

感情認識の差の原因となる表現の獲得と時系列データを活用した感情分析

中川, 翼 / Nakagawa, Tsubasa

(出版者 / Publisher)

法政大学大学院理工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学研究科編

(巻 / Volume)

65

(開始ページ / Start Page)

1

(終了ページ / End Page)

8

(発行年 / Year)

2024-03-24

(URL)

<https://doi.org/10.15002/00030751>

感情認識の差の原因となる表現の獲得と 時系列データを活用した感情分析

ACQUISITION OF EXPRESSIONS CAUSING DIFFERENCES IN EMOTION RECOGNITION AND
SENTIMENT ANALYSIS USING TIME SERIES DATA

中川 翼

Tsubasa Nakagawa

指導教員 彌富 仁

法政大学大学院理工学研究科応用情報工学専攻修士課程

It is often difficult to correctly infer a writer's emotion from text exchanged online, and differences in recognition between writers and readers can be problematic. In this study, we propose a new framework for detecting sentences that create differences in emotion recognition between the writer and the reader and for detecting the kinds of expressions that cause such differences. Put into practical use, this framework would likely have the ability to mitigate problems based on misunderstandings. Furthermore, we constructed an extended dataset by collecting sentences posted within one hour before and after each of the 43,200 tweets in the WRIME dataset and conducted analyses utilizing the surrounding posts. The dataset constructed in this study enables analysis and emotion estimation utilizing the time-series data of posts on social networking service (SNS) and is expected to contribute to the field of emotion analysis.

Keywords : Natural Language Processing, Sentiment Analysis, Text Mining

1. はじめに

感情分析 [1] は与えられた文章から感情を推定する自然言語処理の主要なタスクの一つである。感情を考慮した対話システム [2] や商品レビューからの意見抽出 [3] など様々な用途がある。感情分析の代表的なタスクとして、感情極性の推定 [4] が挙げられる。与えられた文章から肯定的 (ポジティブ) であるか否定的 (ネガティブ) であるかを推定するタスクであり、単語ごとに極性情報が付与された極性辞書などを用いた手法 [5] が一般的であったが、深層学習が台頭した現在では単語が持つ意味や細かな表現を考慮した推定 [6] が可能となった。

より高度な推定が可能となったことで、感情極性 (肯定的・否定的) の推定だけでなく、感情の種類や強度の推定 [7, 8, 9, 10, 11, 12, 13] も試みられるようになった。感情の種類を考慮した推定では、Ekman の 6 つの感情 (喜び・悲しみ・驚き・怒り・恐れ・嫌悪) [14] や Plutchik の 8 つの感情 (喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼) [15] が代表的な基本感情として用いられている。相手の感情をより正確に把握するためには、このような粒度の細かい単位での推定が重要になる。一方で、機械学習を用いた感情分析において、このような多様な感情を対象にした場合、感情は人の経験や背景にも依存するため、文章のみから正確な推定を行うことは難しい。

近年では、social networking service (SNS) 等を中心

としたオンライン上での文章のやりとりの増加を受けて、X (旧 Twitter) の文章を対象にした研究 [7] も行われている。文章から相手の感情を正確に推定することが、オンライン上でのコミュニケーションでは重要である。また、文章中に含まれている感情表現の原因と思われるフレーズを抽出する emotion cause extraction (ECE) [16] や感情とその原因をペアで抽出する emotion cause pair extraction (ECPE) [17, 18] といった、感情の原因抽出に焦点を当てた研究も盛んに行われている。感情の推定だけでなく、入力された文章からその感情に対応する表現や原因を明らかにすることも求められているといえる。

これら多くの先行研究 [8, 9, 10, 11, 12] では文章を読んだ第三者が付与したラベルを用いて推定が行われてきたため、読み手の感情に沿った推定であるといえる。しかし、文章の書き手が持つ感情と読み手が受ける感情には差が存在することが指摘されている [19]。実際に、SNSをはじめとするオンライン上でのコミュニケーションの増加に伴い、書き手と読み手の認識の齟齬が問題となっている [13]。SNS における認識の齟齬を減らすためにも、より書き手の感情に沿った推定が求められる。

近年では、書き手の感情の推定を機械学習技術により実現するためのデータセットの構築 [20] が試みられている。Kajiwara ら [21] は SNS 上の文章を対象に 8 種類の感情に対して、投稿者本人の感情強度ラベルとその投稿を読んだ第三者による感情強度ラベルの両

方が付与された計 43,200 件からなる WRIME データセットを構築・公開した。彼らは bidirectional encoder representations from transformers (BERT) [22] を用いて両方のラベルの推定を行い、読み手の感情ラベルより書き手の感情ラベルの推定の方が難しいことを示した。

