International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Critical Review on Outlier Detection Techniques

Poonam Rana[1], Deepika Pahuja[2], Ritu Gautam[3]

Lecturer, DAVIM Faridabad, India

Abstract: ***Outlier Detection is a Data Mining Application. Outlier contains noisy data which is researched in various domains. The various techniques are already being researched that is more generic. We surveyed on various techniques and applications of outlier detection that provides a novel approach that is more useful for the beginners. The proposed approach helps to clean data at university level in less time with great accuracy. This survey includes the existing outlier techniques and applications where the noisy data exists. Our paper defines critical review on various techniques used in different applications of outlier detection that are to be researched further and they gives a particular type of knowledge based data i.e. more useful in research activities. So where the Anomalies is present it will be detected through outlier detection techniques and monitored accordingly especially in educational Data Mining.***

Keywords: Outlier, Outlier Detection, Anomaly, Self and Card watch.

1. Introduction

An Outlier is a data object that significantly deviates from normal objects as if it were generated by different mechanism. Outlier is different from noise as noise is a random error or measured variance and it should be removed before outlier detection. Outlier detection aims to find patterns in data that do not conform to expected behavior. An example is illustrated in figure 1.Outlier detection is extensively used in a wide variety of applications such as military surveillance for enemy activities to prevent attacks, intrusion detection in cyber security, fraud detection for credit cards, insurance or health care and fault detection in safety critical systems and in various kind of images.

It is important in analyzing the data due to the fact that they can translate into actionable information in a wide variety of applications. An irregular traffic pattern occurrence in a computer network could mean that a hacked computer is sending out sensitive data to an unauthorized destination [1].

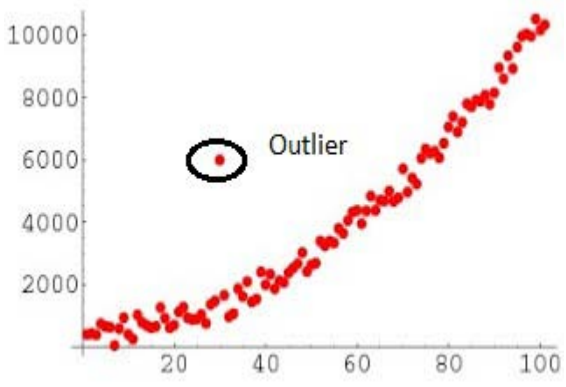

Figure 1

In outlier detection the normal behavior is to characterized by the model and deviation comes from the model is an outlier. The normal behavior model represents the positive patterns that are allowed and also negative patterns that are detecting as an outlier [6]. To detect a subsequence event in the event sequence is very important in various applications like in intrusion detection, doubtful activities monitoring and in molecular biology [18].

Credit card fraud is an open problem in the banking industry outlier detection is used to detect the Fraud. Neural Network is trained over the various credit frauds like lost cards, stolen cards and mail Frauds to detect a fraud [20]. Credit card fraud basically having two characteristics in which time span fraud will happen. The first one in which acceptance and the rejection has been made in very limited span of time and other one huge credit card operations are to be processed at less span of time period [22].

1.1 Types of Outliers

A very important aspect of an outlier detection technique is the nature of the desired outlier. Outlier Classification is done on the basis of their occurrence; generally there are three kinds of outliers which are enumerated as follows:

- Point Outliers
- Contextual Outliers
- Collective Outliers.

Point Outlier: when a data instance is different from set of data then instance is termed as point outlier. It is the simplest form of outlier and used in various researches. For example credit card fraud Detection, the outlier can be detected with respect to amount spent if expenditure is higher compared to normal transactions then it is an outlier.

Contextual Outlier: when a data instance is anomalous with respect to some context (condition), then instance said to be Contextual Outlier. Contextual outliers mostly explored on time series data. For example, in context of age a six feet adult may be a normal person while six feet child is an outlier.

Collective Outlier: When a collection of related data is anomalous from rest of the entire data set, then it is a collective Outlier. They can occur only in data sets where data instances are related. Collective outlier has been explored on graphical data, sequential data and spatial

data. For example Human Electrocardiogram output illustrated in Figure 2. The highlighted region is an outlier because same value exists for the abnormally long time. Collective outliers can be applied for graph data, sequence data and spatial data.

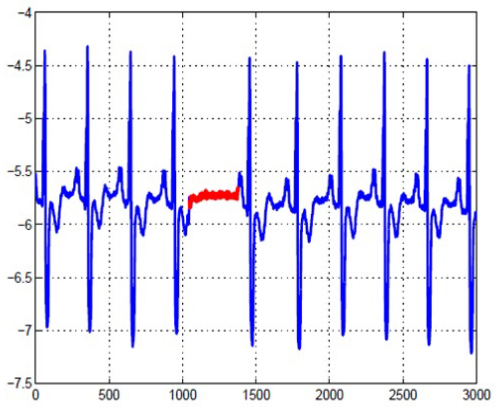

Figure 2

1.2 Reason for occurrence of an Outlier

The main cause for occurrence of an outlier is anomalous data. There may have been an error in data transmission or transcription. Outliers arise due to the change in human error, fraudulent error, system behavior, instrument error or simply by natural deviations in population. Identifying the Outlier is a subjective exercise because it consist all the rigid methods that are used to solve any particular problem. Outlier detection [50] has been used for centuries to detect and remove unusual observations from data where necessary. Outlier detection can recognize system faults and fraud before they can result in potential catastrophic consequences.

