

公的なメールに対する BERT を用いたセンチメント分析

菊池理紗¹ 大関洋平¹¹ 東京大学大学院総合文化研究科

rs.kikuchi.11@gmail.com oseki@g.ecc.u-tokyo.ac.jp

概要

本研究は、公的な場面で書かれたメールを対象に、読み手の抱く感情と、それに影響を与えるだろう文章の特徴を明らかにすることを目的とした。はじめに、コーパスや心理実験から収集したメールを、WRIMEで再訓練したBERTを用いて、ネガティブ群・ニュートラル群・ポジティブ群に分類した。次に、各群のメールに対して形態素解析を行い、特徴量を算出した。分析の結果、ネガティブ群は断りのメールが多く、他の2群よりも総文字数や名詞出現率、カタカナ形態素の出現率が低く、動詞・助動詞の出現率、ひらがな形態素の出現率が高かった。また、ニュートラル群は、記号の出現率が高く、動詞出現率やひらがな形態素の出現率が低かった。

1 はじめに

日常生活において、文章は重要なコミュニケーション手段の一つである。それは手書きのメモや手紙に始まり、インターネットの発達を背景に開発された、電子メール（以下、メール）や Twitter, Facebook, Instagram, Slack といった様々なツールでも書かれている。これらのツールを媒介として産出される文章は、文章を通して読み手とコミュニケーションを行うことを目的としている。

文章は、書き手と読み手の相互作用の結果と捉えることができる [1]。特に、コミュニケーションを前提とした文章では、書き手は、読み手の立場や自分と読み手の現在や未来の関係性、読み手の感情に対する推測を理由に、文章表現を選択している [2, 3, 4]。一方で、読み手は、略語の使用や言葉遣い、文章から推測される書き手の感情などの影響を受けて、様々な感情・印象を抱く [5, 6, 7]。すなわち、書き手と読み手の相互的な影響によって、文章とそれに伴う感情が生み出されているといえる。

しかし、日本語の文章を対象としたこれまでの研究では、コミュニケーションを前提として産出され

た文章を対象とした研究は少なく、読み手との関係性に着目した研究や読み手の抱く印象・感情と文章的特徴の関係を検討した研究はさらに限られる。そこで、本研究では、コミュニケーションを前提とした日本語の文章の中でも、公私共に用いることができるメールを対象に、読み手の抱く感情と、それに影響を与える可能性のある文章的特徴を明らかにすることを目的とする。なお、本研究においては、書き手と読み手の関係性は、大学教員と学生、大学事務職員と学生といった公的な関係性に限定する。メールは大学の初年次教育でも取り上げられる媒体であり [8, 9]、その一方で、大学生の書く公的なメールには不備があるという指摘もある [10, 11]。本研究の成果は、今後の初年次教育における指導やメールの産出に対する一助となるだろう。

2 BERT によるセンチメント分析

2.1 カテゴリー個数の設定

センチメント分析とは、ある文章に含まれた感情極性を推測する分析手法であり、この場合、推測されるのは、書き手の感情と考えられる。しかし、ある文章を読んだとき、読み手の感情は、その読み手が書き手の感情だと推測したものと対応して生じる [6]。すなわち、文章に含まれる感情極性を分析することにより、それを読んだ際の読み手の感情を間接的に推測することが可能であるといえる。

英語においては、The Enron Email Dataset¹⁾を用いたメールのセンチメント分析が存在する [12, 13, 14, 15]。これらの研究から、メールは、ネガティブ・ニュートラル・ポジティブの3種類に分類できると明らかになった。したがって、本研究においても、感情極性が3種類であることを前提に分析を行った。

