

高品質なコミットメッセージの 自動生成に向けたデータ収集手法の構築

オープンソースソフトウェア工学研究室
S2420142 山村慎太郎

コミットメッセージ

Gitにおいて、ファイル・ソースコードの**変更**やその記録の際に付与される説明文



コミットメッセージの役割・利点^[3]

- 開発者間のコミュニケーション手段
 - コード変更の意図・影響理解の促進
- プロジェクトの長期的な進化における、変更に関する情報源

1. <https://github.com/junit-team/junit4/commit/078e45bd31a9920e00633ad6ce658e22529304f2>

[3] Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, and Hui Liu. 2022. What makes a good commit message? In Proceedings of the 44th International Conference on Software Engineering (ICSE '22)

[4] Jiawei Li, Iftekhar Ahmed, Commit Message Matters: Investigating Impact and Evolution of Commit Message Quality, 45th International Conference on Software Engineering (ICSE2023)

コミットメッセージの現状と課題

メッセージには、**変更内容**と**変更理由**の両方含まれるのが開発者にとって好ましい^[3]

変更部分 @Bean
— @Primary
 public Integ

変更差分¹

Remove @Primary from IntegrationMBeanExporter

Commit [3ea84f9](#) has wrongly introduced a `@Primary` marker on `IntegrationMBeanExporter` so any use of both Spring's JMX support and Spring Integration's JMX support leads to an exception. This commit makes sure to remove the unnecessary `@Primary`

コミットメッセージ¹

変更理由

- 必要性：メンテナンス作業の効率化^[3]，欠陥発生率の低減^[5]

一方，これらの情報が記述されないケースも存在^[3]

- 変更内容や理由が欠如しているメッセージ
- 何も記述されず，空白となっているメッセージ

➤ 原因：開発者がメッセージを書くための時間や動機の不足

1. <https://github.com/junit-team/junit4/commit/078e45bd31a9920e00633ad6ce658e22529304f2>

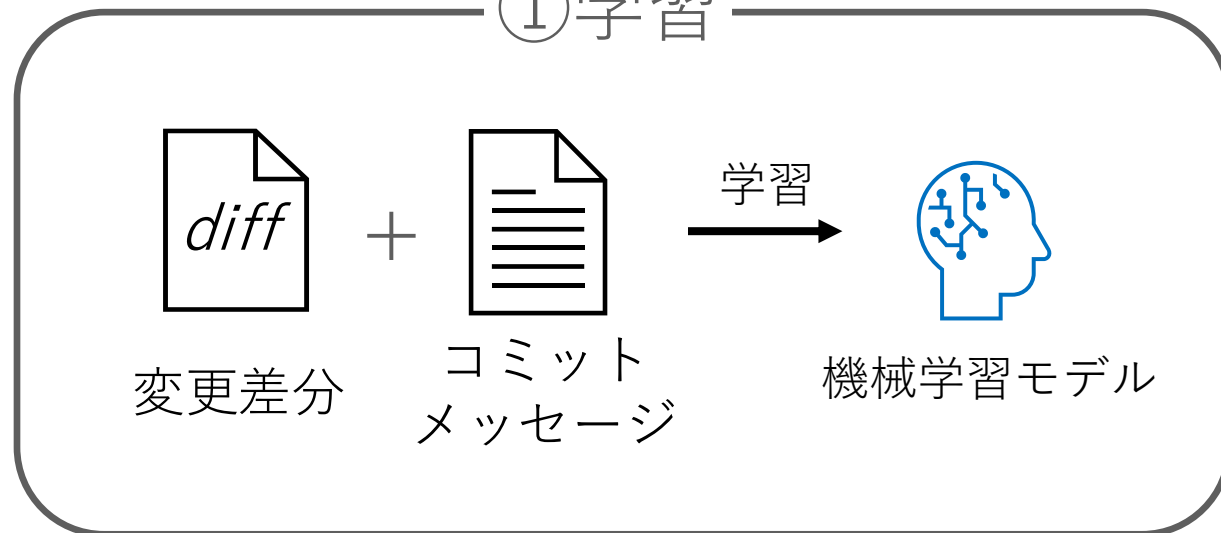
[3] Yingchen Tian, Yuxia Zhang, Klaas-Jan Stol, Lin Jiang, and Hui Liu. 2022. What makes a good commit message? In Proceedings of the 44th International Conference on Software Engineering (ICSE '22)