1.3 Challenges of Outlier detection

Modeling Normal object - Normal objects always creates a problem to identify an outlier from the non-outliers (normal objects). It is very difficult to enumerate all the normal behaviors in a specific application. Outlier detection quality is basically depending on the modeling of normal objects. So this is the challenge to identify the outliers from the normal objects. Sometimes some methods are used to give as a name to the data a normal object or outlier.

Application Specific Outlier-Another challenge of outlier detection is application specific based. The outlier detection methods varies according to the application like medical application which depend on a specific set of data and marketing analysis basically formed with the various fluctuations and daily routines changes rapidly so that outlier detection method will change accordingly. It is very difficult to identify the outlier when application changed so it is required to change the method also.

Noisy Data- Handling the Noisy data in outlier detection is a very big challenge we already define that Noise is different from the outlier but noisy data that disguide the outlier. The noisy data will blur the correct data and deviates point that leads to effect in detecting the outlier from the data set. Sometimes the noisy data often low quality data that deviates the point of outlier detection and it has a huge challenge for outlier detection.

Understandability- Another challenge of outlier detection is understandability basically to understand the concept of outlier and if the outlier is detected so it is to understand that why these points are outliers.

2. Related Work

The survey conducted including techniques of outlier detection on the basis of their application and also provided a broad way of advantages and disadvantages of the techniques in a comparative way [58].Various techniques of outlier detection in data mining already been discussed. The comparative study of distance based outlier detection technique and density based outlier detection technique was given[59].A comparative study of various outliers methods in medical data, which is used in the medical diagnoses. The analysis conducted on the multi-dimensional data set i.e. based on the statistical methods [62].Survey conducted on different kinds of datasets (univariate, bivariate and multivariate), different outlier detection techniques like supervised, unsupervised, parametric and non-parametric techniques even for complex dataset and also explained different outlier detection approaches with their strength, weakness and application[65].

3. Outlier Detection Applications

Outlier detection is used in various domains of applications. It can easily be used with data, image, and software. Basically anomaly detection and misuse is used for removing the noisy data and producing accurate data set. Various applications of outlier detection are enumerated below:

3.1 Intrusion Detection

Intrusion detection identifies all of the suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system.

Types of Intrusion Detection

There are different kinds of intrusion systems based on network, system calls and kind of packets which are enumerated as follows:

- **Network - Based Intrusion Detection**

Network intrusion detection system usually consists of several sensors at different node and also known as Intrusion Detection Prevention system. It checks the network traffic by connecting to a network hub or network tap. Basically it identifies the content of individual packets for malicious traffic. It provides a real time corrective action to correspondence attack.

- **Host – Based Intrusion Detection**

Host intrusion detection system (HIDS) and software applications (agents) installed on workstations. The agent monitors the operating system by writing log files and triggering alarms. It identifies the intrusion by analyzing system calls; file system modification etc. even in critical

servers. Drawbacks of HIDS are difficult to detect intrusion on multiple computers and to maintain in large networks with different operating system and configurations.

- **Knowledge-based Intrusion detection**

Knowledge-based intrusion detection prevents database from previous attack signatures and unknown vulnerabilities to identify future intrusions. Drawback of Knowledge-based intrusion, it may fail to identify unique attacks due to signatures continual updates requirement.

- **Behavior-based Intrusion Detection**

Intrusion Detection Systems based on behavior also known by the name Anomaly Intrusion Detection System uses a learned pattern of normal activity to identify intrusions. If it deviates from learned pattern then alarm will be triggered.

- **Stack-based Intrusion Detection**

Stack-based intrusion detection, examines the packets through the TCP/IP stack to detect an intrusion in a system.

Generally two approaches are used to detect computer security intrusion system in real time: Misuse and Anomaly detection. Misuse detection aims to detect known attacks against computer system while anomaly detection uses knowledge of users normal behavior to detect attempted attacks. Some outlier detection techniques used in Intrusion Detection given below in Table 1.

Table 1: Intrusion Detection

Name of technique	*Type of Intrusion*	*References*
Neural Networks	Network based intrusion[2], Process based intrusion[3], Software based intrusion[4], Geodetic Network (ADALINE algorithm)[68]	[2] - [4], [68]
Statistical Profiling Using Histograms	Stack and Host-based[6], Host based[7], Host-based[9, 10], Network-based[11], Host-based[12] and Stack-based[13]	[6] - [13]
Rule Based Systems	Host and Network based	[15]
Parametric Statistical Modeling	System based	[18]
Non-Parametric Statistical Modeling	Network based	[19]

Anomaly detection algorithms are used to detect intrusion using the program profile behavior [2].A large corpus of data is explored in order to detect novel attacks and misuse detection to detect known attacks and its variations using artificial neural networks[3]. Neural networks with anomaly detection approach used to detect existence and unknown potential intrusions in computer software system [4].

Anomaly intrusion detection scheme is used to detect TCP connections and detect calls in executive process[6].In natural intrusion system "self" plays an important role. The Natural immune system has several properties which are important for computer society. They include : (1)detection is distributed and each copy of system to be unique.(2) detection is probabilistic and online.(3) detection should recognize foreign particles not only seen one. On the basis of these properties a computer virus detection method was implemented at file authentication level, and self was defined statistically in terms of files containing programs and protected data[53].