1) <https://www.kaggle.com/datasets/wcukierski/enron-email-dataset>

2.2 分析のモデル

センチメント分析には、極性辞書を用いて単語に感情極性を付与する [12, 13], TF-IDF 値をもとに k-means 法でグループ化する [14], 単語の極性をまとめたデータセットを用いて Support Vector Machine で分類する [15] といった様々な分析手法がある。しかし、極性辞書は単語の極性ラベルであり、文章全体を包括したラベルではない。また、k-means 法を用いた分類は、後続の研究 [15] でクラスとラベルの間の関連が不明であるという指摘がある。そこで、本研究では、文章の特徴量と極性ラベルの関係を BERT に学習させ、それを用いてメールを分類した。具体的には、「WRIME：主観と客観の感情分析データセット ver.2」²⁾[16] を用いて事前学習済み BERT モデルを再訓練し、分類を行った。

WRIME は、日本語の文章の感情分析を行うためのデータセットであり、35,000 件の SNS の投稿について、投稿者と 3 名の読み手が 8 種類の基本感情の強度 (0: 無~3: 強い 4 段階) と感情極性 (-2: 強いネガティブ~2: 強いポジティブの 5 段階) のラベルを付与したものである。本データセットを学習データとして採用した理由は、ラベルが文章全体に対して付与されており、かつ、SNS の投稿を使用しているためである。SNS は、コミュニケーションを目的として用いられるツールであり、「話し言葉の要素を多く含む新しい書き言葉」 [17] と定義される「打ち言葉」を用いている点が、本研究の分析対象であるメールと共通している。加えて、既に Twitter のセンチメント分析を行う際の学習データセットとして利用されている [18, 19]。したがって、WRIME で再訓練を行うことで、よりの確にセンチメント分析が行えると考えた。

また、読み手が付与したラベルを BERT に学習させた場合に、書き手の感情を最も正しく判断することが指摘されているため [16]、本研究では、3 名の読み手が付与した感情極性の平均値をラベルとして採用した。さらに、前節で述べたように、メールは 3 種類に分類できることから、5 段階の感情極性のうち、強いネガティブと弱いネガティブは「ネガティブ」、強いポジティブと弱いポジティブは「ポジティブ」としてまとめ、ニュートラルを含めた 3 段階のラベルとして扱うこととした。なお、ラベル数は、ネガティブが 12,155 件、ニュートラルが 11,462

件、ポジティブが 11,383 件である。

BERT モデルは、事前学習済みモデル bert-base-japanese-v2³⁾ を用いた。BERT モデルの実装には、Hugging Face Transformers⁴⁾ を使用し、先行研究を参考に、学習率やバッチサイズ、エポック数を検討した。その中で最も accuracy が高かった設定を用い、学習率を 2e-5、バッチサイズを 32、ドロップアウト率を 0.1、最適化手法を AdamW [20] として、3 エポックの訓練を行った (accuracy = 0.674)。

2.3 分析対象

現在、日本では、The Enron Email Dataset のような大規模なメールデータセットは存在しない。そのため、2 種類のメールコーパスと第一著者が心理実験 [21, 22] を通して収集したメール (計 358 通) を分析対象とした。使用したコーパスは、金澤 (2014) の「日本語教育のためのタスク別書き言葉コーパス」 [23] と東 (2020) の「中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス」 [24] である。これらのコーパスには、日本語母語話者と中国語母語話者、韓国語母語話者の作成したメールが収録されている。本研究では、その中から日本語母語話者が産出したメールのみを使用した。

また、「日本語教育のためのタスク別書き言葉コーパス」には、公的なメール以外の文章も含まれていたため、「メール」と明記され、かつ、読み手が大学教員であるタスク 1、タスク 4、タスク 7、タスク 10 のメール (計 120 通) のみを使用した。同様に、「中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス」においても、読み手が教員であるタスク 1、タスク 3、タスク 4、タスク 7、タスク 8 のメール (計 145 通) のみを使用した。

心理実験では、計 93 名の参加者に対して、同じ課題内容で 2 回メールを書かせた。これについては、上述の 2 種類のメールコーパスと条件を合わせるため、1 回目に書かれたメール (93 通) のみを使用した。なお、上述のコーパスを含めた各メールの読み手と課題内容を表 1 に示した。課題内容の「依頼」「断り」「意見」などの分類は、元のコーパスにおけるタスク内容の説明から引用した。

