

話題と感情の可視化に基づくフォロワー推薦

山本 湧輝^{1,a)} 熊本 忠彦^{2,b)} 灘本 明代^{1,c)}

概要: Twitter では任意のユーザをフォローすることで、そのフォローされたユーザ（フォロワーと呼ぶ）がツイートした内容を自分のタイムライン（TL）に表示させることができるようになる。そのため、実世界での友人や知人だけでなく、共通の趣味・嗜好を有するユーザや有益な情報をツイートしているユーザ、好きな有名人（芸能人や政治家、スポーツ選手など）をフォローすることで、自分の TL を充実させることが可能である。しかしながら、Twitter 上には数多くのユーザが存在しているため、その中からユーザがフォローしたくなるようなユーザ（すなわちフォロワー候補）を探し出すのは面倒な作業と言える。この問題を解決するために、最適なフォロワー候補の推薦について多くの先行研究がある。これら先行研究では、ツイートの話題が似ているユーザをフォロワー候補として推薦するケースが多いが、ツイートの話題が類似しているからといって、その話題に対するユーザの感情が異なる人をフォロワー候補として推薦するのは適さない場合も多数存在する。そこで我々は、ユーザの興味のある話題が類似しているかどうかだけでなく、その話題に対してどのような感情を持っているかも考慮してフォロワー候補を推薦するシステムを提案する。このとき、任意の話題に対する感情を表現するために、我々が先行研究で提案した Twitter 向けの 8 次元の感情軸（「喜・好」、「安」、「昂」、「哀」、「怖」、「怒・厭」、「驚」、「恥」）を用いることとし、ツイートからの感情をこの 8 軸からなるレーダーチャートを用いて可視化する。

1. はじめに

近年、Twitter の普及により、人々は Twitter を利用して気軽にそしてリアルタイムに自分の得た情報や思ったことや感じたことを発信できるようになっている。そのため、Twitter 上には数多くのユーザが発信した膨大な量の情報が存在している。Twitter では任意のユーザをフォローすることで、そのフォローされたユーザ（フォロワーと呼ぶ）がツイートした内容を自分のタイムライン（TL）に表示させることができるようになっており、実世界での友人や知人だけでなく、共通の趣味・嗜好を有するユーザや有益な情報をツイートしているユーザ、好きな有名人（芸能人や政治家、スポーツ選手など）をフォローすることで、自分の TL を充実させることが可能である。しかしながら、2015 年 8 月現在、世界には 3 億 200 万人の Twitter 月間アクティブユーザがいる [1] と言われており、Twitter 上には数多くのユーザが存在している。この中からユーザがフォローしたくなるようなユーザ（すなわちフォロワー候補

を探し出すのは困難な作業であり、ユーザにとって実質的かつ潜在的な不利益となっている。

このような問題を解決するために、任意のユーザに対し最適なフォロワー候補を推薦することができるシステムが数多く提案されている。例えば、Pennacchiotti ら [2] は、協調フィルタリング手法を用いてツイートの話題からユーザ間の類似度を計算し、フォロワー候補を推薦するシステムを提案している。一方、Gurini ら [3] は、クラスタリング手法を用いてツイートから話題を抽出し、共通あるいは類似の話題があるかどうかに基づいてフォロワー候補を推薦するシステムを提案している。このように多くの先行研究では、ツイートの話題が似ているユーザをフォロワー候補として推薦しており、その結果、プロ野球の阪神タイガースが好きなユーザに対しては阪神タイガースについて多くツイートしている人を、ジブリ映画が好きなユーザに対してはジブリ映画について多くツイートしている人をフォロワー候補として推薦することが可能となっている。

しかしながら、阪神タイガースについて多くツイートしていても、アンチ阪神ファンで、阪神の悪口ばかりツイートしているユーザもいれば、阪神ファンであっても、阪神のことを野次ってばかりのユーザもいる。同様に、ジブリ映画について多くツイートしていても、感動的で泣けるジブリ映画が好きなユーザもいれば、楽しく陽気なジブリ映

¹ 甲南大学
Konan University

² 千葉工業大学
Chiba Institute of Technology

a) wabisabiwasabin@gmail.com

b) kumamoto@net.it-chiba.ac.jp

c) nadamoto@konan-u.ac.jp

画が好きユーザもいる。このように、ツイートの話題が類似しているからといって、それぞれのユーザのその話題に対する感情を見てみると、必ずしも類似しているとは言えず、フォロワー候補として適していない場合もある。

そこで我々は、ユーザの興味の対象（すなわちツイートの話題）が類似しているだけでなく、その話題に対してユーザがどのような感情を持っているかも考慮してフォロワー候補を推薦するシステムを提案する。

