Analysis of Machine Logs to Detect Patterns and Perform Auto-remediation

[1]Juily A. Kulkarni, [2]Shivani S. Joshi, [3]Shriya S. Bapat, [4]Ketaki V. Jambhali
Department of Computer Engineering,
PVG's COET, Pune – 411009.

Abstract : Log data is an important and valuable resource for understanding system status and performance issues[1]. System logs record system states and significant events at various critical points to help debug performance issues and failures, and perform root cause analysis. The log format is the standard log format which contains timestamp, process name, message, log type, id etc. These logs are analyzed to detect any sequence of events which provide us with the patterns necessary for further implementation. From these patterns future critical situations like memory issues, network failure, machine shutdown etc. are found. After detecting these critical situations auto remediation is done by sending alert messages or notifications stating the error which can occur and remedial measures like system restart, code re-execution etc. are carried out by the system which will help in avoiding these future critical situations and help protect the system.

IndexTerms – **Machine logs, Logstash, Elasticsearch, Anomalies, LSTM, Apriori algorithm, Auto-remediation.**

I. INTRODUCTION

Log data is a definitive record of what's happening in every business, organization or agency and its often an untapped resource when it comes to troubleshooting and supporting broader business objectives[2]. Asking a virtual personal assistant for help in debugging a production system may seem like a far-fetched idea, but the idea of using a machine learning or deep learning approach is actually very feasible and practical. These algorithms have proved very useful in recent years at solving complex problems in many fields. From computer vision to autonomous cars to spam filters to medical diagnosis, these algorithms are providing solutions to problems and solving issues where once expert humans were required[3]. System logs record the states of the system at various stages and important events at various critical points to help understand performance issues and failures. This log data is universally available in all computer systems. As system logs record events from actively running processes, they become an important resource for anomaly detection.

Anomaly detection is an important task if we want the system to be secure and efficient. As the systems get more advanced and complex, anomaly detection becomes an essential task. Our work proposes an intelligent system which uses parsed and filtered system logs to detect anomalies using Artificial Intelligence approaches. Patterns containing error messages are classified and detected from these system logs. Based on the anomaly detection and prediction performed, remedial measures are taken wherein an alert message if sent to the user and system takes appropriate remedial measures. This work helps is detecting and fixing the critical situations which may arise in the future so as to avoid the loss of time, memory, data, etc.

II. RELATED WORK

According to our survey, there exist certain systems which perform frequent pattern mining and anomaly detection. Each system has it's unique features and methodologies. We have tried to improve upon these systems by predicting future anomalies and performing auto-remediation for the same.

Min Du, Feifei Li, Guineng Zheng, Vivek Srikumar have proposed the DeepLog system for HDFS and OpenStack log datasets. This method uses a deep neural network based approach. It performs anomaly detection at per log entry level.This paper uses both deep learning approach using LSTM algorithm and classic mining methods such as clustering. DeepLog system is implemented using Keras with TensorFlow as the backend. System logs record system states and important events at various critical points to help debug performance issues and failures, and perform root cause analysis. Such log data is universally available in nearly all computer systems and is a valuable resource for understanding system status. DeepLog, is appproach for anom-aly detection that handles the large volumes of system logs. log entries are viewed as elements of a sequence that follows certain patterns .DeepLog is a deep neural network that models this sequence of logentries using a Long Short-Term Memory (LSTM). This allows DeepLog to automatically learn a model of log patterns containing execution anomalies. Since it is a learning-driven approach,it is possible to incrementally update the DeepLog model so that it can adapt to new log patterns that emerge over time. Each log entry is parsed to a log key and a parameter value vector. The log key sequence parsed from a training log file is used by DeepLog to train a log key anomaly de-tection model. For each distinct key, DeepLog also trains and maintains a model for detecting system performance anomalies. Consider a log sequence "54→57", and suppose the predictedprobability distribution is "{18: 0.8, 56: 0.2}", which means thatthe next step could be either "18" or "56". this ambiguity couldbe caused by using an insuffcient history sequence length. For example, if two tasks share the same workflow segment "54→57",the first task has a pattern "18→54→57→18" that is executed 80%of the time, and the second task has a pattern "31→54→57→56"that is executed 20% of the time. this will lead to a model that predicts "{18: 0.8, 56: 0.2}" given the sequence "54→57". DeepLog can separate out different tasks from a log file and construct a work-flow model for each task

using both deep learning (LSTM) and classic mining (density clustering) approaches. this enables effective anomaly diagnosis. By incorporating user feedback, DeepLogsupports online update/training to its LSTM models, hence is able to incorporate and adapt to new execution patterns[1].

