

顔文字の役割に着目したツイートの多次元感情抽出手法の提案

山本湧輝^{†,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) *m1424010@center.konan-u.ac.jp* b) *kumamoto@net.it-chiba.ac.jp* c) *nadamoto@konan-u.ac.jp*

概要 近年, Twitter などのマイクロブログはその手軽さから多くのユーザに利用されている. そういったマイクロブログには今起こったことや感じたことをリアルタイムに投稿している場合が多いので, そのユーザの感情が現れやすい傾向がある. また, Twitter には顔文字と呼ばれる表情を持った顔のように見える記号を使用することで表情を表す表現が多く用いられている. しかしながら, 同じ顔文字でも文との関係によってその役割が異なると言える. そこで本論文では, 文の感情と顔文字の感情に着目し, その関係を分類することで, 顔文字の役割を定義するとともに, 顔文字の役割を考慮した感情値抽出手法を提案する.

キーワード 感情抽出, Twitter, 顔文字

1 はじめに

近年, Twitter などのマイクロブログが普及しており, 膨大な情報が投稿されている. その中で, Twitter は, 140 文字以内のツイートと呼ばれる短文を投稿できる為, 従来の SNS やブログと違いリアルタイムな投稿が多く行われている. その為, 今起こったことや感じたことを気軽にツイートをしている場合が多く, その人の感情が現れやすいと言える. しかしながら, 実際の会話と違いテキスト上でのコミュニケーションでは, ジェスチャーや表情のような非言語的コミュニケーションを用いることができないため, 正確に自分の意見や考えを伝えることが難しい. そのため, Twitter ではテキストで正しく伝えることが困難な意見や考えを正確に伝えるために, 顔文字が利用されている. 顔文字とは, 「(^O^)」 「(・ω・)」のような表情などを持った顔のように見える記号を使用することで表情を表す表現である. 顔文字を用いることで, テキスト上で説明することが困難な感情を表現することが出来る. 顔文字が文に付与す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. しかしながら, 同じ顔文字でも文によってそのその役割が異なる場合がある. 例えば, 顔文字「(^O^)」を例に挙げると, 「そういう態度は腹立たしい(^O^)」といったきつい表現を用いているツイートでは, きつい言い回しを和らげているような印象を受けるが, 「今日はすごく嬉しかった(^O^)」では嬉しいという言葉が顔文字が強めているように感じられる. このように同じ顔文字でも文との関係によってその役割が異なる. そこで, 本論文では文の感情と顔文字の感情の関係によってツイートの感情がどう変化するかを分類する. さらに, このような文と顔文字の関係を役割と呼び, その役割に着目した感情抽出手法を提案する.

表 1 顔文字に関する研究の分類

文献番号	顔文字のみ	文+顔文字
[8]	○	
[9]	○	
[10]	○	
[11]	○	
[12]	○	
[13]	○	
[14]	○	
[15]		○
[16]		○
[17]		○
[18]		○

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た [1]. ここでは, 文のみの場合と顔文字を合わせた場合の感情値を比較することで, 顔文字の役割を分類し, 文と顔文字の感情ごとに役割の重みを計算し分析を行った. しかしながら, 文の感情と顔文字の感情の組み合わせごとの重みを決定したが, それぞれの役割が文に与える影響の大きさは求めていなかった. そこで本論文では, 文と顔文字の感情ごとに役割を決定し, それぞれの役割に対して感情を抽出する手法を提案する.

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 章でツイートの感情抽出手法について, 第 6 章でまとめと今後の課題についてについて述べる.

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おり, 多次元の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 [2] がある. 人間の感情は「嫌悪」

表2 感情表現の例

感情軸	感情表現					
喜	喜び	明るい	楽しみ	楽しい	面白い	嬉しい
好	愛	好き	好み	恋しい	愛しい	いい
安	安心	溜め息	ほっと	のんびり	すっきり	あっさり
哀	悲しさ	哀しさ	悲しみ	悲しい	痛い	孤独
厭	厭	嫌い	不快	暗い	憎い	辛い
怖	不気味	怖い	恐い	怖い	不安	震える
怒	怒り	憤り	腹立ち	怒る	腹立たしい	憤怒
恥	恥ずかしい	赤面	あられもない	もじもじ	照れる	はにかむ
昂	焦る	苛立つ	熱狂	奮起	感動	動揺
驚	驚く	魂消る	驚き入る	ぼんやり	歓喜	慌てる

⇔「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[3]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[4][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、ツイートから感情を抽出するという点で異なる。

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と顔文字に注目し、そのツイートから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。表1に顔文字に関する研究の分類を示す。表1に示すように、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、Soranaoka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を対象と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。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2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の文書集合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のレビュアー

