American-Eurasian Journal of Scientific Research 12 (3): 161-171, 2017
ISSN 1818-6785
© IDOSI Publications, 2017
DOI: 10.5829/idosi.aejsr.2017.161.171

A Comprehensive Survey on Outlier Detection Methods

R. Muthukrishnan and G. Poonkuzhali

Department of Statistics, Bharathiar University, Coimbatore-641046, Tamil Nadu, India

Abstract: In ancient days, outlier is viewed as noisy data in statistics, has turned out to be a vital problem which is being researched in different fields of application domains. Outlier detection is a primary step in many scientific research studies, because it has a negative impact on the results. Visual inspection alone cannot always identify an outlier and it can lead to mislabelling an observation as an outlier. Using a specific function of the observations leads to a superior outlier labelling rule. The estimation of parameters with classical measures such as mean is highly sensitive to outliers. Statistical methods were developed to accommodate outliers and to reduce their impact on the analysis. Numerous outlier detection methods have been developed specific to certain application domains, while few methods are more general. Outlier detection has been explored in a much broader area including discriminant analysis, experimental design, multivariate data, linear models etc. There are various approaches to outlier detection depending on the application and number of cases/variables in the data set. An attempt has been made to review the outlier detection methods which are entrenched and commonly used now-a-days. This paper provided a survey on the structure of existing outlier detection methodologies.

Key words: Outliers ・ Nearest neighbour ・ Density ・ Cluster ・ Robust Distance ・ Depth

INTRODUCTION

The outlier problem is as old as statistics. One important task of statistics is the identification and proper handlings of outliers which are often thought to be extreme values which are caused by measurement or transmission errors. Various outlier detection methods have been developed specific to certain application areas, while few methods are more generic.

There are various approaches to outlier detection depending on the application and number of observations in the data set. The detection of outliers can be a very hard problem. Whereas in one dimension, observations that are far away from the main data cloud can easily be detected, this is not necessarily the case in higher dimension, when the outliers are not extreme along the coordinates but in any other direction with increasing dimensionality, multivariate outliers become harder to detect, yet they can heavily influence the statistical results.

The existence of the problem of anomalous values has been recognized and discarded since eighteenth century: [1-4].

Observations that appear apart from the bulk of the data called outliers, noise, maverick, anomaly, glitch, surprising value, exotic, dirty, abnormality, deviants, discordant depends on the opinion of the investigator. The authors have given many definitions for outliers [5-16].

The significance studies of outlier analysis are,

- The investigator desired to check whether the data is normally distributed or any contaminated data are present in the data. Since the noisy data shift the location and scale estimator.
- Outliers can seriously bias or influence estimates that may be of substantive interest.
- The outliers themselves for the purpose of obtaining certain critical information.
- The purpose of examining the outlying observation is the basic probability model may be contaminating due to presence of outliers.
- Outliers which influence assumptions of a statistical test, for example, outliers violating the normal distribution assumption in an ANOVA test and deal

Corresponding Author: R. Muthukrishnan1, Department of Statistics, Bharathiar University,
Coimbatore-641046, Tamil Nadu, India.

with them properly in order to improve statistical analysis. This could be considered as a preliminary step for data analysis.

In some point of view, outlier is an unavoidable one. As of some statisticians one should not consider an outlying observation unless there is an emphasis in the presence of outlier in the data. Consequently the fact is that the people who have to deal with data are forced to make judgement, whether or not to include the outliers or replace the outliers with some other points without affecting the interpretation. [17] said that, there was no way of drawing a dividing line between those that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others. [18] wrote that they had never rejected an observation merely because of its large residual and that all completed observation with equal weight ought to be allowed to contribute to the result. On the other hand, rejection was never predicted now-a-days as being carried out according to any formal procedure, but was merely a matter of the investigators judgement.

The literature on outliers is enormous and its prominence to many other areas. The main classical books relevant to outlier analysis are [6, 19-22]. A number of survey and review articles available in the literature. [7] reviewed work on outliers in circular data, experimental data, discriminant analysis, Bayesian method, time series etc. [23] provided a review in the area of anomaly detection based on statistical approaches. [24] provided an extensive survey of techniques including statistical model, neural networks and machine learning.

[25] reviewed outlier detection techniques for numeric and symbolic data. [26] provided a comprehensive survey of outlier detection systems and hybrid intrusion detection systems. [27] provided overview for the problem of detecting anomalies in discrete sequences. [28] reviewed of multivariate outlier detection methods especially robust distance based methods. A survey based on outlier detection for temporal data has been studied in [29].

The broad range of outlier detection methods are summarised and categorised in the remainder of this paper. In the next five sections, we categorised the outlier detection techniques in the context of (i) nearest neighbour (ii) density (iii) cluster (iv) statistical approach (v) robust distance and (vi) depth based outlier detection techniques. The overall discussion is presented in the last section.

Nearest Neighbour Based Outlier Detection Techniques: Nearest neighbor based anomaly detection techniques require a distance or similarity measure between two data points.

