

ツイートの感情抽出の為の顔文字の役割分類

若井 祐樹^{†,a} 熊本 忠彦^{‡,b} 灘本 明代^{‡,c}

† 甲南大学大学院 自然科学研究科 ‡ 千葉工業大学 情報科学部 ‡ 甲南大学 知能情報学部

a) m1224008@center.konan-u.ac.jp b) kumamoto@net.it-chiba.ac.jp c) nadamoto@konan-u.ac.jp

概要 近年, Twitter の普及に伴いユーザは自らの感情を気軽に発することができるようになった. この中には, 顔文字を用いてユーザ自らの感情を表現しているツイートが多く存在する. しかし, ツイート本文と顔文字の組み合わせによって感情を表現する手法が異なる場合がある. そこで, 本論文では顔文字が付与されたツイートの感情を抽出するために, ツイートにおける顔文字を 3 つの役割に分類し, 顔文字を含んだツイートの感情を抽出することを提案する.

キーワード 感情抽出, 顔文字, Twitter

1 はじめに

現在, Twitter などのマイクロブログが普及しており, ユーザは自らの感情を気軽にツイートすることができる. ツイートの中には, 感情を示す単語や顔文字が多く使われている. 特に, 顔文字はツイート本文から感じる感情が顔文字の種類によって変わる場合がある. 顔文字とは「\ (^o^)/」のように文字や記号等を用いて, 顔や身体表現を表すコミュニケーション技法のことである. 我々はこれまで, 熊本ら [1] が提案している 3 本の感情軸「楽しい 悲しい」「嬉しい 怒り」「のどか 緊迫」を用いて Twitter 用の感情語辞書の拡張を検討した [2]. ここで, 顔文字がそのツイートの感情に左右することが判明した. そこで本論文では, 文だけでなく顔文字にも着目した感情抽出を行うことを提案する.

顔文字を考慮した感情抽出を行うとき, 同じ顔文字でも文によってその顔文字の役割が異なる事がある. 例えば, 顔文字「\ (^o^)/」のみだと喜んでるように感じ取れるが「無理だ\ (^o^)/」はあきらめている様子「きたあああ\ (^o^)/」は喜びをより強めているように感じ取れ「ふざけんなこら\ (^o^)/」はきつい文章を顔文字が和らげていることで冗談で言っているように感じ取れる. つまりは同じ顔文字でもそれぞれ違う文章と一緒に使うと感じ取れる感情は変わることがわかる. そこで本論文では, 文と顔文字との関係から感情値を算出することを提案する. 本論文ではこの関係を「役割」と呼ぶ.

2 顔文字の役割

2.1 役割の定義

任意の複数の顔文字付きツイートを分析した結果, 顔文字にはそれぞれ役割が存在し, 文章に何らかの影響を与えていることが分かった. 村上ら [3] は顔文字の役割

は「強調」と「配慮(弛緩)」に分類できると述べている. 「強調」は顔文字が文の意味を強めているもの. 「弛緩」は顔文字が文の意味を和らげているものと定義されている. 2 つの役割を考慮して分析した結果, 本研究ではさらに「自嘲」が必要と考えた. 「自嘲」は顔文字があることで, 自分に呆れて笑うさまなどといったニュアンスを感じるものと定義する.

2.2 予備実験

上記の役割の特徴を抽出するために被験者 5 名に予備実験を行った. 顔文字を含んだ 270 ツイートを対象に, まず被験者にこれらツイートを提示し, 各々のツイート内の顔文字が「強調」「自嘲」「弛緩」「その他」の中でどの役割が適しているのか分類してもらった. その時, 感情に対しても同様にして選んでもらった. ここで, 感情語とは, 先に述べた 3 本の感情軸を構成する語であり, 具体的には「楽しい」「悲しい」「嬉しい」「怒り」「のどか」「緊迫」である. 被験者 5 人中 2 人以上が選んだ役割をそのツイートに関する役割とした.

結果と考察

それぞれの感情語毎に対して 3 つの役割の分布を図 1 に示す. 図 1 より, どの感情語に対しても「自嘲」「弛緩」に比べて「強調」を示す顔文字が多いことがわかる. 例えば「きりたんぼ美味しかった(*^^*)」や「コナンめっちゃ怖い(((; ° °)))」といった, 文の感情を強めた顔文字が多く使われていることがわかった. 「強調」に関しては, 顔文字の使い方が様々あるため, さらに細かい分類をする必要がある. また「のどか」は他の感情語に比べて, ツイート数が少ないことがわかる. これは, 実験に使用するツイートのデータを収集する際に「のどか」に関するツイート数が他の感情語に比べて少ないこと及び, ユーザ実験の段階で「のどか」の分類が難しかったことがあげられる. 「怒り」に関しては, 他の感情語に比べ「弛緩」の値が突出していることがわかる. 「怒り」に