本研究で我々は WRIME データセットにおいて書き手と読み手の感情認識の差が大きかった「怒り」の感情に着目し、隠れ怒り文章、すなわち読み手が書き手の怒りの感情を十分に読み取れていない文章の検出を行うとともに、その差を生み出している表現の検出を試みた。我々の BERT ベースの検出モデルは、隠れ怒り文章を一定の性能で検出するとともに、検出された文章には識別器の判断材料と推定される頻出単語がいくつか見られた。一方で、それらの単語自体に怒りの感情を表す意味はないため、読み手がその感情を読み取るのは難しく、怒りの感情を持った人間が文章を書く際に無意識に用いている表現であるといえる。こうした隠れ怒りを検出できなかった文章を実際に確認してみると、SNS 特有の極端に短い文章や意味不明な文章が多く見られ、単一の投稿から書き手の感情を正確に推定するのは極めて難しい。一方で、そういった解析が困難な文章が特定の感情を持った人間が連続で投稿した中の 1 つであった場合、前後の投稿に推定する上で有用な情報や感情表現を含んでいる可能性が高い。

こうしたことから、我々は WRIME データセットに含まれている文章の前後 1 時間以内に投稿された文章を収集し、WRIME データセットを拡張した Extended WRIME データセットを構築するとともに、前後の投稿を活用した解析を行った。構築したデータセットは、追加された文章に感情ラベルは付与されておらず、元の WRIME データセットと併用することを想定しており、元のデータセットを置き換えるものではない。

本研究の主な貢献は以下の通りである。

- SNS 文書において書き手と読み手の感情認識の差の原因となる表現の獲得
- 時系列情報を含んだ今後の感情推定研究のためのデータセットの構築
- 負の感情は正の感情と比べて感情の継続性があることを示す解析結果
- 感情推定における時系列データの効果的な活用法の提案と議論

本研究で取り組んだ感情認識の差の原因となる表現の獲得で得られた知見を共有することで、SNS における認識の齟齬を減らすための一助となることが期待できる。また、本研究で構築したデータセットは SNS における文章単位の時系列データを活用した解析や感情推定を実現可能とするものであり、元の WRIME データセットとともに日本語の感情分析分野への貢献が期待できる。

表 1: 感情強度ラベルの例

| | 喜び | 悲しみ | 期待 | 驚き | 怒り | 恐れ | 嫌悪 | 信頼 |
|-------|----|-----|----|----|----|----|----|----|
| 書き手 | 0 | 1 | 2 | 1 | 1 | 0 | 0 | 1 |
| 読み手 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 読み手 2 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| 読み手 3 | 0 | 2 | 0 | 0 | 0 | 1 | 1 | 0 |

2. データセット

(1) WRIME データセット

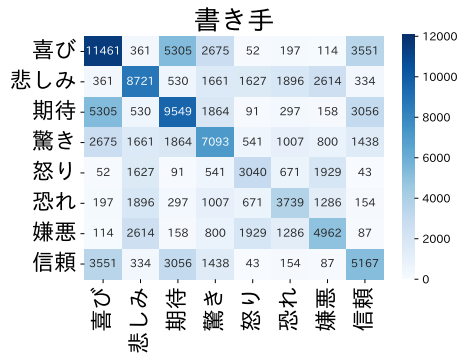
WRIME データセット^{*1}は Kajiwara ら [21] によって構築された日本語の感情分析データセットである。X (旧 Twitter) に投稿された計 43,200 件の文章に対して、投稿者本人とその内容を読んだ第三者 3 人による感情強度が 4 段階 (0 - 3) で付与されている。付与された 8 種類の感情強度は Plutchik の 8 感情 (喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼) [15] に基づく。表 1 に、付与された 8 種類の感情強度の例をそれぞれ示す。文章内に明確な感情表現がない限り、第三者である読み手が書き手の感情を正確に推定することは難しい。

我々は、この書き手と読み手の両方の感情ラベルを持つという特性が双方の感情認識の差を明らかにする上で適していると考え、このデータセットを用いて書き手と読み手のそれぞれの感情ラベルについて分析を行った。図 1 に、書き手と読み手のそれぞれにおける、感情強度が 2 以上である感情ラベルの共起行列を示す。書き手の感情強度は投稿者本人による感情ラベルであり、読み手の感情強度は第三者 3 人による感情ラベルを平均したものである。書き手による感情ラベルの共起行列 1a と読み手による感情ラベルの共起行列 1b を比べると、読み手は文章から一つの感情に絞って読み取る傾向にあるのに対し、書き手は複数の感情が共起する傾向にあり、より複雑であることが分かる。また、各感情のラベル数に着目すると、読み手は書き手の感情を十分に読み取れず、過小評価する傾向にある。特に「怒り」と「信頼」の感情においてその傾向は顕著である。具体的には、怒りのラベルでは書き手が 3,040 件であるのに対し、読み手は 543 件しか認識できていない。信頼のラベルについても同様に書き手が 5,167 件であるのに対し、読み手は 185 件しか認識できていないため、読み手は書き手が持つ「怒り」や「信頼」の感情を十分に読み取れていないと言える。