一方、多くの先行研究でインターネットからの感情分析や感情抽出が行われているが、そのほとんどが Positive/Negative といった 2 クラス（あるいは Neutral を加えた 3 クラス）の感情を対象としている [4],[5],[6],[7],[8],[9]。Positive と Negative に単純化された感情は、レビューに基づく商品の推薦/非推薦の決定や賛成意見/反対意見の抽出といった 2 クラス（あるいは 3 クラス）に分類可能な問題には適しているが、より複雑な人の感情を表現したツイートを対象とする場合には十分とは言えない。例えば、上述の阪神タイガースとジブリ映画の例で示したように、ユーザの怒りや悲しみ、喜びといった感情を Positive/Negative だけでは扱えない。そこで本研究では、ツイートの感情を抽出する際に多次元の感情軸を用いることにする。具体的には、国語学者の中村が提唱している 10 次元の感情軸 [11] を我々が行った被験者実験の結果に基づいて整理し直すことで、ツイートの感情を表現するのに適した 8 次元の感情軸（「喜・好」、「安」、「昂」、「哀」、「怖」、「怒・厭」、「驚」、「恥」）[10] を提案している。本研究では、この 8 次元の感情軸を用いて、対象ユーザとフォロワー候補となりうるユーザのツイートから話題とそれぞれの話題に対する感情を抽出し、話題と感情の両面において類似しているユーザをフォロワー候補として推薦するシステムを提案する。

なお、任意のユーザに対し最も適したフォロワー候補を抽出するためには、Twitter 上のすべてのユーザのツイートから話題と感情を抽出する必要があるが、Twitter 上には非常に多くのユーザが存在しているため現実的ではない。そこで本論文では、本研究のはじめの一歩として、任意のユーザに対し関係の近いユーザを推薦の範囲とすることにする。すなわち、推薦の対象となるユーザ（本論文では「対象ユーザ」と呼ぶ）のフォロワーがフォローしているユーザ（すなわちフォロワーのフォロワーであり、本論文では「 F -ユーザ」と呼ぶ）を推薦の範囲として、この中からフォロワー候補を決定することにする。

提案システムは、対象ユーザの興味のある話題とそれらの話題に対する感情の両方が類似している F -ユーザをフォロワー候補として推薦する。具体的にはまず、対象ユーザのツイートとすべての F -ユーザのツイートを Twitter Rest API^{*1} を用いて取得する。次に、それぞれの F -ユーザごと

に、対象ユーザと各 F -ユーザのツイートをクラスタリング手法を用いて分類し、各クラスに分類された両者のツイートの分布から両者に共通の話題を抽出する。さらに、それぞれの共通の話題に対し、対象ユーザとその F -ユーザのツイートの感情値（感情の強さを定量化したもの）を算出する。以上の結果から、共通の話題が多く、それぞれの共通の話題に対する感情値が近い F -ユーザをフォロワー候補として推薦する。このとき、推薦されたフォロワー候補がどのようなユーザかを直感的に判断できるように、抽出された共通の話題毎に対象ユーザとフォロワー候補の感情値（8次元の感情軸に対する 8個の感情値）をレーダーチャートを用いて表し、この 2つのレーダーチャートを重ね合わせて提示する。

以下、2章では関連研究について述べ、提案システムの特徴を明らかにする。3章では共通の話題の抽出方法について述べ、4章では各共通の話題に対する感情値算出方法について述べる。5章では話題と感情の可視化について述べ、6章では評価実験について述べる。最後に7章では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

Twitter ユーザの増加に伴いフォロワーの推薦に関する研究が多数行われている。北村ら [12] はリプライ等のユーザ間関与に基づく 2-hop ユーザ推薦モデルを提案している。Scott ら [13] はフォロー関係から友人を推薦する手法を提案している。Jilin ら [14] は、ソーシャルネットワークサイトでツイート抽出した興味とフォロー関係からフォロワーを推薦するシステムを提案している。Hannon ら [15] は Twitter でフォローやリプライなどから関連性の高いユーザを見つける Twittomender システムを提案しています。Armentano ら [16] はフォローフォロワー関係からフォロワーの推薦を行っている。これらの研究は、フォロー関係のみを考慮しており、ツイートの話題と感情を考慮している本研究とは異なるため、我々の提案システムと相補的に利用することも可能である。また、ツイートの話題のみを考慮したフォロワー候補推薦の研究も数多く行われている。久米ら [17] は、ツイートから特徴語を抽出し、その特徴語から興味を持っているカテゴリを推定するとともに、カテゴリ（話題）の特徴語をツイートしているユーザを検索することでフォロワー候補を決定している。熊本ら [18] は任意の Twitter ユーザのツイートの話題と感情を可視化するシステムを提案しているが、話題の抽出に Yahoo! キーフレーズ抽出 API [19] を用いているため、話題がフレーズ単位で抽出され、二人のユーザに共通の話題を抽出するのは難しい。そのため、フォロワー候補の推薦は行っていない。本研究では、二人のユーザのツイートをクラスタリング手法を用いて分類し、各クラスに分類された両者のツイートの分布から両者に共通の話題の抽出を可能にして