For general purpose self is dynamic and complex in natural immune system for this method was developed for defining self in UNIX running processes for normal pattern of system calls in online communication intrusion system using the concept of signatures [7].

A method for distinguishing self from other in context of computational systems using negative selection and it is feasible on Change detection method in any computer virus system and faces the problem of computational difficulty which in turn protects antibodies to take modified form of self. The major disadvantage with this method is that it is not able to detect the virus if designed using the same logics of program used for checking which is quite difficult [8].

Generally intrusion system uses normal data and intrusion data for building up classifiers. Intrusion data is scarce and very difficult to collect. They build an intrusion system using normal data with the help of novelty detection approach by using non-parametric density estimation approach based on Parzen –window estimators with Gaussian kernels. The major advantage of approach is that it can easily adapt data changes and doesn't require any training while new integrations and thus best for changing network environments [19].

3.2 Fraud Detection

Fraud detection is at alarming rate and hence becomes a great threaten for the institution and banks using a credit card transactions Fraud is reported under crime activities that includes banks, mobile phones fraud detection, commercial etc. Outlier is basically used to detect a noisy data that is being presented in the original data. The various techniques are applied to detect a fraud these are presented in table 2 enumerated below.

Table 2: Fraud Detection

Name of technique	*Type of Fraud*	*References*
Neural Network	Credit Card Banking[20, 21, 23], financial accounting fraud detection [64]	[20, 21, 23, 64]
Novelty Detection	Online short Detection	[25]
Rule-based	Credit card Banking	[22]
Clustering	Credit card Banking	[24]
Bayesian Classification and decision tree	Financial accounting fraud detection	[64]

Software developed and implemented using neural network on Mellons bank mainframe computers resulting in reduced fraud consistently accurate and timeless of fraud detection [20]. The neural network requires high diagnostic quality, to overcome this an algorithm developed that generalizes the transaction data and obtain higher level of diagnostic rules then conflating with rule based information and a classification results in better fraud detection[21].

A neural network based database mining technique developed for credit card fraud detection, only feed forward network implemented in Cardwatch, the user can manipulate various parameters through GUI. The major advantage of the model is very extensible and can work on variety of commercial databases [23].

Technique developed using Novelty detection method for on-line short detection which basically detects short windings, DC field windings of large synchronous turbine generators has various problems like shorted turns can cause vibrations in machine and can eventually lead to hazardous breakdown, a catastrophic effect even in absence of rotor and operating cost which play an important role in online systems. To overcome this, a twin-signal sensing method is used in which pulses are injected into the each terminal of rotor. Method was tested on running test rotor with voltage excitation and it gives more accurate results compared to others. Advantage of this is it is simple and online and can result in superb detection [25].

A model namely Minerva was constructed using a neural network embedded with a nonlinear version of fisher's discriminant analysis which adequately separates a good operation of fraudulent operations from other closer to normal traffic. Minerva was installed in SEMP's transactional system dominated by disk access and has a quite good rating time and preventing human intervention queries and amount from frauds. Minerva has both technically as well as economically feasibility [22].

3.3 Medical and public health outlier detection

The patient data is to be collected from the various features of patient like blood test, height, weight, patient age. The various methods of outlier detection is used in medical diagnoses which helps to detect critical diseases at early stage for preventing it to become a severe and life-taking disease. Outlier detection plays a major role in detecting various kinds of cancers. Some outlier detection techniques used in Medical and public health are illustrated below in Table 3.

Table 3: Medical and public health

Name of Technique	*Type of Disease*	*References*
Parametric Statistical Modeling	ECG[26]	[26]
Bayesian Approach	MYELOMA CANCER	[5]
Cluster-based	Lympography and Breast cancer	[55]
Fuzzy Logic	Heart Diseases	[66]
TANAGRA (Data mining tool)	Breast Cancer	[70]
Neural Network	Hypoglycemia	[71]
k-Nearest Neighbor Classifier	electroencephalogram (EEG)	[72]

Brute force approach used to find less similar sequences in time series data called discords has various limitations like complexity, using Heuristics with SAX (Symbolic Aggregate Approximation) results in finding discords faster. ECG (electrocardiograms)are a time series of electrical potential between two points on the surface of body caused by heart and it may contain small discords which needs to be discovered for good health of patient, they are detected very fast using the heuristic approach[26].

An approach proposed to discover cluster based local outlier that provides importance to the local data behavior and new find CBLOF algorithm is used for discovering outlier resulting in detecting meaningful and interesting outliers. An approach is used in Lympography to detect a defect in bones acting as an outlier[55].

Using the generation algorithm of rough set theory 91 rules developed then minimized to 72 through domain intelligence and then finally to 65 through validation and threshold values. The minimized rules are explored to identify the characteristics affecting the relationship between heart disease and its attributes by using formal concept analysis. This helps in detecting heart disease at early stage[66].

A new Data mining tool namely TANAGARA was introduced which helps to detect the complex relationship among patient symptoms, diagnosis and resulting behavior from large database of patient to detect breast cancer [70].

