

行動促進の根拠を含むツイートの抽出手法

見塚 圭一[†] 鈴木 優^{††} 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

^{††} 奈良先端科学技術大学院大学情報科学研究科 〒630-0192 奈良県生駒市高山町 8916-5

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

E-mail: [†]m1724005@s.konan-u.ac.jp, ^{††}ysuzuki@is.naist.jp, ^{†††}nadamoto@konan-u.ac.jp

あらまし TwitterなどのSNSには、ユーザに行動を促進させることを目的としたツイートが投稿されることがある。そのようなツイートの中には、なぜその行動をした方が良いか等の行動促進の根拠を含むツイートが多数ある。しかしながら、このような根拠を含むツイートはその情報の真偽が不明であるにも関わらず、ユーザによって信用され、誤った行動や不利益を生む可能性がある。そこで、本研究ではある話題に対する行動促進の根拠を含むツイート内の情報とWeb上の情報を比較し提示する事により、ユーザが根拠部分の真偽を判断するための支援を行う研究を行う。本論文では、この研究のはじめの一歩として、Twitter上から根拠を含む行動促進ツイートの抽出手法を提案する。具体的には、まず、著者らによって設定したルールに基づいて、ユーザの行動を促進している部分を含むツイートを抽出する。次に、機械学習によって、行動促進の根拠となる部分を含むツイートを抽出する。さらに提案手法の有用性を示すため、評価実験を行う。

キーワード Twitter, 情報抽出, 機械学習

1. はじめに

SNS上には、さまざまな話題に対する投稿が存在している。例えば、ある話題に対する意見や感想、解説などが挙げられる。SNSの中でもTwitterは、ユーザが多く、誰でも情報を自由に気軽に発信することができる特徴がある。そのため、Twitter上には一般人による情報や専門家による情報など、さまざまな情報を持つツイートが散見される。それらツイートの中には他人に行動を促進させるような内容を含むツイートが存在する。本研究ではこのように他人に行動を促進させようとしていることを行動促進と呼ぶ。

行動促進させようとしているツイートは明示的な行動促進と暗示的な行動促進の2種類がある。明示的な行動促進を含むツイートとは、ツイート内に明らかに行動を促進している部分を含むツイートである。例えば、「アニサキスは冷凍保存または加熱処理をすることにより死滅するため、きちんと処理を行ってから食べましょう。」といったものである。一方、暗示的な行動促進を含むツイートとは、ツイート内には明らかに行動を促進している部分は存在しないが、ユーザがツイートを読んだときに書かれていることから想像できる行動をさせようとするような、暗示的に行動を促進しているツイートである。例えば、「アニサキスを生で食べると、数時間後から激しい腹痛や嘔吐などの症状が出る。」といったツイートでは、生で食べてしまうことが無いように、対策を行って食べる必要があると想像でき、実際に行動を起こすことがあるのではないかと考えられる。本研究では、明示的な行動促進を含むツイートに着目し、研究を行う。以下、明示的な行動促進を含むツイートの事を行動促進ツイートと呼ぶ。

さらに、行動促進ツイートの中にもなぜその行動を促進して

いるのかの根拠を書いているツイートと、行動促進部分だけが記載されているツイートがある。例えば、「ワクチン接種によるインフルエンザの予防効果はほとんどないとWHO、厚労省も認めています。またワクチンに水銀が添加されており長期投与の副作用も懸念されます。手洗い、うがい、鼻呼吸が予防接種以上に効果があります。という訳で、インフルエンザの予防接種はやめましょう。」と「インフルエンザの予防接種はやめましょう。」では、両方とも予防接種をやめましょうという行動促進をしている。しかしながら、前者はなぜその行動をしたほうがいいのか等の行動促進の根拠を含んでおり、後者は含んでいない。そして、前者のように根拠を含む行動促進ツイートはその情報の真偽が不明であるにも関わらず、根拠の書いていない後者のような行動促進ツイートよりも信じられてしまう可能性が高い。このような根拠を含む行動促進ツイートは多数存在し、その真偽は不確かである。実際に、20代の10名の被験者に対し根拠の有無によりツイートの信じられやすさが変化するのかについての予備実験を行った。予備実験は行動促進ツイートであり且つその根拠を記述してある30個のツイートをを用いた。具体的には、この30個のツイートに対し、根拠部分を含んだままのものと同根拠部分を削除したもののペアを用意する。このペアの内どちらか一方を被験者に提示し、被験者はそのツイートに対し、信じるか信じないかの判定を行った。そして数日後に同じ被験者に最初に見せたツイートのもう一方のツイートを提示し、同様に信じるか信じないかの判定を行った。その結果、80%の被験者が、根拠を含むツイートの方を信じると回答した数が多かった。この結果より、根拠が行動促進ツイート内に含まれることにより、信じられやすくなることが明らかとなった。その為、行動促進ツイートの根拠の真偽を判定することは重要であると考え、根拠を含む行動促進ツイートを抽出し、

