

複数の感情分析ツールを統合した メタ感情分析手法の提案

熊本 忠彦¹ 蔣 承志²

現在、Web 上には様々な感情分析ツールが公開されており、文章からその書き手の感情を推測することも可能となっている。しかしながら、単一の感情分析ツールには得意不得意があり、推測可能な感情の種類が限られていたり、精度良く推測できなかったり、推測不可とエラーを返したりすることもある。そこで本論文では、140 文字以内の短い文章（ツイート）を対象に、複数の感情分析ツールを統合することで、ツイート投稿者の様々な感情をツイートからより頑健に、より高精度に推測できるようになるメタ感情分析手法を提案し、メタ感情分析という考え方の可能性を検証する。なお、本論文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 種類であり、感情分析ツールには複数の感情を対象とするツールを 3 種類、感情極性（ポジティブ/ネガティブ）を対象とするツールを 2 種類の計 5 種類を採用する。

1 はじめに

近年、ソーシャル・ネットワーキング・サービス（SNS）は、情報発信・収集手段の一つとして日常生活に欠かせないものとなっている。特に、Twitter [1] は、代表的なマイクロブログの一つであり、いつでもどこからでも 140 文字以内の短い文章を気軽に投稿・閲覧できるという利点があるため、年齢や性別に関係なく様々な人々に幅広く支持されている [2]。

Twitter には、ジャンルを問わず様々な話題に関する個人の意見や感情が発信されており、様々な問題や行為に対する世論を知るための有益な情報源になっている。と同時に、友人や知人とのコミュニケーション手段としても用いられており、プライベートな情報が流通する場にもなっている。そして、こういった多種多様な話題に対する様々な情報から抽出される感情あるいは感情極性（ポジティブ/ネガティブ）を利用して、別のサービスにつなげようという研究も少なくなく、多岐にわたっている。例えば、炎上（投稿者が意図しない投稿の広がり）の検知 [3] やニュース記事閲覧者のアクセス解析 [4]、相談窓口における苦情処理の支援 [5]、特定銘柄の株価予測 [6]、災害時における被害情報の抽出 [7]、フォロワーの推薦 [8]、サッカー選手の自動採点 [9] などが挙げられる。

文章からその書き手の感情を推測するためのツールとして様々な感情分析ツールが公開されているが、単一の感情分析ツールには得意不得意があり、推測可能な感情の種類が限られていたり、

精度良く推測できなかったり、推測不可とエラーを返したりすることもある。

そこで本論文では、140 文字以内の短い文章（ツイート）を対象に、複数の感情分析ツールを統合することで、ツイート投稿者の様々な感情をツイートからより頑健に、より高精度に推測できるようになるメタ感情分析手法を提案し、メタ感情分析という考え方の可能性を検証する。

なお、本論文で対象とする感情は、国語学者の中村明が提唱した 10 種類の感情分類「哀、厭、安、怖、昂、好、喜、驚、怒、恥」[10] をベースにした「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 種類であり、ツイート投稿者の感情の強さを「感情が強く込められている（4 点）、割と込められている（3 点）、やや込められている（2 点）、あまり込められていない（1 点）、全然込められていない（0 点）」という 5 段階尺度に準じた実数値（0~4）として定量化する。また、感情分析ツールには、感情ラベルを付与するタイプのものと感情の強さを定量的に算出するタイプのものがあるが、本論文では、後者の中から、複数の感情を対象とするツールを 3 種類、感情極性を対象とするツールを 2 種類の計 5 種類を採用する。

ここで、本論文の以下の構成を示す。まず、2 章で関連研究について述べ、本研究の新規な点を示す。次に、3 章でツイートを讀んだときに感じるツイート投稿者の感情をアンケート調査に基づいて定量化する。4 章でツイート投稿者の感情（アンケート調査結果の平均値）を目的変数、5 種類の感情分析ツールによって算出された各ツイートの感情（感情値/感情極性値）を説明変数とする重回帰分析を行い、各感情分析ツールの出力値を統合したメタ感情分析手法を提案する。5 章で提案手法の精度を評価し、メタ感情分析という考え方の可能性を検証する。最後に、6 章で本論文のまとめと今後の課題について述べる。

2 関連研究

本章では、文章から感情あるいは感情極性を推測するための手法をいくつか取り上げる。以下に示すように、様々な感情分析手法が提案されているが、感情あるいは感情極性を定量的に扱える手法であれば、たとえ出力として感情ラベルや感情極性ラベルを付与する手法であっても、メタ感情分析手法に統合することは可能と考えられる。本論文では、いくつかの既存の感情分析ツールの出力値を統合することで、様々な感情をより頑健に、より高精度に推測できるようになることを示し、メタ感情分析という考え方の可能性を検証している点が新規な点と言える。

複数の感情ラベルを対象に、感情ラベルの付与判定を行うための手法がある。例えば、松林らは、「喜」、「怒」、「哀」、「楽」に「無感情」を加えた 5 種類の感情を対象に、Word2Vec [11] を用いてツイートをベクトル化し、ランダムフォレストにより構築した分類器を用いることで、ツイートに感情ラベルを付与する手法 [3] を提案している。一方、本間らは、著者らの研究と同様、中村明の 10 感情 [10] を採用しており、この 10 感情に「無感情」を加えた 11 種類の感情ラベルの中から一つを付与する手法 [4] を提案している。具体的には、一般化ニューラルネットワーク（GRNN: Generalized Regression Neural Network）[12] を用いて