Evolving a new approach to detect an Hypoglyecemia through block based neural network. The block based neural network structure is two dimensional consist of four blocks with input and output node. The weights are associated with the input nodes. To optimize the weight of

block based neural network where a swarm optimization is presented to enhance the performance. This approach gives satisfactory results to detect aHypoglycemia [71].

To detect the features of human stress by the combination of two techniques that is electroencephalogram (EEG) and spectral centroids techniques. Power ban ration is used to create a classifier to produce a high accuracy of k-nearest neighbor classifier [72].

3.4 Image Detection

Images can be of any type main aim of outlier detection is to detect an abnormal behavior of the images. Each data consist of the various features of the image that includes color, brightness, image co-ordinates and texture. The techniques to detect an outlier in images are illustrated in table 4

Table 4: Image Detection

Name of Technique	*Features of image*	*Reference*
Regression	Wavelength	[32]
Clustering	Hyperspectral image	[33]
Neural Networks	Sequential images	[36]
Mixture of Models	Mammogram	[38]
Classification	Spatiotemporal data in image sequence	[34]
Support Vector Machines	Multispectral and hyperspectral images using segmentation	[35]
Hidden Markov Model	General images	[14]

Near-Infrared (NIR) spectroscopy is an analytical tool which is used to measure directly and indirectly physical and chemical properties. In order to get correct results for the quantitative NIR measurement using a new strategy named as CWT-mIPOW-PLS to eliminate the outliers. The Results shows the effectiveness of CWT domain that avoids the background noise and eliminating the useless wavelengths [32].

Table 5: Text Data

Name of technique	*Kind of data/dataset*	*References*
Neural Networks	Synthetic Data[29], BINARY DATA[40]	[29], [40]
Support Vector Machines	Reuters dataset	[41]
Clustering Based Approach	Dynamic Text, graph[61] and time series data[28]	[61], [28]
Machine learning	spatial database	[60]
Wavelet Transform	Large dataset feature	[57]
Non-Parametric Statistical Modeling Approach	Time series data	[27]

Time-efficient method is used to detect an anomaly in hyper spectral images. The hyperspectral data is used Reed–Xiaoli (RX) algorithm to detect an anomaly. RX algorithm compares the pixels spectrum to the local spectrum of its surroundings. Anomaly is detected by using the hyper spectral data and using end member spectra. End member spectra arecomposite spectrum which may or may not be a real spectra[33].

Video Surveillance is an application domain which provides a low- cost monitoring activity in a given environment. Video Surveillance system is provide the integrated view to the environment to respond the ongoing activities, it is used in various sectors like military, real time interpretations so information is more valuable. Classification technique is used to examine the spatial features of the image sequence [34].

Clustering Technique and rejection filtering mechanism is determined to analyze the inter versus intra cluster distances to represent the data in the cluster. This method used for finding the good quality of image segmentation and better methods of feature extraction. Pattern Recognition problem is also analyzed to train the neural network [36].

Mammogram is a photographic film of X-Ray of a breast which is analyzed by experts to detect breast cancer, a novelty detection approach using local parzen estimators helps experts by highlighting the interesting regions in mammogram[38].

A class of image probability models developed namely HIP(Hierarchical image processing models), very similar to hidden Markov tree model with capability of modeling more general image structures and it can be used for variety of image processing tasks like compression, classification, noise suppression, up-sampling and altering errors etc. [14].

3.5 Text data Detection

Noisy data is present in the pile of contents that is to be detected through the outlier techniques. The data can be spatial or can be a temporal means spatial related to the geographical conditions and temporal related to the time aspects. The main aim of outlier detection is to handle the noisy data that is presented in the pile of text. Various techniques for detecting anomalies in Text are enumerated in Table 5.

Table 5

MultiScale Approach	Spatial and Temporal Data	[54]
Statistical Approach	Multivariate Data[56], Text data[39]	[56], [39]
Clustering, K-generation, Thompson's Tau method, max-flow min-cut algorithm.	Multidimensional Data	[67]
Local Outlier Factor	Synthetic and Real time data	[37]

Strong Statistical techniques introduced to face may challenges like: hierarchical shrinkage for rare data,

statistically "garbage collection "for detecting new event, clustering in time to distinguish the different events for same topic, and deterministic annealing for creating the hierarchy. All challenges re solved by bringing together Expectation-Maximization, Shrinkage, vertical word movement and annealing data deterministically. Although the problem with this approach is only suitable for small datasets in detecting and Tracking topics[39].

Giving an approach of new algorithm i.e. LOF(Local Outlier Factor) to detect an outlier from the input data. LOF algorithm proposed less computational time and also dynamically updating the profiles of data streams, also this algorithm computationally very efficient as comparative to other algorithms [37].

Anomaly detection method for spacecraft was developed using two data mining techniques-pattern clustering and association rule mining on time series data. It is used to detect faults and has two additional features – it requires little prior knowledge which is obtained automatically and it uses association rule mining so it can be implemented in space crafts and can even detect anomalies that can't be found by previous fault detection techniques[28].

Novel input data is a great source of error as it significantly differs from trained data so an approach proposed novelty detection through neural network, basically to investigate the relationship between the Novelty as an input data and correspondingly output comes from neural network, synthetic data is generated to analyze multiphase pipeline and performance demonstrated through the oil flow in pipelines. It is very expensive to implement through kernel based estimator as comparison to semi-parametric estimator [29].

