

顔文字がツイート投稿者の感情推測に及ぼす影響の分析

蔣 承志[†] 熊本 忠彦^{††}

[†] 千葉工業大学大学院情報科学研究科情報科学専攻 〒275-0016 千葉県習志野市津田沼 2-17-1

^{††} 千葉工業大学情報科学部情報ネットワーク学科 〒275-0016 千葉県習志野市津田沼 2-17-1

E-mail: [†]tjm327279161@yahoo.com, ^{††}kumamoto@net.it-chiba.ac.jp

あらまし テキストベースのメッセージ交換が主である Twitter では、相手の表情や仕草を見たり、声を聴いたりすることができないため、メッセージの意図（主にメッセージの感性的側面）を正確に捉えることができないことも多い。このような問題を回避するために、顔文字のような非言語表現が用いられることもあるが、その用法は直感的であり、書き手である投稿ユーザと読み手である閲覧ユーザとの間に共通認識が必ずしもあるとは言えず、顔文字がメッセージの感性的側面に対してどのような役割を果たしているのかについては不明な点も多い。そこで本稿では、Twitter 上で投稿・閲覧されるツイートを対象に、(1) 顔文字の付与によりツイートから推測される投稿者の感情がどのように変化するかをアンケート調査に基づいて調べ、(2) ツイート感情（顔文字のないツイートから推測される投稿者の感情）・顔文字感情（顔文字から推測される投稿者の感情）・顔文字付きツイート感情（顔文字付きツイートから推測される投稿者の感情）のそれぞれを定量化する。さらに (3) 感情どうしの関係を重回帰分析により定式化することで明らかにする。最後に (4) 得られた重回帰式の精度を評価し、その有効性と頑健性を検証する。なお、本稿で対象とする感情は、国語学者の中村明が提唱した 10 種類の基本感情「哀、厭、安、怖、昂、好、喜、驚、怒、恥」をベースにしている。

キーワード Twitter, ツイート感情, 顔文字感情, 顔文字付きツイート感情, 重回帰分析

1 ま え が き

インターネットの一般社会への浸透とスマートフォンをはじめとするモバイル機器の普及により、SNS (Social Networking Service) を利用する人が増えている。特にマイクロブログの一つである Twitter [1] は、いつでもどこでも気軽に投稿・閲覧できるという利点のため、年齢・性別に関係なく多くの人が利用しており、現代におけるコミュニケーション手段の一つとして欠かせないものとなっている [2]。しかしながら、テキストベースのメッセージ交換が主である Twitter では、相手の表情や仕草を見たり、声を聴いたりすることができないため、メッセージの意図（主にメッセージの感性的側面）を正確に捉えることができず、齟齬が生じることも少なくない。このような問題を回避するための一つの手段として、顔文字や絵文字のような非言語表現が用いられることがあるが、その用法は直感的であり、顔文字がメッセージの感性的側面に対してどのような役割を果たしているのかについては不明な点も多い。

昨今、テキストベースのコミュニケーションにおける顔文字の役割を明らかにするために、様々な観点から研究が行われている [3] [4]。例えば、メッセージの読み手が推測した書き手の感情を対象とした研究 [5] やメッセージの読み手の感情の変化を対象とした研究 [6]、メッセージ内容に対する読み手の印象を対象とした研究 [7] などがある。このように様々な観点から顔文字がメッセージの感性的側面に及ぼす影響を明らかにするた

めの研究が行われているが、メッセージの感情・顔文字の感情・顔文字が付与されたメッセージの感情という 3 種類の感情を定量的に扱い、お互いの関係を定式化しようという試みはこれまでなかった。

