. IEEE Commun. Surv. Tutor. **15**(4), 2000–2026 (2013)
80. R. Kumar, T. Kaur, Outlier detection in wsn: a survey. Adv. Res. Comput. Sci. Softw. Eng. **3**(7), 609–617 (2013)
81. A. Abduvaliyev, A. Pathan, K. Zhou, J. Roman, R. Wong, On the vital areas of intrusion detection systems in wireless sensor networks. IEEE Commun. Surv. Tutor. **15**(3), 1223–1237 (2013)
82. M. Rassam, A. Zainal, A. Maarof, Advancements of data anomaly detection research in wireless sensor networks: a survey and open issues. Sensors **13**(8), 10087–10122 (2013)
83. A. Abukhalaf, H. Wang, J. Zhang, Outlier detection techniques for localization in wireless sensor networks: a survey. Fut. Gen. Commun. Netw. **8**(6), 99–114 (2015)
84. C. Shannon, A mathematical theory of communication. Bell Tele. Syst. Tech. Publ. **27**(3), 379–423 (1948)
85. M. Marinho, J. Granjal, J. Monteiro, A survey on security attacks and countermeasures with primary user detection in cognitive radio networks. EURASIP Inf. Secur. **1**, 1–14 (2015)
86. O. Reilly, C. Gluhak, A. Imran, M. Rajasegarar, Anomaly detection in wireless sensor networks in a non-stationary environment. IEEE Commun. Surv. Tutor. **16**(3), 1413–1432 (2014)
87. S. Shahid, N. Naqvi, I. Qaisar, Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey. AIR. **43**(2), 193–228 (2015)
88. A. Sodemann, A. Ross, M. Borghetti, A review of anomaly detection in automated surveillance. IEEE Trans. Syst **42**(6), 1257–1272 (2012)
89. S. Archana, N. Pawar, Survey on outlier pattern detection techniques for time-series data. IJSR **1**(1), 1852–1856 (2014)
90. J. Faria, R. Gonçalves, A.G. deCarvalho, Novelty detection in data streams. AI. Rev. **45**(2), 235–269 (2015)
91. A. Deshmukh, M. Kapse, A survey on outlier detection technique in streaming data using data clustering approach. Eng. Comput. Sci. **5**(1), 15453–15456 (2016)
92. J. Zhang, Advancements of outlier detection: a survey. ICST Tran. Scal. Info. Sys. **13**(1), 1–26 (2013)
93. P. Purankar, R. Patil, A survey paper on an effective analytical approaches for detecting outlier in continuous time variant data stream. Eng. Comput. Sci. **4**(11), 14946–14949 (2015)
94. M. Joudaki, H. Rashidian, A. Minaei-Bidgoli, B. Mahmoodi, M. Geraili, B. Nasiri, M. Arab, Using data mining to detect health care fraud and abuse: a review of literature. Heal. Sci. **7**(1), 194–202 (2015)
95. S. Cousineau, D. Chartier, Outliers detection and treatment: a review. Psyc. Res. **3**(1), 58–67 (2015)
96. S. Cho, H. Eo, Outlier detection for mass spectrometric data, Spri. Stat. Anal. Prot., 91–102m (2016)
97. P. Bezerra, G. Costa, B. Guedes, L. Angelov, A comparative study of autonomous learning outlier detection methods applied to fault detection, IEEE Int. Conf. on Fuz. Sys., FUZZ-IEEE, pp. 1–7 (2015)
98. R. Phua, C. Lee, S.Smith-Miles, K. Gayler, A comprehensive survey of data mining-based fraud detection research, pp. 1–14 (2010)
99. S. Pawar, S. Amruta, D. Tambe, A survey on outlier detection techniques for credit card fraud detection. IOSR Comput. Eng. **16**(2), 44–48 (2014)
100. A. Kathiresan, V. Vasanthi, A survey on outlier detection techniques useful for financial card fraud detection. IJIET **6**(1), 226–235 (2015)
