

ツイートの感情の関係に基づく Twitter 感情軸の決定

山本 湧輝[†] 熊本 忠彦^{††} 灘本 明代^{†††}

[†] 甲南大学大学院 自然科学研究科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

^{††} 千葉工業大学 情報科学部 〒275-0016 千葉県習志野市津田沼 2-17-1

^{†††} 甲南大学 知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]m1424010@center.konan-u.ac.jp ^{††}tkumamoto@net.it-chiba.ac.jp ^{†††}nadamoto@konan-u.ac.jp

あらまし マイクロブログの一つである Twitter ではその手軽さから多くのユーザが気軽にツイートを投稿している。このようなマイクロブログサービスには今起こったことや感じたことを投稿している場合が多い為、ユーザの感情が現れやすい。これまでに我々はツイートから抽出する感情の種類を感情軸と呼び、中村明の感情表現辞典で分類されている 10 種類の感情を感情軸と定義するとともに、それぞれの感情軸においてツイートの感情値を求める手法を提案した。しかしながら、システムが算出する感情値と被験者が感じる感情値には一定の誤差があった。その原因として次の 2 点が上げられる。1 つ目の原因として Twitter に使われているツイート特有な単語や顔文字が感情語辞書には含まれていないという問題が上げられる。ツイート特有な単語については今後対応する必要があると考える。2 つ目の原因として感情軸 10 軸は種類が多く、ツイートの感情を 10 軸に分類することは困難であるという問題が上げられる。実際にユーザ実験にて、被験者から「10 軸の感情の種類が多く判断しにくい」といった声があがった。その為、本研究では 10 軸から Twitter 用に次元削減を行うことを提案する。そこで本論文では、既存の 10 軸を用いてツイートの感情を抽出した場合にある軸が他の軸に対してどのような影響を与えているかを「感情の関係」と定義し、この感情の関係から Twitter に適した感情軸を決定する手法を提案する。

キーワード 感情分析, Twitter

1. はじめに

マイクロブログサービスが普及している現在、膨大な情報がインターネット上に投稿されている。中でも Twitter は 140 文字以内のツイートと呼ばれる短文を投稿する為、他の SNS やブログと違い手軽にツイートを投稿することが可能となっている。その為、Twitter には今起こったことや感じたことを気軽に投稿している場合が多い。そして、このようなツイートにはユーザのその時々の感情が現れている場合が多い。これまで我々はツイートから抽出する感情の種類を感情軸と定義し、それぞれの感情軸に対して感情の強さを表す値 (感情値) を求めてきた [1]。ここでは、ツイートから抽出する感情軸を中村明の感情表現辞典 [2] で分類されている「喜」、「好」、「安」、「哀」、「厭」、「怖」、「怒」、「恥」、「昂」、「驚」の 10 軸を用いている。中村の感情表現辞典は 197 人の作家の作品及び 806 点の人間の感情について書かれた文書を分析し、上記の 10 軸に人間の感情を分類している。そして、それらの感情軸各々に感情を示す単語が付与されている。我々はこれらの単語の感情値を求め、ツイートの感情値を求める手法を提案した。ここでは、提案手法の精度を求めするために、ユーザ実験を行った。その結果、システムの判断と被験者の判断には大きな違いがあった。その原因として以下の 2 点が上げられる。

(1) Twitter に使われているツイート特有な単語や顔文字が感情語辞書には含まれていない。

(2) 感情軸 10 軸は種類が多く、ツイートの感情を 10 軸に

分類することは困難である。

(1) については、先の論文 [1] にて顔文字の感情を考慮した感情抽出について提案しているが、ツイート特有な単語については今後対応する必要があると考える。(2) については、実際にユーザ実験にて、被験者から「10 軸の感情の種類が多く判断しにくい」といった声があがった。その為、本研究では 10 軸から Twitter 用に次元削減を行うことを提案する。そこで、既存の 10 軸を用いてツイートの感情を抽出した場合にある軸が他の軸に対してどのような影響を与えているかを「感情の関係」と定義し、この感情の関係から感情軸の次元削減を行い、Twitter に適した感情軸を決定する手法を提案する。