そこで本稿では、Twitter 上で投稿・閲覧されているツイートを対象に、顔文字が付与されることで、ツイートから推測される投稿者の感情がどのように変化するかをアンケート調査 [8] に基づいて調べ、その結果に対し重回帰分析を行うことで、ツイート感情（顔文字のないツイートから推測される投稿者の感情）・顔文字感情（顔文字から推測される投稿者の感情）・顔文字付きツイート感情（顔文字付きツイートから推測される投稿者の感情）の 3 つの関係を定式化する。なお、本稿で対象とする感情は、国語学者の中村明が提唱した 10 種類の基本感情「哀、厭、安、怖、昂、好、喜、驚、怒、恥」[9] をベースに、アンケート回答者が評価しやすいようにアレンジした「悲しい、嫌い、安心、怖れ、高揚、好き、喜び、驚き、怒り、恥ずかしい」の 10 個であり、各回答者に提示刺激（ツイート、顔文字、顔文字付きツイート）から感じる投稿者の感情をそれぞれの感情毎に「感情が強く込められている（4 点）、感情が割と込められている（3 点）、感情がやや込められている（2 点）、感情があまり込められていない（1 点）、感情が全然込められていない（0 点）」の 5 段階で評価してもらうことで、各提示刺激の感情を定量化している。

以下、2 章で関連研究について述べ、本研究の新規性を明らかにする。3 章でツイート感情、顔文字感情、顔文字付きツイー

ト感情に関するデータの取得について述べ、4章で重回帰分析を行い、ツイート感情・顔文字感情・顔文字付きツイート感情の関係を定式化する。5章で重回帰分析による定式化の精度を評価し、その有効性と頑健性を検証する。最後に6章で本稿のまとめと今後の課題について述べる。

2 関連研究

近年、メッセージに顔文字が付与されている場合と付与されていない場合とで、何が変わるのかに関して様々な観点から研究が行われている。本章では、関連する先行研究について述べ、本研究の新規性を明らかにする。

まず、読み手が推測した書き手の感情や読み手の感情の変化を対象とした研究がある。例えば、廣瀬ら[5]は、携帯メールとして受け取ったメッセージにメッセージの内容と一致する顔文字もしくは絵文字が付与されている場合に、メールを受け取った人への感情（喜び、悲しみ、怒り、不安）の伝わり具合がどのように変化するかを、顔文字を付与した場合と絵文字を付与した場合で比較している。一方、荒川ら[6]は、怒りを感じている相手から謝罪のメールを受け取ったときに、顔文字の有無によって読み手の感情（怒り、親しみ、礼儀正しさ、反省の度合い）がどう異なるかを調べている。

一方、メッセージ内容に対する読み手の印象を対象とした研究もある。例えば、田口[7]は、顔文字を付与したり、文頭や文末にあるひらがなを小文字にしたりすることで、メール文（お礼文、挨拶文、依頼文）に対する印象がどのように変わるかをアンケート調査に基づいて調べている。このとき評価対象となった印象は、「積極的な-消極的な、心のひろい-心のせまい、重々しい-軽々しい、責任感の強い-無責任な、感じの良い-感じの悪い、親しみやすい-親しみにくい、真面目な-不真面目な、親切的な-不親切的な、知的な-知的でない、外向的な-内向的な、明るい-暗い、かわいい-かわいくない、わかりやすい-わかりにくい、若い-若くない」からなる14本の形容詞対であった。

本研究では、ツイート感情、顔文字感情、顔文字付きツイート感情のそれぞれを10個の感情値からなる10次元ベクトルとして定量的に扱っている点とこの3種類の感情間の関係を重回帰式という形で定式化している点が先行研究と異なっている。

3 感情データの取得

本章では、まず、Twitterを一定以上の頻度で利用しているヘビー投稿ユーザとヘビー閲覧ユーザを抽出し、ヘビー投稿ユーザが普段使用している顔文字とツイートを収集する。次に、収集した顔文字の中から出現頻度の高い顔文字（54個）を抽出し、ヘビー閲覧ユーザに各顔文字に込められている投稿者の感情を評価してもらうことで、顔文字感情データを取得する。同様に、収集したツイートの中から選出したツイート（200個）をヘビー閲覧ユーザに読んでもらい、各ツイートに込められている投稿者の感情を評価してもらうことで、ツイート感情データを取得する。このようにして得られるツイート感情データと顔文字感情データからツイートと顔文字をそれぞれ8個ずつ選

出し、組み合わせることで、64個の顔文字付きツイートを生成する。さらに、この顔文字付きツイート（64個）をヘビー閲覧ユーザに読んでもらい、各顔文字付きツイートに込められている投稿者の感情を評価してもらうことで、顔文字付きツイート感情データを取得する。

