

変更の要約と変更理由を含む 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用いて複数人で行われる

※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

- 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



コミット

- ソフトウェアへの変更をバージョン管理システムへ反映させるための機能
 - コミットメッセージや変更差分 (diff) が含まれる
 - コミットメッセージ：変更を説明する文
 - 他の開発者が行った変更を理解するために有用

remove terminal data when terminal is disposed of for reconnected ter... [Browse files](#)

コミットメッセージ

main (#155233)

meganrogge committed 6 hours ago Verified 1 parent c1956b8 commit 29a4da129d1a68de066ffa79d7aad2906f50502a

Showing 4 changed files with 64 additions and 44 deletions Split Unified

Filter changed files

src/vs

platform/terminal/common

terminal.ts

terminalProcess.ts

変更差分 (diff)

```
@@ -168,7 +168,7 @@ export interface IPtyHostAttachTarget {
168 168     fixedDimensions: IFixedTerminalDimensions | undefined;
169 169     environmentVariableCollections: ISerializableEnvironmentVariableCollections | undefined;
170 170     reconnectionOwner?: string;
171 -    task?: { label: string; id: string; lastTask?: string; group?: string };
171 +    task?: { label: string; id: string; lastTask: string; group?: string };
```

コミットの例¹

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

良いコミットメッセージ

変更の要約 (What) と変更理由 (Why) が書かれた
コミットメッセージ^[8]

- 経験豊富な開発者の殆どが、コミットメッセージには変更の要約と変更理由を含めるべきであるという認識を持つ^[8]

➡ 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^[9, 10]

What

Update actuator security documentation

Update documentation to align with the new role based method.

See [gh-6889](#)

Why

良いコミットメッセージの例²

[8] Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, Hui Liu, What Makes a Good Commit Message?, 44th International Conference on Software Engineering (ICSE2022)
[9] 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.
[10] 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 Acm 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^[11]
 - 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^[11]

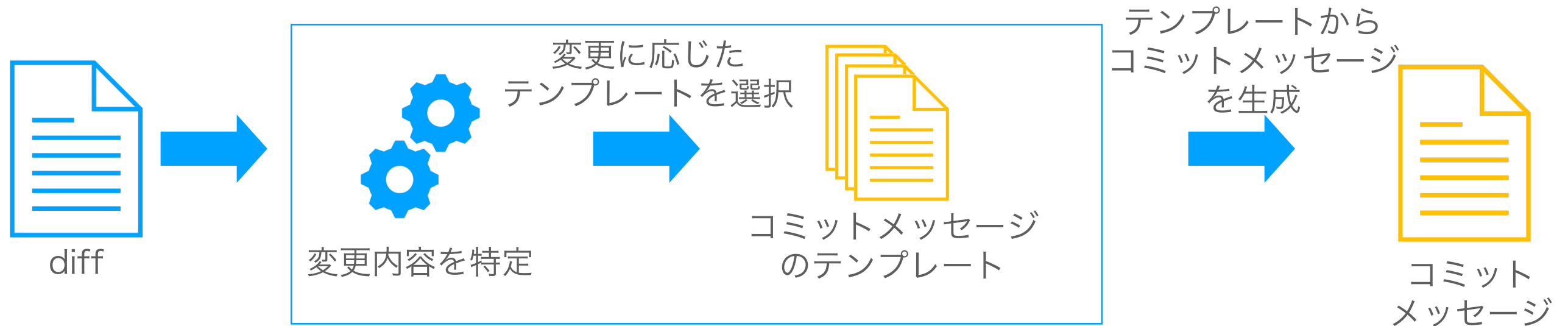


- コミットメッセージを自動生成する研究が数多く行われている
 - テンプレートを用いて生成する手法
 - 機械学習を用いて生成する手法

[11] 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^[12, 13]



