

Shipwright: A Human-in-the-Loop System for Dockerfile Repair

Jordan Henkel, Denini Silva, Leopoldo Teixeira,
Marcelo d' Amorim, Thomas Reps

2021 International Conference on Software Engineering (ICSE)

発表者：辻 裕真

背景と目的

- Dockerは最も広く利用されている仮想化ツールの1つ
- Dockerfileは4分の1以上がビルドに失敗
- Dockerfileのビルドエラーを検出, 修正する研究は行われていない
- Dockerfileを自動修正するShipwrightを提案
- Shipwrightにより開発者を支援することが目的

ビルド失敗の原因

- ・ 外部ファイルの問題

ベースイメージやURLなどが変更された場合

- ・ バージョンの依存関係

ベースイメージのバージョンとリポジトリ内のファイルに記述しているバージョンが異なる場合

- ・ コマンドが存在しない

別途インストールしなければコマンドを使用できない場合

- ・ プロジェクト固有の問題

データセット

- binnacle

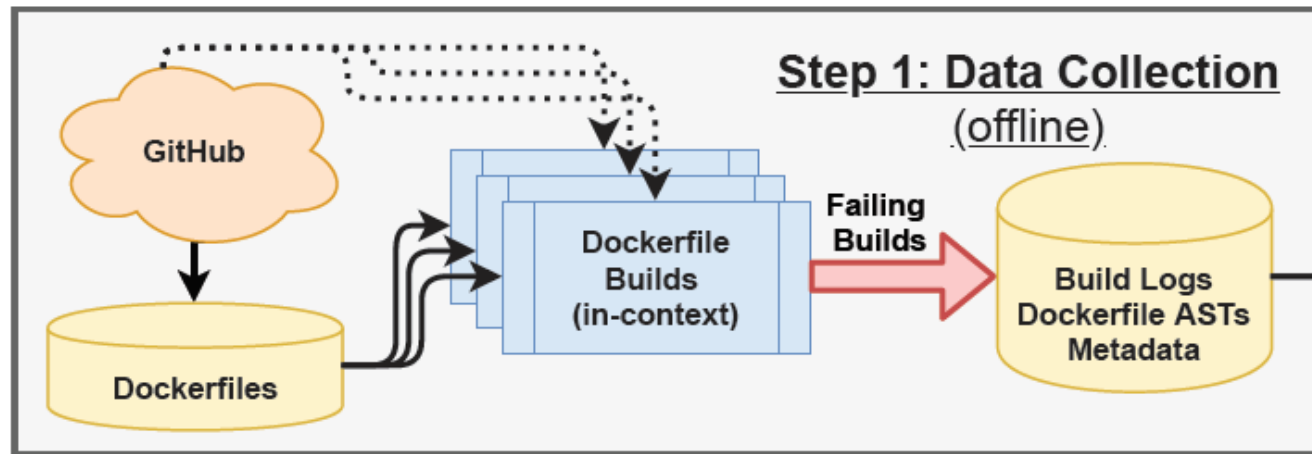
