様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ６ 月 １４ 日現在

機関番号：３４４０６
研究種目：若手研究
研究期間：2019～2023
課題番号：１９Ｋ２０２４４

研究課題名（和文）システムログを対象とした「不吉なログ」の検出による障害検知の試み

研究課題名（英文）A system fault detection using bad smell system log

研究代表者
尾花　将輝（Obana, Masaki）

大阪工業大学・情報科学部・講師

研究者番号：００７１００７１

交付決定額（研究期間全体）：（直接経費）　　1,900,000円

研究成果の概要（和文）：アプリケーションから出力されるログから異常動作を検知するため，アプリケーションログの分析フレームワークおよび分析手法を提案した．教師無し学習を用いてアプリケーションから出力される少数のログメッセージを検出する手法を提案し，運用中のアプリケーションから障害が検知できるかについて検証した．
　まず，安定版のオープンソースソフトウェアをサービスとして運用し，運用中のログデータの分析を行った．次に，出力されるログに対し，複数の分類手法を組み合わせたログの分析手法を提案し，疑わしい挙動を示すログメッセージとバグの関係についての調査を行った．結果，提案手法によって異常動作を示すログを検出することができた．

研究成果の学術的意義や社会的意義
昨今の情報システムの重要さは年々大きくなっており，システム障害が発生する事は社会に大きな影響を与えるため，社会への影響を最小限に収めるためにも早期に障害を検知することが重要である．一方で，昨今の情報システムはログデータが莫大な事や，様々なパッケージソフトウェア，クラウドサービスを利用されているためログのフォーマットの違いやあるためログの分析は容易ではない．そこで，アプリケーションごとにログの特徴を学習し，本手法から出力される特徴的なログデータだけを分析する事でログの分析を容易にし，社会に大きな影響を与える前に障害の対策を実施することができる．

研究成果の概要（英文）：We proposed a framework and methods to detect anomalies with application logs. Using unsupervised learning techniques, rare log patterns were detected in the proposed framework. In addition, the various clustering methods help detect the skeptical anomalies in the application log.
 As a result of applying the open-source system in real running, the proposed methods can detect rare log patterns. We found that the detected rare log includes suspicious anomalies in the system running through various clustering techniques.

研究分野：ソフトウェア工学

キーワード：ソフトウェア　アプリケーションログ　障害　異常動作検知　実績的

様　式　C－１９、F－１９－１（共通）

１．研究開始当初の背景

近年のコンピュータシステムでは，パッケージソフトウェアやクラウドシステム等を複雑に組み合わせたシステム開発が主流である．しかし，ソースコードが公開されないパッケージソフトウェアを用いたシステム開発における障害の検知は，ソースコードに頼らない方法で障害を検知する必要がある．一般的にこのようなソフトウェアに対し障害発生時の原因追及方法として膨大なログを分析し，原因箇所を特定した後にソフトウェアの設定ファイルの見直しや開発者にバグの修正依頼を行う．しかし，障害は複数のソフトの組み合わせで発生する場合が多く，障害内容によってはログにエラーメッセージが出力されない場合があり，一見正常に動作した膨大なログから障害原因を調査する必要がある．このような障害の原因となりうるログメッセージ群を開発者の経験だけで発見することは困難でありコストがかかる事が考えられる．

２．研究の目的

本研究の目的はソフトウェアから出力されるログ，特にアプリケーションから出力されるログを対象に障害発生を迅速に検知し，ソフトウェアの運用者や開発者の支援を行う事が目的である．更にアプリケーションはログフォーマットが統一されない傾向にあったり，新しい機能が実装されるサイクルが早いため，他のアプリケーションで得られた知見を反映しにくい傾向にある．そこで，運用中のアプリケーションから出力されるログから障害の可能性があるログを検出する異常動作検出手法を提案する．これにより，障害が発生から障害対応までの時間を短縮することで社会的影響を少しでも小さくする事が社会的な意義となる．