3.1 Twitter ヘビーユーザの抽出

本節では、アンケート調査（予備調査Aと予備調査B）を行い、一定以上の頻度でつぶやいているヘビー投稿ユーザと一定以上の頻度で閲覧しているヘビー閲覧ユーザを抽出する。

まず、20代～50代のインターネットユーザを対象とする予備調査Aを行い、男性10,808人、女性11,002人からなる21,810人分の回答データを収集した。具体的には、設問Q1でTwitterアカウントの有無を尋ね、アカウントを持っていると回答した人には設問Q2とQ3においてTwitterの閲覧頻度と投稿頻度を尋ねた。閲覧頻度に関しては「ほぼ毎日（1日に4回以上）、ほぼ毎日（1日に2回～3回）、ほぼ毎日（1日に1回）、週4～5日、週2～3日、週1日、たまに閲覧している、ほとんど閲覧していない」の中から1つを、投稿頻度に関しては「ほぼ毎日（1日に4回以上）、ほぼ毎日（1日に2回～3回）、ほぼ毎日（1日に1回）、週4～5日、週2～3日、週1日、たまにつぶやいている、ほとんどつぶやいていない」の中から1つを選択してもらった。その結果、Twitterアカウントを持っていると答えた10,672人のうち、「ほとんど閲覧していない」と回答した1,750人を除く8,922人（83.6%）をヘビー閲覧ユーザとして抽出し、「ほとんどつぶやいていない」と回答した5,062人を除く5,610人（52.6%）をヘビー投稿ユーザとして抽出した。

なお、顔文字付きツイート感情データを取得する際にヘビー閲覧ユーザが不足したため、予備調査Aと同じ要領で、予備調査Bを行っている。すなわち、対象となる回答者は、20代～50代のインターネットユーザであり、男性15,498人、女性12,634人からなる28,132人が参加した。この28,132人の中に予備調査Aに参加した21,810人は含まれていない。このアンケート調査により、Twitterアカウントを持っていると答えた13,906人のうち、「ほとんど閲覧していない」と回答した2,019人を除く11,887人（85.5%）をヘビー閲覧ユーザとして抽出し、「ほとんどつぶやいていない」と回答した7,028人を除く6,878人（49.5%）をヘビー投稿ユーザとして抽出した。

3.2 顔文字の収集

本節では、予備調査Aにより抽出されたヘビー投稿ユーザ891人が参加するアンケート調査（予備調査C）を行い、Twitter上で用いられている顔文字を収集する。

まず、設問Q1でTwitter利用時における顔文字の使用頻度を尋ね、「よく使う、たまに使う、あまり使わない、ほとんど使わない、全く使わない」のうち「全く使わない」と回答した110人を除く781人に対し、設問Q2で普段使っている顔文字を入力してもらった。このとき、回答欄は5つあり、1～5個の顔文字が入力可能であった。以上の結果、1,635個の顔文字を収集することができた。次に、この1,635個の顔文字に対し各顔文

表 1 選出した顔文字（8 個）とその感情値

順位	頻度	顔文字	悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
16	14	^-^	0.40	0.35	2.03	0.35	1.54	2.27	2.37	0.65	0.36	0.78
20	10	(*_*)	1.57	1.31	0.53	1.26	0.71	0.57	0.56	2.06	0.84	1.22
28	7	\ (^~) /	0.29	0.24	1.43	0.28	2.46	2.24	3.34	1.02	0.23	0.44
35	6	(^-^)/	0.31	0.23	2.01	0.21	2.13	2.51	3.01	0.70	0.27	0.46
35	6	(T^T)	3.18	1.67	0.39	1.29	0.50	0.42	0.39	0.75	1.20	0.91
35	6	(-_-)	1.81	1.62	0.45	1.03	0.39	0.36	0.37	0.59	1.41	0.90
35	6	(*^-~*)	0.29	0.32	2.01	0.30	1.98	2.82	3.05	0.70	0.24	1.20
44	5	(; 皿 `)	2.15	1.48	0.40	1.69	0.56	0.43	0.39	1.31	1.01	1.57

