Analysis of System Error Log Using Association Mining

Rachana Mishra(Asst.Prof.),
Truba inst.of sc.& Tech,Bhopal
rachanamishra812@gmail.com
+919425012958

Shirish Mohan Dubey(Asst.Prof.),
Shri Ram inst. Of Tech,Jabalpur
shirish_m_dubey@rediffmail.com
+919827342864

Kamlesh Kumar shrivastava
Shri Ram inst. Of Tech,jabalpur
kamleshkumar.shrivastava@gmail.com
+919827676734

Abstract: **With the growth of communication networks, event logs are increasing in size at a fast rate. Today, our operating systems that generate tens of gigabytes of log data per day. Event Viewer maintains logs about program, security, and system events. You can obtain information about your hardware, software & system components. Apriori is frequently adopted to discover frequent item sets, from which strong association rules can be easily generated, from among massive amounts of transactional or relational data. Association rule mining is the one of the most important technique in data mining. We give you useful information for vendor to resolve errors.**

Keword-Data mining, frequent set event viewer, error logs analysis.

I. INTRODUCTION:

In this paper we discuss the how to use Event Viewer as a troubleshooting tool. Today, event logs contain vast amounts of data. Therefore, the mining of frequent patterns from event logs is an important system and network management task. In this paper we discusses the properties of event log data, analysis of the suitability of popular mining algorithms for processing event error log data and proposes an efficient algorithm for mining frequent patterns from event logs[2,3].

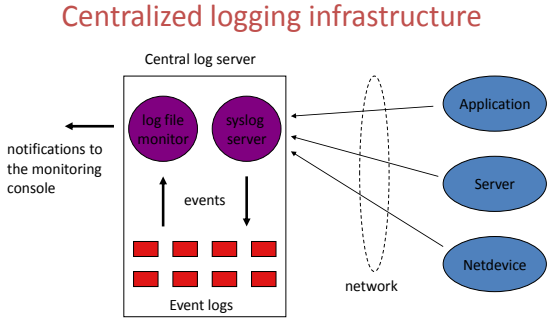
Event Viewer displays detailed information about system events. Event logging and log files are playing an important role in system and network administration. With the growth of communication network, event logs are increasing in size rate.[28] This information includes the event type, the date and time that the event occurred.log data are likely to contain information that deserves closer attention-such as security events, error log and warning. The source of the event, the category for the event, the Event ID, the user who was logged on when the event occurred and the computer on which the event occurred. A component you can use to view and manage event logs, gather information about hardware and software problems, and monitor security events. Event Viewer maintains logs about program, security, and system events. In our invention we use error analysis of event type. Data mining usually required the availability of a large database to be analyzed. This is beginning of error log analysis. There are several ways for defining the frequent event type pattern.

The algorithms assume that each event from the error log has two attributes date of the event occurrence and event type. Various rules and patterns were derived from original database, which could be applied to decision making. Event logs are an important task for network administrators, security analysts and other system management personnel. The obtained knowledge can be useful for reducing the system errors. Recently suggested approaches have been mostly based on the Apriori algorithm for mining frequent itemsets[29] and have been designed for mining frequent event type patters[30,31,32].But in this paper we find frequent errors in our operating system.

A certain combination of event types is considered a frequent pattern. In this paper we can use the data mining [2][3].Our job is helpful for reducing the error in operating system. We are using rule based application. Main aim of this paper is to provide different error type comparisons any way they are occurring in the system. In this paper we will discuss about the application, system and security event, error produced by these three event type. We can formulate the task of mining frequent error event type patterns or frequent error item sets. We use Apriori algorithm for finding frequent sets. The complete event viewer infrastructure is described in figure 1.

Roles of event log.

- ✓ Event logs play an important role in modern IT systems, many system components like applications, servers, and network devices have a builtin support for event logging (with the BSD syslog protocol being a widely accepted standard)[4]



Applications, servers, and network devices use the *syslog* protocol for logging their events to the central log server that runs a *syslog* server. Log monitoring takes place on the central log server and alerts are sent to the monitoring console.

Figure-1

- ✓ In most cases event messages are appended to event logs in real-time, event logs are an excellent source of information for monitoring the system (a number of tools like Swatch and Log surfer have been developed for log monitoring).
- ✓ Information that is stored to event logs can be useful for analysis at a later time, e.g., for audit procedures.