以下、2. で関連研究を整理する。3. で感情軸に基づくツイートの感情抽出手法について、4. で感情の関係について述べ、最後に 5. 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

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数多く提案されている。代表的な感情モデルとして、Plutchik のモデル [3] がある。人間の感情は「嫌悪」⇔「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 [4] 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

表 1 感情語辞書の例

感情軸	感情語	感情値	感情語	感情値	感情語	感情値	感情語	感情値	感情語	感情値
喜	喜び	0.91	楽しい	0.54	面白い	0.85	大切だ	0.54	笑顔	0.70
好	恋しい	0.64	好き	0.71	恋しい	0.62	愛情	0.76	愛	0.80
安	溜め息	0.45	安らぎ	0.81	ほっと	0.82	安心	0.45	平安	0.45
哀	悲しい	0.45	憂鬱	0.82	痛い	0.62	悲痛	0.56	物悲しい	0.67
厭	不快	0.82	嫌いだ	0.75	辛い	0.72	酷い	0.67	汚い	0.34
怖	不気味	0.45	恐怖	0.64	不安	0.55	心配	0.78	恐怖心	0.56
怒	怒る	0.87	怒鳴る	0.54	憤怒	0.72	憤り	0.87	苛立ち	0.67
恥	赤面	0.72	はにかむ	0.87	照れる	0.56	恥ずかしい	0.67	屈辱	0.45
昂	昂る	0.45	発狂	0.57	感動	0.72	興奮	0.34	高揚	0.54
驚	驚く	0.71	歓喜	0.54	慌てる	0.52	ドッキリ	0.65	ショック	0.78

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[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、ツイートに適した感情軸を提案している点で異なる。高岡ら[6]は中村の提案する10次元の感情軸[2]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における感情の関係を分析している点異なるが、高岡らの手法がツイートにも対応できるかを検討する。

テキスト評価分析や文書の極性判定に関する研究では、レビューなどのテキスト文書から評価を抽出し、ポジティブ（肯定）・ネガティブ（否定）に分類する研究が多数ある。中丸ら[7]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。本研究では多次元の感情に対して分析を行っている点異なる。加藤ら[8]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印象の変化を分析している。本研究では怒りの感情だけではなく、多次元の感情に対して分析を行っている点と、Twitterを対象としている点異なる。池川ら[9]はツイート文の感情極性を推定する為に、非言語情報を表現する手段である顔文字に着目して感情極性を推定している。しかしながら、多次元の感情について分析を行っていない為、本研究とは異なる。村上ら[10]は顔文字とその感情情報の辞書を作成する為に顔文字の解釈、分類、使われ方について考察を行っている。しかしながら、多次元的な感情を考慮していない為、本研究とは異なる。

3. 感情軸に基づくツイートの感情抽出手法

これまで我々が提案してきた感情軸とそれを用いた感情語辞書に基づくツイートの感情抽出手法について述べる。その感情抽出手法を用いて求めたツイートの感情値と被験者の判断の違いを比較し、既存の10軸の感情においての問題点を分析する。

3.1 感情語辞書

ツイートから感情を推定する為、まずはじめに単語毎に感情

の定量化を行い、感情語辞書の構築を行う必要がある。ここでは感情語辞書構築の手法として、熊本ら[4]が提案している感情語辞書構築システムを用いる。熊本らの手法は、辞書を構築する為に大量のデータに現れる任意の単語と、あらかじめ定義してある感情語群との共起関係を調べ、その結果に基づき辞書構築を行っている。熊本らが感情語辞書構築システムを用いて構築した感情語辞書は、用いる感情軸を新聞記事を読んだ人々がその記事からどのような印象を受けたかをアンケート調査により調べた結果に基づき、「楽しい⇔悲しい、うれしい⇔怒り、のどか⇔緊迫」の3つに決定している。しかしながら、本研究ではTwitterを対象としているため熊本らの提案している感情語辞書をそのまま用いるのは困難である。その理由は2つあり、1点目は、Twitterの場合はユーザが自由にツイートを投稿している為、様々な感情を持つツイートが多い。その為、熊本らの用いた3つの感情軸で感情を表現することが困難である。そこで本研究では、ツイートに含まれる様々な感情に対応できるように中村の感情表現辞典[2]で分類されている「喜」、「好」、「安」、「哀」、「厭」、「怖」、「怒」、「恥」、「昂」、「驚」の10軸の感情を用い熊本らの感情語辞書構築システムにより感情語辞書を構築する。感情語辞書構築システムでは対極となる感情軸が必要だが、中村の感情軸には対極となる軸が存在しない。その為、Plutchikの感情の輪[3]で向かい合っている感情は対極関係であることを参考にして、対極の感情軸を決定する。具体的には、中村の感情軸をPlutchikの感情に当てはめて「sadness(哀)⇔joy(喜)」、「fear(怖)⇔anger(怒)」、「remorse(厭)⇔love(好)」に決定した。また、それ以外の「安」、「恥」、「昂」、「驚」の感情軸に関してはPlutchikの感情の輪に対応した感情が無かった為、手動で「恥⇔安」、「驚⇔昂」に決定した。

