

Twitterにおける発話者へのリプライを用いた ユーザ感情推定手法

堀宮ありさ[†] 坂野 遼平[†] 佐藤 晴彦[†] 小山 聡[†] 栗原 正仁[†]
沼澤 政信^{††}

[†] 北海道大学大学院情報科学研究科 〒060-0814 札幌市北区北14条西9丁目
^{††} 小樽商科大学商学部社会情報学科 〒047-0814 小樽市緑3丁目5番21号
E-mail: †{horimiya-a@complex. , r_banno@complex. , haru@complex. , oyama@ ,
kurihara@}ist.hokudai.ac.jp, ††numazawa@res.otaru-uc.ac.jp

あらまし 人間とコンピュータのよりインタラクティブなコミュニケーションを実現する為には、高精度な感情推定技術の確立が必要である。特に近年では、コンテキスト・アウェア・コンピューティングへの応用等、ユーザのリアルタイムな感情を推定することが求められている。本研究では、実時間性が高く感情表現が含まれやすい媒体として、マイクロブログサービスの一つである Twitter [2] に注目する。Twitter は投稿文字数が短く制限されていることから文法的に崩れた文章が多用される傾向にあり、従来の感情語や文末表現、係り受け関係を利用した感情推定技術をそのまま適用することが難しい。そこで本稿では、人間の人間に対する推測能力に着目し、ユーザの発言に対する他者の反応であるリプライを利用して感情推定を行う手法を提案する。

キーワード 感情推定, Twitter, Web マイニング

1. はじめに

1.1 背景と目的

人間とコンピュータのインタラクティブなコミュニケーションの実現を目指す研究が、言語処理・音声処理・画像処理等様々な分野にわたって為されている。言語処理の分野においては、特にテキストを対象とした感情推定の研究が盛んである。人間の持つ感情という曖昧な情報をコンピュータで処理することは難しく、現在に至るまで様々な手法が模索されている。

近年では、コンテキスト・アウェア・コンピューティング [1] の概念が広まりつつあり、感情推定についてもユーザのリアルタイムな感情やその変化を捕捉可能な技術が求められていると言える。例えば、コンピュータがユーザの状況や感情を考慮することができるようになれば、ユーザとコンピュータエージェントとのより自然な対話も実現可能であると考えられる。

また、近年インターネットを用いた情報発信の手段はホームページ、ブログ、SNS と多様化しているが、中でも Twitter などのマイクロブログが爆発的に広まりをみせている。Twitter は一度の投稿が 140 文字以内の短文であることから情報発信の敷居が低く、日常的な言葉によって個人のリアルタイムな感情が反映されやすいという特徴がある。

そこで本研究では、感情推定の対象として Twitter 上の情報に焦点を当てる。Twitter に投稿される文章は 1 文が短く文法的に崩れた文章になることが多く、従来の感情語や文末表現、係り受け関係を利用した感情推定技術をそのまま適用することが難しいという問題がある。そこで我々は、Twitter の特徴である他者との会話機能から、人間の人間に対する推測能力に着

目し、ユーザの発言に対する他者の反応であるリプライを利用して感情推定を行う手法を提案する。

1.2 感情と感情表現の集合関係

発話文から感情推定を行う場合、図 1 に見られるように、全発話の中には感情が含まれる発話の集合 (A) と直接的な感情表現を含む発話の集合 (B) が存在する。直接的な感情表現とは感情語 (感情を表す語) や文末表現を指す。それぞれの発話例を表 1 に示す。

$A \cap B$ の例文では、発話者が感情を持っており、また文章にも「嬉しい」という感情語が含まれている。 $A \cap \bar{B}$ の例文は発言者が感情を持っているが、文章には直接的な感情表現が含まれておらず、この文章のみから感情を推定する事はコンピュータにとっても、そして発話者と面識の無い人間にとっても難しい。これは、同じ単語や表現でも個人によって意味が異なるためである。多くの単語と感情の繋がりは一意には定まらず、一般性が存在していない。しかし、フォロワー^(注1)によるリプライとして例えば「良かったね!」「おお、やったじゃん」「チャンスだね、頑張れ!」等の応答があった場合、これらを見る事で「幸福感」という感情を推定することが出来る。これは、人間が「良かったね」「やったじゃん」などの単語が「喜ばしいこと」に対する反応であると認識することが出来るためである。このように、「喜ばしいこと」という感情を含むツイートに対して、他者の反応に一定の傾向があれば、リプライに表れやすい特徴的な語から感情推定を行う事が可能となる。 $\bar{A} \cap B$ の例文

