International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

A Detailed Analysis on Windows Event Log Viewer for Faster Root Cause Detection of Defect using Different Graph Plotting Method

Mr. Akshay Wankhade[1], Prof. Pramila M. Chawan[2]

[1]M.Tech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *Condition checking is a standout among the most imaginative ways that organizations and assembling organizations can set aside extra cash. It spins around the guideline of prescient upkeep, which is a proactive method for settling failing gear before they cause issues. It is very necessary that this conditioning monitoring software should work very accurately. For this various data, sensing techniques are used. A plenty of work has been done to build a model based on data sensing techniques such as Data Through proximity sensor, velometer and accelerometer. As a result, software was very complex in nature this leads to maximum chances of defects. In this Study I am performing root cause identification of defect by data analysis using different machine learning algorithm and developing efficient algorithm-based on pattern recognition to detect root cause of defect. The new planned calculation would be anything but easy to execute and have high precision. Another point is look at the past methods and the enhanced one regarding their forecast, capability and exactness.*

Key Words: Condition Monitoring, Sensing techniques, Proximity Sensors, Defects.

1. INTRODUCTION

Event logging and Event logs assume an essential job in present day IT frameworks. Today, numerous applications, working frameworks, arrange gadgets, and other framework segments can log their Events to a neighbourhood or remote log server. Thus, Event logs are a brilliant hotspot for deciding the wellbeing status of the framework, and various instruments have been produced in the course of the last 10-15 years for checking Event signs continuously. Nevertheless, larger part of these instruments can achieve basic undertakings just, e.g., raise a caution following a blame message has been annexed to a log record. Then again, very numerous fundamental Event handling undertakings include occasion connection – a reasonable understanding methodology where new significance is doled out to many occasions that occur inside a predefined time interim.

Previously the Defects are Manage and solve by the experienced developers. It was a very challenging situation for Managers to continuously hire developers with experience and having the capability of solving the defect. Due to which few machine-learning techniques were introduced in order to process large amounts of data without failing. Pattern Recognition techniques have gained popularity over the years because of their ability in discovering practical knowledge from the database and transforming them into useful information. Accuracy of prediction for each of these available techniques varies according to the methods used and the scenario or which it is built.

Presently, a lot of research has been done to develop models based on artificial neural networks. Many different training techniques have been incorporated and have shown to improve accuracy. Currently the defect is assigned a life cycle, also known as Bug Life cycle is the journey of a defect cycle, which a defect goes through during its lifetime. It varies from organization to organization and from project to project as it is governed by the software testing process and depends upon the tools used. Once the bug is posted by the tester, the lead of the tester approves the bug and assigns the bug to developer team. There can be two scenarios, first that the defect can directly assign to the developer, who owns the functionality of the defect. Second, it can also be assigned to the Dev Lead and once it is approved with the Dev Lead, he or she can further move the defect to the developer. In India, currently, company use traditional way to solve a bug based on severity of bug that already been trained. Little work is done to improve the accuracy of solving the defects.

2. Literature Survey

The motivation behind this examination was to break down Microsoft Windows Event logs for curios that might be relevant to an examination. How are agents utilizing Windows Event signs in scientific examinations? How do specialists approach the different kinds of ruptures when gathering information from Windows Event logs? What are the prescribed procedures to break down Windows Event logs?

According to [1] present day IT frameworks regularly produce vast volumes of Event logs, and Event design revelation is an essential log the executive's errand. For this reason, information mining strategies have been proposed in numerous past works. In this paper, we present the Log Cluster calculation, which executes information grouping and line design digging for literary occasion logs.

In order to analyze large amounts of textual log data without well-defined structure, several data mining methods have been proposed in the past, which focus on the detection of

line patterns from textual event logs. The algorithms assume that a single line in the event log describes each event and each line pattern represents a group of similar events. A novel data-clustering algorithm called Log Cluster, which discovers both frequently occurring line patterns and outlier events from textual event logs.