[4] Jiawei Li, Iftekhar Ahmed, Commit Message Matters: Investigating Impact and Evolution of Commit Message Quality, 45th International Conference on Software Engineering (ICSE2023)

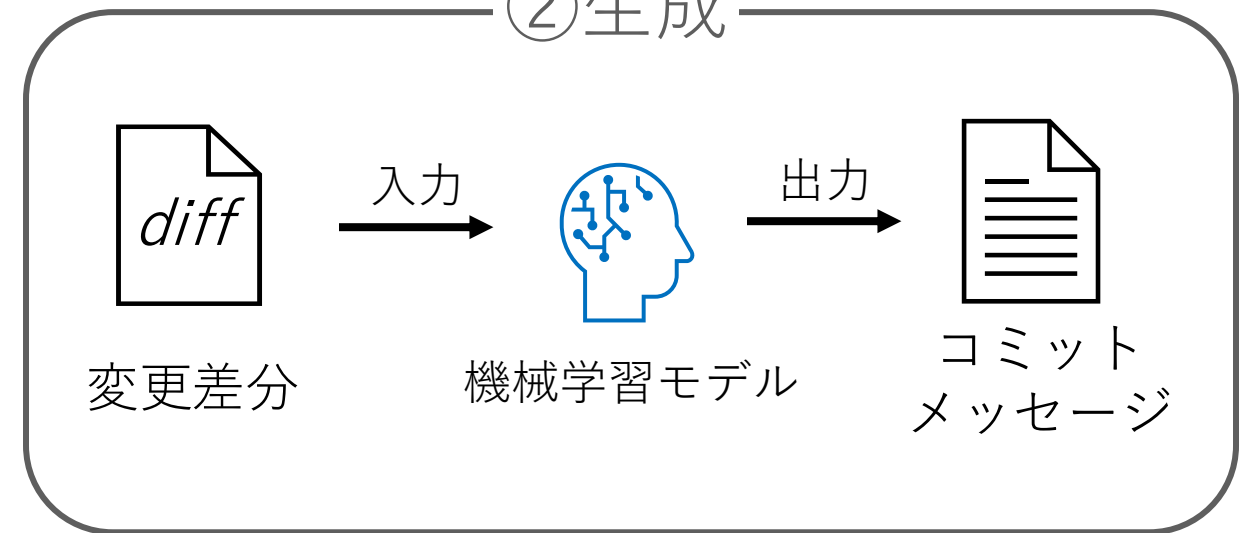
コミットメッセージの自動生成

例：学習ベース^{[5],[6]}

①学習



②生成



- ① 変更差分とメッセージのペアを訓練データとしてモデルを学習
- ② 変更差分からメッセージを自動生成

自動で生成できるようにすることで、開発者の時間・動機不足に対処

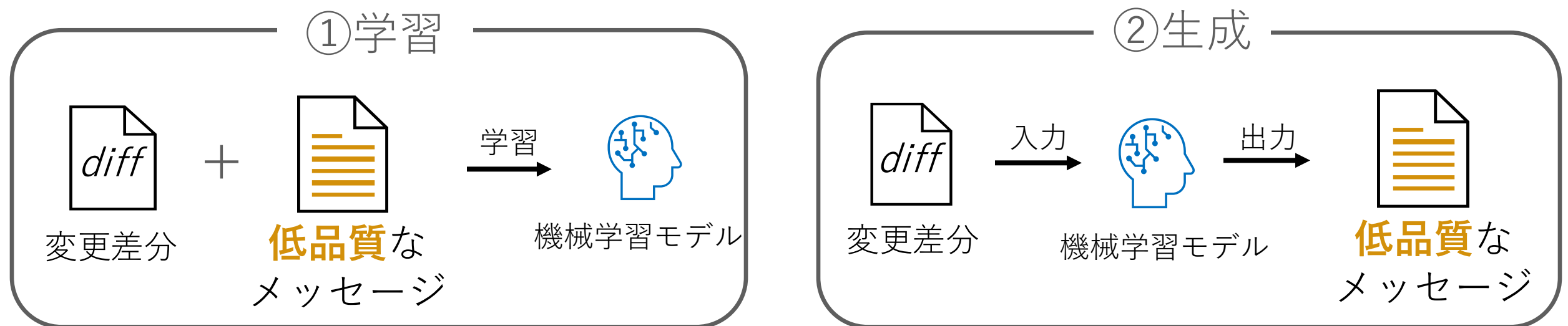
[5] Zhongxin Liu, Xin Xia, Ahmed E. Hassan, David Lo, Zhenchang Xing, and Xinyu Wang. 2018. Neural-machine-translation-based commit message generation: how far are we?. In Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, ASE 2018, Montpellier, France, September 3-7, 2018. ACM, 373–384.
[6] 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).

