

G-01

Twitter 特有表現を考慮したツイートの多次元感情抽出手法の提案

「Extracting of Multi-dimensional Sentiment from tweets Using Twittism」

山本 湧輝† 熊本 忠彦†† 灘本 明代†††
Yuki Yamamoto Tadahiko Kumamoto Akiyo Nadamoto

表 1 感情表現の例

感情軸		感情表現	
喜	喜び	明るい	楽しみ
好	愛	好き	好み
安	安心	溜め息	ほっと
哀	悲しさ	哀しさ	悲しみ
厭	厭	嫌い	不快
怖	不気味	怖い	恐い
怒	怒り	憤り	腹立ち
恥	恥ずかしい	決まりが悪い	あられもない
昂	焦る	苛立つ	気を揉む
驚	驚く	魂消える	驚き入る

1. はじめに

Twitter は、140 文字以内のツイートと呼ばれる短文を投稿する為、従来の SNS やブログと違い手軽にツイートを投稿することが可能となった。その為、Twitter には今起こったことや感じたことを気軽に投稿している場合が多く、そのユーザの感情が現れやすい傾向がある。

また、Twitter 等のマイクロブログにはテキスト上で伝えることの難しい、ジェスチャーや表情といった情報を伝えるための手段として、顔文字が用いられている。これまで我々は顔文字の役割を考慮した感情抽出を行ってきた[1]。本論文では、顔文字だけでなく、ツイートの感情を明示的に表す目的で用いられている叫喚フレーズを加えたものを Twitter 特有表現と定義する。浅井ら[2]はマイクロブログ上で使われる突発的な感情の例である、「○○きたあああ」のような語尾の母音を 3 回以上繰り返す表現を「叫喚フレーズ」と定義している。本研究では、浅井らが定義した叫喚フレーズを Twitter 特有表現の一つとして定義する。

このような Twitter 特有表現が文に付与されることによってツイートの感情が変化すると考えられる。例えば、「そういう態度腹立たしい(´0´)」のように「腹立たしい」というきつい言葉を使っているが、顔文字が入ることで文のきつい言い回しが和らげられているように感じられる。また「やったあああああ」のように「やった」という喜びを表す表現が叫喚フレーズが入ることによってより強められているように感じられる。このことから、Twitter 特有表現が文に付与されることによりツイート全体の感情が変化すると考えられる。

そこで本論文では、Twitter 特有表現を考慮した感情抽出手法を提案する。具体的に顔文字の場合に関しては、ユーザ実験を行いそれによって得られた顔文字の役割を考慮する。叫喚フレーズに関しては、顔文字と同様にユーザ実験を行い叫喚フレーズが文に与える影響を考慮する。

以下、第 2 章で関連研究について、第 3 章で感情語辞書について、第 4 章で顔文字辞書について、第 5 章で Twitter 特有表現を考慮したツイートの感情抽出について、第 6 章でまとめと今後の課題について述べる。

2. 関連研究

近年、感情を抽出する研究が行われている。その中で、感情表現を表す感情モデルが提案されている。代表的な感情モデルとして、Plutchik のモデル[3]がある。人間の感情は「嫌悪」⇔「信頼」、「悲しみ」⇔「喜び」、「驚き」⇔「予測」、「恐れ」⇔「怒り」の 8 つの基本となる感情に分類され 4 次元のベクトルで表されている。熊本ら[4]は新聞記事を読んだ人々が記事に対して受けた印象をアンケート調査で調べ、分析することにより、6 本の感情軸を提

案している。各感情軸を反義語関係となる 2 種類の印象語で構成しており、「楽しい」⇔「悲しい」、「うれしい」⇔「怒り」、「面白い」⇔「つまらない」、「楽観的」⇔「悲観的」、「のどか」⇔「緊迫」、「驚き」⇔「ありふれた」の 6 本が提案されている。また、これらの感情軸をベースにした感情値抽出手法[5]も提案しており、新聞記事に対して有効であることを示している。しかしながら、本論文では Twitter に着目し、ツイートの適した感情軸を提案している点で異なる。