(注1): あるユーザのフォロワーとは、そのユーザを指定してフォロー (受信登録) しているユーザのことである。

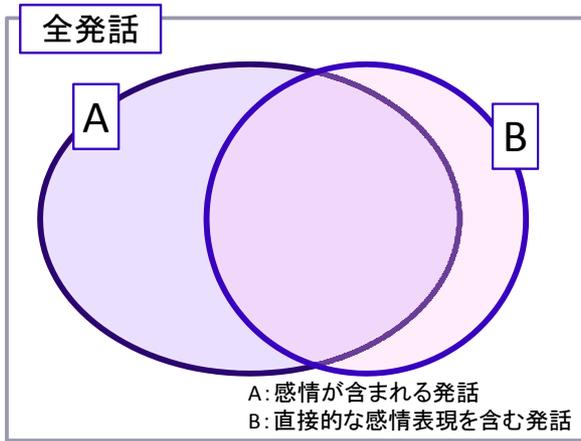


図1 発話における感情と感情表現

表1 発話例

集合	例文
$A \cap B$	今夜はカレーらしい。久しぶりだ、嬉しいなあ
$A \cap \bar{B}$	えええ！今日のバイト店長と二人きりだ！
$\bar{A} \cap B$	『怒り』の葡萄』読み終えた
$\bar{A} \cap \bar{B}$	よし、準備して学校へ行こう

は「怒り」という感情語が含まれているが、特に発言者は特定の感情を持たない。 $\bar{A} \cap \bar{B}$ は感情も感情表現も含まれていない。

従来の手法では、直接的な感情表現から感情を判断していたため主に $A \cap B$ が推定範囲となっており、 $\bar{A} \cap B$ がノイズとして存在していた。しかし、提案手法では発言者の文章ではなく他者の反応である返信から感情推定を行う為、従来の手法では考慮出来なかった $A \cap \bar{B}$ を推定範囲に含む事が出来る。

以降、2章では既存研究を挙げ、3章ではSVMと $tfdif$ を利用した提案手法の詳細を説明する。4章では実データを用いた実験とその結果について報告し、5章で考察した後、6章で総括する。

2. 既存研究

2.1 感情の種類

感情推定とは、ユーザがどのような感情を抱いているのかを判別することであり、特に自然言語文からの感情の自動判定を行うものが多い。例えば、語に好感度(プラス、マイナスの評価)を与え、文毎の事象の望ましさを推定して話者感情を判断する手法[3]を始めとして、感情極性(positive/negative/neutral)の推定手法[4]がいくつか存在している。

更に、感情極性よりも細かい感情推定の研究も行われており、Plutchikの基本感情8種類(怒, 喜, 安, 驚, 恐, 悲, 厭, 昂)[5]に分類するもの、感情表現辞典[6]を元に10種類(喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚)に分類するもの[7]がある。他に日本語語彙大系の一般名詞属性体系中の感情に関する語彙を調べることで22種類の感情を定義したもの[8]があるが、粒度の細かさから精度が充分ではなく、大分類(いくつかの感情をまとめて一つの分類としている)への推定問題に帰着しており、細かい感情推定は現状ではまだ実現していない。

感情極性で3種類にあらかじめ推定した後により詳細に10種類に分類するもの[9]では、多クラス分類問題では上位クラスを分類した後に下位クラスを分類したほうが分類精度が良いことや、先に感情極性に分ける事で致命的な誤回答を避けられる事を示している。

我々人間は、会話において発話内容から快・不快の単純な感情のみならず、喜怒哀楽に代表されるような複雑な感情を理解することが出来る。よって、人とコンピュータのインタラクティブなコミュニケーションを目指すならば、二値の極性のみならず、より細かい粒度で感情を推定することが必要となる。

