

バグの混入および修正に関連したコミットメッセージの多クラス感情分析

嶋田 拓真

京都工芸繊維大学

t-shimada@se.is.kit.ac.jp

西浦 生成

京都工芸繊維大学

k-nishiura@kit.ac.jp

水野 修

京都工芸繊維大学

o-mizuno@kit.ac.jp

要旨

コミットメッセージは対話手段の一つとして利用され、開発者の感情や意図が表現される。本研究では、コミットメッセージに表れる感情に着目し、多クラス感情分析を用いて、バグの混入および修正に関連する感情の傾向を調査する。オープンソースプロジェクトから取得したコミットメッセージに多クラス感情分析を適用し、表れる感情の傾向を調査する。さらに、SZZアルゴリズムに基づいてコミットに付与したバグの混入や修正時点を表す情報を用いてコミットメッセージを分類し、バグの混入および修正の時点における感情の傾向を調査する。調査の結果、コミットメッセージに表れる感情は、感情の強さレベルに基づく閾値の大きさによって異なる傾向を示した。バグ混入時点とその他の時点との比較、およびバグ修正時点とその他の時点との比較においても、閾値によって特徴的な感情傾向を見せた。これらの結果は、コミットメッセージに表れる感情とその強度が、バグの混入および修正と関連していることを示しており、感情分析を通じてバグの予測に対する理解が深まることが期待される。

1. はじめに

複数の開発者が共同で開発を進めるためのツールとしてGitが広く用いられている。Gitにおける変更単位はコミットと呼ばれ、コミットの内容を記録するためのログとしてコミットメッセージがある。コミットメッセージは他者に明確に伝わるように記述することが求められる。また、近年盛んとなっているオープンソースソフトウェア(OSS)の開発において、遠隔地にいる他の開発者とコミュニケーションを行う上で、バグトラッキング

システムやメーリングリストなどのテキストを用いたコミュニケーション手段が重要となる。コミットメッセージも変更の記録だけでなく、開発者間のコミュニケーション手段の一つとして利用されている側面があり、開発者の感情や意図がそこに表現されることも少なくない[1]。

一方で、自然言語処理技術の一つとして感情分析がある。感情分析は、テキストデータなどに含まれる感情を識別し分析するプロセスであり、ソーシャルメディアやレビューサイトの意見分析などに広く用いられている[2][3][4]。感情分析は、一般に感情極性というネガティブ(negative)またはポジティブ(positive)の2つの極で表される属性に基づいて分類される[5]ほか、近年では、ネガティブ・ポジティブの2値でなく多様な感情の種類の有無を分析することができる多クラス感情分析も研究されている[6]。自然言語を用いて記述するコミットメッセージに感情分析を適用し、その極性とソフトウェアのバグや様々な開発要因との関係を探る研究も行われているが[7][8]、多クラス感情分析を用いてその関係を分析した研究は存在していない。

本研究では、オープンソースプロジェクトから取得したコミットメッセージに多クラス感情分析を適用し、コミットメッセージに表れる感情とコミット内容との関連を調査する。まず、コミットメッセージに多クラス感情分析を適用し、表れる感情の傾向を調査する。次に、コミットの種類(バグ混入、バグ修正)によって、そのコミットメッセージに表れる感情の傾向の違いは生じるかどうかを調査する。具体的には、バグ混入コミットとそれ以外のコミット、バグ修正コミットとそれ以外のコミットで、それぞれ表れる感情の傾向の違いが生じるかどうかを比較する。

2 背景

2.1 感情分析

感情分析 (sentiment analysis) とは、様々な製品やサービス、トピックなどに関する人々の意見や考え、態度を収集し、感情を識別して分析するプロセスである。感情分析は、ソーシャルメディアやレビューサイトなどの意見分析 [2] に留まらず、政治 [3] や医療 [4] など、多岐にわたる分野で活用されている。個人の考えや意見から得られる反応を分析することで、企業や社会にとって有益な情報を得ることができる。また、感情分析の研究は英語を対象としたものが多いが、日本語を始めとする他の言語においても研究が進められている。言語によって文化的背景や言語の表現方法が異なるため、それぞれに適した感情分析技術が求められている。テキストや音声、表情など、多様なデータが感情分析の対象となるが、本研究では、特にテキストデータを対象とした自然言語処理技術としての感情分析を扱う。

感情分析は、一般に感情極性に基づいて分類されることが多い [5]。感情極性とは、ネガティブ (negative) またはポジティブ (positive) という2つの極で表される感情の属性である。どちらの極にも分類できない場合はニュートラル (neutral) として処理される。その一方で、感情は単純な二値分類だけでは捉えきれない複雑さを持つ。喜怒哀楽に代表されるように、感情は多様で多次元的な特徴を持っている。そのため、近年では、複数の感情ラベルを扱う多クラス感情分析を使用した研究も進められている [6]。多クラス感情分析の導入により、感情分析はより詳細な感情の傾向を捉えることが可能になっている。

