

甲南大学大学院
自然科学研究科
知能情報学専攻
修士論文 No. 195

Twitter 上の行動促進ツイートの抽出手法

A Method for Extracting Behavioral Facilitation Tweet for Normal Situation and Disaster Situation

2019年3月
見塚 圭一

甲南大学大学院 自然科学研究科

要旨

近年、SNSが普及している。そのことから、SNSを用いて情報を集めるユーザが増加している。そしてTwitter等SNS上では様々な情報が交換されている。その情報の中には、注意喚起やアドバイス、意見などがある。その中でも、注意喚起やアドバイスといった情報は人々の行動に影響を与えることから、重要であると考えられる。本研究では、SNSのうちTwitterを対象とし、人に行動を促す情報を含むツイートを行動促進ツイートと呼ぶ。ユーザがTwitterを利用する際には大量のツイートを眺めることとなり、その中の行動促進ツイートは流し読みをしてしまい、内容を深く考えずに認識され、拡散されてしまう恐れがある。また、災害時にはユーザは正常な判断ができず、行動促進ツイートの内容を深く考えずに拡散することや実際に行動を起こすことがあると考えられる。そのため、デマを含む行動促進ツイートが拡散されてしまう場合がある。そこで、平常時、災害時の行動促進ツイートを抽出し、話題ごとの行動促進ツイートの分析、ユーザへの行動促進ツイートの提示を行う必要があると考えた。本研究では、Twitter上の行動促進ツイートの抽出手法の提案を行う。さらに、行動促進ツイートの分析も行う。具体的には、平常時、災害時、それぞれのツイートからルールベース、SVM、ランダムフォレスト、LSTMの4つの手法を用いて行動促進ツイートの抽出を行い、その抽出精度を比較し、適した抽出手法を提案する。さらに、提案手法の有用性を示すための実験を行う。そして、平常時、災害時の各話題における行動促進ツイートの分析を行う。

Summary

Social Network Services(SNS) becomes popular, especially Twitter is one of the most popular media. Many users get some information from Twitter. There is some information which includes attention, advice or opinion on twitter. The attention and the advice information are important, because the information encourages people to adapt behavior. In this study, we target on only Twitter, and we designate such tweets, which compel users to adopt a behavior, as *behavioral facilitation tweets*.

When users use Twitter, they have to read a lot of tweets. So they can not read behavioral facilitation tweet carefully. Especially, on the disaster situation, they can not read behavioral facilitation tweet more carefully. However, behavioral facilitation tweets include many rumors. Then people spread many rumors on Twitter because it is easy to spread the information.

In this research, we proposed the methods that automatically extracts behavioral facilitation tweets on the normal situation and disaster situation. Specifically, we propose four types of methods which extract behavioral facilitation tweets automatically from numerous tweets: rule-based, support vector machine (SVM), Random Forest, and long short-term memory (LSTM). Moreover, we analyze the behavioral facilitation tweet. We use topics of normal situation and disaster situation to measure the benefits of our proposed methods, comparing rule-based, SVM, Random Forest and LSTM-based methods to extract behavioral facilitation tweets.

目次

1	はじめに	1
2	関連研究	2
3	行動促進ツイートの定義	3
4	ルールベースを用いた行動促進ツイートの抽出	4
5	SVMを用いた行動促進ツイートの抽出	6
6	ランダムフォレストを用いた抽出手法	7
7	LSTMを用いた行動促進ツイートの抽出	9
8	評価実験	10
8.1	実験1	10
8.1.1	実験データ	10
8.1.2	実験方法	11
8.1.3	結果と考察	11
8.2	実験2	15
8.2.1	実験データ	16
8.2.2	実験方法	16
8.2.3	結果と考察	16
9	まとめと今後の課題	26

目 次

1	行動促進ツイートの分類	4
2	LSTM のネットワークの構造	9
3	実験 1 における 2 つの話題の実験結果	14
4	実験 2 における 3 つの話題の実験結果	18
5	実験 2 における地震の話題のツイートのプロット結果	19