S.VijayaKumar, A.S.Kumaresan, U.Jayalakshmi proposed a system to predict new interesting patterns from Web server access log files. They have applied Apriori algorithm for matching new interesting patterns from the log data and applied support and confidence to calculate the measures of the found patterns[4].

III. PROPOSED METHOD

The proposed system is a system for analysing system logs to detect anomalous patterns and predicting future anomalies. Auto-remediation is done based on these predicted anomalies. The 3 important modules of our system are:

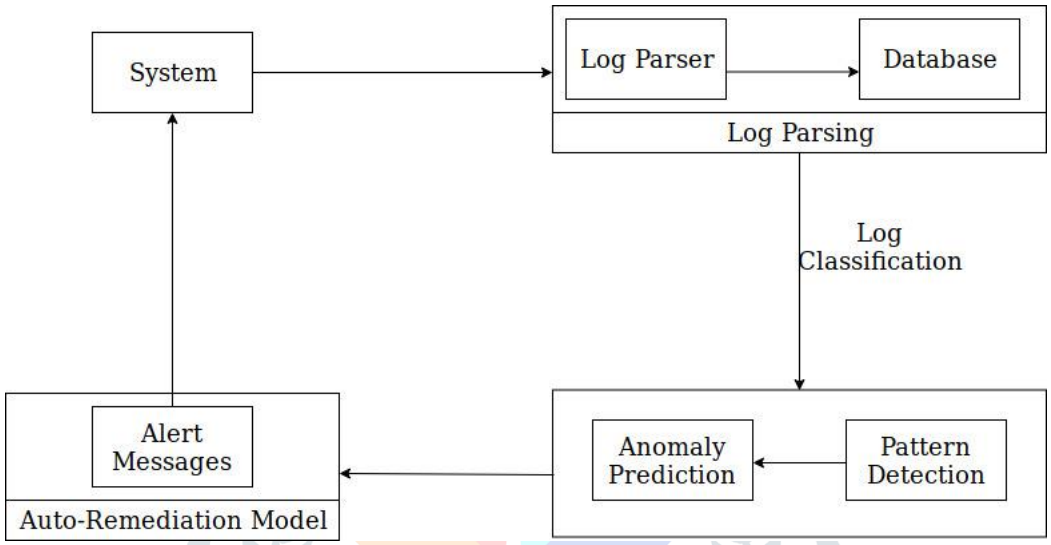

Figure 1: System Architecture

A. Log Parsing: System logs are parsed into their respective fields like timestamp, type, hostname , program name , process id and message using Logstash. These logs are stored onto Elasticsearch in json format which is then used for further log analysis.Elasticsearch is used as database.

B. Pattern Detection and Anomaly Prediction: Logs stored on the Elasticsearch are classified based on their type. Type of the log messages can be error, warning, critical, informational etc . Apriori algorithm is used for detection of frequently occurring patterns. Using these patterns future anomaly detection and prediction is carried out using Long-Short-Term-Memory (LSTM) algorithm which is an advance version of Recurrent Neural Networks.

C. Auto-Remediation: If the probability of occurence of error is more than 50% alert messages are passed to the action-mapping mechanism and remedial measures are taken by the system . This avoids the system crash or the failure due to errors.

IV. ACKNOWLEDGEMENT

We thank Mr. Samir Sood (Harman Connected Service Corp India Pvt. Ltd.) for his support, help and guidance without which this research would not be what it is.

V. REFERENCES

[1] Min Du, Feifei Li, Guineng Zheng, Vivek Srikumar, “DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning” ,2017.

[2] https://www.splunk.com/en_us/solutions/solution-areas/log-management.html

[3] https://logz.io/blog/machine-learning-log-analytics/

[4] S.VijayaKumar, A.S.Kumaresan, U.Jayalakshmi, Frequent Pattern Mining in Web Log Data using Apriori Algorithm, International Journal of Emerging Engineering Research and Technology Volume 3, Issue 10, October 2015.