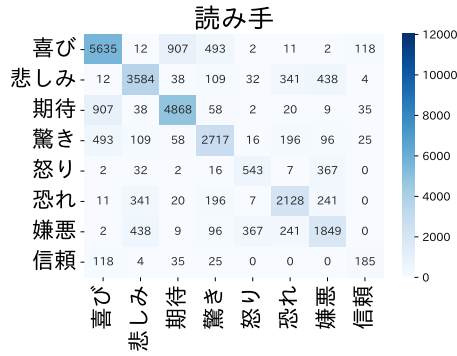
(2) Extended WRIME データセット

WRIME データセットに含まれている文章を実際に確認してみると、SNS 特有の極端に短い文章や意味不明な文章がいくつか見られ、単一の投稿から書き手の感情を正確に推定するのは極めて難しい。こうしたことから、我々は学術研究用の Twitter API を用いて WRIME データセットに含まれている文章の前後 1 時間以内に

^{*1} <https://github.com/ids-cv/wrime>



(a) 書き手による感情ラベルの共起行列



(b) 読み手による感情ラベルの共起行列

図 1: 書き手と読み手のそれぞれにおける感情ラベルの共起行列

投稿された文章を収集し, WRIME データセットを拡張した Extended WRIME データセットを構築した. ただし, Extended WRIME データセットは, 追加された文章に感情ラベルは付与されておらず, 元の WRIME データセットと併用することを想定しており, 元のデータセットを置き換えるものではない. 本研究で構築したデータセットは SNS における文章単位の時系列データを活用した解析や感情推定を実現可能とするものであり, 元の WRIME データセットとともに日本語の感情分析分野への貢献が期待できる.

Extended WRIME データセットの内訳を表 2 に示す. 元となる WRIME データセットには計 43,200 件の文章が含まれているが, その内 Twitter API を用いて当該投稿を特定できたのは計 15,816 件であり, Extended WRIME データセットは, これらを対象に同じ投稿者が前後 1 時間以内に投稿した文章を Twitter API を用いて収集した計 46,986 件の投稿と, 各投稿の投稿時間, 投稿者のユーザー ID やユーザーネームといった付随データから構成される.

表 3 は Extended WRIME データセットに含まれている, 同じユーザーによって連続で投稿された文章群の例である. Extended WRIME データセットで追加された文章は元の文章が投稿された背景を明らかにするものであり, 書き手の感情を推定する上で有用であると考えられる.

表 2: Extended WRIME データセットの内訳

| ツイート数 | |
|-------|--------|
| 過去の投稿 | 23,116 |
| 元の投稿 | 15,816 |
| 未来の投稿 | 23,870 |
| Total | 62,802 |

3. 提案手法

(1) 感情認識の差の原因となる表現の獲得

データセットの分析結果から, 本研究では書き手と読み手の感情認識の差が大きい「怒り」の感情に焦点を当て, その差の原因となり得る表現を検出するための手法を提案する. 書き手の怒りの感情強度が読み手の強度を 2 以上上回っている文章を第三者が怒りの感情を推定しにくい「隠れ怒り文章」と定義し, それらの文章を解析することで書き手と読み手の感情認識の差の原因となり得る「隠れ怒り表現」の検出を試みる. ここで, WRIME データセットのような SNS 文書には表 4 に示すような感情の推定が極めて難しい文章が少なからず存在する. このような文章はノイズとなり, 正確な分析を阻害するため, モデルが識別可能な文章のみに絞って解析することで, ノイズとなる文章を事前に除去する. したがって, 提案手法は (a) 隠れ怒り文章の予測, (b) 隠れ怒り表現の検出の 2 段階で構成される.

(a) 隠れ怒り文章の予測

第 1 段階では, BERT [22] をベースとするモデルを使用し, 隠れ怒り文章の検出器を構築する. 入力された文章が隠れ怒り文章であるかどうかを予測する 2 値分類を行う.

(b) 隠れ怒り表現の検出

第 2 段階では, 第 1 段階で検出された隠れ怒り文章を解析して, 書き手と読み手の感情認識の差の原因となり得る隠れ怒り表現を検出する. 隠れ怒り文章とそれ以外の文章における出現頻度の差が大きい上位 10 単語を抽出する. ここでの出現頻度は, 対象の単語の出現回数を各文章群における全単語の出現回数の和で割った値である.