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

表3 収集した顔文字の例

顔文字	回数	割合	累積%
(^o^)	2832	16.0%	16.0%
(^^)	935	5.3%	21.3%
(>_<)	917	5.2%	26.5%
(´_`)	742	4.2%	30.7%
(^-^)	527	3.0%	33.7%
(´▽`)	504	2.9%	36.6%
(´ω´)	499	2.8%	39.4%
(;ω;)	413	2.3%	41.8%
(^-▽^-)	389	2.2%	44.0%
(^ω^)	361	2.0%	46.0%

表4 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
(≧ ≦)	目	喜 好 昂
(o)	口	怖
(ω)	口	安
(´ `)	眉	怖
()	頬	嬉
(" ")	頬	嬉 昂

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。

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. 人手により教師データの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類し、パーツごとの感情を決定する。
4. 3で抽出したパーツをもつ顔文字 400 個を収集し、顔文字辞書に追加する。

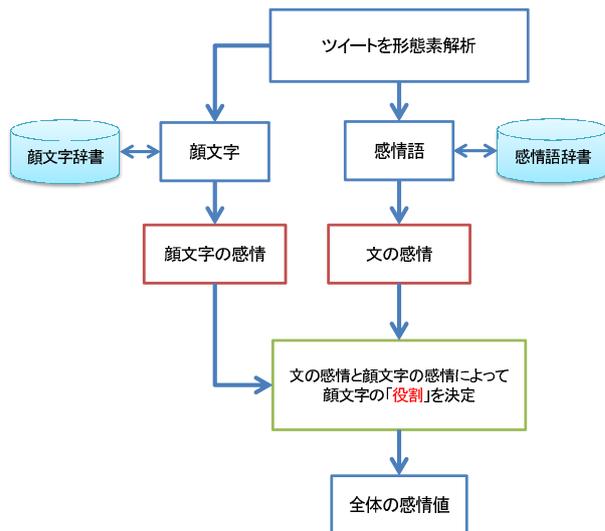


図1 感情抽出の流れ

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。また、今回用いる 100 個の顔文字で収集した顔文字の 92 % の割合を占める。収集した顔文字の例を表 3 に示す。

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。

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わかった。その例を表 4 に示す。抽出したパーツを参考に顔文字辞書に顔文字を手で追加した。その結果、顔文字辞書に登録した顔文字を 500 個まで増やした。

5 顔文字の役割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顔文字辞書を用いて、顔文字の役割を考慮したツイートの感情値算出手法を提案する。以下と図 1 に算出手順を示す。

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

表5 文・顔文字・ツイートの感情極性ごとの分析結果

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6 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	-	-	-

る.

- 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.
- 顔文字を含んでいるツイートでは、文の感情と顔文字の感情から顔文字の役割を決定し、ツイート全体の感情値を算出する. また、顔文字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.

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 Juman² 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

²Juman <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

表7 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	やったあぁ(▽`)/
弛緩	文の感情≠顔文字の感情	勉強ばかり嫌になる(´0´)
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	心配ごと多い(▽`)
付加	文の感情がなく、顔文字の感情がある	学校なう(´0´)

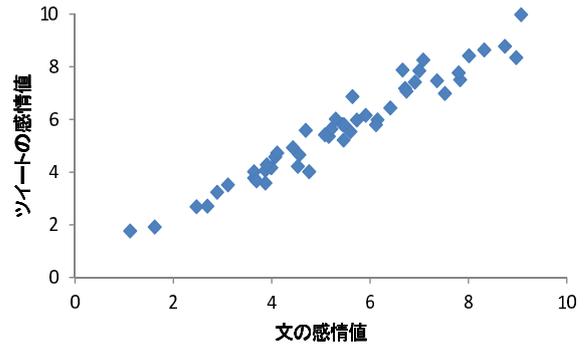


図2 強調の場合の散布図

代表表記が登録されていないものは、手動で Juman の代表表記に追加した. 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「楽しい」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. 本論文ではその中で、否定語についてのルールを適用することで否定語の判定を行い、否定語を含む場合の感情語を打ち消すことで問題を解決する.

5.2 顔文字の役割の決定

5.2.1 ユーザ実験

文の感情と顔文字の感情から役割を決定するために実験を行った. 被験者は10代から50代までの男女各10人の合計100人である.

実験データは顔文字を含んだ任意のツイート100件とそのツイートから顔文字を除いた文字のみのツイート100件の合計200ツイートである. 被験者はこの200件の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、各々10段階で評価した.