理由の2点目は、新聞記事と違いTwitterではユーザが自由にツイートしている為、口語的な表現が存在する。その為、Twitter用に感情語辞書を再構築する必要がある。しかしながら、ツイートは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現れてない場合が多い。そこで、ツイートと同様に口語的な表現が多く、表記ゆれや文法の乱れが少ないYahoo!映画^(注1)のレビューデータ74,000文書を用いて熊本らのシステムにより辞書構築を行った。その結果、今回用いる各

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

表 2 文の感情値の算出例
ツイート:「野球はすごく面白い」

感情語	喜	哀	怒	怖	好	厭	昂	驚	安	恥
凄い	0.13	0	0.01	0	0.01	0	0.7	0	0	0.38
面白い	0.85	0	0	0.01	0.04	0	0.03	0	0	0.32
合計	0.98	0	0.01	0.01	0.05	0	0.73	0	0	0.7

ツイート:「最近すごく面白くない」

感情語	喜	哀	怒	怖	好	厭	昂	驚	安	恥
凄い	0.13	0	0.01	0	0.01	0	0.7	0	0	0.38
面白くない	0	0	0	0	0	0	0	0	0	0
合計	0.13	0	0.01	0	0.01	0	0.7	0	0	0.38

表 3 実験で用いたデータの一部

ツイート
今日はめっちゃ嬉しいことがいっぱいあった 今親元と離れたばっかだから心配ごと多い めっちゃむかむかする みなさんが無事参加できるのかも心配ですね 赤点取ったら怒るからね! 学校から見る景色は好き

感情軸に対して約 5,600 語~7,500 語の感情語とその感情値を感情語辞書に登録した。構築した感情語辞書の一部を表 1 に示す。

3.2 ツイートの感情抽出

次に 10 軸の感情を用いてツイートの感情を抽出する方法を述べる。ここでは、3.1 節で構築した感情語辞書を用いてツイートの感情を抽出する。具体的には、Twitter から取得したツイートに対して形態素解析エンジン Juman^(注2) を用いて形態素解析する。次に、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。マッチングした感情語の感情値を各軸ごとにそれぞれ加算することで、ツイートの感情値を抽出する。例えば、表 2 の「野球はすごく面白い」というツイートの場合は、形態素に分けると「野球 | は | すごく | 面白い」となる。ツイート内のそれぞれの形態素と感情語辞書に登録されている感情語をマッチングすると、「すごく」と「面白い」がそれぞれマッチングする。マッチングした感情語の感情値を各軸ごとにそれぞれ加算すると、10 軸それぞれの感情値を求めることが出来る。閾値を 0.4 とした場合、この「野球はすごく面白い」というツイートは「喜」と「昂」の感情軸の感情値が閾値以上の為、このツイートの感情は「喜」と「昂」になる。

しかしながら、Twitter などのマイクロブログではユーザが自由にツイートすることができ、表記ゆれが多く発生する為、感情語辞書と正しくマッチングしない。例えば、感情語辞書に「楽しい」という言葉が感情語として登録されているが、ツイートが「今日は楽しもう」だと形態素に分けた際に「楽しもう」になるためにマッチングしない。このような問題を解決するために Juman には表記揺れを緩和する目的で代表表記という基

