

SVMを用いた感情抽出に基づくフォロワー推薦

秋山 和寛[†] 熊本 忠彦^{††} 灘本 明代[†]

[†] 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

^{††} 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

E-mail: [†]s1371002@s.konan-u.ac.jp, ^{††}kumamoto@net.it-chiba.ac.jp, ^{†††}nadamoto@konan-u.ac.jp

あらまし Twitterにおいてユーザがどのような人物をフォローするのは様々な判断基準があると考えられる。しかしながら、膨大なアカウントの中から自分がフォローしたいと思うユーザを探し出すのは困難である。その為、様々なフォロワー推薦の研究が行われている。フォロワー推薦は大きく分けて、人のネットワークに基づくフォロワー推薦手法とコンテンツの類似性に基づくフォロワー推薦手法がある。これまで我々はコンテンツの類似性に基づくフォロワー推薦に加え、感情も類似しているフォロワー推薦手法を提案してきた。これまでの研究では感情語辞書に基づいたツイートの感情値を決定してきたが、その精度が十分であるとは言い難い。そこで本論文では新たに機械学習によるツイートの感情値の決定を行い、コンテンツが類似しているだけでなくそのコンテンツに対する感情も考慮したフォロワー推薦手法を提案する。

キーワード 感情抽出, Twitter (ツイッター), SNS, SVM

1. はじめに

マイクロブログを代表する Twitter の機能の一つにフォロー機能がある。ユーザがどのような人物をフォローするのには、現実での友人や知人、共通の趣味を持つユーザ、情報収集、好きな有名人など様々な判断基準があると思われる。しかしながら、現在世界では 3 億 1000 万人^(注1)、日本国内だけでも 4000 万人^(注2) を超える Twitter 月間アクティブユーザが存在しており、この膨大な数のユーザの中から自分がフォローしたいと思うユーザを探し出すのは困難であるといえる。そこでこの問題を解決するために様々なフォロワー推薦の研究が行われている。例えば、Jilin ら [1] は、ソーシャルネットワークサイトでツイートから抽出した興味とフォロー関係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。桑原ら [2] は投稿者のメッセージから共通の話題に対して同じような感情を持つ投稿者の推薦を行っている。フォロワー推薦は大きく分けると、ユーザ同士のフォロワー・フォロワー関係を考慮したネットワークに基づく推薦手法とユーザ同士のツイートした話題の類似性を考慮したコンテンツの類似性に基づく推薦手法の 2 つに分類される。後者において、あるユーザ A が”オリンピック楽しみ!”、別のユーザ B が”オリンピックなんて金の無駄”とつぶやいている場合、どちらもオリンピックの話題についてつぶやいているのでユーザ A にユーザ B が推薦される。しかしながら、オリンピックに肯定的なユーザ A に対して、否定的なユーザ B を推薦するのは適切であるとは言い難い。これは話題の類似性のみに着目しており、個々のユーザが話題に対しどのような感情を抱いているかは考慮していないためと考えられる。

これまで我々はユーザ同士のツイートの話題の類似性だけで

なく感情の類似性も考慮したフォロワー候補の推薦手法を提案している [3]。これまでの提案手法では、ツイートの感情抽出に我々の作成した感情語辞書を用いている。具体的には、我々の提案する Twitter 感情抽出用の 8 次元の感情軸 [4] からなる感情語辞書を用いて、ツイート文中の単語ごとの感情値の和によりツイートの感情値を決定している。そして、その感情値を用いてユーザ同士の感情の類似度を比較し、推薦対象のフォロワーを決定している。しかしながら、実験より感情値計算の精度が十分であるとは言い難い事が判明した。

そこで、本論文では感情の分類にサポートベクターマシン (SVM) を用い、分類した各ツイートの感情からユーザ同士の話題ごとの感情の類似度を求め、フォロワー候補ユーザの推薦を行う。

図 1. と以下に、SVM を用いたツイートの感情抽出に基づくフォロワー推薦の手順を示す。尚、本論文では、フォロワーを推薦されるユーザを「閲覧ユーザ」と呼び、フォロワー推薦の対象となるユーザは閲覧ユーザのフォロワーのフォロワーかつ閲覧ユーザのフォロワーでないユーザを対象とする。フォロワーのフォロワーを本論文では「ff ユーザ」と呼ぶ。