[30] calculated the k-nearest for the new exemplar and it is classified according to the majority classification of the nearest neighbour. Author studied two specific based algorithms, one is the nearest neighbour algorithm and another one is nearest-hyperrectangle algorithm. k nearest neighbour algorithm (kNN) outperform the first-nearest neighbour is found to give predictions that are substantially inferior to those given by kNN in a variety of domains. The advantage of kNN is an extremely powerful and flexible inductive learning algorithm and also easily trained. If the kNN procedure achieves close to the desired results, then it may be worthwhile investing time to train and test other, more complicated classifiers.

[31] proposed the distance based outliers. The author used nested loop algorithm as well as an additional procedure is also developed by dividing the space into a homogenous grid of cells and then using these cells to calculate outliers. A point x in a data set is an outlier with respect to parameter k and d, if no more than k points in the data set are at distance of d. This algorithm has many advantages but even has certain disadvantages: (i) A distance could be complicated to determine since the author required many iterations. (ii) Ranking procedure does not employ to determine outliers. (iii) The cell based algorithm does not scale for higher dimensions.

[32] considered the problem of detecting type in spatial point processes in the presence of substantial clutter. Authors used kth nearest neighbour distance of points in the process to classify them as clutter. The observed kth nearest neighbour distance is modelled as a mixture distribution. This method allows for detection of generally shaped features that need not be path connected. This method works well in high dimension and also used to produce very high breakdown point robust estimators of a covariance matrix. The advantage of this method is that it can be applied without user input about the shapes of the regions. The time it requires and its complexity are significantly less. It outperforms existing methods when the proportion of outliers is very high.

[33] proposed a distance based outliers in the novel formulation method. It is based on the distance of a data point from its kth nearest neighbour. Rank each data points based on the distance. The top n ranking is considered as outliers. In addition, relatively straight forward solutions were developed based on the classical

nested loop join and index join algorithm to detect outliers. This algorithm performs well with respect to both size of data and data dimension.

[34-36] compute the anomaly score of a data instance as the sum of its distances from its k nearest neighbours. [37] show that for adequately randomized data, a simple trimming step could result in the average complexity of the nearest neighbour search to be nearly linear. [38] employed sampling to improve the efficiency of the nearest neighbor based technique. The authors compute the nearest neighbor of every instance within a smaller sample from the data set.

Density Based Outlier Detection Techniques: Density based outlier detection techniques estimate the density of the neighborhood of each data instance. An illustration that lies in a neighborhood with low density is declared to be outliers while an illustration that lies in a dense neighborhood is declared to be normal. Density based techniques perform poorly if the data has regions of varying densities.

Local Outlier Factor (LOF) is an essential concept for local density of the nearest neighbor. [39, 40] assign an outlier score to a given data, known as *Local Outlier Factor* (LOF). For any given data, the LOF score is equal to ratio of average local density of the k nearest neighbors of the instance and the local density of the data instance itself. To find the local density for the data, the author used to find the radius for the smallest hyper sphere centered at the data instance that contains its k nearest neighbours. The local density is then computed by dividing k by the volume of this hyper sphere. Here the anomaly data will get higher LOF score.

[41] discussed Connectivity-based Outlier Factor (COF). Also author studied variation between LOF and COF. In COF, the neighbourhood for an instance is computed in an incremental mode. The author used the procedure to begin; the closest instance to the given instance is added to the neighborhood set. The next instance added to the neighborhood set is such that its distance to the existing neighbourhood set is minimum among all remaining data instances. The distance between an instance and a set of instances is defined as the minimum distance between the given instance and any instance belonging to the given set. The procedure is repeated until it reaches size k. Once the neighborhood is computed, the anomaly score (COF) is computed in the same manner as LOF.

[42] proposed a measure called Multi-granularity deviation Factor (MDEF). The inverse of the standard deviation is the anomaly score for the data. [43] proposed Outlier Detection using In-degree Number (ODIN). It is the simpler version of LOF. ODIN is equal to the number of k nearest neighbors of the data which have the given data in their k nearest neighbor list. [44] proposed a technique called Probabilistic Suffix Trees (PST) to find the nearest neighbours for a given sequence.

Cluster Based Outlier Detection Techniques: Cluster is a collection of data objects similar to one another within the same cluster and dissimilar to the objects in other clusters. Normal data belong to a cluster in the data, while outliers either do not belong to any cluster. Clustering is not a new concept but data clustering together with outlier detection is a recent scientific discipline under fast development. Anomaly detection techniques can operate is any one of the three modes: Supervised, Semi supervised and unsupervised anomaly detection.

[45] proposed WaveCluster algorithm. It is a novel clustering approach based on wavelet transforms. Multi-resolution property of wavelet transform is used. It can effectively identify arbitrary shape clusters at different degrees of accuracy. Also WaveCluster is highly efficient in terms of time complexity. It especially attractive for very large data bases, data mining, efficient information, knowledge recovery, information recovery. Also it is insensitive to the order of input data to be used. [46], [47] emphasized information algorithm for clustering data and interpreting results. These books provided clear picture of cluster validity from the orientation of application. [48] shown that Partitioning Around Medoids algorithm (PAM) provide better class separation than the means produced by the k-means clustering algorithms.

[49] explained two clustering methods and applied these methods to network data. First method is most widely used clustering method called k-means method. It requires number of clusters in the data set. After decided number of clusters author made some algorithm using cluster centers. This algorithm is easy to execute and works quite well in most situations. Secondly, the author utilized Approximate Distance Clustering (ADC) method given by [50]. The procedure is based on the subset of the data. For each data point, determine the distance to each element of the subset and retain the smallest distance. Main strength of this approach is that it does not require a network security to executive it nor does it need perfectly clear data.