*1 <https://dev.twitter.com/rest/public>

いる。

一方、感情分析・感情推定に関する研究が多数行われている中で、感情を表現するための多次元の感情モデルも提案されている。代表的な感情モデルとして、Plutchikのモデル [20] がある。このモデルは人間の感情を「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している。熊本ら [21] 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 つの感情軸を提案している。すなわち、「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 つであり、各感情軸は反義語関係にある 2 種類の印象語で構成されている。Takaoka ら [22] は、中村の提唱した 10 次元の感情軸 [11] の次元削減を行い、名言の感情をモデル化した 6 次元の感情軸を提案している。本研究では、中村の 10 次元の感情軸 [11] を我々が行った被験者実験の結果に基づいて整理し直すことで、ツイートの感情を表現するのに適した 8 次元の感情軸（「喜・好」、「安」、「昂」、「哀」、「怖」、「怒・厭」、「驚」、「恥」）[10] を用いている点が異なっている。

3. 共通の話題の抽出

提案システムは、対象ユーザと各 冨ユーザの共通の話題を二人のツイートから抽出するために、以下の処理を行う。

まず、Twitter Rest API を用いて対象ユーザと 冨ユーザのツイートを n 個ずつ収集し、計 $2n$ 個のツイートを取得する。本論文では n 個のツイートを新着した 200 個のツイートとし、合計 400 個のツイートを取得する。次に、取得したツイートから共通の話題を抽出するために、ツイートのクラスタリングを行う。このクラスタリングには単文のクラスタリングに向いている [24] と報告のあった Repeated Bisection 法 [25] を用いる*2。クラスタリングの結果、それぞれのクラスタには複数のトピック（話題を示す名詞）が含まれているが、本研究ではクラスタの中心ベクトルに最も近いトピックをそのクラスタの話題と定義する。今回用いたクラスタリングツール bayon では分割クラスタ数を指定することができるが、我々の予備実験の結果から分割数は 10 クラスタとした。これら分割されたクラスタから共通の話題を抽出するために、我々は以下の 2 つのポイントに着目する。

- それぞれのクラスタにおける対象ユーザと 冨ユーザのツイート数の比率あるクラスタ内に対象ユーザと 冨ユーザのツイートが混在している場合、そのクラスタの話題は共通の話題となる可能性があるが、その比率が 10:1 のように偏っている場合と 1:1 のように均等な場合を比べてみれば、均等な場合の方が共通の話題と

してより適切であると考えられる。そこで、 i 番目のクラスタにおける両者のツイート数の比率 R_i を以下の式を用いて求め、その値が閾値 T_R 以下のクラスタの話題を両者に共通の話題とする。

$$R_i = \frac{|X_i - Y_i|}{X_i + Y_i}$$

ここで、 X_i は i 番目のクラスタに属するユーザ X のツイートの数を示し、 Y_i は i 番目のクラスタに属するユーザ Y のツイートの数を示す。

- ツイートの凝集性

Repeated Bisection 法はハードクラスタリングであるため、クラスタリングの対象となったツイートは必ずいずれかのクラスタに分類される。そのため、単発的な話題（相互に関連性のない話題）を含むツイートが「その他」のクラスタともいべきクラスタ（本論文ではガベージクラスタと呼ぶ）に集められる傾向がある。このようなガベージクラスタの話題は、たとえ二人のユーザのツイート数が同じであっても、共通の話題とは言いがたいため、事前に削除する必要がある。そこで、ガベージクラスタ内のツイートの話題に関連性が乏しいことを利用してガベージクラスタを選別し、除外することにする。具体的には、 i 番目のクラスタ C_i のセントロイド c_i とそのクラスタに含まれるツイート x のコサイン類似度をツイートごとに求め、その平方和をクラスタ C_i の凝集性 A_i と定義し、その値が閾値 T_A 未満のクラスタをガベージクラスタとして除外することにする。

$$A_i = \sum_{x \in C_i} \left(\frac{x \cdot c_i}{|x||c_i|} \right)^2$$

以上より、本研究では、両者のツイートの比率 R_i が閾値 T_R 以下であり、かつツイートの凝集性 A_i が閾値 T_A 以上であるクラスタを「共通話題クラスタ」と呼び、共通話題クラスタの中心ベクトルに最も近いトピックを共通の話題として扱うこととする。なお、それぞれの閾値は我々の予備実験により、 $T_R = 0.25$ 、 $T_A = 0.60$ とする。

4. 共通の話題に対する感情値算出

提案システムは、抽出された共通の話題毎に対象ユーザと 冨ユーザの感情ベクトル（8 次元の感情軸に対する 8 個の感情値からなるベクトル）を算出し、その類似度を求める。具体的には、共通話題クラスタ内に含まれるそれぞれのツイートから我々が構築した感情語辞書を用いて 8 次元の感情軸に対する 8 個の感情値（8 次元のベクトル）を求め、この感情ベクトルをそれぞれのユーザ毎に足し合わせた結果得られる感情ベクトルを、その共通の話題に対するそれぞれのユーザの感情ベクトルとする。そして、このようにして得られる 2 つの感情ベクトルのコサイン類似度

