

技術的負債の 説明文生成手法の構築

オープンソースソフトウェア工学研究室

学籍番号：S2420102

氏名：野口 優輝

技術的負債 (Technical Debt)

ソフトウェア開発において、納期やコストの制限によって導入される質の低い設計やソースコード[?]

例) スパゲッティコード：構造が複雑で理解が難しいソースコード
コードスメル：深刻な問題を抱えているソースコードの特徴

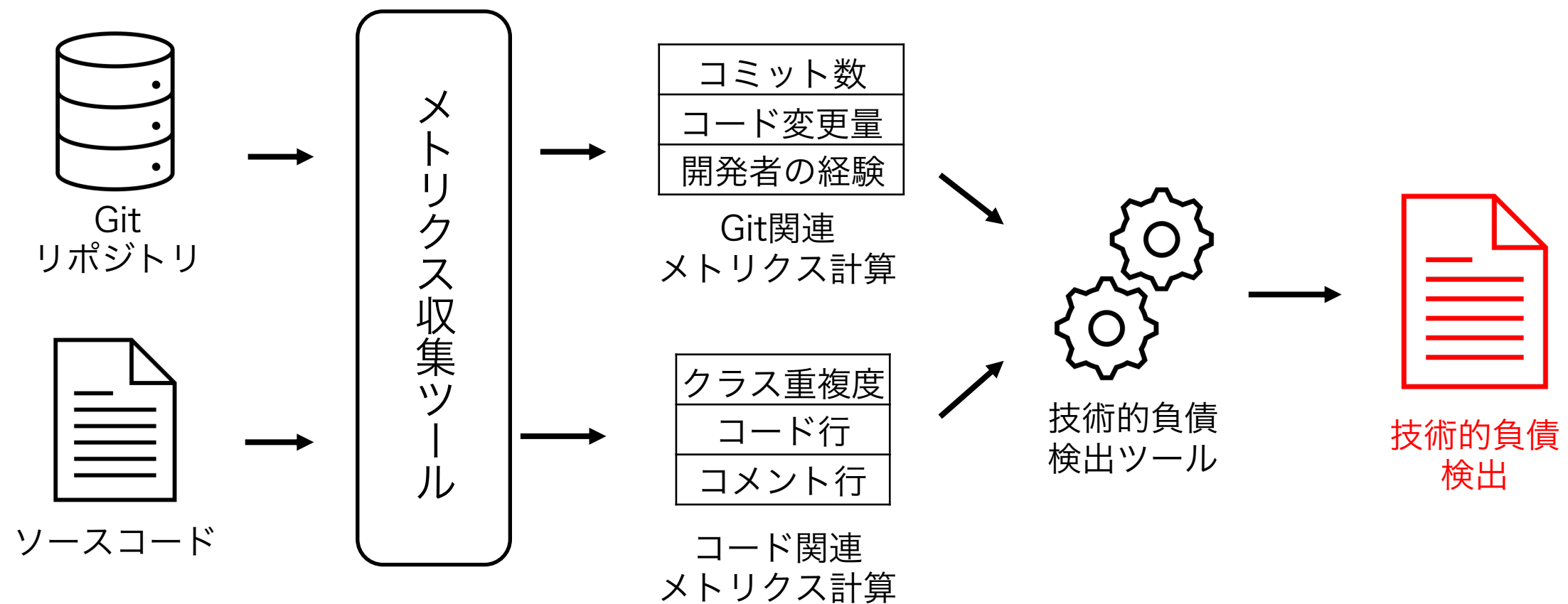
技術的負債が蓄積していくと . . .