2.2 ソフトウェア開発における感情分析

OSSでのソフトウェア開発においては、コードレビューやバグトラッキングシステム、メーリングリストなどのコミュニケーション手段が利用されている。このようなテキストベースのコミュニケーションにはしばしば使用者の感情が反映される [1][9]。さらに、チーム作業において感情はタスクの品質や生産性、グループ内の信頼関係、仕事の満足度などに影響を与えることが知られている [10]。そのため、ソフトウェア開発に感情分析を活用することで、開発者の感情がプロジェクトの進行や成果

に与える影響を明らかにすることができる。例えば、バグレポートに含まれる感情からバグレポートの優先度を予測する [11]、メーリングリストに感情分析を適用し、コード品質を予測する [12] といった研究がある。これらの研究は、感情分析がソフトウェア開発における重要な要素となりうることを示しており、品質向上や開発プロセスの効率化に貢献する可能性がある。

2.3 コミットメッセージ

Gitの機能の一つにコミット (commit) があり、これはソフトウェアに加えられた変更履歴を記録するために用いられる。具体的には、ファイルやディレクトリの追加・変更をリポジトリに記録することを指す。コミットする際には、変更内容を説明するメッセージを記述することが一般的であり、これをコミットメッセージ (commit message) またはコミットコメント (commit comment) と呼ぶ。本論文では統一して、コミットメッセージと呼称する。

コミットメッセージは、ソフトウェア開発の過程でのコミュニケーションの一環として使用される。その多くはコミットにおける技術的な側面を説明しており、基本的にはニュートラルで書かれる傾向があるが [7]、開発者の感情や意図がしばしば表現される [9]。そのため、コミットメッセージに表れる感情を分析することで、ソフトウェア開発において有益な情報を抽出できる可能性がある。

コミットメッセージに表れる感情を調査したこれまでの研究では、使用されるプログラミング言語やコミット日時などの開発要因との関係調査 [7] や、ソフトウェアのバグとの関連分析 [8] などが行われている。これらの研究結果から、コミットメッセージの感情がソフトウェア開発プロセスの改善に貢献することが期待されている。その一方で、これらの研究では感情極性が分析の対象となっており、多クラス感情分析が対象とする感情の種類との関係は未だ明らかになっていない。

3 研究目的

本研究の目的は、コミットメッセージから抽出できる感情の種類が、バグの混入および修正にどのように関連しているかを調査することである。具体的に、以下の2つの研究設問を設ける。

- RQ1: コミットメッセージにはどのような感情が表れやすいのか。
- RQ2: コミットの種類によって、コミットメッセージに表れる感情の傾向は異なるのか。

感情分析を行い、コミットメッセージに表れる感情の傾向を明らかにすることで、ソフトウェア開発における感情の影響を理解することを目指す。単にネガティブ・ポジティブの極性を判定するだけでは捉えきれない、より複雑で微細な感情の傾向を把握するために、多様な感情の可能性を分析できる感情分析ツールを用いて調査を行う。使用する感情分析ツールの説明については4.1節で述べる。

RQ2におけるコミットの種類の判別には、SZZ アルゴリズムに基づいてコミットに付与されたタグを利用する。そのため、本研究ではSZZ アルゴリズムが適用されたデータセットを分析対象として用いる。SZZ アルゴリズムについては4.2節で説明する。

4 分析方法

4.1 使用する感情分析ツール

本研究では、以下の2種類の感情分析ツールを使用する。

4.1.1 Senti4SD

Senti4SD[13]は、ソフトウェア工学の分野に特化した感情極性分類器である。Stack Overflow というコンピュータや情報技術に関するコミュニティから抽出した4,423件の投稿から作られたデータセットを用いて、トレーニングとテストが行われている。辞書ベースの特徴、キーワードベースの特徴、意味的特徴を組み合わせたアプローチを採用しており、ソフトウェア開発に特有の専門的な言語の使用に適応できるよう設計されている。Senti4SDはテキストを入力として受け取り、分析後にテキストに感情極性ラベル(negative, neutral, positive)を割り当てる。

4.1.2 roberta-base-go_emotions

roberta-base-go_emotions[14]は、RoBERTaという大規模言語モデルをベースにしたGo Emotions[15]データ

表 1. RoGoEmo で分析可能な感情ラベル一覧

感情ラベル	感情 (和訳)	分類
admiration	感心	Positive
amusement	楽しみ	
approval	好意	
caring	思いやり	
desire	欲望	
excitement	興奮	
gratitude	感謝	
joy	喜び	
love	愛	
optimism	楽観	
pride	誇り	
relief	安心	
anger	怒り	Negative
annoyance	苛立たしさ	
disappointment	失望	
disapproval	不満	
disgust	嫌悪	
embarrassment	恥ずかしさ	
fear	恐れ	
grief	苦痛	
nervousness	緊張	
remorse	後悔	
sadness	悲しみ	Ambiguous
confusion	混乱	
curiosity	好奇心	
realization	実感	
surprise	驚き	

セットで訓練した多クラス感情分析モデルである。本研究では便宜上、このモデルを用いた感情分析器をRoGoEmoと呼称する。

Go Emotions データセットは、ニュートラル (neutral) を含む28の感情ラベルに分類された58,000件の英語のReddit コメントからなるデータセットである。感情ラベルは12種のポジティブな感情、11種のネガティブな感情、4種のあいまいな感情、そしてニュートラルで構成されている。ニュートラルを除く感情ラベルの一覧を表1に示す。RoGoEmoで分析可能な感情ラベルはGo Emotions データセットの28種類の感情に従う。RoGoEmoは、入力テキストに複数の感情が同時に存在する可能性を考慮し、感情ラベルごとにテキストにその感情が表れている確率(感情値)を出力する。