2) <https://github.com/ids-cv/wrime>

表1 各コーパスのタスク内容とセンチメント分析の結果

出典	タスク番号	読み手	親疎	課題内容	感情極性 (単位は「通」)		
					ネガティブ	ニュートラル	ポジティブ
金澤 (2014)	タスク1	大学教員	疎遠	本の貸し出し依頼	1	13	16
	タスク4	大学の学長	疎遠	奨学金に対する意見	0	18	12
	タスク7	大学教員	親密	故郷の観光地の説明	0	1	29
	タスク10	大学教員	親密	英語教育に対する意見	1	10	19
東 (2020)	タスク1	大学教員	親密	調査協力の依頼	2	15	12
	タスク3	大学教員	親密	本の貸し出し依頼	0	15	14
	タスク4	大学教員	疎遠	本の貸し出し依頼	1	13	15
	タスク7	大学教員	疎遠	翻訳の仕事の断り	17	9	3
	タスク8	大学教員	親密	翻訳の仕事の断り	17	5	7
菊池 (2021)	—	大学事務職員	—	情報提供の依頼	4	35	25
菊池 (2022)	—	大学事務職員	—	情報提供の依頼	2	15	12
合計					45	149	164

2.4 分析結果

358 通のメールは、45 通がネガティブ、149 通がニュートラル、164 通がポジティブに分類された (表1の「感情極性」を参照)。メールの内訳を検討したところ、ネガティブに分類されたメールが特定のタスクに偏っている可能性が考えられた。そのため、同じ読み手に対して異なる内容のメールを書いている東 (2020) の「中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス」のタスク3、タスク4、タスク7、タスク8の計116通について、対数線形モデルを適用して検討した。読み手との親疎 (親密、疎遠)、課題内容 (依頼、断り)、感情極性 (ネガティブ、ニュートラル、ポジティブ) という三つの要因が、各感情極性に分類されたメールの通数に影響を及ぼすと仮定して分析を行った結果、課題内容と感情極性の1次の交互作用とそれぞれの主効果を含んだ飽和モデルが適切であった ($\chi^2(4) = 3.920, df = 6, p = .69$)。また、交互作用のうち、課題内容が「断り」の場合に「ネガティブ」に分類されたメールが多かった ($u = 4.22, z = 3.96, p < .01$)。したがって、感情極性とメールの文章的特徴の関係を検討する際には、メールの内容も考慮に入れて考える必要がある。

3 各感情極性の文章的特徴量の算出

各感情極性に分類されたメールの特徴を検討するため、件名を除いたメールの総文字数と下記の項目を算出した (表2を参照)。これらの特徴量は、テキスト分類や著者推定、文章の難易度判定を行う際に用いられることの多い指標であり [25, 26, 27]、ウェ

ブサイト「文体診断ロゴーン」⁵⁾でも診断の根拠として算出されるものである。

(1-a) 平均文長：句点 (総文字数 ÷ 区点の数)

(1-b) 平均文長：記号 (総文字数 ÷ 句点・疑問符・感嘆符・改行で区切られた回数)

(2) 平均句読点間隔 (総文字数 ÷ 句読点の数)

(3) 特殊語出現率 (記号の数 ÷ 総形態素数)

(4) 名詞出現率 (名詞の数 ÷ 総形態素数)

(5) 動詞出現率 (動詞の数 ÷ 総形態素数)

(6) 助詞出現率 (助詞の数 ÷ 総形態素数)

(7) 助動詞出現率 (助動詞の数 ÷ 総形態素数)

(8) ひらがな出現率 (ひらがなのみで構成される形態素の数 ÷ 総形態素数)

(9) カタカナ出現率 (カタカナのみで構成される形態素の数 ÷ 総形態素数)

(10) 異なり形態素比率 (頻度1の形態素の数 ÷ 総形態素数)

