

変更内容と変更理由を含む 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

ユーザーのニーズ、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

- バージョン管理システムを利用して複数人で開発を行う

※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

リポジトリという単位でソフトウェアを保存している

- ソフトウェアへの変更をリポジトリへコミットすることで開発を進める



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してコミット
 - コミットメッセージを記述するメリット
 - 他者が行った変更を理解することが可能
 - 過去に行った変更を理解することが可能

Try computing windows paths correctly on markdown drop (#166077) ...
mjbvz committed 7 hours ago

Don't fail to send notebook open perf when reopening the same notebook...
roblourens committed 11 hours ago ✓

Commits on Nov 12, 2022

Pickup bundle.l10n.json (#166169) ...
TylerLeonhardt committed yesterday ✓

Merge pull request #166139 from microsoft/connor4312/cli-server-messa...
connor4312 committed yesterday ✓

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

コミットメッセージ language-features/src/languageFeatures/dropIntoEditor.ts

diff

```
80 80
81 81
82 82
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);
```

コミットログの例¹

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

コミットメッセージの現状

- 時間やモチベーションの不足のため、開発者はコミットメッセージを書くことを疎かにしている^[1]
 - コミットメッセージの14%が空白、75%が1-15単語で書かれている^[1]

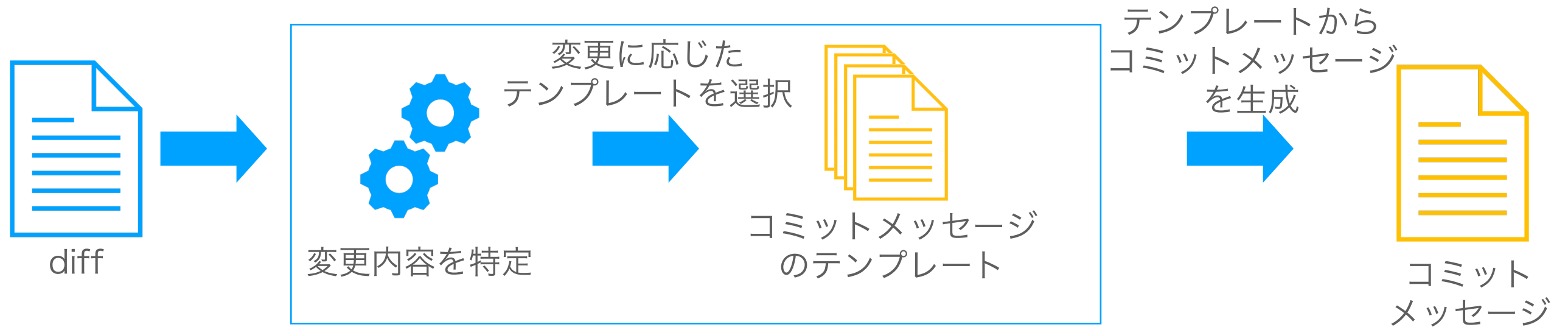


- コミットメッセージを自動生成する研究が数多く行われている
 - テンプレートを用いて生成する手法
 - 機械学習を用いて生成する手法

[1] Robert Dyer, Hoan Anh Nguyen, Hridesh Rajan, and Tien N. Nguyen. 2013. Boa: A Language and Infrastructure for Analyzing Ultra-large-scale Software Repositories. In Proceedings of the 2013 International Conference on Software Engineering (ICSE '13). IEEE Computer Society, 422-431.