A Multiscale approach proposed to detect spatial temporal outlier by analyzing the change between consecutive spatial and temporal scales by discovering unexpected, interesting and implicit knowledge, such as local instability. Spatial-temporal outlier (STO) is defined as a spatial-temporal referenced object whose thematic attribute values are significantly different from those of other spatially and temporally referenced objects in its spatial or/and temporal neighborhood[54].

It is difficult to find missing values in large dataset using formal inspection and graphical displays. A semiautomatic method for outlier detection is proposed that work with continuous multivariate survey data in large datasets. This method relies on explicit probability model continuous data and uses normal distribution to Find Out missing values[56].

Novel deviation Detection approach named Find Out developed on the basis of wavelet transform integrating with clustering leads to highly cost effective technique resulting in detecting outliers efficiently and accurately in huge datasets [57].

Support Vector Machines and neural networks implemented on class data of Reuter dataset in order to carry out the comparison SVM gives the overall best performance compared to NN[41].

WEKA, an algorithm made by collection of different machine learning algorithms for solving real world data mining applications which generally find outliers in non-spatial datasets but when integrated with a nearest neighborhood and statistical approach can detect spatial outlier in Spatial dataset[60].

MSTDBCNFHDC, an algorithm was developed on the basis of Minimum Spanning Tree based clustering and Density-based clustering. The algorithm based on the geometric property of partitioned regions/clusters detects outliers from cluster and then uses a new cluster validation to produce best number of "true" clusters with center for each of them. The inter-cluster distances between centers of clusters/regions are used to find best number of noise-free clusters[61].

3.6 Sensor Networks

Now a days sensor networks are used in the various applications of day to day life activities. A sensor network is a grouping of specialized transducers with ability of communication which helps to monitor and record conditions like humidity, pressure, vibrations, intensity of sound, level of pollution and concentration of chemicals etc. at different locations. A sensor network is a communication system which intends to record conditions and monitor at various locations. A sensor network have multiple detection station called sensor node. Each node is portable, less weighted and very small in size. Basically through the outlier detection techniques faulty sensor networks are detected so that communication level is to be increased. Reliability in wireless sensor networks is affected by the various causes like environment condition, using low quality sensors etc. that leads to corrupted data generations by sensor containing missing values. Some outlier detection techniques used in Sensor Networks Table 6.

Table 6: Sensor Networks

Name of technique	Kinds of networks	References
Bayesian Networks	Wireless sensor networks	[43]
Rule-based	Wireless sensor networks	[44]
Nearest Neighborhood Based Techniques	Multisensor network[45] Event based[47], multisensor network(car and light sensor)[49]	[45, 47, 49]
Parametric Statistical Modeling	Wireless[46]	[46]
Aggregation Tree	Wireless Sensor networks	[48]

Bayesian Belief Network method is used for capturing the correlations among the attributes and how to use it to detect outliers and missing values in the streamed data generated from the sensors in a distributed and on-line fashion. The Naive Bayes technique detects the outliers

and missing values accurately. Naïve Bayesian network is used to classify that whether the corresponding observations belong to it or not. If it does not belong to it and beyond to the position so it is an outlier [43].

An algorithm proposed to overcome problems and restrictions in wireless sensor networks to detect outliers with following properties: generic, in-networks, robust and can be used in all sensors. Applicable to all data cleaning applications and even for critical safety applications. The algorithm has more accuracy with reasonable load and low power consumption [44].

A framework proposed based on approximation of distribution on Multidimensional data in resource constrained sensor networks. The computation effort is distributed among various nodes in network in order to share resources thus reducing the communication and processing costs. It is experimented on real and synthetic data and results are highly accurate and effective [47].

An unsupervised approach using two novel concepts-modifier set and candidate set with use of commit-disseminate-verify in aggregation tree has been proposed to detect global outliers in network. It is answerable to both snapshots queries as well as continuous queries yielding efficient results and extendable to multidimensional data [48].

To address the change analysis in correlated multisensor network (car and light) means to compute the anomaly score of each sensor when potential difference is known priorly. A method using a stochastic neighborhood developed having ability to compute the anomaly scores of each sensor which may be difficult to detect by other naive methods [49].

3.7 Related Application Domains

- Mobile Phones Fraud Detection
- Speech Recognition
- Industrial damage Fraud Detection
- Traffic Monitoring
- Detecting web faults

4. Our Contribution

The main contribution of the paper is description of various techniques in brief with their application domain and also describing the critical review of techniques with their type of input. It gives an overview of detecting outliers in various fields like detecting intrusion attacks in different intrusion systems, detecting online frauds in banks, institutions using credit cards and in financial accounting, in medical field detecting outliers help in analyzing various critical diseases at early stage so it can be cured, detecting anomalies in various kinds of images and videos, finding missing values and wrong data in datasets and streams of data, detecting outlier in kinds of sensor networks help to increase communication level in networks. This survey enhances the outlier detection survey done in[52] which is limited to only to technique and their application domain.

After doing a critical review we are able to propose a new novel approach in Educational Data mining that can be used for university admissions, generally admissions in University takes place on the basis of categories i.e. GENERAL, OBC, SC/ST and Physically Challenged with some kind of other quotas like Sports, State Wise and also achievements helps in evaluating Students previous performance, Decomposing the database on the basis of category, Quota and achievements using a factorization method then applying a K-nearest neighborhood approach to decide whether admission can be given to student or not. The proposed technique can be implemented to mine large databases in universities at time of admission process, giving a new scope for researches to help universities in giving correct results and also providing corrupted free admissions.

