

Online Testing of RESTful APIs: Promises and Challenges

Alberto Martin-Lopez, Sergio Segura, Antonio Ruiz-Cortés

ESEC/FSE 2022 Research Papers“Online Testing of RESTful APIs:Promises and Challenges”

発表者: 田中 智達

背景

- APIが正しく動作しているかどうかは非常に重要
 - APIに1つでもバグがあると、それを利用している他の何十、何百ものサービスに影響を与える可能性.
- APIをテストする業界標準のツールやライブラリはテストの実行を自動化されている.
- 産業界ではAPIを生産中に継続的にテストするオンラインテストのモデルへと移行.
 - テストケースは依然として手作業で実装する必要がある.

先行研究

- RESTful APIのテストケースの自動生成に関する研究が近年盛ん.
- ほとんどのアプローチが以下のラボ環境下で評価.
 - 少数のオープンソースAPI [55, 68, 87]
 - 限られたテストケース数 (最大数千) [70, 72, 90]
 - 限られた時間 (最大数時間) [57, 70, 78]
 - 単一の組織内で使用 [58, 78]

目的とアプローチ

●目的

- オンラインシナリオ下における，さまざまな最新鋭のテストケース生成技術の長所と限界を評価することで，これまでの研究の補完を行う．

●アプローチ

- 世界中の数百万人のユーザが利用する13の産業用APIについて、15日間ノンストップでテストケースを自動生成し，実行．

オンラインテスト

- ソフトウェアがリリースされる前に開発環境で行われるオフラインテストとは対照的に、実稼働中のソフトウェアに対して行われるテストのこと。 [60, 63]
- オンラインテストには2種類存在する。
 - 受動的な監視活動[60, 61]
 - 矛盾や異常の検出を目的とし、実世界の使用状況下でシステムを観察。 [59, 83]
 - 能動的なテストプロセス[65, 80]
 - 以前のテスト結果から得られたフィードバックに基づき、新しいテストケースがシステムに継続的な刺激を与える。 [65, 100]
 - 本論文ではこちらの定義を採用。

RQ

- RQ1 : RESTful APIのオンラインテストの障害検出能力はどの程度か
 - 経時的に検出される失敗数や種類を調査することが目的
- RQ2 : RESTful APIのオンラインテストの障害検出能力はどの程度か
 - 発見されたすべての障害のうち、時間の経過とともに検出された固有の障害の数と種類を調べることを目的
- RQ3 : 異なるテスト間で観察される違いは何か
 - APIカバレッジと障害検出能力の観点から、異なるテスト戦略を比較し、異なるアプローチがどのようにお互いを補完するかを理解することが目的

テスト対象のAPI

- 13のAPIから189の操作と1,694の入力パラメータを取得.
- 先行研究[54, 87, 90, 93]にてテストした大規模で複雑な実世界のAPI から構成.
 - 世界中の数百万人のユーザが利用する商用API
 - Spotify, Yelp
 - 産業規模のオープンソースAPI
 - Ohsome, LanguageTool
 - 公的機関 (アメリカ政府) が開発した複雑なAPIの代表
 - FDIC (連邦預金保険公社)
- 一般的な商用APIでは読み取り専用だが、書き込み専用もしくはその双方の機能を持つAPIもテスト.

テスト対象のAPI

API	カテゴリー	テスト中のサービス	O	P	R	W
Amadeus[2]	Travel	Hotel search	1	27	X	
DHL[23]	Shipping	Service point search	1	9	X	
FDIC[24]	Banking	all	6	73	X	
Foursquare[26]	Social, mapping	Venues search	1	17	X	
LanguageTool[27]	Languages	Text proofread	1	11		X
Marvel[28]	Entertainment	Characters	6	72	X	
Ohsome[31]	Mapping	all	112	1,146	X	
OMDb[32]	Media	all	1	9	X	
RESTcountries[42]	Reference	all	21	42	X	
Spotify[44]	Media	Playlists	12	48	X	X
Stripe[45]	Financial	Products	5	47	X	X
Yelp[48]	Recommendations	Business search	1	14	X	
YouTube[49]	Media	Comments	10	148	X	X
YouTube[51]	Media	Search	1	31	X	
Total			189	1,694	13	4