自動生成の課題

学習データの品質が保証されていない

- 学習には、OSSプロジェクトのメッセージが使用される
- 調査された5つのOSSプロジェクトのメッセージのうち、平均約44%のメッセージに改善の余地あり^[3]

学習データの品質に問題があると、生成されるメッセージの品質にも影響



- 変更**内容**・**理由**を含まないメッセージが生成されてしまう
- テストデータとして扱う際、生成物の評価も適切に行えない^[7]

先行研究^[3]

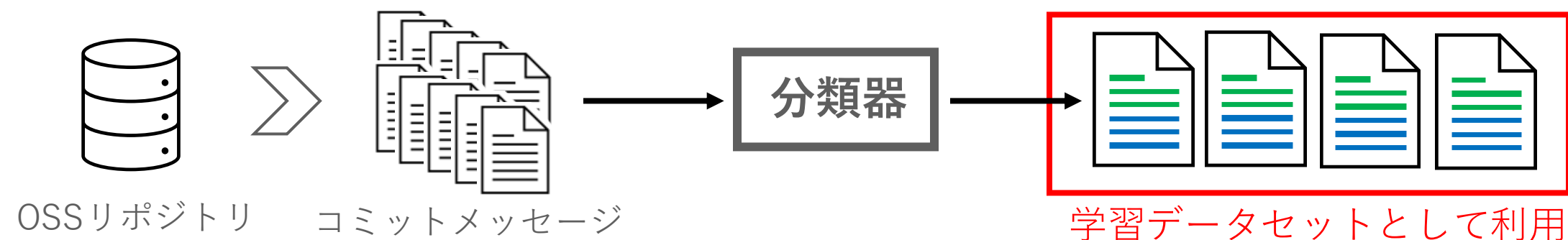
変更内容・理由を含むメッセージを自動で識別する分類器

- 変更内容・理由の有無を手動で分類した約1600件のデータで学習



データセットの作成に活用可能

- 既存のOSSリポジトリから、変更内容・理由を含むメッセージを収集可
- 品質が保証された学習データセットの不足に対処



先行研究の課題点

先行研究の分類器（Bi-LSTM）の分類精度^[3]

変更内容・理由	Precision	Recall	F1
あり（正例）	0.81	0.74	0.78
なし（負例）	0.70	0.78	0.74

- 正例予測のうち、実際に正例だった割合（Precision）は約80%
- 約20%が誤分類（偽陽性）

データセットの作成に利用する場合、変更内容・理由が欠如したメッセージが20%程度混在

- 20%は学習において無視できない割合
- 大規模なデータセットを作成する場合、累積的に大きな影響

本研究の目的・アプローチ

コミットメッセージの自動生成に向けた、**変更内容**と**変更理由**を含む学習データセットの作成

- 高品質なデータセットを作成することで、生成メッセージの品質向上に貢献

アプローチ：事前学習済みLLMでメッセージを分類

- より高い精度での変更内容・理由の分類
- 特に，“Precision”を重視した分類を目指す
 - メッセージのデータは、既存のOSSリポジトリに大量に存在するため