*2 実際には Repeated Bisection 法が実装されたクラスタリングツールである bayon [26] を用いる

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
喜・好	楽しみ	0.91	大切だ	0.54	恋しい	0.64
安	安心	0.45	安らぎ	0.81	休息	0.82
昂	昂る	0.45	すごい	0.57	感動	0.72
哀	悲しい	0.45	憂鬱	0.82	痛い	0.62
怖	怖い	0.45	恐怖	0.64	不安	0.55
怒・厭	嫌だ	0.82	怒る	0.87	辛い	0.72
驚	驚く	0.71	驚愕	0.54	慌てる	0.52
恥	赤面	0.72	恥じ	0.87	照れる	0.56

を、対象ユーザと 冚ユーザのその共通の話題に対する感情の類似度とする。

4.1 感情語辞書

ツイートから感情値を算出する際に、ツイートを構成する単語毎に感情の強さを定量化した感情語辞書が必要となる。本研究では、感情語辞書を構築するための手法として、熊本ら [21] 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、感情語辞書を構築するために、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、その結果に基づいてそれぞれの単語の感情値を数値化している。

しかしながら、熊本らが構築した感情語辞書では、新聞記事の印象を表現するのに適した 3 つの感情軸（「楽しい⇔悲しい」、「うれしい⇔怒り」、「のどか⇔緊迫」）が用いられており、Twitter 上のツイートに対する感情を表現するのに適しているとは言えない。その理由は主に 2 つあり、1 点目として、Twitter 上のツイートは新聞記事よりユーザの日常と深く関係しており、様々な感情を有している場合が多いという点が挙げられる。一方、我々は、これまでの先行研究 [10] において、国語学者の中村が提案している 10 次元の感情軸 [11] を我々が行った被験者実験の結果に基づいて整理し直すことで、ツイートの感情を表現するのに適した 8 次元の感情軸（「喜・好」、「安」、「昂」、「哀」、「怖」、「怒・厭」、「驚」、「恥」）を提案している。そこで本研究では、この 8 つの感情軸を用いて感情語辞書を構築することにする。

また、理由の 2 点目として、Twitter 上のツイートには口語的な表現が多く、さらに表記上の揺れや文法の乱れなどから単語間の共起関係を正しく分析できない場合が多いという点が挙げられる。

そこで本研究では、ツイートと同様に口語的な表現が多いが、表記上の揺れや文法の乱れが少ない Yahoo!映画^{*3} のレビューデータ 74,000 文書から熊本らの感情語辞書構築システムを用いて感情語辞書の構築を行った。その結果、それぞれの感情軸に対して約 5,600 語～7,500 語の感情語

表 2 ツイートの感情値の算出例
ツイート「野球はすごく面白い」

感情語	喜・好	安	昂	哀	怖	怒・厭	驚	恥
凄い	0.14	0	0	0	0	0.01	0.7	0.38
面白い	0.89	0	0	0	0.01	0	0.03	0.32
合計	1.03	0	0	0	0.01	0.01	0.73	0.7

ツイート「最近すごく面白くない」

感情語	喜・好	安	昂	哀	怖	怒・厭	驚	恥
凄い	0.14	0	0	0	0	0.01	0.7	0.38
面白くない	0	0	0	0	0	0	0	0
合計	0.14	0	0	0	0	0.01	0.7	0.38

と対応する感情値を感情語辞書に登録することができた。今回構築した感情語辞書の一部を表 1 に示す。

4.2 ツイートの感情抽出

8 次元の感情軸を用いてツイートの感情を決定するには、4.1 節で構築した感情語辞書を用いてツイートの感情値を求める。具体的には、ツイートに対して形態素解析エンジン Juman^{*4} を用いて形態素解析し、ツイート中の形態素と感情語辞書中の感情語のマッチングを行う。しかしながら、表記上の揺れがある場合には感情語とのマッチングを正しく行うことができない。例えば、感情語辞書に「楽しい」という感情語が登録されていても、「今日は楽しかった」というツイートからは「楽しかった」という形態素が抽出され、「楽しい」とはマッチングしない。そこで我々は、ツイート中の形態素をこの代表表記に統一し、感情語辞書中の感情語とのマッチングを行う。このとき、感情語辞書に記載されている感情語の中には、代表表記を有しているが、代表表記が未登録なものがある。このような単語は手で感情語辞書に追加することとし、実際に 32,326 単語を登録した。一方、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。

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算出することができない。例えば、「嬉しくない」の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

^{*3} Yahoo!movie <http://movies.yahoo.co.jp/>

^{*4} <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

表 3 類似度算出の例

	AS_i	BS_i	S_i
i=阪神	(20,3,12,3,4,3,3,5)	(20,4,11,4,3,3,5,3)	0.99
i=巨人	(6,3,3,10,4,20,3,3)	(3,5,5,12,3,18,5,3)	0.97
	S (合計)		1.96

