

変更理由を含む コミットメッセージの 自動生成手法の構築

オープンソースソフトウェア工学研究室
学籍番号：60230113
氏名：松尾 和輝

ソフトウェア開発

ユーザのニーズ，マーケティング上の目標をソフトウェアとして実現する作業

- 開発者らが行ったソフトウェアへの変更を，バージョン管理システムへコミットすることで開発を進める

バージョン管理システム：コンピュータ上でソフトウェアの変更履歴を管理するためのシステム

コミット：ソフトウェアへの変更（diff）をバージョン管理システムへ反映させるための機能

任意でコミットメッセージを記述することができる



コミットメッセージ

- ソフトウェアへの変更を説明する文
 - 変更内容や課題票へのリンク等が記述される

※課題票：不具合報告や新機能の提案などの課題が記述されたもの

```
● Try computing windows paths correctly on markdown drop (#166077)
Fixes #165352

language-features/src/languageFeatures/dropInto

diff --80,7 @@ export function createUriListSnippet(document: vscode.TextDocument, uris: readon
80      80
81      81      const snippet = new vscode.SnippetString();
82      82      uris.forEach((uri, i) => {
83      -          const mdPath = dir && dir.scheme === uri.scheme && dir.authority === uri.authority
84      -              ? encodeURIComponent(path.posix.relative(dir.path, uri.path))
85      -              : uri.toString(false);
83      +          const mdPath = getMdPath(dir, uri);
```

コミットの例¹

- コミットメッセージの自動生成に関する研究が盛ん^[1, 2]

1. <https://github.com/microsoft/vscode/commit/29a4da129d1a68de066ffa79d7aad2906f50502a>

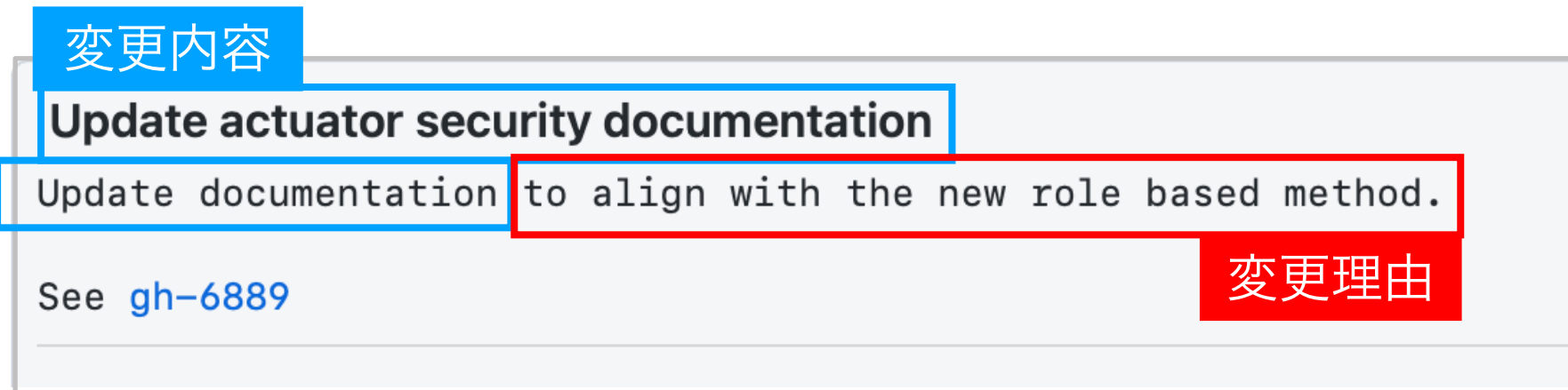
[1] Zhongxin Liu, Xin Xia, Ahmed E. Hassan, David Lo, Zhenchang Xing, and Xinyu Wang. 2018. Neural-machine-translation-based commit message generation: how far are we?. In Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE 2018, Montpellier, France, September 3-7, 2018. ACM, 373-384.

[2] Lun Yiu Nie, Cuiyun Gao, Zhicong Zhong, Wai Lam, Yang Liu, and Zenglin Xu. 2021. CoreGen: Contextualized Code Representation Learning for Commit Message Generation. Neurocomputing (2021).

