

# 顔文字がツイート印象に及ぼす影響の分析

## An Analysis of Influence of Emoticons on Affective Impressions Feeling from Tweets

中平 鴻司†  
Koji Nakahira†

熊本 忠彦†  
Tadahiko Kumamoto†

### 1. はじめに

インターネット技術の急速な発展に伴い、ソーシャルメディアはネット利用者にとって重要な情報源の1つとなっており、今なお急成長している。Oxford Internet Institute[1]によると、ソーシャルメディアは米国のニュースソースにとって重要なプラットフォームへと進化しており、ソーシャルメディアを使用してニュースを入手する人が増えている。現在のソーシャルメディアと従来のマスメディアには明らかな違いがあり、従来のマスメディアは多くの人に伝達可能である反面、一方的であり、ソーシャルメディアはインターネット上で使用され、コンテンツはユーザが選択または編集することができ、短くて軽い、いつでもどこでも保存、管理、アクセス、修正などの編集をできる。イベントやニュースに関する報道の不十分なところを強力に補完し、誰もが情報発信できる。主にソーシャルメディアを通じ、Twitter[2]などの普及により、その時々起きたことや感じたことをほぼリアルタイムにTwitterに投稿する人が多く存在している。個人間の情報発信が可視化されやすくなったことにより従来のコミュニケーションの概念を超えて、新たなメディアとして浸透しつつある。

また、Twitter Japan[3]は2018年12月26日に事業戦略説明会を開催した。その場でTwitter社は全世界での月間アクティブユーザ数が2018年10月現在で3億2600万人を超えていると発表した。ほかのSNSやメッセージングサービスと比較して、Twitterがどのような特色を持ったサービスであるかということ、Twitterは他のSNSと比べて重大事件、交通機関の遅延、旬なトピックの情報がすぐに手に入ると言われており、もっともリアルタイム性の高いSNSであると感じる。Twitterは、ユーザが意見を表明する非常に簡潔なマイクロブログであり、様々な事件、新しい政府の方針、新しい話題、テキストからユーザの意見や評価を分析することができる。これらは、すべてのユーザの意見とは言っても個人の意見であることが多い。Twitterデータのセンチメント分析であれば、社交行事に対する社会の

態度、製品のユーザ評価、その他の情報を掘り下げることが可能であり、世論の動向、製品の評判などの用途にとって重要な意味を持つ。

ツイート内容から印象の分析は、出来事や製品に関する意見をSNSなどから効率的に収集することができる。しかしながら、メッセージ交換を主とするテキストベースのコミュニケーションでは身振り、手振りを伝えられないために、メッセージの感性的側面が伝わらないことも多い、このような時に、顔文字を追加することで様々な感性情報を伝達できるようになり、コミュニケーションをより豊かにすることが可能となる。

ツイートの印象などツイートに内在する心的な情報に着目した研究もあり、注目を浴び始めている。例えば、熊本ら[4]は、任意のTwitterユーザの印象選好(そのユーザが普段、どのような話題の、どのような印象のツイートを見たり投稿したりしているか)を円グラフあるいは散布図を用いて可視化するシステムを提案している。Guriniら[5]は、任意のTwitterユーザが興味を有する話題を同定するとともに、それぞれの話題に関するツイートのセンチメントをPositive、Negative、Neutralの中から決定し、ユーザ間の類似度を算出することで、フォローすべきユーザを推薦するシステムを提案している。しかしながら、いずれかのシステムも顔文字付与によるツイートの印象や感情の変化には対応していない。

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、ツイートの印象と顔文字が付与されたツイートの印象だけでなく、顔文字の印象も定量化する。そして、顔文字の印象およびツイートの印象と顔文字付きツイートの印象との関係を重回帰分析により調べることで、顔文字の印象がツイートの印象に及ぼす影響を明らかにする。

なお、本稿で対象とする印象は、熊本ら[6]が提案した8種類の基本印象「攻撃的・不愉快、ネガティブ、感じの良い、楽しい・愉快、ポジティブ、ほのぼの、鬱陶しい、怖い」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)から感じるツイートの印象をそれぞれの印象毎に「印象を強く感じる(4点)、印象を割と感じる(3点)、印象をやや感じる(2点)、印象をあまり感じない(1点)、印象を全然感じ

† 千葉工業大学大学院 情報科学研究科  
Chiba Institute of Technology,  
2-17-1, Tsudanuma, Narashino, 275-0016 Chiba, Japan