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」を否定しているので、「喜」の感情にはならない。この問題を解決する為に、熊本ら [23] は Juman の出力結果を変換することで、否定語を正しく扱うため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して問題を解決する。このようにして求めたツイートの感情の例を表 2 に示す。

4.3 感情に基づく類似度計算

次にツイートの感情値からユーザどうしの類似度を計算する。まず、算出したツイートの感情値を用いて、そのツイートが含まれる共通話題クラスターの感情値を求める。4 節で決定した共通話題クラスターに対して以下の式を用いて、対象ユーザと 冨ユーザの類似度 S を求める。

$$S = \sum_{i=1}^k \left(\frac{AS_i \cdot BS_i}{|AS_i| |BS_i|} \right)^2$$

ここで、二人のユーザの共通話題クラスターと判断されたクラスター数を k とする。 AS_i は i 番目の共通話題クラスターにおける対象ユーザの感情ベクトルを示し、 BS_i は i 番目のクラスターにおける 冨ユーザの感情ベクトルを示す。この 2 つの感情ベクトルの類似度の計算にはコサイン類似度を用い、共通の話題と判断されたクラスター全てのコサイン類似度の平方和を計算することで、類似度 S を求める。この類似度 S が閾値 T_S 以上のときに、その 冨ユーザをフォロー候補として推薦する。ある対象ユーザと 冨ユーザの阪神と巨人各々の話題に対する類似度計算の例を表 3 に示す。ここでの感情値は(喜・好, 安, 昂, 哀, 怖, 怒・厭, 驚, 恥)を示す。表 3 の結果より、対象ユーザと 冨ユーザは阪神という話題に対して「喜・好」のような好意的な感情を持ち、巨人には「怒・厭」のような反感を抱いている事がわかり、その結果、この対象ユーザと 冨ユーザは阪神という話題の感情及び巨人という話題の感情において、高い類似度を示す。

5. トピックと感情の可視化

本研究では対象ユーザのフォロー支援を目的にフォロイを推薦している。しかし、実際にどういった理由で推

薦されているのかが分からないと実際にフォローするのは困難である。そこで、対象ユーザが直感的に推薦された新しいフォロイの候補の話題と感情がわかるようにレーダーチャートを用いて可視化する。レーダーチャートを用いて 2 ユーザのデータを重ねて表示することで話題に対する感情の違いを直感的に比較することが可能となる。レーダーチャートは共通の話題毎に作成する。図 1 にレーダーチャートの例を示す。レーダーチャートのタイトルには共通の話題を示す。また、レーダーチャートの軸は 8 軸の感情を用いる。この時、対角の軸は感情軸の関係が対極になるように配置する。対極の感情を決定するにあたり、Plutchik の感情の輪 [20] を参考に対極の感情軸を「喜・好 ⇄ 怒・厭」、「安 ⇄ 恥」、「怖 ⇄ 哀」、「驚 ⇄ 昂」とする。次に、そのクラスターの感情軸毎の感情値をデータとしてレーダーチャートを生成する。図 1 では、表 3 に示した、ある対象ユーザとある 冨ユーザとの感情値を阪神と巨人の話題について各々レーダーチャートで示している。青色が対象ユーザであり、赤色が 冨ユーザである。この結果から、対象ユーザと 冨ユーザの話題に対する感情が類似していることが直感的に把握できる事が分かる。レーダーチャートで可視化することでツイートを見ることなく直感的に容易にその推薦された 冨ユーザをフォローするか否かの指標になる。

6. 評価実験

提案手法の有用性を示すために 2 つの実験を行った。1 つ目の実験では提案手法におけるクラスタリングの有用性を示す。2 つ目の実験では我々の提案する感情を考慮したフォロイ推薦手法の有用性を示す。また、2 つの実験とも、対象とする対象ユーザは 5 名でそれぞれ別の趣味を持った対象ユーザである。それぞれ対象ユーザの 冨ユーザはランダムで各々 100 名を抽出し実験を行った。各々の対象ユーザの特徴は以下の通りである。

User1 「阪神タイガース」と「読売ジャイアンツ」に対して興味を持っている対象ユーザで、「阪神タイガース」に対しては「喜・好」が一番多く、「昂」もある程度多くチームを応援するようなツイートを多く投稿している。また、「読売ジャイアンツ」に対しては「怒・厭」が一番多く、「哀」もある程度多く批判的な内容のツイートを多く投稿している。

