一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
DE2018-17 (2018-09)

Twitter 上の行動促進ツイート抽出手法の比較検討

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
〒 658-8501 兵庫県神戸市東灘区岡本 8-9-1
E-mail: †mizuka000k1@gmail.com

あらまし　Twitter 上には，様々な情報が溢れている．その中には，他人に行動を促しているような内容のツイートがある．本研究では，このようなツイートを「行動促進ツイート」と呼ぶ．行動促進ツイートはその真偽にかかわらず，ユーザがツイートを見てそのツイートが促すように行動して不利益を被る場合がある．そこで，我々は行動促進ツイートを抽出し，ユーザにアラートを出して行動促進を行っていることを伝える必要があると考えた．本論文ではその内，行動促進ツイートの抽出手法を提案する．具体的には，行動促進ツイート抽出手法としてルールベース，SVM，ランダムフォレスト，LSTM の 4 つの手法を用いて実験を行い，どの手法が適しているかを比較検討する．そして，最適な行動促進ツイートの抽出手法を提案する．

キーワード　Twitter，情報抽出，機械学習

A Comparison of The Method for Extracting Behavioral Facilitation Sentences from the Twitter

Keiichi MIZUKA†, Yu SUZUKI ††, and Akiyo NADAMOTO†††

† Graduate School of Natural Science, Konan University
8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo, 658-8501, Japan
†† Graduate School of Information Science,NARA INSTITUTE of SCIENCE and TECHNOLOGY
8916-5 Takayama-cho，Ikoma，Nara，630-0192 Japan
††† Faculty of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada-ku, Kobe, Hyogo, 658-8501, Japan
E-mail: †mizuka000k1@gmail.com

Key words　Twitter, Extracting Information, Machine Learning

1. はじめに

代表的なマイクロブログである Twitter 上には，ある話題に対する意見や説明など，様々な情報が存在している．また，Twitter の特徴として，ユーザ数が多く，情報の発信が容易で，自由に内容を記述できるという点が挙げられる．そのため，膨大な量の情報が Twitter 上に散見される．そして，ある話題に対する情報の中には，他人に行動を促進しているツイートがある．例えば，ダイエットの話題では，“ダイエットのときは，よく噛んで食べろ！” といった強く行動を促進しているツイートや，“ダイエットの時は，サラダから食べよう！” といった提案のように弱く行動を促進しているツイートがある．本研究では，このようにユーザに行動を促進しているツイートを「行動促進ツイート」と呼ぶ．また，ユーザは日々大量のツイートを見る場合が多い．そのため，多数のツイートの中に埋もれている行動促進ツイートの流し読みをして，無意識にその内容を信じて行動する場合がある．そして，誤った行動促進を行っているツイートを信じ行動することにより，不利益を被る可能性が

ある．

10 代のスマートフォン利用に関する実態調査(注1)によると，10 代の男女が最も好きな情報源は Twitter が 1 位となっている．そして，最も信頼する情報源として，Twitter と回答した被験者が 5.3%存在している．このことから，Twitter から情報を得て，その内容を信じている若いユーザが少なからず存在していることがわかる．

そこで我々は，行動促進ツイートに対して「このツイートはあなたに行動を促進しています」というようなアラートを出す事により，ユーザに内容を考える機会を提供することができると考える．さらに，誤った情報が容易に拡散されてしまうことを防ぐことができると考えられる．そこで，本研究では行動促進ツイートに対して，アラートを提示する手法の提案を行う．本研究により，誤った情報を鵜呑みにし，不利益を被るユーザを減少させられると考えられる．本研究の手順は以下の通りである．