ない(0点)」の5段階で評価してもらうことで、各提示刺激の印象を定量化している。

本稿の構成は以下のとおりである。2.章で関連研究を紹介し、本研究の新規性を明らかにする。3.章で顔文字印象、ツイート印象、顔文字付きツイート印象に関するデータの取得について説明する。4.章で重回帰分析を行い、ツイート印象・顔文字印象・顔文字付きツイート印象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.章で本稿の結論と今後の課題を述べる。

## 2. 関連研究

Twitterを利用した研究は数多くなされており、特にツイートメッセージや顔文字、顔文字付きツイートを対象とした研究では、メッセージに顔文字を付与することで、メッセージの感情的側面がどう変わるかが様々な観点から研究されてきた。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、メッセージ中の印象、感情、コメントの抽出、分析は、感情分析、印象分析、テキスト分類、情報推薦や情報検索などに関連する重要な研究トピックの1つであり、盛んに研究されている。例えば、目良ら[7]は、発話文中に出現する単語の好感度を用いて、話者の感情(喜び、楽しみ、期待、受容、悲しみ、嫌悪、怒り、恐れ、後悔、不安)を推定するシステムを提案している。熊本ら[8]は、「楽しい⇔悲しい」、「うれしい⇔怒り」、「のどか⇔緊迫」という3種類の印象を対象に、新聞記事を読んだ人々が感じる印象の強さを数値的に求めるための手法を提案している。

次に、顔文字付きツイートの感情、印象分析を対象とする先行研究では、メッセージ内容に対して、顔文字の有無によりメッセージの印象や感情がどう変わるかを調べている。例えば、荒川ら[9]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。田口ら[10]は、顔文字および小文字がメール文(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。廣瀬ら[11]の携帯メールとして受信したメッセージにメッセージの内容と同じ顔文字またはピクトグラムが与えられている場合、メールを受け取った人への感情(喜び、悲しみ、怒り、不安)の伝わり具合がどう変化するかを、顔文字(emoticon)の場合と絵文字(pictogram)の場合で比較している。

以上で述べたように、先行研究では、顔文字の有無により、ツイートから感じる印象がどのような変化するかを調べているが、顔文字の印象を定量化していない。すなわち、ツイート印象に対する顔文字の影響を調べるにあたり、ツイート印象と顔文字付きツイート印象を比較し、有意な差のあった印象を特定しているに過ぎず、個々の顔文字における印象の違いについては考えていない。これに対し、本研究では、ツイート印象、顔文字印象、顔文字付きツイート印象のそれぞれを8個の印象値からなる8次元ベクトルとして定量的に扱っている点とこの3種類の印象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. 印象データの取得

日常的にツイートを投稿しているユーザ(以下、投稿ユーザと呼ぶ)と日常的にツイートを閲覧しているユーザ(以下、閲覧ユーザと呼ぶ)を抽出する。投稿ユーザが普段から使用するツイートや顔文字をアンケート調査に基づいて調べ、顔文字の使用頻度を調査する。アンケート調査に顔文字の中から出現頻度の高い顔文字(54個)を抽出し、閲覧ユーザに各顔文字の印象を評価してもらい、顔文字印象データを取得する。次に、収集したツイートの中から抽出したツイートを閲覧ユーザに読んでもらい、各ツイートの印象を評価してもらうことで、ツイート印象データを獲得する。最後は、得られる顔文字印象データとツイート印象データから顔文字とツイートをそれぞれ8個ずつ抽出し、組み合わせで、64個の顔文字付きツイートを作成する。この64個の顔文字付きツイートを閲覧ユーザに読んでもらい、その印象を評価してもらうことで、顔文字付きツイート印象データを取得する。

### 3.1 Twitter ユーザの抽出(予備調査A)

アンケート調査(予備調査A)を行い、一定以上の頻度で利用している投稿ユーザと閲覧ユーザを抽出する。

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26,306人、女性23,636人からなる49,942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24,578人の

うち、「ほとんど閲覧していない」と回答した3,769人を除く20,809人(84.7%)を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12,090人を除く12,488人(50.9%)を投稿ユーザとして抽出した。

### 3.2 顔文字の収集 (予備調査 B)

予備調査により抽出された投稿ユーザのうち891人が参加するアンケート調査(予備調査B)を行い、Twitter上で用いられている顔文字を収集する。

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 顔文字印象データの取得 (本調査 1)