本語彙に付与されている ID のようなものが存在する。そこで、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為、感情語辞書に記載されている語の中で代表表記を有していて且つ、代表表記が未登録なものを辞書に追加した。また、形態素に同形が存在する場合は曖昧な単語なので考慮しないものとする。

次にツイートに否定語が入っている場合に正しい感情値が抽出することができないという問題が発生する。例えば、「嬉しくない」というような語の場合だと形態素解析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「嬉しい」という形容詞は感情語辞書では「喜」の為、この文の感情は「喜」になってしまう。しかしながら、この「嬉しくない」は「嬉しい」を否定している為、「喜」の感情にはならない。この問題を解決する為に、熊本ら [5] は Juman の出力結果を変換する事により否定語についてのルールを提案している。本論文ではこの熊本らの提案している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はその感情を持たないとする事で問題を解決する。このように求めたツイートの文の感情の例を表 2 に示す。表 2 のツイート「野球はすごく面白い」の場合、各感情軸の感情値の合計が閾値の 0.4 以上となる感情軸は「喜」と「昂」なので、このツイートの感情は「喜」と「昂」となる。また、表 2 のツイート「最近すごく面白くない」の場合、各感情軸の感情値の合計が閾値の 0.4 以上となる感情軸が存在しないためこのツイートは感情なしとなる。

4. 感情の関係

3.2 節で示した既存の 10 次元の感情軸を Twitter に用いた時の問題点を洗い出す為にユーザ実験を行い、その結果から、感情の関係を分析する。具体的には、共起分析とクラスタ分析の 2 つの手法を用いた感情軸決定手法に基づいて適切な感情軸を決定する。

4.1 ユーザ実験

ユーザ実験手順

被験者は 10 代から 50 代までの各年代男性 10 人、女性 10 人からなる 100 人である。実験データは任意の 100 ツイートを用い、被験者はこの 100 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。実験で用いたデータの一部を表 3 に示す。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ツイートの各々の感情候補の感情値とする。そして、感情値がある閾値以上のものをそのツイートの感情とする。また、このときすべての感情値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。本実験では、実験で用いた全ツイートの感情の感情値の平均値である 48 を閾値とした。本実験の結果得られた感情値の一部を表 4 に示す。

ユーザ実験結果と考察

実験で得られた被験者が感じた感情と 3.2 節で提案したシステムが判定した感情を比較し適合率を求めた。また、3.2 節で提

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

表 4 実験の結果の一部

ツイート	喜	哀	怒	怖	好	厭	昂	驚	安	恥
今親元と離れたばっかだから心配ごと多い	18	243	15	88	30	31	13	27	31	19
めっちゃむかむかする	0	36	205	41	35	65	22	13	7	0
赤点取ったら怒るからね!	13	50	127	76	29	33	38	8	4	5
学校から見る景色は好き	141	3	21	29	110	13	5	0	40	2

表 5 感情軸別の実験結果

	喜	哀	怒	怖	好	厭	昂	驚	安	恥
適合率	38%	27%	16%	15%	20%	15%	15%	29%	19%	4%
再現率	53%	29%	27%	21%	27%	15%	22%	50%	26%	8%
F 値	44%	28%	20%	17%	23%	15%	18%	37%	22%	5%

表 6 喜をベースにした共起頻度数と共起確率

	喜	哀	怒	怖	好	厭	昂	驚	安	恥
$ X \cap Y $	-	2	0	1	24	3	4	3	11	4
$ X $	38	34	15	18	27	24	4	14	11	8
共起頻度	-	0.06	0	0.06	0.89	0.13	1	0.21	1	0.5

案したシステムにおいて感情値が 0.3 以上のものをそのツイートの感情とした。その結果の適合率、再現率、F 値を表 5 に示す。結果より、システムの判定と被験者の判定には感情軸によって大きな違いがあった。また、感情軸によって他の感情に影響を与えている軸が存在することがわかった。例えば、「喜」や「哀」の感情軸はツイートに単独で存在することもあるが、「安」は単独では存在せず他の感情と必ず同時に存在する。そこで、ツイートの感情抽出の為の感情軸が必要であると考えた。また、被験者が選んだ感情が 1 つのツイートにおいて複数存在している場合がある。例えば、「めっちゃむかむかする」というツイートの場合、被験者の判定は、「怒」、「怖」、「哀」であった。このような場合我々は一つのツイートにおいて感情が共起しているといい、この共起している感情を本論文では共起感情と呼ぶ。例えば先の例において「怒」と「怖」と「哀」は共起感情であることがわかる。その為、既存の 10 軸の感情より次元削減を行いツイートの感情軸を決定する手法を提案する。具体的には共起関係からグラフを用いて次元削減をする方法と、ワード法による階層型クラスタリングを用いて次元削減をする方法の 2 つの方法で次元削減を行い Twitter にとって最適な感情軸を決定する。

