

絵文字のマッピングを用いた 感情表現の分析手法の検討

津田恵佑^{†1} 深草理貴^{†1} 森辰則^{†1}

^{†1}横浜国立大学大学院

E-mail: tsuda-keisuke-mj@ynu.jp, fukagusa-riki-wf@ynu.jp
tmori@ynu.ac.jp

概要

テキストと付随する絵文字の表す感情は一致することが普通であるが、そうでないものも存在する。テキストと絵文字の感情の組み合わせには様々な類型があり、それらがどのような感情の機微を示しているのかを自動的に読み取るのは現状困難である。本論文は、絵文字付きテキストのテキストと絵文字の感情を同一平面上で可視化するマッピングを行うことで、その分布からレトリカル絵文字を定量的に検出し、それらの持つ感情の機微の分析を行った。この結果を用いることにより、絵文字付きテキストの感情分析をより詳細に行えるようになることが期待できる。

1 はじめに

絵文字は文字では伝えにくい意味や感情を表現することができるため、感情表現の簡略化やコミュニケーションの充実のために広く採用されている。絵文字は主に、テキストが表す感情を可視的に強調するために用いられるため、テキストと付随する絵文字の表す感情は一致することが普通である。しかし、稀にテキストと絵文字の持つ感情が異なる場合が見られ、それらは「皮肉」「自虐」などの特殊な感情表現や感情の機微を示すために用いられる。このように使用される絵文字を我々は「レトリカル絵文字」と呼んでいる [1]。テキストと絵文字の感情の組み合わせには様々な類型があり、それらがどのような感情の機微を示しているのかを自動的に読み取るのは現状困難である。そこで本論文は、絵文字付きテキストのテキストと絵文字の感情を同一平面上で可視化するマッピングを行うことで、その分布からレトリカル絵文字を定量的に検出し、それらの持つ感情の機微の分析を行った。テキストの感情ベクトルを取得し、付随する絵文字をそのベクトルの座標

に配置することで、テキストの感情情報に沿った絵文字の分布が形成され、テキストの感情と絵文字の感情を同一平面上で可視化することができる。マップ上ではテキストの感情に応じた分布領域が形成され、この分布を見ることで、テキストと絵文字の感情の差異を定量的に特定することができ、レトリカル絵文字の検出を行うことができる。また、この様にして収集したレトリカル絵文字付きテキストが持つ感情の機微を、2名の被験者によって判断し、命名(ラベリング)を行った。この結果を用いることにより、絵文字付きテキストの感情分析をより詳細に行えるようになることが期待できる。

2 関連研究

廣瀬ら [1] は、テキストと異なる感情を表す絵文字をレトリカル絵文字とした。例えば、「試験落ちちゃった」に「笑顔」の絵文字がつく場合、テキストは「悲しみ」を示すが、絵文字付きテキスト全体では「自虐」の表現をしている。レトリカル絵文字の検出手法として、入力されたテキストに対し、2種の絵文字からどちらが適しているかを予測するモデルを作成した。その予測結果と実際に付属していた絵文字が異なっていればレトリカル絵文字であるとして検出を行っていた。しかし、この手法では検出対象である絵文字の種類ごとに1つのモデルを作成する必要がある。本研究では、テキストの感情情報を基に付属の絵文字のマッピングを行い、その分布特徴を分析し、実際にレトリカル絵文字の検出を行う。また、それらがどのような感情の機微を持つのかをラベリングまでを行う。本研究は、未知のレトリカルな絵文字付きテキストに対しても、同手法を用いることで、その感情の機微を予測することに貢献できると考えられる。

3 絵文字付きテキストの感情分析手法の概略

本研究では、絵文字付きテキストに対し、テキストの持つ感情と絵文字の感情を同一平面上で可視化するためのマッピングを行う。その際、テキストを感情強度分析をした結果得られる、喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼の8次元で表されるテキストの感情ベクトルを2次元に圧縮し、それが表す平面座標上に付属する絵文字を配置する。その結果、テキストの持つ感情のうち、期待・喜び・悲しみ・怒り・驚きの感情強度が強いテキストが密集する領域が可視化され、その領域に分布する絵文字の種類を見ることができ、テキストと絵文字のレトリカルな組み合わせを見出すことができる。