(1) 閲覧ユーザと ff ユーザのツイートの収集。

過去 200 件のツイートを対象とし、閲覧ユーザと ff ユーザのツイートを収集する。

(2) 閲覧ユーザと ff ユーザの話題の類似性の抽出

閲覧ユーザと各推薦候補ユーザのツイートを併せてクラスタリングを行う。そして、各クラスタの両者のツイート数の比率と凝集性を考慮して両者に共通の話題を抽出する。

(3) 類似した話題の感情の比較

抽出した各話題に対する両者の感情を求め、それぞれの話題に属するツイートを機械学習により、感情ごとに分類する。本研究では機械学習に SVM を用いる。その後、各クラスタ内の両者の感情の割合から両者の感情の類似度を求め、類似度が

(注1) : <https://about.twitter.com/ja/company>

(注2) : <https://twitter.com/TwitterJP/status/793649186935742465>

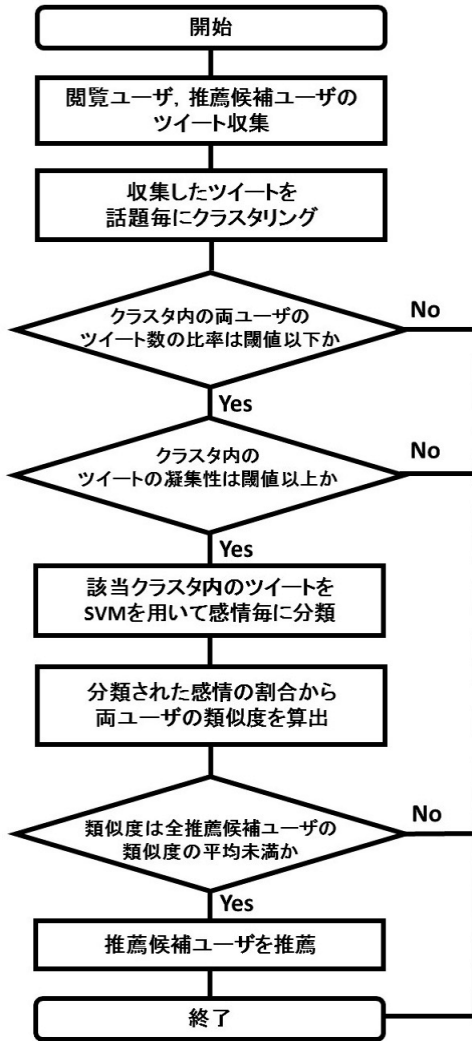


図1 フロー図

全推薦候補ユーザの類似度の平均未満である推薦候補ユーザを決定する。

以上の結果より、共通の話題が多く、それぞれの話題に対する感情が似ている推薦候補ユーザをフォロワー候補として推薦する。

以下、2.章では関連研究について述べ、提案手法と従来手法の違いとその立ち位置を示す。3.章ではこれまでの提案手法である話題の類似性の抽出手法について述べ、4.章ではSVMによる感情抽出を用いた提案手法について述べ、5.章では実験により提案手法の有用性を測る。最後に6.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をはじめとしたソーシャルネットワーキングサービス(SNS)の普及に伴い、フォロワー推薦に関する研究が数多く行われている。また、SVMを用いてテキストから感情を抽出する研究も行われている。フォロワー推薦手法は大きく分けて、ネットワークに基づく推薦手法とコンテンツの類似性に基づく推薦手法の2つに分類される。以下で2種類の手法及びSVMによる感情抽出について説明する。

2.1 ネットワークに基づく推薦手法

ネットワークに基づく推薦手法ではユーザ同士のフォロワー・フォロワー関係に基づいてフォロワー候補が推薦される。北村ら[5]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[6]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。これらの研究ではフォロワー・フォロワー関係を考慮してフォロワー候補を推薦しているため、趣味や興味についてツイートしていないユーザでも、フォロワー・フォロワーの関係から興味のあるユーザを推薦することが可能となっている。しかし、閲覧ユーザとフォロワー・フォロワーのフォロー傾向が異なる場合、興味のないユーザが推薦されてしまうということもありえる。これに対し本研究では閲覧ユーザの投稿した話題と類似しているユーザを推薦しているため、興味のあるユーザを推薦することができる。