User2 「政治」に対して興味を持っている対象ユーザで、「維新の党」に対して「怒・厭」「哀」等の感情が多く、批判的なツイートを多く投稿している。

User3 「読売ジャイアンツ」に対して興味を持っている対象ユーザで、「読売ジャイアンツ」に対して「喜・好」「昂」のツイートを多く投稿している。

User4 「アニメ」に対して興味を持っている対象ユーザで、「アイドルマスター(アニメのタイトル)」のファンで

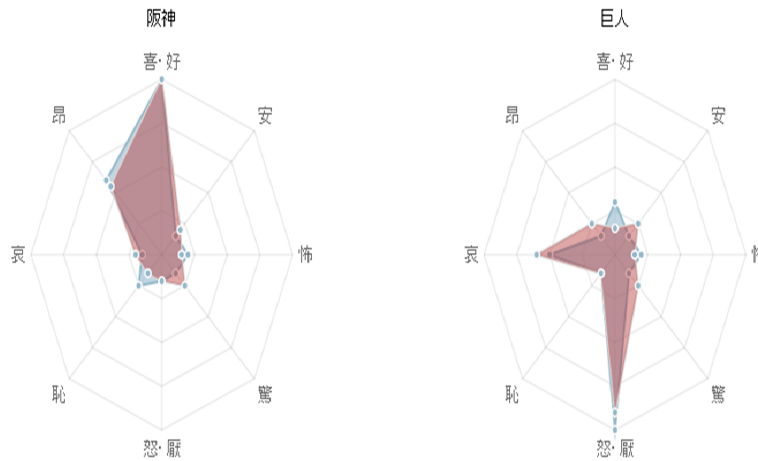


図 1 レーダーチャートの例

あり、「喜・好」「驚」のツイートを多数している。User5「サッカー」に対して興味を持っている対象ユーザーで、「浦和レッズ」に対して「喜・好」の感情が多くチームを応援するようなツイートを多く投稿している。

表 4 実験 1 の精度 (適合率)

	ベースライン手法	提案手法
User1	0.25	0.72
User2	0.20	0.79
User3	0.19	0.55
User4	0.20	0.45
User5	0.19	0.47

6.1 実験 1 : クラスタリングの有用性

実験条件

我々の提案する共通話題クラスタの決定手法の有用性を示す為に比較実験を行った。ここでは、bayon によってクラスタリングされたそのままの結果をベースラインとし、話題の比率と凝集性を用いた我々の提案手法と適合率による比較実験を行い、提案手法の有用性を示す。正解データは人手で対象ユーザーと 冨ユーザーのツイートを見比べて共通の話題を選び、その共通の話題とクラスタの話題を比較して正しいと感じたものを選んだ。

結果と考察実験結果を表 4 に示す。表 4 の全ての対象ユーザーにおいてベースライン手法より提案手法の精度は格段に上がっている。精度が上がった理由として 2 つ考えられる。1 つめの理由として、2 ユーザーのツイートを混ぜた状態でクラスタリングしているため、ベースラインでは、2 ユーザー共通の話題を表すクラスタではなく片方のツイートのみでクラスタを形成するケースが多く見られ、そのようなクラスタは、共通の話題ではなく片方のユーザーの話題となってしまう不適切な結果となっていた。これに対し、我々の提案手法はクラスタ内のツイートの比率を考慮しているため、これらユーザーの偏ったクラスタが削除されることにより精

度が上がったと考えられる。2 つめの理由として、bayon の結果のみのベースライン bayon がハードクラスタリングであるため必ずどこかのクラスタにツイートが含まれてしまう。反面、我々の提案手法ではクラスタ内に含まれる凝集性を考慮しているため、これらのクラスタに含まれる話題が削除され、精度が上がったと考えられる。この 2 つの理由から、ベースライン手法より提案手法の精度は格段に上がったと考えられる。例えば、User1 において bayon の結果では候補ユーザーのみがツイートしている話題である「テニス」や「ゲーム」など対象ユーザーがツイートしていない話題のクラスタが多数形成されている。それに対して、提案手法ではこれらのクラスタは削除され、共通の話題である「阪神」や「読売」などのクラスタが多く形成された。

6.2 実験 2 : 感情を考慮したフォロワー推薦手法の有用性

実験条件

提案手法の共通の話題と感情を考慮したフォロワー推薦と、話題のみのフォロワー推薦との比較実験を行い、提案手法の有用性を示す実験を行った。ベースラインは、話題のみの類似度からフォロワーを推薦している手法を用いた。正解データは人手により対象ユーザーと 冨ユーザーのツイートを見比べてフォロワーとして適切と感じたユーザーを選んだ。

結果と考察

実験結果を表 5 に示す。

User1 の場合は適合率・再現率・F 値の全ての値でベースラインを上回った。ベースラインでは、阪神と巨人についてツイートしている 冨ユーザーがフォロワー候補として推薦対象となっており、中には、阪神ファンのみならず巨人ファンと思われる 冨ユーザーもいた。我々の手法では、「阪神」に対して「喜・好」や「昂」といった感情をツイートしていて、「巨人」に対して「厭・怒」といった感情をツイートしているユーザーが推薦され、巨人ファンと思われ