その真偽を判定する研究を行う。

具体的には、ある話題に対する行動促進の根拠を含む行動促進ツイート内の情報と Web 上の情報を比較し、根拠部分に対する追加情報を提示する手法を提案する。本手法により、ユーザが情報の真偽を判断するための支援を行えることが期待できる。本研究の全体の流れを以下に示す。

- (1) Twitter 上からある話題に関するツイートの取得
- (2) 行動促進ツイートの抽出
- (3) 根拠を含む行動促進ツイートの抽出
- (4) 根拠部分と Web 上の情報を比較し、追加情報を取得
- (5) ユーザに対し、追加情報を提示

本論文では、この研究のはじめの一步として、(2)、(3) の Twitter 上の行動促進の根拠を含むツイートの抽出手法を提案する。具体的には、まず、行動促進部分を含むツイートをルールベースを用いて抽出する。次に抽出したツイート群から行動促進の根拠を含むツイートを機械学習を用いて抽出を行う。

以下、2章で関連研究について述べる。3章で行動促進ツイートの抽出方法について述べ、4章で行動促進ツイートの根拠部分の抽出方法について述べる。5章で手法に対する実験、考察について述べる。最後に6章で、まとめと今後の課題について述べる。

2. 関連研究

Twitter 上から情報を抽出する研究は数多くされている。古川 [1] らの研究では、Twitter 上に存在している犯罪に関する情報の分析を行い、Twitter 上にしか存在していないユーザの経験に基づく犯罪情報が存在していることを確認し、それらが有益な情報となることを示している。Popescu [2] [3] らの研究では、自動的にイベントに関連する情報を Twitter 上から抽出を行い、さらにイベントに対する意見をも抽出する手法を提案している。Phuvipadawat [4] らの研究では、最新のニュースに関するツイートをリアルタイムに抽出する。そして、それらの内容を基に分類し、ニュースに関する出来事を追跡し、提示する手法を提案している。これらの研究は、Twitter 上から情報を抽出するという点では類似しているが、本研究では、ある話題に関する行動促進の根拠を含むツイートを抽出するという点で異なる。

また、web 上の情報に追加情報を付与し、ユーザの判断支援を行うといった研究も多く行われている。柿本 [5] らの研究では、Twitter やブログなど Web 上の流言情報に対し、ユーザがページ閲覧時に流言情報や流言に対する訂正情報、詳細な情報へのリンクなどを提示し、ユーザに気づきを与え、情報の拡散防止を行おうとしている。藤川ら [6] は、Twitter 上で話題となっているツイートを抽出し、そのツイートに関連するツイートも取得する。そして、関連するツイートが話題のツイートに対し、信用しているか、疑問を持っているか、または、根拠が含まれているかを分類する。最後に抽出されたツイートの提示を行い、ユーザに話題となっているツイートが流言かどうかの判断の手助けをする手法を提案している。これらの研究では、ユーザに対し情報を与える際には、ツイートを利用してい

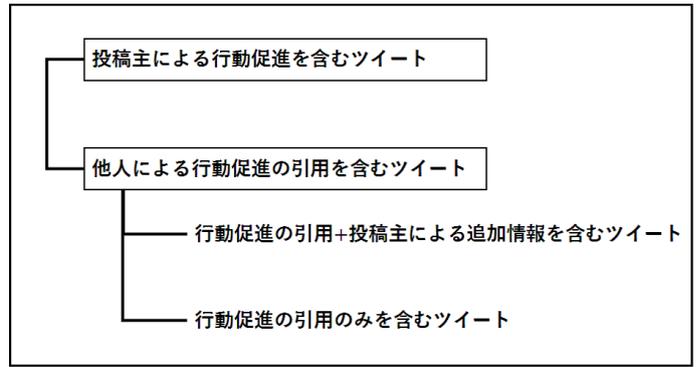


図1 行動促進ツイートの分類

るが、本研究は、Twitter 上のある話題の行動促進の根拠を含むツイートの根拠部分と Web 上の根拠に対する情報を比較し、追加情報の取得、提示を行い、ユーザが情報の真偽を判断する支援を行うという点で異なる。