Dockerfileが含まれているGitHub上のリポジトリのデータセット

- i) リポジトリ内にDockerfileが1つのみ存在する
- ii) Dockerfileがリポジトリのルートディレクトリに存在する

32,466個のリポジトリのDockerfileが抽出された

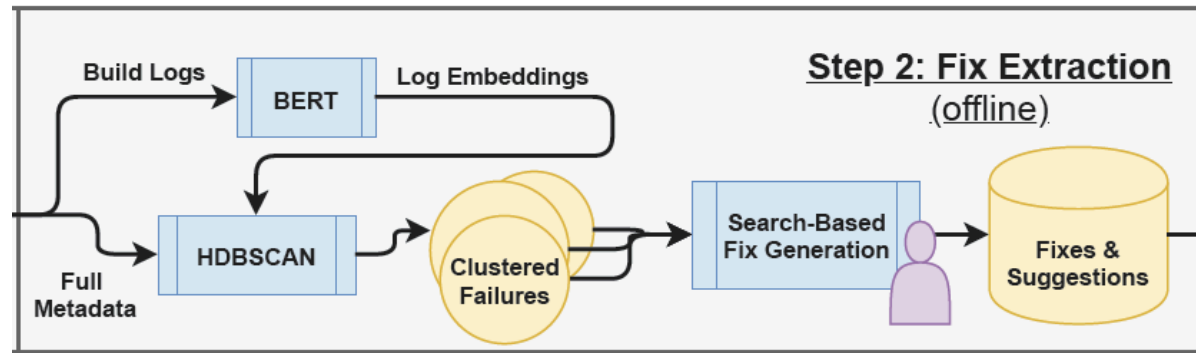
時間とリソースの都合上20,526個のDockerfileをビルドした

Shipwright

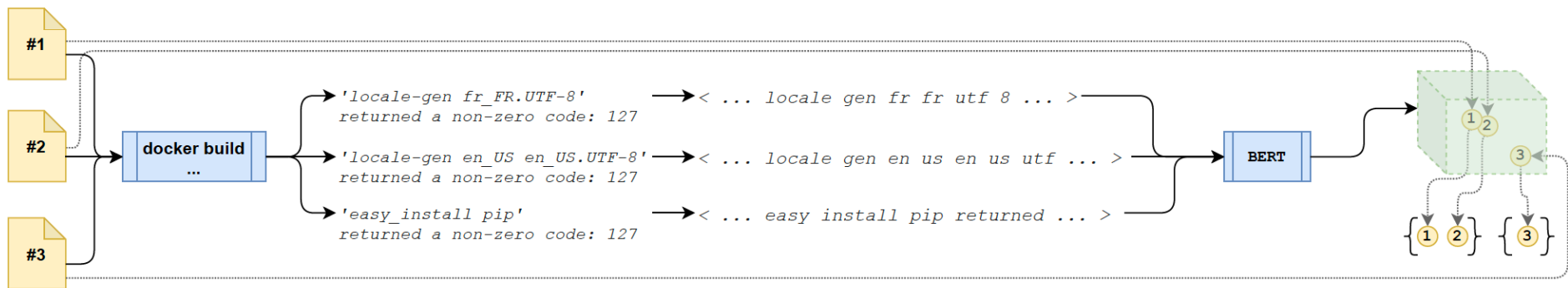


- リポジトリをクローンし、Dockerfileをビルド
- ビルドが失敗した際のログとメタデータをデータベースに保存

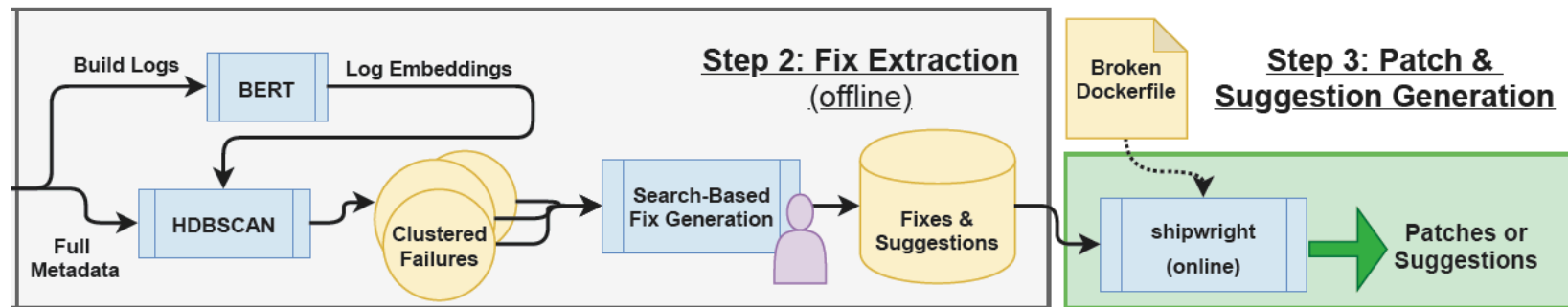
Shipwright



- ビルドログの最後の数行にBERT (自然言語処理モデル)を適用
- HDBSCANを用いてDockerfileのデータとBERTの結果をクラスタリング



Shipwright



- クラスタリングされたビルド失敗の原因から人間が修正を一般化し，13の修正パターンを作成
- 修正できない問題の場合，修正の提案を行う（50の提案を作成）
- ビルドが失敗するDockerfileに対して，必要な修正を適用

Shipwright

修正のパターン

TABLE I: Selected Repairs.