図: テスト中のRESTful API

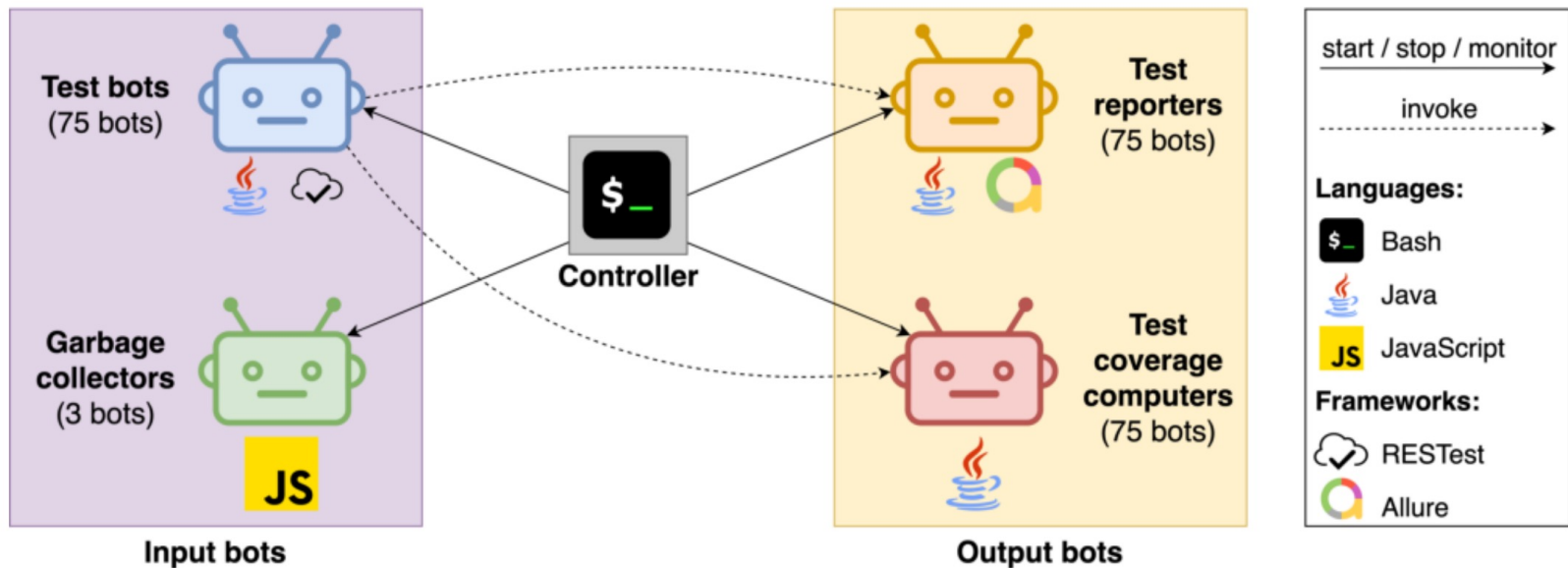
(O=操作, P=パラメータ, R=読み取り, W=書き込み)

実験概要

- テストケースの生成と実行にはRESTest[91]を使用.
 - 多様なテストケース生成ストラテジーを統合したオープンソースフレームワーク.
 - しかし予備調査において、テストケースの生成、実行、レポートを同じプロセスで行うのは希望する規模でのアプリケーションの開発と展開を妨げることになることが判明.
- テストプロセスの主要な部分をbotと呼ばれる凝縮性の高い自律的なプログラムに分離したものを考案.

テスト環境の構造

- bot部は以下の2種類に分類。
 - Input bots : テストケースの生成と実行を行う.
 - Output bots : テスト出力の解析と活用を行う.
- botはControllerコンポーネントにより自動で起動, 停止, 監視される.



RQ1 (オンラインテストの故障検出能力)

- 合計389,216個の失敗が発見.
- 失敗の63%は、APIのレスポンスと仕様の不一致が原因.
- すべてのAPIで失敗が明らかになり、故障検出率(FDR)は0.04から0.79の範囲である.
- FDRの進化によりDHL, Ohsome, Yelpで一時的な停止やAPIの使用禁止といった異常が確認.

RQ2 (オフラインテストの故障検出能力)

- 11のAPIで147件の障害が発見.
- 約半数の73件はAPIのレスポンスと仕様の不一致が原因.
- 約3分の1 (52件) は無効なAPI入力を有効なものとして処理したことが原因.
- 最初の3日間で87% (128件) の不具合, 10日間で全ての不具合を発見.

RQ3 (異なるテスト間の比較)

- 有効なテストデータ生成と無効なテストデータ生成は同程度の数の固有のバグを発見しており，明らかに相補的.
- 全てのテストデータ生成手法は，何らかのAPI において固有のバグを検出したり，カバレッジを向上させることに寄与.
 - ただしセマンティックデータだけは例外.
- 制約条件付きテストがテストケースの生成において最も効果的であり，すべてのテストケースが欠陥の検出やカバレッジの向上に寄与.
- APIのパラメータや入力値に関するテストケースの多様性は，新たなバグを発見したり，APIのカバレッジを向上させることには役立たない.

課題

- RESTful APIのオンラインテストにテストケース生成技術を採用する上での課題として以下のものが挙げられる.
 - 不具合発生の原因の自動特定
 - botと人間との効果的な相互作用
 - 最適なテスト手法の決定
 - 最適なテスト実行スケジュール
 - 計算機資源の最適化

結論

- RESTful APIのオンラインテストが産業界で次に必要とされる機能である可能性を提示.
- 実際に完全に採用されるために克服すべき重要な課題のいくつかを提示.