表 2 選出したツイート（8 個）とその感情値

ツイート	悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
寒いなあ。風邪引きそう	1.68	1.29	0.72	1.43	0.77	0.79	0.60	0.94	0.85	0.61
消えてなくなりたい	3.19	2.07	0.15	2.20	0.80	0.19	0.17	0.70	1.55	2.06
娘が歩いた！	0.18	0.18	2.10	0.27	3.20	2.75	3.57	3.01	0.18	0.25
今日のライブは本当に楽しかったなあ～。またライブに来たい。	0.29	0.20	1.38	0.25	3.12	3.16	3.29	1.06	0.23	0.33
天気がいいので気分も最高 今日頑張ろう	0.43	0.45	2.07	0.40	2.65	2.23	2.85	0.66	0.48	0.50
3歳の娘が進んでお手伝いしてくれた。優しさに感謝。	0.24	0.22	2.33	0.25	2.72	3.14	3.40	2.15	0.25	0.48
何回も同じこと質問してウザい！	1.34	3.19	0.32	0.97	1.24	0.30	0.30	0.94	3.29	0.47
今日の前で、火災が起きました。	1.59	0.92	0.28	2.30	1.71	0.38	0.39	2.68	0.95	0.33

表 3 顔文字付きツイート（データセット 1 の 8 個）とその感情値

顔文字付きツイート	悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
寒いなあ。風邪引きそう (-_-)	1.79	1.35	0.71	1.29	0.78	0.66	0.57	0.92	0.78	0.62
消えてなくなりたい (*_*)	2.85	1.73	0.36	2.14	0.90	0.39	0.44	0.88	1.32	1.42
娘が歩いた！(; 皿 `)	0.36	0.34	1.85	0.54	2.82	2.43	3.09	2.78	0.29	0.50
今日のライブは本当に楽しかったなあ～。またライブに来たい。^-^	0.35	0.33	1.62	0.37	2.89	3.04	3.23	1.17	0.28	0.43
天気がいいので気分も最高 今日頑張ろう \ (^~) /	0.31	0.32	2.05	0.31	2.61	2.19	2.88	0.72	0.32	0.44
3歳の娘が進んでお手伝いしてくれた。優しさに感謝。(*^-~*)	0.31	0.28	2.26	0.32	2.63	3.02	3.45	2.08	0.27	0.50
何回も同じこと質問してウザい！(^-^)/	1.24	2.60	0.43	0.90	1.04	0.46	0.45	0.95	2.61	0.53
今日の前で、火災が起きました。(T^T)	2.28	1.27	0.24	2.39	1.54	0.30	0.30	2.48	1.05	0.37

字の出現頻度を求め、出現頻度上位の顔文字 54 個を **3.3** 節のアンケート調査で提示する刺激として採用した。

3.3 顔文字感情データの取得

本節では、予備調査 A により抽出されたヘビー閲覧ユーザ 528 人が参加するアンケート調査（本調査 1）を行い、顔文字から推測される書き手の感情に関するデータ（顔文字感情データ）を取得する。

まず、**3.2** 節で選出された 54 個の顔文字を均等に分け、18 個の顔文字からなるデータセットを 3 つ作成するとともに、各データセット内の 18 個の顔文字を 174～177 人の回答者に見てもらい、それぞれの顔文字に書き手のどのような感情が込められているかを評価してもらった。このとき、各回答者には、提示された 10 個の感情「悲しい、嫌い、安心、怖れ、高揚、好き、喜び、驚き、怒り、恥ずかしい」のそれぞれに対し、「感情が強く込められている（4 点）、感情が割と込められている（3 点）、感情がやや込められている（2 点）、感情があまり込められていない（1 点）、感情が全然込められていない（0 点）」の中から最も近いものを 1 つ選んでもらった。なお、予備調査 C に参加したユーザはあらかじめ回答者から除外するとともに、

データセット間で回答者が重複しないよう調整した。

以上の結果得られた「54 個の顔文字 × 10 種類の感情に対する 5 段階評価値」からなるデータに対し、それぞれの顔文字の各感情において 5 段階評価値の平均を求め、顔文字感情データ（54 個の顔文字 × 10 個の感情値）とした。なお、感情評価結果に偏りがある回答者のデータは事前に削除した。