4.2 使用するデータセット

本研究では、Apache James、Apache MINA、Apache OpenJPA、Eclipse BIRT の 4 つのオープンソースプロジェクトから取得したコミットメッセージをデータセットとして使用し、感情分析を行う。取得したコミットメッセージには、SZZ アルゴリズム [16] に基づき、2 種類のタグ（BUG タグ、FIX タグ）が付与されている。SZZ アルゴリズムとは、不具合混入コミットを推定するために広く使用されている手法である。このアルゴリズムでは、ソースコードの変更履歴から、バグがどのコミットで修正されたかを特定し、その修正箇所の変更が行われたコミットを追跡することで、バグが混入したコミットを特定する。その結果、バグが混入したと推定されるコミットには BUG タグが、バグを修正したと推定されるコミットには FIX タグが付与される。また、本研究では、BUG タグまたは FIX タグが付いていないコミットメッセージには新たに OTHER タグを付与する。

4.3 データセットに対する前処理

データセットに対して以下の前処理を行う。

1. 4 つのオープンソースプロジェクトから収集した全てのコミットメッセージの中から、トークン数が 514 以下のコミットメッセージを取得する。
2. 1. で抽出したコミットメッセージに Senti4SD を適用し、negative または positive と判定されたコミットメッセージを抽出する。

この前処理を経て得られた前処理済みコミットメッセージを RoGoEmo の分析対象として使用する。最初にコミットメッセージをトークン数でフィルタリングするのは、RoGoEmo で分析可能なテキストのトークン数に上限があり、514 トークンを超えるテキストに対しては分析を行うことができないためである。negative または positive と判定されるコミットメッセージのみを抽出するのは、RoGoEmo が一般的なメッセージに基づいたデータセットで学習された感情分析ツールであり、ソフトウェア開発の分野特有の用語や言い回しに対してうまく対応できない場合があるためである。Senti4SD を用いて negative または positive と判定されるコミットメッセージを抽出し、その上で RoGoEmo を適用することで、RoGoEmo とコミットメッセージ間のギャップを埋めることを目的としている。

4.4 RQ1 に関する分析手順

前処理で抽出した全ての前処理済みコミットメッセージに対して、RoGoEmo を適用し、多クラス感情分析を行う。得られる結果を基に、コミットメッセージにおける感情の表れやすさを「全ての前処理済みコミットメッセージに対する感情コミットメッセージの割合」として算出し、感情ラベルごとの出現割合を調査する。ここで言う感情コミットメッセージとは、RoGoEmo によって分析された感情値が、設定した閾値を超える値を持つコミットメッセージを指す。感情コミットメッセージを定義するために閾値を設定し、それぞれの閾値に基づいて感情コミットメッセージの出現割合を算出する。

RoGoEmo には感情が表れているかどうかを判断するための明確な基準となる閾値がないため、独自に閾値を設定する必要がある。広い視点で分析を行うために、以下の 4 つの閾値を設定し、閾値ごとに感情コミットメッセージの割合を算出する。

- 閾値①: 全コミットメッセージにおける外れ値
- 閾値②: 0.1 以上
- 閾値③: 0.2 以上
- 閾値④: 0.5 以上

閾値①は、Senti4SD で neutral と判別されたものも含めた全てのコミットメッセージに対して RoGoEmo を適用し、得られた感情値の四分位範囲から算出できる外れ値を基に設定する相対的な指標であり、感情ラベルごとに算出する。この閾値により、基本的な分布から逸脱した、特に感情が表れているコミットメッセージを特定できる。閾値②から④は、それぞれ定数で、0.1、0.2、0.5 に設定する。閾値①から順に閾値が大きくなっており、感情値の大きさ、すなわち感情の強さにおける傾向を比較することができる。閾値ごとに感情の表れやすいコミットメッセージの分布がどのように変化するかを調査することで、コミットメッセージにおける感情の傾向を明らかにする。

4.5 RQ2 に関する分析手順

まず、本研究で使用する 4 つのプロジェクトにおいて、取得できる全てのコミットメッセージに付与されたタグを調べ、BUG タグ、FIX タグ、OTHER タグの 3 つに

表 2. 前処理によるコミットメッセージ (CM) の抽出結果

プロジェクト	除外 CM 数	取得 CM 数	Senti4SD による分析結果		
			neutral	negative	positive
Apache James	0	11,866	10,985	334	547
Apache MINA	0	2,014	1,739	145	130
Apache OpenJPA	1	4,382	3,971	154	257
Eclipse BIRT	14	25,882	23,166	1,062	1,655
合計	15	44,144	39,861	1,695	2,589

分類し、それぞれの件数を調べる。また、RoGoEmoの対象となる前処理済みコミットメッセージに対しても同様の分類と調査を行う。次に、コミットの種類によってコミットメッセージに表れる感情の傾向が異なるのかどうかを具体的に調査するために、以下の2つのパターンで比較を行う。