5. Conclusion

We conclude that critical analysis on applications of outlier detection will help in further research approaches. Outlier information is very useful when data is compared with the original data. The above critical review will help in the further research. Outlier detection approaches gives a simple and concrete output for the given data. Our research work includes the critical analysis on the various application domains and techniques of the outlier detection. It has been a great work for those who want to start the research on outlier detection and its domain. The entire work consist different phases and lots of theoretical concepts regarding the Anomalies.

References

[1] Kumar, V. 2005. Parallel and Distributed Computing for Cyber security. Distributed Systems Online, IEEE 6, 10.
[2] Ghosh, A. K., Schwartzbard, A., and Schatz, M. 1999a. Learning program behavior profiles for intrusion detection. In Proceedings of 1st USENIX Workshop on Intrusion Detection and Network Monitoring. 51–62.
[3] Ghosh, A. K., Schwartzbard, A., and Schatz, M. 1999b. Using program behavior profiles for intrusion detection. In Proceedings of SANS Third Conference and Workshop on Intrusion Detection and Response.
[4] Ghosh, A. K., Wanken, J., and Charron, F. 1998. Detecting anomalous and unknown intrusions against programs. In Proceedings of the 14th Annual Computer Security Applications Conference. IEEE Computer Society, 259.
[5] Grzegorz M. Boratyn, Tomasz G, Smolinski, Mariofanna Milanova, Jaeek M, Zurada, SudeepaBhattacharyya, and Larry J. Suva. BAYESIAN APPROACH TO ANALYSIS OF PROTEIN PATTERNS FOR IDENTIFICATION OF MYELOMA CANCER. Proceedqs of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2-5 November 2003.
[6] Forrest, S., Esponda, F., and Helman, P. 2004. Aformal framework for positive and negative detection schemes. In IEEE Transactions on Systems, Man and Cybernetics, Part B. IEEE, 357 - 373.

[7] Forrest, S., Hofmeyr, S. A., Somayaji, A., and Longstaff, T. A. 1996. A sense of self for Unixprocesses. In Proceedinges of the ISRSP96. 120 - 128.
[8] Forrest, S., Perelson, A. S., Allen, L., and Cherukuri, R. 1994. Self nonself discrimination in a computer. In Proceedings of the 1994 IEEE Symposium on Security and Privacy. IEEE Computer Society, Washington, DC, USA, 202.
[9] Forrest, S., Warrender, C., and Pearlmutter, B. 1999Detecting intrusions using system calls: Alternate datamodels. In Proceedings of the 1999 IEEE ISRSP. IEEE Computer Society, Washington, DC, USA, 133 - 145.
[10] Hofmeyr, S. A., Forrest, S., and Somayaji, A. 1998. Intrusion detection using sequences of system calls. Journal of Computer Security 6, 3, 151 - 180.
[11] Jagadish, H. V., Koudas, N., and Muthukrishnan, S. 1999. Mining deviants in a time series database. In Proceedings of the 25th International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 102- 113.
[12] Cabrera, J. B. D., Lewis, L., and Mehra, R. K. 2001. Detection and classification of intrusions and faults using sequences of system calls. SIGMOD Records 30, 4, 25 - 34.
[13] Gonzalez, F. A. and Dasgupta, D. 2003. Outlierdetection using real-valued negative selection. Genetic Programming and Evolvable Machines 4, 4, 383- 403.
[14] Spence, C., Parra, L., and Sajda, P. 2001. Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model. In Proceedings of the IEEE Workshop on Mathematica Methods in Biomedical Image Analysis. IEEE Computer Society, Washington, DC, USA, 3.
[15] Lee, W. and Stolfo, S. 1998. Data mining approaches for intrusion detection. In Proceedings of the 7th USENIX Security Symposium. San Antonio, TX.
[16] Lee, W., Stolfo, S., and Chan, P. 1997. Learning patterns from Unix process execution traces for intrusion detection. In Proceedings of the AAAI 97 workshop on AI methods in Fraud and risk management.
[17] Lee, W., Stolfo, S. J., and Mok, K. W. 2000. Adaptive intrusion detection: A data mining approach. Artificial Intelligence Review 14, 6, 533 – 567.
[18] Gwadera, R., Atallah, M. J., and Szpankowski, W. 2005b. Reliable detection of episodes in event sequences. Knowledge and Information Systems 7, 4, 415 - 437.
[19] Chow, C. and Yeung, D. -Y. 2002. Parzen-window network intrusion detectors. In Proceedings of the 16th International Conference on Pattern Recognition. Vol. 4. IEEE Computer Society, Washington, DC, USA, 40385.
[20] Ghosh, S. and Reilly, D. L. 1994. Credit card fraud detection with a neural-network. In Proceedings of the 27th Annual Hawaii International Conference on System Science. Vol. 3. Los Alamitos, CA.
[21] Brause, R., Langsdorf, T., and Hepp, M. 1999. Neural data mining for credit card fraud detection. In Proceedings of IEEE International Conference on Tools with Artificial Intelligence. 103 - 106.
[22] Dorronsoro, J. R., Ginel, F., Sanchez, C., and Cruz, C. S. 1997. Neural fraud detection in credit card operations. IEEE Transactions On Neural Networks 8, 4 (July), 827 - 834.
[23] Aleskerov, E., Freisleben, B., and Rao, B. 1997. Cardwatch: A neural network based database mining system for credit card fraud detection. In Proceedings of IEEE Computational Intelligence for Financial Engineering. 220-226.
[24] Bolton, R. and Hand, D. 1999. Unsupervised profiling methods for fraud detection. In Credit Scoring and Credit Control VII.
[25] Guttormsson, S., II, R. M., and El Sharkawi, M. 1999. Elliptical novelty grouping for on-line short-turn detection of excited running rotors. IEEE Transactions on Energy Conversion 14, 1 (March).
[26] Lin, J., Keogh, E., Fu, A., and Herle, H. V. 2005. Approximations to magic: Finding unusual medical time series. In Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems. IEEE Computer Society, Washington, DC, USA, 329 - 334.
[27] Desforges, M., Jacob, P., and Cooper, J. 1998. Applications of probability density estimation to the detection of abnormal conditions in engineering. In Proceedings of Institute of Mechanical Engineers. Vol. 212. 687 - 703.
[28] Yairi, T., Kato, Y., and Hori, K. 2001. Fault detection by mining association rules from house-keeping data. In In Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space.
[29] Bishop, C. 1994. Novelty detection and neural network validation. In Proceedings of IEEE Vision, Image and Signal Processing. Vol. 141. 217 - 222.
[30] Diaz, I. and Hollmen, J. 2002. Residual generation and visualization for understanding novel process conditions. In Proceedings of IEEE International Joint Conference on Neural Networks. IEEE, Honolulu, HI, 2070 - 2075.
[31] Parra, L., Deco, G., and Miesbach, S. 1996. Statistical independence and novelty detection with information preserving nonlinear maps. Neural Computing 8, 2, 260 - 269.
[32] Chen, D., Shao, X., Hu, B., and Su, Q. 2005. Simultaneous wavelength selection and outlier detection in multivariate regression of near-infrared spectra. Analytical Sciences 21, 2, 161 - 167.
[33] Olga Duran and Maria Petrou, A Time-Efficient Method for Anomaly Detection in Hyperspectral Images. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 45, NO. 12, DECEMBER 2007
[34] Diehl, C. and Hampshire, J. 2002. Real-time object classification and novelty detection for collaborative video surveillance. In Proceedings of IEEE International Joint Conference on Neural Networks. IEEE, Honolulu, HI.
[35] Hazel, G. G. 2000. Multivariate Gaussian MRF for multispectral scene segmentation and outlier detection. GeoRS 38, 3 (May), 1199 - 1211.