(2) 文章単位の時系列データを活用した感情推定

Extended WRIME データセットが持つ文章単位の時系列データの活用した感情推定手法を提案する. 以下の 6 つの手法を用いて, 活用方の違いによる感情推定性能の比較を行う.

(i) only original tweets (baseline)

この手法では Extended WRIME データセットで追加された文章は使用せず, WRIME データセットに含まれている文章のみで学習・推定を行う. 過去の投稿を考慮した手法との比較に用いる.

(ii) concatenate with [SEP] token (concat)

表 3: 同じユーザーによって連続で投稿された文章群の例

| | 投稿時間 (UTC) | 文章 |
|-------|------------------|--|
| 元の投稿 | 2020/09/27 01:36 | ええ……………芸能界で何が起きているんだ…… |
| 過去の投稿 | 2020/09/27 01:34 | え! ???、!!!! |
| 過去の投稿 | 2022/09/27 01:34 | RT @YahooNewsTopics: 【女優の竹内結子さん死亡 40 歳】 https://t.co/kM7TNkYqt7 女優の竹内結さんが死亡したことが 27 日、分かった。40 歳。都内の自宅で遺体で見つかった。 |

表 4: 感情の推定が極めて難しい文章の例

| 文章 | 書き手 | 読み手 1 | 読み手 2 | 読み手 3 |
|--|-----|-------|-------|-------|
| 雨の日 1 日 1 回は見るよねこれ | 3 | 0 | 0 | 1 |
| あーー | 2 | 0 | 0 | 0 |
| ア ッ ヒ ョ ヒ ョ ! ! ! フ ァ????????!!!!!! みたいな鳴き声しとる | 3 | 0 | 0 | 0 |

過去の投稿と元の投稿を結合し 1 つの文章として学習・推定を行う。BERT [22] には入力できる単語数に上限があるため、入力できる投稿の数には制限がある。

(iii) vector average (average)

過去の投稿と元の投稿を 1 つずつモデルに入力することで投稿ごとの文章ベクトルを取得し、それらの平均から学習・推定を行う。入力できる投稿の数に制限はなく、過去の投稿も元の投稿も同等に扱われる。

(iv) time weighted average (time-weighted)

手法 (iii) と同様に投稿ごとの文章ベクトルを取得し、それらの平均から学習・推定を行う。ここでの平均は、元の投稿との投稿時間の差が少ない、すなわちより直近の投稿であるほど重みが大きくなるよう以下の式で重み付けされた加重平均である。

$$w(t) = e^{-\alpha|t|} \quad (1)$$

t は元の投稿からの時間差、 α はハイパーパラメータである。入力できる投稿の数に制限はなく、元の投稿との投稿時間の差が考慮される。

(v) sentence-level attention weighted average (attention-weighted)

手法 (iii) と同様に投稿ごとの文章ベクトルを取得し、それらの平均から学習・推定を行う。ここでの平均は、文章単位の Self-Attention [23] を算出し、元の投稿から見た attention weight で重み付けされた加重平均である。入力できる投稿の数に制限はなく、元の投稿との関連度が考慮される。

(vi) time-weighted × attention-weighted (pro-

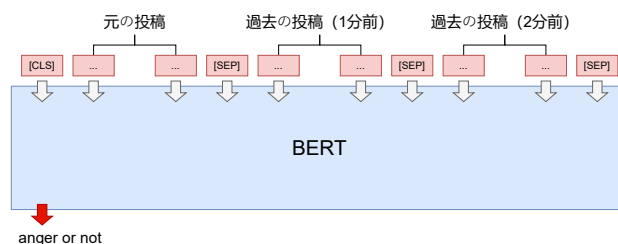


図 2: 手法 (ii) の概要

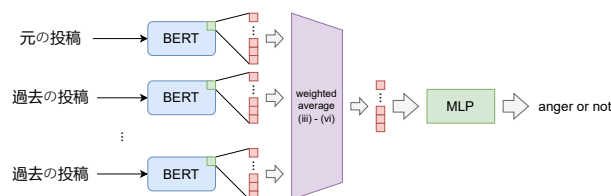


図 3: 手法 (iii) - (vi) の概要

posed)

手法 (iii) と同様に投稿ごとの文章ベクトルを取得し、それらの平均から学習・推定を行う。ここでの平均は、手法 (iv) の元の投稿との投稿時間の差を考慮した重みと手法 (v) の元の投稿から見た attention weight を掛け合わせたもので重み付けされた加重平均である。入力できる投稿の数に制限はなく、元の投稿との投稿時間の差と関連度の両方が考慮される。

手法 (ii) の概要を図 2 に、手法 (iii) から (vi) の概要を図 3 にそれぞれ示す。

4. 実験

(1) 怒りの感情認識の差を生み出している表現の検出

書き手の怒りの感情強度が読み手の強度を 2 以上上回っている隠れ怒り文章を検出する 2 値分類を行った。感情強度の差を求める際、書き手の強度が読み手の強度を上回るケースと読み手の強度が書き手の強度を上回るケースの 2 つの可能性が考えられるが、怒りの感情ラベルのみを見た場合、後者のケースはほぼ見られなかったため、今回は考慮しなかった。評価指標には receiver operating characteristic (ROC) 曲線下の面積である area under the curve (AUC) を用いた。

検出された隠れ怒り文章を解析して、書き手と読み手の感情認識の差の原因となり得る隠れ怒り表現を検出した。隠れ怒り文章とそれ以外の文章における出現頻度の差が大きい上位 10 単語を抽出した。ここでの出現頻度は、対象の単語の出現回数を各文章群における全単語の出現回数之和で割った値である。

(2) SNS 文書における時間的な継続性の検証

本研究で我々が構築した Extended WRIME データセットは、注目する単一の投稿情報のみならず、過去の投稿情報との併用が書き手の感情予測に効果的であること、つまり投稿者の感情には時間的な継続性があり、それが投稿に観測可能な形で現れているという前提のもと構築されている。この前提の妥当性を検証するため、Extended WRIME データセットに含まれている過去の投稿を用いて WRIME データセットに付与されている書き手の感情ラベル、すなわち投稿者の感情がどの程度前から発生したものなのかを分析した。過去の投稿には感情ラベルは付与されていないため、感情ラベルを持つ元の文章のみで学習した BERT [22] ベースのモデルで過去の投稿の感情をそれぞれ推定し、擬似ラベルを付与した。元の投稿と同じ感情が検出された過去の投稿の数を、元の投稿との投稿時間の差で分けて集計し、その割合を求めることで 1 分ごとの感情の検出率を求め、各感情に時間的な継続性があるのかを確認した。

(3) 時系列データを活用した感情推定

(2) に示した 6 つの手法で WRIME データセットが持つ書き手の感情ラベルを推定し、Extended WRIME データセットが持つ文章単位の時系列データの活用方法の違いによる感情推定性能の比較を行った。Extended WRIME データセットには元の文章の前後 1 時間以内に投稿された過去と未来の両方の文章が含まれているが、本実験では実用性を考慮した上で過去に投稿された文章のみを用いた。評価指標には ROC 曲線下の面積である AUC を用いた。手法 (iv) time-weighted と手法 (vi) proposed で用いるハイパーパラメータ α は各 fold の検証データでの AUC が最大となるように 0 から 1 の範囲でパラメータチューニングを行った。最適化フレームワークには Optuna^{*2} [24] を使用し、(tree-structured parzen estimator) TPE [25] によるベイズ

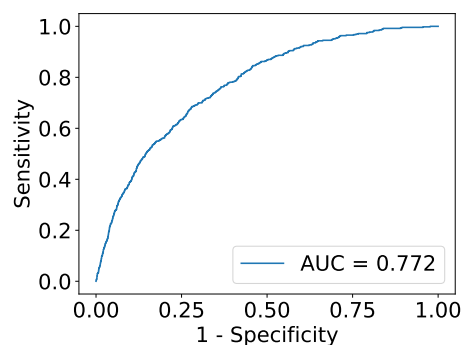


図 4: 隠れ怒り文章を検出する 2 値分類の ROC 曲線

最適化を行った。

(4) 実験設定

全体の実験を通して、計 43,200 件の WRIME データセットによる 5 fold のクロスバリデーションで学習・評価を行った。このとき、モデルの汎化性能をより正確に評価するため、WRIME データセットに含まれている投稿者のユーザー ID が各 fold で被らないよう分割した。また、BERT の事前学習モデルには、日本語の SNS テキストで事前学習されたモデル [26] を使用し、モデルの実装には Transformers^{*3} [27] を用いた。損失関数には交差エントロピー誤差、最適化手法に Adam [28] を使用し、バッチサイズは 32、ドロップアウト率は 0.1、学習率は $2e^{-5}$ 、学習回数は 3 回とした。また、クラス間の偏りの影響が減るよう損失関数への重み付けとオーバーサンプリングを行った。

5. 結果と考察

(1) 怒りの感情認識の差を生み出している表現の検出結果

図 4 に、隠れ怒り文章を予測する 2 値分類の ROC 曲線とその曲線下の面積である AUC を示す。隠れ怒り文章を予測する 2 値分類の AUC が 0.772 (検出感度が 80.1% のとき、特異度が 58.6%) であることから、隠れ怒り文章の検出器として一定の効果があることが確認できた。

図 5 に、ラベル情報のみで検出された隠れ怒り表現と提案手法で検出された隠れ怒り表現のそれぞれ上位 10 単語を示す。横軸は隠れ怒り文章とそれ以外の文章における出現頻度の差であり、隠れ怒り表現としての強さを表している。提案手法では検出器が識別可能な文章のみに絞り解析したことで、出現頻度の差がより明確に見られるようになった。また、解析する上で不適切な文章を事前に除くことでラベル情報のみでは検出することができなかった単語を検出することができた。

(2) SNS 文書における時間的な継続性の検証結果

図 6 に、元の投稿が投稿される 10 分前からの各感情の検出率の推移を示す。図 6a は「喜び」「期待」「驚き」「信頼」といった比較的ポジティブな感情の推移を表

^{*2} <https://github.com/optuna/optuna>

^{*3} <https://github.com/huggingface/transformers>

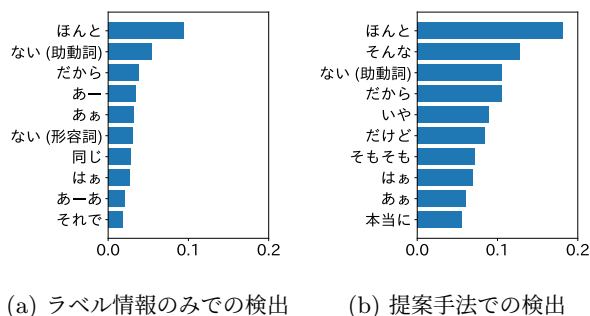


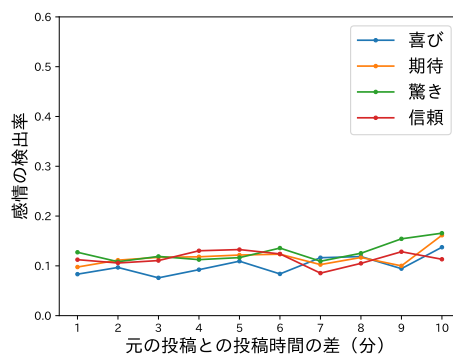
図 5: 検出された隠れ怒り表現（上位 10 単語）

し、Fig. 6b は「悲しみ」「怒り」「恐れ」「嫌悪」といったネガティブな感情の推移を表す。図 6a より、過去 10 分以内の検出率が約 10 %でほぼ一定であり、比較的ポジティブな感情の時間的な継続性は低いといえる。一方で、図 6b を見ると全体を通して図 6a より検出率が高く、元の投稿との投稿時間の差が少ないほど検出率が高い傾向にあることから、ネガティブな感情の方が時間的な継続性は高く、特に直近の投稿には書き手の感情を推定する上で有用な情報や感情表現が含まれている可能性が高いといえる。

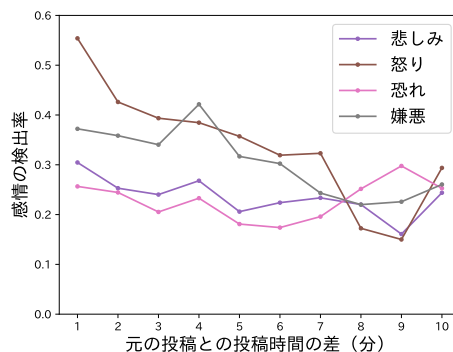
この検証結果は、過去の感情をモデルによって推定した結果に基づくという制限があるが、感情の継続性について一定の知見を与える結果を示しており、これは本研究で構築した Extended WRIME データセットの意義を支持するものである。Web コンテンツなどの負の感情を持つ記事を読者が継続的に見続けてしまう doomscrolling [29] は近年知られるようになった現象であり、読み手の健康に悪影響を与えることなどが近年指摘されている [30]。一方、負の感情の方が発信の継続性が高いという結果は、著者らの知る限り新しい知見である。これは、負の感情に対しては、受信側も発信者も執着しやすいという共通の傾向があることを示しており、興味深い結果といえる。

(3) 時系列データを活用した感情推定モデルの性能

表 5 に、書き手が持つ「喜び」「怒り」「恐れ」の感情を検出する二値分類の AUC をそれぞれ示す。 α は元の投稿との投稿時間の差を考慮した重み付けをする際に用いるハイパーパラメータである。Extended WRIME データセットを用いた、元の投稿からの投稿時間の差を考慮した手法 (iv) time-weighted, 関連度を考慮した手法 (v) attention-weighted, それらの両方を考慮した手法 (vi) proposed では、元の WRIME データセットのみを用いた手法 (i) baseline の結果と比べていずれも一定の精度向上が確認できた。一方で、過去の投稿を単純に連結する手法 (ii) concat では、それを行わない場合と比べ有意な差が見られず、過去の投稿と元の投稿を同等に扱う手法 (iii) average では、手法 (i) baseline の結果と比べて、感情の種類によらず精度が落ちている。手法 (iv) - (vi) について、感情ごとの結果の違いに着目すると、「怒り」や「恐れ」といったネガティブな感情の方が「喜び」といったポジティブな感情より精度向上



(a) ポジティブな感情



(b) ネガティブな感情

図 6: 過去の投稿における感情の検出率の推移

の度合いが大きい傾向が見られた。これは (2) の検証結果にてネガティブな感情の方が、関連する内容の文章が連続で投稿される傾向にあったことから、Extended WRIME データセットに含まれている過去の投稿を考慮した推定はそういったネガティブな感情においてより効果的であると考えられる。

6. 結論

本研究では書き手と読み手の両方による 8 種類の感情強度が付与された文章から、感情認識の差が大きい「怒り」の感情に焦点を当て、BERT [22] を用いてその差が大きい文章の予測を行うとともに、その差を生み出していると思われる表現の検出を試みた。BERT ベースの検出器を用いて文脈を考慮した分析を行うことで、「ほんと」や「そんな」といった、書き手と読み手の怒りの感情認識の差の原因となり得る表現を検出することができた。

さらに、学術研究用の Twitter API を用いて WRIME データセットに含まれている文章の前後 1 時間以内に投稿された文章を収集し、WRIME データセットを拡張した Extended WRIME データセットを構築するとともに、前後の投稿を活用した解析を行った。分析から、ネガティブな感情を含む文章の方がポジティブなものより連続で投稿される傾向にあり、投稿者の感情に時間的な継続性が見られた。Extended WRIME データセットの過去の投稿も用いた感情推定では、投稿時間の差や文章同士の関連度を考慮することで感情推定性能の向上

表 5: 書き手が持つ感情を検出する二値分類の AUC

| (a) 喜び | | | | | | |
|---|--------------|--------------|--------------|--------------|--------------|--------------|
| | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Mean |
| (i) baseline | 0.843 | 0.824 | 0.826 | 0.844 | 0.825 | 0.832 |
| (ii) concat | 0.843 | 0.817 | 0.824 | 0.844 | 0.849 | 0.835 |
| (iii) average | 0.825 | 0.807 | 0.800 | 0.802 | 0.828 | 0.812 |
| (iv) time-weighted ($\alpha = 0.731$) | 0.839 | 0.824 | 0.826 | 0.838 | 0.855 | 0.836 |
| (v) attention-weighted | 0.830 | 0.820 | 0.829 | 0.841 | 0.851 | 0.834 |
| (vi) proposed ($\alpha = 0.601$) | 0.837 | 0.814 | 0.836 | 0.844 | 0.851 | 0.837 |

| (b) 怒り | | | | | | |
|---|--------------|--------------|--------------|--------------|--------------|--------------|
| | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Mean |
| (i) baseline | 0.835 | 0.811 | 0.772 | 0.806 | 0.715 | 0.788 |
| (ii) concat | 0.830 | 0.815 | 0.782 | 0.808 | 0.726 | 0.792 |
| (iii) average | 0.800 | 0.714 | 0.772 | 0.794 | 0.694 | 0.755 |
| (iv) time-weighted ($\alpha = 0.950$) | 0.838 | 0.812 | 0.779 | 0.815 | 0.728 | 0.795 |
| (v) attention-weighted | 0.833 | 0.823 | 0.779 | 0.805 | 0.738 | 0.796 |
| (vi) proposed ($\alpha = 0.601$) | 0.839 | 0.821 | 0.795 | 0.827 | 0.737 | 0.804 |

| (c) 恐れ | | | | | | |
|---|--------------|--------------|--------------|--------------|--------------|--------------|
| | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 | Mean |
| (i) baseline | 0.757 | 0.763 | 0.755 | 0.760 | 0.740 | 0.755 |
| (ii) concat | 0.764 | 0.742 | 0.762 | 0.750 | 0.734 | 0.750 |
| (iii) average | 0.747 | 0.737 | 0.743 | 0.734 | 0.722 | 0.737 |
| (iv) time-weighted ($\alpha = 0.866$) | 0.763 | 0.757 | 0.784 | 0.758 | 0.746 | 0.762 |
| (v) attention-weighted | 0.766 | 0.748 | 0.772 | 0.771 | 0.748 | 0.761 |
| (vi) proposed ($\alpha = 0.731$) | 0.774 | 0.755 | 0.780 | 0.770 | 0.748 | 0.765 |

できることが示唆された。本研究で構築した Extended WRIME データセットは日本語 SNS における文章単位の時系列データを活用した解析や感情推定を実現可能とするものであり、感情分析分野への貢献が期待できる。

謝辞

研究で日々ご指導くださいました彌富仁教授, 研究の助言を数多くいただきました北田俊輔先輩, 並びに知的情報処理研究室の皆様深く御礼申し上げます。

参考文献

- [1] L. Zhang and B. Liu, “Sentiment analysis and opinion mining,” in *Synthesis Lectures on Human Language Technologies*, 2012.
- [2] M. Firdaus, H. Chauhan, A. Ekbal, and P. Bhattacharyya, “EmoSen: Generating sentiment and emotion controlled responses in a multimodal dialogue system,” *IEEE Transactions on Affective Computing*, vol. 13, pp. 1555–1566, 2022.
- [3] X. Fang and J. Z. Zhan, “Sentiment analysis using product review data,” *Journal of Big Data*, vol. 2, pp. 1–14, 2015.
- [4] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” in *Conference on Empirical Methods in Natural Language Processing*, 2013.
- [5] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity in phrase-level sentiment analysis,” in *Human Language Technology - The Baltic Perspective*, 2005.
- [6] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, 2018.
- [7] L. A. M. Bostan and R. Klinger, “An analysis of annotated corpora for emotion classification in text,” in *International Conference on Computational Linguistics*, 2018.
- [8] S. Aman and S. Szpakowicz, “Identifying expres-

- sions of emotion in text,” in *International Conference on Text, Speech and Dialogue*, 2007.
- [9] S. M. Mohammad and F. Bravo-Marquez, “Emotion intensities in tweets,” 2017.
- [10] S. M. Mohammad and S. Kiritchenko, “Understanding emotions: A dataset of tweets to study interactions between affect categories,” in *International Conference on Language Resources and Evaluation*, 2018.
- [11] L. A. M. Bostan, E. Kim, and R. Klinger, “Good-NewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception,” 2020.
- [12] C. Strapparava and R. Mihalcea, “SemEval-2007 task 14: Affective text,” in *International Workshop on Semantic Evaluation*, 2007.
- [13] C. Yang, K. H.-Y. Lin, and H.-H. Chen, “Writer meets reader: Emotion analysis of social media from both the writer’s and reader’s perspectives,” *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, pp. 287–290, 2009.
- [14] P. Ekman, “An argument for basic emotions,” *Cognition & Emotion*, vol. 6, pp. 169–200, 1992.
- [15] R. Plutchik, “A general psychoevolutionary theory of emotion,” in *Theories of emotion*, 1980.
- [16] Y. Chen, S. Y.-M. Lee, S. Li, and C.-R. Huang, “Emotion cause detection with linguistic constructions,” in *International Conference on Computational Linguistics*, 2010.
- [17] R. Xia and Z. Ding, “Emotion-cause pair extraction: A new task to emotion analysis in texts,” in *Annual Meeting of the Association for Computational Linguistics*, 2019.
- [18] W. Li, V. Pandelea, M. Ge, L. Zhu, and E. Cambria, “Espec: Emotion-cause pair extraction in conversations,” *IEEE Transactions on Affective Computing*, vol. 14, pp. 1754–1765, 2023.
- [19] S. Buechel and U. Hahn, “Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation,” in *Proceedings of the 11th Linguistic Annotation Workshop*, 2017.
- [20] S. Buechel and U. Hahn, “EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis,” in *Conference of the European Chapter of the Association for Computational Linguistics*, 2017.
- [21] T. Kajiwara, C. Chu, N. Takemura, Y. Nakashima, and H. Nagahara, “WRIME: A new dataset for emotional intensity estimation with subjective and objective annotations,” in *North American Chapter of the Association for Computational Linguistics*, 2021.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *North American Chapter of the Association for Computational Linguistics*, 2019.
- [23] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Neural Information Processing Systems*, 2017.
- [24] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
- [25] S. Watanabe, “Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance,” *ArXiv*, vol. abs/2304.11127, 2023.
- [26] Sakaki, Takeshi, S. Mizuki, and N. Gunji, “BERT pre-trained model trained on large-scale japanese social media corpus,” 2019.
- [27] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, “Transformers: State-of-the-art natural language processing,” in *Conference on Empirical Methods in Natural Language Processing*, 2019.
- [28] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2014.
- [29] R. Jennings, “Doomscrolling, explained,” 2020.
- [30] A. Watercutter, “Doomscrolling is slowly eroding your mental health,” 2020.