- ソフトウェアの保守性の低下, 機能欠陥を引き起こす
→ 除去するための保守コストが増大

技術的負債は早期に除去されることが望まれる

機械学習を使ったTDの検出

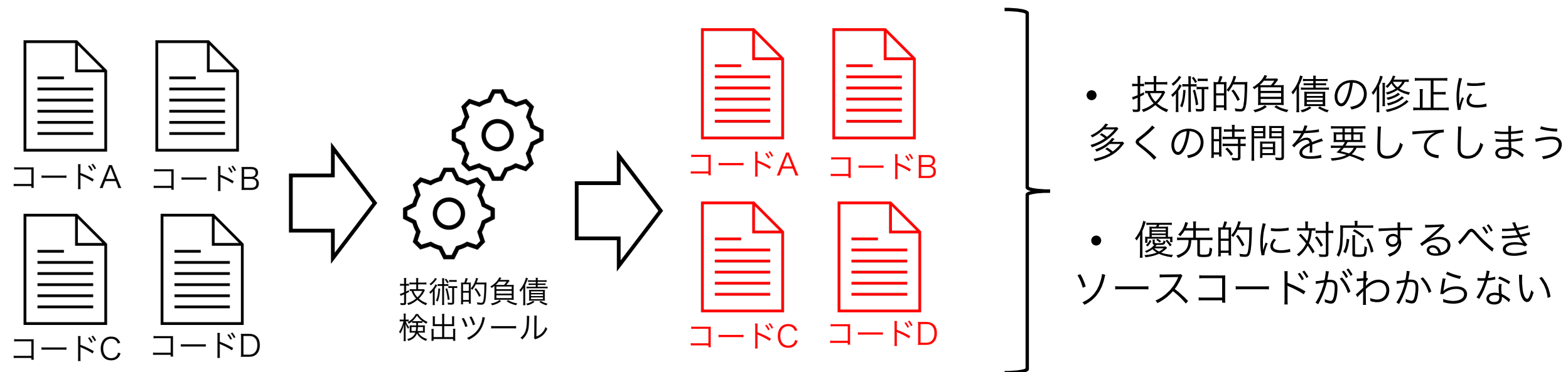
機械学習を使って、ソフトウェアクラスに含まれている技術的負債を検出[?]



クラス単位やメソッド単位で技術的負債を検出することが可能

機械学習を使ったTDの検出：課題点

ソースコード中の技術的負債を検出するだけで、開発者が理解するのは難しい



技術的負債についての説明があると

- 技術的負債を理解し、プロジェクトの健全性が向上
- 技術的負債を除去する上での無駄なコストや遅延を防ぐ

⇒ 技術的負債についての説明文が求められる

課題管理システムを用いた報告

TD Issue

- 技術的負債のラベルが割り当てられた課題票 (Issue)

Issue

不具合や機能欠陥などのさまざまな課題が管理されている

tests: ensure "expected" is first value, "actual" is second value #101

New issue

Closed

ndushay opened this issue on Jun 1, 2017 · 0 comments · Fixed by #108



ndushay commented on Jun 1, 2017

Contributor

Know this is wrong for DownloaderSettings tests; perhaps for others as well.



ndushay added tech debt in progress labels on Jun 1, 2017

ndushay self-assigned this on Jun 2, 2017

ndushay mentioned this issue on Jun 2, 2017

Settings updates #108

ndushay added this to the 2017-06-02 automate crawls sprint 5 milestone on Jun 2, 2017

tingulfen closed this as completed in #108 on Jun 3, 2017

Assignees

ndushay

Labels

tech debt

Projects

None yet

Milestone

2017-06-02 automate crawls sprint 5

Merged

Development

Successfully merging a pull request may close this issue.

Settings updates

sul-dlss/wasapi-downloader

Labels

tech debt

本研究の着眼点

着眼点

- TD Issueの本文は技術的負債について示されている

Settings updates #108

Merged tingulfinsen merged 6 commits into master from settings-updates on Jun 3, 2017

本文 Commits 6 Checks 0 Files changed 18

ndushay commented on Jun 2, 2017 Contributor

user facing:

- add setting for "retries" - how many times to retry a download for each file (retries + 1 = total tries) (Connects to [setting to set the number of retries per file download #93](#))
- params rename to crawlId and crawlIdLowerBound (was jobId) for consistency with wasapi terminology (Resolves [rename "jobId" in command line to "crawlId" ? #98](#))

tech debt:

- fix assertEquals so expected and actual args in right position (resolves [tests: ensure "expected" is first value, "actual" is second value #101](#))
- max line length of 120 for better readability (don't worry about it for tests)

本文内容

技術的負債の簡単な説明

発生している状況や環境など

- 検出された技術的負債を理解するために、TD issueの本文を自動生成することでより良い開発者支援になるのでは？
 - 技術的負債が明文化され、適切に管理することができる