表目次

1	明示的行動促進ツイートの例	4
2	暗示的行動促進ツイートの例	5
3	ルールベースに使用する行動促進フレーズのタイプ一覧	5
4	タイプ1を満たすツイートの例	6
5	タイプ2を満たすツイートの例	6
6	タイプ3を満たすツイートの例	7
7	タイプ4を満たすツイートの例	7
8	ランダムフォレストにおける素性の組み合わせ一覧	8
9	ランダムフォレストに用いる素性の例	8
10	実験1における話題ごとの実験データの数	11
11	実験1における話題ごとのSVMのパラメータ	11
12	実験1におけるダイエットの話題の抽出結果	12
13	実験1における熱中症の抽出結果	12
14	実験1におけるダイエットに関する提案手法を用いた行動促進ツイートの抽出結果	13
15	実験1における2つの話題のLSTMを用いた抽出結果	15
16	実験2における話題ごとのSVMのパラメータ	17
17	実験2における話題ごとの実験データの数	17
18	実験2における地震の抽出結果	18
19	実験2におけるダイエットの抽出結果	20
20	実験2における熱中症の抽出結果	20
21	実験2におけるルールベースによる誤判定の例	21
22	実験2におけるSVMの誤抽出の例	22
23	実験2におけるLSTMの誤抽出の例	23
24	実験2における地震に関する提案手法を用いた行動促進ツイートの抽出結果	24
25	各話題における行動促進ツイート	25

1 はじめに

近年、SNS が普及し、それに伴い SNS を用いて情報を収集するユーザも増加している。そして SNS 上では様々な情報が交換されている。SNS の中で特に Twitter は気軽にツイートできることから、様々な情報が交換されている。そこで本研究では、SNS のうち Twitter を対象とする。Twitter には、注意喚起やアドバイス、意見など様々な情報がある。その中でも、注意喚起やアドバイスといった情報は人々の行動に影響を与えている。例えば、「ダイエットの時には、水をたくさん飲みましょう。」や「地震が発生した時は、すぐに机の下に隠れよう!」といったツイートが挙げられる。このような人に行動を促すツイートを本研究では、行動促進ツイートと呼ぶ。

ユーザが Twitter を利用する際には大量のツイートを眺めることとなり、その中の行動促進ツイートは流し読みをされてしまい、内容を深く考えずに認識されてしまう。そのため、内容が誤っていたとしてもツイート内の行動を起こしてしまうと考えられる。

一方、今日様々な災害が多数発生している。災害時に Twitter は情報を取得する手段の 1 つとなっている。そのため、災害時において、冷静な判断が行えない状態でユーザが内容を深く考えずに行動促進ツイートを拡散したり、ユーザが内容を実際に行動に移す可能性がある。そのため、誤った情報が行動促進ツイート内に出現していても拡散したり、行動を起こすなどし、ユーザが不利益を被る場合がある。また、東日本大震災時のツイートを分析した研究 [1] によると、デマツイート 7177 件中、行動促進ツイートは 3773 件となっており、デマツイート中の行動促進ツイートの割合は 53% となっている。このことから、災害時に流れる行動促進ツイート内にデマ情報が現れる可能性が高いといえる。災害時の冷静ではない状態において、行動促進ツイートの内容を吟味して実行すべきかを判断し、行動することは困難であると考えられる。

これらのことから、行動促進ツイートを抽出し、ユーザに提示することはデマ情報への注意喚起の 1 つとして重要であることがわかる。そこで、行動促進ツイートを抽出してアラートを出し、ユーザに内容を考える機会を提供することが必要であると考えた。本研究では行動促進ツイートを抽出する手法の提案を行う。これにより、ユーザが平常時のみならず、災害時の正確な判断ができない状況において、誤った情報を信じることにより、不利益を被る事を減少できると考えられる。

具体的には、ルールベースの手法と Support Vector Machine (SVM), ランダムフォレスト (RF), Long Short-term Memory (LSTM) の 3 つの機械学習の手法を用いる。この 4 つの手法を用いた行動促進ツイートの抽出の比較実験を行い、行動促進ツイートの抽出に最も適した手法がどの手法であるのかを検証する。まず、平常時の話題である「ダイエット」「熱中症」の話題に対して比較実験を行い、平常時の話題において適した行動促進ツイートの抽出手法がどの手法であるのかを検証する。次に、災害時の話題である「地震」の話題に対しても、平常時の話題にて使用した手法を改善をしたものを用いて比較実験を行い、災害時の話題において適した行動促進ツイートの抽出手法の比較検証を行う。さらに、ダイエットや熱中症といった平常時の話題に対して同様の手法を用いて行動促進ツイートの抽出を行い、災害時と平常時の行動促進ツイートの比較、分析を行う。

以下、2 章で関連研究について述べる。3 章で提案手法の概要を述べる。そして、4 章でルールベースを用いた抽出方法について述べ。5 章 SVM を用いた抽出手法、6 章でランダムフォレストを用いた抽出手法、7 章で LSTM を用いた抽出手法について述べる。8 章で平常時、災害時それぞれの状況において、提案手法を用いた行動促進ツイートの抽出手法の比較実験を行い、