[5] noted that outliers may be considered as noise points lying outside a set of defined clusters. [51] integrated the knowledge of labels to improve on their unsupervised clustering based anomaly detection technique by calculating a measure called semantic

anomaly factor. [52] introduced the FindOut algorithm is an extension of the WaveCluster algorithm. The main procedure in FindOut algorithm is to remove the clusters from the original data and then identify the outliers. Author used hash table structure to represent the data set. FindOut creates to efficiently perform wavelet transform on high dimensional data sets. This hash based technique can be used for any grid based data processing approach. Authors felt that previous research showed that such techniques may not be effectual because of the nature of the clustering. FindOut can successfully identify outliers in large data set. This method is highly cost effective and efficiency.

[53] modified PAM proposed by [54]. PAM does not recognize relatively small clusters in situations where good partitions around medoids obviously exist. To overcome this problem [63] proposed to partition around medoids by maximizing a criteria Average Silhouette defined by [54]. Also authors proposed a fast-to-compute approximation of Average Silhouette. [55] proposed a technique FindCBLOF, which assigns an anomaly score called as cluster based Local Outlier Factor (CBLOF) for each data. To find CBLOF, Squeezer algorithm is used which can produce good clustering results and at the same time it deserves good scalability.

[56] used clustering and Mahalanobis distance to detect outliers. The basic idea of this method is first: using a partitioning clustering method, split the n points cloud in k smaller sub clouds. Secondly, apply Mahalanobis distance to observations and to all clusters. In each cluster if it is an outlier then the observation is considered as an outlier. Remove the observation detected as an outlier and repeat the process until no outliers are found. The final decision is whether all the observations belonging to a given cluster are outliers is based on a table of cluster of Mahalanobis distance. [57] proposed a clustering based approach to detect outliers. Author has used the k-means clustering algorithm. [58] provided outlier detection using distance distribution clustering (ODDC). [59] provided a survey of partition based clustering algorithms in the context of data mining.

Statistical Approach Based Outlier Detection Techniques: Statistical approaches were the earliest algorithm used for outlier detection. Some of the earliest are applicable only for single dimensional data sets [11, 12, 60-65]. Statistical models are generally appropriated to quantitative real valued data sets or at the very least quantitative ordinal data distributions where the ordinal data can be transformed to suitable numerical values for statistical processing. [66] used the simplest statistical outlier detection techniques viz informal box plots to pinpoint outliers in both univariate and multivariate case.

Histogram based techniques for the outlier by [67-70] introduced a non parametric approach for outlier detection in machinery operation. [71, 72] used regression based anomaly detection has been extensively investigated for time series data. [73, 74], proposed certain technique to detect the presence of outliers in a data set by the Akaike Information Criterion (AIC) during model fitting. For multivariate data, a basic technique is to construct attribute-wise histograms. During testing, for each test instance, the outlier score for each attribute value of the test instance is calculated as the height of the bin that contains the attribute value. The per-attribute outlier scores are aggregated to obtain an overall outlier score for the test instance.

[75] described a simple statistical technique for novelty detection can be based on determining whether the test samples come from the same distribution. Authors used t-test to find damaged beams. [76, 77] provided the students t-test applied for outlier detection in structured beams. [78, 79] proposed Extreme Value Theory (EVT) for outlier detection that concerns abnormally low or high values in the tails of the data distribution. [80] used a technique Packet Header Anomaly Detection (PHAD) and Application Layer Anomaly Detection (ALAD) applied to network intrusion detection.

[81] described the additional complexity multivariate time series data over the univariate series using multivariate ARIMA model. [82] used a Chi Square statistic to determine outliers in operating system called data. [83-85] provided robust outlier detection approach has been applied in Autoregressive Integrated Moving Average (ARIMA) model.

[86] described outlier detection method based on Gaussian mixture models (GMM) for sensor fault detection. [87] developed a class of models for probability distribution of images called hierarchical image probability (HIP) models. [88] proposed Hidden Markov Models (HMM) are stochastic models for sequential data. Gaussian mixture models have been frequently used models as a mixture of parametric distribution. [32, 89] used a mixture of poisson distribution to model the normal data and then detect outliers.

Robust Distance Based Outlier Detection Techniques: Outlier observations can be determined by using various distance based methods. One can find outliers by distance for each observation using Location and Scale Estimator. The following are the methods survived for detection of anomaly observation using robust distance.

M-estimators were originally proposed by [90]. The M-estimators are robust generalizations of the classical Maximum Likelihood Estimator (MLE) and are obtained iteratively. Weight functions are assigned to estimate location and scatter. Several authors have used robustified Mahalanobis distance based upon the robust estimators of location and scale to identify multiple outliers. Chi-square value is used to obtain the critical value for the distribution of the Mahalnobis distance. Then compared the Mahalanobis distance value with the Chi-square value to detect outliers. The drawback of this method is having low breakdown points in high dimensions.