本研究の目的とアプローチ

目的

- 検出された技術的負債の説明文を自動生成するモデルの構築

アプローチ（仮）

1. 事前学習済みモデル（FTモデル）を使用

- 事前学習済みモデルを使い，Issue本文と対応するソースコードを学習
- ソースコードを入力に，Issue本文を生成

2. プロンプト学習を使用

- Issue本文と対応するソースコードからプロンプトを生成
- プロンプトを使って，Issue本文を生成

アプローチ概要（FTモデル）

使用するFTモデル

- CodeBERT, CodeT5, GraphCodeBERT
→ ソースコードと自然言語で学習されている

ファインチューニングさせるもの

- Issue本文+ソースコード

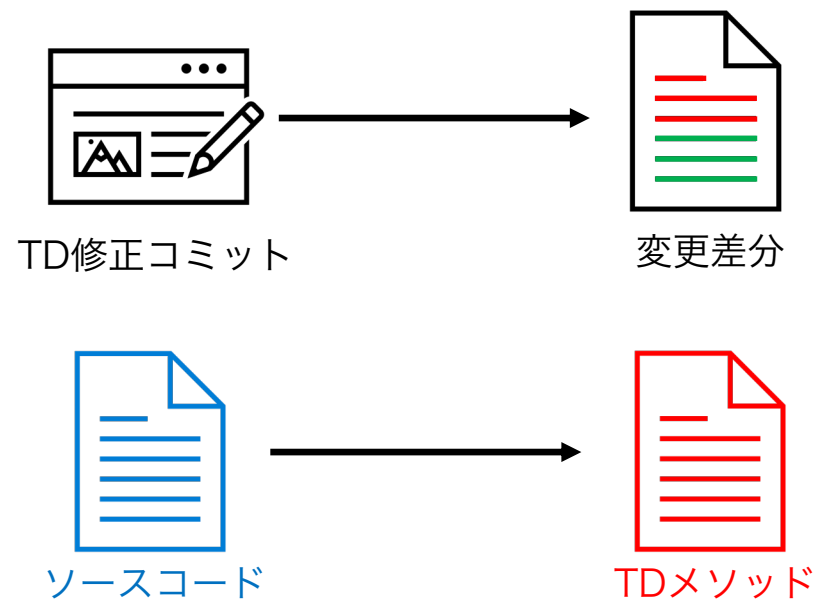
Pattern1. ソースコードの変更部分（diff）

Pattern2. TDを含むソースコード中のメソッド

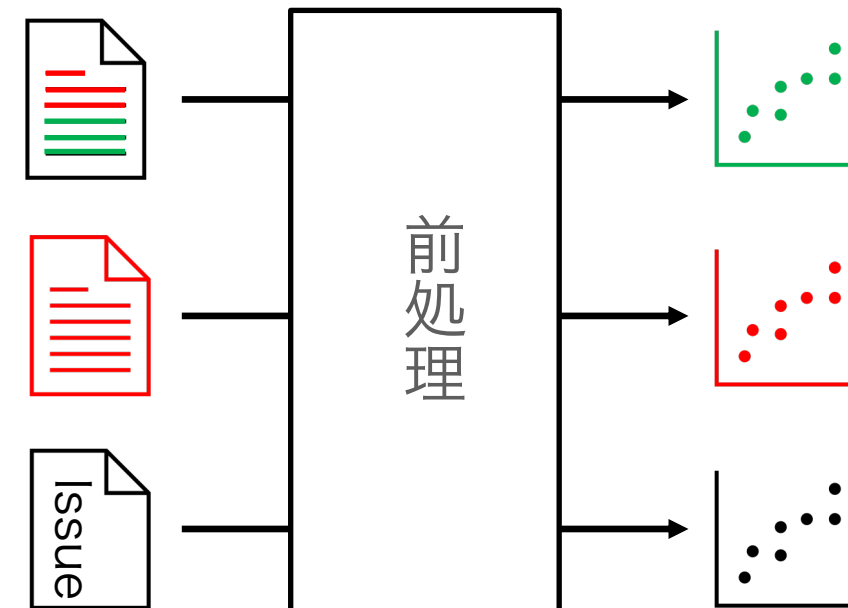
⇒ ソースコードをどの形でFTさせるかは検討中…

イメージ図 (FTモデル)