アンケート調査(本調査1)を行い、抽出された閲覧ユーザ496人から顔文字の印象に関するデータ(顔文字印象データ)を取得する。

3.2節で選択された54個の顔文字が均等に分割され、18個の顔文字、各データセットに18個の顔文字がある3つのデータセットに作成される。160~176人の回答者に見てもらい、それぞれの顔文字に閲覧ユーザがどのような印象を感じているかを評価してもらった。このとき、各閲覧ユーザには、提示された8個の印象「攻撃的・不愉快、ネガティブ、感じの良い、楽しい・愉快、ポジティブ、ほのぼの、鬱陶しい、怖い」のそれぞれに対し、「印象を強く感じる(4点)、印象を割と感じる(3点)、印象をやや感じる(2点)、印象をあまり感じない(1点)、印象を全然感じない(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Bに参加した閲覧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×8種類の印象に対する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印象において5段階評価値の平均を求め、顔文字印象データ(54個の顔文字×8個の印象値)とした。なお、印象評価結果に偏りがある閲覧ユーザのデータは事前に削除した。

### 3.4 ツイートの収集 (予備調査 C)

本節では、3.1節により抽出された投稿ユーザのうち1,108人が参加するアンケート調査(予備調査C)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ユーザがどのような印象を込めたかをデータとして取得する。

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印象が込められていますか?」とツイートに込めた印象を回答してもらった。ここで、各投稿ユーザには、提示された8種類の印象のそれぞれに対し、「印象を強く感じる(4点)、印象を割と感じる(3点)、印象をやや感じる(2点)、印象をあまり感じない(1点)、印象を全然感じない(0点)」の中から最も近いものを1つ選んでもらった。なお、新たなツイートを思いつく投稿ユーザにはこの2つの設問を最大3回繰り返してもらった。その結果として、「2,786個ツイートデータ×8種類の印象に対する5段階評価値」からなるデータを得ることができた。2,786個のデータが得られたが、その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)印象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でデータからそれぞれの印象において、比較的強い印象を含む200個のツイートを均等に分けることで、20個のツイートとその印象値からなるデータセットを10セット作成した。すなわち、各データセットには、それぞれの印象において強い印象を持つツイート2個と各ツイートに対する8種類の印象の5段階評価値が含まれている。

### 3.5 ツイート印象データの取得 (本調査 2)

本節では、3.1節の予備調査Aにより抽出された閲覧ユーザのうち1,999人が参加するアンケート調査(本調査2)を行い、各閲覧ユーザにデータセット内の各ツイートの印象を評価してもらうことで、ツイート印象データを取得する。

具体的には、それぞれのデータセット内の20ツイートを196~225人の閲覧ユーザに読んでもらい、その印象を評価してもらった。このとき、同じ閲覧ユーザが

表 1: 抽出した顔文字 (8 個) とその印象値

順位	頻度	顔文字	攻撃的・ 不愉快	ネガ ティブ	感じ の良い	楽しい ・愉快	ポジ ティブ	ほの ぼの	鬱陶 しい	怖い
16	14	^-^	0.98	0.90	2.30	2.51	2.21	2.46	1.01	0.91
23	10	(*_*)	1.10	1.90	0.81	0.87	0.76	0.90	1.42	0.99
29	7	\ (^ ^ ) /	0.43	0.47	2.50	2.99	2.80	2.06	0.63	0.34
39	6	(^-^)/	0.49	0.55	2.71	2.84	2.64	2.46	0.60	0.39
40	6	(T^T)	1.07	2.39	0.77	0.66	0.60	0.84	1.47	0.88
41	6	(--)	1.21	2.05	0.81	0.59	0.67	0.85	1.37	0.84
43	6	(*^-^*)	0.44	0.56	2.59	2.87	2.56	2.40	0.55	0.33
49	5	( ; 丱 ^ )	1.10	2.24	0.82	0.76	0.65	0.83	1.49	0.99