Research of [2] Indicates, Event logs contain immense measures of information that can undoubtedly overpower a human. Along these lines, mining designs from Event logs is a vital framework the board assignment. This paper exhibits a novel bunching calculation for log record informational collections, which causes one to recognize visit designs from log documents, to fabricate log record profiles, and to distinguish abnormal log record lines.

Log file monitoring techniques can be categorized into fault detection and anomaly detection. Because of fault discovery, the space master makes a database of blame message designs. In the event that a line is attached to a log record that coordinates an example, the log document screen makes a specific move. This ordinarily utilized methodology has one genuine blemish - just those deficiencies that are as of now known to the space master can be recognized. In the event that a formerly obscure blame condition happens, the log document screen overlooks the comparing message in the log record, since there is no counterpart for it in the example database. Likewise, usually hard to discover an individual with adequate learning about the framework. Because of inconsistency location, a framework profile is made which reflects ordinary framework movement. In the event that messages are logged that do not fit the profile, a caution is raised. With this methodology, already obscure blame conditions are identified, yet then again, making the framework profile by hand is tedious and blunder inclined. Although association rule algorithms are powerful, they often cannot be directly applied to log files, because log file lines do not have a common format. Furthermore, log file lines seldom have all the attributes that are needed by the association rule algorithms.

Author of [3] Describes Classification is widely used technique in the data-mining domain, where scalability and efficiency are the immediate problems in classification algorithms for large databases. Now a day is large amount of data is generated, that need to be analyses, and pattern must be extracted from that to get some knowledge. Classification is a supervised machine-learning task which builds a model from labelled training data. The model is used for determining the class; there are many types of classification algorithms such as tree-based algorithms, naive Bayes and many more. These classification algorithms have their own pros and cons, depending on many factors such as the characteristics of the data.

According to A Data Classification Method Using Genetic Algorithm and K-Means Algorithm with Optimizing Initial Cluster Center [4] Aiming at the problems of the classical data classification method, this paper proposes a method using genetic algorithm and K-means algorithm to classify data. To improve the effectiveness of data analysis, considering that the classical K-means algorithm is easy to be influenced by the initial cluster center with random selection, this paper improves the K-means algorithm by using the method of optimizing the initial cluster center. This paper first uses the sorted neighborhood method (SNM) to preprocess the data, and then the K-means algorithm is used to cluster data. To improve the accuracy of the K-means algorithm, this paper optimizes the initial cluster center, and unifies the genetic algorithm for the data dimensionality reduction. The experimental results show that the proposed method has higher classification accuracy than the classical data classification method has.

Study of [5] describes some of the primary event logs are Application, Security and System. In addition to several others, Forwarded Events and Setup under the Windows log file allow for the monitoring of remote event logs and alerting of events of interest. There are other logs in the Applications and Services folder including Hardware Events, and Media Center (See Figure 1). To set up the Operational Log, the system administrator, after selecting the Properties, can enable and adjust the size of the log and how often it is updated or cleared. If the computer is not on a corporate network, the individual owner of the system can setup these logs

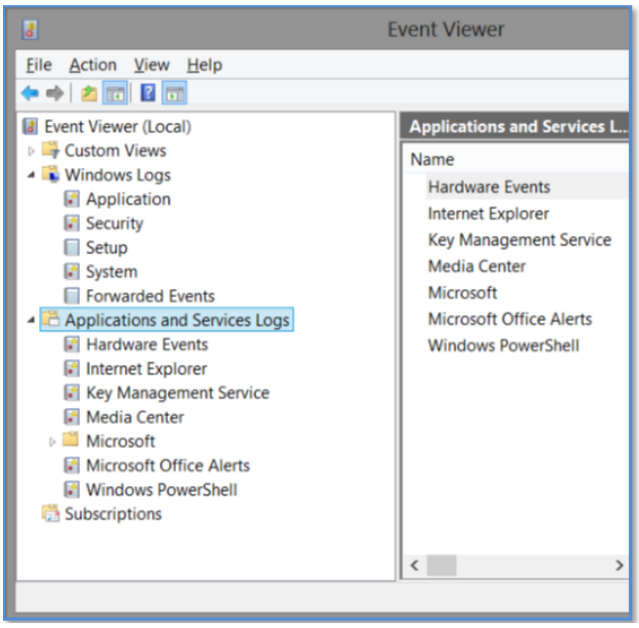

Fig 1: Application Event Viewer

One way event logs can be helpful during an investigation is to search for certain artifacts that will help identify the source of an attack. The IP address is an important key when trying to locate a specific individual or group. This is one of the reasons various event logs are so important for the collaboration and confirmation of other hard data found on the system or network. IP Addresses by themselves cannot be