SEEK DIGITAL LIBRARY

II. Data Mining and Event Viewer Descriptions

Data Mining or Knowledge Discovery in Databases (KDD) as it is also known is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. This encompasses a number of different technical approaches, such as clustering, data summarization, learning classification rules, finding dependency networks, analyzing changes, and detecting anomalies. Data mining is the search for relationships and global patterns that exist in large databases but are `hidden' among the vast amount of data, such as a relationship between patient data and their medical diagnosis. These relationships represent valuable knowledge about the database and the objects in the database and, if the database is a faithful mirror, of the real world registered by the database[6].

Data mining refers to "using a variety of techniques to identify nuggets of information or decision-making knowledge in bodies of data, and extracting these in such a way that they can be put to use in the areas such as decision support, prediction, forecasting and estimation. The data is often voluminous, but as it stands of low value as no direct use can be made of it; it is the hidden information in the data that is useful"

Basically data mining is concerned with the analysis of data and the use of software techniques for finding patterns and regularities in sets of data. It is the computer which is responsible for finding the patterns by identifying the underlying rules and features in the data. The idea is that it is possible to strike gold in unexpected places as the data mining software extracts patterns not previously discernable or so obvious that no-one has noticed them before.

Data mining analysis tends to work from the data up and the best techniques are those developed with an orientation towards large volumes of data, making use of as much of the collected data as possible to arrive at reliable conclusions and decisions. The analysis process starts with a set of data, uses a methodology to develop an optimal representation of the structure of the data during which time knowledge is acquired. Once knowledge has been acquired this can be extended to larger sets of data working on the assumption that the larger data set has a structure similar to the sample data. Again this is analogous to a mining operation where large amounts of low grade materials are sifted through in order to find something of value.

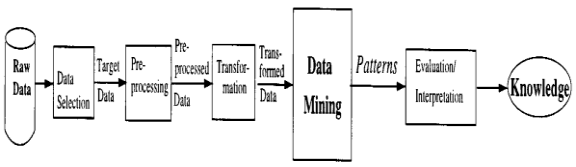


Figure2: All steps in KDD process

The phases depicted start with the raw data and finish with the extracted knowledge which was acquired as a result of the following stages [8][9]:

- *Selection:* - Selecting or segmenting the data according to some criteria e.g. all those people who own a car, in this way subsets of the data can be determined.
- Preprocessing: - This is the data cleansing stage where certain information is removed which is deemed unnecessary and may slow down queries for example unnecessary to note the sex of a patient when studying pregnancy. Also the data is reconfigured to ensure a consistent format as there is a possibility of inconsistent formats because the data is drawn from several sources e.g. sex may recorded as f or m and also as 1 or 0.
- Transformation :- The data is not merely transferred across but transformed in that overlays may added such as the demographic overlays commonly used in market research. The data is made useable and navigable.
- Data Mining: - This stage is concerned with the extraction of patterns from the data. A pattern can be defined as given a set of facts(data) F, a language L, and some measure of certainty C a pattern is a statement S in L that describes relationships among a subset Fs of F with a certainty c such that S is simpler in some sense than the enumeration of all the facts in Fs.
- Interpretation and Evaluation:- The patterns identified by the system are interpreted into knowledge which can then be used to support human decision-making e.g. prediction and classification tasks, summarizing the contents of a database or explaining observed phenomena.

III .INTRODUCTION TO EVENT VIEWER

In Windows XP, an event is any significant occurrence in the system or in a program that requires users to be notified, or an entry added to a log. The Event Log Service records application, security, and system events in Event Viewer.

With the event logs in Event Viewer, you can obtain information about your hardware, software, and system components, and monitor security events on a local or remote computer. Event logs can help you to identify and diagnose the source of current system problems, or help you to predict potential system problems [25][2].

Event log types:-

Microsoft windows server2003,windows XP, windows 2000 server and window NT record event in three kinds of logs:

- Application log : -The application log contains events logged by programs. For example, a database program may record a file error in the application log. Events that are written to the application log are determined by the developers of the software program.
- Security log :-The security log records events such as valid and invalid logon attempts, as well as events related to resource use, such as the creating, opening, or deleting of files. For example, when logon auditing is enabled, an event is recorded in the security log each time a user attempts to log on to the computer. You must be logged on as administrator or as a member of the administrators group in order to turn on, use, and specify which events are recorded in the security log.
- System log :-The system log contains events logged by Windows XP system components. For example, if a driver

SEEK DIGITAL LIBRARY

fails to load during startup, an event is recorded in the system log. Windows XP predetermines the events that are logged by system components.[25] Servers running windows server2003 and window2000 server that are domain controllers might have the following additional logs in event viewer:

- Directory service log:-This includes any information regarding the Active Directory® directory service and Active Directory database maintenance.
- File replication log:-This service is used for replication of files, such as domain policies, between domain controllers.
- DNS server services log:-This log includes events related to the Domain Name System (DNS) Server service running on Windows Server 2003 and Windows 2000 Server.