¹ 正会員 千葉工業大学 情報科学部 情報ネットワーク学科
kumamoto@net.it-chiba.ac.jp

² 非会員 フリー
jmq327279161@gmail.com

ツイートに対する各感情のスコアを求め、最大値をとる感情のラベルを付与するが、その最大値が閾値以下の場合には感情ラベルとして「無感情」を付与する。ただし、GRNN を用いてツイートの感情極性値も算出し、その結果、ポジティブと判定されたにもかかわらず、ネガティブな感情ラベル「怒、哀、怖、厭、恥」が付与されている場合や、ネガティブと判定されたにもかかわらず、ポジティブな感情ラベル「喜、好、安」が付与されている場合は、感情ラベルを「無感情」に変更する。

感情極性（ポジティブ/ニュートラル/ネガティブ）を判定するための手法がある。例えば、秋山らは、節単位で各単語の原形や品詞名、極性値 [13]、および極性反転語（逆接の接続詞や接続助詞など）の出現数を抽出し、条件付き確率場（CRF: Conditional Random Field）を用いて感情極性を判定する手法 [5] を提案している。

感情極性の強さを判定するための手法がある。例えば、三和らは、単語の感情極性を「ポジティブ」(1.0)、「少しポジティブ」(0.4)、「どちらでもない」(0.0)、「少しネガティブ」(-0.4)、「ネガティブ」(-1.0) の 5 段階に分けるとともに、単語の感情極性への影響を「強調」(1.5)、「補足」(1.1)、「弱化」(0.2)、「否定(打消)」(-0.8) の 4 段階に分け、独自に構築した極性値辞書（単語の極性値を表す辞書）と影響値辞書（感情極性への影響度を数値化した辞書）を用いてツイートの感情極性値を算出し、その感情極性を判定する手法 [14] を提案している。

本問らの手法 [4] や秋山らの手法 [5]、三和らの手法 [14] では、出力として感情ラベルあるいは感情極性ラベルを付与しているが、その計算過程で、感情や感情極性を数値化しているため、メタ感情分析手法に統合することが可能と考えられる。

一方、複数の学習器を用意し、それぞれの学習器に学習データを与え学習させた上で融合することで、精度の向上を目指すアンサンブル学習 [15] [16] と呼ばれる学習方法がある。Akhtar らの研究 [15] では、3 つの深層学習手法（CNN: Convolution Neural Network [17]、LSTM: Long Short Term Memory Network [18]、GRU: Gated Recurrent Unit [19]）と 1 つの機械学習手法（SVR: Support Vector Regression [20]）を用い、経済に関するニュースやマイクロブログ上のテキスト情報から株式市場のセンチメントを -1 （弱気） $\sim +1$ （強気）の範囲で推定する手法を提案している。また、Han らの研究 [16] では、4 つの回帰モデル（Linear Regression [21]、Huber Regression [22]、GBDT: Gradient Boost Decision Trees [23]、XGBoost [24]）を組み合わせて、ツイート投稿者の 4 種類の感情（anger, sadness, joy, fear）の強度を $0\sim 1$ の範囲で推定する手法を提案している。このアンサンブル学習という考え方と本研究のアプローチは似ているが、本研究で用いた感情分析ツールはいずれも個別に設計・開発された既存品であるため、提案手法が対象とするツイートを用いて学習させることができない点や提案手法が対象とする感情と完全には一致しない感情（あるいは感情極性）を対象に設計されている点が異なっている。加えて、本研究では、対象とする感情に合わせて設計されたわけではない既存の感情分析ツールの出力値を統合することで、様々な感情をより頑健に、より高精度に推測できるようになることを示している点が新規な点と言える。

表 1 予備調査 2 における設問内容

設問番号	設問内容
Q1, Q3, Q5	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、 <u>今ツイートをするようなイメージ</u> で入力してください。
Q2, Q4, Q6	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？以下の選択肢の中からお答えください。

3 ツイート感情データの取得

本章では、3.1 節で 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、3.2 節でヘビー投稿ユーザが普段使用しているツイートを収集する。3.3 節で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ツイート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。

3.1 ヘビーユーザの抽出（予備調査 1）

本節では、アンケート調査（予備調査 1）を行い、一定以上の頻度で Twitter 上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 Twitter を閲覧しているヘビー閲覧ユーザを抽出する。なお、本論文ではすべてのアンケート調査をインターネット調査会社 [25] を介して Web 上で実施している。

まず、20 代～50 代のインターネットユーザを対象に予備調査 1 を実施し、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、「はい」、「いいえ」、「ツイッターが何かわからない」の三択に対し「はい」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 8 つの中から、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 8 つの中から最も近いものをそれぞれ 1 つずつ選んでもらった。

最終的には、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。

3.2 ツイートの収集（予備調査 2）

本節では、アンケート調査（予備調査 2）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ツイート投稿者が込めた感情をデータとして取得する。なお、ツイート投稿者の感情に

表 2 本調査における設問例

設問内容
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、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？
健康診断受診しました。健康でした。

関するデータは、本調査（3.3 節）で用いるツイートを取捨選択する際に用いられる。

まず、前節で抽出したヘビー投稿ユーザを対象に予備調査 2 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、表 1 に示したように、設問 Q1, Q3, Q5 でツイートを入力してもらい、続く設問 Q2, Q4, Q6 で前問で入力したツイートに込めた感情を回答してもらった。このとき、Q1 と Q2 への回答は必須としたが、Q3 以降は任意とし、回答できる人のみが回答した。また、感情評価のために回答者に提示した評価項目は、「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、それぞれの評価項目に対し、「感情を強く込めた」(4 点)、「割と込めた」(3 点)、「やや込めた」(2 点)、「あまり込めていない」(1 点)、「全然込めていない」(0 点)の中から最も近いものを 1 つ選んでもらった。