3.4 ツイートの収集と投稿者感情データの取得

本節では、アンケート調査（予備調査 D）を行い、ツイートを収集するとともに、それぞれのツイートに投稿者がどのような感情を込めたかをデータとして取得する。

まず、**3.1** 節の予備調査 A により抽出されたヘビー投稿ユーザを対象とする予備調査 D を実施し、20 代～50 代の男性 554 人、女性 554 人からなる 1,108 人分の回答データを収集した。具体的には、設問 Q1 で「あなたは普段どのようなツイートを投稿していますか？あなたが普段しているようなツイートを、今ツイートをするようなイメージで入力してください。」と尋ね、ツイートを入力してもらい、続く設問 Q2 で「今、入力したツイートにはどのような感情が込められていますか？」とツイートに込めた感情を回答してもらった。このとき、回答者に

提示された感情は「悲しい、嫌い、安心、怖れ、高揚、好き、喜び、驚き、怒り、恥ずかしい」の10個であり、それぞれの感情に対し「感情を強く込めた（4点）、感情を割と込めた（3点）、感情をやや込めた（2点）、感情をあまり込めていない（1点）、感情を全然込めていない（0点）」の中から最も近いものを1つ選択してもらった。なお、新たなツイートを思いつく回答者にはこの2つの設問を最大3回繰り返してもらった。結果、回答者1人につき「1～4個のツイート×10種類の感情に対する5段階評価値」からなるデータを得ることができた。以上の結果、全部で2,786個のデータが得られたが、中にはノイズと思われるデータも含まれていたため、(1)ランダムな英数字などツイートではないと思われるデータや伏せ字、顔文字等を含むデータを削除する、(2)感情評価結果に偏りがあるデータを削除する、(3)重複ツイートを削除する、(4)差別的な発言や侮蔑的な発言、性的な発言など不適切な内容を含むデータを削除する、という手順でデータクリーニングを行った。

このデータクリーニングの結果、838個のツイートからなる投稿者感情データ（838個のツイート×10種類の感情に対する5段階評価値）が得られた。最終的には、この投稿者感情データからそれぞれの感情において強い感情を持つ20個のツイートを選出し、全部で10個のデータセットを作成した。すなわち、各データセットには、それぞれの感情において強い感情を持つツイート2個と各ツイートに対する10種類の感情の5段階評価値が含まれている。

3.5 ツイート感情データの取得

本節では、アンケート調査（本調査2）を行い、3.4節で作成された10個のデータセット内の各ツイートを回答者に読んでもらい、ツイートに込められている投稿者の感情を評価してもらうことで、ツイート感情データを取得する。

まず、3.1節の予備調査Aにより抽出されたヘビー閲覧ユーザを対象とする本調査2を実施し、20代～50代の男性1,131人、女性1,129人からなる2,260人分の回答データを収集した。具体的には、それぞれのデータセット内の20ツイートを224～228人の回答者に読んでもらい、ツイート投稿者の感情を評価してもらった。このとき、同じ回答者が複数のデータセットを評価しないよう調整した。

回答者には「あなたがフォローしている人が以下のようなツイートをしましたとします。そのツイートを読んで、ツイートを投稿した人のどのような感情がそのツイートに込められているとあなたは感じますか？」という設問を提示し、「悲しい、嫌い、安心、怖れ、高揚、好き、喜び、驚き、怒り、恥ずかしい」のそれぞれに対し、「感情が強く込められている（4点）、感情が割と込められている（3点）、感情がやや込められている（2点）、感情があまり込められていない（1点）、感情が全然込められていない（0点）」の中から最も近いものを1つ選択してもらった。

以上の結果得られた「200個のツイート×10種類の感情に対する5段階評価値」からなるデータに対し、それぞれのツイートの各感情において5段階評価値の平均を求め、ツイート感情

表4 重回帰分析の結果（自由度修正済み決定係数 R^2 '）

感情	R^2 '	感情	R^2 '
悲しい	0.95	好き	0.98
嫌い	0.98	喜び	0.98
安心	0.97	驚き	0.96
怖れ	0.93	怒り	0.99
高揚	0.96	恥ずかしい	0.93