高岡ら[6]は中村の提案する 10 次元の感情軸[7]から 6 次元の感情軸へ次元削減を行い、この 6 次元の感情軸を用いて、名言から感情語を抽出している。そして、これら感情軸と感情語を用いて、ユーザの気分にあった名言を検索するシステムを提案している。本研究ではツイートに対する Twitter 特有表現の役割に着目し、その役割を考慮したうえでツイートの感情を抽出する点が異なる。

また、顔文字に関する研究も多数存在する。顔文字の感情分析には顔文字のみの感情分析と、顔文字と文を考慮した感情分析に分けられる。文献[8], [9], [10], [11], [12], [13], [14] は顔文字のみを対象としているため本研究とは異なる。文献[15], [16], [17], [18]は文と顔文字を対象として感情分析を行っている点で本論文と類似している。しかしながら、しかしながら、Sorana ら[15]は Twitter を用いて文の感情と顔文字の感情の共起関係を調べ、文の感情と顔文字の感情の関係性を分析している。本論文ではユーザ実験を行い分析を行っている点で異なる。中丸ら[16]は顔文字を感情極性に分けて文との関係により感情分析を行っている。本論文では顔文字も多次元の感情を有しているとして分析をしている点で異なる。加藤ら[17]は携帯メールにおけるコミュニケーションの中で怒りの感情に着目し、顔文字が付与されることによって受け手の受ける印象の変化を分析している。本論文では顔文字の及ぼす影響を他の感情に対しても分析している点と、Twitter を対象にしている点が異なる。篠山ら[18]は顔文字が文に付与されることで文の感情がどのように変化するかを分析した。

† 甲南大学大学院, Konan University

†† 千葉工業大学, Chiba Institute of Technology

††† 甲南大学, Konan University

表2 感情語辞書の例

感情軸	感情語	感情値	感情語	感情値
喜	楽しみだ	0.91	大切だ	0.54
好	にこやかだ	0.64	好き	0.51
安	安心	0.45	安らぎ	0.81
哀	悲しい	0.45	憂鬱	0.82
厭	嫌だ	0.82	嫌いだ	0.75
怖	怖い	0.45	恐怖	0.64
怒	怒る	0.87	怒鳴る	0.54
恥	恥ずかしい	0.72	恥じらい	0.87
昂	昂る	0.45	すごい	0.57
驚	驚く	0.72	驚愕	0.54

しかしながら、感情の強弱の変化についての分析をしていないため、本論文とは異なる。

3. 感情語辞書

感情推定に関する研究には多くの感情軸が提案されているが、本論文では中村明の感情表現辞典[7]で分類されている「喜」、「好」、「安」、「哀」、「厭」、「怖」、「怒」、「恥」、「昂」、「驚」の感情軸を用いる。中村明の感情表現辞典では各感情軸を表現する語を感情表現としている。各感情軸の感情表現の例を表1に示す。しかしながら、感情表現辞典に登録されている語は定量化されていないため、そのままでは用いることができない。そこで、本研究では熊本ら[4]が提案している感情辞書構築システムを用いて定量化された感情語辞書を構築する。熊本らのシステムでは、ある文書集合を入力としてそれぞれの文書内に現れる単語間の共起関係から対極の軸により感情語を定量化している。

そこで本研究ではこの入力の文書集合を決定する。本研究ではツイートからの感情分析を対象としているが、ツイートは短文のため崩れた表現や文法の乱れが多くあり、単語間の共起関係が正しくなっていないと考え、ツイートと同様に口語的な表現が多くありながら、表記ゆれや文法の乱れが少ないYahoo!映画¹のレビューデータ74,000文書を用いて辞書構築を行った。その結果今回用いる各感情軸に対して32,000個の感情語と感情値のペアを感情語辞書に登録した。構築した感情語辞書の例を表2に示す。

4. 顔文字辞書

次に、顔文字辞書の構築を行う。顔文字の役割は顔文字のみで決まっているのではなく、文の感情と顔文字の感情の関係から顔文字の役割が決定する。したがって、感情軸ごとに顔文字を追加する必要がある。しかし、既存の顔文字辞書では本論文の感情軸ごとに顔文字を分類していないため、顔文字辞書の構築を行った。以下に顔文字辞書構築の手順を示す。