101. R. Ahmed, M. Mahmood, N. Islam, A survey of anomaly detection techniques in financial domain. Fut. Gen. Comput. Syst. **55**, 278–288 (2016)
102. M. West, J. Bhattacharya, Intelligent financial fraud detection: a comprehensive review. Comput. Sec. **57**, 47–66 (2016)
103. R. Dhimmar, J. Chauhan, A survey on profile-injection attacks in recommender systems using outlier analysis. Comput. Sci. Man. Stud. **2**(12), 356–359 (2014)
104. P. Dokas, P. Ertoz, L. Kumar, V. Lazarevic, A. Srivastava, J. Tan, Data mining for network intrusion detection, NSF Work. Data Min., pp. 21–30 (2002)
105. S. Kaur, R. Singh, A survey of data mining and social network analysis based anomaly detection techniques. Egypt. Inf. J. **39**, 1–18 (2015)
106. E. Ajitha, P. Chandra, A survey on outliers detection in distributed data mining for big data. Basic Appl. Sci. Res. **5**(2), 31–38 (2015)
107. S. Markou, M. Singh, Novelty detection: a review-part 2: neural network based approaches. Sig. Proc. **83**, 2499–2521 (2003)
108. S. Ankur, Y. Singh, Oulier analysis using frequent pattern mining: a review. Comput. Sci. Inf. Technol. **5**(1), 47–50 (2014)
109. B. Said, A. Dominic, D. Samir, Outlier detection scoring measurements based on frequent pattern technique. Appl. Sci. Eng. Technol. **6**(8), 1340–2134 (2013)
110. L. Pimentel, M. Clifton, D. Clifton, L. Tarassenko, A review of novelty detection. Sig. Proc. **99**, 215–249 (2014)
111. S. Markou, M. Singh, Novelty detection: a review-part 1: statistical approaches. Sig. Proc. **83**, 2481–2497 (2003)
112. M. Hadi, S. Imon, A. Werner, Etection of outliers. Wiley Inter. Rev Comput. Stat. **1**, 57–70 (2009)
113. K. Ge, Y. Xiong, H. Zhou, Z.-H. Ozdemir, H. Yu, J. Lee, Top-eye: top-k evolving trajectory outlier detection, ACM, CIKM, pp. 1–4 (2010)
114. T. Dave, D. Varma, A review of various statistical methods for outlier detection. IJCSET **5**(2), 137–140 (2014)
115. H. Cho, S. Eo, Outlier detection for mass spectrometric data. Methods Mol. Biol. **1362**, 91–102 (2016)
116. T. Rezaei, A. Kasirun, M. Rohani, V. Khodadadi, Anomaly detection in online social networks using structure-based technique, ICITST, pp. 619–622 (2013)
117. Q. Savage, D. Zhang, X. Yu, X. Chou, P. Wang, Anomaly detection in online social networks. Soc. Netw. **39**, 62–70 (2014)
118. Y. Yu, R. Qiu, H. Wen, Z. Lin, C. Liu, A survey on social media anomaly detection, pp. 1–24, (2016)
119. S. Sarma, S. Sarma, A survey on different graph based anomaly detection techniques. India J. Sci. Technol. **8**(31), 1–7 (2015)
120. F. Ranshous, S. Shen, S. Koutra, D. Harenberg, S. Faloutsos, C. Samatova, Anomaly detection in dynamic networks: a survey. Wiley. Inter. Rev. Comput. Stat. **7**(3), 223–247 (2015)

121. D. Akoglu, L. Tong, H. Koutra, Graph based anomaly detection and description: a survey. Data Min. Knowl. Discov. **29**(3), 626–688 (2015)
122. J. Gupta, M. Gao, J. Aggarwal, C. Han, Outlier detection for temporal data: a survey. IEEE Trans. Knowl. Data Eng. **26**(9), 2250–2267 (2014)
123. J. Gupta, M. Gao, J. Aggarwal, C. Han, Outlier detection for temporal data. SLDMKD **5**(1), 1–129 (2014)