コミットメッセージの自動生成

- テンプレートを用いて生成する手法[2, 3]



– 利点

- 簡潔なコミットメッセージを生成することが可能

– 欠点

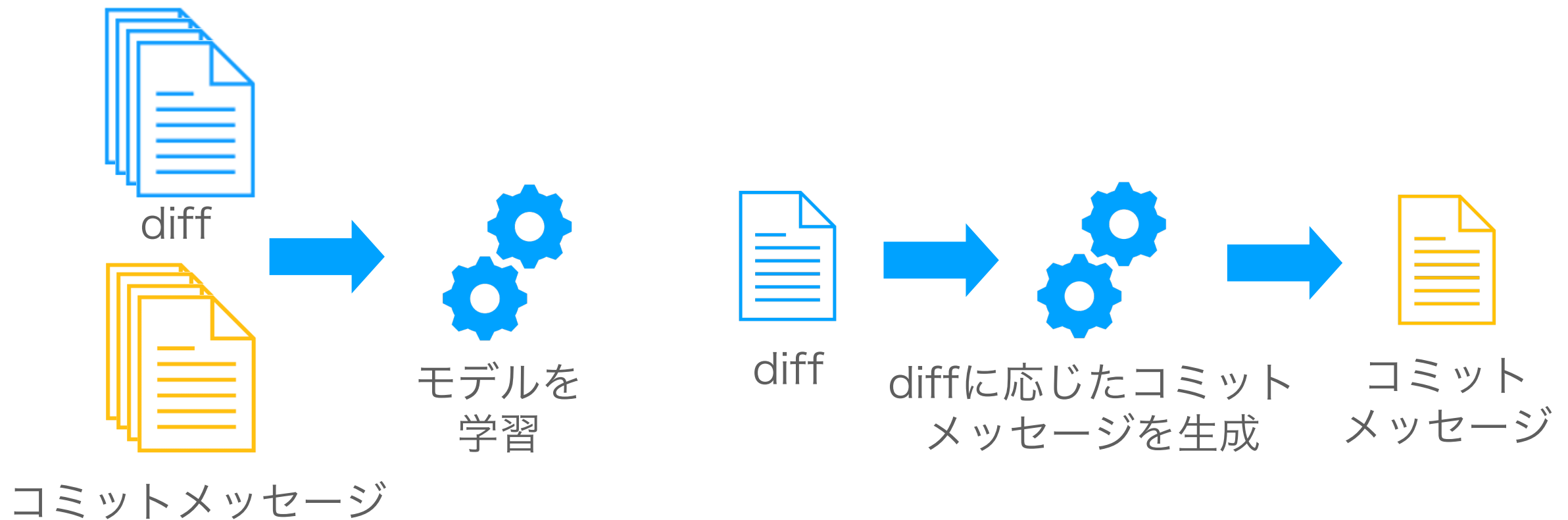
- テンプレートは自作する必要がある
- テンプレートに沿ったコミットメッセージしか生成できない

[2] Rpl Buse and W. R. Weimer. 2010. Automatically documenting program changes. In ASE 2010, 25th IEEE/ACM International Conference on Automated Software Engineering. ACM, 33–42. <https://doi.org/10.1145/1858996.1859005>.

[3] Luis Fernando Cortés-Coy, Mario Linares-Vásquez, Jairo Aponte, and Denys Poshyvanyk. 2014. On automatically generating commit messages via summarization of source code changes. In 2014 IEEE 14th International Working Conference on Source Code Analysis and Manipulation. IEEE, 275–284. <https://doi.org/10.1109/SCAM.2014.14>

コミットメッセージの自動生成

- 機械学習を用いて生成する手法^[4, 5]



– 利点

- テンプレートでは生成できないコミットメッセージも生成可能

– 欠点

- 生成されるコミットメッセージの品質は、学習用データに依存する

[4] 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, Marianne Huchard, Christian Kästner, and Gordon Fraser (Eds.). ACM, 373–384.

[5] 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).

コミットメッセージ自動生成手法の問題

- コミットメッセージには、変更内容（What）と変更理由（Why）を含める必要がある^[6]
 - なぜ変更する必要があるのかを理解できない



- 既存手法では、Whyを含むコミットメッセージが生成できない^[2, 3, 4, 5]

Whatのみのコミットメッセージと、diffで学習させてしまっている^[4, 5]