1. 教師データとする顔文字を集める。
2. 人手により各顔文字の感情を決定する。
3. 教師データの顔文字をパーツに分解し、パーツごとの感情を決定する。
4. 3で定義したパーツ持つ顔文字400個を収集し、顔文字辞書に追加する。

表3 パーツごとに割り当てられた感情の例

パーツ	パーツの種類	感情
(^^)	目	喜 好 安
(> <)	目	哀 厭 驚
(. .)	目	哀 厭 恥
(° °)	目	驚 好 昂
(≥ ≤)	目	喜 好 昂
(o)	口	怖
(ω)	口	安
(´ `)	眉	怖
(。.)	頬	喜
(// //)	頬	喜 昂

まず、教師データ作成のため、Twitterから収集した顔文字付きの17,647ツイートの中で出現頻度が高い100個の顔文字を教師データに用いる。

次に、教師データの感情を決定するためにユーザ実験を行った。被験者は顔文字を利用したことがある20代の男性10人である。実験の方法は、顔文字を見てもらいその顔文字がどの感情軸に当てはまるか最大3つの感情から選んでもらい、1番に選択した感情を3点、2番に選択した感情を2点、3番に選択した感情を1点としてアンケート結果を収集した。また、今回実験で用いる顔文字100件で収集した顔文字の92%を占める。

次に、ユーザ実験で各感情に割り当てられた顔文字のパーツに着目し分析を行った。顔文字は人の表情を表現するという性質上、人の顔のパーツを表していることが多い。そこで、本論文では顔文字のパーツを「目」、「口」、「眉」、「頬」の4つとし分析を行った。その結果、パーツの中には各感情と強い共起関係にあるものがあることがわかった。その一部を表3に示す。各感情との共起関係により各々のパーツに感情を付与した。付与した感情を参考に顔文字辞書に顔文字を人手で追加した。その結果、実験で用いた顔文字100個と新たに追加した顔文字400個の合計500個の顔文字を顔文字辞書に追加した。

5. Twitter 特有表現を考慮したツイートの感情抽出手法

ツイートから感情を抽出するために、構築した感情語辞書とTwitter特有表現(顔文字と叫喚フレーズ)を用いて、Twitter特有表現を考慮したツイートの感情値算出手法を提案する。以下に示す。

1. Twitterからツイートを取得する。
2. 取得したツイートを形態素解析し、単語を抽出する。
3. 感情語辞書とのマッチングにより各単語の感情値を算出する。
4. Twitter特有表現を含んでいるツイートでは、文の感情とTwitter特有表現との関係から、ツイート全体の感情値を算出する。また、Twitter特有表現を含んでいない場合は3の感情値をツイート全体の感情値とする。

¹ Yahoo!映画, <http://movies.yahoo.co.jp>

表4 文・顔文字・ツイートの感情極性の分析結果

文+顔文字	ツイートの感情		
	positive	negative	neutral
positive+positive	14	0	0
positive+negative	12	5	3
positive+neutral	0	0	0
negative+positive	1	19	0
negative+negative	1	35	0
negative+neutral	0	0	0
neutral+positive	3	0	1
neutral+negative	0	3	1
neutral+neutral	0	0	0

表5 顔文字の役割

文+顔文字	ツイートの感情		
	positive	negative	neutral
positive+positive	強調	-	-
positive+negative	弛緩	転換	-
positive+neutral	-	-	-
negative+positive	転換	弛緩	-
negative+negative	-	強調	-
negative+neutral	-	-	-
neutral+positive	付加	-	-
neutral+negative	-	付加	-
neutral+neutral	-	-	-

5.1 文の感情値

Twitter から取得したツイートに対して形態素解析エンジン Juman を用いて形態素解析し、ツイート内の形態素と感情語辞書に登録されている感情語のマッチングを行う。しかしながら、表記ゆれが発生する場合は感情語が正しく抽出できない。表記ゆれを解決するために Juman の代表表記を用いる。具体的には、ツイート中の形態素を代表表記に統一し感情語辞書との単語のマッチングを行う。この時感情語辞書中の単語で Juman の代表表記が登録されていないものは、手動で Juman の代表表記を追加した。また、形態素に同形が存在する場合は曖昧な単語と考え考慮しないものとする。