4.2 共起分析

高岡ら [6] は 10 次元の感情ベクトルを用いて、名言から感情の自動抽出を行った。ここでは、10 次元の感情は名言には適切でなかったため次元を結合し感情ベクトルの次元数を削減している。そこで、高岡らの提案した次元を結合する手法を用いて共起関係からツイートの感情軸を決定することを行う。

4.2.1 感情の共起確率

感情の関係を分析するにあたり、共起感情は類似している感情である可能性が高いと考え、4.1 節の実験を用いてそれぞれのツイート数を求める。ここで比較元となる感情を X 、この X と比較する感情を Y とする。次に比較元となる感情 X と比較する感情 Y が共起しているツイート数を求める。このツイートの数を共起頻度数 $|X \cap Y|$ とする。一例として「喜」をベースにした場合の各々の結果を表 6 に示す。例えば、表 6 の場合、「喜」と「好」が共起しているツイートが 100 ツイート中 24 ツイートあった事を示す。よって、「好」と「喜」の共起頻度数 $|X \cap Y|$ は 24 となる。このように、共起頻度数を用いて、

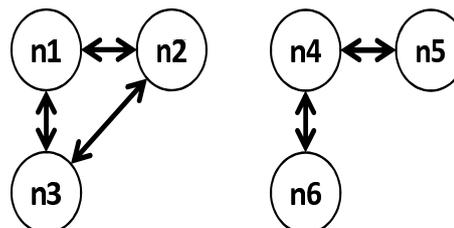


図 1 次元結合の例

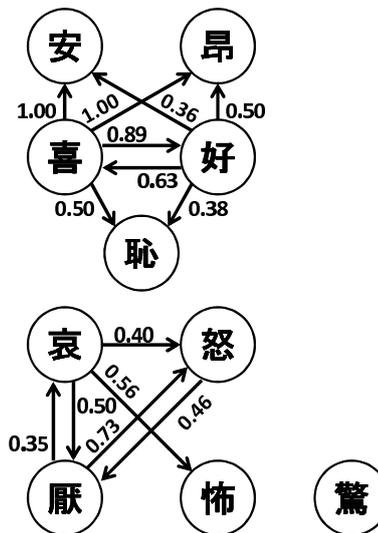


図 2 感情グラフ

各感情がツイート 100 件の中に出現する確率 C_{XY} を以下の式で求める。

$$C_{XY} = \frac{|X \cap Y|}{|X|} \quad (1)$$

4.1 節の実験データを用いて 10 軸すべての感情をベースとして共起頻度数と共起確率を求める。

4.2.2 感情の結合手法

感情の関係を分析するために、感情をノードとして共起確率が閾値 β 以上の感情を接続してその値をラベルとする、ラベル付き有向グラフを作成する。このラベル付き有向グラフを感情グラフと呼び、この感情グラフを用いて感情の関係を分析する。まず、2 つのノード間を比較し、双方向に共起確率が閾値 β 以上の場合、共起関係として接続する。例えば、図 1 の左図

表7 感情の結合結果

	第1結合	第2結合	第3結合
$\beta=0.35\sim 0.30$	喜+好	厭+怒	
$\beta=0.25\sim 0.15$	喜+好	厭+怒	哀+怖
$\beta=0.10$	喜+好+恥+安	哀+怖+厭+怒	