[91] introduced the robust minimum volume ellipsoid (MVE) method for detection of outliers in multidimensional data. Subsets of approximately 50 per cent of the observations are examined. The best subset is then used to calculate the covariance matrix. An appropriate cut-off value is then estimated and the observations with distances that exceed that cut-off are declared to be outliers.

The minimum covariance determinant (MCD) estimator was proposed by [91]. MCD is a robust estimator to estimate the location and shape of the clusters. Points that are outliers with respect to a particular cluster will not be involved in the location and shape calculations of that cluster and points that are outliers with respect to all clusters will not be involved in the calculations of any cluster. The difference between the single population case and the multiple cluster case is that, in the latter, MCD samples need to be computed for each cluster.

The SDE procedure was developed independently by [92, 93] and is mentioned in [94]. Simplistically, the idea is that an outlier or high leverage point will separate out and away from the bulk of the data when viewed from the right perspective. There are two stages to the information of the robust multivariate location and dispersion estimators. First, robust distances are determined via a projection computation. These distances become the arguments in a weight function that is used to calculate a weighted mean vector and weighted covariance matrix. While definition of the Stahel-Donoho estimator requires the supremum over all possible directional vectors, [95] proposed a shortcut method which uses just n directional vectors, one vector in the direction of each centered observation. The projections of the original data on these n directional vectors produce the robust distances.

[96] proposed the Feasible Solution Algorithm (FSA) for obtaining approximations to Rousseeuw's MCD estimator. Also the author described that the MCD estimate resulting from the FSA can be used to detect outliers using the usual robust distance scheme. The FSA begins by first assuming that there are at most k outliers in the data. A random sample of (n-k) observations is then selected from the original sample of n observations, with the remaining k observations trimmed from the data. The randomly selected observations are used to form an initial mean vector and covariance estimate along with the respective covariance determinant.

[97] proposed BACON (Blocked Adaptive Computationally Efficient Outlier Nominator) method. The desire is to find an outlier detection method that is applicable to very large datasets. The first observation is that the added computational complexity of trying to find optimal robust estimators may not be justified by significantly better outlier detection. The second observation is that insisting upon a completely affine equivariant method may add substantial computational complexity to an algorithm without a proportional improvement in the detection of outliers. Using these two observations, the authors developed BACON as a method that abandons optimality conditions in favour of a very fast outlier detection strategy that can be run in a non-robust, affine equivariant mode with breakdown point of 20 per cent, or in a robust, near-affine equivariant mode with a breakdown point of 40 per cent.

[2] proposed an Orthogonalized Gnanadesikan-Kettenring (OGK) estimator by a general method to obtain positive-definite and approximately affine-equivariant robust scatter matrices starting from any pair-wise robust scatter matrix. This method was applied to the robust covariance estimate of [28, 98] proposed an alternative method to detect outliers based on the comedian (Median Absolute Deviation turns to be comedian) which is introduced by [99]. They used to compute eigen values and eigen vectors to find the location and scatter and then calculated Mahalanobis distance to detect outliers using some cut-off value.

Depth Based Outlier Detection Techniques: Data depth is an important concept to Multivariate data analysis. Using the different notion of data depth, one can compute depth values for all sample points in the data cloud. Order the depth values based on center outward ranking. It means that the data points with the highest depth called the deepest or central point or it simply called center. The data points with lowest depth values are called outliers. Based on this ordering of depth one can compute Multivariate location, scale, skewness and kurtosis and Graphical methods such as Contour plot, Bag plot, Sunburst plot, Perspective plot, DD plot, Blotched bag plots for analyzing the distributional characteristics of the Multivariate data cloud and detect outliers.

A survey of depth function found in [100-104] introduced a notion of depth in the regression setting. The various notions of depth based techniques are proposed: Mahalanobis depth proposed by [105]. This depth depends on the location and covariance matrix. It satisfies the properties alline invariant, maximality at center, vanishing at infinity. Halfspace depth was proposed by [106]. It reflects a generalization of the notion of ranks to multivariate data. For univariate case, given some number x, all values less than or equal to x is a closed halfspace and all points less than x is an open halfspace. Similarly, all points greater than or equal to x form a closed halfspace and all points greater than x is an open halfspace.

The convex hull peeling depth was proposed by [107]. It has been constructed by drawing the minimum convex set which enclosed all sample points. Those points on the boundary are considered as group1 and discarded. The convex hull of the remaining is formed; those on the boundary are taken as group 2. The process is repeated, providing an entirely sample-based method dividing the data into order groups. Oja depth was introduced by [108]. It is based on average volumes of simplices not based on distances. It satisfies the properties affine invariant, maximality at center, vanishing at infinity. The Simplicial depth was introduced by [109]. The depth emerges naturally out of a fundamental concept underlying affine geometry namely that of a simplex and it satisfies the requirements one would expect from a notion of data depth.

The Likelihood depth was proposed by [110]. The main idea is to order the observations corresponding to their likelihood. If the density happens to be ellipsoidal, then the ordering is similar to the rankings that are derived from a density estimate using a fixed bandwidth. Projection depth was introduced by [111]. It induced estimators are very favourable because they can enjoy very high breakdown point robustness without having to pay the price of low efficiency, meanwhile providing a promising center-outward ordering of multidimensional data. The computation of projection depth seems intractable since it involves supremum over infinitely many direction vectors. Rayleigh Projection depth was introduced by [112]. The traditional projection depth has many good properties but it is indeed not practical due to its difficult computation especially for the high dimensional data sets. Defined on the mean and variance of the data sets, the new depth Rayleigh Projection Depth can be computed directly by solving a problem of generalized eigen value.