研究の計画

(P1) 変更内容・理由を含むメッセージの分類

- Precisionの高い分類手法の確立

(P2) 変更内容・理由を含むメッセージの収集

- 既存のOSSリポジトリのデータに対し分類・収集

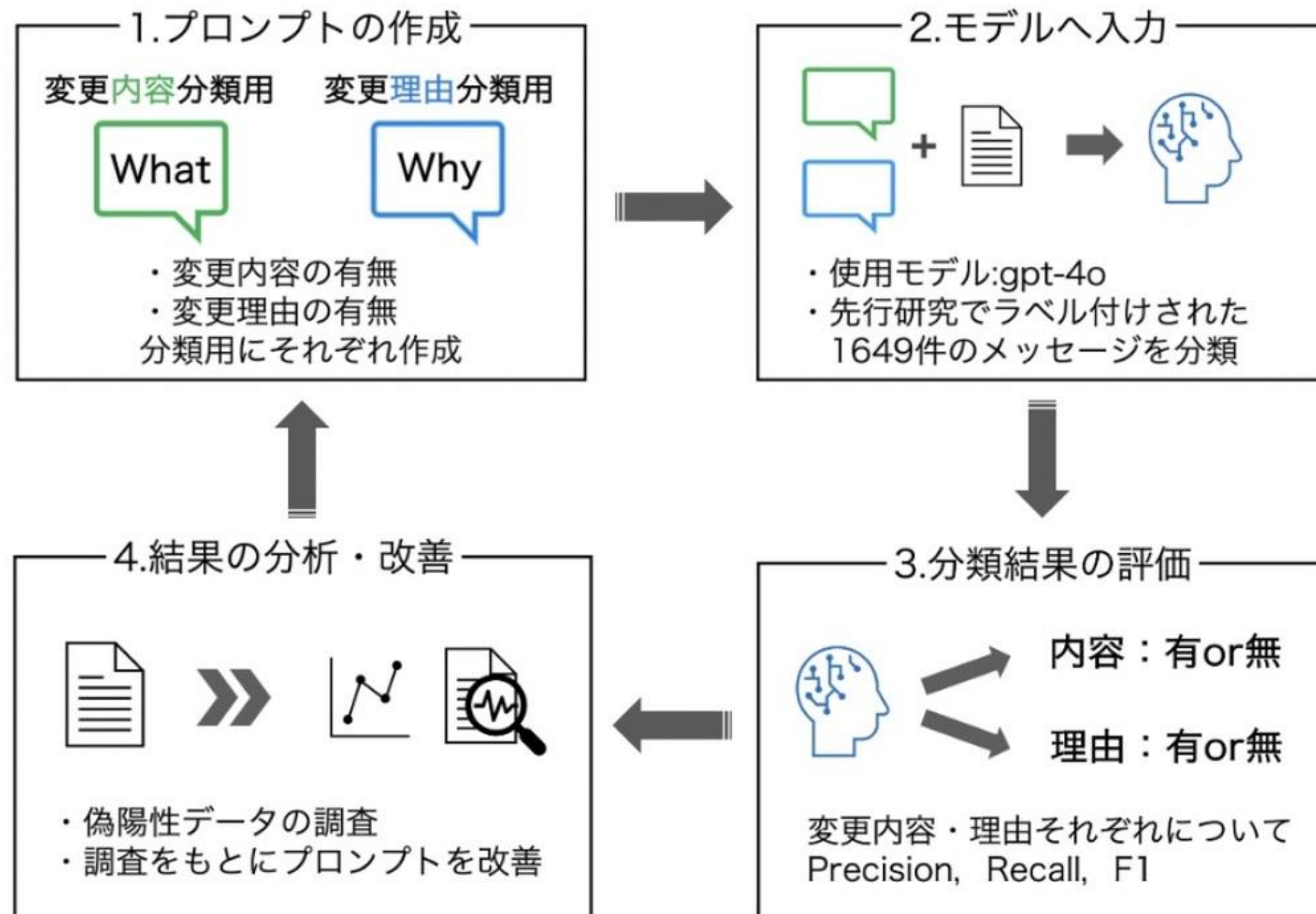
(P3) コミットメッセージの自動生成

- 収集したメッセージで学習・生成

➤ 今回は、それぞれのフェーズの更新内容を説明

(P1) Precisionの高い分類手法の確立

前回：Precisionの高い分類手法を構築



(P1) Precisionの高い分類手法の確立

課題点：Recallが低すぎる（理由分類）

変更理由分類	Precision	Recall	F1
先行研究 (10分割交差検証)	0.765	0.709	0.731
本手法 (10回平均)	0.924	0.085	0.156