Event Types:-

The description of each event that is logged depends on the type of event. Each event in a log can be classified into one of the following types [4][3][25]:

- Information :-An event that describes the successful operation of a task, such as an application, driver, or service. For example, an Information event is logged when a network driver loads successfully.
- Warning :-An event that is not necessarily significant, however, may indicate the possible occurrence of a future problem. For example, a Warning message is logged when disk space starts to run low.
- Error :-An event that describes a significant problem, such as the failure of a critical task. Error events may involve data loss or loss of functionality. For example, an Error event is logged if a service fails to load during startup.
- Success Audit (Security log):-An event that describes the successful completion of an audited security event. For example, a Success Audit event is logged when a user logs on to the computer.
- Failure Audit (Security log) :-An event that describes an audited security event that did not complete successfully. For example, a Failure Audit may be logged when a user cannot access a network drive.

IV. DATA MINING FOR EVENT LOGS

Data mining for event logs has been identified as an important system . Management task – detected knowledge can be used for building rules for event correlation systems or event log monitoring tools, improving the design of web sites, etc[6][3].

Recently proposed mining algorithms – mostly based on the Apriori algorithm for mining frequent itemsets, designed for mining frequent patterns of event types. Event log is viewed as a sequence {E1,...,En}, where Ei = (ti, ei), ti – time of occurrence of Ei, ei – type of Ei, and if $i < j$ then $ti \leq tj$. Frequent pattern can be defined in several ways, with most common definitions being window- and slice-based[25].

Shortcomings of existing mining approaches:

- Apriori is known to be inefficient for mining longer patterns,
- Infrequent events remain undetected but are often interesting (e.g., fault events are normally infrequent but highly interesting),
- Focused on mining event type patterns from preprocessed event logs, ignoring patterns of other sorts (in particular, line patterns from raw event logs help one to find event types or write rules for log monitoring tools).

V. IMPLEMENTATION OF APRIORI ALGORITHM:

The obtained knowledge can be useful for various purposes, like writing new rules for event log monitoring tools, handling security incidents, etc. However, today's large networks and IT systems can generate large amounts of log data which makes the manual review of event logs infeasible. Therefore, the automation of event log analysis is an important research problem in network and system management [4].

Our operating system generates number of logs, like Application, System and Security [2]. In this paper we have analyzed different event logs, Event – a change in the system state, e.g., a disk failure; when a system component (application, network device, etc.) encounters an event, it could emit an event message that describes the event.And event logging – a procedure of storing event messages to a local or remote (usually flat-file) event log & log generates by these events (Informational, Warning, Error, Success Audit, Failure Audit etc.).Here we use only error log files for finding the frequent error sets.

Our Project analysis can be useful for a vendor to resolve system software errors because every vendor wants to identify whether its software is working properly or not, sometimes it may be possible that software is not able to identify and recognize the hardware, or software is not having proper driver for that hardware. Then it generates Error. If software is generating errors, then the user wants the solution of the cause. But we know that end user cannot update the software and if this is not reported to the vendor then generating solution for the error is very difficult.

For this purpose we are identifying and generating the Error Events from the System Log and setting priorities between those events for the Vendor which will help the vendor for the future up gradation of the software. Transaction database and itemsets are generating by candidate generation and pruning of itemsets. Itemsets are pruned for comparing the threshold value. Transaction is described in table 1.1

Transaction ID	Itemset
10-01-11	9,11,7
11-01-11	9,7

SEEK DIGITAL LIBRARY

12-01-11	11,26
13-01-11	1002,9
14-01-11	1003,11

Table-1.1. A sample Transaction database and itemset

Due to its excellent performance Apriori is frequently adopted to discover frequent itemsets .The Apriori Algorithms an influential algorithm for mining frequent itemsets for Boolean association rules. Apriori needs several iterations of the data. The Apriori property is based on the following observation. By definition, if an itemset I does not satisfy the minimum support threshold, min sup, then I is not frequent; that is P(l) < min sup. If an item A is added to the itemset I, then the resulting itemset (i.e., I Ụ A) cannot occur more frequently than I. Therefore, I u A is not frequent either; that Uses a uniform minimum support threshold. Difficulties to find rarely occurring events [20].

Apriori property:

All nonempty subsets of a frequent itemset must also be frequent.