3. 行動促進ツイートの抽出

3.1 行動促進ツイートの定義

本研究において、我々の定義する行動促進ツイートとは、他人に行動を促進するような内容を含むツイートの内、明示的に行動を促進させようとしている部分を含むツイートの事である。例えば、「地震や火事の際には落ち着いて行動しましょう。」や「食事の時にサラダを先に食べると脂肪がつきにくくなるので、先に食べましょう。」といったツイートである。

行動促進ツイートはいくつかのタイプに分類される。図1に我々が分類した行動促進ツイートの分類を示す。図1に示すように行動促進ツイートの中には、投稿主による行動促進を含むものと他人による行動促進の内容を引用したツイートが存在する。例えば、「医学協会の先生が「躊躇ってるうちに助かる確率どんどん下がるからね！怖くてもとにかく心肺蘇生と AED やりなさい！相手が後遺症なく社会復帰するためにも！」って感じのこと言ってた」といったように、他人の発言を引用しているツイートやリツイートに対し行動促進の部分を追加したようなツイートである。本研究では投稿主による行動促進ツイートのみならず、他人による行動促進の内容を引用したツイートも行動促進ツイートとして抽出する。

3.2 行動促進ツイートの抽出

本論文では、行動促進ツイートを抽出するはじめの一步として、ルールベースを用いる手法を提案する。なぜなら、行動促進ツイートは、人に行動を促すために、ツイート内に特徴的な語が現れやすいと考えた為である。行動促進ツイートの抽出に用いるルールを以下に示す。尚、ルールは文末のフレーズに着目しているが、文末の判断には、「。」「,」「!」「改行」の4種類を用いる。

ルール：文末に行動促進フレーズを含む。

このルールを満たすものを行動促進ツイートとして抽出を行う。表1に抽出された行動促進ツイートの例を示す。

3.2.1 行動促進フレーズ

明示的に行動促進を行う際には文末に特定の表現を含む傾向

表 1 行動促進ツイートの例（行動促進は太字で示している）

AEDの話はあれで終わると片手落ちなんで一応追記。「女性にを使って訴えられる」というのは都市伝説ですし、法的にも守られるはず（断言したい…）なんで躊躇無く使しましょう。一方、被救護者の尊厳も大事なんで、周囲の人に声を掛けて女性は内向きに、男性は外向きに立って壁になってもらうとか。
AEDは心肺停止状態でないと使えない仕様になってます、よく人に電流を流すの怖いって人いますが全責任はAEDが持ってくれるので怯まずバンバン使っていきましょう、あと人工呼吸時に直接口を付けずに出来る小道具などがAEDのポケットに入っている筈なので人工呼吸する時はそれを使おう！
AEDは正しく使しましょう。周りの方は出来る限り協力しましょう。
わたしがAEDの講習を受けたときに教えられたのは、最近人は人が倒れていても平気でスマホを向ける人がいるから、とくに女性の場合は目隠しのために誰かに人を呼んできもらって壁を作るとよい、周囲の人に協力を仰ぐときは必ずその人の目を見て、相手を指差して「あなたは110番してください」

表 2 行動促進フレーズ一覧

タイプ	内容
タイプ 1	文末に動詞の「意志形」を含む。
タイプ 2	文末に動詞の「命令形」を含む。
タイプ 3	文末が「動詞+と+良い」という形を含む。
タイプ 4	文末が「動詞+良い」という形を含む。
タイプ 5	文末が「〇〇した方が良い」という形を含む。

がある。本論文では、この行動促進を行う際に用いられる表現を行動促進フレーズと呼ぶ。表 2 に我々の提案する行動促進フレーズを示す。行動促進フレーズは、無作為に抽出したツイート群のうち、行動促進を行っているツイートを手動で抽出し、分析を行うことにより決定した。例えば、「呼吸が正常じゃないと思ったらすぐ AED を使おう。」といった文や「インフルエンザは部屋の湿度を高く保つことにより防止できるので、湿度には気をつけるとよい。」といった文の「使おう」や「気をつけるとよい」といった表現である。行動促進フレーズを含むツイートを行動促進ツイートとして抽出する。

以下、我々の提案する行動促進フレーズについて説明する。

タイプ 1： 文末に動詞の「意志形」を含む。

明示的に行動促進を行う際には動詞の意志形 [7] が多く用いられている。動詞の意志形とは、動詞が意志形の助動詞「(よ)う」に接続した形のことである。これは、他人に行動を誘いかける際に用いられることが多い。例えば、「呼吸が正常じゃないと思ったらすぐ AED を使おう。」といったツイートや「高血圧を予防することで、脳卒中や心筋梗塞などの病気の予防をしましょう。」といったツイートの「使おう」や「しましょう」といった表現である。