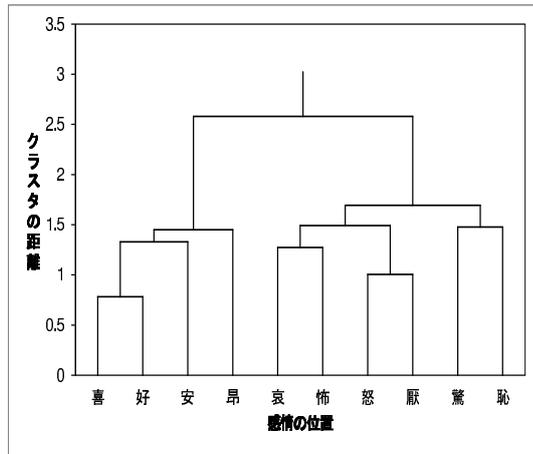


図3 クラスタ分析によって得られたデンドログラム

のノード n1 と n2 の場合、n1 → n2 の共起確率と、n2 → n1 の共起確率がそれぞれ閾値以上の為、共起関係として接続している。また、すべてのノードの組み合わせに対して同様の作業を行う。次に、このように作成した感情グラフが双方向に接続されており、且つ閉路の場合に感情を結合する。具体的には、図1の左図の場合、n1 と n2、n1 と n3、n2 と n3 は互いに共起確率が閾値以上の為、感情グラフにおいて双方向に接続されている。また、n1、n2、n3 が閉路を形成している。このような場合に n1、n2、n3 が双方向に接続されており、且つ閉路になっているので、n1、n2、n3 を結合する。図1の右図の n4、n5、n6 の場合、n5 と n6 が双方に接続されていない。つまりは、n4、n5、n6 が閉路を形成していないため、結合しない。また、閉路が複数存在する場合、共起確率が低いノードを持つ部分グラフは、そのノード間の連結が弱いと考える。つまり、そのノードの感情の関係が弱いと考える。そこで、共起確率が高いノードを持つ閉路を構成しているノードを結合する。

4.2.3 共起分析の結果

共起頻度の閾値 β を 0.35 として 10 軸の感情グラフを作成した。その結果を図2に示す。図2の感情グラフを用いて感情の関係を分析する。まず、「喜」と「好」の感情のペアを例に感情の結合の流れを説明する。「喜」を対象とすると、「好」との共起確率が 0.89 となり、閾値以上なので、共起関係「喜」→「好」として接続する。逆に、「好」を対象とした場合、「喜」との共起確率が 0.63 となり、閾値以上のため、共起関係として接続する。よって、「喜」と「好」は双方向に接続し、このような場合に共起感情として結合する。同様の処理をすべてのノードに対して行った結果、「喜」と「好」の感情のペアと「厭」と「怒」の感情のペアが結合する。この感情のペアが結合したのは、ツイートという短い文において複雑な感情を読み解くことが困難な為と考えられる。例えば、「今日は楽しかった」というツイ

トの場合、ツイートが短すぎるため複雑な感情が表現できずポジティブな感情である「喜」や「好」が共起したと考えられる。また、「安」と「昂」の感情は「喜」の感情と片方向のみ強い接続があった。具体的には、「喜」→「安」と「喜」→「昂」は共起確率が 1.0 となったが、「安」→「喜」と「昂」→「喜」は共起確率が 0.35 以下となった。これは安らいでいるツイートは必ず喜んでいるが、喜んでいるツイートは必ずしも安らいでいるわけではないという結果である。このことから「安」と「昂」の感情は単独で存在することがなく、他の感情を補助する役割があるのではないかと考えられる。また、共起頻度の閾値 β が 0.35~0.10 までの感情の結合結果を表7に示す。表7の結果から、 β の値を小さくすると多くの感情が結合している。また、 β の値を 0.10 にすると大きく分けてポジティブとネガティブに感情が2分割されている。しかし、 $\beta=0.10$ の第1結合はポジティブを表す感情が結合しているが、感情「恥」が第1結合に含まれている。これは閾値を下げすぎたため不要な感情まで結合してしまったのではないかと考えられる。

4.3 クラスタ分析

感情に対してクラスタリングを行う事により、類似している感情が同じクラスタに分類できると考え、4.1節の実験結果を元にクラスタ分析を行う。その際、階層型クラスタリングを行う。本研究ではウォード法を用いる。ウォード法によるクラスタ分析によって得られたデンドログラムを図3に示す。クラスタ分析によって得られた図3のデンドログラムのクラスタの距離に基づいて2, 3, 6, 8つにクラスタを分類し、検証を行った。その結果を表8に示し、以下にその考察を示す。