（1） ある話題に含まれる行動促進ツイートの抽出．

（2） 抽出した行動促進ツイートに対しての真偽を判定．

（3） 行動促進ツイートに対してアラートの提示．

本論文では，その内（1）の Twitter 上の行動促進ツイートの抽出手法を提案する．

これまで我々はツイートを分析し，行動促進ツイートには文末に特徴的な表現が存在していることを確認した．そこで，ルールを用いて行動促進ツイートの抽出を行った [1]．しかしながら，ルールを用いて行動促進ツイートを抽出した結果，ルールだけでは誤って抽出された例が多数あった．花粉症を例にすると “花粉症の人は，病院に行こう！” と “花粉症になったので病院に行こう！” といったツイートがある．両ツイートは共に “行こう” という行動を促進している表現が含まれているが，前者は行動促進をしており，後者は投稿者の意志を示している．このように同じ表現であっても，行動を促進している場合と，促進していない場合がある．そこで，その問題に対応するため，本論文では，機械学習を用いた手法を提案する．具体的には，Support Vector Machine(SVM)，ランダムフォレスト，Long Short-term Memory(LSTM) の 3 つの機械学習を用いた．さらに，この 3 つの機械学習の手法とこれまで提案したルールベースの手法の 4 つの手法の比較実験を行う．そして，行動促進ツイートの抽出に最も適した手法はどの手法であるのかを検証する．

以下，2 章で関連研究について述べる．3 章で 4 つの行動促進ツイートの抽出方法について述べ．4 章で 4 つの提案手法を用いた比較実験を行い，その考察について述べる．最後に 5 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行われている．矢野 [2] [3] らの研究では，ユーザの行動をツイートから抽出する手法を提案し，感情と行動の関係の分析を行っている．古川 [4] 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犯罪の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 [5] [6] らの研究では，Twitter 上からあるイベントに関連する情報を自動的に抽出を行い，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 [7] らは，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し，それらの内容を基に分類し，提示する手法を提案している．Miyabe [8] ら 梅島 [9] らの研究では，マイクロブログ上の流言の分析を行い，流言の拡散を防ぐために，Twitter 上の流言情報の抽出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本研究と類似しているが，本研究では，ある話題に関する行動促進ツイートを抽出するという点で異なる．

表 1 ルールベースに使用する行動促進タイプ一覧

タイプ	内容
タイプ 1	ツイート内に動詞の「意志形」を含む．
タイプ 2	ツイート内に動詞の「命令形」を含む．
タイプ 3	ツイート内に「動詞+と+良い」という形を含む．
タイプ 4	ツイート内に「○○した方が良い」という形を含む．

3. 提 案 手 法

これまで我々の提案した (1) ルールベースによる行動促進ツイートの抽出手法 [1] に変更を加えたもの，さらに (2)SVM と (3) ランダムフォレスト，(4)LSTM の 3 種類の機械学習を用いたもの合計 4 種類の行動促進ツイートの抽出手法を提案する．

3.1 ルールベースを用いた抽出手法

我々が行ったツイートの分析 [1] では，行動促進ツイートの場合，「～ましょう」や「～しよう」といったように，行動促進をしている部分の末尾にその特徴が現れていることがわかった．ツイートは様々な人が自由に記述しているため，行動促進をしている部分が必ずしも文の末尾であるとは限らない．例えば，「正しいダイエットをすることで、筋力低下や内臓疾患を防ぐことができることを知っておきましょう！」や，「憂鬱な朝にはダイエット食材としても有名なバナナを食べましょう\nバナナは精神を安定させます。」のように，感嘆符の直前であったり，改行の直前である場合が多数ある．そこで，本研究では「。」，「、」，「！」，「\n(改行)」の直前の行動促進を行っているフレーズを「行動促進フレーズ」と呼ぶ．そして，その行動促進フレーズ内の形態素の品詞構成に着目し，表 1 に示す 4 つのタイプを提案する．以下 4 つのパターンについて説明する．

タイプ 1 とタイプ 2

「食べよう」や「食べろ」といった表現は行動促進を行う際に，「。」，「、」，「！」，「\n(改行)」の直前に出現することが多い傾向があるためタイプ 1 とタイプ 2 を提案する．その内，タイプ 1 は「ダイエットのために，食前 30 分は運動しよう。」といった意志形を含むタイプである．タイプ 2 は「ダイエットのときは食後 4 時間以上経ってからは睡眠を取れ！」といった命令形を

(注1)：10 代のスマートフォン利用に関する実態調査
https://marketing-rc.com/report/report-teensmartphone-20151120.html

表 2　ランダムフォレストにおける素性の組み合わせ一覧
(n=名詞 , v=動詞 ,adj=形容詞, p=助詞, p_end=終助詞, aux=助動詞)