ただし、ツイート内に「どうしよう。」や「そうしよう。」といった「指示詞+動詞の意志形」、または「それにしよう。」や「どれにしよう。」といった「指示詞+助詞+動詞の意志形」という表現のみが出現しているツイートは削除する。これは、このような表現がツイート内に単体で出現している場合は対象の行動が不明確であるので、行動促進を行っているとはいえないためである。例えば、「週末にみんなが旅行へ行くようなので、僕は家で一人で遊ぶ。そうしよう。」や「AEDを使う相手が心停止していたらどうしよう。」といったツイートである。

タイプ 2： 文末に動詞の「命令形」を含む。

明示的に行動促進を行う際には命令形を含む場合が多い。こ

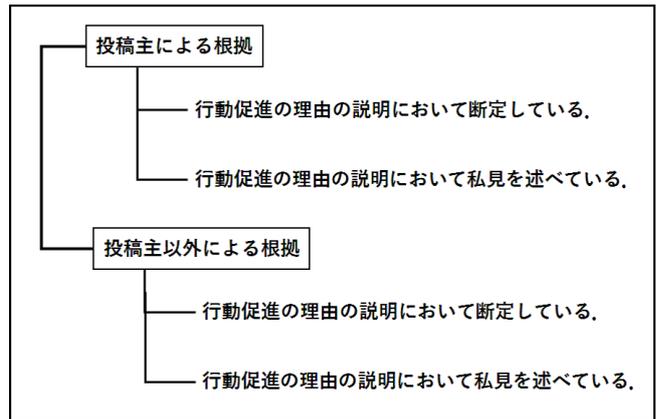


図 2 行動促進の根拠の分類

れは、他人に行動をしてもらう、もしくは強制的に行動させる際に用いられることが多いためである。例えば、「必要だと少しでも思うことがあれば、AEDはひるまずに使って下さい。」といったツイートや「塩分の取りすぎは高血圧の原因になり、動脈硬化・脳卒中・心筋梗塞・腎障害のリスクを増やします。適切な塩分接種を心掛けてください。」といったツイートである。
タイプ 3： 文末が「動詞+と+良い」という形を含む。
タイプ 4： 文末が「動詞+良い」という形を含む。
タイプ 5： 文末が「〇〇した方が良い」という形を含む。

タイプ 3~5 で用いられている表現は、明示的に行動をする事により何か効果が得られることが表現されている為、行動促進をしている表現であると考えた。例えば、「心肺蘇生で胸骨圧迫する際は、上から衣服をかけてあげるとよい。」といったツイートや「AED が出来ないなら救急車を呼べば良い。」、「AED を探して持ってくる間は心臓マッサージした方がいい。」といったツイートである。

以上のいずれかの行動促進フレーズを含むツイートを行動促進ツイートとして抽出する。

4. 行動促進ツイートの根拠部分の抽出

4.1 行動促進ツイートの根拠部分の定義

本研究では、行動促進を行う理由の説明を含むものを行動促進ツイートの根拠部分とする。行動促進の根拠部分は図 2 に示すように、4 種類に分類する。本研究では、投稿主による行動促進の根拠が含まれておりかつその根拠を断定しているものと、

表 3 行動促進の根拠を含むツイートの例（行動促進は太字，根拠は下線で示している）

AED を健康な人に使ったらどうなる？ 答えは「何も起こらない」AED は自動で患者がショックを必要としているかどうかを診断して必要に応じてショックを与える機械なので、健康な人にショックを与えないようになっているんだ。呼吸が正常じゃないと思ったらすぐ AED を使おう！
一応言っておこう。AED を女性に対し使う為に服を脱がせたとしても、刑法上緊急避難が成立する可能性が高いです。使用する側の認識云々なので、たとえ AED 要らんかったとしても問題ありません。むしろ邪魔した方がアウトになる可能性があります。なので安心して使しましょう。
AED はメーカー色々ありますが多分ほとんど蓋を開けたらしゃべって指示してくれます。不要だったら不要だったでいいので（作動しなかった=致死性不整脈はなかった、というのもひとつの情報となる）、使用を迷わないでください。
AED を使う際、相手が女性で周りに人がいる場合は外側を向いて囲むように壁を作るように教えてもらう。女性への配慮もあるし、そういうことで免疫のない人が目にしないで済むことにもなるだろうし、最近では写真を撮られるらしいからそれの対策にもなる。だから女性でなくても壁は作ったほうがいい。