表 2: 抽出したツイート (8 個) とその印象値

ツイート	攻撃的・ 不愉快	ネガ ティブ	感じ の良い	楽しい ・愉快	ポジ ティブ	ほの ぼの	鬱陶 しい	怖い
寒いなあ。風邪引きそう	0.81	1.75	1.11	0.82	0.91	1.44	1.25	0.66
消えてなくなりたい	1.96	3.33	0.29	0.25	0.17	0.25	2.26	2.39
娘が歩いた!	0.32	0.27	2.75	2.75	2.77	3.12	0.55	0.25
今日のライブは本当に楽しかったなあ〜。またライブにきたい。	0.32	0.31	2.56	2.80	2.65	2.50	0.47	0.26
天気がいいので気分も最高 今日頑張ろう	0.55	0.45	2.57	2.24	2.89	2.36	0.89	0.41
3歳の娘が進んでお手伝いしてくれた。優しさに感謝。	0.34	0.32	2.83	2.59	2.53	3.13	0.61	0.28
何回も同じこと質問してウザい!	2.92	2.50	0.31	0.23	0.40	0.22	2.66	1.89
今日の前で、火災が起きました。	1.24	1.56	0.36	0.29	0.42	0.27	1.04	2.55

表 3: 顔文字付きツイート (データセット 1 の 8 個) とその印象値

顔文字付きツイート	攻撃的・ 不愉快	ネガ ティブ	感じ の良い	楽しい ・愉快	ポジ ティブ	ほの ぼの	鬱陶 しい	怖い
寒いなあ。風邪引きそう (-_-)	0.82	1.86	1.14	0.85	0.87	1.32	1.26	0.71
消えてなくなりたい (*_*)	1.79	3.10	0.51	0.44	0.44	0.48	2.08	2.07
娘が歩いた! ( ; 丱 ^ )	0.48	0.53	2.27	2.37	2.37	2.81	0.88	0.34
今日のライブは本当に楽しかったなあ〜。またライブにきたい。^-^	0.35	0.42	2.42	2.92	2.72	2.25	0.69	0.24
天気がいいので気分も最高 今日頑張ろう \ (^ ^ ) /	0.37	0.38	2.61	2.58	2.83	2.32	0.70	0.34
3歳の娘が進んでお手伝いしてくれた。優しさに感謝。 (*^-^*)	0.40	0.35	2.59	2.46	2.43	2.94	0.68	0.26
何回も同じこと質問してウザい! (^-^)/	2.52	2.30	0.45	0.46	0.51	0.38	2.38	1.74
今日の前で、火災が起きました。(T^T)	1.27	1.62	0.45	0.33	0.48	0.30	1.14	2.26

表 4: 顔文字付きツイート (データセット 2 の 8 個) とその印象値

顔文字付きツイート	攻撃的・ 不愉快	ネガ ティブ	感じ の良い	楽しい ・愉快	ポジ ティブ	ほの ぼの	鬱陶 しい	怖い
寒いなあ。風邪引きそう (T^T)	0.95	1.76	1.22	0.98	0.87	1.35	1.41	0.80
消えてなくなりたい (-_-)	1.84	3.12	0.52	0.49	0.52	0.50	2.29	1.92
娘が歩いた! (*_*)	0.46	0.52	2.56	2.62	2.51	2.83	0.79	0.45
今日のライブは本当に楽しかったなあ〜。またライブにきたい。( ; 丱 ^ )	0.59	0.71	2.19	2.63	2.38	2.04	0.84	0.49
天気がいいので気分も最高 今日頑張ろう^-^	0.55	0.45	2.52	2.34	2.77	2.37	0.75	0.41
3歳の娘が進んでお手伝いしてくれた。優しさに感謝。 \ (^ ^ ) /	0.46	0.51	2.72	2.57	2.53	2.87	0.83	0.38
何回も同じこと質問してウザい! (*^-^*)	2.56	2.20	0.51	0.54	0.68	0.54	2.33	1.72
今日の前で、火災が起きました。(^-^)/	2.45	1.73	0.49	0.53	0.55	0.45	1.83	2.75

複数のデータセットを評価しないよう調整した。各閲覧ユーザには「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たとします。そのツイートを読んで、あなたはどのような印象を感じますか?」という設問を提示し、「攻撃的・不愉快、ネガティブ、感じの

良い、楽しい・愉快、ポジティブ、ほのぼの、鬱陶しい、怖い」のそれぞれに対し、「印象を強く感じる (4点)、印象を割と感じる (3点)、印象をやや感じる (2点)、印象をあまり感じない (1点)、印象を全然感じない (0点)」の中から最も近いものを1つ選択しても

表 5: 重回帰分析の結果 (自由度修正済み決定係数  $R^2$ )