- BO (Bug-Others) 比較: BUG タグ付きコミットメッセージと、それ以外のコミットメッセージを比較する。
- FO (Fix-Others) 比較: FIX タグ付きコミットメッセージと、それ以外のコミットメッセージを比較する。

これらの比較では、バグが混入したコミットまたはバグを修正したコミットと、そうでない一般コミットにおける、コミットメッセージに表れる感情の傾向の違いを明らかにする。RQ1 と同じく4つの閾値 (閾値①から④) を用いて、コミットの種類における感情の表れやすさを調査する。

5 結果

4つのプロジェクトから取得したコミットメッセージに対する前処理結果を表2に示す。表2より、Senti4SDでnegativeまたはpositiveと判別された合計4,284件のコミットメッセージがRoGoEmoの分析対象となる。

調査結果は閾値ごとの棒グラフとそれらをまとめた表で示す。グラフにおける感情ラベルは、表1の感情ラベル一覧と同じ順番で並べる。そのため、グラフの左側にポジティブ (positive) な感情、その右にネガティブ (negative) な感情、右端には曖昧 (ambiguous) な感情の出現割合が表示される。

表 3. 閾値ごとのコミットメッセージへの感情出現数および割合

感情ラベル	閾値①	閾値②	閾値③	閾値④
admiration	297 (6.93%)	31 (0.72%)	27 (0.63%)	17 (0.40%)
amusement	262 (6.12%)	1 (0.02%)	1 (0.02%)	0 (0.00%)
approval	872 (20.35%)	429 (10.01%)	158 (3.69%)	21 (0.49%)
caring	1,096 (25.58%)	14 (0.33%)	3 (0.07%)	0 (0.00%)
desire	825 (19.26%)	14 (0.33%)	8 (0.10%)	0 (0.00%)
excitement	100 (2.33%)	4 (0.09%)	0 (0.00%)	0 (0.00%)
gratitude	453 (10.57%)	133 (3.10%)	129 (3.01%)	129 (3.01%)
joy	384 (8.96%)	7 (0.16%)	6 (0.14%)	2 (0.05%)
love	171 (3.99%)	6 (0.14%)	4 (0.09%)	2 (0.05%)
optimism	894 (20.87%)	38 (0.89%)	26 (0.61%)	22 (0.51%)
pride	693 (16.18%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
relief	1,251 (29.20%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
anger	408 (9.52%)	12 (0.28%)	9 (0.21%)	2 (0.05%)
annoyance	1,041 (24.30%)	240 (5.60%)	69 (1.61%)	15 (0.35%)
disappointment	1,205 (28.13%)	102 (2.38%)	39 (0.91%)	12 (0.28%)
disapproval	1,059 (24.72%)	202 (4.72%)	94 (2.19%)	19 (0.44%)
disgust	608 (14.19%)	3 (0.07%)	0 (0.00%)	0 (0.00%)
embarrassment	1,032 (24.09%)	6 (0.14%)	0 (0.00%)	0 (0.00%)
fear	120 (2.80%)	2 (0.05%)	1 (0.02%)	0 (0.00%)
grief	470 (10.97%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
nervousness	1,028 (24.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
remorse	1,293 (30.18%)	13 (0.30%)	13 (0.30%)	12 (0.28%)
sadness	813 (18.98%)	19 (0.44%)	17 (0.40%)	0 (0.00%)
confusion	368 (8.59%)	29 (0.68%)	21 (0.49%)	14 (0.33%)
curiosity	179 (4.18%)	1 (0.02%)	0 (0.00%)	0 (0.00%)
realization	843 (19.68%)	90 (2.10%)	24 (0.56%)	1 (0.02%)
surprise	152 (3.55%)	0 (0.00%)	0 (0.00%)	0 (0.00%)

5.1 RQ1 に関する分析結果

4,284件の前処理済みコミットメッセージに対してRoGoEmoを適用し、多クラス感情分析を行った結果を表3および図1に示す。図中の横軸は感情ラベル、縦軸は割合を示しており、上から順に閾値①、②、③、④における結果を示している。また、閾値①のグラフを指すときには「図1 (Outlier)」, 閾値②のグラフを指すときには「図1 (0.1)」と表記する。

5.2 RQ2 に関する分析結果

4つのオープンソースプロジェクトから収集した全てのコミットメッセージまたは前処理済みコミットメッセージをタグで分類した結果を、それぞれ表4、表5に示す。表中の割合は、合計コミットメッセージ数に対するタグごとの合計数の割合を表しており、小数第二位までとしている (小数第三位を四捨五入)。4,284件の前処理済みコミットメッセージにBO比較を行った結果を図2に、FO比較を行った結果を図3にそれぞれ示す。図中の横

