Abstract

Detecting and Clustering Anomalous Log Profiles Using Interpretable Machine Learning Techniques

In this study, we apply a method that employs a cooperative game-theoretic approach to enhance the interpretation of anomaly detection models for log sequences, a well-explored area of research. Specifically, we utilize Shapley values to cluster log sequences and identify anomalous subgroups characterized by similar log templates.

Our approach models anomaly label predictions based on the presence of log templates within a log sequence. We developed the "template count representation", where each log sequence is represented by a vector that counts the occurrences of each log template. Anomaly scores and predictions are then modeled based on this representation. Shapley values are used to interpret the contributions of individual log templates to the model's decisions. In the context of anomaly predictions, templates with high Shapley values are those that significantly influence the model towards classifying a record as an anomaly.

By clustering log sequences based on their Shapley values, we can identify groups where predictions are driven by similar templates. This clustering provides insights into which templates lead to anomalous behavior in specific clusters. Identifying these subgroups has numerous practical applications, as various activities can cause abnormal behavior in large computer systems, ranging from simple component failures to cyberattacks. Each type of event may require different remediation strategies.

We tested our methodology on two widely-referenced datasets, where log lines are generated by different computer systems and anomaly labels were obtained by domain experts. Our experiments demonstrated that transforming to the Shapley space significantly reduced the dimensionality of the original template count representation and our process successfully identified many anomalous clusters where similar templates caused the anomalous behavior. This enables domain experts to link these subgroups to potential causes of abnormal behavior and automate the appropriate response steps to restore system functionality. Furthermore, clustering in the space of Shapley values obtained from a classification model predicting the anomaly labels yielded by the original anomaly detection model proved to be an effective post-processing step, capable of improving detection performance according to commonly used precision metrics.