3.1 絵文字付きテキストの収集

2022年10月に、Twitter APIを用い、「泣き・怒り・笑顔・爆笑」の表情を持つ絵文字付きテキストを1,000件ずつ収集した。その実例の一部が表1である。

表1 絵文字付きテキストの実例

泣き	え、待って、雨降ってきた😭
怒り	なんで保存できないんだよ😡
笑顔	めちゃくちゃ優雅な朝や😊
爆笑	よし！優勝や-----!!!!!!😂

3.2 テキストの感情強度ベクトルを取得

テキストの持つ感情ベクトルを取得するために、入力されたテキストが持つPlutchik[4]の基本感情（喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼）の強度を予測する感情強度予測モデルを作成する。BERTを用いて作成し、8つの感情強度の総和が1になる様に正規化して出力する。このモデルの出力の実例が表2である。

表2 感情強度予測モデルの出力の実例

	喜び	悲しみ	期待	驚き	怒り	恐れ	嫌悪	信頼
え、待って、雨降ってきた	0	0.27	0	0.48	0	0.15	0.1	0
なんで保存できないんだよ	0	0.95	0	0	0	0.01	0.04	0
めちゃくちゃ優雅な朝や	0.99	0	0.01	0	0	0	0	0
よし！優勝や-----!!!!!!	0.83	0	0	0.16	0	0	0	0.01

次に、モデルによって出力された8次元ベクトルに対し、次元削減アルゴリズムT-SNE[2]を用いて、2次元ベクトルに圧縮する。その結果の実例が表3である。

表3. 圧縮した結果の2次元ベクトルの実例

テキスト	絵文字	X座標	Y座標
え、待って、雨降ってきた	😭	0.96	5.7
なんで保存できないんだよ	😡	38	1.7
めちゃくちゃ優雅な朝や	😊	18	-54.2
よし！優勝や-----!!!!!!	😂	-2.2	-26.3

3.3 テキストの感情強度ベクトルに基づく絵文字のマッピング

これによって取得した2次元ベクトルを座標として、テキストに付属していた絵文字をプロットする。その結果得られたマップが図1である。これを見ると、テキストの感情強度ベクトルは空間上に一様に分布しているのではなく、列島状のクラスタを形成しつつ分布していることがわかる。

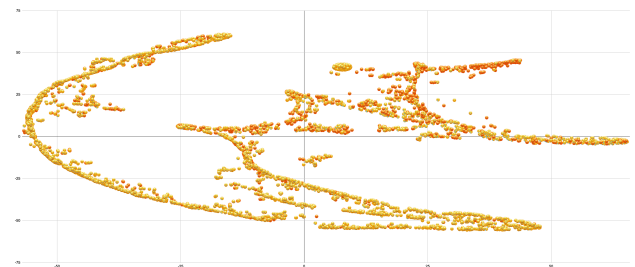


図1 絵文字のマッピング結果

3.4 分布密度の可視化

感情ベクトルの圧縮ならびに2次元平面への写像の際には、すべての絵文字付きテキストを対象とする。一方で、分布密度の可視化に際しては、絵文字の種類ごとに行う。本論文では4種類の絵文字を対象としているので、4種類の可視化結果が得られる。各可視化において密度が高い領域がその絵文字の典型的な使用状況であり、テキストの感情と絵文字の感情が整合していることが期待される。一方で、ある絵文字について、分布密度が低いところに存在する絵文字は何らかのレトリカルな使用であることが期待される。例えば、悲しみ(泣き)のテキストの密度が高いところに、笑顔の絵文字が配置されているとすれば、それは、その絵文字とそれが付随するテキストの組が自虐などのレトリカルな意味を表すであろう。絵文字の分布に対してカーネル密度推定を行い、その結果をヒートマップにより可視化した。「泣き」絵文字に対して行なった結果が図2である。