表 8 顔文字の役割ごとの回帰式

役割	回帰式
強調	$y=0.9886x+0.3215$
弛緩	$y=0.8728x+0.3083$
転換	$y=-0.8002x+0.001$

表 9 顔文字の役割ごとの決定係数と相関係数

役割	決定係数	相関係数
強調	0.95	0.97
弛緩	0.91	0.95
転換	0.95	0.98

5.2.2 実験結果

実験において、あるツイートのある感情における被験者のつけた点数の合計値をそ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をそのツイートの感情とする。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、文のみのツイートの感情と顔文字を含んだツイートの感情の変化より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）との違いを見て役割を決定する。

表 5 に結果を示す。表 5 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、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 5 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 6 に示し、以下に顔文字の役割を示す。

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と negative の文に negative の顔文字が付加された場合

弛緩 顔文字が文の感情を和らげている。

positive の文に negative の顔文字が付加されたと

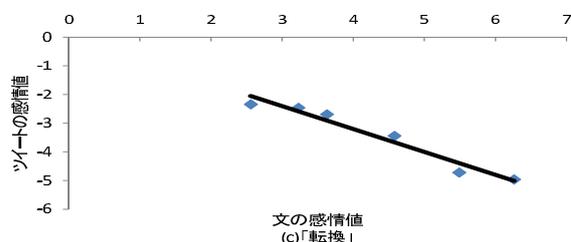
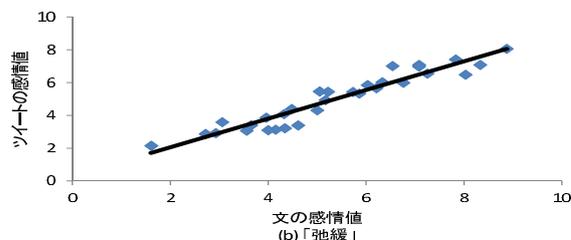
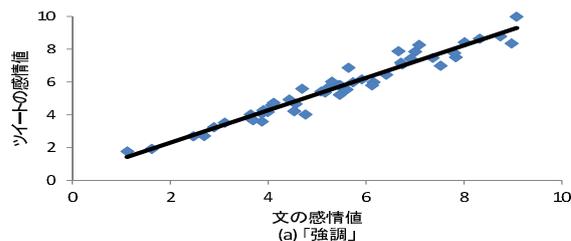


図 3 回帰分析の結果

きツイートの感情が弱い positive になる場合と、negative の文に positive の顔文字が付加された付加されたときツイートの感情が弱い negative になる場合

転換 顔文字の感情が文の感情と入れ替わる

positive の文に negative の顔文字が付加されたときツイートの感情が弱い negative になる場合と、negative の文に positive の顔文字が付加されたときツイートの感情が弱い positive になる場合

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と neutral の文に negative の顔文字が付加された場合

顔文字の役割と特徴を表 7 に示す。また、役割「強調」の場合を例に、文の感情値とツイートの感情値の散布図を図 2 に示す。図 2 に示した散布図から分かるように、顔文字の役割が強調の場合は文の感情値とツイートの感情値には正の相関があることが分かる。そこで、すべての役割に対してそれぞれ相関係数を求めた。その結果、相関係数が 0.95~0.98 という高い数値が得られた。このことから文の感情とツイートの感情の対応関係を回帰分析により定式化できることが判明した。しかし、役割「付加」に関しては文の感情が存在しないため、回帰分析による定式化を行うことができない。したがって、役割「強調」、「弛緩」、「転換」において回帰分析を用いて定式化を行う。

そこで、文の感情値を説明変数、ツイートの感情を目的変数とする単回帰分析を行った。求めた回帰式の結果を表8に示す。この回帰式に文の感情値を代入することにより、ツイートの感情値を算出する。

ここで、各回帰分析の精度を表9に示す。また、各役割における回帰分析の結果を図3にまとめる。表9によると、どの役割においても決定係数の値が0.5より高く、回帰分析の結果が良好であったことを示している。

次に、役割「付加」については文の感情値が存在しないため、顔文字の感情値を重みとして加算することでツイートの感情値を求める。ここで、重みは役割が付加と判断されたツイートの感情値の平均とする。その結果、重みは5.8となった。

6 まとめと今後の課題

本論文では、顔文字の役割を考慮してツイートの感情値を抽出するための第一歩として、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。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。そして、それぞれの役割に対して回帰分析で定式化することにより、顔文字の役割を考慮した感情抽出手法を提案した。

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。

謝辞

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

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] R.Plutchik. The nature of emotions. American Scientist, Vol.89, pp. 344-355, 2001
- [3] 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.
- [4] 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, Vol.J94-D, No.3, pp. 540-548, 2011.
- [5] 熊本忠彦, 河合由起子, 張建偉. ユーザ印象評価データの分析に基づく印象マイニング手法の設計と評価, 情報処理学会論文誌データベース, Vol. 6, No. 2, pp. 1-15, 2013.
- [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関連性の分析, 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.