また、ツイートに否定語が入っている場合は正しい感情値を抽出することが出来ない。例えば、「嬉しくない」のような語の場合だと Juman による出力結果は形容詞「嬉しい」と形容詞性述語接尾辞「ない」に分けられる。感情語辞書とのマッチングを行うと「嬉しい」という形容詞が感情語辞書とマッチしこの文の感情は「喜」になる。しかしながら、否定されているので感情語を打ち消さなければならない。この問題を解決するために、熊本ら[4] は単語の印象を考慮するためと品詞を変換する目的で、Juman の出力結果を変換するルールを用いている。その中に否定語についてのルールがあり、これを用いることで否定語と連結する語に関しての感情語の感情値を 0 にすることで問題を解決する。

5.2 顔文字の役割

我々の提案する顔文字の役割は、文の感情値と顔文字の感情値から決定する。そこで、顔文字の役割を決定する為

表6 顔文字の役割の特徴

役割	特徴	ツイートの例
強調	文の感情=顔文字の感情	やったああ(・∀・)ノ
弛緩	文の感情⇔顔文字の感情	勉強ばかり嫌になる(^o)
転換	文の感情⇔顔文字の感情かつ、 文の感情より顔文字の感情が大きい	心配事多いo(▽)o
付加	文の感情がなく、顔文字の感情がある	学校なう(^o)

に、文字のみの場合の感情値と顔文字を含む場合の感情値を比較する実験を行う。

5.2.1 ユーザ実験

文の感情値と顔文字の感情値から顔文字の役割を決定する為の実験を行った。被験者は 10 代から 50 代までの 100 人であり、各年代男性 10 人、女性 10 人からなる。実験データは顔文字を含んだ任意の 100 ツイートと、そのツイートから顔文字を除いた文字のみの 100 ツイートの合計 200 ツイートである。被験者はこの 200 ツイートを見て各々のツイートに対して当てはまると感じた感情を 10 軸の感情の中から最大 3 つの感情を選び、当てはまると感じた感情を 10 段階 (1~10) で評価してもらった。

5.2.2 実験結果

実験において、それぞれのツイートに対し被験者らが付与した点数を感情毎に合計し、それぞれの感情の感情値とする。感情値がある閾値以上で且つ最大のものをそのツイートの感情とする。また、ツイートごとの各感情値において最大のものは複数存在しなかったため、今回の実験に用いたツイートの感情は一意に定まった。このときすべての感情が閾値以下である場合は、そのツイートの感情は無感情として扱う。実験結果より、顔文字を除いた文字のみのツイートの感情と顔文字を含んだ元のツイートの感情を比べ、その顔文字の役割を決定するここで実験では 10 軸の感情によりツイートの感情を判定したが、顔文字の役割を決定するに当たり、感情の変化が大きいほど顔文字の役割が決定しやすいため 10 軸の感情を「positive」、 「negative」、 「neutral」の 3 軸に分類し、この 3 軸の感情の変化により役割を決定する。具体的には、「喜」、「好」、「安」、「昂」を「positive」、「哀」、「厭」、「恥」、「怖」、「驚」、「怒」を「negative」、10 軸に判定されなかったツイートを「neutral」とする。文字のみで判定した感情 (実験より) とすでに求めてある顔文字の感情 (辞書より) と顔文字を付加したツイートで判定した感情 (実験より) との違いを比較して役割を決定する。

表 4 に結果を示す。表 4 では、実験より求めたツイートの文の感情と顔文字辞書から取得した顔文字の感情の組み合わせを縦軸にし、そして実験結果の顔文字を含んだツイートの感情を横軸にし、各々のツイート数を示している。表 4 の結果より文の感情と顔文字の感情の組み合わせが、ツイートの感情極性に与える影響について分類した。その結果を表 5 に示し、表 6 に顔文字の役割と特徴とその例を示し、以下に、4 つの顔文字の役割について説明する。

