Available Online at http://www.recentscientific.com

ISSN: 0976-3031

International Journal of Recent Scientific Research
Vol. 4, Issue, 10, pp.1548-1550, October, 2013

International Journal of Recent Scientific Research

RESEARCH ARTICLE

ANOMALY DETECTION AND APPLICATIONS – A STUDY

Visalakshi. S and Radha.V

Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India

ARTICLE INFO

Article History:

Received 18th, September, 2013
Received in revised form 28th, September, 2013
ted 14th, October, 2013
hed online 28th October, 2013

Key words:
Anomaly detection, discrete sequence, outlier, sequence based.

ABSTRACT

Anomaly detection is a major area in different ground of Research and Applications. Anomaly Detection (AD) is to find entire objects which are different from other objects. Various techniques of anomaly detection are developed for individual domain. This paper describes the concept of anomaly, causes of anomalies, classification of Anomaly Detection, various domains with techniques of anomaly detection.

INTRODUCTION

Anomaly detection refers to the problem of different pattern in data which are conventional to estimated behavior. These unusual behavior patterns are often referred to as anomalies or outliers or contaminants depends on application domains. The importance of anomaly detection is to detect problems, new event and abnormal activities in data. The applications of anomaly detection are: Fraud Detection, Network Intrusion detection, Ecosystem Disturbance, Public Health, Medicine, and Fault Detection etc. This study tells about the broad discussion about anomaly detection. Section II covers the concept of anomalies and different phases of anomaly detection, Section III tells about the different types of anomalies. Section IV tells about the causes & techniques of anomaly etection. Section V describes application of anomaly detection. ection VI concludes the paper.

Anomalies

An anomaly means deviation from the usual behavior. This section includes the phases of anomaly detection, types of anomalies, cause of anomalies and techniques of anomaly detection.

Phases of Anomaly detection

The different phase of anomaly detection includes Nature of Input data, Availability of labels and Output of anomaly detection.

1) *Nature of Input data*

Data input is a collection of data instances and they are referred as point, vectors, patterns, or event. Input data can be classified based on the relationship of instances. Each data instances are described using a set of attributes. Data instance consists of univariant attribute or multivariate attributes. Each technique of anomaly detection needs different data types, for example statistical technique use continuous and categorical data, nearest neighbor technique use distance measure of data, original data use classification and clustering technique.

2) *Data Labels*

A label associated with data instance specifies whether that instance belongs to normal or anomalous. Obtained labeled data tells about the behavior and also tells whether the labeled data is accurate or difficult. Labeling of data is carried out manually through labeled training data sets. Based on labels, anomaly detection techniques will operate on the following three modes:

i. **Supervised Anomaly Detection** - The training dataset labeled is normal, as well as anomaly class in the supervised model and they classify predictive model for normal vs. anomaly classes.
ii. **Unsupervised Anomaly detection** - The training data is not necessary and the technique in this group provides assumption that normal instance is far more frequent than anomalies in rest data.
iii. **Semi-Supervised anomaly detection** - It assumes that the training data contains labeled instances only for normal class and not necessary for anomaly class. It builds model for normal and use the model to find anomalies in the test data.

3) *Output of anomaly detection*

The output of anomaly detection technique is the output of anomalous behavior from the given dataset. The anomaly detection outputs are:

- **Scores** – It assigns an anomaly score for each instance in the test data and the analyst choose specific threshold to select the anomalies.
- **Labels** - It each test instance is assigns a separate label for normal or anomalous.

Types of Anomaly

The different type of anomalies is charted in below Fig.1.

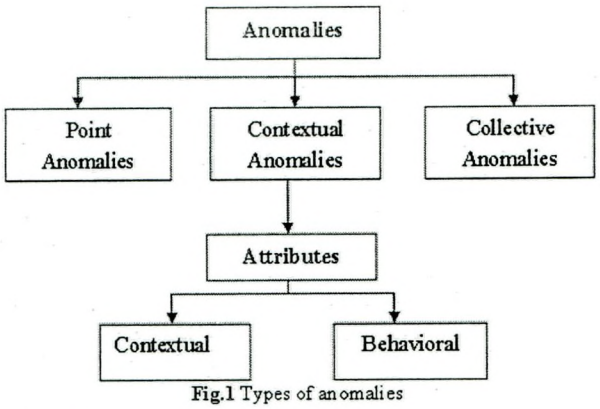

Fig.1 Types of anomalies

Point Anomalies

Among group of data's, the unusual case can be considered as anomalous, which is termed as a point anomaly. Point anomaly is almost used in all research when compared to other anomalies.

Contextual Anomalies

A ding to contextual anomalies or conditional anomaly, the data instance is anomalous in a particular context. The data instance is described using two sets of attributes:

1. **Contextual attributes:** Determining the context or neighborhood for that particular anomalous instance. Example: parameters of water will determine the quality of drinking water.
2. **Behavioral attributes:** It defines only non-contextual characteristics of instances. Example: pH, Temperature will tell the average of pH and temperature value.