DISCUSSION

In this article an attempt is made to bring the gist of the outlier detection methodologies. There are many universally applicable generic outlier detection methodologies available in the literature. This comprehensive survey broadly categorised based on statistical, density, cluster, depth and robust nature. All these are varying based on the features like data type, size of data, nature of data etc. The brief summary of the various outlier detection techniques discussed in the previous sections under different categories is given below:

The main advantage of nearest neighbor based outlier detection technique is that, it do not based on any assumed distribution to fit the data. This technique is fairly straightforward and easy to detect the outliers. The main drawback is the most of the nearest neighbor based methods are not effective in high dimensional data set. Since each data points are equally distance with each other as the number of dimensions increases, as a result, deviation of each data point cannot be observed. Hence it is difficult to view outliers in high dimensional data sets. These methods are taken computationally more time.

The density based outlier detection technique is more efficient as compared to nearest neighbor based outlier detection technique. On the other hand, in order to improve the efficiency, the density based methods are more complexity and computationally very expensive. This method does not only look at its local density but also explore its neighbours.

Clustering based outlier detection technique is quite intuitive and consistent. The computational time requires depends on the clustering algorithm used to produce clusters from the data. Among the other clustering based techniques, the test phase method is fast, since this method compares a test case with a tiny number of clusters. This technique can frequently be modified to complex data types by simply plugging in a clustering procedure for a particular data type. The disadvantage of clustering outlier detection technique is the performance is highly dependent on the effectiveness of clustering procedure in capturing the cluster structure of normal cases. Many techniques detect outliers as a consequence of clustering and hence are not effective for outlier detection. Several clustering algorithms force every occurrence to be assigned to some cluster. This might effect in outliers getting allot to a large cluster, thereby being considered as normal cases by techniques that operate under the statement that outliers do not belong to any cluster. Several clustering based techniques are effective only when the outliers do not form significant clusters among themselves.

Statistical approach based outlier detection methods have some advantages and disadvantages: if a probabilistic model is known, the methods are very efficient and it is possible to found outliers. The computational time of statistical approach based outlier detection technique depends on the nature of statistical model that is essential to be fitted on the data. If the fundamental data distribution hold true, statistical techniques provide a statistically justifiable solution for outlier detection. Moreover, the model is presented in a dense form, it possible to identify outliers without storing the original data. This category is usually not applied in a multi-dimensional data because most of the distribution models apply to the univariate data. Hence these methods is greatly restrictions their applicability in most practical application in high dimensional data. Therefore, statistical methods are limited to real data set with large size to certain extend and it is difficult to apply in problem of multivariate distribution.

Non-robust Mahalanobis distance based outlier detection methods can be affected in the presence of a outliers, however, Robust Mahalanobis distance based methods gives reliable results and detects maximum number of outlier points. Robust distance based techniques are taken computationally more time since it needs more iterations when compared with other procedures.

Depth based outlier detection techniques are emerging now-a-days in various fields of researches. The different notions of depth procedures are well established to detect outliers. This method is entirely non parametric and also moment free. Some of the procedures are based on projection pursuit and hence computational time is very less and more efficient in high dimensional data set.

REFERENCES

1. Glaisher, J., 1872. On the rejection of discordant observations. Monthly Notices Roy. Astr. Soc., 33: 391-402.
2. Maronna, R. and R. Zamar, 2002. Robust estimates of location and dispersion for highdimensional datasets. Technometrics, 44: 307-317.
3. Stone, E., 1868. On the rejection of discordant observations. Monthly Notices of the Royal Astronomical Society, 34: 9-14.
4. Wright, T., 1884. Treatise on the Adjustment Observations by the Method of Least Squares, New York.
5. Aggarwal, C.C. and P.S. Yu, 2001. Outlier detection for high dimensional data, Proceedings of ACM SIGMOD International Conference on Management of Data, pp: 37-46.
6. Barnett, V. and T. Lewis, 1994. Outliers in statistical data. John Wiley and Sons, New York.
7. Beckman, R. and R. Cook, 1983. Outlier..........s. Technometrics, 25: 119-149.
8. Collett, D. and T. Lewis, 1976. The subjective nature of outlier rejection procedures. Applied statistics, 25: 228-237.
9. Edgeworth, F., 1887. On discordant observations. Philosophical Magazine Series, 5(23): 364-375.
10. Ferguson, T., 1961. On the rejection of outliers. Proceedings of the Fourth Berkeley symposium on Mathematical Statistics and Probability, 1: 253-28.
11. Grubbs, F., 1950. Sample criteria for testing outlying observations. The Annals of Mathematical Statistics, 21: 27-58.
12. Grubbs, F.E., 1969. Procedures for detecting outlying observations in samples. Technometrics, 11: 1-21.
13. Gumbel, E.J., 1960. Discussion on rejection of outliers by anscombe, F.J. Technometrics, 2: 165-166.
14. Guttman, I. and D.E. Smith, 1969. Investigation of rules for dealing with outliers in small samples from the normal distribution: I: Estimation of the mean. Technometrics, 11: 527-550.
15. Peirce, B., 1852. Criterion for the rejection of doubtful observations. The Astronomical Journal, 2: 161-163.