No	組み合わせ			No	組み合わせ			No	組み合わせ			No	組み合わせ			No	組み合わせ		
1	n	v	v	10	n	aux	adj	19	v	v	aux	28	v	p	v	37	aux	adj	p_end
2	n	v	adj	11	n	aux	aux	20	v	v	p_end	29	v	p	adj	38	aux	aux	v
3	n	v	aux	12	n	aux	p_end	21	v	adj	v	30	v	p	aux	39	aux	aux	adj
4	n	v	p_end	13	n	p	v	22	v	adj	aux	31	v	p	p_end	40	aux	aux	aux
5	n	adj	v	14	n	p	adj	23	v	adj	p_end	32	aux	v	v	41	aux	aux	p_end
6	n	adj	adj	15	n	p	aux	24	v	aux	v	33	aux	v	aux	42	aux	p	v
7	n	adj	aux	16	n	p	p_end	25	v	aux	adj	34	aux	v	p_end	43	aux	p	adj
8	n	adj	p_end	17	v	v	v	26	v	aux	aux	35	aux	adj	v	44	aux	p	aux
9	n	aux	v	18	v	v	adj	27	v	aux	p_end	36	aux	adj	aux	45	aux	p	p_end

含むタイプである．

タイプ 3 とタイプ 4

「食べると良い」や「避けた方がいい」といった表現が他人に物事を勧める際に使用されることが多い傾向があるため，タイプ 3 とタイプ 4 を提案する．タイプ 3 は「ダイエットのときは，アサイーを食べると良い。」といったように「動詞＋と＋良い」というタイプである．タイプ 4 は「ダイエットのときは，間食は避けた方が良い。」というように「〜した方が良い」というタイプである．

このように，4 つの行動促進タイプを提案する．そして，これら行動促進タイプに当てはまる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．

3.2　SVM を用いた抽出手法

行動促進ツイートの抽出は，行動促進ツイートで「あるか」，「ないか」の 2 クラス分類によりできると考え，2 クラス分類の代表である SVM を用いて学習する．SVM の学習には機械学習ライブラリである Scikit-learn を使用する．SVM のカーネルは RBF カーネルを用いる．コストパラメータ c は 1000，カーネルパラメータ g は 0.001 とする．

使用する素性は，ツイート中の動詞，助動詞に行動促進ツイートの特徴が現れていると考えられることから，教師データに対し形態素解析を行い得られた動詞，助動詞を対象とする．動詞と助動詞は，活用形が行動促進であるかどうかに影響を与えていると考えられるため，終止形には戻さずにそのまま使用する．これらの各語を Word2Vec を用いてベクトルに変換し，ツイート中の各単語のベクトルの平均値を素性とする．Word2Vec の学習済みモデルデータとして，東北大学乾研究室の日本語 Wikipedia エンティティベクトル(注2)を利用する．

3.3　ランダムフォレストを用いた抽出手法

機械学習を用いることにより，ルールベースで決定したルール以外の行動促進ツイートが抽出できると考えた．そこで，ランダムフォレストを用いて行動促進ツイートの抽出を行う．学習には機械学習のライブラリである Scikit-learn を使用する．決定木の個数は 500，決定木の個数以外のパラメータはデフォルトのものを使用する．

(注2)：日本語 Wikipedia エンティティベクトル
http:www.cl.ecei.tohoku.ac.jp/ m-suzuki/jawiki_vector/

表 3　各提案手法の実験結果

提案手法	適合率	再現率	F 値	10 交差検定
ルールベース	0.6940	0.8661	0.7775	——
SVM	0.5130	0.5275	0.5630	0.5447
ランダムフォレスト	0.5687	0.8464	0.6803	0.4726
LSTM	0.8340	0.7913	0.8121	0.4768

使用する素性は，表 2 における品詞の組み合わせとなっているツイート中のある部分の単語ベクトルである．単語ベクトルには one-hot ベクトルを用いる．ツイートに表 2 の品詞の組み合わせが含まれていれば 1，含まれていなければ 0 とした．