印象	$R^2$	印象	$R^2$
攻撃的・不愉快	0.94	ポジティブ	0.98
ネガティブ	0.99	ほのぼの	0.99
感じの良い	0.98	鬱陶しい	0.96
楽しい・愉快	0.98	怖い	0.99

らった。

以上の結果得られた「200 個のツイート × 8 種類の印象に対する 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印象において 5 段階評価値の平均を求め、ツイート印象データ (200 個のツイート × 8 個の印象値) とした。なお、印象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印象データの取得 (本調査 3)

アンケート調査 (本調査 3) を行い、顔文字付きツイートの印象を評価してもらうことで、顔文字付きツイート印象データを取得する。

まず、3.3 節において得られた顔文字印象データ (54 個の顔文字 × 8 個の印象値) と 3.5 節において得られたツイート印象データ (200 個のツイート × 8 個の印象値) からそれぞれ 8 個のデータを抽出した。抽出した顔文字印象データをそれぞれの出現順位・頻度とともに表 1 に示すとともに、ツイート印象データを表 2 に示す。

次に、この抽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作成した。この 64 個の顔文字付きツイートを同じツイート・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 A により抽出され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 865 人、女性 800 人からなる 1,665 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 103～112 人の閲覧ユーザに読んでもらい、その印象を評価してもらった。このとき、同じ閲覧ユーザ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印象評価の仕方も 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印象データ (64 個の顔文字付きツイ

ト × 8 個の印象値) 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印象データを表 3 に示す。データセット 2 に対する顔文字付きツイート印象データを表 4 に示す。

## 4. 重回帰分析による印象どうしの関係の定式化

### 4.1 重回帰分析

重回帰分析を使用して印象どうしの関係を定式化する。ここで、説明変数としてツイート印象データと顔文字印象データを用い、顔文字付きツイート印象データは目的変数として、重回帰分析を行う。

重回帰分析は以下のように行われる。すなわち、任意の印象 (1 種類) の顔文字付きツイート印象データを目的変数、すべての印象 (各 8 種類) に対するツイート印象データと顔文字印象データを説明変数とする重回帰分析を合計 8 回を行った。ここで、目的変数の算出に貢献しない変数を削除するために、変数選択法として変数増加法を採用する。表 5 は、各重回帰分析において算出された自由度修正済みの決定係数を示している。

表 5 によると、自由度修正済み決定係数は、いずれの印象においても、0.90 以上であり、得られた重回帰式の近似が非常に良いことを示している。表 6 は、各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) をまとめたものである。目的変数と説明変数が同じ印象の場合には色を付けた。

### 4.2 重回帰式の意味

4.1 節において生成された重回帰式の意味を説明する。具体的には、目的変数「ネガティブ」を例に、重回帰式を用いて表 3 に示した顔文字付きツイート「消えてなくなりたい (\*\_\*)」のネガティブ印象値  $f$  を求める方法を示す。ある顔文字付きツイート「消えてなくなりたい (\*\_\*)」におけるネガティブ印象値  $f$  は、表 6 に示すようにそのツイートから算出される説明変数「ツイートの印象値－ネガティブ」、「ツイートの印象値－怖い」、「顔文字の印象値－楽しい・愉快」、「顔文字の印象値－ポジティブ」の値をそれぞれ  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$  とすると

$$f = 0.764 x_1 + 0.085 x_2 - 0.473 x_3 + 0.408 x_4 + 0.394$$

という式で表される。

ここで、表 1 と表 2 の対応する印象値を  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$  に代入すると、顔文字付きツイート「消えてなくなりたい (\*\_\*)」のネガティブ印象値  $f$  は 3.04 に

表 6: 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印象値							
		攻撃的・ 不愉快	ネガ ティブ	感じ の良い	楽しい ・愉快	ポジ ティブ	ほの ぼの	鬱陶 しい	怖い
ツイ ー ト の 印 象 値	攻撃的・不愉快	1.207							
	ネガティブ		0.764	0.097	0.093	0.205	0.184		-0.233
	感じの良い			0.824		0.245	-0.697	-0.644	-0.421
	楽しい・愉快	-0.213			0.955		-0.255	0.275	
	ポジティブ					0.735			0.032
	ほのぼの				-0.158		1.510	0.216	0.173
	鬱陶しい	-0.769				-0.084		0.575	
	怖い	0.068	0.085	-0.102	-0.191		0.133	0.085	0.856
顔 文 字 の 印 象 値	攻撃的・不愉快								
	ネガティブ								
	感じの良い								
	楽しい・愉快		-0.473		0.149	0.180			
	ポジティブ		0.408						
	ほのぼの	0.054		0.111			0.136		
	鬱陶しい					0.217			
	怖い				0.159				
定数項		0.952	0.394	0.027	0.091	-0.698	-0.773	0.797	0.770