16. Weisberg, S., 1985. Applied Linear Regression. John Wiley and Sons.
17. Bernoulli, D., 1777. Dijudicatio maxime probabilis plurium observationum discrepantium atque verisimillima induction inde formanda. Acta Academiae Scientiorum Petropolitanae, 1: 3-33.
18. Bessel, F. and J. Baeuer, 1838. Gradmessung in Ostpreussen und ihre Verbindung mit Preussischen und Russischen Dreiecksketten. Berlin.
19. Aggarwal, C., 2013. Outlier Analysis. Springer.
20. Cook, R.D. and S. Weisberg, 1982. Residuals and influence in regression, Chapman and Hall, New York.
21. Hawkins, D., 1980. Identification of Outliers. Springer, Netherlands.
22. Rousseeuw, P. and A. Leroy, 2003. Robust Regression and Outlier Detection. Wiley.
23. Markou, M. and S. Singh, 2003a. Novelty detection: a review part 1: statistical approaches. Signal Processing, 83: 2481-2497.
24. Hodge, V. and J. Austin, 2004. A survey of outlier detection methodologies. Artificial Intelligence Review, 22: 85-126.
25. Agyemang, M., K. Barker and R. Alhajj, 2006. A comprehensive survey of numeric and symbolic outlier mining techniques. Intelligent Data Analysis, 10: 521-538.
26. Patcha, A. and J.M. Park, 2007. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51: 3448-3470.
27. Chandola, V., A. Banerjee and V. Kumar, 2012. Anomaly detection for discrete sequences: A survey. IEEE Transactions on Knowledge and Data Engineering, 24: 823-839.
28. Sajesh, T. and M. Srinivasan, 2012. Outlier detection for high dimensional data using the comedian approach. Journal of Statistical Computation and Simulation, 82: 745-757.
29. Gupta, M., J. Geo, C. Aggarwal and J. Han, 2014. Outlier detection for temporal data: A survey. IEEE Transactions on Knowledge and Data Engineering, 25: 1.
30. Wettschereck, D., 1994. A study of distance-based machine learning algorithms. PhD thesis, Oregon State University, Corvallis.
31. Knorr, E. and R. Ng, 1998. Algorithms for mining distance based outliers in large datasets. In: Proceedings of the International Conference on Very Large Data Bases, pp: 392-403.
32. Byers, S.D. and A.E. Raftery, 1998. Nearest-neighbor clutter removal for estimating features in spatial point processes. Journal of the American Statistical Association, 93: 577-584.
33. Ramaswamy, S., R. Rastogi and K. Shim, 2000. Efficient algorithms for mining outliers from large data sets. ACM SIGMOD Record, 29: 427-438.
34. Angiulli, F. and C. Pizzuti, 2002. Fast outlier detection in high dimensional spaces. In: In Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery, Springer-Verlag, pp: 15-26.
35. Eskin, E., A. Arnold, M. Prerau, L. Portnoy and S. Stolfo, 2002. A geometric framework for unsupervised anomaly detection. In: Applications of data mining in computer security, Springer, pp: 77-101.
36. Zhang, J. and H. Wang, 2006. Detecting outlying subspaces for high-dimensional data:the new task, algorithms and performance. Knowledge and Information Systems, 10: 333-355.
37. Bay, S.D. and M. Schwabacher, 2003. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM press, pp: 29-38.
38. Wu, M. and C. Jermaine, 2006. Outlier detection by sampling with accuracy guarantees. In: In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, pp: 767-772.
39. Breunig, M.M., H.P. Kriegel, R.T. Ng and J. Sander, 1999. Optics-of: Identifying local outliers. In: In Proceedings of the Third European Conference on Principles of Data Mining and Knowledge Discovery, Springer-Verlag, pp: 262-270.
40. Breunig, M.M., H.P. Kriegel, R.T. Ng and J. Sander, 2000. Lof: identifying density-based local outliers. In: In Proceedings of 2000 ACM SIGMOD International Conference on Management of Data, pp: 93-104.
41. Tang, J., Z. Chen, A.W. chee Fu and D.W. Cheung, 2002. Enhancing effectiveness of outlier detections for low density patterns. In: Advances in Knowledge Discovery and Data Mining, Springer, pp: 535-548.
42. Papadimitriou, S., H. Kitagawa, P.B. Gibbons and C. Faloutsos, 2002 Loci: fast outlier detection using the local correlation integral. In: Tech. Rep, Pittsburgh.

43. Hautamaki, V., I. Karkkainen and P. Franti, 2004. Outlier detection using k-nearest neighbour graph. In: Proceedings of 17th International Conference on Pattern Recognition, EEE Computer Society, 3: 430-433.

44. Sun, P., S. Chawla and B. Arunasalam, 2006. Mining for outliers in sequential databases. In: SIAM International Conference on Data Mining, SIAM, pp: 94-105.

45. Sheikholeslami, G., S. Chatterjee and A. Zhang, 1998. Wavecluster: A multi- resolution clustering approach for very large spatial databases. In: Proceedings of the 24rd International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc, 98: 428-439.