以上の結果、2,786 件分のツイートと投稿者の感情に関するデータが得られた。ただし、このデータ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) 「なし」や「日頃の生活」、「サッカーについて」といったツイートではないと思われるデータやハッシュタグ付きのデータ、伏せ字、絵文字、顔文字等を含むデータを削除する、(2) 10 個の評価項目のうち 7 個以上において「感情を強く込めた」もしくは「割と込めた」が選ばれているような感情評価の結果に偏りがあるデータを削除する、(3) 重複した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、という手順で著者らの判断により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 件分のツイートと投稿者感情データ (10 感情 × 5 段階評価値) が得られた。最終的には、この 838 件分のデータから、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 200 ツイート分のデータからなる 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 (全部で 20 ツイート) とその投稿者感情データが含まれている。

3.3 ツイート感情データの取得 (本調査)

本節では、アンケート調査 (本調査) を行い、ツイートを読んで感じるツイート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 節で抽出したヘビー閲覧ユーザを対象に本調査を実

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、一人の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。

回答者に提示した設問の例を表 2 に示す。感情評価の方法は投稿者感情データ取得時とほぼ同じであり、各回答者には、提示された 10 個の評価項目の各々に対し、「感情が強く込められている (4 点)」、「割と込められている (3 点)」、「やや込められている (2 点)」、「あまり込められていない (1 点)」、「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、20 ツイート × 10 感情のデータに対する 5 段階評価において一度も評価されなかった点数が 3 つ以上ある回答者のデータを評価結果に偏りのあるノイズデータと見做し、削除した。そのため、1 ツイート当たりの回答者数は 208～219 人になった。

以上の結果、200 ツイートに対するツイート感情データ (200 ツイート × 10 感情 × 5 段階評価の平均値) が得られた。ツイート感情データの例を表 3 に示す。

4 メタ感情分析手法の提案

本章では、4.1 節でメタ感情分析手法のベースとなる各感情分析ツールの概要を示す。4.2 節でツイート感情データ (200 ツイートに対する 5 段階評価の平均値) を目的変数、それぞれの感情分析ツールによって 200 ツイートから算出された出力値を説明変数とする重回帰分析を感情毎に行い、各感情分析ツールの出力値を統合する重回帰式を作成する。本論文で提案するメタ感情分析手法は、感情毎に用意された重回帰式に各感情分析ツールの出力値を代入することで、その感情の感情値を算出するというものである。

4.1 採用した感情分析ツールの概要

本節では、メタ感情分析手法のベースとなる各感情分析ツールの概要を示すとともに、各感情分析ツールの出力に欠損値が生じた場合の処理についても述べる。なお、本研究では、複数の感情を推測するためのツールとして、著者らが提案した印象マイニングシステム [26] と User Local 社のテキスト感情認識 AI [27]、および Metadata 社の感情解析 API [28] を採用し、感情極性を求めるためのツールとして、Google Cloud の Natural Language (NL) API [29] と Metadata 社の高精度ネガポジ API [30] を採用する。

ここで、参考のために、表 3 に示した 6 つのツイートを入力とした場合の各感情分析ツールによる出力値を表 4 にまとめる。なお、表 4 において記号「—」は欠損値であることを示している。

4.1.1 印象マイニングシステム

著者らが提案した印象マイニングシステム [26] は、新聞記事データベースを解析することで、それぞれの記事に現れた各単語がその記事の印象に及ぼす影響を数値化し、印象辞書を構築することができる。具体的には、「ある印象を有する単語はその印象を表現する印象語群と共に^{*1} しやすく、逆の印象を表現する印象

^{*1} 文献 [26] では、2 つの語が同じ記事内で用いられていることを共起と定義している。

表3 ツイート感情データの例

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
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
息子が大学に合格しました～！！	0.21	0.21	2.70	0.22	3.27	2.29	3.58	1.96	0.14	0.30
電車が満員 くさい 最悪	2.22	3.37	0.20	1.55	0.80	0.17	0.14	1.46	3.30	0.51
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
お店に行ってみただけ売り切れだったー。	2.98	1.20	0.33	0.71	0.86	0.37	0.33	1.29	1.66	0.51
今日の歌番組、好きな歌手出てる～！	0.22	0.25	1.63	0.23	3.08	3.25	3.37	2.12	0.18	0.36

表4 各感情分析ツールの出力値 (表3で示したツイートを対象に)

No.	ツイート	印象マイニングシステム							
		嬉しい ⇔ 悲しい	好き ⇔ 嫌い	安心 ⇔ 不安	笑い ⇔ 怒り	誉れ ⇔ 恥	高揚 ⇔ 冷静	驚く ⇔ 平気	
1	娘が歩いた！	0.40	0.60	0.46	0.55	0.30	0.51	0.46	
2	息子が大学に合格しました～！！	0.47	0.58	0.41	0.54	0.33	0.46	0.54	
3	電車が満員 くさい 最悪	0.44	0.63	0.42	0.52	0.36	0.57	0.49	
4	今日の前で、火災が起きました。	0.30	0.74	0.58	0.61	0.28	0.33	0.45	
5	お店に行ってみただけ売り切れだったー。	0.50	0.70	0.39	0.55	0.50	0.43	0.56	
6	今日の歌番組、好きな歌手出てる～！	0.52	0.68	0.48	0.65	0.37	0.59	0.45	

