6ZC-06

npmパッケージにおける脆弱性修正期間とリポジトリ情報の分析

曽我　恵太†　　佐藤　将也‡

岡山県立大学大学院情報系工学研究科†　　岡山県立大学情報工学部‡

1. はじめに

オープンソースソフトウェア(以下，OSS) は様々なソフトウェア開発で使われている． OSSの脆弱性について，種類や数は研究されているものの，修正に要する期間は充分に考慮されていない．OSS を継続して開発に利用するには，脆弱性が活発に修正されていることが望ましい．そこで本研究では，脆弱性の発見から修正までの期間を調査し，OSS のリポジトリ情報との相関を分析する．これにより，リポジトリ情報が OSSを選択する判断材料となるか npm を対象に調査した結果を示す．

2. OSSにおける脆弱性の修正

2.1 脆弱性の修正手順

OSS とは再配布の自由やソースコードの無償配布などを含むライセンスを持つソフトウェアである． OSS は複数の開発者によって GitHub を用いて開発されることが多い． GitHub はバージョン管理システムである Git の仕組みを利用したソフトウェア開発のプラットフォームである．本研究では GitHub を用いて開発されている OSSを調査分析の対象とする． GitHub では Issuesを用いて機能追加や脆弱性について議論が行われる．開発者は脆弱性の通知を受け取ると，Issues 等を用い，修正について議論する．修正はコミットとして作成する．コミットを適用することにより脆弱性を修正する．

2.2 脆弱性修正期間

パッケージの修正が活発に行われているか判断するための指標として脆弱性の修正期間に注目した．脆弱性の修正期間が短いほど当該 OSSは脆弱性の修正を重視していると言える．本稿では，脆弱性の修正期間を次のように定義した．National Vulnerability Database(以下，NVD)における報告日を脆弱性の発見日，脆弱性に対応する Issue が登録された日を Issue 開始日，修正のコミットが適用された日をコミット日とする．また，脆弱性の発見日からコミット日までを全期間とした．ここで，発見日から Issue開始日，Issue 開始日からコミット日，および全期間の 3 つの期間に着目して分析を行う．

2.3 リポジトリ情報

本研究では GitHub における Star 数，Contributors 数，リポジトリに含まれるファイル数，コード行数，およびコメント行数をリポジトリ情報とする．これらの情報と脆弱性修正期間との関連を分析することで，パッケージのメタデータであるリポジトリ情報からパッケージの開発が活発に行われているか判断することを目的とする．Chinthanet らは脆弱性修正に関連する情報として GitHub の Issue やコミットなどを抽出した[1]．Issue やコミットの情報は脆弱性修正に関して分析するために有用である

3. 調査

3.1 方法

npm パッケージを対象に脆弱性の修正期間とリポジトリ情報の相関を調査した．npm は Node.jsのパッケージ管理システムである．調査対象のnpm パッケージは，GitHub のリポジトリにてStar 数が 1,000 を超えるものから無作為に 50 個を選択した．

調査対象のパッケージについて Google Open Source Insights(以下，OSI)[2]から情報を取得する．取得する情報は依存関係のパッケージの数，Security Advisories から脆弱性の有無，および脆弱性の内容である．この中で脆弱性の報告があったパッケージを問題パッケージとする．OSI の Security Advisories には対象パッケージの依存関係のパッケージの脆弱性についても記載がある．問題パッケージには依存関係のパッケージも含む．問題パッケージのリポジトリ情報として Star 数，Contributors 数，Common Vulnerability Scoring System(以下，CVSS)スコア，ファイル数，リポジトリ内のコード行数とコメント行数を取得した．さらに，問題パッケージについて脆弱性修正期間の情報を取得す

Analysis of time duration for fixing vulnerabilities and repository information in npm packages

Keita Soga†, Masaya Sato‡, †Graduate School of Computer Science and Systems Engineering, Okayama Prefectural University, ‡Faculty of Computer Science and Systems Engineering Okayama Prefectural University

表 1　リポジトリ情報と脆弱性修正期間の相関係数

	発見から Issue	Issue からコミット	全期間
Star	-0.214	-0.339	-0.312
Contributors	-0.188	-0.381	-0.318
CVSS スコア	-0.492	0.192	-0.191
ファイル数	-0.211	-0.376	-0.330
コード行	-0.444	-0.490	-0.531
コメント行	-0.341	0.258	-0.065