Collective Anomalies

When a collection of associated information is anomalous then it is termed as collective anomalies. Normally the collective anomalies explored only in sequence and spatial data.
The data's are detected as anomalies will belongs to either point or contextual or context anomalies.

Causes of anomalies

e various causes of anomalies [3] include:

- Data from Various Module - The objects that are dissimilar from the right place to different type or difference class.
- Normal Dissimilarity - Datasets models by statistical distributions.
- Data Measurement and Error Collection - Error arises when data collected from devices or during entering the data in database.

Techniques to solve anomalies

They are 3 different problem formulations to handle the existing anomalies [4]. They are

- **Sequence-based anomaly detection** – It detects anomalous chain from a database of investigation sequences.
- **Contiguous subsequence-based anomaly detection** – It detects anomalous closest subsequences within an extended sequence.
- **Frequency-based anomaly detection** – It detects patterns in an examination series with anomalous frequency of occurrence.

Applications of Anomaly detection

This section tells the different application of anomaly detection. The different applications are shown in Fig.2 below:

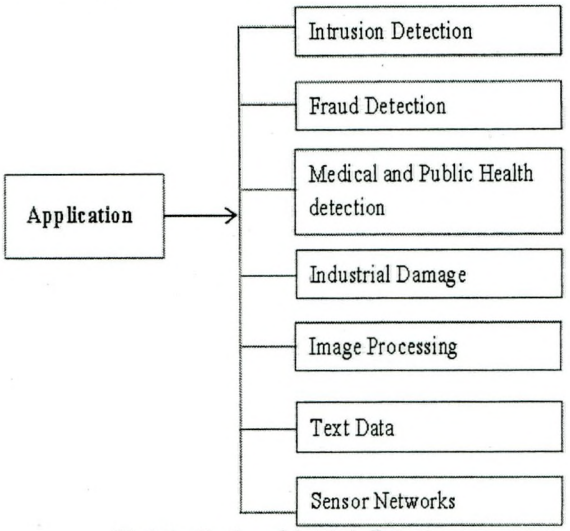

Fig.2 Application of Anomaly Detection

1) **Intrusion Detection** - refers to detecting malicious activity like break-ins, penetration, hacking network and any kind of computer related abuse. [2].
2) **Fraud Detection** - Detects offense activities in commercial organization such as credit card companies, banks, insurance agencies, and cell phone companies to prevent users from creation prohibited earnings.
3) **Medical and Public Health Anomaly Detection** - It is used to maintain patient records and the techniques includes Parametric Statistical Modeling, Neural Networks, Bayesian Networks, Rule-Based System and Nearest Neighbor based techniques [1].
4) **Industrial Damage Detection** - It uses sensor data to protect from damage due to permanent usage, normal wear and tear and prevent earlier from lots of breakages.
5) **Image Processing** - Anomalies causes by motion or insertion of foreign object or by instrumentation errors. The techniques used are mixture of models, Neural Networks, Support Vector Machines, Bayesian Networks, Clustering Based and Spectral.
6) **Text Data** - Anomalies caused due to new fascinating event which are irregular topic are handled in large disparity and the techniques for text data include Statistical Profiling using Histograms, Mixture of Models, Neural Networks, Support Vector Machines and Clustering based.
7) **Sensor Networks** - It detect sensor faults in senor networks and techniques used in this domain are Bayesian networks, Rule Based System, Parametric Statistics Modeling, Nearest Neighbor Based Techniques and Spectral.

CONCLUSION

The current survey provides a broad review about the anomaly detection and techniques. Anomaly detection can detect anomalies in various applications. This paper concludes with overall concept

of anomalies, different phases of anomaly, types of anomalies, reason for arising anomalies, techniques to solve the detected anomalies and the applications of anomaly detection.

Acknowledgement

The author expresses their gratitude to Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India for the progress of research work.

References

1) Martijn Bakke., Tibor Lapikas., Ben Tangena., and Jan Vreeburg. (2012) Monitoring water supply systems for anomaly detection and response.
2) Varun Chandola., Arindam Banerjee., and Vipin Kumar. (2009)Anomaly Detection: Survey", University of Minnesota, ACM.
3) Francesco Tamberi.(2007)Anomaly Detection Data Mining Techniques, Department of Computer Science University of Pisa.
4) Varun Chandola., Arindam Banerjee., and Vipin Kumar. (2012) Anomaly Detection for Discrete Sequences: A Survey, IEEE.
5) Neda Noori1., Leila Boti., and Ebrahim Nowzarpoor Shami.(2012) Surveying Different Aspects of Anomaly Detection and Its Applications.

****