３．研究の方法

本研究は運用中，つまりリリースされたアプリケーションを対象にした異常動作検出を行う．そのため，複数の安定版（最新版）のオープンソースソフトウェアを運用し，その後に報告されるバグを本研究手法で検知できるかが重要となる．そこで，Redmine，Jitsi Videobridge，OpenNebula といった様々なオープンソースソフトウェアを運用し，ログデータの収集を行った．また，図１に示すソースコードからログメッセージのパターンを抽出し，正常動作が記録されたログの中に障害を示すログメッセージが含まれるかを分析した．

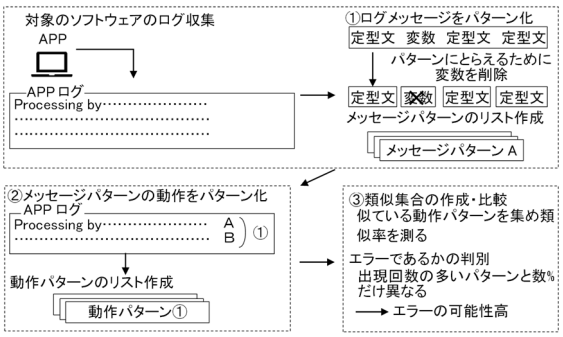

図１．ソースコードからログメッセージパターンの抽出手法

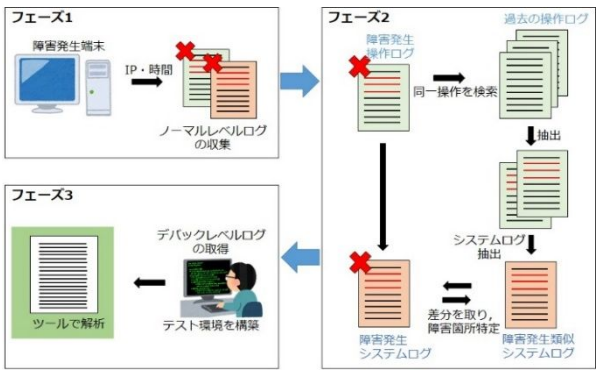

図２．アプリケーションログを用いた障害特定フレームワークいの提案

次に，正常に動作したログメッセージに比べ，特徴のあるログメッセージに障害が含まれているかどうかについて分析を行った．障害原因を追究するためのフレームワークを提案した．アプリケーションログから類似操作または同一操作を抽出し，これらの操作がシステムログとの類似度を求めることで障害を検知する手法である（図２参照）．これにより，アプリケーションログからでも障害が検出できるかどうかについての分析を行った．さらに，error や fault 等のログメッセージで出力されないアプリケーションやシステムログからでも障害を検知できることがわかった．

最後に，アプリケーションログから障害の可能性のあるログメッセージを抽出する手法の提案を行った．過去の実績より障害を示すログメッセージは少数であることから，教師無し学習を用いることで少数のログメッセージを抽出する異常動作検知手法を提案した（図３参照）．これにより，膨大なログを分析せずとも，本手法によって一部のログメッセージのみをレビューすることにより，社会的影響のある障害をより早く検知，対応することができる．