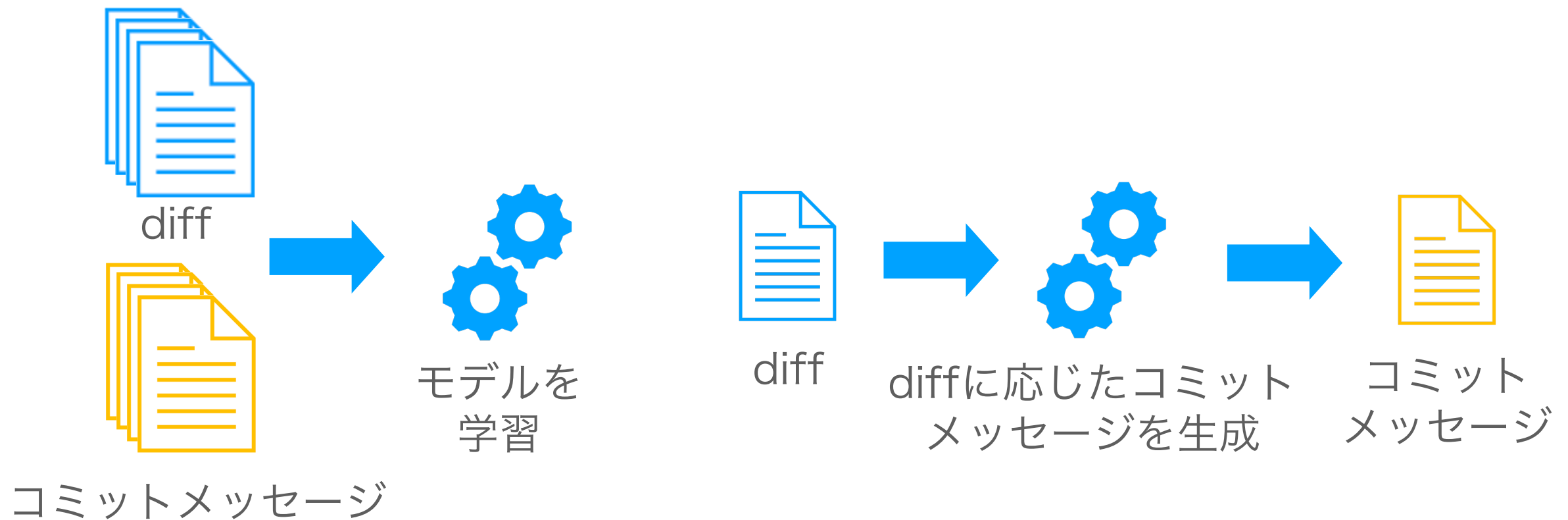
- － 利点
 - ・ 簡潔なコミットメッセージを生成することが可能
- － 欠点
 - ・ テンプレートは自作する必要がある
 - ・ 同じコミットメッセージが生成される可能性がある
 - ・ テンプレートに沿ったコミットメッセージしか生成できない

[12] 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>.

[13] 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^[14, 15]



－ 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

[14] 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.

[15] 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の自動生成手法の問題

- コミットメッセージには、変更の要約だけでなく変更理由も含める必要がある^[8]
 - 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、変更理由を含むコミットメッセージが生成できない^[12, 13, 14, 15]
 - 変更の要約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^[14, 15]
 - 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

目的とアプローチ

- 目的

- 機械学習ベースで、変更の要約と変更理由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

- 変更の要約と変更理由を特定
- これらの情報を用いて、モデルを学習

進捗

- 既存手法で、変更理由を含むコミットメッセージを生成できるかどうか不明
 - 仮説：普通のコミットメッセージを用いて学習させているため、良いコミットメッセージが生成できないのでは^[8]



- 良いコミットメッセージを用いて、既存の機械学習ベースの手法を用いて学習させる

進捗

- 既存手法
 - CoreGen^[15]
 - Transformerベースのモデル
- データセット
 - ①普通のコミットメッセージ（データ数：27,144）
 - ②普通のコミットメッセージ（データ数：756）
 - ③良いコミットメッセージ（データ数：756）

のそれぞれで学習

進捗

- ①普通のコミットメッセージ（データ数：27,144）
- ②普通のコミットメッセージ（データ数：756）
- ③良いコミットメッセージ（データ数：756）

	BLEU	LOUGE-1	LOUGE-2	LOUGE-L	METEOR
①	20.78	32.46	20.19	32.11	16.18
②※1	3.91	20.09	6.28	19.82	7.81
③	0.00※2	17.17	8.60	16.75	9.67

※1 データを変えて5回試し、最も良い精度を出したものを記載

※2 BLEU-1からBLEU-4の結果のlogを平均して計算しているが、いずれかが0になると0になってしまう

- ③の生成例
 - 文章として成り立っている (≠良いコミットメッセージ)
 - 生成文 : `update readme.md`
 - 正解文 : `Update README.md - > <iden> moved to Github`
 - 文章として成り立っていない
 - `remove unnecessary <file_name> :: remove remove ( <pr_link> )`
`method method method method method method method`
`method method method method method method method`
`method method method method method` (以下略)
 - 意味が真逆 + 文章として成り立っていない
 - 生成文 : `remove <file_name> ( ) method method method`
`method method method on <method_name> ( <pr_link> )`
 - 正解文 : `Add <iden> annotation .`

進捗

- ②について
 - 文章として成り立っているメッセージもあれば、成り立っていないメッセージもある
 - 学習させるデータによって変わる？
 - ①について
 - 文章として成り立っていないメッセージは少ない
- 
- データが少ないためうまく生成できない？
 - 良いコミットメッセージのデータを増やして学習させればうまくいく？

今後の予定

- 良いコミットメッセージのデータを増やす
 - 先行研究で提案された, 良いコミットメッセージ分類器を使う
- WhatとWhyをコミットメッセージに含める方法を考える
 - 良いコミットメッセージに含まれるWhyは, 何をもとに書かれているのかを調べる (Whatも?)
- 他の生成手法でも良いコミットメッセージが生成できないか検証する?