なお、平均文長は2種類を算出した。メールでは、宛名や署名が書かれることが多く、それらの末尾には句点を使用しないためである。本研究では、宛名や署名、メールアドレス、電話番号などは改行を基準に各1文と数え、かつ、疑問符や感嘆符が使用された文も1文と数えた場合の平均文長も算出した。

算出した各特徴量について、感情極性を要因とした1要因3水準の参加者間分散分析を行い、有意差がみられた場合は、Shaffer法を用いた多重比較を行った。分析の結果、総文字数は、ネガティブ群が他の群に比べて有意に少なく ($F(2,355) = 23.22, p < .01$)、平均文長は、句点のみで計算した場合のみ、ネガティブ群が最も短く、ニュートラル群が最も長いという結果がえられた ($F(2,355) = 13.60$,

3) <https://huggingface.co/cl-tohoku/bert-base-japanese-v2>

4) <https://huggingface.co/docs/transformers/index>

5) <http://logoon.org/>

表2 各感情極性に分類されたメールの文章的特徴量

感情極性	通数	総文字数	平均文長		平均句読点間隔	出現率	
			句点	記号		特殊語	名詞
ネガティブ群	45 通	97.2 (53.38)	27.92 (10.39)	25.3 (9.56)	19.3 (11.92)	0.15 (0.05)	0.27 (0.04)
ニュートラル群	149 通	190.4 (98.72)	36.2 (9.87)	27.3 (7.87)	20.4 (6.60)	0.18 (0.06)	0.31 (0.04)
ポジティブ群	164 通	206.8 (102.37)	33.1 (9.91)	26.0 (6.12)	19.5 (7.87)	0.16 (0.05)	0.31 (0.05)

感情極性	通数	出現率					異なり形態素比率
		動詞	助詞	助動詞	ひらがな	カタカナ	
ネガティブ群	45 通	0.15 (0.04)	0.22 (0.05)	0.16 (0.04)	0.56 (0.07)	0.008 (0.02)	0.60 (0.13)
ニュートラル群	149 通	0.12 (0.03)	0.22 (0.04)	0.13 (0.04)	0.48 (0.07)	0.016 (0.02)	0.42 (0.12)
ポジティブ群	164 通	0.13 (0.03)	0.22 (0.05)	0.12 (0.03)	0.51 (0.07)	0.015 (0.02)	0.44 (0.11)

注：表の数字は平均値。()は標準偏差。

$p < .01$). 記号や改行も含めて計算した場合の平均文長, また, 平均句読点間隔では, 群間に差はみられなかった (平均文長: 記号: $F(2,355) = 1.93$, n.s., 平均句読点間隔: $F(2,355) = 0.63$, n.s.). これらに加えて, 特殊語出現率は, ネガティブ群が最も低く, ニュートラル群が最も高かった ($F(2,355) = 9.47$, $p < .01$). すなわち, ネガティブ群は, 文章全体の長さが短く, 句点で文を区切ることが多いと推察できる. 一方で, ニュートラル群は, ポジティブ群と比較すると, より多くの記号を用いており, 文の区切りでも句点を使うことがより少ないと考えられる.

続いて, 各形態素の出現率をみると, 名詞出現率は, ネガティブ群が他の2群よりも低く ($F(2,355) = 13.79$, $p < .01$), 動詞出現率は, ニュートラル群が最も低く, ネガティブ群が最も高かった ($F(2,355) = 18.36$, $p < .01$). 助詞出現率は, 有意差がみられなかった ($F(2,355) = 2.85$, n.s.). しかし, 助動詞出現率は, ポジティブ群が最も低く, ネガティブ群が最も高かった ($F(2,355) = 26.30$, $p < .01$). 前章の対数線形モデルの結果をふまえると, ネガティブ群は断りのメールが多かったために「～できない」といった「動詞+助動詞:ない」といった組み合わせがみられた可能性が推察される. 対照的に, ニュートラル群やポジティブ群には, 金澤 (2014) の「日本語教育のためのタスク別書き言葉コーパス」のタスク7で書かれた, 故郷の観光地や名物を説明するメールが含まれるため, 名詞の数が多くなったと考えられる. この点については, 今後, 各群で使用されている語彙を詳しく検討することで確認したい.