[36] Singh, S. and Markou, M. 2004. An approach to novelty detection applied to the classification of image regions. IEEE Transactions on Knowledge and Data Engineering 16, 4, 396 - 407. To Appear in ACM Computing Surveys, 09 2009
[37] Pokrajac, D., Lazarevic, A., and Latecki, L. J. 2007. Incremental local outlier detection for data streams. In Proceedings of IEEE Symposium on Computational Intelligence and Data Mining(data)
[38] Tarassenko, L. 1995. Novelty detection for the identification of masses in mammograms. In Proceedings of the 4th IEEE International Conference on Artificial Neural Networks. Vol. 4. Cambridge, UK, 442 - 447.
[39] Baker, D., Hofmann, T., McCallum, A., and Yang, Y. 1999. A hierarchical probabilistic model for novelty detection in text. In Proceedings of International Conference on Machine Learning.
[40] Manevitz, L. M. and Yousef, M. 2000. Learning from positive data for document classification using neural networks. In Proceedings of Second Bar-Ilan Workshop on Knowledge Discovery and Learning. Jerusalem.
[41] Manevitz, L. M. and Yousef, M. 2002. One-class SVMS for document classification. Journal of Machine Learning Research 2, 139 - 154.
[42] Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y. 1998. Topic detecion and tracking pilot study. In Proceedings of DARPA Broadcast News Transcription and Understanding Workshop. 194 - 218.
[43] Janakiram, D., Reddy, V., and Kumar, A. 2006. Outlier detection in wireless sensor networks using Bayesian belief networks. In First International Conference on Communication System Software and Middleware. 1 - 6.
[44] Branch, J., Szymanski, B., Giannella, C., Wolff, R., and Kargupta, H. 2006. In-network outlier detection in wireless sensor networks. In 26th IEEE International Conference on Distributed Computing Systems.
[45] Phuong, T. V., Hung, L. X., Cho, S. J., Lee, Y., and Lee, S. 2006. An outlier detection algorithm for detecting attacks in wireless sensor networks. Intelligence and Security Informatics 3975, 735 - 736.
[46] Du, W., Fang, L., and Peng, N. 2006. Lad: localization outlier detection for wireless sensor networks. J. Parallel Distrib. Comput. 66, 7, 874 - 886.
[47] Subramaniam, S., Palpanas, T., Papadopoulos, D., Kalogeraki, V., and Gunopulos, D. 2006. Online outlier detection in sensor data using non-parametric models. In VLDB '06: Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment, 187 - 198.
[48] Kejia Zhang, Shengfei Shi, H. G. and Li, J. 2007. Unsupervised outlier detection in sensor networks using aggregation tree. Advanced Data Mining and Applications 4632, 158 - 169.
[49] Ide, T., Papadimitriou, S., and Vlachos, M. 2007. Computing correlation outlier scores using stochastic nearest neighbors. In Proceedings of International Conference Data Mining. 523 - 528.
[50] Rousseeuw, P. and Leroy, A. : 1996, Robust Regression and Outlier Detection. John Wiley &Sons., 3rd edition.
[51] VARUN CHANDOLA University of Minnesota ARINDAM BANERJEE University of Minnesota and VIPIN KUMAR University of Minnesota, Outlier Detection : A Survey
[52] Karanjit Singh and Dr. ShuchitaUpadhyaya, Outlier Detection: Applications And Techniques IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012 ISSN (Online): 1694-0814
[53] **S.** Forrest, A. **S.** Perelson, L. Allen, and R. Cherukuri. Selfnonself discrimination in a computer. In Proceedings of the1994 IEEE Symposium on Research in Security and Privacy, Los Alamos, CA, 1994. IEEE Computer Society Press.
[54] Tao Cheng Zhilin Li, A MULTISCALE APPROACH TO DETECT SPATIAL- TEMPORAL OUTLIERS.
[55] Zengyou He, XiaofeiXu, Shengchundeng, Discovering Cluster Based Local Outliers.
[56] Bonnie Ghosh-Dastider RAND, Santa Monica OPR, J. L. Schafer, Outlier Detection and Editing Procedures for Continuous Multivariate Data.
[57] Dantong Yu*, Gholamhosein Sheikholeslamiy and Aidong Zhang, Find Out: Finding Outliers in Very Large Datasets. Knowledge and Information Systems(2002) 4: 387{412
[58] Victoria J. Hodge and Jim Austin, A Survey of Outlier Detection Methodologies.
[59] M. O. Mansur, Mohd. Noor Md. Sap, Outlier Detection Technique in Data Mining: A Research Perspective. Proceedings of the Postgraduate Annual Research Seminar 2005.
[60] Shan Huang, Jisu Oh, Spatial Outlier Detection. Computer Science Department, University of Minnesota, 200 Union Street SE, 2004.
[61] T. Karthikeyan1, S. John Peter2 and S. Chidambaranathan3. Hybrid Algorithm for Noise-free High Density Clusters with Self-Detection of Best Number of Clusters. International Journal of Hybrid Information Technology Vol. 4, No. 2, April, 2011
[62] Kornel CHROMIŃSKI, Magdalena TKACZ, COMPARISON OF OUTLIER DETECTION METHODS IN BIOMEDICAL DATA. JOURNAL OF MEDICAL INFORMATICS & TECHNOLOGIES Vol. 16/2010, ISSN 1642-6037.
[62] Jay B. Simha1 and S. S. Iyengar, Customer Value Analysis with Fuzzy Data Mining. Proceedings of the International Conference on Cognition and Recognition.
[63] Anuj Sharma, Prabin Kumar Panigrahi. A Review of Financial Accounting Fraud Detection based on Data Mining Techniques. International Journal of Computer Applications (0975 – 8887) Volume 39– No. 1, February 2012
[64] V. Ilango, R. Subramanian, V. Vasudevan. A Five Step Procedure for Outlier Analysis in Data Mining. European Journal of Scientific Research ISSN 1450-216X Vol. 75 No. 3 (2012), pp. 327-339.
[65] B. K. Tripathy1, D. P. Acharjya1 and V. Cynthya. A FRAMEWORK FOR INTELLIGENT MEDICAL DIAGNOSIS USING ROUGH SET WITH FORMAL