投稿主以外が行っている行動促進の根拠を引用しておりかつその根拠を断定しているものという 2 種類を行動促進の根拠として抽出を行う。例えば、「インフルエンザにかかった場合、若くて健康な人は数日間自然回復します。無理な受診が感染拡大と医療費増大を招きます。自宅で水分と栄養を補給して十分に睡眠を取りましょう。」や「冬になって、気温が低下すると体温が逃げないよう血管が収縮します。血管が収縮することで血圧が上がり、危険な状況となりやすくなります。」と医者が言っていたので外と室内の温度差にも注意しましょう。」といったツイートである。

ここで、投稿主または投稿主以外が私見を述べているものを根拠に含めない理由としては、明確に個人の考えを含むと判断できるツイートは信じられる可能性が低い傾向にあると考えたためである。例えば、「病院にインフルエンザの予防接種を受けに行くときに感染する可能性があると思うので、感染拡大を防ぐためにも予防接種を受けに行かないようにしましょう。」や、「友達の話によると、風邪の原因は汚れた手で目や鼻を触ることだと思うとのことなので、触らないようにするといい。」といったツイートである。このようなツイートは本研究では根拠として扱わない。表 3 に行動促進の根拠を含むツイートの例を示す。

4.2 行動促進の根拠部分の抽出

抽出した行動促進ツイートから行動促進の根拠部分を含むツイートを機械学習を用いて抽出する。機械学習には、文脈を考慮することのできる、再帰型ニューラルネットワーク (RNN) を利用する。根拠部分の抽出時には、行動促進をしている部分を含むツイートを行動促進部分と根拠候補部分に分割する。これにより、根拠候補部分に行動促進に対する理由の説明を含むかどうかを判定できると考える。

4.2.1 RNN の構造

文の単語の順番を考慮することの出来る LSTM を再帰的ニューラルネットワークに適用したモデルを利用し、学習を行う。深層学習の実装には、Python の機械学習ライブラリである Chainer を用いる。

4.2.2 教師データ

本研究では、“根拠”を“行動促進を行っている理由の説明を含むもの”と定義している。そこで、根拠を抽出する際、文章の書き方に着目する。“根拠”となりうる文章の書き方として、ある物事の説明を行っている書き方の文章が根拠部分の抽

出に適していると考えた。その為、根拠データの正解データとして、辞書の書き方を用いる。本論文では、辞書的な書き方がなされている Wikipedia を利用する。不正解データとして、TwitterAPI を用いて無作為に抽出したツイートを用いる。ツイートを引く理由は、ある話題に関するツイートを収集した場合は、辞書的な書き方がなされているツイートが複数出現するが、無作為にツイートを抽出した際には、殆どのツイートは投稿主の主観を述べているものが多く、説明を含むものが少ないと考えたためである。これらを教師データとして用い、学習を行う。Wikipedia を利用する際には Wikipedia の記事本文を文単位に区切ったものを利用する。これは、Twitter では 1 ツイート当たり 280 字と制限があるため、同じ条件で学習させる必要があると考えたためである。

4.2.3 入力データの作成

RNN の学習時の入力には、単語を分散表現を用いてベクトル化したものを用いる。単語の分散表現には Word2Vec を用いる。Word2Vec で学習を行う際の教師データは、Wikipedia の全文書を用いる。形態素解析器には Mecab を用いる。Word2Vec の実装には、Python ライブラリの gensim を用い、学習方式には CBOW を使用した。ベクトルサイズは 200 次元である。

4.2.4 根拠部分と行動促進部分への分割

行動促進ツイート内に根拠を含むかどうかの判定時には、行動促進ツイートから行動促進部を除いて、残った部分に対して根拠を含むかどうかの判定を行う。これは、行動促進部の書かれ方と根拠部分の書かれ方が異なるため、行動促進ツイート内に行動促進部を含んだ状態では、根拠部分を含むかどうかの判定を正確に行うことが出来ないと考えたためである。

5. 実験

我々の提案する行動促進ツイートの抽出手法の有用性を調べるためにルールベースを用いた行動促進ツイートの抽出の評価実験を行った。

5.1 実験条件

実験データ

ある話題語を用いて取得したツイートから行動促進ツイートをルールを用いて抽出する。使用した話題語は表 4 に示す 5 つである。ツイートは各々の話題語に対して、無作為に抽出した 200 件のツイート、合計 1000 件のツイートである。話題語の

表 4 使用した話題語一覧

AED	心筋梗塞	インフルエンザ	震災	大雪
-----	------	---------	----	----

表 5 各話題語ごとの評価

話題語	適合率	再現率	F 値
AED	0.77	0.81	0.79
心筋梗塞	0.75	0.92	0.82
インフルエンザ	0.74	0.88	0.83
震災	0.63	0.96	0.74
大雪	0.66	0.84	0.74