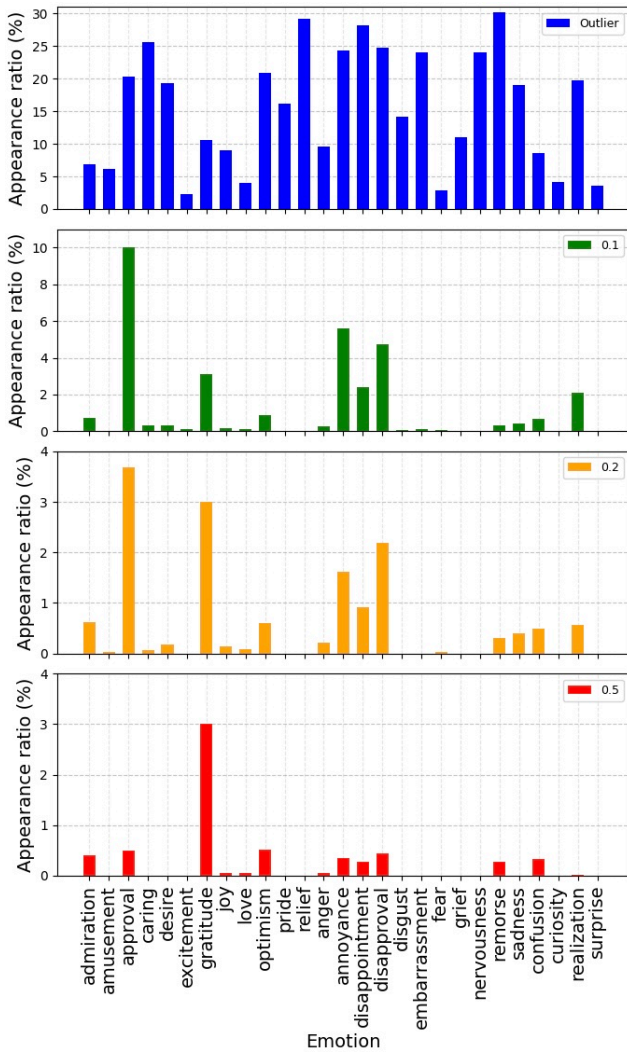


図 1. コミットメッセージに表れる感情の出現割合

軸は感情ラベル、縦軸は割合を示しており、上から順に閾値①、②、③、④における結果を示している。また、閾値①のグラフを指すときには「図 2 (Outlier)」、閾値②のグラフを指すときには「図 2 (0.1)」というように表記する。

6 考察

表 2 より、4つのプロジェクトから取得した CM に対して前処理を行った結果、そのほとんど (90.30%) がニュートラルであり、CM における一般的な感情傾向と同じであると言える。

表 4. 全てのコミットメッセージ (CM) に対するタグ分類結果

プロジェクト名	CM 数	BUG	FIX	OTHER
Apache James	11,866	1,513	751	9,787
Apache MINA	2,014	710	296	1,164
Apache OpenJPA	4,383	1,828	1,652	1,768
Eclipse BIRT	25,896	5,400	3,476	18,083
合計	44,159	9,451	6,175	30,802
(割合)		(21.4%)	(14.0%)	(69.8%)

表 5. 前処理済みコミットメッセージ (CM) に対するタグ分類結果

プロジェクト名	CM 数	BUG	FIX	OTHER
Apache James	881	112	82	704
Apache MINA	275	128	84	116
Apache OpenJPA	211	189	166	156
Eclipse BIRT	2,717	664	416	1,770
合計	4,284	1,093	748	2,746
(割合)		(25.5%)	(17.5%)	(64.1%)

6.1 RQ1: コミットメッセージに表れる感情についての考察

6.1.1 閾値①: 外れ値における考察

図 1 (Outlier) より、CM に表れる感情はネガティブに分類される感情がやや出現しやすいことが分かる。しかし、表 2 より、前処理済み CM 全体にはポジティブな感情が多い傾向があり、RoGoEmo がソフトウェア開発の分野にうまく対応できない可能性があることも考慮する必要がある。反対に、CM に表れる割合が低い感情は、具体的には、excitement, love, fear, curiosity, surprise が挙げられる。これらの感情は、ソフトウェア開発の場においてはあまり重要な役割を果たさない可能性があると考えられる。しかし、例えば fear の感情が表れた CM の内容に着目してみると、「fix 268001: Failures and errors occurs in junit test of org.eclipse.birt.report.tests.model in 2.5.0 daily build v20090311-0630」や「[268037]Manipulate signed plugin will lead to installation failure」など、バグ修正の報告や不具合箇所の報告など、無視できないものも含まれている。一概に割合だけで議論することはできないが、一つの指標として感情の出現割合を見ることで、CM の特徴を把握することができると思われる。