最後に, 表記については, ひらがな出現率は, ニュートラル群が最も低く, ネガティブ群が最も高かった ($F(2,355) = 19.07$, $p < .01$). カタカナ出現率は, ネガティブ群が他の2群より低かった ($F(2,355) = 3.59$, $p < .01$). 加えて, 頻度1の形態素の割合を示す, 異なり形態素比率は, ネガティブ群が他の2

群よりも高かった ($F(2,355) = 42.50$, $p < .01$). これらについても, どのような語彙が使用されていたのかを中心に, 今後の研究で改めて検討したい.

4 おわりに

本研究で得られた結果からは, 「断り」を目的とするメールがネガティブ群に多く分類されたことから, 内容と感情極性の関係性が示唆された. また, ネガティブ群に分類されたメールは, 他の2群とは異なり, 総文字数や句点のみを用いた平均文長が少なく, 特殊語出現率, 名詞出現率, カタカナ出現率が低く, 動詞出現率, 助動詞出現率, ひらがな出現率が高かった. 一方で, ニュートラル群に分類されたメールは, ポジティブ群と比較して, 句点のみを用いた平均文長が長く, 特殊語出現率と助動詞出現率が高く, 動詞出現率とひらがな出現率が低かった. 今後, 具体的な使用語彙も加えて, 更なる検討を行う必要があるが, 感情極性と文章の特徴量との関係について, 一定の法則性を示すことができた.

先行研究では, 名詞出現率が高い文章は, 低い文章よりもわかりやすいという指摘がある [28]. また, 語彙の選び方や, ひらがな出現率・カタカナ出現率は, 文章の難易度判定にも用いられる指標である [29, 30] ことから, 今後の研究として, 文章難易度と感情極性の関連を検討することもできるだろう.

謝辞

本研究は, JST さきがけ JPMJPR21C2 および JSPS 科研費 JP21H05061 の支援を受けたものです.