データ（200個のツイート×10個の感情値）とした。なお、感情評価結果に偏りがある回答者のデータは事前に削除した。

3.6 顔文字付きツイート感情データの取得

本節では、アンケート調査（本調査3）を行い、顔文字付きツイートから感じる投稿者の感情を評価してもらうことで、顔文字付きツイート感情データを取得する。

まず、3.3節において得られた顔文字感情データ（54個の顔文字×10個の感情値）と3.5節において得られたツイート感情データ（200個のツイート×10個の感情値）からそれぞれ8個のデータを抽出した。選出した顔文字感情データをそれぞれの出現順位・出現頻度とともに表1に示すとともに、ツイート感情データを表2に示す。なお、顔文字・ツイートともに感情分布がなるべくバランスよくなるよう選ばれている。

次に、この選出されたツイートと顔文字を総当たりに組み合わせることで、64個の顔文字付きツイートを生成した。この64個の顔文字付きツイートを同じツイート・同じ顔文字が重複しないように8個のデータセットに均等に分類した。

ここで、3.1節の予備調査Bにより抽出されたヘビー閲覧ユーザを対象として本調査3を実施し、20代～50代の男性837人、女性882人からなる1,719人分の回答データを収集した。具体的には、それぞれのデータセット内の8個の顔文字付きツイートを95～113人の回答者に読んでもらい、ツイート投稿者の感情を評価してもらった。このとき、同じ回答者が複数のデータセットを評価しないよう調整した。なお、設問の提示の仕方も感情評価の仕方も3.5節で述べた本調査2と同じなので、ここでは割愛する。

以上の結果、64個の顔文字付きツイートからなる顔文字付きツイート感情データ（64個の顔文字付きツイート×10個の感情値）が得られた。参考のために、データセット1に対する顔文字付きツイート感情データを表3に示す。

4 重回帰分析による感情どうしの関係の定式化

4.1 重回帰分析

本節では、顔文字付きツイート感情データを目的変数、ツイート感情データと顔文字感情データを説明変数とする重回帰分析を行い、感情どうしの関係を定式化する。

重回帰分析は以下の要領で行った。すなわち、任意の感情（1種類）に対する顔文字付きツイート感情データを目的変数、すべての感情（それぞれ10種類）に対するツイート感情データと顔文字感情データを説明変数とし、計10回の重回帰分析を行った。このとき、変数選択法には変数増加法を採用し、目的

表 5 全データに対する重回帰分析の結果得られた重回帰式（偏回帰係数と定数項）

説明変数		目的変数／顔文字付きツイートの感情値									
		悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
ツイートの感情値	悲しい	0.604		0.047			0.272	0.281	0.060		
	嫌い		0.839			-0.070				0.424	
	安心			0.413		-0.413	-0.115				-0.240
	怖れ				0.769		-0.232	-0.219			
	高揚			-0.297		0.365			-0.109	0.073	-0.335
	好き	-0.117		-0.164			0.927	0.504	0.149		-0.160
	喜び			0.580				0.421			0.432
	驚き	-0.040		0.037		0.129	0.134	0.211	0.803		0.024
	怒り					-0.017				0.453	
	恥ずかしい		-0.143			-0.035				-0.124	0.479
顔文字の感情値	悲しい					0.141	0.062	0.089	0.166		
	嫌い	0.387							-0.243		
	安心										
	怖れ										
	高揚					0.167					
	好き			0.126							
	喜び				-0.122						
	驚き						-0.097	-0.090	0.180		
	怒り		0.296			-0.421	-0.535	-0.713		0.176	
	恥ずかしい										0.047
定数項		0.290	-0.019	0.212	0.345	0.610	0.303	0.252	-0.014	-0.217	0.638

表 6 全データに対する平均二乗誤差（RMSE）

感情	RMSE	感情	RMSE
悲しい	0.17	好き	0.13
嫌い	0.11	喜び	0.16
安心	0.12	驚き	0.14
怖れ	0.17	怒り	0.09
高揚	0.15	恥ずかしい	0.08
		平均 RMSE	0.13