選択の理由としては、健康に関する話題や災害に関する話題は情報を手に入れることが容易であり、それを発信し、行動を促進するツイートが多く存在していると考えたためである。

被験者

被験者はクラウドソーシングで集めた 20 人である。ツイート内に行動促進に関する部分が含まれているかどうかを、被験者により人手で判定した。

5.2 実験手順

以下に実験手順を示す。

- (1) 話題語を含むツイート無作為に抽出し、被験者に提示。
- (2) 行動促進を行っているツイートであるかを被験者が「○」、「×」で評価。ここで、そのツイートが行動促進を行っていれば「○」、行動促進を行っていなければ「×」と評価する。
- (3) 各話題語に対し、(1) から (2) の処理を実施。

5.3 結果と考察

正解データは行動促進ツイートであると判断した被験者が過半数を超えたものとする。実験結果より、適合率は 0.71、再現率は 0.88、F 値は 0.79 となった。ここで、正解データの数は 1000 ツイート中、333 個であった。また、ルールベースにより抽出された行動促進ツイートの数は 414 個で、その中に含まれる正解データは 294 個であった。このことから、ルールベースにより、行動促進ツイートの抽出を行うことができる程度でできていることがわかる。

各話題語ごとの適合率、再現率、F 値を表 5 に示す。結果より、どの話題語に対しても行動促進ツイートのある程度抽出できていることがわかる。これは、今回選択した話題語が、健康に関連するものや、災害に関するものと、注意喚起や行動促進を行いやすい話題語を選択したためと考えられる。

以下に使用したルールについての考察を述べる。

各行動促進フレーズのタイプごとのシステムによる抽出結果の正解ツイート数と適合率を表 6 に示す。表 6 より、タイプ 1、タイプ 2 の数が多いことがわかる。これは、行動促進を行う際に、文末に意志系、命令形を含む事が基本的な形であるからだと考えられる。一方、タイプ 3~5 の数は非常に少なくなっており、行動促進を行うような表現であるがあまり使われない表現であることがわかる。これは、明示的に行動を促進する際には、明らかに行動を促進しているとわかりやすいような表現が好まれるからではないかと考えられる。

次に、各行動促進フレーズのタイプごとの適合率については、

表 6 タイプごとのツイート数と適合率

タイプ	タイプ 1	タイプ 2	タイプ 3	タイプ 4	タイプ 5
正解ツイート数	124	152	4	2	12
適合率	0.71	0.74	0.80	0.20	0.63

タイプ 4 以外のタイプにおける適合率はある程度高いものとなっており、有用性があると考えられる。タイプ 4 は適合率が低いいため、不要な行動促進フレーズであると考えられる。

以下に使用した各行動促進フレーズごとの結果、考察について述べる。

タイプ 1: 文末に動詞の「意志形」を含む。

表 7 にタイプ 1 を満たすツイートの抽出例を示す。行動促進ツイートと判定されているツイートのように、文末に意志系を含むツイートは大半が行動促進ツイートであると判定された。一方、行動促進ツイートではないと判定されたツイートの特徴としては、同じ表現であるが、投稿主の意志や願望を示している表現となっている、このため、行動を促進しているとは言えず、行動促進ツイートではないと評価した被験者が多くなったと考えられる。

タイプ 2: 文末に動詞の「命令形」を含む。

表 8 にタイプ 2 を満たすツイートの抽出例を示す。行動促進ツイートと判定されているツイートのように、文末に命令系を含むツイートも大半が行動促進ツイートであると判定された。一方、行動促進ツイートではないと判定されたツイートの特徴としては、タイプ 1 と同様に、投稿主の期待や願望を示している表現となっている、そのため、行動を促進しているとは言えず、行動促進ツイートではないと評価した被験者が多くなったと考えられる。

タイプ 1 とタイプ 2 に関しては、そのツイートが投稿主の行動に関する表現であるかを考慮することにより、同じ表現でも行動を促進しているかの判別を行うことができるのではないかと考えられる。

タイプ 3: 文末が「動詞+と+良い」という形を含む。

表 9 にタイプ 3 を満たすツイートの抽出例を示す。タイプ 3 によって抽出されたツイートは 5 件と少ないながらも、1 件を除き、残りすべてが行動促進ツイートとして判定された。行動促進ツイートと評価されたツイートは、具体的な行動を伴った行動促進の内容が含まれている。そのため、行動促進ツイートであると判定されたと考えられる。行動促進ツイートではないと評価されたツイートは、個人の意見を示す表現であると判断されたため、行動促進ツイートではないと判断されたと考えられる。