強調：顔文字が文の感情をより強めている。

positive の文に positive の顔文字が付加された場合はより positive を強めていて、negative の文に negative の顔文字が付加された場合はより negative が強められていると考え、文と顔文字の感情極性が同じ場合は強調とする。

表7 叫喚フレーズによって感情の変化があった時の
ツイートの割合

感情	感情(強)	感情(弱)
喜	67%	2%
哀	35%	10%
好	80%	2%
厭	33%	14%
安	8%	41%
恥	14%	0%
怒	27%	2%
怖	29%	10%
昂	100%	0%
驚	84%	4%

弛緩：顔文字が文の感情を和らげている。

強い positive の文に negative の顔文字が付加された場合は positive を弱めていて、強い negative の文に positive の顔文字が付加された場合は negative を弱めていると考え。文と顔文字の感情極性が異なりかつ、文の感情値が大きい場合は弛緩とする。

転換：顔文字が文の感情（の感情極性）を変化させる。

弱い positive の文に negative の顔文字が付加された場合は negative になり、弱い negative の文に positive の顔文字が付加された場合は positive になると考え。文と顔文字の感情極性が異なりかつ、文の感情値が小さい場合は転換とする。

付加：文の感情がないときに顔文字の感情が付加される。

neutral の文に positive の顔文字が付加された場合 positive になり、neutral の文に negative の顔文字が付加された場合は negative になると考え、文が無感情かつ、顔文字の感情がある場合は付加とする。

5.3 叫喚フレーズの重み

叫喚フレーズがツイートの感情に与える影響を分析するためにユーザ実験を行った。データセットは、映画の実況ツイート中から、叫喚フレーズが含まれている任意の 50 ツイートを用い、被験者は 20 代男性 11 名である。実験手順は以下に示す。

(1) 抽出した 50 のツイートから叫喚フレーズを削除したツイートを見せ、感情表現辞典の 10 軸の感情の中からそのツイートの感情に適した感情軸を最大 3 軸選んでもらい、選択した感情軸に対して 10 段階評価を行った。

(2) (1) と同じツイートから叫喚フレーズを削除せず、ツイートそのままを見せ、(1) と同様の評価をしてもらった。

実験の結果を表 3 に示す。表 3 では叫喚フレーズがある場合とない場合を比較し、叫喚フレーズがあることで感情が強くなったときの割合を感情(強)で示し、弱くなったときの割合を感情(弱)で示している。ここで、叫喚フレーズが文に付与されることによって変化した感情の変化量を叫喚フレーズの重みとして求める。

表 3 の結果より、ある感情*i*の叫喚フレーズの重み TW_i を以下の式で求める。

$$TW_i = (SS_i - SW_i) / \max(SS_i, SW_i) \quad (1)$$

SS_i は、表 3 の結果の感情*i*の感情(強)の値を示し、 SW_i は、表 3 の結果の感情*i*の感情(弱)の値を示す。上記の式で求めた重み TW_i は、喜：0.97、哀：0.71、好：0.98、厭：0.58、安：-0.80、恥：1.00、怒：0.93、怖：0.66、昂：1.00、驚：0.95 である。

ツイートの語尾に母音が 3 回以上繰り返されている場合に、叫喚フレーズが付与されていると判断し、文の感情値に実験によって求めた共感フレーズの重み TW_i を加算することでツイートの感情を求める。

6. まとめと今後の課題

本論文では、Twitter 特有表現を考慮してツイートの感情値を抽出するため、感情語辞書の構築と顔文字辞書の作成のためのユーザ実験を行った。また、叫喚フレーズのツイートに与える重みを求めるためのユーザ実験も行った。その結果感情語辞書に 32,322 個の感情語と感情値、顔文字辞書には 500 個の顔文字を登録した。また、顔文字の役割を決定するために感情軸を感情極性に分類しユーザ実験を行った。その結果、文の感情極性、顔文字の感情極性、ツイートの感情極性から役割を分類した。また、叫喚フレーズを考慮するために、各感情に特有表現の感情値の重みを設定し、感情の抽出を行った。