filtered using the event log as the filtering mechanism. However, a specific IP can be found by using Microsoft's Power Shell Get NetIPAddress Cmdlet (Microsoft, 2012). Another way to get IP addresses would be to create a custom view and use the XML tab to write the query (Paper Cut, 2013). One case where IP addresses were a key point in the investigation is the APT1 unit Cyber Espionage attack. Mandiant, one of the leading security companies, released a report concerning one of the Cyber Espionage units in China (Mandiant, 2013). Mandiant named the unit "APT1". 8 An Advanced Persistent Threat (APT) has become the most common attack on enterprise level network systems. Well-funded and educated Nation States such as China or other criminal groups conducting cyber espionage (Cloppert, 2009) usually carry out APT attacks. Mandiant responded to 150 victims in seven years that APT1 had stolen vast amounts of data from application.

[6]Clustering is an important tool, which has seen an explosive growth in Machine Learning Algorithms. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm is one of the most primary methods for clustering in data mining. DBSCAN has ability to find the clusters of variable sizes and shapes and it will detect the noise. The two important parameters Epsilon (Eps) and Minimum point (Mints) are required to be inputted manually in DBSCAN algorithm and on the basis these parameters the algorithm is calculated such as number of clusters, un-clustered instances as well as incorrectly clustered instances and also evaluate the performance on the basic of parameters selection and calculate the time taken by the datasets. Experimental evaluation based on different datasets in ARFF format with help of WEKA tool, which shows that quality of clusters of our proposed algorithm, is efficient in clustering result and more accurate. This improved work on DBSCAN have used in a large scope.

[7]Among the various methods of supervised statistical pattern recognition, the Nearest Neighbor rule achieves consistently high performance. A new sample is classified by calculating the distance to the nearest training case; the sign of that point then determines the classification of the sample. The k-NN classifier extends this idea by taking the k nearest points and assigning the sign of the majority. It is common to select k small and odd to break ties (typically 1, 3 or 5).Larger k values help reduce the effects of noisy points within the training data set, and the choice of k is often performed through cross-validation. There are many techniques available for improving the performance and speed of a nearest neighbor classification. One approach to this problem is to pre-sort the training sets in some way (such as kd-trees or Voronoi cells). Another solution is to choose a subset of the training data such that classification by the 1-NN rule (using the subset) approximates the Bayes classifier. This can result in significant speed improvements k can now be limited to 1 and redundant data points have been removed from the training set. The nearest neighbor rule is quite simple, but very computationally intensive.

On Weighting Clustering [8], this paper is the first attempt at its formalization. More precisely, we handle clustering as a constrained minimization of a Bregman divergence. Weight modifications rely on the local variations of the expected complete log-likelihoods. Theoretical results show benefits resembling those of boosting algorithms and bring modified (weighted) versions of clustering algorithms such as k-means, fuzzy c-means, Expectation Maximization (EM), and k-harmonic means. Experiments are provided for all these algorithms, with a readily available code. They display the advantages that subtle data reweighting may bring to clustering. The main contribution of this paper is to adopt an insight from classification to improve the performance of unsupervised learning algorithms by making more precise this analogy to boosting algorithms. The raw data is extracted by meter which install on the main electric entrance of the resident, the mean-shift clustering is based on the data after events detector processing, and the result of clustering is used as input data of final identification. An actual residential trail is given to show the clustering is available in the NILM system.