2.2 コンテンツの類似性に基づく推薦手法

大村ら[7]は、ツイートやプロフィールの内容から被推薦ユーザが興味を示している語を抽出し、その語についてつぶやいたユーザを推薦候補としている。黒柳ら[8]は共通の著名人をフォローしているということは、その著名人のファン、若しくは何かしらの興味、関心がある可能性が高いと考え、著名人を共通にフォローしているユーザをフォロー候補として提示している。これらの研究では閲覧ユーザと似た興味、関心を持つユーザをフォロワー候補として推薦しているが、興味、関心の対象に対しての感情まで同じであるとは限らないため不適切なユーザを推薦する可能性がある。これに対し、本研究では話題に対する感情の類似性も考慮している点が異なる。

2.3 SVMによる感情抽出手法

高次元の特徴量を効率よく扱うことができるSVMは、特徴量が多くなる傾向にあるテキストの分類に適していると言える。江村ら[9]は、感情語を含まないテキストから感情を抽出するための手法として文末に存在する顔文字が感情を表すという仮定に基づき、SVMを用いて学習、感情モデルを構築し、書き手の感情を抽出する手法を提案している。江村らの研究は学習に使用しているのが文ではなく顔文字という点が本研究と異なる。岡村ら[10]は、ソーシャルブックマークサービスのタグを用いて教師データを収集し、SVMで印象の分類機を構築し、それを用いてWebページのURLが含まれるツイートの印象を推定しているが、分類を娯楽目的のポジティブ、娯楽目的でないポジティブ、ネガティブの三種類としている点が本研究とは異なる。

3. 話題の類似性の抽出手法

3.1 話題の類似性

話題の類似性の決定手法はこれまで我々が提案してきた手法を用いる。本章ではこれまで我々の提案した話題の類似性の決定手法の概要を述べる。

3.2 共通の話題の抽出

話題の類似性の決定手法では、初めに閲覧ユーザと推薦候補ユーザに対する共通の話題を抽出する。処理の流れは以下の通りである。まず閲覧ユーザと推薦候補ユーザのツイートを200

件ずつ収集し、計 400 件のツイートを取得する。次に、共通の話題を抽出するために、取得したツイートに対しクラスタリングを行う。この際、クラスタリング手法には短文のクラスタリングに優れているといわれる [11] Repeated Bisection 法 [12] を用いる。このとき、ツイートの名詞のみを対象としてクラスタリングを行う。また、クラスタリングツールには Repeated Bisection 法が実装されたツールである bayon [13] を用いる。クラスタリングを行うと、それぞれのクラスタには話題を示す名詞が複数存在しうるが、本研究ではクラスタの中心ベクトルに最も近い名詞をそのクラスタの話題とする。なお、クラスタの分割数については我々の予備実験の結果より 10 クラスタとする。また、共通の話題を抽出するに際し、以下の 2 つの点を考慮する。

- 話題の非共通度

あるクラスタ内に閲覧ユーザと推薦候補ユーザのツイートが混在している場合、そのクラスタは両者にとっての共通の話題となる可能性がある。しかしながら、一口に混在していると言っても、どちらかのツイートがそのクラスタ内の大半を占めているような場合もあれば、そうでない場合も考えられる。この場合、共通の話題として相応しいのは後者であると考えられるため、以下の式を用い、 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 を求め、その値が閾値 T_R 以下であるクラスタの話題を両者にとって共通の話題とする。

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

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 のツイート数を示す。

- ツイートの凝集性

クラスタリングに用いた Repeated Bisection 法はハードクラスタリングであるため、相互に関連性のない話題を含むツイートで構成されたクラスタが生じる場合がある。本研究ではこのようなクラスタをガベージクラスタと呼ぶ。ガベージクラスタの話題については、たとえ両者のツイート数が同じだとしても共通の話題とは言い難いため、取り除く必要がある。この際、ガベージクラスタ内のツイート同士の話題の関連性が低いという点を利用する。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_i 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取り除く。

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

以上より、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名詞を共通の話題として扱うことにする。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。

表 1 ツイートの感情値の算出例
ツイート「梅雨はジメジメして嫌い」

感情語	喜・好	安	昂	哀	怖	怒・厭	驚	恥
梅雨	0.14	0	0.248	0	0	0.09	0	0
ジメジメ	0.276	0	0	0	0	1.29	0	0
嫌い	0	0	0	0	0	1	0	0
合計	0.416	0	0.248	0	0	2.38	0	0