大ゼミにて指摘

問題・影響

- 流石に低すぎる，論文等で見たこと無い数値
 - 特定のパターンのデータしか拾えない可能性
 - データセット作成時に時間・コストがかかる
- 目的ベースでは問題無いように見えたが，様々な悪影響あり

(P1) Precisionの高い分類手法の確立

対応：更なる試行錯誤により， Recallの向上を狙う

- プロンプトの見直し（偽陰性の分析も追加， ChatGPTで分析）
- コスト面を考慮し， モデルを追加（GPT-4o-mini）

改善結果

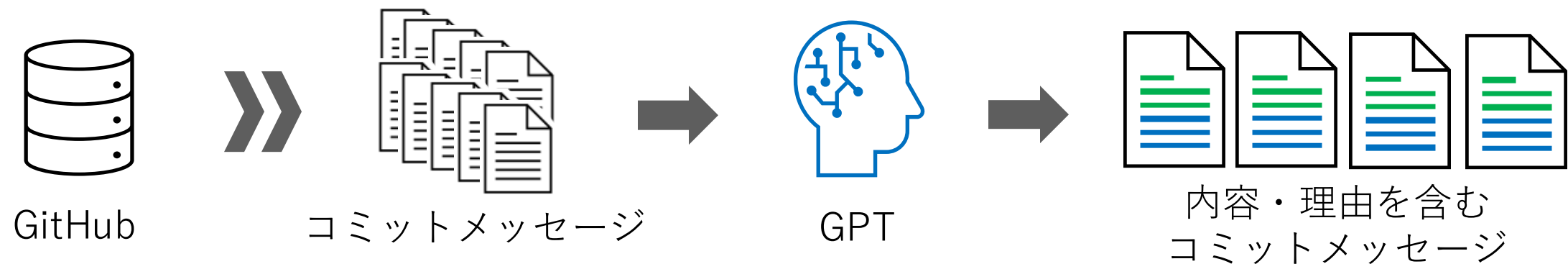
- 理由分類において， GPT-4o-miniが予想外に威力を発揮
 - Precision・Recall共に向上

変更理由分類	Precision	Recall	F1
前まで(4o)	0.924	0.085	0.156
改善後(4o-mini)	0.979	0.399	0.567

➤ 少なくとも理由分類についてはminiを採用？

(P2) データセットの作成

既存のOSSプロジェクトから新たにメッセージを収集



- (P1) の手法により変更内容・理由を含むメッセージを収集し、学習データセットとして活用

予定：収集物を目視評価

- 既存の分類器（Bi-LSTM）では低品質なデータが含まれる^[8]
 - Bi-LSTM分類器をベースラインとして、本手法の優位性を示す
 - つまり、2種類のデータセットを作成
 - “変更内容・理由を含むか”を重視して評価・比較

(P3) コミットメッセージの自動生成

(P2) で作成したデータセットで学習・生成

- 本手法で作成したデータセット
- Bi-LSTM分類器で作成したデータセット

それぞれで生成を行い，“内容・理由を含むか”を目視評価

- もちろんBLEUスコアなど客観評価も

➤ **学習データ**と**生成物**の両面で品質への好影響を示したい

検討事項：few-shotでの生成も評価に？

- 元々の計画では，fine-tuningによる生成のみを想定
- これまでは想定していなかった方向性の追加

(P3) コミットメッセージの自動生成

データ品質がFew-shotでの生成物にも影響^[7]

- Few-shotでのコミットメッセージ自動生成の実証調査
 - 不適切な生成物のうち、約半数は低品質な提示例が起因
 - また、客観評価の際、評価用データの品質が原因で評価が不正確に
- few-shotでも **高品質データセットの必要性**を明示的に主張

トレンドも加味し、few-shotでの生成・評価も検討

- 最近は、プロンプトベースの手法が主流
 - タスクにもよるが、fine-tuningによる手法は下火
- few-shotでも実験を行うことで、研究的価値を向上

進捗状況

(P1) 変更内容・理由を含むメッセージの分類

- Precisionの高い分類手法の確立

(ほぼ)
完了

(P2) 変更内容・理由を含むメッセージの収集

- 既存のOSSリポジトリのデータに対し分類・収集

進行中

(P3) コミットメッセージの生成

- 収集したメッセージで学習・生成

➤ (夏休み中にメイン実験は終わらせておきたい...)