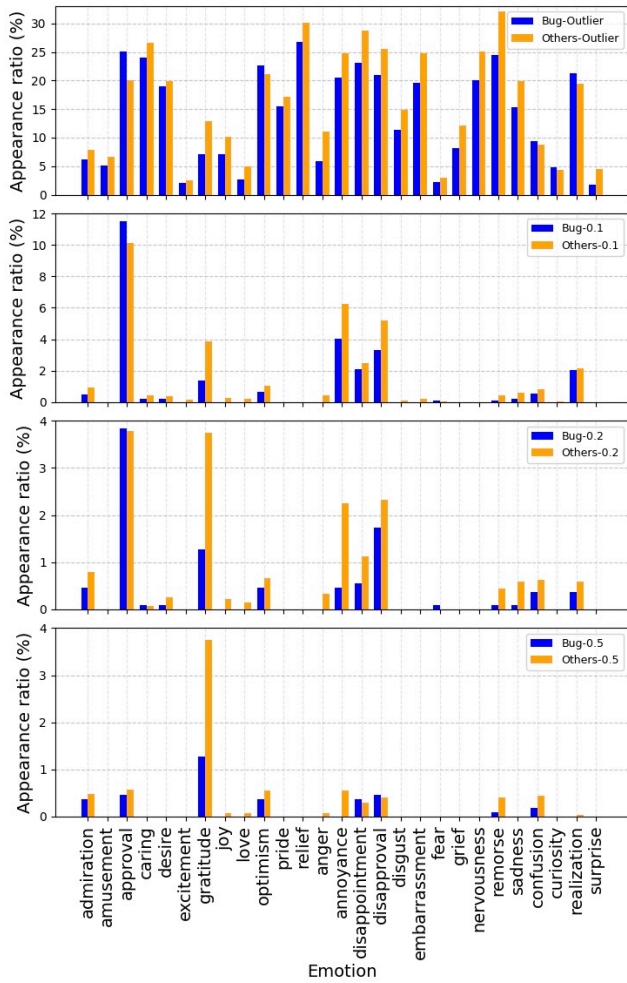


図 2. 感情コミットメッセージの BO 比較結果

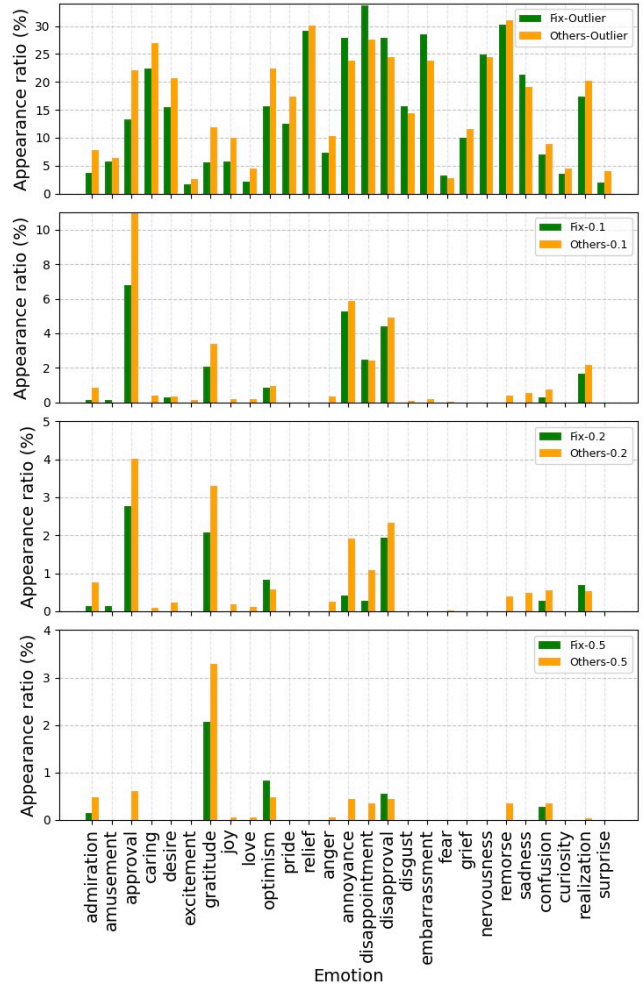


図 3. 感情コミットメッセージの FO 比較結果

6.1.2 閾値②: 0.1 (10%) における考察

図 1 (0.1) より、閾値①における分析結果とは感情の傾向が変化しており、approval がやや突出する形になっている。また、閾値①では、caring, relief, embarrassment, remorse, nervousness といった感情の割合が高かったが、大幅にその割合を減らしており、表 3 より、その他の感情もその割合を大きく減らしている。

6.1.3 閾値③: 0.2 (20%) における考察

図 1 (0.2) より、閾値②のときと大きく傾向は変わっておらず、閾値②と閾値③の間に大きな差はないと言える。

6.1.4 閾値④: 0.5 (50%) における考察

図 1 (0.5) より、CM に表れる感情の割合は gratitude が突出する形になっている。閾値④は、かなり強く表れている感情を抽出する閾値であり、gratitude の感情が突出していることは、感情の強度とその表現の明確さが非常に高いことを意味する。gratitude の突出は、開発者の積極的なコミュニケーションやチームワークの結果として理解できる。メッセージ内容を見ると、報告者やツール制作者への感謝が表現されていた。また、例えば「NioDatagramConnector works just fine thanks to my last refactoring with AbstractIoProcessor (一部抜粋)」のように、作業に有益なものに言及するメッセージにも gratitude が表れており、CM を閾値と感情で分析することがデータマイニングに繋がると考えられる。

6.1.5 RQ1における考察のまとめ

閾値を①から④へ変化させるにつれてより強い感情が強調されており、感情の強度がより明確に反映されていることが確認できる。特に、小さい閾値ではネガティブな感情が優勢であったのに対し、閾値を大きくすると approval や gratitude といったポジティブな感情が強く表れる傾向が見られた。このことから、CM 上では前向きなコミュニケーションや円滑なやり取りが行われていると考えることができる。さらに、感情と閾値で CM をフィルタリングすることによる、データマイニングの可能性が示唆された。gratitude が作業に有益だったものについて言及するメッセージに強く表れていたように、過去の開発者の行動を感情分析することで、有益なツールや情報を得られる可能性がある。