例えば，「控えましょう」であれば，「控え（動詞)」＋「ましょ（助動詞)」＋「う（助動詞)」の組み合わせとなっているおり，表 2 の No.26 に当たるため，この部分の素性ベクトルは 1 となる．

3.4　LSTM を用いた抽出手法

行動促進ツイートの抽出時に，ツイート内の単語の出現順序を考慮し，学習させることにより，正しく判定が行えると考え，文の単語の順番を考慮することの出来る LSTM を用いて学習を行う．LSTM の実装には，Python の機械学習ライブラリである Chainer を用いる．

素性は，ツイート内に出現する単語を Word2Vec を用いてベクトル化を行ったものを使用する．使用する単語は，教師データ内から，url，Twitter のユーザ名を取り除いた後のデータに対して形態素解析を行い，得られた全品詞を用いる．全品詞を用いる理由としては，文の単語の流れを捉えることにより，同様の表現でも違う意味が存在するという問題に対応することができると考えたためである．LSTM の各種パラメータは，隠れ層の数は 1，ユニット数は 200，バッチサイズは 250，エポック数は 300，学習率は 0.001，オプティマイザーは Adam を利用する．

4.　実　　験

提案した 4 つの手法による行動促進ツイートの抽出精度の比較実験を行う．

実験データ

実験データはアンケートにより行動促進ツイートであるかそうでないかを判定したツイートを用いる．アンケートにはクラ

表 4　ルールベースによる行動促進ツイートの抽出例

No.	正解	システム	ツイート
1	○	○	糖質制限ダイエットをしていて\n よくある失敗談は\n ご飯やパンなどの主食を抜くことで\n 無性にお腹が空いてその結果\n お菓子をドカ食いしてしまうことだ\n これは結果的に主食を普通に\n 食べることよりも太ることに\n なるため、\n 変に制限するくらいなら\n 食べてしまう方がいいです
2	○	×	【豆知識】\n ダイエットのポイント!\n あなたは 1 日に何回トイレに行く？\n 8 回以下の人は、老廃物や毒素が完全に出ていないので、\n 代謝をよくしたり、利尿作用があるものをとって回数を増やすことが大切。
3	×	○	ダイエットサポートきたーーー！ 新しい服も可愛い。着てほしい服ばかり。\n……よし。まずは健康的な体重を目指そう。
4	×	×	あさごはん\n おかし食べたので\n カロリー様子見のためにアーモンドは♀\n 飽和脂肪酸すごすぎる\n ついに豆乳の特濃がおいしく\n 飲めるようになった〜笑笑\n ダイエットがはかどる

表 5　SVM による行動促進ツイートの抽出例

No.	正解	システム	ツイート
1	○	○	健康に減量しましょう！無理はダメ！\n 飲み物はノンカロリー。\n やはりタンパク質を多めに摂る。\n 脂身ののない肉は欠かせない栄養素。\n 野菜とバランスよくとりましょう。\n 睡眠も大事です。睡眠はカロリーを消費するので、寝ながらのダイエット。\n ホルモン分泌の関係がありそう！
2	○	×	【基礎代謝 UP】\n 筋肉量が少ない人はエネルギー生産量が少なく基礎代謝量が低いです\n 体重の増減よりも体脂肪率の増減に注目しましょう\n 体脂肪率は筋トレやスクワットなどで減らす事ができます #ダイエット
3	×	○	【時間が無いからダイエットできない！】・・・そんなあなたに魔法をかけてみせましょう。心の準備はいいですか？！”Ready Go!!!”
4	×	×	・週 1 加圧トレーニング\n・週 1 ジムでマシン\n・空いた時間に緩く筋トレ\n・食事制限\n・ご飯食べなそうな日は朝 MCT オイル\n 海外ダイエット薬&サプリ試し中\n#ダイエッターさんと繋がりたい\n#ダイエット仲間募集中\n#アナボリックステロイド

ウドソーシングを利用した．具体的には，“ダイエット” の話題に関して，ランダムに 2,222 件のツイートを収集する．その後，それらのツイートに対し，アンケートを用いて行動促進ツイートであるか，そうでないかのタグ付けを行う．タグ付けの方法は，1 ツイートあたり 3 人から 5 人の被験者が判定する．3 人以上が行動促進と答えたツイートを正例とし，同様に 3 人以上が違うと答えたツイートを負例とする．その結果，正例 829 件，負例 703 件の合計 1,532 件のツイートを正解付きデータとして決定した．ここでダイエットの話題を選択した理由はツイートをしている人が健康についての行動を人に勧める傾向があると考えたためである．