No.	User Local テキスト感情認識 AI					Metadata 感情解析 API			Google Cloud Natural Language API	Metadata 高精度ネガポジ API
	喜び	好き	恐れ	悲しみ	怒り	好ましい ⇔ 嫌い	嬉しい ⇔ 悲しい	怒り ⇔ 怖れ	ポジティブ ⇔ ネガティブ	ポジティブ ⇔ ネガティブ
1	—	—	—	—	—	0	0	0	0.2	0
2	0.85	0.32	0.60	0.16	0.49	0	0	0	0.8	0
3	0.01	0.02	0.65	0.15	0.43	0	-3	0	-0.9	-1
4	0.40	0.43	0.67	0.12	0.45	0	0	0	0.3	0
5	0.15	0.15	0.51	0.49	0.11	0	0	0	-0.1	0
6	0.65	0.88	0.05	0.12	0.38	1	0	0	0.6	2

語群とは共起しにくい」という仮定のもと、新聞記事データベースに現れる任意の単語と印象軸を構成する対比的な印象を有する2つの印象語群との共起の仕方を調べ、どちらの印象語群とより共起しやすいかを数値化することにより、印象辞書(単語が記事の印象に及ぼす影響を数値化したもの)を構築している。このとき、文献[26]では、「楽しい⇔悲しい」、「嬉しい⇔怒り」、「のどか⇔緊迫」という3種類の印象軸を定義し、それぞれの印象軸に対比的な印象を有する2つの印象語群を用意している。そこで本論文では、本研究で対象とする10種類の感情を念頭に、「嬉しい⇔悲しい」、「好き⇔嫌い」、「安心⇔不安」、「笑い⇔怒り」、「誉れ⇔恥」、「高揚⇔冷静」、「驚く⇔平気」という7種類の感情軸を定義し、それぞれの感情軸に対比的な感情を有する2つの感情語群を用意した上で、印象マイニングシステムを用いて感情辞書を構築した。ここで、それぞれの感情軸に用意した感情語群を表5に示す。なお、この感情辞書を用いた場合、出力となる感情値は、0~1の実数値であり、感情軸の左側の感情が強

ければ1に近づき、右側の感情が強ければ0に近づく。

4.1.2 User Local テキスト感情認識 AI

User Local 社のテキスト感情認識 AI [27] は、深層学習ベースのツールであり、文章を入力すると、「喜び」、「好き」、「恐れ」、「悲しみ」、「怒り」の各感情に対し、0~1の実数値を出力する。このとき、数値が1に近づくほどその感情が強いことを表し、0に近づくほどその感情が弱いことを表している。なお、入力する文章の文字数は20~300文字が推奨されており、10文字未満が入力された場合はエラーとなる。そのため、表4に示したように、6文字からなるツイート「娘が歩いた！」に対しては欠損値が生じている。

4.1.3 Metadata 感情解析 API

Metadata 社の感情解析 API [28] は、文章を「好ましい⇔嫌い」、「嬉しい⇔悲しい」、「怒り⇔怖れ」という3種類の感情軸で評価し、それぞれの感情軸に対し+3~-3の整数値(7段階尺度値)を出力する。それぞれの感情軸の左側の感情が強ければ

表 5 各感情軸を構成する感情語群

感情軸	感情語
嬉しい ⇔ 悲しい	うれしい, 嬉しい, 喜ぶ, 歓喜する 悲しい, 哀しい, 悲しむ, 哀しむ, 嘆く
好き ⇔ 嫌い	好きだ, 大好きだ, 愛好する 嫌いだ, 大嫌いだ, 嫌悪する
安心 ⇔ 不安	安心だ, 安心する, 平気だ, 安堵する 不安だ, 心配だ, 心配する, 恐れる, 怖い, 恐怖する, 危惧する, 憂慮する
笑い ⇔ 怒り	笑う, 大笑いする, 喜ぶ 怒る, 憤る, 激怒する
誉れ ⇔ 恥	誉れだ, 名誉だ, 栄誉だ, 誇る 恥ずかしい, 恥だ, 気恥ずかしい, 不名誉だ, 恥じる
高揚 ⇔ 冷静	高揚する, 高ぶる, 熱烈だ, 興奮する 冷静だ, 落ち着く
驚く ⇔ 平気	驚異的だ, 驚く, 驚愕する, 驚嘆する, びっくりする 平気だ, 平然とする, 平然たる

+3, 右側の感情が強ければ -3 となり, 中立な感情の場合は 0 となる。

4.1.4 Google Cloud Natural Language API

Google Cloud の Natural Language (NL) API [29] は, 入力された文章の感情極性を数値化し, +1~-1 の実数値を出力する。すなわち, この感情分析ツールの出力値は, 入力された文章の背景にある感情が全体的にポジティブであれば +1 に近づき, 逆にネガティブであれば -1 に近づく。出力値が 0 に近いときは, ニュートラルと捉えることができる。

4.1.5 Metadata 高精度ネガポジ API

Metadata 社の高精度ネガポジ API [30] は, 文章を解析し, その内容がどの程度肯定的かあるいは否定的かに応じて +3~-3 の整数値 (7 段階尺度値) を出力する。すなわち, とても肯定的と判断すれば +3 を出力し, 逆にとても否定的と判断すれば -3 を出力する。なお, この高精度ネガポジ API は, 辞書ベースのツールであり, 数万以上の形容表現・価値判断表現類に対しあらかじめ付与された 7 段階のネガポジ属性「excellent, very good, good, neutral, no good, bad, very bad」を用いて肯定的・否定的の度合いを算出している。

4.1.6 各感情分析ツールの出力に欠損値が生じ場合の処理

メタ感情分析手法は, 各感情分析ツールの出力に欠損値が生じた場合, 中間的な値で代替する。すなわち, 著者らの印象マイニングシステムと User Local 社のテキスト感情認識 AI では, 0~1 の範囲で実数値が出力されるので, 0.5 という値で代替し, Metadata 社の感情分析 API と高精度ネガポジ API および Google Cloud の NL API では, $+x \sim -x$ ($x = 1$ or 3) の範囲で値が出力されるので, 0 という値で代替する。なお, 後述のベースライン手法でも同じ処理が行われる。