124. R. Devi, R. Amalraj, Hubness in unsupervised outlier detection techniques for high dimensional data–a survey. Comput. Appl. Tech. Res. **4**(11), 797–801 (2015)
125. J. Fanaee-T, H. Gama, Tensor-based anomaly detection: an interdisciplinary survey. Knowl. Based. Syst. **98**, 130–147 (2016)
126. S. Aggarwal, C. Yu, Outlier detection for high dimensional data, ACM, SIGMOD, pp. 37–46, (2001)
127. H. Zimek, A. Schubert, E. Kriegel, A survey on unsupervised outlier detection in high-dimensional numerical data. Stat. Anal. Data. Min. **5**(5), 363–387 (2012)
128. M. Supriya, G. Shinde, Outliers detection using subspace method: a survey. Comput. Appl. **112**(16), 20–22 (2015)
129. A. Otey, E. Parthasarathy, S. Ghoting, An empirical comparison of outlier detection algorithms, ACM, SIGKDD, pp. 1–8 (2005)
130. E. Campos, O. Zimek, A. Sander, J. Campello, J. Micenková, B. Schubert, E. Assent, I. Houle, On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. Data Min. Knowl. Disc. **30**(4), 891–927 (2016)
131. Y. Wang, Statistical techniques for network security: modern statistically-based intrusion detection and protection, IGI Glo (2008)
132. A. Sari, A review of anomaly detection systems in cloud networks and survey of cloud security measures in cloud storage applications. Info. Sec. **6**(02), 142–154 (2015)
133. X. Liu, W. Zheng, Y. Chawla, S. Yuan, J. Xie, Discovering spatio-temporal causal interactions in traffic data streams, ACM, SIGKDD, pp. 1010–1018 (2011)
134. M. Ieva, F. Paganoni, Detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models. Heal. Care Man. Sci. **18**(2), 166–172 (2015)
135. S. Aggarwal, C. Zhao, Y. Yu, Outlier detection in graph streams, ACM IEEE, ICDE, pp. 399–409, (2011)
136. J. Gao, J. Liang, F. Fan, W. Wang, C. Sun, Y. Han, On community outliers and their efficient detection in information networks, ACM, SIGKDD, pp. 813–822 (2010)
137. W. Wang, Y. Xu, Leveraging deep learning with lda-based text analytics to detect automobile insurance fraud. Dec. Sup. Sys. **105**, 87–95 (2018)
138. K. Zijlstra, W. van der Ark, L. Sijtsma, Outliers in questionnaire data: Can they be detected and should they be removed. Edu. and Beh. Stat. **36**, 186–212 (2011)
139. C. Liu, X. Chen, F. Lu, On detecting spatial categorical outliers. Geo. Inf. **18**(3), 501–536 (2014)
140. C. Aggarwal, *Outlier Analysis*, 2nd edn. (Springer, Berlin, 2017)
141. P. Billor, N. Hadi, A. Velleman, Blocked adaptive computationally-efficient outlier nominators. Comput. Stat. Data Anal. **34**, 279–298 (2000)
142. V. Boriah, S. Chandola, V. Kumar, Similarity measures for categorical data: a comparative evaluation, SIAM, SDM, pp. 243–254 (2008)
143. V. Chandola, V. Boriah, S. Kumar, A framework for exploring categorical data, SIAM, SDM, 187–198 (2009)
144. S. Wu, S. Wang, Parameter-free anomaly detection for categorical data. Machine learning and data mining in pattern recognition. Lecture Notes Comput. Sci. **6871**, 112–126 (2011)
145. S. Taha, A. Hadi, A general approach for automating outliers identification in categorical data, ACS/IEEE (AICCSA), pp 1–8 (2013)
146. T. Shyu, M. Sarinnapakorn, K. Kuruppu-Appuhamilage, I. Chen, S. Chang, W. Goldring, Handling nominal features in anomaly intrusion detection problems, Work. Res. Iss. Dat. Eng. STDMA., 55–62 (2005)