4. SVM を用いた感情の類似性の抽出手法

本研究では、話題毎に感情が類似している n ユーザをフォロイーとして推薦する。つまりは、共通話題クラスタを構成するツイートの感情が類似している n ユーザがフォロイーの推薦対象となる。これまでの我々の提案手法では、ツイートの感情抽出に我々の作成した感情語辞書を用いて、ツイート文中の単語ごとの感情値の和によりユーザ同士の感情の類似度の比較を行っていた。ツイートの感情値の算出例を表 1 に示す。しかしながら、実験より感情値の算出精度が十分であるとは言い難い。そこで、本論文では感情の分類に機械学習を用い、分類した各ツイートの感情からユーザ同士の話題ごとの類似度を求め、フォロイー候補ユーザの推薦を行う。機械学習には SVM のライブラリである LIBSVM [14] を用い、カーネルは RBF カーネルを用いる。尚、SVM による感情の分類の際には感情軸は中村明の感情語辞書 [15] に基づいた 10 軸の感情軸「喜」、「好」、「哀」、「怒」、「厭」、「安」、「恥」、「怖」、「昂」、「驚」を対象とし、類似度算出の際には我々が提案している Twitter 用の 8 軸の感情軸「喜・好」、「哀」、「怒・厭」、「安」、「恥」、「怖」、「昂」、「驚」を対象とする。

4.1 SVM を用いた感情の分類

抽出された共通の話題毎に共通話題クラスタ内の各ツイートの感情を SVM を用いて分類する。分類された感情を比較し、両ユーザのその話題に対する感情の類似度を求める。教師データとは以下の通りである。

教師データ

SVM の教師データには、以下我々の行ったアンケートにより感情毎に分類されたツイートをを用いる。具体的には、1 セットにつき、無作為に抽出した 25 件のツイートに対し、10 人の被験者による 11 軸 5 段階評価を 58 セット、計 1450 件のツイートに対し 10 人の被験者による評価を行った。被験者はツイートを見て、10 軸の感情軸に、感情を持たない軸を表す「その他」を加えた 11 の軸それぞれにどの程度当てはまると感じたかを 5 段階で評価した。取得した評価データの内、被験者毎に各軸の評価値を求め、最大、最小の被験者のデータを削除し 8 名の実験結果を教師データとして用いる。ツイートの感情を決定するにあたり、各ツイートに対し以下の処理を行う。

- (1) 評価値の二値化

評価データ内で評価値が 3 以下のものは 0 に置き換え、4 以上のものは 1 に置き換える。

- (2) ツイートの感情の決定

各感情軸毎に 8 名の被験者の評価値の和を求め、ある軸の和が

他の軸の和に対し2倍以上大きければ、その軸をそのツイートの感情とする。ただし、我々のこれまでの研究[16]から、「安」と「昂」の感情軸は単独で存在することがなく、他の感情軸を補助する役割があることが判明している。このことから、「安」「昂」の2軸に関しては、他の軸よりも評価値の和が大きければ該当ツイートの感情とする。

素性

素性は、全教師データに形態素解析を行い、得られた名詞・動詞・形容詞全てを対象とし、各々の単語がツイート内に含まれている(1)か否か(0)とする。また、各データのクラスは10軸の感情を1-10の数字に置き換えたものを使用する。

教師データとなるツイートと感情分類の例を表2、表3に示す。

性能評価及び考察

SVMの精度を調べるために性能評価を行った。評価には交差検定を用い、分割数は10とし、パラメータは $c = 2.0$ 、 $g = 0.0315$ とする。また、新規にアンケートを行い得られた未知のデータに対する分類も行った。

性能評価の結果を表4に示す。

表4より、交差検定の精度は61.8%と低い値となっているが、これは教師データの量がまだまだ少ないためであると考えられる。未知データの分類の精度が同様に低い値であるのも同様の理由によるものだと考えられる。