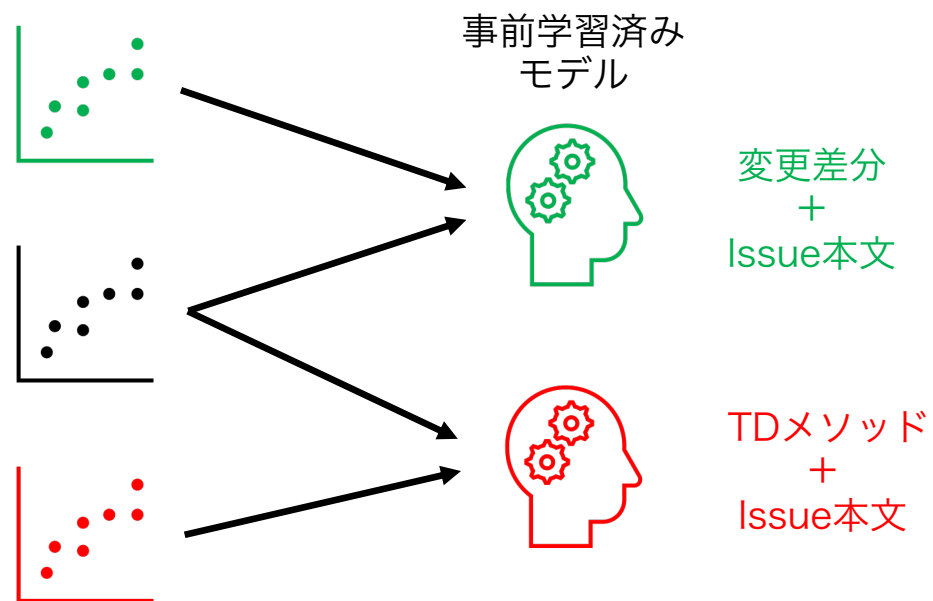
1. TDを含むメソッドと変更差分を取得



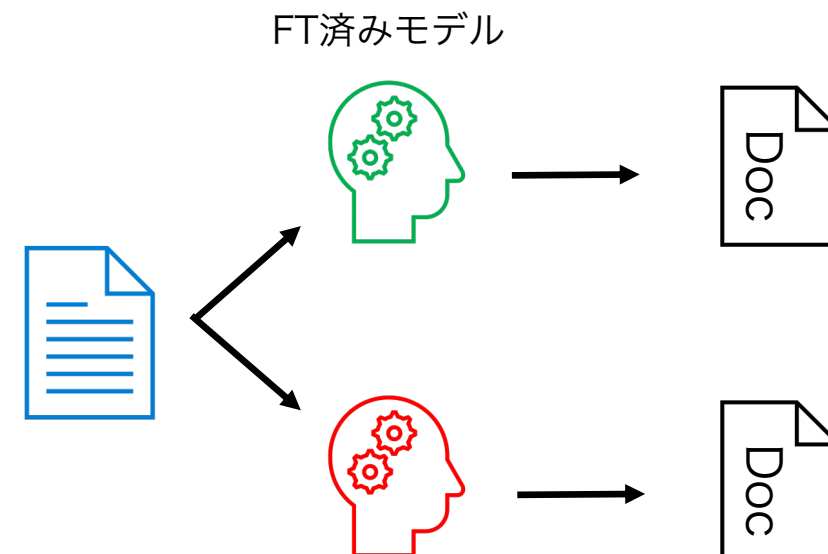
2. ソースコードとIssue本文のベクトル化



3. 事前学習済みモデルをファインチューニング

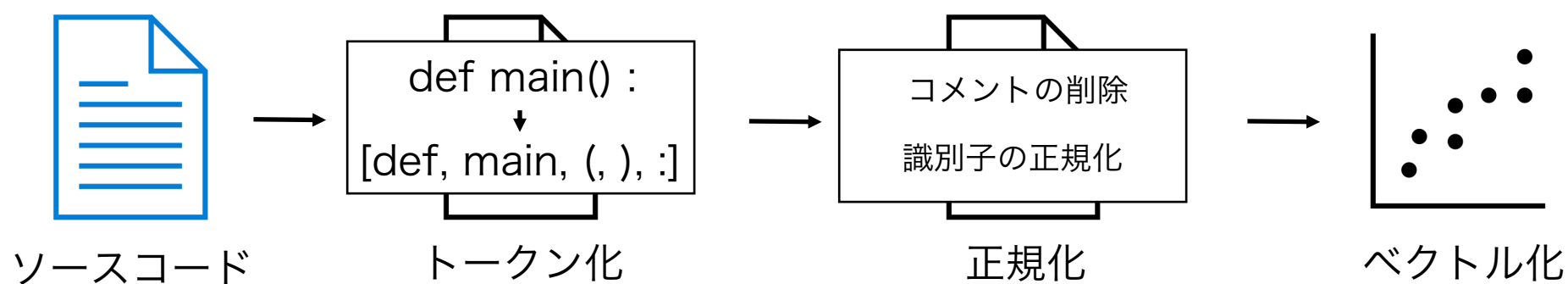


4. FTモデルを使ってIssue本文を生成

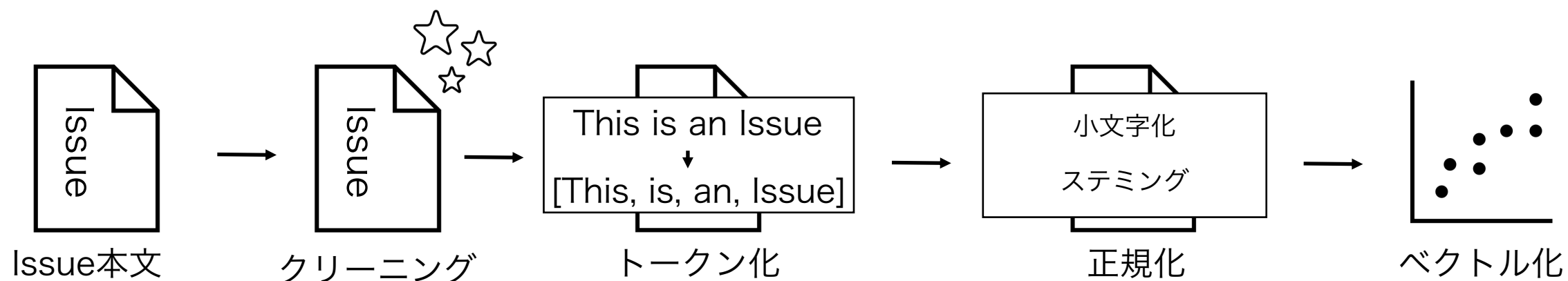


前処理 (仮)

ソースコード



Issue本文



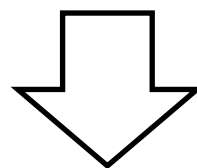
データセット

Javaプロジェクトのリポジトリを対象

- Java：多くの大規模プロジェクトがあり，技術的負債を多く含む
- 既存研究^[?]でのデータセットをフィルタリングして使用

フィルタリング

- 公開されているリポジトリ
- 過去に技術的負債の除去が行われている



プロジェクト数	技術的負債を含むソースコードファイル
140	28,424

実験データセット

[?] D. Skryseth, K. Shivashankar, I. Piln and A. Martini, "Technical Debt Classification in Issue Trackers using Natural Language Processing based on Transformers," In Proceedings of 6th International Conference on Technical Debt (TechDebt), pp. 92-101, 2023

TD Issueに紐づいたソースコードの特定

1. Pull Requestに変更情報が記載

tests: ensure "expected" is first value, "actual" is second value #101

New issue

Closed ndushay opened this issue on Jun 1, 2017 · 0 comments · Fixed by #108

ndushay commented on Jun 1, 2017 Contributor

Know this is wrong for DownloaderSettings tests; perhaps for others as well.

ndushay added **tech debt** **in progress** labels on Jun 1, 2017

ndushay self-assigned this on Jun 2, 2017

ndushay mentioned this issue on Jun 2, 2017
Settings updates #108

ndushay added this to the **2017-06-02 automate crawls sprint 5** milestone on Jun 2, 2017

tingulfen closed this as completed in #108 on Jun 3, 2017

Assignees
 ndushay

Labels
tech debt

Projects
None yet

Milestone
2017-06-02 automate crawls sprint 5

Development
Successfully merging a pull request may close this issue.