表 5 実験 2 の精度

手法	適合率	再現率	F 値	
User1	ベースライン手法	0.70	0.54	0.61
	提案手法	0.90	0.77	0.83
User2	ベースライン手法	0.81	0.65	0.72
	提案手法	0.87	0.65	0.74
User3	ベースライン手法	0.76	0.60	0.68
	提案手法	0.77	0.73	0.75
User4	ベースライン手法	0.86	0.80	0.83
	提案手法	0.92	0.73	0.81
User5	ベースライン手法	0.82	0.60	0.70
	提案手法	0.67	0.52	0.59

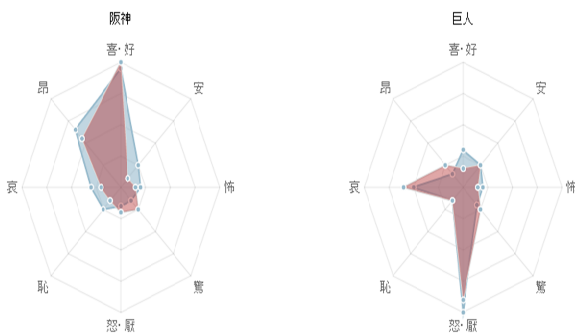


図 2 User1 と類似度が高い 冨ユーザのレーダーチャート

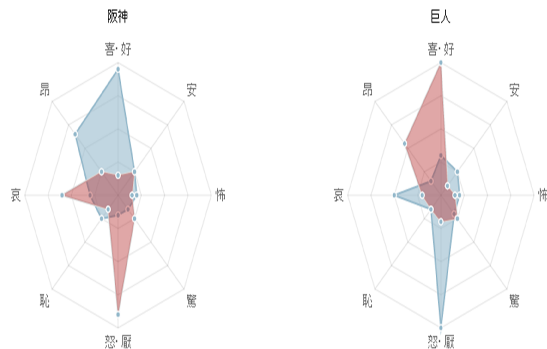


図 3 User1 と類似度が低い 冨ユーザのレーダーチャート

る 冨ユーザや、阪神を野次っている 冨ユーザをフォロワー推薦対象から排除することができた。その理由の一つとして、対象ユーザと 冨ユーザが「阪神」と「巨人」に対して感情的なツイートを多くしており、感情抽出が成功していると共に、被験者からもこれらツイートの感情が分かりやすかった為と考えられる。

User2 に関しては、適合率、再現率共にベースラインと提案手法との差が見られなかった。これは、対象ユーザである User2 は政治やニュースについてのツイートが多いため、あまり感情的なツイートが多くされていないことで精度に差が見られなかったと考えられる。

User3 は、ベースラインと提案手法を比較すると提案手法の方が良い結果となっているが、同じ野球を共通の話題としている User1 の結果と比較すると、それほどベースラインと提案手法の結果の差が大きくない。その理由として、

User3 とその 冨ユーザは巨人の選手に対してのツイートを多く投稿しており、その為、選手毎に評価が分散してしまい User1 と比較してベースラインと提案手法の差が小さくなっていると思われる。

User4 に関しては、適合率、再現率共にベースラインと提案手法との差は見られなかった。その理由として、対象ユーザとその 冨ユーザはアニメのタイトルであるアイドルマスター内に出て来る特定のキャラクターに対するツイートを多く投稿しており、その為、キャラクター毎に評価が分散してしまいベースラインと提案手法との差が見られなかったと考えられる。

User5 は、ベースラインと提案手法を比較すると提案手法がベースラインを下回る結果となった。その理由として、User5 があまり感情的なツイートをしていなかった為と考えられる。User5 と同じように感情的なツイートをあまりしていない User2 は政治やニュースについてのツイートであるため 冨ユーザも同様に感情的なツイートをあまりしていない傾向がある。しかし、User5 の 冨ユーザはサッカーや浦和レッズに対して感情的なツイートを多くしている。その為、推薦されるべき 冨ユーザと User5 の感情がかけ離れてしまいそれに伴い類似度が低くなり、提案手法の精度が良くない結果となった。

また今回の実験結果の中で、User1 に対して、最も類似度が高く推薦フォロワー推薦となった 冨ユーザのレーダーチャートを図 2 に示す。この図 2 から対象ユーザと同様に阪神に対して「喜・好」と「昂」の感情を強く持っていて且つ、巨人に対して「怒・厭」の感情を強く抱いていることが分かる。実際にツイートを見ることなくレーダーチャートを見るだけで、直感的に自分と同じような話題とその話題に対して持っている感情が理解できる。反対に、今回の実験結果の中で図 2 と同様に阪神と巨人についてツイートしているが類似度が低く話題は同じであるが感情が異なるため、推薦フォロワー候補となっていないユーザを図 3 に示す。図 3 では対象ユーザが阪神に対して「喜・好」「昂」という感情が強く、「怒・厭」「哀」という感情が弱いが、逆に 冨ユーザは「怒・厭」「哀」という感情が弱く「喜・好」「昂」という感情が強い。また、巨人に対しても同様に感情の差が見られ、その為、この 冨ユーザは User1 のフォロワー推薦候補として適さないことがわかる。このように、実際にツイートを見ることなくレーダーチャートを見るだけで、話題に対して違う感情であることが容易に分かる。