4.2 感情の類似に基づく推薦候補の決定

クラスタリングによる共通の話題の抽出及び、SVMによる感情分類の結果、話題が類似し、かつ感情が類似しているユーザを推薦候補として閲覧ユーザに提示する。感情の類似度の算出には共通話題クラスタ内の両者の各感情の割合を用いる。各共通話題クラスタ i における両ユーザの感情の割合に基づく類似度の計算には、2つの確率分布間の距離尺度を測ることができるJSダイバージェンス(Jensen-Shannon divergence)[17]を用いる。

$$JS_i = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$$

$$KL(P||Q) = \sum_{i=1}^8 P(i) \log(2) \frac{P(i)}{Q(i)}$$

$$M = \frac{1}{2}(P(i) + Q(i))$$

ここで、 $P(i)$ は共通話題クラスタ i 内における閲覧ユーザの8軸の感情それぞれの割合を表し、 $Q(i)$ は推薦候補ユーザの8軸の感情それぞれの割合を表している。上記の式より、両ユーザのある共通話題クラスタ i の感情の割合の類似度 JS_i を求める。次に以下の式を用いて、閲覧ユーザ a とffユーザ b の感情の類似度を求める。

$$DS_{ab} = \sum_{i=1}^k JS_i$$

ここで k は共通話題クラスタの数とする。各共通話題クラスタに対する感情の割合の類似度を求め、その和を両者の感情の割

合の類似度 DS_{ab} とする。この感情の割合の類似度 DS_{ab} が全推薦候補ユーザの類似度の平均未満であるとき、推薦候補ユーザを閲覧ユーザに推薦する。

5. 実験

提案手法の有用性を測るために、評価実験を行った。実験は、7名の被験者に自身のffユーザ200名それぞれのツイートを見てもらい、実際にフォローしたいと被験者が評価したffユーザを正解データとし、適合率、再現率、F値を求める。

7人のユーザの特徴は以下の通りである。

- ユーザ1

ツイート内容は日常生活に関するものやフォロワーへのリプライが多く、フォロワーは現実世界での友人・知人が多い。

- ユーザ2

ツイート内容は日常生活に関するものが多く、フォロワーは現実世界での友人・知人及び情報収集のためのユーザが多い。

- ユーザ3

ツイート内容は趣味のゲームに関するものが多く、フォロワーもそのゲームに関するユーザで占められている。

- ユーザ4

ツイート内容は日常生活に関するものが多く、フォロワーは現実世界での友人・知人が多い。

- ユーザ5

ツイート内容は日常生活に関するものやフォロワーへのリプライが多く、フォロワーは現実世界での友人・知人が多い。

- ユーザ6

ツイート内容は日常生活に関するものが多く、フォロワーは現実世界での友人・知人及び情報収集のためのユーザが多い。

- ユーザ7

ツイート内容は趣味のゲームに関するもので占められており、フォロワーはそれとは別の趣味であるスポーツに関するユーザで占められている。

5.1 結果と考察

実験結果を表5に示すとともに、実験結果に対する考察を行う。表5より、ユーザ1-6に対しては、全般的に適合率が低くなっているが、これは4.1章で述べたSVMの分類精度が低いという理由に加え、現在SVMでの分類時にはどれか1種類の感情に分類を行っているため、複数の感情を併せ持っているツイートや、挨拶などの感情を持たないツイートの分類を上手く行えていないということが考えられる。この問題については、学習データに感情を持たない軸を加える、感情分類の際に各感情の推定確率を求め、各感情の値からそのツイートが単一の感情を持つのか、複数の感情を持つのかを判定するというような方法で解決を図る。推定確率の例を表6に示す。

ユーザ1、5の値が特に低くなっているのは、両ユーザが様々なタイプのユーザと交流しており、ツイートの内容に話題の偏りが少なく、結果として個々のユーザそれぞれとの類似度が低くなっているためだと考えられる。このように、日常的なコミュニケーションをTwitterの主な利用目的としているユーザに対しては共通の話題が抽出しづらく、ひいては感情を算出す