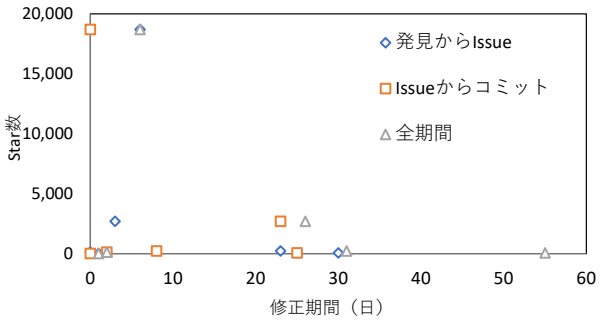

図 1　Star 数と脆弱性修正期間の散布図

る．問題パッケージについてのリポジトリ情報と脆弱性修正期間の情報を利用して相関係数を計算し，分析する．

3.2 結果

調査対象のパッケージ 50 個から取得した問題パッケージは 6 個であった．問題パッケージのリポジトリ情報と当該脆弱性修正期間の情報との相関係数を表 1 に示す．

4. 考察

4.1 相関分析

Star 数と Contributors 数は弱い負の相関がある．負の相関は修正までの期間が長いほどリポジトリ情報の値が小さいという傾向を示す．CVSS スコアは発見日から Issue 開始日と負の相関がある．Issue 開始日からコミット日と全期間では相関係数の絶対値は小さい．ファイル数は，Star 数と Contributors 数と近い相関がある．コード行数は，脆弱性修正期間の全ての項目と負の相関がある．これは，コード行数が多いリポジトリは，脆弱性修正期間が短い傾向にあることを示す．コメント行数は，相関係数が最大でも-0.3 前後であり，相関は弱い．

4.2 外れ値の考察

図 1 に問題パッケージにおける Star 数と脆弱性修正期間の散布図を示す．図 1 より，弱い負の相関が見られるが，Star 数 20,000 付近の点が他の点より大きく離れている．これが相関係数に大きく影響した可能性がある．Contributors 数，ファイル数，およびコード行数でも同様の傾向があった．そこで，外れ値の点を除外し，相関係数を再度計算した．結果を表 2 に示す．

表 2　外れ値を除外した相関係数

	発見から Issue	Issue からコミット	全期間
Star	-0.315	0.538	0.084
Contributors	-0.268	0.577	0.134
CVSS スコア	-0.698	-0.034	-0.454
ファイル数	-0.205	0.140	-0.055
コード行数	-0.549	-0.292	-0.496
コメント行数	-0.298	0.587	0.120

表 2 より，Star 数と Contributors 数について，Issue 開始日からコミット日との相関係数が 0.5 程度であり，表 1 から大きく変化した．CVSS スコアでは，発見日から Issue 開始日は 0.7 となっており，強い負の相関がある．ファイル数では，相関係数の変化は小さいが，相関係数の絶対値はより小さくなり，ほとんど相関がない．コード行数は，表 1 と大きく変わらず負の相関がある．コメント行数は，特に Issue 開始日からコミット日との相関が大きくなり，正の相関がある．

以上のことから，CVSS スコアが高いほど発見から Issue 作成までの期間が短い傾向にあることが分かる．また，Star 数，Contributors 数，およびコメント行数は値が大きいほど修正までの期間が長い傾向にあることが分かる．

5. おわりに

npm パッケージを対象にリポジトリ情報と脆弱性修正期間について相関分析を行った結果を述べた．分析結果より，コード行数が脆弱性修正期間と負の相関を持つことを示した．課題として，脆弱性への対処内容の詳細な分析，およびより多くのパッケージを対象とした分析がある．

参考文献

[1] Chinthanet, B., Kula, R.G., McIntosh, S., et al.: Lags in the release, adoption, and propagation of npm vulnerability fixes, Empirical Software Engineering, Vol.26, pp.1-28 (2021).
[2] Google LLC.: Open Source Insights, （オンライン），入手先〈https://deps.dev/〉，（参照 2021-11-01）．