● 2つのクラスタに分けた場合

クラスタの距離2で分割し、2つのクラスタに分けた場合、第1クラスタに「喜、好、安、昂」、第2クラスタに「哀、怖、怒、厭、驚、恥」の感情が分類された。第1クラスタはポジティブな印象を受ける感情が分類され、第2クラスタはネガティブな印象を受ける感情が分類されている。このことから、ツイートの感情分析においてポジティブやネガティブに判別したい場合はこの2クラスタを用いればいいことが分かる。しかし、「驚」や「恥」などの感情はネガティブな感情であると言い切ることができない。これはユーザによってネガティブだと感じる人と感じない人がいる。よって少し強引なクラスタの分割であると考えられる。

● 3つのクラスタに分けた場合

クラスタの距離1.5で分割し、3つのクラスタに分けた場合は、第1クラスタに「喜、好、安、昂」、第2クラスタに「哀、怖、怒、厭」、第3クラスタに「驚、恥」の感情が分類された。2つのクラスタに分けた場合はポジティブとネガティブに感情を分類していたが、3つのクラスタに分けた場合は第3クラスタにポジティブでもネガティブでもないニュートラルな印象を受ける感情が分類されている。このように分類された理由に関しては、「驚、恥」は直感的にネガティブな印象を受けるが、状況によって使われ方が異なる。例えば「恥」という感情の場合、恥じている時はネガティブな印象を受けるが、恥じらっている場合はネガティブな印象を受けない。よって単純なネガティブ

表 8 クラスタリングの結果

2つのクラスタに分けた場合	
第1クラスタ	喜+好+安+昂
第2クラスタ	哀+怖+怒+厭+驚+恥
3つのクラスタに分けた場合	
第1クラスタ	喜+好+安+昂
第2クラスタ	哀+怖+怒+厭
第3クラスタ	驚+恥
6つのクラスタに分けた場合	
第1クラスタ	喜+好+安
第2クラスタ	昂
第3クラスタ	哀+怖
第4クラスタ	怒+厭
第5クラスタ	驚
第6クラスタ	恥
8つのクラスタに分けた場合	
第1クラスタ	喜+好
第2クラスタ	安
第3クラスタ	昂
第4クラスタ	哀
第5クラスタ	怖
第6クラスタ	怒+厭
第7クラスタ	驚
第8クラスタ	恥

なクラスタから別のクラスタに分類されたのではないかと考えられる。

- 6つのクラスタに分けた場合

クラスタの距離 1.4 で分割し、6つのクラスタに分けた場合は、第1クラスタに「喜、好、安」、第2クラスタに「昂」、第3クラスタに「哀、怖」、第4クラスタに「怒、厭」、第5クラスタに「驚」、第6クラスタに「恥」が分類された。3つのクラスタに分けた場合と比較すると第1クラスタから「昂」が別のクラスタに分類された。理由としては「驚」や「恥」の感情と同様に、状況によって「昂」の感情も使われ方が異なることが考えられる。例えば、「昂」には楽しい時にツイートが昂ぶっている場合や怒っている時にツイートが昂ぶっている場合が考えられる。よって単純なポジティブなクラスタから独立したクラスタに分類されたのではないかと考えられる。

- 8つのクラスタに分けた場合

クラスタの距離 1.2 で分割し、8つのクラスタに分けた場合、第1クラスタに「喜、好」の感情が分類され、第6クラスタに「怒、厭」の感情が分類された。その他の感情に関しては、他の感情と同じクラスタに分類されることはなかった。この結果は 4.2 節の共起頻度 β の値が 0.35~0.30 の感情の結合と同様の結果となった。

- クラスタ分析のまとめ

クラスタ分析を用いた感情の結合ではクラスタ数をいくつにするかによって結果は大きく異なる。クラスタ数が少ない場合

はポジティブやネガティブなどの抽象的なクラスタに分割された。また、8つのクラスタに分けた場合、4.2 節の共起頻度 β の値が 0.35~0.30 の感情の結合と同様の結果となった為、適切な感情軸であると考えられる。