147. K. Koufakou, A. Ortiz, E. Georgiopoulos, M. Anagnostopoulos, G. Reynolds, A scalable and efficient outlier detection strategy for categorical data, IEEE, ICTAI, pp. 210–217 (2007)
148. G. Koufakou, A. Georgiopoulos, M. Anagnostopoulos, Detecting outliers in high-dimensional datasets with mixed attributes, DMIN (2008)
149. O. Taha, A. Hegazy, A proposed outliers identification algorithm for categorical data sets, INFOS, 1–5, (2010)
150. E. Rokhman, N. Subanar, Winarko, Improving the performance of outlier detection methods for categorical data by using weighting function. Theor. App. Inf. Technol. **83**, 327–336 (2016)
151. F. Zhao, X. Liang, J. Cao, A simple and effective outlier detection algorithm for categorical data. Mach. Learn. Cybern. **5**, 469–477 (2014)
152. L. Lei, D. Zhang, L. Zhang, Cloud model-based outlier detect algorithm for categorical data. Int. J. Database Theory Appl. **6**(14), 199–213 (2013)
153. M. Bouguessa, A practical outlier detection approach for mixed-attribute data. Exp. Sys. Appl. **42**, 8637–8649 (2015)
154. A. Reddy, S. Babu, B. Govardhan, Outlier analysis of categorical data using NAVF. Inf. Econ. **17**(1), 1–5 (2013)
155. Y. Liang, J. Chin, S. Dang, A new method for measuring uncertainty and fuzziness in rough set theory. Int J Gen Syst **31**, 331–342 (2002)
156. S. He, Z. Xu, X. Huang, Z. Deng, Fp-outlier: Frequent pattern based outlier detection. ComSIS **2**, 726–732 (2005)
157. S. Ghoting, A. Otey, M. Parthasarathy, Loaded: link-based outlier and anomaly detection in evolving data sets, IEEE, ICDM, pp. 387–390 (2004)
158. S. Otey, E. Ghoting, A. Parthasarathy, Fast distributed outlier detection in mixed-attribute data sets. Data Min. Knowl. Discov. **12**(2–3), 203–228 (2006)
159. L. Pang, G. Cao, L. Chen, Outlier detection in complex categorical data by modeling the feature value couplings, in *25th International Conference on Artificial Intelligence and Statistics,*,pp. 1902–1908 (2016)
160. J. Das, K. Schneider, Detecting anomalous records in categorical datasets, ACM, SIGKDD, 220–229 (2007)
161. H. Narita, K. Kitagawa, Detecting outliers in categorical record databases based on attribute associations, Progress in WWW Research and Development, pp. 111–123 (2008)
162. A. Rashidi, L. Hashemi, S. Hamzeh, Anomaly detection in categorical datasets using Bayesian networks, Part II, AICI, pp. 610–619 (2011)
163. B. Das, K. Schneider, J. Neill, Anomaly pattern detection in categorical datasets, ACM, SIGKDD, pp. 169–176 (2008)
164. E. Castillo, J.M. Gutiérrez, A.S. Hadi, *Expert Systems and Probabilistic Network Models* (Springer, Berlin, 1997)
165. K. Moore, A. Wong, Optimal reinsertion: a new search operator for accelerated and more accurate bayesian network structure learning, in *20th IEEE International Conference on Machine Learning and Applications - ICMLA 2021*, pp. 552–559 (2003)
166. J. Breunig, M. Kriegel, H. Ng, R. Sander, LOF: identifying density–based local outliers, ACM, SIGMOD, pp. 93–104 (2000)
167. A. Yu, X. Qian, W. Lu, H. Zhou, Finding centric local outliers in categorical/numerical spaces. Knowl. Inf. Syst. **9**, 309–338 (2006)