[9]Clustering is an important tool which has seen an explosive growth in Machine Learning Algorithms. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm is one of the most primary methods for clustering in data mining. DBSCAN has ability to find the clusters of variable sizes and shapes and it will also detect the noise. The two important parameters Epsilon (Eps) and Minimum point (Mints) are required to be inputted manually in DBSCAN algorithm and on the basis these parameters the algorithm is calculated such as number of clusters, un-clustered instances as well as incorrectly clustered instances and also evaluate the performance on the basic of parameters selection and calculate the time taken by the datasets. Experimental evaluation based on different datasets in ARFF format with help of WEKA tool, which shows that quality of clusters of our proposed algorithm, is efficient in clustering result and more accurate. This improved work on DBSCAN have used in a large scope.

A Comparative analysis of single pattern matching algorithms in text mining [10] Text Mining is an emerging area of research where the necessary information of user needs to be provided from large amount of information. The user wants to find a text P in the search box from the group of text information T. A match needs to be found in the information then only the search is successful. Many String-matching algorithms available for this search. This paper discusses three algorithms in unique pattern searching in which only one occurrence of the pattern is searched. Knuth Morris Pratt, Naive and Boyer Moore algorithms implemented in Python and compared their execution time for different Text length and Pattern length. This paper also gives you a brief idea about time Complexity, Characteristics given by other authors. The paper is concluded with the best algorithm for increase in text length and pattern length.

An Improved Algorithm for Decision-Tree-Based SVM [11] Decision-tree-based support vector machine which combines support vector machines and decision tree is an effective way for solving multi-class problems. A problem exists in this method is that the division of the feature space depends on

the structure of a decision tree, and the structure of the tree relate closely to the performance of the classifier. To maintain high generalization ability, the most separable classes should be separated at the upper nodes of a decision tree. Distance measure is often used as a separability measure between classes, but the distance between class centres cannot reflect the distribution of the classes. After analysing the tree structure and the classification performance of the decision-tree-based support vector machine, a new separability measure is defined based on the distribution of the training samples in the feature space, the defined separability measure was used in the formation of the decision tree, and an improved algorithm for decision-tree-based support vector machine is proposed. Classification experiments prove the effectiveness of the improved algorithm for decision tree-based support vector machine.

In this paper, author discussed decision-tree-based SVM and the separability measure between lasses based on the distribution of the classes. To improve the generalization ability of SVM decision tree, a novel separability measure is given based on the distribution of the training samples in the feature space. Based on the idea that the most easily separated classes are separated firstly during the decision tree is formed, and by introducing the defined separability measure in feature space into the formation of the decision tree, an improved algorithm for decision-tree-based SVM is obtained. Classification experiments for different data sets prove the performance improvement of the improved algorithm for decision-tree- based SVM.

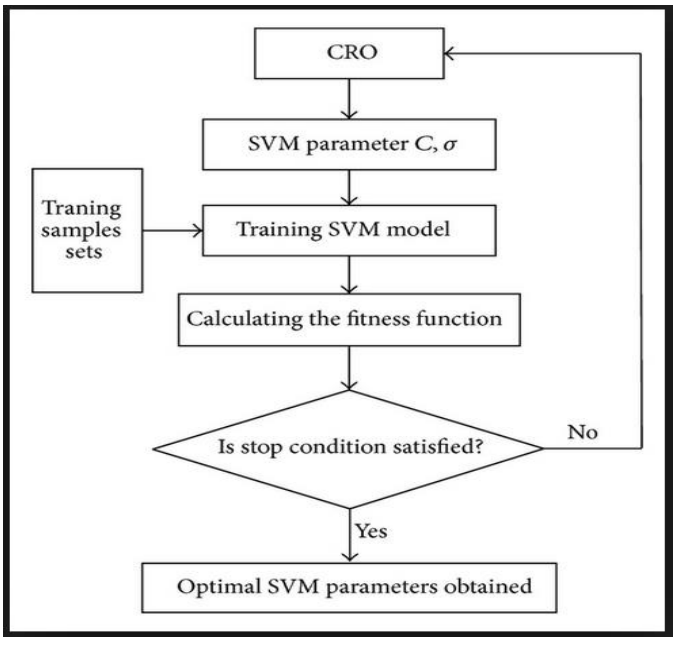

Fig -2: Support Vector Machine

Event Log Analyzer has an agentless design that utilizes worked in syslog and occasion log server to store the occasion logs and syslog is got from all the arranged gadgets, and gives exhaustive occasion, consistence, and custom reports. This helps arrange overseers break down framework issues, enhance organize security, and lessen downtime of servers, workstations, area controllers, switches, and switches of big business systems. The gathered logs are parsed and put away in the inbuilt PostgreSQL database for examination and report age.