4.4 共起分析とクラスタ分析の結果の比較と考察

4.2 節の共起分析の結果と 4.3 節のクラスタ分析の結果を比較し考察を行う。共起分析の共起頻度 β の閾値が 0.10 の場合とクラスタリング分析のクラスタ数が 3 つの場合を比較すると、両方の結果ともポジティブ、ネガティブ、ニュートラルの 3 種類に感情軸が分類されている。ネガティブな感情に関しては「哀、怖、厭、怒」の 4 つの感情が同様に結合した。しかし、ポジティブな感情に関しては異なる結果となった。ポジティブな感情は「喜、好、安」に関しては高い結合を示したが、「恥、昂、驚」の感情はあまり強い結合を示さなかった。その理由としては、「恥、昂、驚」の感情は人によってポジティブに感じる人と感じない人にばらつきがあった。従って、共起分析とクラスタ分析のポジティブな感情の結果が異なると考えられる。

また、共起分析の共起頻度 β の閾値 0.35~0.30 の場合とクラスタリング分析のクラスタ数が 8 つの場合を比較すると、この場合は共起分析とクラスタ分析のどちらも同様の結合結果となった。この結果より「喜、好」と「厭、怒」に関して強い結合を示した為、喜と好の感情の組み合わせと厭と怒の感情の組み合わせは結合する必要があると考えられる。よって、既存の 10 軸から喜と好の感情の組み合わせと厭と怒の感情の組み合わせを結合した「喜、好」、「安」、「昂」、「哀」、「怖」、「怒、厭」、「驚」、「恥」の 8 軸を Twitter 用の感情軸として決定する。

5. まとめと今後の課題

本論文では、これまでに我々が構築した 10 軸の感情を用いてツイートの感情を求める手法の問題点を洗い出す為にユーザ実験を行った。その結果、システムの判定と被験者の判定には感情軸によって大きな違いがあった。そこで、ツイートの感情抽出の為に感情軸が必要であることがわかった。また、被験者が選んだ感情が 1 つのツイートにおいて複数存在している場合があることに着目し、既存の 10 軸の感情の共起関係の分析とクラスタ分析の 2 種類の手法から、ツイートの感情軸を決定する手法を提案した。共起分析では、「喜」と「好」の感情のペアと「厭」と「怒」の感情のペアは、共起している傾向があった。また、「安」と「昂」の感情は単独で存在することがなく、他の感情を補助する役割があることがわかった。最後にクラスタ分析と共起分析の結果を比較を行った。その結果、共起分析の結果とクラスタ分析の結果が同じとなった喜と好の感情の組み合わせと厭と怒の感情の組み合わせは結合する必要があると考えられた。今後は、共起している感情軸を結合してそれに対応した感情語辞書を構築し、感情抽出を行いたい。また、共起分析の結果から「安」と「昂」の感情軸は他の感情軸を補助する役割があることがわかったので、その役割を考慮した感情抽出を行いたい。

謝辞

本研究の一部は JSPS 科研費 24500134, 26330347 及び、私

学助成金 (大学間連携研究補助金) の助成によるものです。ここに記して謝意を表します。

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 (DEIM 2014), No.E6-2, 2014.
- [2] 中村明, 感情表現辞典, 東京堂出版, 1993.
- [3] R. Plutchik, “The nature of emotions, American Scientist”, Vol.89, pp.344-355, 2001.
- [4] 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.
- [5] 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”, 信学論, Vol.J94-D, No.3, pp.540-548, 2011.
- [6] K. Kouichi, and A. Nadamoto, “Words-of-Wisdom Search based on Multi-dimensional Sentiment Vector”, International Journal of Business Intelligence and Data Mining(IJBIDM), pp.172-185, 2012.
- [7] 中丸茂, “顔文字が文章の信頼度に及ぼす影響”, 人工知能学会研究会資料 (言語・音声理解と対話処理研究会), 37, pp.173-176, 2003.
- [8] 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.
- [9] 池川知里, 新妻弘崇, 太田学, “顔文字の役割を利用したツイートの感情極性推定”, 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 (DEIM 2014), No.E6-4, 2014.
- [10] 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.