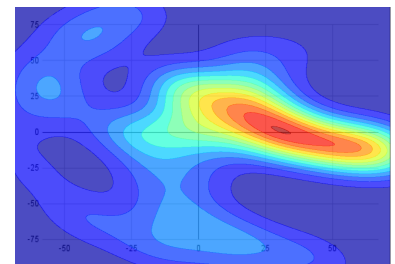


図2 「泣き」絵文字の密度分布の結果の可視化

実際に密度分布の小さい箇所で見られたレトリカル絵文字付きテキストを表4に示す。

表4 レトリカル絵文字の実例

レトリカル絵文字付きテキスト	テキストの感情	絵文字の感情
はあ...なんて可愛い...😊	喜び	悲しみ
楽しすぎて寝れなくなった😴	喜び	怒り
政権批判ばかりする人は、投票行ってから政権批判してくださいね😊	怒り	喜び
めちゃくちゃよくしてもらってたのに残念😞	悲しみ	喜び

4 感情強度予測モデル

BERT を用いた感情強度予測モデルを作成する。作成するモデルはニューラル言語モデルのオープンソースライブラリである Transformers[5] を用いて、東北大学の研究チームによって作成された BERT の日本語の事前学習モデル [6] を用いる。また、Transformers の分類タスク用のクラスを用いることで、提案するモデルを作成する。学習には、感情強度推定のための日本語データセットである WRIME[7] を用いる。WRIME では、1つのテキストに対して、8つの感情それぞれの強度（0～3）が振られている。本研究では、WRIME の Train ラベルがついたデータを学習用データとした。学習用データは 17,104 件であった。この学習用データを用いて BERT をファインチューニングした。

5 カーネル密度推定を用いた絵文字の密度分布の可視化

感情強度予測によって出力された 8 次元ベクトルを、次元圧縮アルゴリズムである t-SNE[2] を用いて 2 次元ベクトルに圧縮した後、カーネル密度推定 [3] を行い、絵文字の分布の密度も同時に可視化する。

5.1 マッピング結果と分布分析

「泣き」「怒り」「笑顔」「爆笑」の各絵文字ごとにマッピングした結果に対し分析対象の領域をナンバリングした図が、それぞれ図 3～6 である。

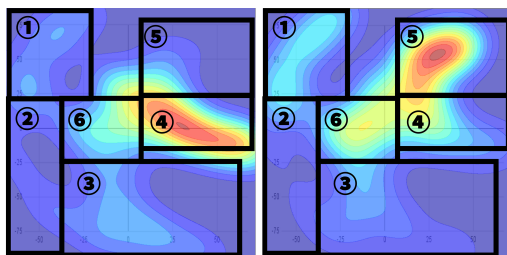


図3 「泣き」絵文字のマップ

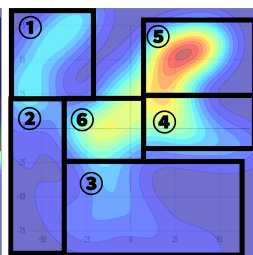


図4 「怒り」絵文字のマップ

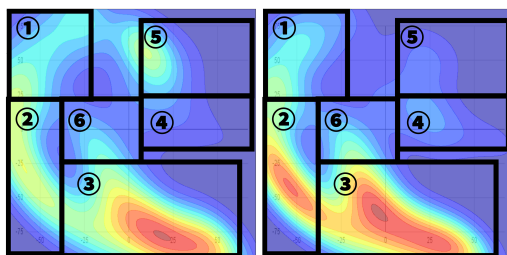


図5 「笑顔」絵文字のマップ

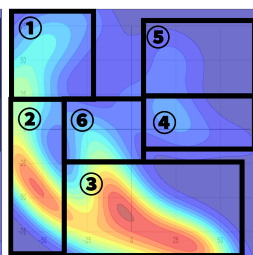


図6 「爆笑」絵文字のマップ

図内にある 1～6 の各領域において、各感情強度の平均と標準偏差を表 5 にまとめた。領域 1～6 はそれぞれ、「期待」「喜び・期待」「喜び」「怒り・嫌悪」「悲しみ」「驚き」が強いテキストが集まっていることが分かる。以降、領域 1 を「期待」、領域 2 を「喜び&期待」、領域 3 を「喜び」、領域 4 を「怒り&嫌悪」、領域 5 を「悲しみ」、領域 6 を「驚き」領域と呼ぶ。