Settings updates
sul-dlss/wasapi-downloader

Development

Successfully merging a pull request may close this issue.

Settings updates
sul-dlss/wasapi-downloader

2. Issueにコミット履歴が紐づいている

Opdd commented on Jun 9, 2019 · edited by php-coder

The puzzle 927-57fa0b9d from #927 has to be resolved:

```
mystamps/src/main/java/ru/mystamps/web/feature/category/CategoryValidation.java
Line 22 in 84316e6
22 // @todo #927 CategoryValidation: remove CATEGORY_ prefix from the constants
```

The puzzle was created by Slava Semushin on 08-Jun-19.

Opdd added the **tech debt** label on Jun 9, 2019

Opdd referenced this issue on Jun 9, 2019
 refactor(SeriesValidation): move series-related values to the corresp...

Opdd mentioned this issue on Jun 9, 2019
Split packages by features #927
14 tasks

php-coder added this to the **0.4.1** milestone on Jun 9, 2019

php-coder added **refactoring** **trivial** labels on Jun 9, 2019

php-coder self-assigned this on Jun 21, 2019

php-coder closed this as completed in 12acc5a on Jun 21, 2019

Assignees
 php-coder

Labels
refactoring **tech debt** **trivial**

Projects
None yet

Milestone
0.4.1

Development
No branches or pull requests

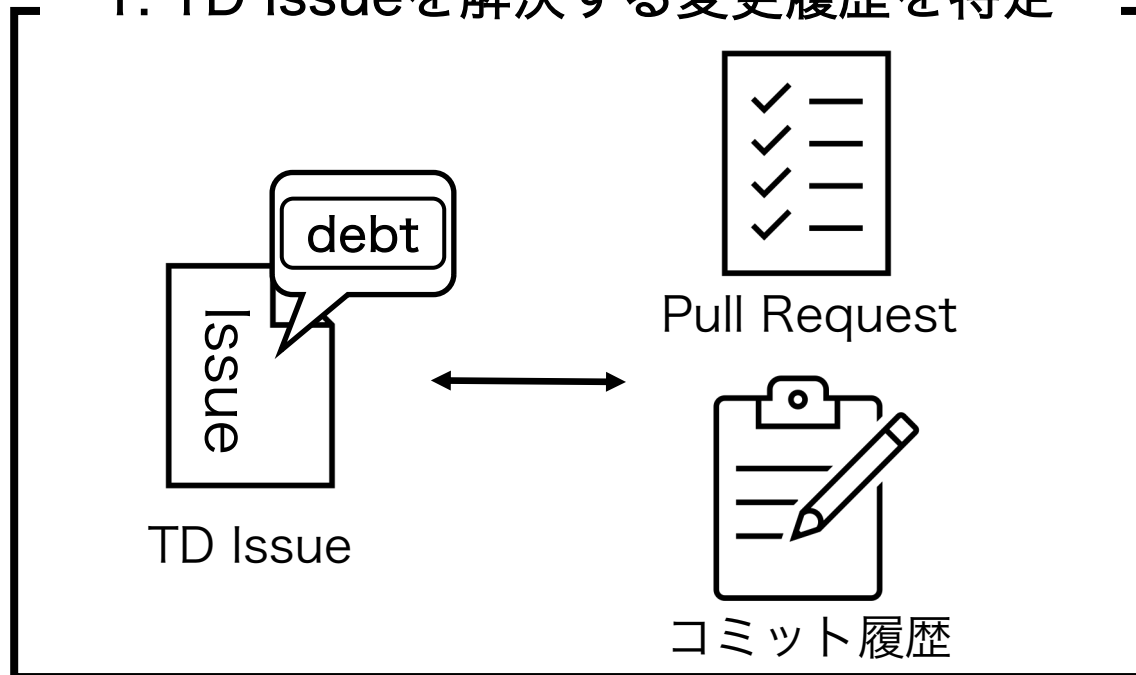
Notifications
Customize
 Subscribe
You're not receiving notifications from this thread.