2.2 推定手段

感情推定の前処理として、文章内の単語に対して感情ラベルを付与するものが多く存在する[7]。文章内に感情語と呼ばれる感情を示す語が含まれている場合には、これらの感情語辞書を用いるアプローチは有効であるが、直接的な感情語が含まれていない文章に対しては適用する事が出来ない。このような問題に対して、文章内の単語に加えて文末表現を考慮したり[3]、文型や文脈を考慮した研究[8]も行われている。また「うれしい」を生起する「遊園地に行く」や「腹立たしい」を生起する「自分だけ処分される」など係り受け関係を利用して感情生起表現を抽出し「感情生起表現辞書」を作成するというものも存在する[7]。

また、コーパスに感情ラベルを付与するものも存在する[9]。コーパスとは自然言語処理に用いるために自然言語の文章を構造化し大規模に集積したものであり、感情ラベルを付与したコーパスは感情コーパスと呼ばれる。これらの手法では感情コーパスを学習データとしてSVMやナイーブベイズ分類器などを使用し、未知データの分類を行う。

しかし、感情コーパスの作成には感情は常に一意に決まるわけではないという問題点がある。感情の判断は、判断する人間の主観に依存する事が考えられ、また周辺状況によっても判断は変化し得るからである。この問題の解決の為に、通常は数名での主観の評価を行い、大多数の意見が一致した感情を正解とする事が多い。しかしながら、コーパスは膨大な量になる為このような人手による感情タグ付けにはコストが掛かり過ぎるという問題がある。これに対し、感情コーパスを自動生成する手法が提案されている[9],[10]。感情コーパスの自動生成では、webテキストからランダムに取りだしたサンプルに対し、明示的な感情表現を利用して自動的に感情タグを獲得する[10]。

以上に述べた既存研究は、いずれも自然言語文を対象としたものであるが、近年では、文体の崩れた文章を対象とした感情推定が注目を集めつつある[12]。このような口語表現、異表記、表記ゆれ、変則的な固有名詞、顔文字、絵文字、未知語などが含まれた文章は不自然言語と呼ばれている。これらの文章は形態素解析器にかけても正しく分割することが出来ないことが多い為、従来は推定に適さないものと見なされていた。しかし、近年では顔文字を正解データとして学習を行うなどの研究も現れている[13]

3. 提案手法

本研究では、短文投稿サービスである Twitter の文章を対象とし、それぞれの発言を Ekman の定義した基本 6 感情「幸福感」「驚き」「恐れ」「悲しみ」「怒り」「嫌悪」に分類して投稿者の感情を推定する。また、Twitter における感情推定の対象者を「発話者」、発話者の発言を「ツイート」、他者から発話者への返信を「リプライ」と呼ぶ。Twitter の文章から感情推定を行おうとすると、一文が短く文法的に崩れているという特徴から、自発的なツイート内容を直接分析するだけでは推定に不十分である。そこで我々は、ツイートの持つ感情ごとにそれに対するリプライにおける言語表現にパターンが存在するという仮説を立て、リプライ文を単語ベクトルと見なして、ツイートの感情推定を行う手法を提案する。提案手法の流れは図 2 となる。

まず、学習データの作成と学習の流れを示す。

- (1) Twitter から感情表現語を元にツイートとリプライを抽出
- (2) 人手によるラベル付け
- (3) ツイートやリプライを素性に分割
- (4) TF-IDF で重みを算出
- (5) SVM で学習

次に、新しく入ってきた文章から感情を推定する流れを示す。

- (1) 新しく入力されたツイート文に対するリプライを抽出
- (2) 素性に分割
- (3) 重みを算出
- (4) 分類器によりツイート文の感情を判定

以下それぞれのフェーズについて詳細を述べる。

3.1 学習データ

本研究では、学習データのラベル付けを人手によって行っている。この手法における感情表現語は、人手による感情ラベル付けの為に補助の役割であり、数人のアノテータによって選ばれた『Twitter においてツイートに感情を込める場合によく用

いられる言葉』を指す。6 感情について、それぞれ 3 つの感情表現語を設定し、各感情表現語毎に 300 ずつを収集する形で、合計 1800 のツイートを集めた。また、それらツイートに対するリプライ 2131 個も同時に収集した。