CONCEPT ANALYSIS. International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 2, No. 2, April 2011.

[66] Anbarasi. M. S, Ghaayathri. S, Kamaleswari. R, Abirami. I. Outlier Detection for Multidimensional Medical Data. International Journal of Computer Science and Information Technologies, Vol. 2 (1), 2011, 512-516.

[67] MevlutGullu and Ibrahim Yilmaz, Outlier detection for geodetic nets using ADALINE learning algorithm. Scientific Research and Essays Vol. 5 (5), pp. 440-447, 4 March, 2010.

[68] Samiran Ghosh, Saptarsi Goswami, Amlan Chakrabarti, Outlier detection from ETL Execution trace. 2011 IEEE.

[69] Varun Kumar, Dharminder Kumar, R. K. Singh, Outlier Mining in Medical Databases: An Application of Data Mining in Health Care Management to Detect Abnormal Values Presented In Medical Databases. IJCSNS International Journal of 272 Computer Science and Network Security, VOL. 8 No. 8, August 2008.

[70] PhyoPhyo San, Sai Ho Ling, Member, IEEE, and Hung T. Nguyen,. Block Based Neural Network for Hypoglycemia Detection. 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30 - September 3, 2011.

[71] NorizamSulaiman, MohdNasirTaib, SahrimLias, ZunairahHj Murat, SitiArmizaMohdAris, Noor Hayatee Abdul Hamid,. EEG-based Stress Features Using Spectral Centroids Technique and k-Nearest Neighbor Classifier. 2011 UKSim 13th International Conference on Modelling and Simulation.