2 participants

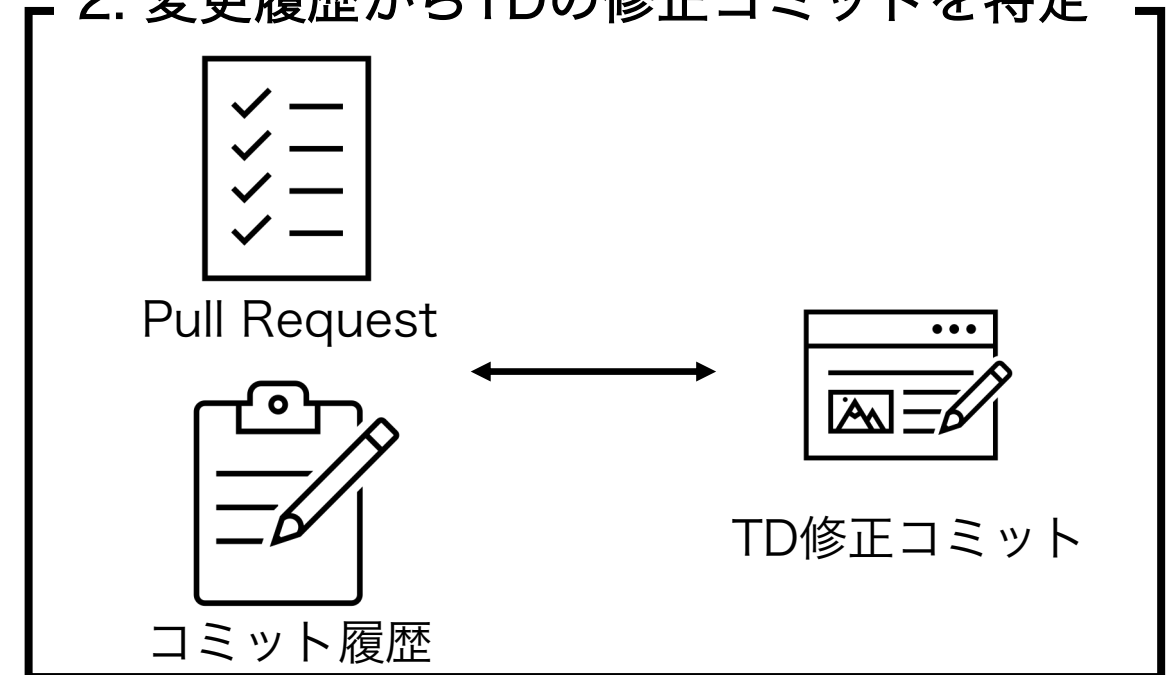
php-coder closed this as completed in 12acc5a on Jun 21, 2019