Id	Pattern	Repair	Src.
1	$\text{"ERROR: Error installing bundler"} \in \log \wedge \text{"bundler requires Ruby version } \geq ([0-9]+.[0-9]+.[0-9]+)" \in \log$	<u>replace</u> FROM ruby:(.*) <u>with</u> FROM ruby:\$0	[19]
2	$\text{"Rpmdb checksum is invalid: dCDPT"} \in \log$	<u>add</u> RUN yum install -y yum-plugin-ovl <u>after</u> FROM(*)	[20]
3	$\text{"E: Some index files failed to download. They have been ignored, or old ones used instead"} \in \log$	<u>replace</u> base image <u>with</u> latest release from hub.docker.com	[21]
4	$\text{"E: Package 'libpng12-dev' has no installation candidate"} \in \log$	<u>replace</u> libpng-12dev <u>with</u> libpng-dev	[22]
5	$\text{"FROM ubuntu(elt:latest:20.04)" } \wedge \text{"Unable to locate package (.*)"} \in \log \in \text{Dockerfile}$	<u>replace</u> \$0 <u>with</u> :18.04, ...	[23]
6	$\text{"but your Gemfile specified ([0-9\\.\.]+)" } \in \log \wedge \text{"FROM ruby(.*)"} \in \text{Dockerfile}$	<u>replace</u> FROM ruby:(.*) <u>with</u> FROM ruby:\$0	[24]
7	$\text{"invalid byte sequence in US-ASCII"} \in \log \wedge \text{"FROM ruby(.*)"} \in \text{Dockerfile}$	<u>add</u> ENV LANG C.UTF-8 <u>after</u> FROM(*)	[25]
8	$\text{"sh: (.*): not found"} \in \log$	<u>if</u> "FROM alpine(*)" $\in \text{Dockerfile}$ <u>then</u> <u>add</u> RUN apk add -no-cache \$0. <u>else</u> <u>add</u> RUN apt-get -y update && apt-get -y install \$0	[26]
9	$\text{"ERROR: unsatisfiable constraints: bzip (missing):"} \in \log \wedge \text{"FROM alpine(.*)"} \in \text{Dockerfile}$	<u>remove</u> bzip <u>in</u> "apk add" command	[27]
10	$\text{"conda: not found"} \in \log \wedge \text{"RUN curl (https://repo.continuum.(.*))"} \in \text{Dockerfile}$	<u>add</u> -L <u>before</u> \$0	[28]

TABLE II: Selected Suggestions.

Id	Pattern	Suggestion
1	$\text{"mix"} \in \log \wedge \text{"Code.LoadError"} \in \log$	Problem running mix on the Elixir project...
2	$\text{"tsc"} \in \log \wedge \text{"error TS"} \in \log$	Error during TypeScript compilation with tsc. Please check your .ts files.
3	$\text{"wget: server returned error"} \in \log \vee \text{"wget: unable to resolve host address"} \in \log$	wget error, you have a broken URL. Please check the log.
4	$\text{"npm ERR"} \in \log$	NPM Error, check your files, in particular, "npm install" commands.
5	$\text{"curl:([0-9]+).(*)"} \in \log$	curl error, you have a broken URL. Please check the log.

評価

クラスタ化

- ・ ビルドに失敗したDockerfile (5,405)の内, 34% (1,814)をクラスタ化できた (144個のクラスタ)
- ・ 全てのクラスタの内36.5%が同じ原因でビルドに失敗していた

評価

Shipwright

- 1,814個のクラスタ化されたDockerfileの内, 20.34%に修正が適用され, 69.63%に提案が行われた
- 3,591のクラスタ化されなかったDockerfileの内, 18.18%に修正が適用され, 64.81%に提案が行われた
- 修正されていないDockerfileに対してShipwrightを使用してプルリクエストを送信した結果, 45のうち19のプルリクエストが受け入れられた

まとめ

- Dockerfileの自動修正ツールであるShipwrightを提案した
- Shipwrightによる修正のプルリクエストの42%が受け入れられた

所感

- ・ やろうとしていたDockerfileの自動修正が行われていた…
- ・ Shipwrightによって修正されたDockerfileを再度ビルドして成功するかどうか気がになった.