4.2 重回帰分析による統合

本節では, 各感情分析ツールの出力値を統合するために, 重回帰分析を行い, 提案するメタ感情分析手法の核となる重回帰式を作成する。

まず, 3.3 節で取得したツイート感情データ (200 ツイート × 対象とする感情 × 5 段階評価の平均値) を目的変数, それぞれの感情分析ツールによって算出された出力値 (200 ツイート × (15 感情値 + 2 感情極性値)) を説明変数とする重回帰分析を感情毎に行った。その結果得られた重回帰式を表 6 にまとめるとともに, このときの自由度修正済み決定係数を表 7 に示す。なお, 重回帰分析にはエスミ社のソフトウェア「EXCEL 多変量解析 Ver.7.0」を用い, 変数選択には変数増減法を採用している。

表 7 によれば, 自由度修正済み決定係数は概ね 0.5 以上となっており, 得られた重回帰式の元のデータに対する当てはまり具合が良いことを示している。しかしながら, 「驚き」と「恥ずかしい」に対しては, それぞれ 0.05, 0.21 という小さい値となっており, 採用した説明変数がそれらの感情を表現するのに向かなかったことを示している。

なお, 本研究では, 提案手法の出力として, 「感情が強く込められている (4 点), 割と込められている (3 点), やや込められている (2 点), あまり込められていない (1 点), 全然込められていない (0 点)」という 5 段階尺度に準じた 0~4 の間の実数値を想定しているが, 重回帰式を用いる性質上, 0 より小さい数値や 4 より大きい数値が算出されてしまう場合がある。そこで, 提案手法では, 0 より小さい数値が算出された場合は 0 に, 4 より大きい数値が算出された場合は 4 に丸めることにする。

5 精度評価によるメタ感情分析の可能性検証

本章では, 複数の感情分析ツールを統合するというアプローチ (メタ感情分析という考え方) の可能性を検証する。具体的には, まず, 5.1 節でそれぞれの感情分析ツールをベースにベースライン手法 (計 5 種類) を作成する。そして, 5.2 節で重回帰分析用の学習データ (200 ツイート分のツイート感情データ) を用いて, 5 種類のベースライン手法と提案手法であるメタ感情分析手法の精度を評価し, 比較するとともに, 5.3 節で学習データを用いて 5 分割交差検定を行い, 未知データに対する提案手法の精度を評価することで, メタ感情分析という考え方の可能性を検証する。

5.1 ベースライン手法の作成

本論文で用いる各感情分析ツールは, 本論文で対象としている 10 種類の感情に合わせて設計されているわけではない。そのため, 感情分析ツールをそのままの形でベースライン手法として用いても十分な精度が見込めない。そこで本節では, 以下の手順で 1 つの感情分析ツールからベースライン手法を 1 つ作成し, 計 5 種類のベースライン手法を用意する。

まず, それぞれの感情分析ツールを用いて, 学習データとなるツイート感情データ (200 ツイート × 対象とする感情 × 5 段階評価の平均値) を目的変数, 当該感情分析ツールによって 200 ツイートから算出された出力値を説明変数とする重回帰分析 (変数増減法) を感情毎に行い, ベースライン手法を作成した。なお, 複数の感情を対象とする印象マイニングシステムや User Local テ

表 6 提案手法であるメタ感情分析手法の核となる重回帰式（各説明変数に対する偏回帰係数と定数項）

説明変数	目的変数									
	悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
嬉しい ⇔ 悲しい			0.910							0.706
好き ⇔ 嫌い				-0.716						
印象 安心 ⇔ 不安	-1.253		0.918	-1.011		0.921				-0.984
マイニング 笑い ⇔ 怒り	-0.887	-1.372			0.755	0.744	1.167		-2.194	
システム 誉れ ⇔ 恥	-0.660	-0.616							-0.638	-0.536
高揚 ⇔ 冷静				-0.545	0.964	0.972				
驚く ⇔ 平気		-0.950							-0.729	
User Local 喜び	-0.878	-0.768	0.828	-0.753	1.111	0.751	1.616	-0.388	-0.718	-0.362
好き	-0.520	-0.369				0.438				
テキスト 恐れ			-0.413	0.787	-0.414	-0.752	-0.732			
感情認識 悲しみ	0.737	-0.480								0.243
AI 怒り		0.688	-0.445			-0.494	-0.550		1.131	
Metadata 好ましい ⇔ 嫌い					0.192					
感情解析 嬉しい ⇔ 悲しい								0.153		
API 怒り ⇔ 怖れ										
Google Cloud NL API										
ポジティブ ⇔ ネガティブ	-0.442	-0.804	0.374	-0.237	0.499	0.653	0.726	0.215	-0.825	-0.119
Metadata 高精度ネガポジ API										
ポジティブ ⇔ ネガティブ	-0.075					0.120	0.124			
定数項	3.043	3.114	0.095	2.076	0.392	-0.098	0.521	1.402	3.040	1.012

表 7 提案手法の自由度修正済み決定係数

悲しい	0.61	好き	0.70
嫌い	0.67	喜び	0.67
安心	0.49	驚き	0.05
怖れ	0.53	怒り	0.69
高揚	0.56	恥ずかしい	0.21

キスト感情認識 AI, Metadata 感情解析 API を用いたときは、それぞれ複数個の説明変数がいれたため、提案手法と同様、感情毎に重回帰式が作成されたが、感情極性を対象とする Google Cloud NL API と Metadata 高精度ネガポジ API を用いたときは、いずれも説明変数が 1 つしかなかったため、各感情に合わせてその説明変数の値を補正するための回帰式（1 次関数）が作成された。