表 2 教師データの例

感情軸	ツイート
喜	一緒に面接した人がここはこうしたほうがいいってアドバイスくれて嬉しい
哀	対抗戦参加はしてるけど戦力になれなくてすごく悲しい
好	最近気がついたんだけど、私ストライプがけっこう好きみたい
厭	自分を大切にできない奴は、大嫌いだよ
安	大阪に帰ってきて一週間たったけど、実家って安心するね
恥	この間買ったサンダルが歩く度にキュッキュウしてる。ちょっと恥ずかしいな。
怒	まじ洒落ならんわ腹立つー。どない騒いどんねんカス
怖	夜一人でふらふらするのやめよ。ちょっと怖い
昂	明日一限体育!! しかもバスケできる!!! 興奮しすぎて寝れないぞ!!
驚	衝撃的すぎる事実を聞かされてどー反応していいかわからへん

表 3 SVM による感情分類の例

ツイート	分類された感情
13 年前の仲間たちが、やがてそれぞれの社会で花ひらいていくその美しさや希望。そんなのを感じていられるだけで、しあわせである。 何か…俺…めっちゃ馬鹿だから後味悪く感じちゃって…やっぱ幸せなエンドを期待してたから…何というか… きっと誰より君を想ってるのは今日も明日も僕だから そして改めて勉強嫌いなあとと思った。よくテスト期間マジメに勉強したな、びっくり。 今日はゆっくり…ゆっくりするんだ…おお心の平穏… ド素人ということを抜いて考えたとしても基礎の基礎からやらかしてて恥ずかしいです… 残業だっつってんだからなんか買っておいてくれればいいのに ピアスはよ開けよ。痛いかな。こわ。あー怖い BBやるのになんでこんなときどきしてんの俺? 芹那ってザワ出身だったっけってふと思っってヤフーで検索してみたら、カールセルまきも北海道出身だって事がわかってビックリ!	喜 哀 好 厭 安 恥 怒 怖 昂 驚

表 4 評価結果

評価手法	精度
10 交差検定	0.618
未知データの分類	0.469

表 5 実験結果

被験者	適合率	再現率	F 値
ユーザ 1	0.072	0.450	0.125
ユーザ 2	0.106	0.667	0.183
ユーザ 3	0.129	0.653	0.215
ユーザ 4	0.102	0.778	0.180
ユーザ 5	0.071	0.435	0.122
ユーザ 6	0.103	0.400	0.164
ユーザ 7	-	-	-

することも難しくなっている。故に、コミュニケーションを重視しているユーザに対しては話題の類似性に基づいた推薦ではなく、ユーザ間のネットワーク関係に基づいた推薦が適していると考えられる。

ユーザ 7 については、本人がつぶやいている内容と推薦候補ユーザのつぶやいている内容が全く異なっており、共通の話題が抽出されなかったため、類似度の算出をすることができなかった。このような閲覧ユーザに対しては、閲覧ユーザがどのような内容のツイートをしているかだけではなく、その閲覧ユーザのフォロワーがどのような傾向のユーザで構成されているのかなどといったことも予め考慮しておく必要があると考え

られる。

推薦が上手く行われた例としては、サッカーが好きな閲覧ユーザに、サッカーに関するツイートをしており、閲覧ユーザが気が合うと感じるユーザを推薦したというものが挙げられる。推薦が上手く行われなかった例としては、声優に興味のない閲覧ユーザに声優のアカウントを推薦したというものが挙げられる。これは、閲覧ユーザのフォロワーに声優が好きで声優のアカウントをフォローしているユーザがいたため、推薦候補ユーザである ff ユーザの中に声優のアカウントが存在しており、“予約”という語が閲覧ユーザと声優のアカウントどちらにも共通して出現していたため、共通話題クラスタが抽出されたからであると考えられる。また、このようなことが起こった理由としては、現在推薦候補ユーザは閲覧ユーザの ff ユーザの中からランダムに取得している。そのため、閲覧ユーザが友人・知人といった理由でフォローしているフォロワーがツイート内容に興味があるからという理由でフォローしているユーザも推薦候補ユーザとなり得るためだと考えられる。この問題に対しては、推薦候補ユーザを取得する際に、閲覧ユーザとフォロワーの関係性及び、フォロワーと ff ユーザの関係性を考慮することで、推薦候補ユーザの中に閲覧ユーザとの関係性が薄いユーザが混ざるのを抑制できるのではないかと考えられる。

以上の結果をまとめると、提案手法では十分な精度を得ることはできなかった。また、他者とのコミュニケーションのために Twitter を利用しているようなユーザに対しては話題や感情の類似性ではなく、ユーザ同士のつながりを考慮した推薦が効