タイプ 4: 文末が「動詞+良い」という形を含む。

表 10 にタイプ 4 を満たすツイートの抽出例を示す。タイプ 4 は適合率が他のタイプに比べ、極端に低い結果となった。行動促進ツイートでないとして評価されたツイートを確認すると、“しとけばいい” や、“すればいい” といった個人の願望や意見を示す表現が多数見られた。その為、行動促進を行っていないと判断され、適合率の低下に繋がったと考えられる。

タイプ 5: 文末が「○○した方が良い」という形を含む。

表 7 タイプ 1 を満たすツイートの例（行動促進は太字で示している）

行動促進かどうか	ツイート
○	てか AED 使うときって一秒でも無駄に出来ないし、救急車来るまで心肺蘇生も並行してやらないといかんし配慮だなんだ考えてる時間なんて全くない。AED 使うときは躊躇なく脱がせましょう。
○	阪神淡路大震災から今日で 23 年。あの震災の記憶や教訓を後世へと繋いでいこう。#阪神淡路大震災から 23 年
×	だいぶ熱が下がってきた。一応インフルエンザじゃないか、明日病院行ってこよう。
×	明日は休みだったけど、インフルエンザで欠員が出たので、そのまま連続勤務する事となった。まあ仕方ないだろう。いつ立場が逆転するか分からないから文句は言わず出勤しよう。(´ 〇 `)

表 8 タイプ 2 を満たすツイートの例（行動促進は太字で示している）

行動促進かどうか	ツイート
○	東京が 4 年ぶりの大雪とかいってるけど、ほんとに車乗る人は冬タイヤにしてください。まじで、5mm 積もったくらいどうでもいいでしょとか、3cm 積もったくらいでとかじゃなくて。雪ふったら冬タイヤにしてください。雪積もってんの普通のタイヤの人の気が知れませんか。怖すぎかよ。
○	本日で阪神大震災から 23 年被災者の 1 人として、あの震災を忘れないようにしたい。皆様も地震災への備えを今一度ご確認ください。
×	あと上岡龍太郎さんが報道のヘリがうるさすぎて、あいつら何を考えとるんだ！と怒ってらっしゃったのと、小林カツ代さんとケンタロウさんが震災で飼い主さんとはぐれてしまった犬や猫の保護活動をお仕事の合間を縫って長くなさったのは毎年呟かせて下さい。
×	同僚がインフルエンザかもつってインスタに体温計の写真アップしてるのを見るとイライラする。体調管理くらいちゃんとしてください。アホですか？

表 9 タイプ 3 を満たすツイートの例（行動促進は太字で示している）

行動促進かどうか	ツイート
○	11 個目水分補給はシッカリと。これを欠かすと脳卒中や心筋梗塞の原因になったりする。特にメ切前にカフェインを摂取するなら水分補給は重要となる。スポーツ飲料はあまりオススメしないが飲みたいなら水で半分は薄めて飲むといい。
○	前に仕事で AED の話題がちょうど出ていたので。女性に AED で電気ショックをする際はは、ブラは外すこと。金具で感電するからです。ネックレスは首から上に避けること。心肺蘇生で胸骨圧迫する際は、上から衣服をかけてあげるとよい。
×	震災の日はサイモン&ガーファングルの明日にかける橋を聞くと良い。

表 11 にタイプ 5 を満たすツイートの抽出例を示す。タイプ 5 では、行動促進ツイートと判定されているツイートは、理由とともに使用されている場合が多く見られた。一方、行動促進ツイートではないと判定されたツイートの特徴としては、投稿主の意見や経験を示している表現となっているため、行動を促進しているとは言えず、行動促進ツイートではないと評価されていると考えられる。

タイプ 3 からタイプ 5 に関しては、タイプ 4 においてのみ、適合率が極端に低くなっているため、タイプ 4 は行動促進ツイートの抽出に適していないといえる。

これらの結果より、提案するルールベースによる行動促進ツイートの抽出は有用であることがわかった、しかしながら、「インフルエンザだと知りました。ゆっくり休んで、お大事にしてくださいね。また面白いツイート待っています^^」や「心筋梗塞、心身のバランスの乱れなどがあります、一度精密検査か、ゆっくり休む事をお勧めします(´ω´)」といった提案したルールのみでは抽出することの出来ないツイートがある。このように取りこぼした行動促進ツイートが存在しているため、今後、ルールの追加や機械学習を用いた行動促進ツイートの抽出手法を考える必要がある。

