International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 6 554 – 556

Implementation of Anomaly Detection Using Data Mining Technique

Ragini[1], Mahesh Kumar[2]
[1]M.Tech. Student ,Computer Science & Engineering
Ganga Institute of Technology and Management Kablana, Jhajjar, Haryana, India
[2]*Assistant Professor*, Computer Science & Engineering
Ganga Institute of Technology and Management Kablana, Jhajjar, Haryana, India
[1]*raginipreetpanghal@gmail.com;* [2]*maheshmalkani@gmail.com*

Abstract: The Purpose of data mining is extracting vital information from huge databases or the data warehouses. Many Data mining applications have used for commercial & scientific sides. This type of study emphatically discusses Data Mining applications into scientific side. Here Scientific data mining differentiates itself and explores that nature of datasets is various from present market concentrated data mining applications. Most people use pattern matching in some form. Search engines on Web use pattern matching to locate information of interest.

Keyword: *Data mining, KMP, BM, Pattern matching, Anomaly detection*

*****

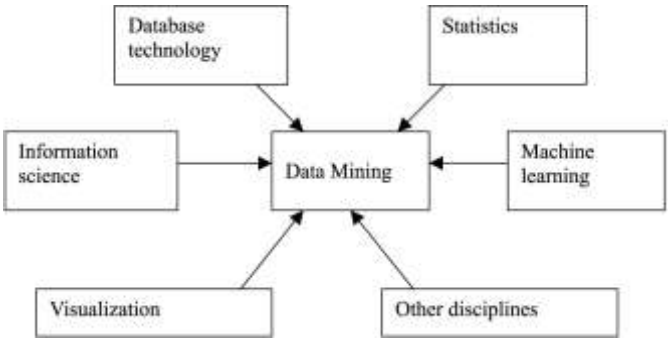
I. INTRODUCTION

Typically anomalous piece would translate to similar kind of trouble like bank fraud, a structural defect, medical problems or errors in a text. Anomalies are also referred to as outliers, novelties, noise, deviations & exceptions.

In actual, in context of abuse and network intrusion detection, interesting objects are often not rare objects, but unexpected bursts in activity. Many outlier detection methods would fail on such data, unless it had been aggregated appropriately. Instead, a cluster study algorithm might be bright to detect micro clusters formed by these patterns.

Three broad categories of anomaly detection techniques exist. **Unsupported anomaly detection** combines detect anomalies in an unlabeled test data set inside assumption that majority of instances in data set are normal by looking for instances that seem to fit least to remainder of data set.

Classification of Data Mining System



II. REVIEW OF LITERATURE

Mohammadjafar Esmaeili (2011) Stream Data Mining & Anomaly Detection in this research attempts to introduce a model that utilizes stream data mining to actively monitor network traffic for anomaly detection.

Sushil Kumar (2012) Anomaly Detection in Network using Data mining Techniques

Implementing new techniques for detecting attacks in network would be examined in further study. Furthermore, data mining techniques could be applied in other domains such as data warehousing in order to improve quality of data.

Harshna (2013) Survey paper on Data Mining techniques of Intrusion Detection

This research had presented a survey of various data mining techniques like feature selection, machine learning & statistical techniques. This paper presents ways in which data mining had been known to aid process of Intrusion Detection & ways in which various techniques have been applied & evaluated by researchers.

Murad A. Rassam (2013) Advancements of Data Anomaly Detection Research in Wireless Sensor Networks

Furthermore, reviewed approaches are compared & evaluated based on how well they meet stated requirements. Finally, general limitations of current approaches are mentioned & further research opportunities are suggested & discussed.

III. PATTERN MATCHING CLASSIFICATION ALGORITHM

Developers usually try to find & rectify bugs by doing an in depth code review along within testing methods & classical debugging strategies. Since such reviews are very

 ISSN: 2321-8169
554 – 556

expensive, there is a need for tools which localize pieces of code that are more likely to contain a bug. Here they have used combination of Hierarchical clustering algorithm & APRIORI algorithm. Based complexity program error would different but every language has predefined syntax & rules & regulation to type code using that compiler trace code line by line. Suppose if any problem occurs in object code, it won't execute properly. To avoid these problems they use an online solution to overcome existing problem.

Finite automata algorithm:- A finite automaton (FA) is a simple idealized machine used to recognize patterns within input taken from some character set (or alphabet) C. job of an FA is to *accept* or *reject* an input depending on whether pattern defined by FA occurs in input.

A finite automaton consists of:

- a finite set S of N states
- a special start state
- a set of final (or accepting) states
- a set of transitions T from one state to another, labeled within chars in C

As noted above, we could represent a FA graphically, within nodes for states, & arcs for transitions. We execute our FA on an input sequence as follows:

- Begin in start state
- If next input char matches label on a transition from current state to a new state, go to that new state
- Continue making transitions on each input char

Our proposed work consists of following steps:

1. First we would create a function to implement KMP pattern matching using MATLAB & test it.
2. In second step we would create Booyer Moore pattern matching using Matlab & test it.
3. study of limitation of KMP & Booyer Moore pattern would be done.
4. Then we would develop a Graphic user interface environment to implement existing KMP & Booyer Moore pattern matching function & get time consumption to perform pattern matching using tic toc function in MATLAB.
5. Then we would develop a new function to implement proposed pattern matching in lesser time.
6. performance chart of existing & new algorithm would be developed.
7. study of performance of proposed pattern matcher would be done using real data set.