TD Issueとソースコードの紐付け

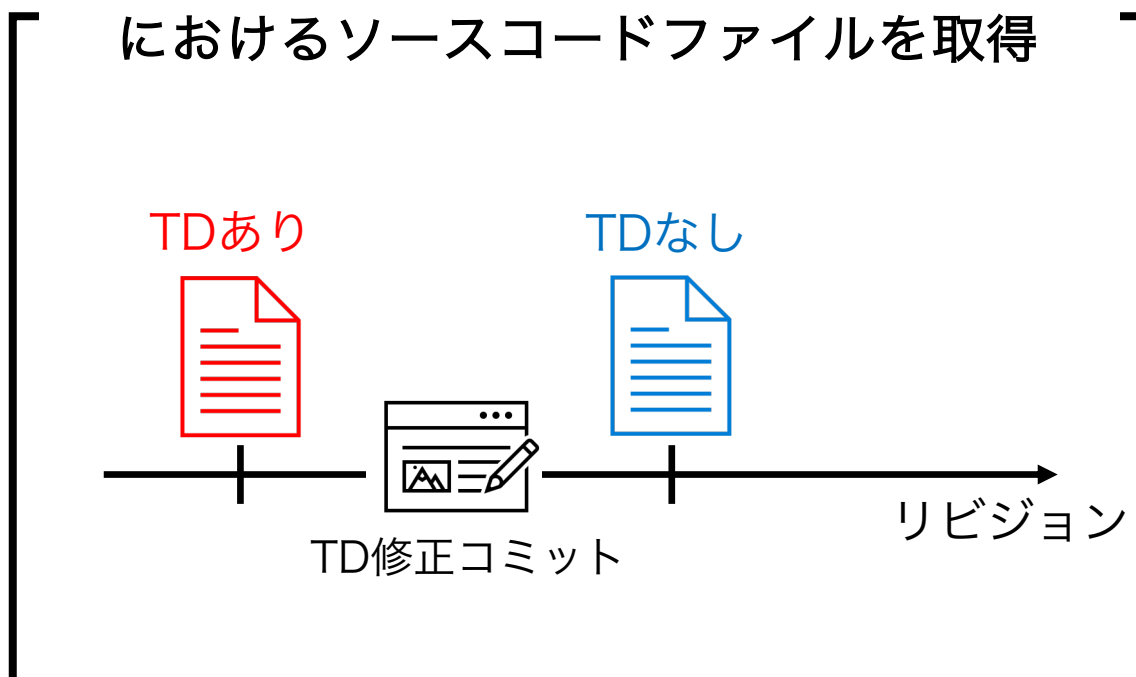
1. TD Issueを解決する変更履歴を特定



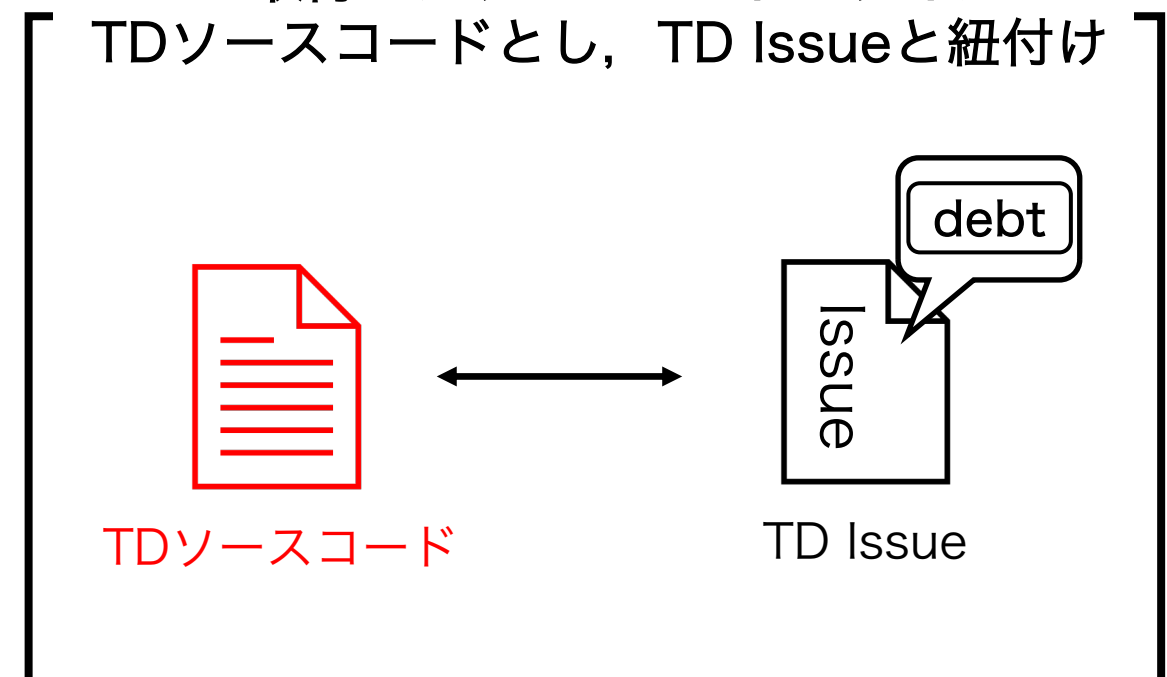
2. 変更履歴からTDの修正コミットを特定



3. 特定したコミットから一つ前のリビジョンにおけるソースコードファイルを取得



4. 3で取得したソースコードファイルをTDソースコードとし、TD Issueと紐付け



今後の予定

とりあえずIssue本文を生成してみる？

1. 持っているデータセットからIssue本文を取り出す

→スクリプトを作成中

2. TDソースコードとIssue本文を元にFT

アプローチの独自性（本研究の工夫点）を考える

実験設定の決定