6.2 RQ2: バグ混入および修正と表れる感情についての考察

表 4 より、全 CM において、BUG タグが 21.40%、FIX タグが 13.98%、OTHER タグが 69.75% となった。また、表 5 より、前処理済み CM において、BUG タグが 25.51%、FIX タグが 17.46%、OTHER タグが 64.10% となった。これらの結果を比較すると、前処理済み CM では BUG タグと FIX タグがそれぞれ約 4% 程度大きくなっている。このことから、BUG タグおよび FIX タグが付与された CM は、全体的な傾向と比較して感情を持つ傾向が強いことが考えられる。

以降、本節では、BUG タグ付き CM を BUG、FIX タグ付き CM を FIX と表記する。

6.2.1 閾値①: 外れ値における考察

図 2 (Outlier) より、閾値①における BO 比較では、approval, optimism, confusion, curiosity, realization で、BUG の割合が大きくなった。これらの感情のうち、approval, optimism はポジティブな感情に分類され、その他の感情は曖昧な感情に分類される。図 3 (Outlier) より、閾値①における FO 比較では、annoyance, disappointment, disapproval, disgust, embarrassment, fear, nervousness, sadness で、FIX の割合が大きくなった。これらの感情は全てネガティブに分類される感情である。これらの結果から、閾値①において、BUG にはポジティブな感情が、FIX にはネガティ

ブな感情が優位に表れており、BUG と FIX において優位に表れる感情の傾向は全く異なることが分かった。一方で、「楽観的がゆえにバグが混入しやすいコードも許容した」「バグを修正する過程で負担や不満を感じた」とも考えられる。

6.2.2 閾値②: 0.1 (10%) における考察

図 2 (0.1) より、閾値②における BO 比較では、approval と fear において、BUG の割合が大きくなった。approval はポジティブ、fear はネガティブな感情に分類される。図 3 (0.1) より、閾値②における FO 比較では、amusement と disappointment において、FIX がの割合が大きくなった。amusement はポジティブ、disappointment はネガティブな感情に分類される。

これらの結果から、閾値②においても、BUG と FIX で優位な感情は異なるが、閾値①で見られた BUG と FIX の明確な傾向の差が小さくなっていることが分かる。具体的には、BUG と FIX の両方でポジティブな感情とネガティブな感情が表れており、感情の性質がより複雑になったことが確認できる。これにより、BUG と FIX の感情の違いは単純なポジティブ・ネガティブの対立ではなく、細かな感情の混合として表れることが分かる。

BUG で優位となった fear のメッセージのひとつに「restructuring for maven; adding some more utility code. This is part 1 of a two-part commit, since I didn't put enough directory structures in place just yet and svn complained and I'm an svn newbie afraid to tempt the gods of source control.」があり、自信のなさが fear として表れている。この恐れを通り、このコミットの時点では実際にバグが混入している。コミットの内容と CM に表れる感情が関係していることがここでも示された。

6.2.3 閾値③: 0.2 (20%) における考察

図 2 (0.2) より、閾値③における BO 比較では、approval, caring, fear で BUG の割合が大きくなった。approval と caring はポジティブ、fear はネガティブな感情に分類される。図 3 (0.2) より、閾値③における FO 比較では、amusement, optimism, realization でにおいて、FIX の割合が大きくなった。amusement と optimism はポジティブ、realization は曖昧な感情に分類される。これらの結果を比較すると、閾値③においても、BUG と

FIX で優位な感情が異なることが分かる。BUG では、閾値②で見られたポジティブな感情とネガティブ感情の混合が引き続き確認できるが、FIX では少し傾向に変化が見られる。閾値①ではネガティブな感情が優位だったのに対し、閾値③ではポジティブな感情が優位に表れるようになっている。また、optimism と realization は、閾値①の段階では BUG において優位な感情であった。optimism と realization の感情が表れた CM を見ると、optimism の感情が表れたコミットのうち、FIX は 8 件存在したが、その中の 5 件は BUG でもあった。realization では 8 件中 4 件が同様の状態であった。optimism と realization が表れたコミットはバグ修正コミットではあるが、同時にバグが混入している可能性があると考えられる。

6.2.4 閾値④: 0.5 (50%) における考察

図2 (0.5) より、閾値④における BO 比較で BUG の割合が大きくなった感情は、disappointment と disapproval であり、共にネガティブな感情に分類される。優位となる感情が閾値③から大きく変化した。図3 (0.5) より、閾値④における FO 比較で FIX の割合が大きくなった感情は、optimism と disapproval である。optimism はポジティブ、disapproval はネガティブな感情に分類される。

これらの結果から、閾値④においては、BUG と FIX で優位な感情が異なるという特徴がなくなったことが分かる。具体的には、disapproval が両者の共通の感情となり、BUG と FIX で共有される感情ができた。この共通の感情である disapproval を除くと、BUG ではネガティブな感情が、FIX ではポジティブな感情が優位となっており。これは閾値①における結果とは反対の傾向である。