ここで、それぞれの重回帰分析における自由度修正済み決定係数を表 8 に示す。表 8 によれば、自由度修正済み決定係数が 0.5 以上であったのは、User Local テキスト感情認識 AI ベースのベースライン手法を用いた際の「悲しい」、「嫌い」、「好き」、「喜び」、「怒り」と Google Cloud NL API ベースのベースライン手法を用いた際の「嫌い」、「好き」、「怒り」のみであり、多くの場合において重回帰分析があまりうまくいっていないことがわかる。表 7 に示した提案手法作成時の自由度修正済み決定係数と比

べてみると、すべての感情において提案手法作成時の自由度修正済み決定係数の方が良いことがわかる。

5.2 学習データに対する精度評価

本節では、学習データに対する提案手法とベースライン手法の精度を評価し、比較する。

まず、重回帰分析に用いた学習データ（200 ツイート分のツイート感情データ）を用いて、提案手法であるメタ感情分析手法と 5.1 節で作成した 5 種類のベースライン手法の精度を評価した。結果を表 9 にまとめる。なお、精度を評価するための指標としては RMSE（Root-Mean-Square Error）を採用し、次の式により求めた。

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}$$

ただし、 x_i は学習データにおける i 番目のツイートの感情値であり、 y_i は当該手法によってそのツイートから算出された感情値を表している。 N は全ツイート数であり、ここでは $N = 200$ となる。一方、表 9 に示した改善率は、提案手法に対する各ベースライン手法からの RMSE の改善率であり、次の式により求めた。

$$\text{改善率} = \frac{RMSE_{\text{ベースライン手法}} - RMSE_{\text{提案手法}}}{RMSE_{\text{ベースライン手法}}} \times 100$$

表 9 によれば、すべての感情において提案手法の精度が最も良

表 8 ベースライン手法の自由度修正済み決定係数

感情	ベースライン手法				
	印象マイニングシステム	User Local テキスト感情認識 AI	Metadata 感情解析 API	Google Cloud NL API	Metadata 高精度ネガポジ API
悲しい	0.34	0.52	0.18	0.42	0.24
嫌い	0.39	0.50	0.15	0.55	0.19
安心	0.31	0.41	0.11	0.36	0.18
怖れ	0.26	0.48	0.13	0.31	0.18
高揚	0.30	0.44	0.22	0.41	0.26
好き	0.37	0.57	0.23	0.53	0.30
喜び	0.36	0.57	0.20	0.49	0.27
驚き	0.02	0.01	0.03	0.03	0.02
怒り	0.41	0.52	0.13	0.54	0.18
恥ずかしい	0.14	0.16	0.04	0.11	0.03
平均	0.29	0.42	0.14	0.37	0.19

表 9 学習データに対する各手法の精度 (RMSE) と改善率

感情	提案手法	ベースライン手法					
	メタ感情分析手法	印象マイニングシステム	User Local テキスト感情認識 AI	Metadata 感情解析 API	Google Cloud NL API	Metadata 高精度ネガポジ API	
悲しい	0.556	0.734 (24.3%)	0.624 (10.9%)	0.827 (32.8%)	0.694 (20.0%)	0.796 (30.2%)	
嫌い	0.531	0.736 (27.8%)	0.672 (20.9%)	0.876 (39.4%)	0.639 (16.9%)	0.855 (37.9%)	
安心	0.519	0.609 (14.8%)	0.563 (7.8%)	0.696 (25.4%)	0.589 (11.9%)	0.669 (22.5%)	
怖れ	0.409	0.511 (20.0%)	0.432 (5.3%)	0.559 (26.8%)	0.501 (18.3%)	0.544 (24.9%)	
高揚	0.556	0.705 (21.1%)	0.632 (12.0%)	0.748 (25.7%)	0.654 (15.0%)	0.732 (24.0%)	
好き	0.547	0.809 (32.4%)	0.669 (18.2%)	0.900 (39.2%)	0.708 (22.8%)	0.860 (36.4%)	
喜び	0.676	0.949 (28.8%)	0.778 (13.1%)	1.066 (36.6%)	0.853 (20.8%)	1.020 (33.7%)	
驚き	0.613	0.627 (2.2%)	0.629 (2.5%)	0.622 (1.4%)	0.625 (1.8%)	0.626 (2.0%)	
怒り	0.566	0.795 (28.8%)	0.722 (21.6%)	0.974 (41.9%)	0.706 (19.8%)	0.946 (40.2%)	
恥ずかしい	0.338	0.354 (4.7%)	0.351 (3.9%)	0.378 (10.6%)	0.364 (7.1%)	0.379 (10.9%)	
平均 RMSE	0.531	0.683 (20.5%)	0.607 (11.6%)	0.765 (28.0%)	0.633 (15.4%)	0.743 (26.3%)	

括弧内のパーセンテージは提案手法に対する各ベースライン手法からの RMSE の改善率を示す。

く、0.34~0.68 の範囲であった。精度が最も良かったのは、「恥ずかしい」の 0.34 であり、ベースライン手法から 3.9%~10.9% の改善であった。しかしながら、「恥ずかしい」に対する自由度修正済み決定係数は、表 7 に示したように、0.21 と小さい値であった。にもかかわらず、精度が良かったのは、ツイート感情データにおける標準偏差が「恥ずかしい」以外の感情では 0.61 (「怖い」)~1.20 (「喜び」) の範囲であったのに対し、「恥ずかしい」では 0.39 と小さく、元々の感情値の振れ幅が小さかったためと考えられる。逆に、精度が最も悪かったのは、「喜び」の 0.68 であったが、標準偏差が 1.20 と最も大きかったためであり、ベースライン手法の精度 (0.85~1.07) と比べると、13.1%~36.6% の改善となっている。なお、自由度修正済み決定係数が最も小さい 0.05 であった「驚き」の精度は 0.61 と良くなかったが、ベースライン手法の精度 (0.62~0.63) よりは良く、1.4%~2.5% の改善であった。