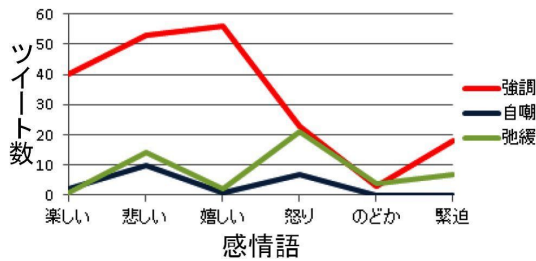


図1 感情語毎の3つの役割のツイート数

おける「弛緩」の例として、「やっぱ態度違うやつ腹立つわ(^^)」のように「腹立つ」といったきつい言い回しを和らげるために「(^^)」のような顔文字を使っていると考えられる。

2.3 顔文字の役割毎の感情値

図1より、感情語が「楽しい」「嬉しい」の場合、「自嘲」「弛緩」がほとんど見られず、「強調」が多いことが分かる。よって、本研究ではツイート本文の感情が「楽しい」「嬉しい」の場合、ツイートに含まれる顔文字の役割は「強調」とみなすこととする。また[2]で示した役割ごとの特徴例より、ツイートの文の感情語と顔文字の感情語が同じならば「強調」と判断し、ツイート文の感情語と顔文字の感情語が対の関係となっていれば「弛緩」と判断する。又ツイートの文が「～ない」と否定形であり、且つ顔文字の感情語が「楽しい」「嬉しい」「のどか」であれば「自嘲」と判断することとする。上記より、本研究ではツイート中の顔文字と文の感情語との関係が以下になる場合、顔文字の役割毎に感情の重みを付与することを行う。

$$TIM_{e,i} = DIM_i \times \alpha \quad (1)$$

$$TIM_{s,j} = DIM_j \times \beta \quad (2)$$

$$TIM_{r,k} = DIM_k \times \gamma \quad (3)$$

ここで i, j, k は各々ある1ツイートを示し、 $TIM_{e,i}$ は顔文字が強調の役割であるツイート i における感情値を、 $TIM_{s,j}$ は顔文字が自嘲の役割であるツイート j における感情値を、 $TIM_{r,k}$ は顔文字が弛緩の役割であるツイート k における感情値を示す。 DIM_i は i における文のみの感情値を、 DIM_j は j における文のみの感情値を、 DIM_k は k における文のみの感情値を示す。ここで文のみの感情値は[2]で提案した手法で求める。また、 α, β, γ 各々の役割の重みを示し予備実験より、 $\alpha > \gamma > \beta$ とする。

3 評価実験

提案手法の有用性を測るために、2章の予備実験の結果の上位3位を正解データとして、顔文字を考慮しないで感情値を算出した場合と、顔文字を考慮して感情値を算出した場合の、各々に対する適合率、再現率、F値を求めた。結果を表1に示す。ここで、各々の役割の重みは

表1 感情語毎の感情値算出システムの精度

		楽しい	悲しい	嬉しい	怒り	のどか	緊迫	平均
顔文字考慮なし	適合率	33%	46%	40%	38%	11%	25%	32%
	再現率	47%	20%	43%	34%	53%	53%	42%
	F値	39%	28%	42%	36%	18%	34%	33%
顔文字考慮あり	適合率	38%	64%	44%	36%	13%	25%	37%
	再現率	63%	22%	58%	21%	56%	47%	45%
	F値	48%	33%	50%	27%	21%	32%	35%

実験より $\alpha = 2.0$, $\beta = 0.2$, $\gamma = 0.3$ とする。

顔文字を考慮しない場合の平均のF値が33%に対して、顔文字を考慮した場合は35%であったことから、顔文字を考慮した方が良い結果となった。しかし感情語によってはF値が下がっているものがあった。特に、「怒り」は顔文字を考慮しない方がF値が高くなっている。考えられる原因として、「弛緩」の重み付けが文の感情値を大きく下げてしまったため、適合率、再現率が共に下がったと考えられる。これにより「怒り」の場合には「弛緩」の文の感情値に対する重み付けの影響の大きさを他の感情語に比べて弱めた方が良かった。

4 おわりに

本研究では、文章と顔文字との感情の関係からツイートの感情値を抽出し、分析を行った。その際に、顔文字の役割をそれぞれ定義した。

今後の課題は、顔文字の感情語辞書に登録している感情語と顔文字の数が少ない為さらに増やす必要がある。さらに役割を定義し、それぞれの場合について重み付けを行ったが、重み付けの値をより最適な値にする必要がある。

謝辞

本研究の一部はJSPS科研費24500134の助成によるものである。ここに記して謝意を表す。

参考文献

- [1] 熊本忠彦, 河合由起子, 田中克己: 新聞記事を対象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, No.3, pp. 540-548, 2011.
- [2] 若井祐樹, 田中美羽, 熊本忠彦, 瀬本明代: 顔文字を考慮したニュースに対するツイートの感情抽出手法の提案, 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 (DEIM2013), 2013.
- [3] 村上浩司, 山田薫, 萩原正人: 顔文字情報と文の評価表現の関連性についての一考察, 第17回言語処理学会発表論文集, pp. 1155-1158, 2012.