DeepLog: Anomaly Detection and Diagnosis from System Logs Through Deep Learning

Min Du; Feifei Li; Guineng Zheng; Vivek Srikumar

{mind, lifeifei, guineng, sivvek}@cs.utah.edu

University of Utah

Abstract

Anomaly detection is a critical step towards building a secure and trustworthy system. Automatically generated system log data is an important and valuable resource for understanding system status. We propose DeepLog, which automatically learns log patterns from normal execution, and detects anomalies when log patterns deviate from the trained model. Furthermore, DeepLog constructs workflows for user to diagnose the detected anomaly.

Two-step procedure:

1. Log Key Anomaly Detection
   - Each log entry = log key + parameter value vector
   - Example log key sequence: 25 18 54 57 18 56 ...
   - Use LSTM:
     - Input: recent log keys up to \( m \)
     - Output: conditional probability of next log key given the input recent sequence
   - Training: \( b = 3 \)
   - Detection: Is the actual next log key among the top \( g \) probable predictions?

2. Parameter Value Anomaly Detection
   - Train a separate model for each distinct log key
   - Example:
     - \( k_1 \): Took 0.61 seconds to build instance.\( ^a \)
     - \( k_2 \): Took 1.13 seconds to deallocate network ...\( ^a \)
     - Parameter value vectors overtime: \( [t_2 - t_1, 0.61], [t_2 - t_1, 1.13] \)
   - Multi-variate time series data anomaly detection problem.
   - Use LSTM:
     - A parameter value vector is given as input at each time step.
     - Anomaly detection

Preliminary: Log Parsing

For log entry \( e = \text{"Took 10 seconds to build instance."} \), printed by printf(""Took %f seconds to build instance."", t), the log key \( k = \text{"Took * seconds to build instance."} \)

- log message
- log key parameters
- Took 0.61 seconds to ...
- Took * seconds to ...
- Took 1.13 seconds to ...

Evaluation results on JDFS log data \( ^{a} \):

\( ^a \) PCA (SOSP’99), BM (Uмия/ARC’10), N-gram (baseline language model)

DeepLog Architecture

Log Key Anomaly Detection

Parameter Value Anomaly Detection

Workflow Diagnosis

Summary

DeepLog is a general-purpose framework for online log anomaly detection and diagnosis. DeepLog learns and encodes entire log message including timestamp, log key, and parameter values. The constructed workflows enable effective anomaly diagnosis. DeepLog is able to adapt to new execution patterns by incorporating user feedback for online update/training to its LSTM models. Extensive evaluation on large system logs has clearly demonstrated the effectiveness of DeepLog.