IV. EXPERIMENTAL RESULTS

Implementation of KMP

KMP

- Compares text left-to-right
- Uses a failure array to shift intelligently
- takes O(m), where m is length of pattern, to compute failure array
- takes O(m), space
- takes O(n), time to search a string

```
>> booremoore('abc','ghhgghabchjhhj')
Elapsed time is 0.000542 seconds.

ans =

     1
```

Fig:1 Pattern Is Found In Booremoore Search 1 Would Be Returned

Proposed implementation

In proposed algorithm we have reduce time consumption secondly there are no of limited character support in Boore moore algorithm here we have increase probability of number of character & reduced time consumption to perform search. Following are result that represent patter matching in case of KMP , BOOYER MOORE And our proposed work.

Implementation of BM in matlab

BM

- Compares pattern from last character
- Uses bad character jumps & good suffix jumps
- takes O(m + size of alphabet) to compute tables
- takes O(m + size of alphabet), space

```
>> booremoore('xyz','ghhgghabchjhhj')
Elapsed time is 0.000574 seconds.

ans =

     0
```

d

Fig 2 Pattern is not found in booremoore search 0 would be returned

```
>> booremoore('abc','ghhgghabchjhhj')
Elapsed time is 0.000542 seconds.

ans =

     1
```

Fig 3 model is found in booremoore search 1 would be returned

 ISSN: 2321-8169
554 – 556

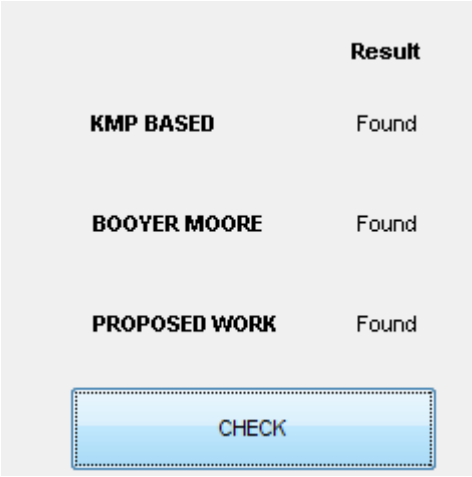


FIG: 4 Result during pattern matching

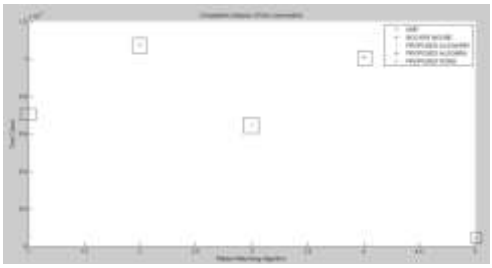


Fig: 5 Mat lab Code To Represent Comparative Analysis

[5] FUTURE SCOPE

In this type of work finding a particular sequence of the tokens is constituted for the presence of constituents of some type of pattern in to find anomaly. The Proposed algorithm will be integrated with a new approach which would surely reduce the consumption of time during algorithm tradition and new algorithms .Here Patterns are generally in the form of sequence or structures of trees. The advantages of this pattern matching is consist of out coming locations of the pattern in a token sequence, to get any component with matched pattern, & for substituting the matching pattern with any other token sequence.

REFERENCES

[1] M. O. Mansur, 2 Mohd. Noor Md. Sap 2005 Outlier Detection Technique in Data Mining: A Research Perspective

[2] Wahidah Husain1, Pey Ven Low2, Lee Koon Ng3, Zhen Li Ong4, 2011 Application of Data Mining Techniques for Improving Software Engineering

[3] Ashish Sureka 2011 Detecting Duplicate Bug Report Using Character N-Gram-Based Features we present an approach to identify duplicate bug

[4] V. Neelima1, 2013 Annapurna Bug Detection through Text Data Mining Volume 3, Issue 5, might 2013 ISSN: 2277 128X International Journal of Advanced Research in Computer Science & Software Engineering

[5] Promila Devi1, Rajiv Ranjan 2014 Enhanced Bug Detection by Data Mining Techniques

[6] Safia Yasmeen 2014 Software Bug Detection Algorithm using Data mining Technique

[7] Dhyan Chandra Yadav April 2015 Software Bug Detection using Data Mining International Journal of Computer Applications (0975 – 8887) Volume 115

[8] Damini.V.S1, Rabindranath2, Priyashree.K3, .Jayashubhaj.K4 10, might 2016 Optimized Error Detection Analytics within Big data on Cloud International Journal of Innovative Research in Science, Engineering & Technology Vol. 5, Special Issue 10, might 2016

[9] Tao Xie & Suresh Thummalapenta, North Carolina State University, David Lo, Singapore Management University, Chao Liu, Microsoft Research ―Data Mining in Software Engineering‖, August, 2009, pp. 55-60

[10] S.Suyambu Kesavan, Dr.K.Alagarsamy ―SE Code Optimization using Data Mining Aproach‖, International Journal of Computer & Organisation – Volume 2, Issue – 3, 2012, pp. 65-68

[11] Abhay Bhatia, Shashikant, Robin Choudhary ―Comparative Study of Pattern Matching Using Text Mining‖, IJRREST: International Journal of Research Review in Engineering Science & Technology (ISSN 2278- 6643) Volume-1 Issue-1, June 2012, pp 58-61

[12] P. Adragna, Queen Mary, University of London, ―Software Debugging Techniques", pp. 72-80

[13] Andreas Hotho, Andreas Nurnberger, Gerhard PaaB, ―A brief of Text mining‖, might 2005

[14] Sarah K Kummerfeld & Judy Kay, ―The neglected battle fields of Syntax Errors‖, 2006

[15] Yiannis Kanellopoulos & Christos Tjortjis, ―Data Mining Source Code to Facilitate Program Comprehension:Experiments on Clustering Data Retrieved from C++ Programs‖