168. P. Chawla, S. Sun, Slom: a new measure for local spatial outliers. Knowl. Inf. Syst. **9**, 412–429 (2006)
169. R. Joshi, V. Bhatnagar, Cbof: Cohesiveness-based outlier factor a novel definition of outlier-ness, Mach. Learn. Data Min. Pattern Recognit., 175–189 (2014)

170. G. Suri, R. Murty, M. Athithan, A rough clustering algorithm for mining outliers in categorical data, *4th Int. Con. PReMI*, pp. 170–175 (2013)
171. G. Suri, R. Murty, M. Athithan, Detecting outliers in categorical data through rough clustering. Nat. Comput. **15**, 385–394 (2016)
172. G. Suri, R. Murty, M. Athithan, An algorithm for mining outliers in categorical data through ranking, in *12th Int. Conf. (HIS)*, IEEE, pp. 247–252 (2012)
173. G. Suri, R. Murty, N. Athithan, A ranking-based algorithm for detection of outliers in categorical data. Int. J. Hybrid Intell. Syst. **11**, 1–11 (2014)
174. Z. Huang, A fast clustering algorithm to cluster very large categorical data sets in data mining. DMKM, ACM, SIGKDD, pp. 1–8 (1997)
175. T. Knorr, E. Ng, A unified approach for mining outliers. CASCON, pp. 236–248 (1997)
176. V. Knorr, E. Ng, R. Tucakov, Distance-based outliers: algorithms and applications. VLDB J **8**, 237–253 (2000)
177. K. Ramaswamy, S. Rastogi, R. Shim, Efficient algorithms for mining outliers from large data sets, ACM, SIGMOD, pp. 427–438 (2000)
178. C. Angiulli, F. Basta, S. Pizzuti, Distance-based detection and prediction of outliers. IEEE Trans. Knowl. Data Eng. **18**(2), 145–160 (2006)
179. F. Angiulli, F. Fassetti, Fast outlier detection in high dimensional spaces, in *Euro. Conf. on the Prin. of Data Min and Kno Disc.*, pp. 19–26 (2002)
180. D. Ebdon, *Statistics in Geography: A Practical Approach-Revised with 17 Programs* (Wiley-Blackwell, Hoboken, 1991)
181. R. Bhaduri, K. Matthews, B. Giannella, Algorithms for speeding up distance-based outlier detection, ACM, SIGKDD, pp. 895–867 (2011)
182. S. Li, S. Lee, R. Lang, Mining distance-based outliers from categorical data, IEEE, ICDM, pp. 225–230 (2007)
183. E. Ghoting, A. Parthasarathy, S. Otey, Fast mining of distance-based outliers in high dimensional datasets. DMKD **16**(3), 349–364 (2008)
184. C. Böhm, C. Haegler, K. Müller, N. Plant, Coco: coding cost for parameter-free outlier detection, ACM, SIGKDD, pp. 149–158 (2009)
185. J. Smets, K. Vreeken, The odd one out: identifying and characterising anomalies, SIAM, SDM, pp. 804–815 (2011)
186. C. Akoglu, L. Tong, H. Vreeken, J. Faloutsos, Fast and reliable anomaly detection in categorical data, ACM, CIKM, pp. 415–424 (2012)
187. A. Taha, A. S. Hadi, Anomaly detection methods for categorical data: a review. ACM Comput. Surv. **52**(2), 1–35 (2019). https://doi.org/10.1145/3312739
188. G. Desrosiers, C. Karypis, A comprehensive survey of neighborhood-based recommendation methods, in *Recommender Systems Handbook Recommender Systems Handbook*, pp. 107–144 (2011)
189. V. Chandola, V. Banerjee, A. Kumar, Anomaly detection for discrete sequences: a survey. IEEE Trans. Knowl. Data Eng. **24**(5), 823–839 (2012)
190. K. Ge, Y. Xiong, H. Zhou, Z. Ozdemir, H. Yu, J. Lee, Top-eye: top-k evolving trajectory outlier detection, ACM, CIKM, pp. 1–4 (2010)