なお、今回用いた 200 ツイートに対し、推測不可とエラーを返したツイートの数は、印象マイニングシステムで 1 件、User Local テキスト感情認識 AI で 15 件、Metadata 高精度ネガポジ API で 2 件あったが、提案手法では 0 件であった。したがって、複数の感情分析ツールを統合することで、より頑健な感情分析を実現できると言える。

5.3 未知データに対する精度評価 (5 分割交差検定)

本節では、重回帰分析に用いた学習データ (200 ツイート分のツイート感情データ) を用いて 5 分割交差検定を行い、未知データに対する提案手法の精度を評価する。

まず、学習データを均等に分割することで 40 ツイート分のツイート感情データからなるデータセットを 5 つ作成した。次に、任意の 4 つのデータセットを用いて 4.2 節に示した方法で重回帰分析を行い、メタ感情分析手法を作成した。このメタ感情分析手法を用いて残りのデータセット (40 ツイート) から感情値を算出

表 10 未知データに対する提案手法の精度 (RMSE)

未知データセット	悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
CV1	0.57	0.57	0.58	0.38	0.59	0.56	0.67	0.62	0.66	0.35
CV2	0.61	0.61	0.62	0.50	0.59	0.63	0.83	0.75	0.67	0.36
CV3	0.46	0.52	0.46	0.37	0.64	0.49	0.57	0.61	0.55	0.24
CV4	0.69	0.54	0.54	0.54	0.62	0.66	0.74	0.61	0.60	0.44
CV5	0.68	0.74	0.66	0.43	0.63	0.72	0.84	0.70	0.68	0.50
平均 RMSE	0.60	0.60	0.57	0.44	0.61	0.61	0.73	0.66	0.63	0.38

表 11 未知データに対する提案手法の精度と学習データに対するベースライン手法の精度を比較した場合の改善率

ベースライン手法	悲しい	嫌い	安心	怖れ	高揚	好き	喜び	驚き	怒り	恥ずかしい
印象マイニングシステム	17.8%	19.0%	6.2%	13.3%	13.0%	24.3%	23.1%	-4.8%	20.6%	-6.8%
UserLocal テキスト感情認識 AI	3.3%	11.3%	-1.5%	-2.7%	2.9%	8.3%	6.2%	-4.5%	12.5%	-7.7%
Metadata 感情分析 API	27.1%	32.0%	17.9%	20.6%	18.0%	31.9%	31.6%	-5.7%	35.1%	-0.2%
Google Cloud NL API	13.1%	6.8%	3.0%	11.5%	6.2%	13.4%	14.5%	-5.2%	10.5%	-4.1%
Metada 高精度ネガポジ API	24.2%	30.3%	14.7%	18.6%	16.2%	28.8%	28.4%	-5.0%	33.2%	0.2%

し、そのデータセットの感情値との間で精度 (RMSE) を求めた ($N = 40$)。その結果を表 10 にまとめる。

表 10 によれば、各感情に対する平均 RMSE は「恥ずかしい」の 0.38～「喜び」の 0.73 の範囲であり、表 9 に示した学習データに対する提案手法の精度と比べてみると、7.1%～12.2% ほど悪くなっているのがわかる。しかしながら、ここで、未知データに対する提案手法の精度 (表 10 における平均 RMSE) と学習データに対する各ベースライン手法の精度 (表 9 における RMSE) を比較し、改善率を求めてみたところ、表 11 に示すような結果となった。

表 11 において、マイナスの数値は未知データに対する提案手法の精度が学習データに対する当該ベースライン手法の精度より悪いことを意味しているが、マイナスの数値となったのはほとんどが「驚き」と「恥ずかしい」に関してであり、他の感情に対してはプラスの数値となっている。すなわち、未知データに対する提案手法の精度が学習データに対する当該ベースライン手法の精度より良かったことを示している。したがって、複数の感情分析ツールを統合するというアプローチにより、様々な感情を対象にできるようになるだけでなく、個々の感情をより高精度に推測できるようになると言える。

6 まとめ

本論文では、複数の感情分析ツールの出力値を重回帰分析により統合することで、ツイート投稿者の様々な感情をツイートからより頑健に、より高精度に推測できるようになるメタ感情分析手法を提案し、メタ感情分析という考え方の可能性を検証した。具体的には、感情分析ツールとして、複数の感情の強さを定量的に算出することができるツール 3 種類と感情極性の度合いを定量的に算出することができるツール 2 種類の計 5 種類を採用し、重回帰式を用いて各感情分析ツールの出力値を統合するメタ感情分析手法を提案した。提案手法の精度を学習データと未知データを用

いて評価した結果、「悲しい、嫌い、安心、怖れ、高揚、好き、喜び、驚き、怒り、恥ずかしい」という 10 種類の感情のうち「驚き」と「恥ずかしい」を除く 8 種類の感情に対し、既存の感情分析ツールをベースにして作成されたベースライン手法より高い精度で感情の強さを推測できるようになったことから、メタ感情分析という考え方の可能性が示されたと言える。