46. Jain, A.K. and R. Dubes, 1988. Algorithms for clustering data. Prentice Hall College Div.

47. Tan, P.N., M. Steinbach and V. Kumar, 2005. Introduction to Data Mining. Addison-Wesley.

48. Bradley, P., U. Fayyad and U. Mangasarian, 1999. Mathematical programming for data mining: formulations and challenges. INFORMS Journal on Computing, 11: 217-238.

49. Marchette, D., 1999. A statistical method for profiling network traffic. In: Proceedings of 1st USENIX Workshop on Intrusion Detection and Network Monitoring, pp: 119-128.

50. Cowen, L. and C. Priebe, 1997. Approximate distance clustering. Applied Statistics, 29: 337-346.

51. He, Z., S. Deng and X. Xu, 2002. Outlier detection integrating semantic knowledge. In: Proceedings of the Third International Conference on Advances in Web-Age Information Management, Springer-Verlag, UK, pp: 126-131.

52. Yu, D., G. Sheikholeslami and A. Zhang, 2002. Findout : Finding outliers in very large datasets. Knowledge and Information Systems, 4: 387-412.

53. Laan, M., K. Pollard and J. Bryan, 2003. A new partitioning around medoids algorithms. Journal of Statistical Computation and Simulation, 73: 575-584.

54. Kaufman, L. and P. Rousseeuw, 1990. Finding groups in data: an introduction to cluster analysis. John Wiley and Sons, New York.

55. He, Z., X. Xu and S. Deng, 2003. Discovering cluster-based local outliers. Pattern Recognition Letters, 24: 1641-1650.

56. Pires, A. and C. Santos-Pereira, 2005. Using clustering and robust estimators to detect outliers in multivariate data. In: Proceedings of the International Conference on Robust Statistics, Finland.

57. Yoon, K., O. Kwon and D. Bae, 2007. An approach to outlier detection of software measurement data using the k-means clustering method. In: First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007), pp: 443-445.

58. Niu, K., C. Huang, S. Zhang, J. Chen and T. Washio, 2007. ODDC: Outlier Detection Using Distance Distribution Clustering, (PAKDD 2007) Workshops, Lecture Notes in Artificial Intelligence (LNAI), Springer-Verlag, pp: 332-343.

59. Velmurugan, T. and T. Santhanam, 2011. A survey of partition based clustering algorithms in data mining: An experimental approach. Information Technology Journal, 10: 478-484.

60. Dixon, W., 1950. Analysis of extreme values. The Annals of Mathematical Statistics, 21: 488-506.

61. Dixon, W., 1951. Ratios involving extreme values. The Annals of Mathematical Statistics, 22: 68-78.

62. Dixon, W., 1953. Processing data for outliers. Biometrics, 9: 74-89.

63. Iglewicz, B. and D. Hoaglin, 1993. How to detect and handle outliers. ASQC Quality Press.

64. Rosner, B., 1983. Percentage points for a generalized ESD many-outlier procedure. Technometrics, 25: 165-172.

65. Tietjen, G.L. and R.H. Moore, 1972. Some grubbs-type statistics for the detection of several outliers. Technometrics, 14: 583-597.

66. Laurikkala, J., M. Juhola1 and E. Kentala, 2000. Informal identification of outliers in medical data. In: Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology, pp: 20-24.

67. Dasgupta, D. and F. Nino, 2000. A comparison of negative and positive selection algorithms in novel pattern detection. In: In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 1: 125-130.

68. Helman, P. and J. Bhangoo, 1997. A statistically based system for prioritizing information exploration under uncertainty. IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans, 27: 0-466.

69. Javitz, H.S. and A. Valdes, 1991. The sri ides statistical anomaly detector. In: Proceedings of the 1991 IEEE Symposium on Research in Security and Privacy, IEEE Computer Society, pp: 316-326.

70. Dasgupta, D. and S. Forrest, 1996. Novelty detection in time series data using ideas from immunology. In: Proceedings of the international conference on intelligent systems, pp. 82-87.