変数の算出に貢献しない変数を削除した。それぞれの重回帰分析において算出された自由度修正済み決定係数を表 4 に示す。

表 4 によれば、自由度修正済み決定係数は、いずれの感情においても 0.90 以上となっており、得られた重回帰式の当てはまり具合が極めて良いことを示している。ここで、それぞれの重回帰分析において得られた重回帰式（各説明変数の偏回帰係数と定数項）を表 5 にまとめる。

5 重回帰式の精度評価

本章では、顔文字付きツイート感情データとツイート感情データ・顔文字感情データとの関係が重回帰式によりどの程度うまく表現されているかを調べるために、全データに対する重回帰式の精度を評価する。次に全データに対する 8 分割交差検定を行うことで、得られた重回帰式が未知なデータに対しても頑健かどうかを検証する。

5.1 全データに対する重回帰式の精度評価

本節では、4.1 節で求めた重回帰式を用いて全データに対する重回帰式の精度を評価する。

まず、重回帰式（表 5 参照）を用いてツイート感情データと顔文字感情データから顔文字付きツイートの感情値を算出し、顔文字付きツイート感情データと比べることで、全データに対する重回帰式の精度を評価した。このとき、評価指標としては以下の式で求められる平均二乗誤差（Root-Mean-Square-Error: RMSE）を採用した。結果を表 6 に示す。

$$RMSE = \sqrt{\sum_{i=1}^{64} (V_i - V'_i)^2 / 64}$$

但し、 V_i は i 個目の目的変数の値を、 V'_i は i 個目の説明変数の値を用いて重回帰式により算出された値を示している。

表 6 によれば、各感情における RMSE は 0.08～0.17 の範囲であり、その平均は 0.13 であった。感情値が取りうる値が 0.0～4.0 ということを考え、この幅（＝4.0－0.0＝4.0）で正規化してみると、感情に対する平均誤差（＝平均 RMSE/4.0）は 3.3%と十分小さいことがわかる。このことから、4.1 節で得られた重回帰式（表 5 参照）は、ツイート感情および顔文字感情と顔文字付きツイート感情との関係を十分高い精度で表現できていると言える。

5.2 未知データに対する重回帰式の頑健性の検証

本節では、全データに対し 8 分割交差検定を行うことで、重回帰分析により得られる重回帰式が未知なデータに対しても頑健かどうかを検証する。

まず、ツイート感情データ、顔文字感情データ、顔文字付きツイート感情をツイートや顔文字が重複しないよう均等に分割し、8 個のデータセットを作成した。すなわち、各データセッ

表 7 8 分割交差検定の結果 (RMSE)

感情	未知データセット								平均 RMSE
	1	2	3	4	5	6	7	8	
悲しい	0.20	0.20	0.17	0.23	0.16	0.20	0.19	0.24	0.20
嫌い	0.16	0.11	0.11	0.12	0.07	0.17	0.11	0.10	0.12
安心	0.15	0.15	0.13	0.18	0.14	0.18	0.19	0.09	0.15
怖れ	0.16	0.19	0.20	0.18	0.20	0.25	0.12	0.16	0.18
高揚	0.41	0.21	0.19	0.14	0.22	0.17	0.23	0.16	0.22
好き	0.10	0.15	0.14	0.16	0.19	0.21	0.19	0.09	0.15
喜び	0.15	0.15	0.18	0.23	0.25	0.25	0.20	0.15	0.20
驚き	0.16	0.21	0.10	0.23	0.17	0.21	0.15	0.20	0.18
怒り	0.13	0.07	0.11	0.10	0.11	0.16	0.09	0.12	0.11
恥ずかしい	0.11	0.08	0.09	0.06	0.10	0.13	0.11	0.13	0.10

トにはツイート感情データと顔文字感情データが 8 個ずつ、各ツイートと各顔文字の組み合わせに対応する顔文字付きツイート感情データが 8 個含まれている。次に、任意の 1 個のデータセットを未知データセットとし、残り 7 個のデータセットを用いて 4.1 節と同じ要領で重回帰分析を行った。さらに、7 個のデータセットから得られた重回帰式を用いて未知データセット内のツイート感情データと顔文字感情データから顔文字付きツイートの感情値を求め、未知データセット内の顔文字付きツイート感情データと比べることで、未知なデータに対する精度 (RMSE) を求めた。以上の操作をすべてのデータセットに対し、感情毎に行った。結果を表 7 にまとめる。