その考察について述べる。最後に9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。Jinら[2]の研究では、Twitterなどのマイクロブログ上から様々な種類のイベントの抽出、解析を行う手法を提案している。Phuvipadawatら[3]は、最新のニュースに関連するツイートを収集し、グループ分けを行い、各グループにランク付けを行い、それを追いかける手法を提案した。さらに、提案手法を用いて、“Hoststream”というシステムを開発している。古川ら[4]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っている。その結果、Twitter上のみ存在しているユーザーの経験に基づく犯罪情報があることを確認し、それらが有益な情報となることを示している。矢野ら[5][6]の研究では、ユーザー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。Ritterら[7]は、“TwiCent”と呼ばれるTwitterを用いたオープンドメインのイベントの抽出と分類を行うシステムを提案している。彼らの手法では、情報を追加するためにイベントに関連するツイートを抽出している。Sellamら[8]はツイッター上からイベントに関する定量的なデータを含むツイートを抽出する手法を提案している。彼らは“Raimond”と呼ばれる提案手法を用いたシステムを作成した。これらの研究は、Twitter上からニュースやイベント、犯罪情報といった特定の話題に対する有益な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、様々な話題を対象にしてTwitter上から重要な情報を含むツイートの抽出を行うおうとしているという点で異なる。

Popescuら[9][10]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。Sankaranarayananら[11]は、ニュースと一致するツイートを抽出する手法を提案している。さらに、Twitterを用いたニュース処理のシステムである、“TwitterStand”というシステムを開発している。彼らの提案する手法では、ツイートの中に意見、感情、または解説のどれかが入っているものが抽出される。そのため、これらの手法では、ツイートにはどのようなものが入っているのかを考慮していない。一方、本研究では、ツイートの中に行動促進を行っているツイートを抽出する手法を提案している。

さらに、Twitterのユーザーが増えるにつれて、Twitter上から噂を抽出する研究が数多くなされるようになった。Miyabeら[12]は、マイクロブログ上の噂の広がりについて分析を行い、噂の拡散を防ぐために噂を収集するためのシステムの評価手法を提案している。Tripathyら[13]は、ソーシャルネットワーク上の噂の広がりに対抗するために噂のようなものの分析、評価を行っている。また、噂の拡散防止に、“anti-rumor”を使用する手法を提案している。Qazvinianら[14]は、マイクロブログ上での噂の検出の問題に取り組み、噂を正しく認識するための3つの特徴の有効性を調査した。Shaoら[15]は、“Hoaxy”というシステムを提案した。これはインターネット上の誤った情報とそれに関連する事実を収集、発見、分析を行うシステムである。彼らの手法では、Twitter上のすべてのタイプの噂情報を抽出しようとしている。一方、本研究では、まず行動促進を含むツイートを抽出し、その中の情報において噂情報の抽出を行う。

さらに、Twitter上の投稿を利用して災害時の有益な情報を抽出しようとしている研究がなされている。Cameronら[16]は、ESA-AWTMと呼ばれるTwitter上に存在している危機管理のための情報を検出、評価、要約、報告を行うシステムを構築した。このシステムにより、緊急時に一般の人による重要な情報を抽出することができる。Rudraら[17]は、Twitterから災害時

の状況情報を抽出し、要約する手法を提案している。榊ら [18] の研究では、災害時のツイートの位置情報を推定し、災害支援活動や避難活動に活用するために、GPS を用いる手法や投稿者のプロフィールを用いる手法などの計 4 つの手法を組み合わせてツイートの位置情報の推定を行う手法を提案している。坂巻ら [19] は、災害時の Twitter 上に存在している何が起きているのかという情報や、被災地で必要とされている物資などの有益な情報を時系列と単語の出現頻度を組み合わせることにより、抽出する手法の提案を行っている。これらの研究は、Twitter 上から災害時の重要な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、災害時の情報の中でも、重要であると考えられる他人に行動を促進するツイートに着目し研究を行っているという点で異なる。