実験方法

学習時の教師データとして，実験データからランダムに抽出した正例 575 件，負例 497 件の合計 1,072 件のツイートを用いる．残りの正例 254 件，負例 206 件の合計 460 件のツイートをテストデータとしてそれぞれの手法の評価に用いる．これらの教師データを用いて 3 つの提案手法を用いて学習を行う，また，10 交差検定を行い，その精度を算出する．尚，ルールベースにおけるテストデータは機械学習の評価に用いるテストデータと同じものを用いて評価を行う．そして，ルールベースと 3 つの学習したモデルを用いてテストデータを判定し，適合率，再現率，F 値を算出，比較する．その結果，最も優れた抽出手法を行動促進ツイートの抽出手法として決定する．SVM，ランダムフォレスト，LSTM の各種パラメータに関しては，第 3 章で説明したものを用いる．

結果と考察

表 3 に各々の手法の結果の一覧を示す．以下に，それぞれの手法により抽出された結果について考察を行い，最後に総評をする．

ルールベース

適合率が低く，再現率が高い結果となった．適合率が低い原因としては，行動促進タイプが文中に含まれているにもかかわらず，ユーザに行動促進をしていないツイートが多いためである．例えば，表 4 の No.3 の場合は「目指そう」という行動促進タイプ 1 が含まれているが，投稿主の意志を示しているツイートである．一方，再現率が高いことより，行動促進タイプにより，ある程度行動促進ツイートが抽出できていることがわかる．

SVM

適合率，再現率とも低い結果となった．適合率が悪い要因は，1 つのツイートに種々の行動促進を行う際に出現する活用形の動詞，助動詞が複数出現する場合が誤って行動促進ツイートになった例が多いためと考えられる．例えば，表 5 の No.3 では，「ましょ」や「みせ」，「かけ」といった行動促進を行う際に出現する単語が頻出している場合，行動促進ツイートとなっている．このように，Word2Vec により作成したツイートの素性がツイートを明確に 2 クラスに分類することができていないため，適合率，再現率がともに低い結果となったと考えられる．

そこで，単純にツイートに出現した単語のベクトルの平均を用いるのではなく，各動詞の活用形に対して重みをつけ，ツイートのベクトルを作成することにより，これらの問題は改善

表 6　ランダムフォレストによる行動促進ツイートの抽出例

No.	正解	システム	ツイート
1	○	○	糖質制限ダイエットをしていて\n よくある失敗談は\n ご飯やパンなどの主食を抜くことで\n 無性にお腹が空いてその結果\n お菓子をドカ食いしてしまうことだ\n これは結果的に主食を普通に\n 食べることよりも太ることに\n なるため、\n 変に制限するくらいなら\n 食べてしまう方がいいです
2	○	×	【ギムネマ葉】\n 主成分はギムネマ酸で砂糖の甘みを感じさせなくなるので、甘いものに対する食欲を減退させたりなど、ダイエットサポートに優れたハーブです。またノンカフェインですので寝る前などお飲みいただけますが、妊娠中、授乳中の方は医師にご相談の上、使用してください。
3	×	○	7 月のライブまでにダイエットしようと思って (まだチケット取れてない w) お昼ご飯をさつまいもにしたけど、あまりにもお腹すいてスタバでスコーン食べてしまった
4	×	×	【ゴマの美容と健康効果】\n〜美容〜\n・シミ・シワ改善\n・ダイエット効果\n・むくみ改善\n・乾燥肌防止\n・老化防止\n・美白効果\n・肌の保護\n・美肌\n〜健康〜\n・骨粗しょう症予防\n・更年期障害予防\n・肝臓機能回復\n・生理不順緩和…