3.2 使用素性

文章をベクトル化する際、素性として形態素や N-gram^(注2)などが考えられる。今回対象となる文章は口語に近く略語が多用されるなど文法的に崩れている為形態素解析は困難であることから、未知語や言語に依存しない N-gram が適していると考えられる。予備実験により 1-gram, 2-gram, 3-gram 及び形態素を素性とした場合について比較したところ、実際に N-gram の精度が平均的に高く、特に高かった 2-gram を採用した。

3.3 TF-IDF による特徴ベクトルの生成

それぞれの素性に重みを与える為に、TF-IDF を利用する。

TF-IDF(Term Frequency - Inverse Document Frequency) は、文書中の単語に重みを与える手法の一種であり、文書中に出現する特徴的な単語に対して、高い重要度を与える。具体的には、ある特定の文書で多く出現する固有名詞のような単語が重要度が高くなり、逆に一般的な単語の重要度は低くなる。例えば、多くの文書に出現する”です”や”ます”のような頻出単語は重要度が低くなりやすい。

計算式は TF (単語の出現頻度) と IDF (逆文書頻度) の二つの指標に基づいて計算される。これは対象とする文書に頻繁に出現し(網羅性)、コーパス中のほかの文書にあまり出現しない(特定性)ような語を重要とみなすものである。ここでは Harman らによる TF [15] 及び、Jones らによる IDF [16] の式 1 を利用する。

$$TFIDF = TF \cdot IDF \quad (1)$$

$$TF(i, j) = \frac{\log(n_{i,j} + 1)}{\log(\sum_k n_{k,j})} \quad (2)$$

$$IDF(i) = \log \frac{|D|}{|\{d : d \in t_i\}| + 1} \quad (3)$$

本研究においては、一つひとつのリプライを文書とみなしている。重みとして TF-IDF を使用することで、リプライの特徴的な素性を他の素性よりも大きく評価する事が出来る。すなわち、全てのリプライに頻繁に出てきやすい素性の特徴量は小さくなり、そのリプライに特徴的に出現する素性の特徴量は大きくなる。

3.4 学習方法

ツイートとリプライのセットに人手で付けた感情ラベルを教師データとして、6 感情+無感情という 7 種類のカテゴリに分けて分類器を生成する。入力はそのクラスに該当するか否かのラベル付きリプライのセット、出力は 2 値分類器となる。本研究では、高次元の特徴量を効率よく扱う事が出来る分類器であるサポートベクターマシン (SVM: Support Vector Machine) [14] を使用する。SVM は線形しきい素子を用いて、2 クラスのパターン識別機を構成する手法である。

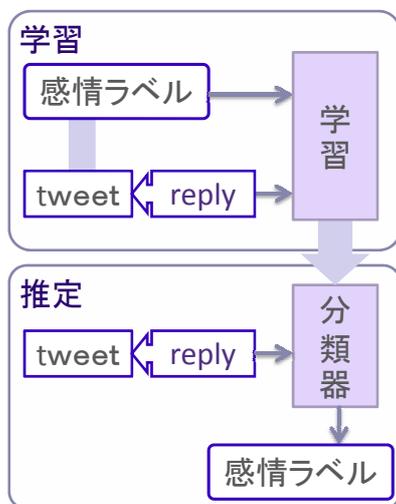


図 2 学習データ作成フロー

(注2): 例えば 2-gram を用いる場合「よかったね!」という文章であれば「よ」「か」「っ」「...」「ね!」という 5 個の素性に分割される

表 2 頻出感情表現語

感情	表現語
幸福感	幸せ, 嬉し, 楽し
驚き	驚いた, びっくり, すごい
恐れ	怖い, こわい, 不安
悲しみ	悲しい, つらい, 虚しい
怒り	腹立つ, 腹が立つ, イライラ
嫌悪	嫌い, キモい, 好きじゃない

表 3 各感情のデータ数

幸福感	驚き	恐れ	悲しみ	怒り	嫌悪	感情無し
274	163	205	245	267	123	595

新しく入ってきたツイートの感情を推定する為には、まずツイート文とリプライをセットで抽出する。次に学習データ作成の場合と同様に N-gram を素性として特徴ベクトルを生成する。その後分類器により各感情カテゴリについて該当するか否かを判定する。