3. CONCLUSION

As we understand from the literature survey, there are many algorithms are available for pattern recognition, however there is lot of manual work involved in creating the data set for building log investigator using these algorithms. Since there is a scarcity of data, the model trained is not suitable for real world task. In addition, some drawbacks, which include proper prediction of the defect, are solved by using event viewer. To device a robust policy, the model needs to be trained rigorously with the help of Machine learning. The algorithm will be able to effectively predict the Root cause of defect based on previous data. It will be able to estimate the severity level that should be granted to a developer if it is customized. I stress on building a general algorithm in order to estimate its accuracy and its improvement over other available techniques. This algorithm will be trained on available datasets and tested on the same. Companies will be able to customize it as per their requirements.

REFERENCES

[1] Risto Vaarandi and Mauno Pihelgas, "LogCluster - A Data Clustering and Pattern Mining Algorithm for Event Logs," M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] Zhai Chun-yan, "An Iterative Learning Control Algorithm Based on Predictive Model," M School of Information Science& Engineering, Northeastern University , Shenyang, 110004

[3] Qingshang Guo, Xiaojuan Chang, Hongxia Chu, "Clustering Analysis Based on the Mean Shift," Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation August 5 - 8, 2007, Harbin, China

[4] Ubaid S. Alzoabia, Naser M. Alosaimia, Abdullah S. Bedaiwia, Abdullatif M. Alabdullatifa, "Classification Algorithm On A Large Continues Random Dataset Using Rapid Miner Tool" IEEE Sponsored 2nd International Conference On Electronics And Commutations System(ICECS 2015).

[5] Richard Nock and Frank Nielsen, "A Comparative Study analysis On Weighting Clustering," IEEE Transactions On Pattern Analysis And Machine Intelligence, VOL. 28, NO. 8, AUGUST 200

[6] C. B. Guevara, M. Santos and V. López, "Negative Selection and Knuth Morris Pratt Algorithm for Anomaly Detection," IEEE Trans. Knowl. Data Eng., vol. 24, no. 5, pp. 823–839, 2012.

[7] Chengguo Chang and Hui Wang, "Comparison of two-dimensional string matching algorithms," Department of

Information Engineering, North China University of Water Resources and Electric Power Zhengzhou, China changchengguo@ncwu.edu.cn

[8] Tosin Daniel Oyetoyan Daniela Soares Cruzes, Reidar Conradi1, "Transition and Defect Patterns of Components in Dependency Cycles during Software Evolution," Department of Computer and Information Science Norwegian University of Science and Technology Trondheim, Norway 2SINTEF, Trondheim, Norway.

[9] Wang Min, "Improved K-means Clustering Based on Genetic Algorithm," North University of China Software School Taiyuan City, Shanxi Province, China.

[10] WANG Zhenyu, ZHENG Guilin, "The Application of Mean-Shift Clustering Residential Appliance Identification," Proceedings of the 30th Chinese Control Conference July 22-24, 2011, Yantai, China.

[11] Xiaodan Wang and Zhaohui Shi, Chongming Wu, "An Improved Algorithm for Decision-Tree-Based SVM," Department of Computer Engineering

BIOGRAPHIES:

Mr. Akshay A. Wankhade
Mtech Student, Dept. Of Computer Engineering and IT,VJTI College, Mumbai, Maharashtra, India

Prof. Pramila M. Chawan
Associate Prof, Dept. Of Computer Engineering and IT,VJTI College, Mumbai, Maharashtra, India