ABSTRACT

Data-Intensive Scalable Computing (DISC) systems are essential for managing large datasets with an emphasis on fault tolerance, cost-effectiveness, and user accessibility. However, input errors in processed data pose significant challenges to programmers. This research uses the Snowfall Analysis program, notorious for its anomalous data leading to forecasting inaccuracies, as a case study. To address these challenges, we employ Titian, an extended library that facilitates debugging by systematically tracing the provenance of erroneous data to its origin. Our analysis demonstrated that Titian accurately identified data errors with a precision of 100%, moreover the average of program runtime which implement Titian is only 0.505 seconds across various sizes of dataset, significantly outperforming standard manual debugging methods. These findings highlight Titian's potential to enhance data provenance in DISC systems, offering theoretical insights into debugging processes and practical applications for improving data integrity in large-scale computational environments.

Keywords: Automated-debugging, DISC, Snowfall, Titian