今後の課題として、テキストの感情辞書の拡張において、出現頻度の低い重要な単語の抽出手法の検討、最後に、評価実験を行い本提案手法の精度について検証したい。

参考文献

- [1] 山本湧輝, 若井祐樹, 熊本忠彦, 灘本明代, “顔文字の役割に着目したツイートの文の感情抽出手法の提案”, 第 6 回データ工学と情報マネジメントに関するフォーラム, No. E6-2, 2014.
- [2] 浅井洋樹, 秋岡明香, 山名早人. “きたああああああああああ!!!!!! 1 1 : マイクログを用いた教師なし叫喚フレーズ抽出”, 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2013), No. A4-4, 2013.
- [3] R.Plutchik. The nature of emotions. American Scientist, Vol. 89, pp. 344-355, 2001
- [4] T.Kumamoto. Design of Impression Scales for Assessing Impressions of News Articles, Lecture Notes in Computer Science, LNCS6193, Springer, pp. 285-295, In International Workshop on Social Networks and Social Media Mining on the Web (SNSMW' 10), 2010.
- [5] 熊本忠彦, 河合由起子, 田中克己. 新聞記事を対とするテキスト印象マイニング手法の設計と評価, 電子情報通信学会論文誌, Vol. J94-D, No. 3, pp. 540-548, 2011.
- [6] 高岡幸一, 灘本明代. “名言のための多次元感情ベクトルの生成”, 第 4 回 Web とデータベースに関するフォーラム (WebDB2011), No. 3G-2, 2011.
- [7] 中村明. 感情表現辞典. 東京堂出版, 1993.
- [8] 風間一洋, 榊剛史, 鳥海不二夫, 篠田孝祐, 栗原聡, 野田五十樹. 顔文字に着目したツイートの感情変化の分析, 第 6 回 Web とデータベースに関するフォーラム, 2013

- [9] 中村純平, 池田剛, 乾伸雄, 小谷善行. 対話システムにおける顔文字の学習, 情報処理学会研究報告. 自然言語処理研究会報告 2003(23), pp. 169-176, 2003.
- [10] M. Ptaszynski. 顔文字処理-取るに足らない表現をコンピュータに理解させるには-, 情報処理 53(3), pp. 204-210, 2012.
- [11] CHO HEERYON, 稲葉利江子, 石田亨, 高崎俊之, 森由美子. 絵文字コミュニケーションにおけるセマンティクス, 情報処理学会研究報告. ICS, [知能と複雑系]2006(110), pp. 1-8, 2006.
- [12] 川上正浩. 顔文字が表す感情と強調に関するデータベース, 大阪樟蔭女子大学人間科学研究紀要 7, pp. 67-82, 2008.
- [13] 伊藤淳子, 宗森純. 擬人化エージェントを介したチャットにおける顔文字と対話雰囲気の関係性の分析, 情報処理学会研究報告. GN, [グループウェアとネットワークサービス] 2008(31), pp. 127-132, 2008.
- [14] 加藤尚吾, 加藤由樹, 小林まゆ, 柳沢昌義. 電子メールで使用される顔文字から解釈される感情の種類に関する分析, 教育情報研究: 日本教育情報学会学会誌 22(4), pp. 31-39, 2007.
- [15] Soranaka.K. Matsushita.M. Relationship Between Emotional Words and Emoticons in Tweets, Technologies and Applications of Artificial Intelligence(TAAI), 2012 Conference on, pp. 262-265, 2012
- [16] 中丸茂. 顔文字が文章の信頼度に及ぼす影響, 言語・音声理解と対話処理研究会 37, pp. 173-176, 2003.
- [17] 加藤由樹, 加藤尚吾, 赤堀侃司. 携帯メールを使用したコミュニケーションにおける怒りの感情の喚起に関する調査, 教育情報研究: 日本教育情報学会学会誌 22(2), pp. 35-43, 2006.
- [18] 篠山学, 松尾朋子. 顔文字を考慮した対話テキストの感情推定に関する研究, 香川高等専門学校研究紀要 1, pp. 151-153, 2010.