表 7　LSTM による行動促進ツイートの抽出例

No.	正解	システム	ツイート
1	○	○	塩分控えて下さい。は\n 塩控えて下さい。とは違います。\n 化学調味料の塩でなく\n 天然塩であれば摂るべきです。\n 岩塩とかですね！\n これから汗をかく季節です。\n 汗書いた時に塩を摂らないと\n どんどん代謝が落ちます。\n しっかり身体に良い塩で\n ダイエットライフを乗り切りましょう！
2	○	○	【豆知識】\n ダイエットのポイント!\n あなたは 1 日に何回トイレに行く？\n 8 回以下の人は、老廃物や毒素が完全に出ていないので、\n 代謝をよくしたり、利尿作用があるものをとって回数を増やすことが大切。
3	○	×	「時間遺伝子ダイエット」まずは 3 食きちんと食べましょう。 食事の時間は毎日一定に！ 夕食は 21 時までに食べましょう。 この 3 つを実践で、ダイエット成功♪※ 朝、昼、夜の食事量を 3 ： 2 ： 1 の割合にするのがおすすめ
4	×	○	【添加物】\n 添加物はダイエットにも大切な酵素の働きを妨げ、脂肪を溜めやすくします。\n 着色料…コチニール色素、カラメル色素\n 保存料…サッカリン、アスパルテーム\n 人工甘味料…ソルビン、安息香酸\n 主に使われる食品は弁当、ガム、清涼飲料水、菓子類、ハム、ソーセージなど。
5	×	×	あんま言ってたらネタバレしそうなんでごはん食べよう。\n 昨日はラーメン食べたので今日は大根と豆腐にします (唐突なダイエットアピール

されると考えられ，今後の課題である．

ランダムフォレスト

適合率は低いが，再現率は高い結果になった．適合率が低い理由は，素性に用いた品詞の組み合わせが，45 パターンと多く，行動促進ではないツイートも行動促進ツイートと判断したためと考える．また，10 交差検定の結果が悪いことより，教師データが少ないために，教師データに含まれていない行動促進をしている単語がテストデータに含まれていることが予測される．例えば，表 6 の No.2 や No.3 では，ツイート内に「サポートに優れた」や「食べてしまっ」などの複数の表現が教師データに出現しておらず，一部の表現のみを用いて判別を行っている状態となっている．そのため，1 つでも行動促進ツイートに出現している表現が存在すると行動促進ツイートとして判定されてしまう．その結果，再現率が高く，適合率が低いという結果となった．教師データに含まれていない表現がテストデータに出現する場合が多数存在していたため，今後の課題として，教師データを増やすことが挙げられる．これにより，教師データの表現が増え，テストデータにおいて判別のための材料が増加するため，適合率が向上すると見込まれる．

LSTM

適合率，再現率，F 値ともに高い値となった．これによりツイートに出現する単語の順序を考慮することにより，適合率，再現率が高くなることがわかる．表 7 の No.1 のように，「控えてください」や「乗り切りましょう」といった行動促進フレーズが含まれているものに関しては正しく評価を行えている．表 7 の No.2 のように，「回数を増やすことが大切。」というツイートのように行動促進フレーズが存在していなくても抽出を行うことができている．これは，教師データにおいて，同様の表現が含まれていたためであると考えられる．表 7 の No.3 は実際には「食べましょう」といった行動促進フレーズが含まれているが，誤って行動促進ツイートではないと判定されている．表 7 の No.4 は実際には行動促進フレーズが含まれていないが，誤って行動促進ツイートとして，判定されている．これは,,「脂肪をためやすくします」という表現が行動促進ツイート内の「〜ますよ」といった表現と類似していることから，誤って抽出されたと考えられる．このことから，表 7 の No.3，No.4 については，学習時の教師データが少ないため，対応できていないと考えられる．しかしながら，教師データが少ないにもかかわらず，適合率，再現率がともに高い値となっていることより，文末のみならず，単語の流れを考慮することが必要であることがわかった．

総評

テストデータの判定におけるそれぞれの手法の適合率，再現率，F 値，10 交差検定の結果を表 3 に示す．この結果より，LSTM

表 8　LSTM のみ正しく判定された例

No.	正解	ルールベース	SVM	ランダムフォレスト	LSTM	ツイート
1	×	○	○	○	×	153 センチで 45 キロとかばりデブやん笑って言われてんけどそーなんかな？\n 確かに前と比べると 4 キロも増えてたけど、デブかどうか自分じゃ分からん... やっぱダイエットした方がいいんかな（´ -‘）.ïjąo
2	×	○	○	○	×	【時間が無いからダイエットできない！】・・・そんなあなたに魔法をかけてみせましょう。心の準備はいいですか？！”Ready Go!!!”