参考文献

- [1] Martin Nystrand. A social-interactive model of writing. **Written Communication**, Vol. 6, No. 1, pp. 66–85, 1986.
- [2] Lee Odell and Dixie Goswami. Writing in a non-academic setting. **Research in the Teaching of English**, Vol. 26, No. 3, pp. 201–223, 1982.
- [3] 大浦理恵子, 安永悟. 読み手を特定することが文章産出におよぼす効果. 久留米大学心理学研究, Vol. 6, pp. 11–20, 2007.
- [4] 菊池理紗. 好ましいメールの産出において考慮される要因の関係. 読書科学, Vol. 62, No. 1, pp. 26–42, 2021.
- [5] WGaray Lewandowski Jr. and Samantha Harrington. The influence of phonetic abbreviations on evaluation of student performance. **Current Research in Social Psychology**, Vol. 11, No. 15, pp. 215–226, 2006.
- [6] 加藤尚吾加藤由樹, 杉村和枝, 赤堀侃司. テキストコミュニケーションにおける受信者の感情面に及ぼす感情特性の影響—電子メールを用いた実験による検討—. 日本教育工学会論文誌, Vol. 31, No. 4, pp. 403–414, 2008.
- [7] 坂本真樹紺野倭人. メール印象を考慮した文章提案システムの構築に関する研究. 人工知能学会第31回全国大会論文集, pp. 1–2, 2017.
- [8] 高松正毅. 初年次教育におけるアカデミック・リテラシー教育の位置と大学教育の問題点. 高崎経済大学論集, Vol. 51, No. 3, pp. 51–65, 2008.
- [9] 杉本あゆみ. 初年次教育におけるアクティブラーニング実践報告—グローバルキャンパスでの試行—. 千葉経済大学短期大学部研究紀要, Vol. 14, pp. 105–117, 2018.
- [10] 布川博士佐藤賢. リアリティのある環境の中に埋め込んだ大学生のメールリテラシー教育の実験—研究室の事例第2報(2014年度実験)—. 情報処理学会研究報告, Vol. 2015, pp. 26–42.
- [11] 太田陽子. 依頼メール作成タスクに見る大学生のポライトネス意識. 相模国文, Vol. 42, pp. 125–136, 2015.
- [12] Sudheendra Hangal and Monica S Lam. Sentiment analysis on personal email archives. **Proceedings of the 24th annual ACM symposium on User interface software and technology**, 2011.
- [13] Saif M Mohammad and Wenda Tony Yang. Tracking sentiment in mail: How fgenders differ on emotional axes. **Proceedings of the ACL 2011 Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis**, pp. 70–79, 2013.
- [14] Sisi Liu and Ickjai Lee. Email sentiment analysis through k-means labeling and support vector machine classification. **Cybernetics Systems**, Vol. 49, No. 3, pp. 181–199, 2018.
- [15] Salah Hag Rayan Ali and Neamat El Gayer. Sentiment analysis using unlabeled email data. **2019 International Conference on Computational Intelligence and Knowledge Economy**, pp. 328–333, 2019.
- [16] Noriko Takemura Yuta Nakashima Tomoyuki Kajiwara, Chenhui Chu and Hajime Nagahara. Wrieme: A new dataset for emotional intensity estimation with subjective and objective annotations. **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2095–2104, 2021.
- [17] 文化審議会国語分科会.(編). 分かり合うための言語コミュニケーション(報告).
- [18] Hidetsugu Nanba Satoshi Fukuda and Hiroko Shoji. Analysis of international public emotional responses toward the covid-19 vaccine. **2022 12th International Congress on Advanced Applied Informatics.**, 2022.
- [19] Kongmeng Liew Eiji Aramaki Partick Ramos, Kiki Ferawati and Shoko Wakamiya. Emotion analysis of writers and readers of japanese tweet on vaccinations. **Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis**, pp. 95–103, 2022.
- [20] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. **arXiv:1711.05101**, 2017.
- [21] 菊池理紗. 文章を用いたやり取りにおける好ましさの解明—公的な電子メールを対象として—. 2021.
- [22] 菊池理紗. 提示情報がメールの産出に与える影響—好ましさと特徴語に着目して—. 日本教育心理学会第64回総会発表論文集, 2022. https://www.jstage.jst.go.jp/article/pamjaep/64/0/64.163/_article/-char/ja/.
- [23] 金澤裕之. 日本語教育のためのタスク別書き言葉コーパス. 2014.
- [24] 東会娟. 中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス. 2020. <http://emailtest.sakura.ne.jp/data/>.
- [25] 浅石卓真. テキストの特徴を計量する指標の概観. 日本図書館情報学会誌, Vol. 63, No. 3, pp. 159–169, 2017.
- [26] 野松道子 久野高志 上田修一石田栄美. 文体からみた学術的文献の特徴分析. 2004年度三田図書館情報学会研究大会発表論文集, pp. 33–36, 2004.
- [27] 亀田雅之. 石岡恒憲. コンピュータによる日本語小論文の自動採点システム. 日本計算機統計学会シンポジウム論文集, pp. 153–156, 2002.
- [28] 蓮見陽子. 同一情報に基づく文章表現の印象・評価の差異について. 学習院大学人文科学論集, Vol. 2, pp. 43–67, 1993.
- [29] 山田尚勇建石由佳. 日本文の読みやすさの評価式. 文書処理とヒューマンインターフェース, Vol. 18, pp. 1–8, 1988.
- [30] Suguru Matsuyoshi Satoshi Sato and Yohsuke Kondoh. Automatic assessment of japanese text readability based on a textbook corpus. **Proceedings of the International Conference on Language Resources and Evaluation**, Vol. 8, pp. 654–660, 2008.