7. まとめと今後の課題

本論文では、話題とその話題に対しての感情を考慮したフォロワー推薦手法を提案した。具体的には、(1) クラスタリングを用いて共通の話題の抽出手法を提案した。具体的には、クラスタに含まれているツイートの比率と凝集性を考慮したクラスタの選別を行う手法を提案した。(2) 共

通な話題トピックの感情値算出手法を提案した。感情値の算出には我々が構築した感情語辞書を用いた。(3) 共通の話題とその話題の感情値から類似度を算出し推薦ユーザを決定する手法を提案した。さらに、評価実験を行い、感情を考慮したフォロワー推薦の有用性を示した。

今後の課題としては、今回は実際の対象ユーザを被験者として、レーダーチャートによるユーザ実験を行う予定である。さらに、対象ユーザが実際にフォローしているユーザからフォロー傾向を分析して、その結果をフォロワー推薦手法に組み込む予定である。

謝辞 本研究の一部はJSPS 科研費 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。

参考文献

- [1] Twitter Reports First Quarter 2015 Results, <https://investor.twitterinc.com/releasedetail.cfm?ReleaseID=909177>.
- [2] M. Pennacchiotti and S. Gurumurthy, "Investigating Topic Models for Social Media User Recommendation", Proceedings of the 20th International Conference Companion on World Wide Web, pp.101-102, 2011.
- [3] D. F. Gurini, F. Gaspiretti, A. Micarelli and G. Sansonetti, "A Sentiment-Based Approach to Twitter User Recommendation", Proceedings of the 5th ACM RecSys Workshop on Recommender Systems, 2013.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] M. Kobayashi, T. Inui and K. Inui, "Dictionary-based acquisition of the lexical knowledge for p/n analysis", SIG-SLUD, pp.45-50, 2001.
- [9] S. Fujimura, M. Toyoda, and M. Kitsuregawa, "A reputation extracting method considering structure of sentence", Institute of Electronics, Information and Communication Engineers, Data Engineering Workshop, 2005.
- [10] 山本湧輝, 熊本忠彦, 灘本明代, "ツイートの感情の関係に基づく Twitter 感情軸の決定", 第7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 2015), No.E5-2, 2015.
- [11] 中村明, 感情表現辞典, 東京堂出版, 1993.
- [12] 北村太一, 小川祐樹, 諏訪博彦, 太田敏澄, "コミュニケーションに着目した Twitter フォロワーユーザ推薦", 人工知能学会全国大会論文集 (CD-ROM), Vol. 26, 2012.
- [13] SA. Golder, S. Yardi, A. Marwick and D. Body, "A structural approach to contact recommendations in online social networks", Proceedings of Workshop on Search in Social Media at ACM SIGIR Conference on Information Retrieval, 2009.
- [14] C. Jilin, G. Werner, D. Casey, M. Michael and G. Ido, "Make New Friends, but Keep the Old: Recommending People on Social Networking Sites", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.201-210, 2009.
- [15] J. Hannon, K. McCarthy and B. Smyth, "Finding useful users on twitter: twittomender the followee recommender", In Advances in Information Retrieval 33rd European Conference on IR Research, pp.18-21, 2011.
- [16] MG. Armentano, DL. Godoy, and AA. Amandi, "A topology-based approach for followees recommendation in twitter", In: The 9th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, ITWP, p.22, 2011.
- [17] 久米雄介, 打矢隆弘, 内匠逸, "興味領域を考慮した Twitter ユーザ推薦手法の提案と評価", 情報処理学会研究報告, pp.1-8, 2015.
- [18] 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.
- [19] テキスト解析: キーフレーズ抽出 API. <http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html>
- [20] R. Plutchik, "The nature of emotions", American Scientist, Vol.89, pp.344-355, 2001.
- [21] T. Kumamoto, "Design of Impression Scales for Assessing Impressions of News Articles", LNCS6193, Springer, In International Workshop on Social Networks and Social Media Mining on the Web (SNSMW'10), pp.285-295, 2010.
- [22] K. Takaoka 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.
- [23] 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.
- [24] 花井俊介, 灘本明代, "酷似レシビ抽出のためのクラスターリング手法の提案", 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 (DEIM2014), No.E5-2, 2014.
- [25] Y. Zhao and G. Karypis, "Comparison of Agglomerative and Partitional Document Clustering Algorithms", University of Minnesota, pp.2-14, 2002.
- [26] M. Fujisawa, "Bayon - a simple and fast clustering tool - Google Project Hosting", 2012, <https://code.google.com/p/bayon/wiki/Tutorial.en> [Online; accessed 11-August-2015].