6.2.5 RQ2 における考察のまとめ

閾値を変化させて、BUG と FIX のそれぞれに着目した比較を行うと、非常に興味深い傾向の変化が見られた。BUG に割合が大きく表れる感情は、閾値①ではポジティブな感情を主として構成されていたが、ネガティブな感情が混ざり、閾値④ではネガティブな感情のみで構成された。対して、FIX に割合が大きく表れる感情は、閾値①ではネガティブな感情のみで構成されていたが、ポジティブな感情が混ざり、閾値③ではポジティブ

な感情を主として構成された。閾値を変えて観察すると、BUG と FIX で優位に表れた感情の入れ替わりのような結果が生じることが分かった。一方で、実際のメッセージ内容を見ると、閾値②における fear や閾値③における optimism と realization のように、CM に表れる感情とコミット内容との関連が見られた。

これらの分析結果から、開発者は CM の中で感情を表現すべきであると考えられる。CM に表れる感情がバグ混入やバグ修正を示す重要な手掛かりになる可能性が十分にあると考えられる。傾向の変化や優位な感情の入れ替わりは、多クラス感情分析だからこそ見つけられたものであると考えている。多クラス分析によって、感情極性の分析だけでは捉えることができない感情の多様性が明らかになり、感情とコミット内容のより詳細な関係を見出すことができたと考えている。バグ混入コミットとその CM に表れる感情を関連づける特徴が明らかになれば、バグ混入を CM に表れる感情の種類から予測することも可能だと考える。

7. おわりに

本研究では、コミットメッセージに表れる感情とバグの混入および修正との関連を分析するため、オープンソースプロジェクトから取得したコミットメッセージを対象として、多クラス感情分析を適用した調査を行った。その結果、コミットメッセージに表れる感情は、感情値の強さに基づく閾値によって特徴的に表れる感情の傾向が変化することを示した。特に、閾値を変化させると、存在が顕著になる感情と表れなくなる感情があり、コミットメッセージにおいて注目すべき感情を抽出できることが分かった。また、バグ混入時点とその他の時点との比較、およびバグ修正時点とその他の時点との比較においても、閾値によって特徴的な感情の傾向が見られた。

しかしながら、適用した多クラス感情分析モデルのコミットメッセージに対する分析精度には課題が残る。今回適用したモデルは一般的なコメントを基に訓練されており、コミットメッセージという特定の領域においてどれだけ効果的に分析できたかは不確かである。今後、ソフトウェア開発に特化した多クラス感情分析を適用することができれば、分析精度を向上させることができ、よりの確な感情傾向の分析が期待できる。

参考文献

- [1] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver. Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 929–932, 2007.
- [2] Yili Wang, Jiakuan Guo, Chengsheng Yuan, and Baozhu Li. Sentiment analysis of twitter data. *Applied Sciences*, Vol. 12, No. 22, p. 11775, 2022.
- [3] Martin Haselmayer and Marcelo Jenny. Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & quantity*, Vol. 51, pp. 2623–2646, 2017.
- [4] Nalini Chintalapudi, Gopi Battineni, Marzio Di Canio, Getu Gamo Sagaro, and Francesco Amenta. Text mining with sentiment analysis on seafarers’ medical documents. *International Journal of Information Management Data Insights*, Vol. 1, No. 1, p. 100005, 2021.
- [5] Bo Pang, Lillian Lee, et al. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, Vol. 2, No. 1–2, pp. 1–135, 2008.
- [6] Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. Latent emotion memory for multi-label emotion classification. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34, pp. 7692–7699, 2020.
- [7] Emitza Guzman, David Azócar, and Yang Li. Sentiment analysis of commit comments in github: an empirical study. In *Proceedings of the 11th working conference on mining software repositories*, pp. 352–355, 2014.
- [8] Syed Fatiul Huq, Ali Zafar Sadiq, and Kazi Sakib. Is developer sentiment related to software bugs: An exploratory study on github commits. In *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pp. 527–531. IEEE, 2020.
- [9] Jin Ding, Hailong Sun, Xu Wang, and Xudong Liu. Entity-level sentiment analysis of issue comments. In *Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering*, pp. 7–13, 2018.
- [10] Munmun De Choudhury and Scott Counts. Understanding affect in the workplace via social media. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 303–316, 2013.
- [11] Anisha Singh, PK Kapur, and VB Singh. Developing classifiers by considering sentiment analysis of reported bugs for priority prediction. *International Journal of System Assurance Engineering and Management*, Vol. 15, No. 5, pp. 1888–1899, 2024.
- [12] Parastou Tourani, Yujuan Jiang, and Bram Adams. Monitoring sentiment in open source mailing lists: exploratory study on the apache ecosystem. In *CASCAN*, Vol. 14, pp. 34–44, 2014.
- [13] Fabio Calefato, Filippo Lanubile, Federico Maiorano, and Nicole Novielli. Sentiment polarity detection for software development. In *Proceedings of the 40th International Conference on Software Engineering*, pp. 128–128, 2018.
- [14] Sam Lowe. roberta-base-go_emotions (revision 58b6c5b), 2024.
- [15] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*, p. 15, 2020.
- [16] Jacek Śliwerski, Thomas Zimmermann, and Andreas Zeller. When do changes induce fixes? *ACM sigsoft software engineering notes*, Vol. 30, No. 4, pp. 1–5, 2005.