今後の課題として、各感情に対する精度の改善が挙げられる。メタ感情分析というアプローチにより、既存の感情分析ツールより精度は向上したが、個々の感情に対する精度としては十分とは言えない。表 6 に示した提案手法の重回帰式を見てみると、感情分析に貢献している説明変数と貢献していない説明変数があることがわかる。さらに様々な感情分析ツールを加えてみるとともに、よりよい組み合わせ方を探求することで、より高精度なメタ感情分析手法を作成可能と考えられる。一方、今回の統合では、重回帰分析を用いたが、深層学習手法を導入することで、非線形な統合も可能になると考えられ、精度の向上が見込まれる。また、同じツイートを読んでも感じる感情は人によって異なるという個人差の問題がある。今回取得したツイート感情データを用いてどのような個人差が生じうるのかをパターン化するとともに、各パターンに適した感情分析手法を用意することで、より実用的な感情分析を実現できるものと考えられる。今後の課題としたい。

謝辞

本研究の一部は、JSPS 科研費 17K00430, 19H04221, 20K12085, 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://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] 松林圭, 五味京祐, 古川和祈, 松尾祐佳, 松原良和, 日諸マルセロ 優次, 中村拓哉, 山下晃弘, 松林勝志, Twitter 上に投稿された文章に基づく感情推定法とその応用に関する検討, 第 78 回情報処理学会全国大会講演論文集, Vol. 2016, No. 1, pp. 79–80, 2016.
- [4] 本間幸徳, 萩原将文, Twitter を用いたニュース記事における読み手の感情推定手法の提案とアクセス解析への応用, 日本感性工学会論文誌, Vol. 12, No. 1, pp.167–174, 2013.
- [5] 秋山和寛, 三澤賢祐, 成田和弥, 熊本忠彦, 灘本明代, CRF を用いたレビューにおける節単位毎の感情推定, 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), No. J2-4, 7 pages, 2018.
- [6] 増井佑亮, 藤野巖, Twitter の多軸的感情情報を利用した株価の予測, 言語処理学会第 24 回年次大会発表論文集, pp. 69–72, 2018.
- [7] 横田尚己, 山田圭二郎, 阿部倫之, つぶやきと感情極性の関係に着目した災害時の被害情報発見手法の検討 –平成 28 年熊本地震を対象に-, 第 56 回土木計画学研究発表会・講演集, No. P8, 6 pages, 2017.
- [8] 熊本忠彦, 灘本明代, 共通話題に対する感情的態度の類似度に基づくフォロワー推薦, 電子情報通信学会論文誌, Vol. J100-D, No. 4, pp. 500–509, 2017.
- [9] 若狭孟, 横山昌平, Twitter を用いたサッカー選手採点のための感情値辞書構築に向けて, 第 14 回 ARG Web インテリジェンスとインタラクション研究会資料, WI2-2019-06, 4 pages, 2019.
- [10] 中村明, 感情表現辞典, 東京堂出版, 東京, 1993.
- [11] Word2Vec, <https://deeplearning4j.org/ja/ja-word2vec>
- [12] Donald F. Specht, “A General Regression Neural Network,” IEEE Transactions on Neural Networks, Vol. 2, No. 6, pp. 568–576, 1991
- [13] 高村大也, 乾孝司, 奥村学, スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, Vol. 47, No. 2, pp. 627–637, 2006.
- [14] 三和未佐希, 立間淳司, 青野雅樹, 単語位置と強弱表現に着目したツイートの感情分析, 第 13 回情報科学技術フォーラム (FIT2014) 講演論文集, 第 2 分冊, No. E-011, pp. 227–228, 2014.
- [15] Md Shad Akhtar, Abhishek Kumar, Deepanway Ghosal, Asif Ekbal, and Pushpak Bhattacharyya, “A Multilayer Perceptron based Ensemble Technique for Fine-grained Financial Sentiment Analysis,” Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 540–546, Copenhagen, Denmark, 2017.
- [16] Xiaotian Han and Song Jiang, “DMGroup at EmoInt-2017: Emotion Intensity Using Ensemble Method,” Proc. of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 243–248, Copenhagen, Denmark, 2017.
- [17] Yoon Kim, “Convolutional Neural Networks for Sentence Classification,” Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1746–1751, Doha, Qatar, 2014.
- [18] Sepp Hochreiter and Jurgen Schmidhuber, “Long Short-Term Memory,” Neural computation, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [19] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, Yoshua Bengio, “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches,” Proc. of the Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, pp. 103–111, Doha, Qatar, 2014.
- [20] Alex J. Smola and Bernhard Scholkopf, “A Tutorial on Support Vector Regression,” Statistics and Computing, Vol. 14, No. 3, pp. 199–222, 2004.
- [21] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay, “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, Vol. 12, pp. 2825–2830, 2011.
- [22] Jyh-Horng Jeng, Chun-Chieh Tseng, and Jer-Guang Hsieh, “Study on Huber Fractal Image Compression,” IEEE Transactions on Image Processing, Vol. 18, No. 5, pp. 995–1003, 2009.
- [23] Harris Drucker and Corinna Cortes, “Boosting Decision Trees,” In Advances in Neural Information Processing Systems, pp. 479–485, 1995.
- [24] Tianqi Chen and Carlos Guestrin, “XGBoost: A Scalable Tree Boosting System,” Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, San Francisco, USA, 2016
- [25] <https://monitor.fast-ask.com/>
- [26] 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 D, Vol. J94-D, No. 3, pp. 540–548, 2011.
- [27] <https://emotion-ai.userlocal.jp/>
- [28] <https://metadata.co.jp/apis/emotion-analyzer.html>
- [29] <https://cloud.google.com/natural-language?hl=ja>
- [30] <https://metadata.co.jp/apis/negaposi-analyzer.html>