4. 評価実験

提案手法の有用性を確かめるため、計算機実験を行った。

4.1 実験方法

3.1 で述べた感情表現語について、5 人の Twitter ユーザを対象とした事前アンケートの結果をもとに表 2 のように設定し、Twitter REST API を用いて、感情表現語 1 つにつき 100 個ずつリプライ付きのツイートをリプライとセットで取得した。

取得したツイート及びリプライは 2011 年 12 月 28 日から 29 日のものである。1800 個のツイート-リプライペアについて、15 人のアナテータがツイートとリプライを参照しながらラベル付けを行った。ツイート自体が@付きであるもの、つまり連続する会話の一部であるようなものを除外して抽出した。一つの発言に大して 3 人が正解ラベルを付与し、2 人以上が一致したものを採用した。ラベル付けに際しては複数選択を認めており、感情無しも含めた 7 クラスのラベル数は表 3 のようになった。

4.2 実験結果

5-fold cross validation により、推定精度の評価を行った。

評価基準として、再現率 (Recall) と適合率 (Precision)、二つの調和平均を取った F 値、そして正答率 (Cross Validation Accuracy) を用いた。

各感情について、あるリプライがその感情であるか否かを判断する 2 クラス分類における Confusion Matrix は、分類器がその感情だと判定したか (Positive) それ以外の感情と判定したか (Negative) と、判定が正しかったか (True) 間違っていたか (False) かの軸を持つ。4 つの事象の意味は以下の通りとなる。

- TP: 実際の値がその感情であり、判定もその感情とした数
- FP: 実際の値が他の感情であり、判定はその感情とした数
- TN: 実際の値が他の感情であり、判定も他の感情であるとした数
- FN: 実際の値がその感情であり、判定は他の感情であるとした数

ここで、それぞれの式は以下で表される。

$$\text{precision} = \frac{TP}{(TP+FP)} \quad (4)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (5)$$

$$F1 = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})} \quad (6)$$

$$\text{Cross Validation Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (7)$$

表 4 に実験で得られた感情毎の再現率、適合率、F 値、正答率を示す。

表 4 再現率、適合率、F 値、正答率

	precision	recall	F1 value	Cross Valication Accuracy
幸福感	0.57	0.71	0.63	66.97%
驚き	0.59	0.64	0.61	62.58%
恐れ	0.67	0.59	0.63	60.24%
悲しみ	0.56	0.58	0.57	57.76%
怒り	0.51	0.61	0.55	58.99%
嫌悪	0.66	0.61	0.55	59.35%
無感情	0.63	0.57	0.60	57.56%

リプライのみを用いて特徴ベクトルをつくり SVM で学習した分類器は、幸福感などある程度精度が出ているが、全体的に十分に良いとは言えない結果となった。しかし、例えば「幸福感」についてより細かく見てみると、ツイートのみを用いて特徴ベクトルを作成した場合の分類器では False Negative だった事例 58 個のうち、27 個は提案手法において正例と判定されていた。このことは、ツイートのみでは判断出来なかった事例のうち、半数近くを提案手法が正しく分類できていることを意味している。

5. 考察

提案手法における実験結果を踏まえ、以下表 5 の事例に沿って考察する。これらは、感情ラベルが幸福感であるもののうち、ツイートのみから特徴ベクトルを生成した場合には負例と判断され、提案手法では正しく正例と判断できたものの例である。例えば、1 つ目の例において、ツイートのみで分類した場合に「寒い」「不安」や顔文字から、ネガティブな感情だと判断したと推測できる。しかしリプライは「無事」「良かった」などの典型的な反応から「幸福感」であると判断できていると考えられる。2 つ目や 3 つ目も同様にして「泣ける」「悲しい」「心が折れる」などネガティブな単語が含まれているが、リプライのポジティブな共感反応から正しく判断することが出来ている。また、最後の 4 つ目の例はツイートのみを見ると何が起きているか他人には分からないが、リプライを見る事で祝いという反応から発言者に嬉しい事があったと推測出来る。このように、特にツイートに複数の感情表現が含まれている場合に、リプライを見る提案手法が有効に働く事例が存在することが確認できた。

6. まとめ

本稿では、発言に対する他者反応を利用して感情推定を行う手法を提案し、マイクロブログサービスである Twitter から抽出したデータに対して評価実験を行った。ツイートのみの学習