表5 各領域の最強感情要素のベクトルの平均と標準偏差

領域	最強感情	平均	標準偏差
領域 1	期待	0.89	0.13
領域 2	期待 喜び	0.35 0.60	0.27 0.25
領域 3	喜び	0.95	0.07
領域 4	悲しみ	0.87	0.15
領域 5	怒り 嫌悪	0.46 0.41	0.21 0.18
領域 6	驚き	0.69	0.25

6 レトリカルな用途が表す感情の機微のラベリング

6.1 レトリカル絵文字の収集

各領域には各感情を強く表すテキストが集まることが分かった。この特徴を用いてレトリカル絵文字付きテキストを収集する。例えば、「悲しみ」のテキストは領域 4 に分布する傾向があり、それに付属する絵文字は「泣き」が多いため、領域 4 には「泣き」絵文字が多く分布する。しかし、領域 4 に分布する「笑顔」「爆笑」の絵文字は、テキストが「悲しみ」であるため、レトリカル絵文字であると判断できる。この手法を用いて、各絵文字におけるレトリカル絵文字付きテキストを収集した。次に各絵文字・各領域ごとにどのような感情の機微を表しているのかをラベリングした。主に付けられたラベルと絵文字付きテキストの実例を表 6 に示す。

表6 レトリカル絵文字とその感情ラベルの実例

レトリカル絵文字付きテキスト	感情ラベル
今年のクリスマスはどうやって過ごせばいいんやろ😊	不安/悩み/困惑
1日が25時間だったらいいのに...😞	叶わぬ期待を嘆く/残念
初の単独/パーソナリティ😊	感動
母ちゃんにおこらした😊	同情を買う
FF復帰したけどギャザクラの装備更新めんどすぎる😞	不満
ゴールシーン最後まで見せる😊	期待外れ
いつでも話しかけてきて!!!😊	強い願望
明日朝イチでコンビニ行ってモンスター買う😊	決意
おい可愛すぎるだろ😊	喜びの強調
ずっと期待してたのに😞	悲しみの強調
みんなちょっと浮かれすぎだぞ😊	怒り
腰やりました腰痛再発😊	自虐
黄チャって誰がやるん？普通青チャでしょ😊	煽り

6.2 「泣き」絵文字に見られた感情の機微

「期待」領域では、「泣き絵文字」が「期待」のテキストと組み合わせられ、「不安・悩み・困惑」「叶わぬ期待を嘆く・残念」という2つの感情のラベルが付けられた。「喜び」領域では「喜び」のテキストと組み合わせられることで、主に「感動」のラベルが付けられた。「喜び&期待」領域では「期待」「喜び」領域で見られた感情ラベルが混在していた。「怒り」領域では「怒り」のテキストと組み合わせられることで、「同情を買う」「不満」のラベルが付けられた。「驚き」領域では他の領域で付けられたラベルが混在していた。

6.3 「怒り」絵文字に見られた感情の機微

「期待」領域では、「怒り」絵文字が「期待」のテキストと組み合わせられることで、「期待外れ」「強い願望」「決意」という3つのラベルが付けられた。「喜び」領域では「喜び」のテキストと組み合わせられることで、「喜びの強調」というラベルが主に付けられた。「期待&喜び」領域では「期待」「喜び」領域で見られた感情ラベルが混在していた。「悲しみ」領域では「悲しみ」のテキストと組み合わせられることで、「悲しみの強調」「怒り」というラベルが付けられた。「驚き」領域では他の領域で付けられたラベルが混在して見られた。