なり、表 3 に示した正しい値 3.10 に接近していることが見られる。他の 7 種類の印象値もこの方式で求められる。

## 5. 重回帰式の精度評価

顔文字付きツイート印象データ、ツイート印象データ、および顔文字印象データの関係と重回帰式がどの程度正確に表しているかを調べるために、すべてのデータを用いて重回帰式の精度を評価する。次に、全データに対して 8 回分割交差検定を行うことにより、得られた重回帰式が未知データに対しても適用可能である

表 7: 全データに対する平均二乗誤差 RMSE (提案手法)

印象	RMSE	印象	RMSE
攻撃的・不愉快	0.18	ポジティブ	0.12
ネガティブ	0.10	ほのぼの	0.11
感じの良い	0.12	鬱陶しい	0.12
楽しい・愉快	0.12	怖い	0.09
		平均 RMSE	0.12

ことを検証する。

### 5.1 全データに対する重回帰式の精度評価

4.1 節で求めた重回帰式 (表 6 参照) を用いて全データに対する重回帰式の精度を評価する。重回帰式を用いてツイート印象データと顔文字印象データから顔文字付きツイートの印象値を算出し、顔文字付きツイート印象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。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。その結果を表 7 に示す。

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

ここで、 $V_i$  は  $i$  個目の目的変数の値を表し、 $V'_i$  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表している。 $N$  はデータの個数であり、ここでは、 $N = 64$  となる。表 7 によると、各印象における RMSE の範囲は 0.09 から 0.18 であり、平均 RMSE は 0.12 であった。印象値が 0.0 から 4.0 までの値を取りうるということを考え、この幅 (=  $4.0 - 0.0 = 4.0$ ) で正規化すると、印象に対する平均誤差 (= 平均 RMSE/4.0) は

3%と小さいことが判明した。したがって、4.1 節で得られた重回帰式 (表 6 参照) は、ツイート印象および顔文字印象と顔文字付きツイート印象との関係をより高い精度で表現することができる。

更に、顔文字印象値の有効性を考えるために、顔文字の存在を無視して、ツイート印象値を顔文字付きツイート印象値と見做した場合 (ベースライン手法) の印象別 RMSE を求めてみた。その結果を表 8 に示す。表 8 によれば、各印象における RMSE は、0.14(怖い)~0.28(攻撃的・不愉快)の範囲であり、平均 RMSE は 0.19 であった。この結果を表 7 の結果と比較するために、以下に示す式を用いて改善率を求めてみたところ、表 9 に示すような結果となった。

$$\text{改善率} = \frac{(\text{ベースライン手法の RMSE} - \text{提案手法の RMSE})}{\text{ベースライン手法の RMSE}}$$

表 9 によれば、各印象における改善率は 24.8%(感じの良い)~41.1%(ほのぼの)の範囲であり、平均改善率は 34.9%であった。いずれの印象でも精度が改善しているのがわかる。これは、複数個の顔文字印象とツイート印象を考えることの有効性を示している。

## 5.2 未知データに対する重回帰式の頑健性の検証

重回帰分析によって得られた重回帰式が未知なデータに対して頑健かどうかを検証するために、すべてのデータを用いて 8 分割交差検定を行う。まず、64 個の顔文字付きツイート印象データを同じツイートや顔文字が含まれないように分割し、8 つのデータセットを作成する。各データセットに各顔文字付きツイートを構成するツイートと顔文字の印象データを追加する。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、残りの 7 つのデータセットを用いて 4.1 節と同じ方法で重回帰分析を実行する。この 7 つ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印象データと顔文字印象データから顔文字付きツイートの印象値を求め、未知データセット内の顔文字付きツイート印象データと比べることで、未知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、印象毎に行った。