の適合率，F 値が他の手法と比べて，最も高く，抽出結果が最も良いことがわかる．これは，LSTM はツイート内の単語の出現の順序を考慮することにより，ツイートの行動促進フレーズが投稿主の意志を示しているのかを判断できているためであると考えられる．ルールベースに関しては，行動促進を行っている表現が含まれていれば抽出を行っているため，他の要素に左右されずに抽出できる．その為，再現率が高くなっていると考えられる．しかしながら，行動促進フレーズが意志を示している場合も抽出してしまっているため，適合率が低下している．SVM，ランダムフォレストに関しては，文末の表現のみに着目し，判定を行っているため，出現している行動促進フレーズが意志であるかの判定が正しく行えず，F 値が低くなっていると考えられる．

また，表 8 のツイートのように，他の手法だと行動促進として抽出され，誤判定されているが，LSTM の場合のみ正しく行動促進ではないと判定できているツイートがあった．これは表 8 の No.1 のツイートの「いいんかな？」や表 8 の No.2 のツイートの「いいですか？」といった表現がツイート内の行動促進フレーズが行動促進を示しているのか，投稿主の意志を示しているのかの判定に影響を与えたためであると考えられる．今後はさらに学習データを増加させ，様々なパターンを学習させる必要があると考えられる．

以上の結果より，LSTM を用いた手法が行動促進ツイートの抽出に最も適していることがわかった．また, データ量が少ないながらも，LSTM においては，高い適合率，再現率となっていることから，行動促進ツイートの抽出においては，教師データが少ない状態であったとしても，学習が行える可能性があるといえる．そこで，今後データ量を変化させて実験を行い，検証する予定である．

5. まとめと今後の課題

本論文では，Twitter 上に存在する行動促進を行っているツイートの 4 つの抽出手法を提案し，それぞれの手法に関して比較実験を行い，最も抽出に適した手法が LSTM であることを示した．具体的には，行動促進ツイートの抽出手法として，ルールベース，SVM，ランダムフォレスト，LSTM の 4 つの手法を提案した．その後，実験データを用いて，学習を行い，それぞれの手法を用いて行動促進ツイートの抽出の精度の評価を行い，最も優れた抽出手法が LSTM であることを確認した．

今後の課題としては，今回，1 つの話題のみを用いて実験を行っているが，他の話題を用いて実験を行い，抽出手法が他の話題にも適用できるのかを確認することである．また，実験に用いたデータが少ないため，精度の評価が正しく行えていない可能性があるため，データ量を増やして実験を行う予定である．さらに，現在は「ダイエット」のドメインのみであるが，今後複数のドメインによる実験を行いたい．

謝　　辞

本論文の一部は JSPS 科研費 17K00430，16K07973，18H03342 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] 見塚 圭一, 鈴木 優, 灘本 明代,”行動促進の根拠を含むツイートの抽出手法,” 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), F5-2, 7pages, 2018.

[2] 矢野 裕司, 横井 健, 橋山 智訓, “行動を表す単語に着目した Twitter からの行動抽出”. 情報科学技術フォーラム講演論文集, 12(4), pp157-164, 2013.

[3] 矢野 裕司, 横井 健, 橋山 智訓, “Tweet からの行動と感情の抽出法とその分析”. 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集 30(0), pp798-803, 2014.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] M. Miyabe, A. Nadamoto, and E. Aramaki. How do rumors spread during a crisis?: Analysis of rumor expansion and disaffirmation on twitter after 3.11 in japan. International Journal of Web Information Systems, 10:394-412, 2014.

[9] 梅島 彩奈, 宮部 真衣, 荒牧 英治, 灘本 明代,”マイクロブログにおける流言マーカー自動抽出のための特徴分析,” 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), F3-2, 8 pages, 2012.