6.4 「笑顔・爆笑」絵文字に見られた感情の機微

「笑顔」と「爆笑」のレトリカル絵文字付きテキストに付けられた感情ラベルはほとんど同様であった。「悲しみ」「怒り」「驚き」の領域で、「自虐」「煽り」というラベルが付けられた。

7 結果と考察

7.1 結果

レトリカル絵文字付きテキストに対して、テキストの感情ベクトルを基に付属の絵文字をマッピングすることによって、実際にレトリカル絵文字を検出することができた。そして、テキストと絵文字の感情的な組み合わせによって、どのような感情の機微を持つのかを効率よくラベリングすることができた。

7.2 考察

7.2.1 「笑顔」「爆笑」の絵文字における用途の違い
「喜び」領域における感情強度ベクトルの「喜び」要素の平均値は、「笑顔」絵文字付きテキストでは0.97、「爆笑」絵文字付きテキストでは0.91であった。この結果から、「笑顔」絵文字の方が「爆笑」絵文字に比

べて多少「喜び」が強いテキストに組み合わせられていられると言える。また、「悲しみ」「怒り」「驚き」領域において、レトリカルな絵文字として使われた件数は「笑顔」絵文字が77件、「爆笑」絵文字が125件であった。これらの結果から、「笑顔」絵文字は「爆笑」絵文字に比べて、純粋に「喜び」を表現する際に使用され、逆に「爆笑」絵文字はレトリカルな表現に用いられやすいと言える。

7.2.2 混在が可能な感情の組み合わせ

図1のマッピング結果を見ると、左上（期待領域）→左下（喜び・期待領域）→下（喜び領域）に連続的な分布が見られることから、「喜び」と「期待」の感情は混ざり合うことができると言える。同様に、下（喜び領域）→中央（驚き領域）での連続的な分布から、「喜び」と「驚き」の感情も混ざり合うことができる。「驚き」と「悲しみ」「怒り」も混ざり合うことができるが、中央（驚き領域）から右（悲しみ領域）と右上（怒り領域）に枝分かれする点から、「悲しみ」と「怒り」は混ざり合うこともあるが、どちらかの感情を極端に持つ場合も存在すると言える。これらとは逆に、「期待」は「悲しみ・怒り・驚き」とは混ざり合えず、「喜び」は「悲しみ・怒り」とは混ざり合えないということが分かる。

8 まとめ

レトリカル絵文字の収集手法として、テキストが持つ感情と絵文字が持つ感情、また絵文字の分布密度の分析結果も同一平面上で可視化するマッピングを行ない、収集したレトリカル絵文字付きテキストが持つ感情の機微をラベリングした。これにより、今後未知のレトリカル絵文字付きテキストに対する感情分析に役立つと考えられる。今後の展望としては、本稿の対象以外であった絵文字の種類に対しても感情の機微のラベリングを行うことと、感情強度予測モデルの精度を向上させ、より正確なマッピングを行うことが考えられる。

参考文献

- [1] 廣瀬将也, 森辰則. 発話者の発話意図推定に影響を及ぼす表情の絵文字の検出. 言語処理学会第28回年次大会発表論文集, (2022)
- [2] L. van der Maaten, G. Hinton. Visualizing Data using t-SNE, Journal of Machine Learning Research, 9(86):2579 – 2605, 2008.
- [3] Yen-Chi Chen. A Tutorial on Kernel Density Estimation and Recent Advances. Biostatistics &

Epidemiology, 1(1):161–187, 2017.

- [4] R. Plutchik. The nature of Emotions. American Scientist, 89:344-350, 2001.
- [5] Transformers. <https://huggingface.co/docs/transformers/indexbert>
- [6] Pretrained Japanese BERT models.
<https://github.com/cl-tohoku/bert-japanese>
- [7] WRIME: 主観と客観の感情分析データセット.
<https://github.com/ids-cv/wrime>