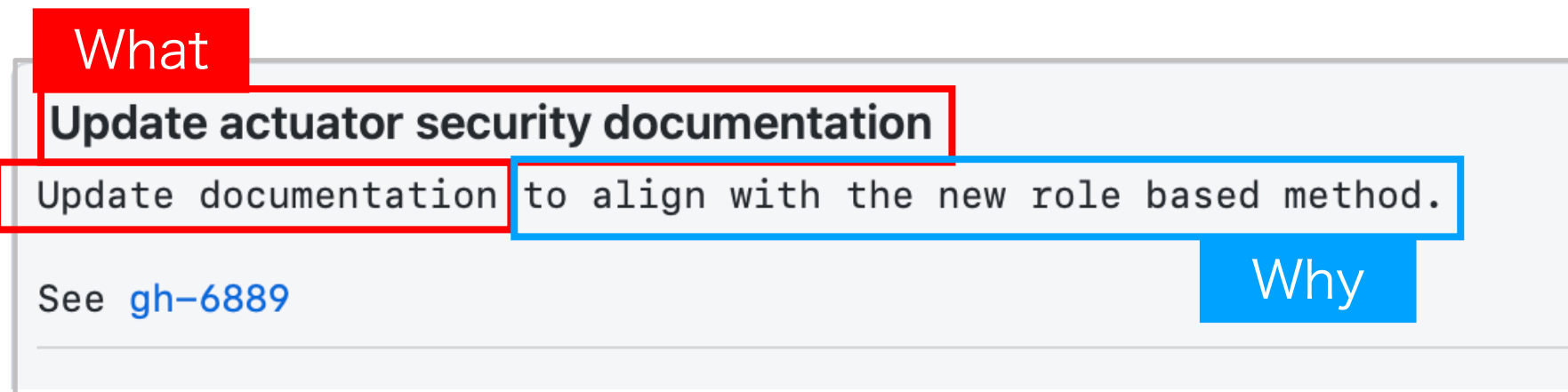
- diffには、コードをどのように変更したのかという情報のみ記載され、なぜ変更したのかという情報は得られない

良いコミットメッセージ

WhatとWhyが書かれたコミットメッセージ^[6]

- 経験豊富な開発者の殆どが、コミットメッセージにはWhatとWhyを含めるべきであるという認識を持つ^[6]

➡ 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である^[7, 8]



良いコミットメッセージの例²

[7] 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.

[8] 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>

目的とアプローチ

- 目的
 - WhatとWhyを含むコミットメッセージを自動生成するモデルを構築する
- アプローチ
 - WhatとWhyをそれぞれ生成して結合
 - What部分
 - diffを学習させてWhatに相当するコミットメッセージを生成
 - Why部分
 - Issue/prへのリンクを活用, リンクがなければLink Recovery手法を用いてリンクを生成
 - リンク先の情報を要約し, Whyに相当するコミットメッセージを生成

進捗

- 方向性
 - What部分にはdiffを活用, Why部分にはLink Recovery手法を用いてそれぞれの生成したコミットメッセージを結合するようなモデルをベースとする
 - 各々を改良してより良いコミットメッセージを生成できるようにする



- 先ずはLink Recovery手法を使えるようにする！！
 - 既存のLink Recovery手法の動作確認を行った

- Link Recovery手法
 - Issueとそれに関連するコミットをリンク付けする技術
 - 1対1, 1対多, 多対1のリンク付け
 - Hybrid Linker^[9]
 - テキスト情報と非テキスト情報の両方を用いてリンク付けを行うモデルを構築した研究
- テキスト情報：Issueタイトル, 説明, diff, コミットメッセージなど
- 非テキスト情報：Issueの著者, コミッター, コミット時間, Issueのタイプ（バグ, 機能, タスク）, Issueの状態（open, close）など

[9] Pooya Rostami Mazrae, Maliheh Izadi and Abbas Heydarnoori. 2021. Automated Recovery of Issue-Commit Links Leveraging Both Textual and Non-textual Data. 2021 IEEE International Conference on Software Maintenance and Evolution (ICSME), 263-273.

今後の予定

- Whatに相当するコミットメッセージを生成する方法を考える
- Issueを要約する方法を考える
- WhatとWhyをそれぞれ結合して最低ラインの良いコミットメッセージを生成できるようにする
- できれば年内に. . .