

Anomaly Detection for Big Data Technologies

Ahmad Alnafessah

Department of Computing, Imperial College London, United Kingdom
a.alfanfessah17@imperial.ac.uk

Giuliano Casale

Department of Computing, Imperial College London, United Kingdom
g.casale@imperial.ac.uk

Abstract

The main goal of this research is to contribute to automated performance anomaly detection for large-scale and complex distributed systems, especially for Big Data applications within cloud computing. The main points that we will investigate are:

- Automated detection of anomalous performance behaviors by finding the relevant performance metrics with which to characterize behavior of systems.
- Performance anomaly localization: To pinpoint the cause of a performance anomaly due to internal or external faults.
- Investigation of the possibility of anomaly prediction. Failure prediction aims to determine the possible occurrences of catastrophic events in the near future and will enable system developers to utilize effective monitoring solutions to guarantee system availability.
- Assessment for the potential of hybrid methods that combine machine learning with traditional methods used in performance for anomaly detection.

The topic of this research proposal will offer me the opportunity to more deeply apply my interest in the field of performance anomaly detection and prediction by investigating and using novel optimization strategies. In addition, this research provides a very interesting case of utilizing the anomaly detection techniques in a large-scale Big Data and cloud computing environment. Among the various Big Data technologies, in-memory processing technology like Apache Spark has become widely adopted by industries as result of its speed, generality, ease of use, and compatibility with other Big Data systems. Although Spark is developing gradually, currently there are still shortages in comprehensive performance analyses that specifically build for Spark and are used to detect performance anomalies. Therefore, this raises my interest in addressing this challenge by investigating new hybrid learning techniques for anomaly detection in large-scale and complex systems, especially for in-memory processing Big Data platforms within cloud computing.

2012 ACM Subject Classification Computing methodologies → Anomaly detection

Keywords and phrases Performance anomalies, Apache Spark, Neural Network, Resilient Distributed Dataset (RDD)

Digital Object Identifier 10.4230/OASISs.ICCSW.2018.8

Category Poster Track



© Ahmad Alnafessah and Giuliano Casale;
licensed under Creative Commons License CC-BY
2018 Imperial College Computing Student Workshop (ICCSW 2018).

Editors: Edoardo Pirovano and Eva Graversen; Article No. 8; pp. 8:1–8:1

OpenAccess Series in Informatics



OASIS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany