書き手と読み手の怒りの感情認識の差の原因となる表現の獲得

中川 翼 [1]　北田 俊輔 [2]　彌冨 仁 [1,2]
[1] 法政大学理工学部応用情報工学科　[2] 法政大学理工学研究科応用情報工学専攻
{tsubasa.nakagawa.5p, shunsuke.kitada.8y}@stu.hosei.ac.jp
iyatomi@hosei.ac.jp

概要

SNS 等，オンライン上でやりとりされる文章から相手の感情を正しく推定するのは時に難しく，書き手と読み手の間での認識・解釈の違いはトラブルにつながる可能性がある．本研究では，書き手と読み手の感情認識の差に着目した上で，感情認識の差が特に大きい「怒り」の感情に焦点を当て，その差が大きい文章を予測する識別器を構築するとともに，感情認識の差を生み出している表現を明らかにするための分析を行った．検出された表現は，それ自体に怒りの感情を表す意味は持たないため，読み手が読み取るのは難しいが，書いた本人は確かに怒りの感情を持つ傾向にあった．

1　はじめに

感情分析は，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 [1]．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．さらに 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，特に Twitter の文章を対象に感情極性を考慮した研究 [4] も盛んに行われている．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．

近年では，感情極性（肯定的・否定的）の推定 [5] だけでなく，感情の種類や強度の推定 [6] も試みられている．感情の種類を考慮した推定では，Ekman の 6 感情（喜び・悲しみ・驚き・怒り・恐れ・嫌悪）[7] や Plutchik の 8 感情（喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼）[8] 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．

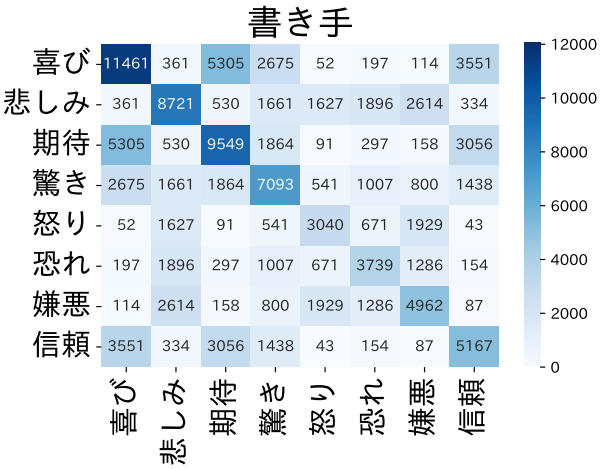
これらの多くの先行研究では文章の読み手が付与したラベルを用いて読み手の感情に沿った推定が行われてきたが，文章の書き手が持つ感情と読み手が受ける感情には差が存在すると考えられる．Kajiwara ら [9] は SNS 上の文章を対象にテキストの書き手による主観的な感情強度ラベルとテキストの読み手による客観的な感情強度ラベルの両方を収集したデータセット WRIME を構築・公開した．彼らは書き手と読み手の感情強度の比較を行い，怒りや信頼の感情を中心に，読み手は書き手の感情を十分に読み取れず，過小評価する傾向にあると述べている．我々は，こうした書き手と読み手の感情認識の差を生み出す表現を明らかにすることが相手の感情をより正確に推定する上で重要になると考えた．

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，今回は感情認識の差が大きい「怒り」の感情に焦点を当て，BERT [10] 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．検出された単語は，それ自体に怒りの感情を表す意味は持たないため，読み手が読み取るのは難しいが，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言える．

2　データセットと実験手法

本研究で我々は，書き手と読み手の感情強度が付与された WRIME データセット[1)] を用いて書き手と読み手の感情認識の差の分析を行った．次にその分析結果から，差の大きい「怒り」の感情に着目し，その差が大きい文章を予測する識別器を構築した．さらに，その差を生み出している表現の検出を試みた．

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

2.1 データセットとその分析

WRIME データセットは Kajiwara ら [9] 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．SNS に投稿された計 43,200 件の文章に対して，その投稿者本人とその内容を読んだ第三者 3 人による 8 種類の感情強度が 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 [8] に基づく．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．

2.2 隠れ怒り文章の予測モデル

上述の分析結果から，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，その差が大きい文章を予測する識別器を構築した．我々は，4 段階（0 ~ 3）で付与された怒りの感情強度において，書き手の強度が読み手の強度を 2 以上上回っている文章は，読み手が書き手の怒りを十分に読み取れておらず，第三者が怒りの感情を推定しにくい “隠れ怒り” 文章であると考えた．計 43,200 件の WRIME データセットを 4 : 1 の割合で訓練用と評価用に分割し，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．

“隠れ怒り” 文章の予測モデルには，BERT [10] をベースとするモデルを用いた．日本語 Wikipedia で事前学習された WholeWordMasking モデル[2] を使用し，モデルの実装には Transformers[3] [11] を用いた．BERT の最終層の [CLS] トークンから得られた 768 次元の分散表現を全結合層に入力し，“隠れ怒り” 文章を検出する 2 値分類を行った．損失関数に交差エ

2) https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking
3) https://github.com/huggingface/transformers

表 1: “隠れ怒り” 文章を検出する 2 値分類の性能

Precision	Recall	F1 score	文章の数
0.139	0.623	0.228	493

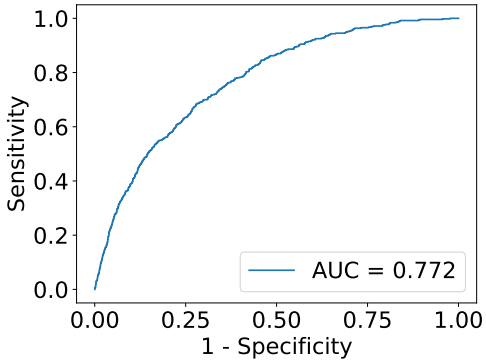


図 2: “隠れ怒り” 文章を検出する 2 値分類の ROC 曲線

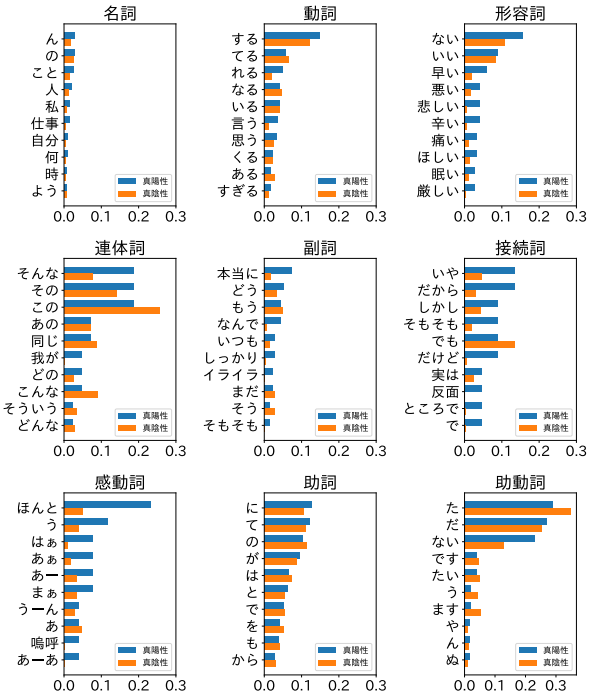


図 3: “隠れ怒り” 文章の予測で真陽性であった文章における頻出単語（上位 10 単語）

ントロピー誤差，最適化手法に Adam [12] 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，計 34,560 件の訓練用データのうち，検出対象である “隠れ怒り” 文章が 1,910 件と全体と比べて少ないため，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．評価指標には，ROC 曲線下の面積である AUC を用いた．

2.3 怒りの感情認識の差を生み出している単語の検出方法

“隠れ怒り” 文章の予測で真陽性であった文章において出現頻度が高い上位 10 単語を品詞別で求めた．次に，求めた上位 10 単語を対象に真陽性であった文章における出現頻度と真陰性であった文章における出現頻度の両方を求め，真陽性での出現頻度が真陰性での出現頻度を特に大きく上回っている単語を抽出した．ここでの出現頻度は，対象の単語の出現回数を真陽（陰）性の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．さらに，抽出された単語を含む文章における書き手と読み手 3 人の怒りの感情強度の平均をそれぞれ求め，抽出された単語が書き手と読み手の怒りの感情認識に差を生み出しているか検証した．

3 実験結果

3.1 隠れ怒り文章の予測モデルの性能

表 1 に，“隠れ怒り” 文章を検出する 2 値分類の性能を示す．計 8,640 件の評価用データのうち，検出対象である “隠れ怒り” 文章が 493 件とデータ数に偏りがあるが，“隠れ怒り” 文章を一定数検出できていると言える．

図 2 に，“隠れ怒り” 文章を検出する 2 値分類の ROC 曲線とその曲線下の面積である AUC を示す．AUC が 0.771（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．

3.2 怒りの感情認識の差を生み出している単語の検出結果

図 3 に，“隠れ怒り” 文章の予測で真陽性であった文章において出現頻度の高かった上位 10 単語を品詞別で求めたものを示す．横軸は真陽性と真陰性のそれぞれの文章における各単語の出現頻度を表している．真陽性での出現頻度が真陰性での出現頻度を大きく上回っている単語は，書き手と読み手の怒りの感情認識に差を生み出している表現であると言える．名詞や助詞には出現頻度の差はほぼ見られないが，接続詞や感動詞には真陽性での出現頻度が真陰性での出現頻度を大きく上回っている単語がいく

表 2: 検出された単語とそれらを含む文章での怒りラベルの平均

単語（品詞）	書き手	読み手 1	読み手 2	読み手 3	差
ほんと（感動詞）	0.731	0.161	0.218	0.207	0.535
そんな（連体詞）	0.415	0.071	0.162	0.099	0.304
だから（接続詞）	0.587	0.198	0.215	0.157	0.397
ない（助動詞）	0.351	0.083	0.092	0.087	0.264
いや（接続詞）	0.365	0.066	0.131	0.109	0.263
だけど（接続詞）	0.789	0.211	0.105	0.000	0.684
全体	0.234	0.047	0.057	0.051	0.182

つか見られる．

表 2 に，図 3 の頻出単語の中で真陽性での出現頻度が真陰性での出現頻度を特に大きく上回っている単語とその単語を含む文章での怒りラベルの平均を示す．差の欄の数値は，書き手の感情ラベルの平均から読み手 3 人の感情ラベルの平均を差し引いた値であり，この値が大きいほど書き手と読み手の怒りの感情認識に差があると言える．検出された単語を含む文章での差の方がデータセット全体での差よりも大きい．よって，検出された単語は，それ自体に怒りの感情を表す意味は持たないため，読み手が読み取るのは難しいが，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言える．

検出された「ほんと」や「そんな」といった単語は，あとに続く単語の意味を強調するために用いられることが多く，文脈次第で意味が変化する．データセットの文章とラベルのみを用いた分析では，こうした単語同士の関係性を考慮した分析は難しい．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，第三者が読み取りにくい“隠れ怒り”表現と思われる単語を検出することができたと考えられる．

4 おわりに

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 [10] 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．

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ったため，今後は複数の感情を考慮し，より書き手の感情に沿った推定モデルを構築したい．

参考文献

[1] Bing Liu. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, Vol. 5, No. 1, pp. 1–167, 2012.

[2] Mauajama Firdaus, Hardik Chauhan, Asif Ekbal, and Pushpak Bhattacharyya. Emosen: Generating sentiment and emotion controlled responses in a multimodal dialogue system. **IEEE Transactions on Affective Computing**, 2020.

[3] Xing Fang and Justin Zhan. Sentiment analysis using product review data. **Journal of Big Data**, Vol. 2, No. 1, pp. 1–14, 2015.

[4] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J Passonneau. Sentiment analysis of twitter data. In **Proceedings of the workshop on language in social media (LSM 2011)**, pp. 30–38, 2011.

[5] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In **Proceedings of the 2013 conference on empirical methods in natural language processing**, pp. 1631–1642, 2013.

[6] Laura Ana Maria Oberländer and Roman Klinger. An analysis of annotated corpora for emotion classification in text. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 2104–2119, 2018.

[7] Paul Ekman. An argument for basic emotions. **Cognition & emotion**, Vol. 6, No. 3-4, pp. 169–200, 1992.

[8] Robert Plutchik. A general psychoevolutionary theory of emotion. In **Theories of emotion**, pp. 3–33. Elsevier, 1980.

[9] Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. Wrime: A new dataset for emotional intensity estimation with subjective and objective annotations. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2095–2104, 2021.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.

[11] Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, et al. Transformers: State-of-the-art natural language processing. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 38–45, 2020.

[12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.