良いコミットメッセージとは？

変更内容と変更理由が書かれたコミットメッセージ^[3]

- ソフトウェア開発は複数人で行うため、良いコミットメッセージはメンバー間の相互理解のために重要^[4, 5]
- 経験豊富な開発者の殆どが、コミットメッセージに**変更内容**だけでなく**変更理由**も記述するべきと認識^[3]



良いコミットメッセージの例²

[3] Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu, What Makes a Good Commit Message?, 44th International Conference on Software Engineering (ICSE2022)
[4] Kapil Agrawal, Sadika Amreen, and Audris Mockus. 2015. Commit Quality in Five High Performance Computing Projects. In 1st IEEE/ACM International Workshop on Software Engineering for High Performance Computing in Science, SE4HPCS 2015, Jeffrey C. Carver, Paolo Ciancarini, and Neil P. Chue Hong (Eds.). IEEE Computer Society, 24–29.
[5] Y. Tao, Y. Dang, X. Tao, D Zhang, and S. Kim. 2012. How do software engineers understand code changes? An exploratory study in industry. In Acn Sigsoft International Symposium on the Foundations of Software Engineering.
2. <https://github.com/spring-projects/spring-boot/commit/f8a53cf775b03c9258b42be194e05e0ee1779f35>

コミットメッセージ自動生成に関する従来研究

- diffとコミットメッセージのペアを訓練データとして、diffからコミットメッセージを自動生成するモデルを学習[1, 2]



- 問題点：生成されるコミットメッセージは**変更理由が含まれにくい**

変更理由を自動生成しにくい理由

- 約32%のコミットに変更理由が記述されていない^[3]
 - － 原因
 - 時間やモチベーションの不足
 - どのように変更理由を記述すれば良いかが分からない
 - － 問題
 - 他者が行った変更の意図を理解できず余分なコストが発生
 - 変更の意図を他者に尋ねる
 - 変更するにいたった議論をまとめた議事録等を探す
 - 変更の意図を自分で見当づける等行う

diff, コミットメッセージ以外の情報を用いて
変更理由を含むコミットメッセージを
自動生成する手法が必要

具体例

- 課題票#4949：READMEをreStructuredTextからMarkdownに変更したのはなぜ？
 - コミット0bc568fで行われた変更が理解できない
 - 課題票#4888で変更理由を説明していることを示唆

CLN: move README.rst to markdown

コミット0bc568fで
記述されたコミットメッセージ³

The screenshot shows a GitHub issue titled "Please explain: why changed to markdown? #4949". The issue is marked as "Closed". The title is highlighted with an orange box and labeled "件名" (Subject). The main content of the issue is highlighted with an orange box and labeled "本文" (Body). A blue arrow points from the commit message "CLN: move README.rst to markdown" to the commit hash "0bc568f" in the issue body. The issue body contains the following text: "In 0bc568f #diff-04c6e90faac2675aa89e2176d2eec7d8 you changed to markdown. Why? It had stopped to daily Ubuntu builds. So I am just interested in the reason. After a short explanation, you may close here. Thanks." The comment section is highlighted with an orange box and labeled "コメント" (Comment). The comment contains the text: "#4888".

Please explain: why changed to markdown? #4949

Closed 件名 this issue on Sep 23, 2013 · 2 comments

本文 commented on Sep 23, 2013 Contributor

In 0bc568f #diff-04c6e90faac2675aa89e2176d2eec7d8 you changed to markdown.

Why?
It had stopped to daily Ubuntu builds. So I am just interested in the reason.
After a short explanation, you may close here.

Thanks.

コメント commented on Sep 23, 2013 Contributor

#4888

課題票#4949⁴

3. <https://github.com/pandas-dev/pandas/commit/0bc568fb5e19208b30255d771dbf08552f47a4ba#diff-04c6e90faac2675aa89e2176d2eec7d8>
4. <https://github.com/pandas-dev/pandas/issues/4949>

目的とアプローチ

- 目的

- コミットメッセージにリンク付けされた課題票を用いて、変更理由を含むコミットメッセージの自動生成手法を構築

- アプローチ

- 変更理由の自動生成
 - コミットメッセージにリンク付けされた課題票から、変更理由を特定するモデルを作成
 - 特定した変更理由を要約しコミットメッセージに付与

進捗

- 方向性

- 課題票から変更理由を要約するためには、変更理由がどこに記述されているかを特定する必要がある
- そのためには変更理由を表す文章か、そうでないかを分類するモデルが必要
- まずは、変更理由を表す文章かそうでないかを分類するモデルを作成する

今後の予定

- 分類のための特徴量を選定
 - テキスト情報
 - 文章
 - 単語数
 - 余り違いはない？
 - 文法的, 文脈的情報
 - どうやって取り入れる？
 - データを見て考える