表6 推定確率

クラスラベル	1	2	3	4	5	6	7	8	9	10
分類されたクラス										
4	0.053	0.153	0.174	0.235	0.058	0.025	0.217	0.029	0.095	0.155
10	0.067	0.114	0.114	0.196	0.124	0.015	0.015	0.024	0.103	0.227
3	0.119	0.118	0.247	0.128	0.122	0.022	0.070	0.020	0.101	0.052
5	0.112	0.080	0.157	0.073	0.325	0.007	0.009	0.022	0.163	0.053
5	0.046	0.025	0.065	0.022	0.460	0.020	0.012	0.014	0.048	0.290

果的であるといえる。

6. まとめと今後の課題

本論文では、SVMを用いたフォロワー推薦手法の提案を行い、その有用性を測った。結果として、推薦の精度は良いとは言えなかった。また、複数のユーザとコミュニケーションをとっているようなユーザに対しては特に効果が薄いということが判明した。

今後の課題として、SVMの精度の向上、感情の分類に用いる感情軸に感情を持たない軸である「その他」の追加及び、推薦の際にユーザ毎のTwitterの利用傾向などを考慮していきたい。

謝 辞

本論文の一部はJSPS科研費26330347, 16K07973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ある。ここに記して謹んで感謝の意を表する。

文 献

- [1] 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.
- [2] 桑原雄, 稲垣陽一, 草野奉章, 中島伸介, 張建偉, "マイクロブログを対象としたユーザ特性分析に基づく類似ユーザの発見および推薦方式", 情報処理学会研究報告(データベースシステム研究会), Vol.149, No. 18, pp. 2B-2, 2009.
- [3] 山本湧輝, 熊本忠彦, 灘本明代, "話題と感情の可視化に基づくフォロワー推薦", 第8回Webとデータベースに関するフォーラム論文集(WebDB 2015), pp.174-181, 2015.
- [4] Yuki Yamamoto, Tadahiko Kumamoto, Akiyo Naadmoto, "Role of Emoticons for Multidimensional Sentiment Analysis of Twitter", International Conference on Information Integration and Web-based Applications & Services(iiWAS2014), pp.107-115, 2014.
- [5] 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フォローユーザ推薦", 人工知能学会全国大会論文集(CD-ROM), Vol. 26, 2012.
- [6] Scott A. Golder, Sarita Yardi, Alice Marwick, and Danah Boyd, "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.
- [7] 大村涼, 赤石美奈, 佐藤健, "語彙構造を用いたTwitterユーザ推薦手法の提案", 全国大会講演論文集2013(1), pp.609-611, 2013.
- [8] 黒柳智士, 山田泰宏, 鈴木浩, 服部哲, 速水治夫, "著名人情報に基づいたTwitterフォローユーザ推薦システム", 研究報告グループウェアとネットワークサービス(GN)2013-GN-86(25), pp.1-4, 2013.
- [9] 江村恒一, 安木慎, 宮崎誠也, 久保山哲二, 青木輝勝, 安田浩, "SVMを用いたモブログテキストからの感情抽出", 電子情報通信学会技術研究報告.KBSE, 知能ソフトウェア工学106(473), pp.61-66, 2007.
- [10] 岡村康行, 湯本高行, 新居学, 上浦尚武, "ソーシャルタグを用いたツイートの印象の推定", 研究報告データベースシステム(DBS), pp.1-8, 2015.
- [11] 花井 俊介, 灘本 明代, "酷似レシピ抽出のためのクラスタリング手法の提案", 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), F8-6, 2014.
- [12] Y. Zhao and G. Karypis, "Comparison of Agglomerative and Partitional Document Clustering Algorithms", University of Minnesota, pp.2-14, 2002.
- [13] M. Fujisawa, "Bayon - a simple and fast clustering tool - Google Project Hosting", 2012, <https://code.google.com/p/bayon/wiki/Tutorial/> [Online; accessed 11-August-2015].
- [14] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM - A Library for Support Vector Machines", 2014, <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [15] 中村明, 感情表現辞典, 東京堂出版, 1993.
- [16] 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(DEIM2015), F5-2, 7 pages., 2015.
- [17] Fuglede, B., Topsoe, F., "Jensen-Shannon Divergence and Hilbert space embedding", International Symposium on Information Theory, 2004