表 7 によれば、各感情における平均 RMSE は 0.10~0.22 の範囲であり、表 6 に示された結果 (全データに対する RMSE) と比べても遜色なく十分高い精度であることがわかる。このことは、4.1 節で得られた重回帰式 (表 5 参照) が未知なデータに対しても頑健に適用可能であることを示唆しており、ツイート感情データおよび顔文字感情データと顔文字付きツイート感情との関係を適切に表現していることがわかる。

6 ま と め

本稿では、Twitter 上で投稿・閲覧されているツイートを対象に、顔文字が付与されることで、ツイートから推測される投稿者の感情がどのように変化するかをアンケート調査に基づいて調べ、その結果に対し重回帰分析を行うことで、ツイート感情 (顔文字のないツイートから推測される投稿者の感情) や顔文字感情 (顔文字から推測される投稿者の感情) と顔文字付きツイート感情 (顔文字付きツイートから推測される投稿者の感情) との関係を重回帰式という形で定式化した。また、精度評価を行うことにより、得られた重回帰式がこの 3 種類の感情の関係を十分高い精度で表現していることを示した。さらに、8 分割交差検定を行った結果、得られた重回帰式が未知なデータに対しても頑健であることを示した。なお、本稿で対象とする感情は「悲しい、嫌い、安心、怖れ、高揚、好き、喜び、驚き、怒り、恥ずかしい」の 10 個であり、国語学者の中村明が提唱した 10 種類の基本感情 [9] をベースにアレンジしたものとなっている。また、各回答者には提示刺激 (ツイート、顔文字、顔

文字付きツイート) から感じる投稿者の感情を評価してもらう際に、それぞれの感情がどのくらい込められているかにより 5 段階 (4 点~0 点) で評価してもらった。

今後の課題として以下のようなものが挙げられる。(1) 本稿では、重回帰分析により重回帰式を求めたが、その式の意味については言及していない。この点についてまず考察する必要がある。(2) 今回の調査・分析により、ツイート感情および顔文字感情と顔文字付きツイート感情との関係を重回帰式を用いて高精度に表現できることが分かった。そこで、ツイートや顔文字から投稿者の感情 (10 個の感情値) を推測するための手法を考案し、重回帰式と組み合わせることで、顔文字付きツイートの感情値を高精度に算出するための手法を開発したい。

謝 辞

本研究の一部は、JSPS 科研費 17K00430 ならびに福田将治 奨学寄付金による研究助成の成果であり、ここに記して感謝の意を表す。

文 献

- [1] <https://twitter.com/>
- [2] 熊本忠彦：Twitter の利用状況とツイートの印象に関する分析，情処研報 (自然言語処理)，Vol. 2016-NL-227，No. 16，pp. 1-13，2016.
- [3] ブタシンスキ ミハウ，奥村紀之，ジェブカ ラファウ：顔文字の現象および研究の概観，人工知能学会誌，Vol. 32，No. 3，pp. 333-341，2017.
- [4] 奥村紀之，ブタシンスキ ミハウ，ジェブカ ラファウ：工学的・心理学的側面から見た顔文字の分析に関する研究動向，人工知能学会誌，Vol. 32，No. 3，pp. 342-349，2017.
- [5] 廣瀬信之，牛島悠介，森周司：携帯電話メールによる感情の伝達に顔文字と絵文字が及ぼす影響，感情心理学研究，Vol. 22，No. 1，pp. 20-27，2014.
- [6] 荒川歩，鈴木直人：謝罪文に付与された顔文字が受け手の感情に与える効果，対人社会心理学研究，Vol. 4，pp. 128-133，2004.
- [7] 田口雅徳：顔文字の付与および文頭・文末の小文字化がメール文の印象に与える影響，情報学研究 (獨協大学情報学研究所)，Vol. 3，pp. 105-111，2014.
- [8] <https://monitor.fast-ask.com/>
- [9] 中村明，感情表現辞典，東京堂出版，東京，1993.