6. まとめと今後の課題

本論文では、行動促進の根拠を含むツイート内の情報と Web 上の情報を比較し提示する事により、ユーザが根拠部分の真偽を判断するための支援を行う研究のはじめの一歩として、Twitter 上から根拠を含む行動促進ツイートの抽出手法を提案した。具体的には、まず、行動促進の内容を含むツイートをルールを用いて抽出する。次に、抽出したツイート群から行動促進の根拠を含むツイートを機械学習を用いて抽出する手法の提案を行った。さらに実験を行い、ルールを用いた行動促進部分の抽出手法の有用性を示した。

今後の課題としては、行動促進ツイートを機械学習を用いて抽出することにより、更なる抽出精度の向上を目指すことである。また、行動促進の根拠部分を含むかどうかに関する実験を行い、提案手法の有用性を測ることが挙げられる。さらに、根拠部分とウェブ上の情報を比較し、ユーザに真偽の判断を支援する情報を提示する手法の提案を行う予定である。

謝 辞

本論文の一部は JSPS 科研費 17K00430, 16K07973 及び、

表 10 タイプ 4 を満たすツイートの例（行動促進は太字で示している）

行動促進かどうか	ツイート
○	AED やるとき心臓マッサージするでしょ？肋骨が折れるくらい強く押していい。骨が折れてもいい。違法性は阻却されるからセクハラにも傷害罪にもならないのは明らか。みんな AED 訓練したことないの？
×	救急外来でインフルエンザの異常行動について説明する先生としない先生がいるのは仕方ないと思う。重症患者の処置のほうを優先すべきだ。薬剤師がきちんと説明すればいい。
×	震災や事故が起こる前から、代理店や巨大メディアを通じた「クリーンでエコ、コストかからない」といったプロパガンダが大嫌い。小泉だってプロパガンダの政治家なんですから、今さら何言ってるんだよって思わないほうがそもそもおかしい。信用するしないのレベルじゃありません。ほっとけばいい。

表 11 タイプ 5 を満たすツイートの例（行動促進は太字で示している）

行動促進かどうか	ツイート
○	昔そういう義務がある職業についてたけど、上半身脱がせるのはやむを得ないし裁判でも勝てるからやった方が 良い 。何より AED 使わなきゃいけないレベルだと一刻を争うレベルだから。
○	んとほんと。AED は少しでも遅れると救命率落ちるから、余計なこと考えず、声かけに反応がなければすぐに やった方が 良い。上半身全裸にしないといけないので 勇氣 はあるけど、人のいるところで機械を横に置いてるのにわいせつ目的だなんて思う馬… https://t.co/2vS7wZddvX
×	水道水で作れるウォーターサーバーって、結構デカイんだけど、ウリは宅配や水の買い置きなどいらぬ… だけ けど、それ、どデカイ浄水器、というのではないんだらうか？最近、災害対策にウォーターサーバーを検討中。震災の時、ウォーターサーバーに大分助けられたから。買い置きは あった方が 良い。

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] 古川 忠延, 阿部 修也, 安藤 剛寿, 岩倉 友哉, 志賀 聡子, 高橋 哲朗, 井形 伸之, “Twitter からの犯罪情報抽出の可能性調査”. 研究報告 デジタルドキュメント (DD), 2011-DD-82 巻, 3 号, pp1-6, 2011.
- [2] A-M. Popescu, M. Pennacchiotti, . Detecting Controversial Events from Twitter. In Proceeding CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management, pp1873-1876, 2010.
- [3] A-M. Popescu, M. Pennacchiotti, D. Paranjpe. Extracting events and event descriptions from Twitter. In Proceeding WWW '11 Proceedings of the 20th international conference companion on World wide web, pp105-106, 2011.
- [4] S. Phuvipadawat, T. Murata. Breaking News Detection and Tracking in Twitter. In Proceeding WI-IAT '10 Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03, pp120-123, 2010.
- [5] 柿本 大輔, 荒牧 英治, 宮部 真衣,, “流言拡散防止のための情報確認行動促進システムの開発”. ライフインテリジェンスとオフィス情報システム研究会 (LOIS), IEICE-LOIS2016-87, pp141-146, 2016.
- [6] 藤川 智英, 鍛冶 伸裕, 吉永 直樹, 喜連川 優, ” マイクロブログ上の話題抽出とユーザの態度の分類に基づく流言検出支援システム, ” 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), F7-1, 8 pages, 2012.
- [7] 加藤 重広, 日本語文法入門ハンドブック, 研究社, 2006.