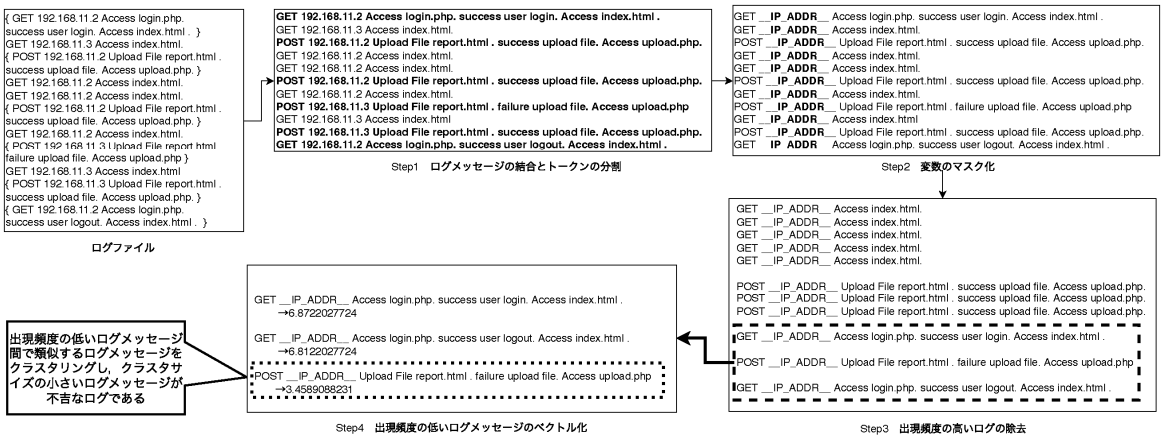

図３．アプリケーションログを用いた異常動作検知手法の提案

４．研究成果

アプリケーションログと障害の関係を調査し，調査結果よりアプリケーションログを対象とした異常動作検知の手法を提案した．これらの手法とサービスとして展開したオープンソースソフトウェアのログデータに適用した結果が表１である．結果，手法により莫大な量であったログから少数のログメッセージ数に絞り込むことができた．また，リリース後に報告や修正されたバグが抽出ログメッセージの中に現象として記録されている事も確認できた．これらにより，莫大なログメッセージを調査せずとも，一部のログメッセージを精査することで障害が検知できる可能性があることがわかった．

表１．手法による抽出したログメッセージ数と，バグとの関係

アプリケーション	ログ行数	リリース後に報告，修正されたバグ件数	手法で抽出したログメッセージ数	抽出したログメッセージにバグが含まれていた件数
Redmine	38,430	2	35	2
Jitsi Videobridge	3,664,285	2	14	2
OpenNebula	24,031,912	3	45	3

また，本手法をまとめてツールへと実装したのが図４である．アプリケーションログから出力されるログを入力とし，障害に関連する可能性のあるログメッセージを一覧で表示するものである．また，出力されるログメッセージの比率が視覚的にわかる，さらに，他の手法の実装を行い精度の比較実験も行った．結果，アプリケーションログに対しは従来の手法よりも本手法の方が高い精度である事も示す事ができた．

図４．本手法を実装したログ分析ツール

５．主な発表論文等

〔雑誌論文〕　計3件（うち査読付論文　3件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 尾花 将輝、花川 典子	４．巻 38
２．論文標題 システムログとアプリケーションログを用いた障害起因特定のためのフレームワーク提案	５．発行年 2021年
３．雑誌名 コンピュータ ソフトウェア	６．最初と最後の頁 3_58～3_74
掲載論文のDOI（デジタルオブジェクト識別子） 10.11309/jssst.38.3_58	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 上田 晃義 ， 尾花 将輝 ， 花川 典子	４．巻 29
２．論文標題 定形的ログメッセージの除去とクラスタリングによる異常動作ログの検出方法の提案	５．発行年 2022年
３．雑誌名 ソフトウェア工学の基礎 29	６．最初と最後の頁 113～122
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 上田 晃義 ， 尾花 将輝 ， 花川 典子	４．巻 30
２．論文標題 自然言語処理とクラスタリングを用いたログレビュー支援のためのアプリケーションログ分析ツールの開発	５．発行年 2023年
３．雑誌名 ソフトウェア工学の基礎 30	６．最初と最後の頁 201～202
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計2件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 上田晃義，吉兼史崇，尾花将輝
２．発表標題 自然言語処理とクラスタリングを用いて非構造なログの分析を支援するため手法の提案
３．学会等名 第28回　ソフトウェア工学の基礎ワークショップ
４．発表年 2021年～2022年

１．発表者名 山田 百華，林原 成那，尾花 将輝
２．発表標題 アプリケーションログの解析による障害検知の試み
３．学会等名 第27回 ソフトウェア工学の基礎ワークショップ
４．発表年 2020年～2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関