結果を表 10 にまとめて示す。表 10 によれば、各印象における平均 RMSE は 0.10~0.21 の範囲であり、表 7 に示された結果 (全データに対する RMSE) と並び立つことの出来る精度であることが判明した。このことは、4.1 節で得られた重回帰式 (表 6 参照) が未知なデータに対しても頑健であることを示しており、顔文字付きツイート印象と顔文字印象データおよびツイ

表 8: ツイート印象値を顔文字付きツイート印象値と見做した場合の印象別 RMSE (ベースライン手法)

印象	RMSE	印象	RMSE
攻撃的・不愉快	0.28	ポジティブ	0.18
ネガティブ	0.16	ほのぼの	0.19
感じの良い	0.16	鬱陶しい	0.20
楽しい・愉快	0.19	怖い	0.14
		平均 RMSE	0.19

表 9: 印象別 RMSE の改善率 (ベースライン手法)

印象	改善率	印象	改善率
攻撃的・不愉快	36.6%	ポジティブ	33.4%
ネガティブ	34.5%	ほのぼの	41.1%
感じの良い	24.8%	鬱陶しい	38.9%
楽しい・愉快	36.1%	怖い	33.6%
		平均改善率	34.9%

ト印象データ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、ツイートの顔文字を追加すると印象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、その結果を得られたツイート印象 (顔文字のないツイートから推測される印象) や顔文字印象 (顔文字から推測される印象) と顔文字付きツイート印象 (顔文字付きツイートから推測される印象) に関するデータを用いて重回帰分析を行うことで、この 3 つの印象の関係を定式化した。また、得られた重回帰式の精度評価を行うことにより、重回帰式がこれら 3 つの印象の関係を適切に表しており、精度が十分に高いことを示した。

さらに、8 分割交差検定を行った結果、未知データに対しても有効であり、重回帰式が頑健であることを示した。なお、本稿で対象とする印象は、熊本ら [6] が提案した 8 種類の基本印象「攻撃的・不愉快、ネガティブ、感じの良い、楽しい、愉快、ポジティブ、ほのぼの、鬱陶しい、怖い」である。

今後の課題としては、顔文字やツイートからの印象 (8 個の印象値) 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印象値を高精度に算出するための手法を開発したい。

表 10: 8 分割交差検定の結果 (RMSE)

印象	未知データセット								平均
	1	2	3	4	5	6	7	8	RMSE
攻撃的・不愉快	0.28	0.20	0.17	0.20	0.20	0.23	0.14	0.25	0.21
ネガティブ	0.15	0.17	0.09	0.16	0.16	0.09	0.11	0.08	0.12
感じの良い	0.17	0.13	0.15	0.12	0.14	0.14	0.13	0.14	0.14
楽しい・愉快	0.18	0.09	0.14	0.17	0.18	0.14	0.11	0.16	0.15
ポジティブ	0.20	0.11	0.19	0.17	0.15	0.14	0.11	0.13	0.15
ほのぼの	0.17	0.09	0.15	0.16	0.13	0.10	0.13	0.14	0.13
鬱陶しい	0.16	0.18	0.14	0.10	0.14	0.11	0.12	0.17	0.14
怖い	0.14	0.09	0.07	0.09	0.08	0.10	0.10	0.13	0.10

## 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。

## 参考文献

- [1] <https://www.oii.ox.ac.uk/>
- [2] <https://twitter.com/>
- [3] [https://www.huffingtonpost.jp/2018/12/27/twitter-japan-user\\_a\\_23628286/](https://www.huffingtonpost.jp/2018/12/27/twitter-japan-user_a_23628286/)
- [4] 熊本忠彦, 鈴木智也, Twitter ユーザの印象選好を可視化するシステムの設計と評価, 電子情報通信学会論文誌, Vol. J98-D, No. 5, pp. 788–801, 2015.
- [5] Davide F. Gurini, Fabio Gaspiretti, Alessandro Micarelli and Giuseppe Sansonetti: A Sentiment-Based Approach to Twitter User Recommendation, Proc. of the 5th ACM RecSys Workshop on Recommender Systems and the Social Web, Hong
- [6] 熊本忠彦: Twitter の利用状況とツイートの印象に関する分析, 情処研報学会 (自然言語処理), Vol. 2016-NL-227, No. 16, pp. 1–13, 2016.
- [7] 目良和也, 市村匠, 相沢輝昭, 山下利之: 語の好感度に基づく自然言語発話からの情緒生起手法, 人工知能学会論文誌, vol.17, no.3, pp.186–195, 2002.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.