Let's define creating frequent sets

C_k as a candidate itemset of size k

L_k as a frequent item set of size k

Main steps of iteration are::-

- Find frequent set L_{k-1}.
- Join step: C_k is generated by joining Lk-1 with itself (Cartesian product L_{k-1} X L_{k-1}).
- Prune step (Apriori property): Any (k − 1) size itemset that is not frequent cannot be a subset of a frequent k size itemset, hence should be removed.
- Frequent set L_k has been achieved.

Objective measures:-

- Based on threshold values controlled by the user
- Some typical measures
- Simplicity
- Support (utility)
- Confidence (certainty)

Determining Attribute Domains:-

To introduce a Apriori for association ,one first has to determine the domains of the table columns. The following attributes domains are required for error analysis [2]

- *Event Type*
- *Event Source*
- *Event Category*
- *Event ID*
- *Date*
- *Time*
- *User*
- *Computer*
- *Description*

VI. SAMPLE OF EVENT ERROR LOG BY TRANSACTION:

Frequent Sets after implementing Apriori when our minimum support is 2 or greater

20-3-2011	[10016, 1002, 32003, 6004, 9]
19-3-2011	[10010, 1002, 4199]
18-3-2011	[1002, 6004, 7, 9]
17-3-2011	[1002]
16-3-2011	[1002, 7]
14-3-2011	[1002]
13-3-2011	[1002]
10-3-2011	[1002]
9-3-2011	[1002]
8-3-2011	[1002]
7-3-2011	[1002]
6-3-2011	[1002, 4319, 8032]
5-3-2011	[11, 4199, 7]
4-3-2011	[11, 7, 9]
20-3-2011	[7]
19-3-2011	[7, 9]
16-3-2011	[7, 7901]
10-3-2011	[7901]

Apriori implementation of error log:-

We select transaction on date wise error. then we find frequent set for different error id.

Frequent Sets after implementing Apriori:-

Item Sets	**Support Count**
1002, 8003, 9	4

Table -2(result of 4 support count)

After minimum confidence threshold, say :-

Frequent Item Set	**Confidence**	**Status**
1002^8003 ->9	66	Rejected
1002^9 ->8003	44	Rejected
8003^9 ->1002	80	Selected
1002->8003^9	14	Rejected
8003->1002^9	57	Rejected
9->1002^8003	23	Rejected

Table-3(result for confidence)

CONCLUSION

We tried to give detail information to industry that worked on operating systems to remove errors, so we do firstly

SEEK DIGITAL LIBRARY

thorough study on event viewer of windows operating systems(client, server). Do analysis on every error which is occurred in runtime or time in installation of any application. Also analysis the type of error which is frequently occurred when operating system is run. We also save the log files from different type of operating systems in different format for analysis and making approach for appropriate result, and graph. Our Project analysis can be useful for a vendor to resolve System Software errors because every vendor wants to identify whether its software is working properly or not, sometimes it may be possible that software is not able to identify and recognize the hardware or software is not having proper driver for that hardware.

Then it generates Error. We do analysis database part and some R &D also done in this field. We also develop an interface to store raw data to database in useful format. We also do R&D to apply mining rules & algorithm in this data for extracting useful information. Our result is based on Apriori are shown in table 4

S.No.	Frequent itemset	Support Count	confidence
1	8003,9,1002	4	80

Table 4 :Result for Apriori

Means if we resolve these three errors then most of the problem will be solve.

FUTURE WORK

There are many areas where this information, analysis and research is useful for improving their product and resolve problem because we provide the information in bulk and give reports and graph. And also give more description on that analysis. So if relevant industry want to use this information they can improved their product for all customer. But this is not the end of event viewer error analysis.

We can continue this analysis for the different operating system. For a future work, we plan to investigate various association rule algorithms, in order to create a set of tools for building log file profiles. We will be focusing on algorithms for detecting error .we developed different tool for finding errors which are frequently occurs.

REFERENCES

[1] Risto Vaarandi, A Breadth-first algorithm for mining frequent patterns from event logs, IEEE international conference, 2003-04.

[2] R.Agrawal,T.lmielinski,and A.Swami," Mining Association Rules between sets of items in Large Database", in proceeding of the ACM SIGMOD International Conference on Management of data ,1993,pp.207-206.

[3] Risto Vaarandi "A Data Clustering Algorithm for Mining Patterns From Event Logs", IEEE international conference,2003.

[4] Risto Vaarandi, "SEC - a Lightweight Event Correlation Tool",Proceedings of the 2nd IEEE Workshop on IP Operations and Management, 2002.