71. Abraham, B. and G.E.P. Box, (1979). Bayesian analysis of some outlier problems in time series. Biometrika, 66: 229-236.
72. Abraham, B. and A. Chuang, 1989. Outlier detection and time series modeling. Technometrics, 31: 241-248.
73. Kadota, K., D. Tominaga, Y. Akiyama and K. Takahashi, 2003. Detecting outlying samples in microarray data: A critical assessment of the effect of outliers on sample classification. Chem-Bio Informatics, 1: 30-45.
74. Kitagawa, G., 1979. On the use of aic for the detection of outliers. Technometrics, 21: 193-199.
75. Ruotolo, R. and C. Surace, 1997. A statistical approach to damage detection through vibration monitoring. Applied mechanics in the Americas, pp: 314-317.
76. Surace, C. and K. Worden, 1998. Novelty detection method to diagnose damage in structures: an application to an offshore platform. In: Proceedings of Eighth International Conference Off-shore and Polar Engineering, pp: 64-70.
77. Surace, C., K. Worden and G. Tomlinson, 1997. A novelty detection approach to diagnose damage in a cracked beam. In: Proceedings of SPIE, pp: 947-953.
78. Roberts, S., 1999. Novelty detection using extreme value statistics. IEE Proceedings - Vision Image and Signal Processing, 146: 124-129.
79. Roberts, S., 2002. Extreme value statistics for novelty detection in biomedical signal processing. In: Proceedings of the 1st International Conference on Advances in Medical Signal and Information Processing, pp: 166-172.
80. Mahoney, M.V. and P.K. Chan, 2002. Learning nonstationary models of normal network traffic for detecting novel attacks. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge Discovery and Data Mining, pp: 376-385.
81. Tsay, R.S., D. Pea and A.E. Pankratz, 2000. Outliers in multivariate time series. Biometrika, 87: 789-804.
82. Ye, N. and Q. Chen, 2001. An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. Quality and Reliability Engineering International, 17: 105-112.
83. Bianco, A.M., M.G. Ben, E.J. Martinez and V.J. Yohai, 2001. Outlier detection in regression models with arima errors using robust estimates. Journal of Forecasting, 20: 565-579.
84. Chen, D., X. Shao, B. Hu and Q. Su, 2005. Simultaneous wavelength selection and outlier detection in multivariate regression of near-infrared spectra. Analytical Sciences, 21: 161-166.
85. Galeano, P., D. Pea and R.S. Tsay, 2006. Outlier detection in multivariate time series by projection pursuit. Journal of the American Statistical Association, 101: 654-669.
86. Hickinbotham, S. and J. Austin, 2000. Neural networks for novelty detection in airframe strain data. In: Proceedings of IEEE IJCNN, Italy.
87. Spence, C., L. Parra and P. Sajda, 2001. Detection, synthesis and compression in Mammographic image analysis with a hierarchical image probability model. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, pp: 310.
88. Duda, R., P. Hart and D. Stork, 2001. Pattern classification. Wiley, New York.
89. Agarwal, D., 2007. Detecting anomalies in cross-classified streams: a bayesian approach. Knowledge and Information Systems, 11: 29-44.
90. Jain, A.K. and R. Dubes, 1988. Algorithms for clustering data. Prentice Hall College Div.
91. Rousseeuw, P.J., 1984. Least median of squares regression. Journal of the American Statistical Association, 79: 871-880.
92. Donoho, D., 1982. Breakdown properties of multivariate location estimators. In: Ph.D. Qualifying Paper, Harvard University, Cambridge, MA.
93. Stahel, W., 1981. Robust schatzungen: Infinitesimale optimalitat und schatzungen von kovarianzmatrizen. In: Ph.D. thesis, Switzerland.
94. Rousseeuw, P. and A. Leroy, 1987. Robust Regression and Outlier Detection. JohnWiley and Sons, England.
95. Rousseeuw, P. and B. van Zomeren, 1990. Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85: 633-639.
96. Hawkins, D., 1994. The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data. Computational Statistics and Data Analysis, 17: 197-210.
97. Billor, N., A.S. Hadi and P.F. Velleman, 2000. Bacon: Blocked adaptive computationally efficient outlier nominators. Computational Statistics and Data Analysis, 34: 279-298.
98. Gnanadesikan, R. and J. Kettenring, 1972. Robust estimates, residuals and outlier detection with multi-response data. Biometrics, 28: 81-124.

99. Falk, M., 1997. On mad and comedian. Annals of the Institute of Statistical Mathematics, 49: 615-644.
100. Cascos, I., 2009. Data depth: multivariate statistics and geometry. New Perspectives in Stochastic Geometry Clarendon Press, Oxford University Press, Oxford.
101. Liu, R., J. Parelius and K. Singh, 1999. Multivariate analysis by data depth: descriptive statistics, graphics and inference. The annals of Statistics, 27: 783-858.
102. Mosler, K., 2013. Depth statistics (Becker C., Fried, R., Kuhnt, S., eds.), Robustness and Complex Data Structures, Festschrift in Honour of Ursula Gather, Springer, Berlin.
103. Serfling, R., 2006. Depth functions in nonparametric multivariate inference. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 72: 1.
104. Rousseeuw, P.J. and M. Hubert, 1999. Regression depth. Journal of the American Statistical Association, 94: 388-402.
105. Mahalanobis, P., 1936. On the generalized distance in statistics. Proceedings of the National Institute of Sciences (Calcutta), 2: 49-55.
106. Tukey, J.W., 1975. Mathematics and the picturing of data. In: Proceedings of the international congress of mathematicians, 2: 523-531.
107. Barnett, V., 1976. The ordering of multivariate data. Journal of the Royal Statistical Society Series A (General), 139: 318-355.
108. Oja, H., 1983. Descriptive statistics for multivariate distributions. Statistics and Probability Letter, 1: 327-332.
109. Liu, R.Y., 1990. On a notion of data depth based on random simplices. The Annals of Statistics, 18: 405-414.
110. Fraiman, R. and J. Meloche, 1999. Multivariate L-estimation. *Sociedad* de Estad?stica e Investigación Operativa Test, 8: 255-317.
111. Zuo, Y. and R. Serfling, 2000. General notions of statistical depth function. The Annals of Statistics, 28: 461-82.
112. Hu, Y., Q. Li, Y. Wang and Y. Wu, 2012. Rayleigh Projection depth, Comput. Stat., 27: 523-530.