[5] H. Mannila, H. Toivonen, and A. I. Verkamo, "Discovery of frequent episodes in event sequences", *Data Mining and Knowledge Discovery* Vol. 1(3), 1997.

[6] [SM2003]C.M.Sperberg-McQueen, "Web Services and W3C",Aug2003http:/w3c.dstc.edu.au/presentations/2003-08-21-web-services-interop/msm-ws.html/

[7] Springer, "Google LaunchesNewsService", PC World, September 23, 2002, http://www.computerworld.com/ developmenttopics /websitemgmt/ story/0,10801, 74470,00.html.[SGB2001] Spiekermann .

[8] U. Fayyad and R. Uthurusamy, "Data mining and knowledge discovery in databases," *Commun. ACM*, vol. 39, pp. 24–27, 1996.

[9] W. H. Inmon, "The data warehouse and data mining," *Commun. ACM*,vol. 39, pp. 49–50, 1996.

[10] "Special issue on knowledge discovery in data- and knowledge bases,"*Int. J. Intell. Syst.*, vol. 7, 1992.

[11] K. J. Cios, W. Pedrycz, and R. Swiniarski, *Data Mining Methods for Knowledge Discovery*. Dordrecht, The Netherlands: Kluwer, 1998.

[12] R.Agrawal and J.C.Shafer,"Parallel Mining of Association Rules",In IEEE Transactions on Knowledge and Data Engineering,vol.8,no.6,pp.962-969,1996.

[13] C. Lonvick, "The BSD syslog Protocol", *RFC3164*, 2001.

[14] H. Mannila, H. Toivonen, and A. I. Verkamo, "Discovery of frequent episodes in event sequences", *Data Mining and Knowledge* Discovery,Vol. 1(3), 1997.

[15] M. Klemettinen, H. Mannila, and H. Toivonen, "Rule Discovery in Telecommunication Alarm Data", Journal of Network and SystemsManagement, Vol. 7(4), 1999.

[16] R.Agrawal and R.Srikant ,"Fast Algorithms for Mining Association Rules", in proceedings of the 20th international conference on Very Large Databases,1994,pp.487-499.

[17] M.J.Zaki Scalable algorithm for association mining.IEEE Transactions on knowledge and data engineering,12930:372-390,2000.

[18] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In Proc. of VLDB Conference, pages 487–499, 1994.

[19] Christian Borgelt and Rudolf Kruse. Induction of association rules: Apriori implementation. In Proceedings of the fifteenth conference on computational statistics, pages 395–400, 2002.

[20] Sergey Brin, Rajeev Motwani, Jeffrey D. Ullman, and Shalom Tsur. Dynamic itemset counting and implication rules for market basket data. SIGMOD Rec., 26(2):255–264, 1997.

[21] David Hand, Heikki Mannila, and Padhraic Smyth, Principles of Data Mining, The MIT Press, 2001.

[22] Ferenc Bodon. A fast apriori implementation. In Bart Goethals and Mohammed J. Zaki, editors, Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI'03), volume 90 of 57 CEUR Workshop Proceedings, Melbourne, Florida, USA, 19. November 2003.

[23] Christian Borgelt. Recursion pruning for the apriori algorithm. In Jr. et al. (JGZ04).

[24] http://cna.org.uk/is/345.02/acctmanagement/tutorial10.htm.

[25] Jiawei Han, Jian Pei, and Yiwen Yin, "Mining Frequent Patterns without Candidate Generation", Proceedings of the ACM SIGMOD International Conference on Management of Data, 2000.

[26] Seminar of Popular Algorithms in Data Mining and Machine Learning, TKK, 12.3.2008,Lauri Lahti.

[27] Mining event logs with SLCT and Loghound,Risto Vaarandi,ieee,2008.

[28] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 478–499.

SEEK DIGITAL LIBRARY

[29] Q. Zheng, K. Xu, W. Lv, and S. Ma, "Intelligent Search of Correlated Alarms from Database Containing Noise Data," in Proceedings of the 8th IEEE/IFIP Network Operations and Management Symposium (NOMS),2002, pp. 405–419.

[30] S. Ma and J. L. Hellerstein"Mining Partially Periodic Event Patterns with Unknown Periods," in Proceedings of the 16th International Conference on Data Engineering, 2000, pp. 205–214.

[31] M. Klemettinen, "A Knowledge Discovery Methodology for

[32] Telecommunication Network Alarm Databases," PhD Paper, University of Helsinki, 1999.

SEEK DIGITAL LIBRARY