そして、Twitter 上でのユーザの行動を分析、予測しようという研究も行われている。Silva ら [20] は、ソーシャルネットワークにおける個人の行動情報が関係にどのように影響を与えるのかを分析している。Xu ら [21] は Twitter におけるツイートの投稿行動に関して分析を行っている。彼らは、ユーザの投稿行動が主に 3 つの要因から行われていると特定した。それは、ニュース速報による情報、友人による投稿、そしてユーザの興味である。これらの研究では、個人の行動が周りに及ぼす影響の分析を行っている。一方、本研究では、Twitter ユーザに行動促進を含むツイートが行動を及ぼすと考え、抽出を行っている。Mogadala ら [22] は、Twitter ユーザの時間によって移りゆく気分をツイートから予測する手法を提案している。Yamamoto ら [23] は、ユーザの行動を抽出し、ユーザの将来の行動を予測する手法を提案している。彼らは、ユーザの行動を抽出し、ユーザの将来の行動を予測している。一方、本研究では他人への行動促進を目的としたツイートの抽出手法を提案する。

3 行動促進ツイートの定義

行動促進ツイートは大きく分けて 2 つのタイプに分類される。図 1 に本研究にて分類した行動促進ツイートの分類を示す。1 つは明示的行動促進ツイート、もう一方は暗示的行動促進ツイートである。

明示的行動促進ツイートとは、ツイート内に明示的に行動を促進している部分を含むツイートのことである。例えば、「地震のときは机の下に隠れましょう。」というものが挙げられる。これは、ツイート内に「隠れましょう」という明らかに行動を促進している部分を含んでいるため、明示的行動促進ツイートである。明示的行動促進ツイートの例を表 1 に示す。

一方、暗示的行動促進ツイートとは、ツイート内に行動を促進している部分を含まないが、他人に行動を促しているツイートのことである。例えば、「地震のときは、机の下に隠れないと上から物が落ちてきて危険だ。」というツイートが挙げられる、このツイートは、ツイート内には行動を促している部分は存在していないが、ユーザが閲覧した際には、「机の下に隠れる」という行動を起こすと考えられる。これは、ツイート内の行動を起こさなければ不利益を被ると判断されたためである。このように行動を促進している部分は存在していないが、暗示的に行動を促進しているツイートを暗示的行動促進ツイートと定義する。暗示的行動促進ツイートの例を表 2 に示す。

本研究では、明示的行動促進ツイートのみを対象とする。本研究ではこの明示的行動促進ツイートを単に行動促進ツイートと呼ぶ。さらに、図 1 に示すように明示的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、「地震で停電になると困るので、陽の光あるうちに冷水でもシャワー浴びて

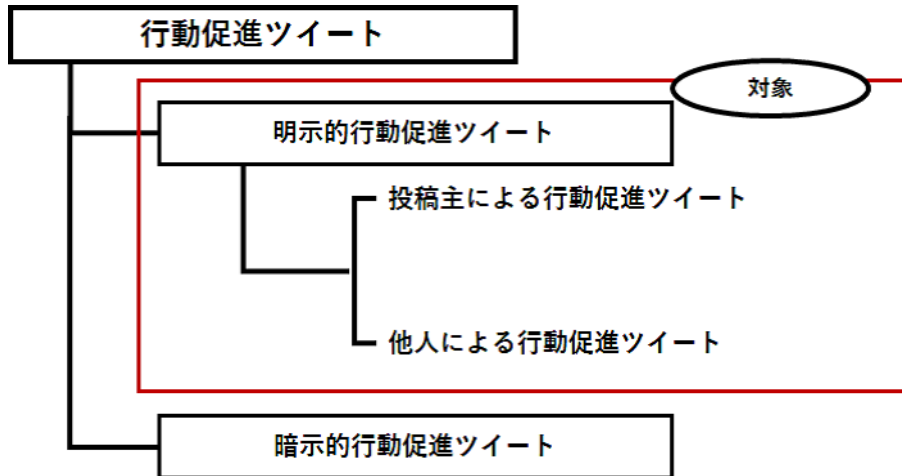


図 1: 行動促進ツイートの分類

表 1: 明示的行動促進ツイートの例

ツイート
【効果的な 25 のダイエット法】 その 2・1 度に食べる量を減らすデザートを食べるときは 3 口ルール。3 口食べたなら 1 度フォークを <u>置こう</u> 。
SOS を発信する時は、場所や目印になる建物、住所など具体的に。できれば状況説明のための写真などを添えて。そして、SOS が解決したらツイートを削除 <u>しましょう</u> ！#北海道#地震
怖いね…地震の怖さを思い知ったからマジで避難防災グッズ用意しといた方が <u>いいよ</u>

※行動促進は下線で示している

おきましょう。」は投稿主が知っている情報を伝えようとする行動促進の内容であるため、投稿主による行動促進ツイートといえる。また、「ニュースで災害時には、水がにごる可能性があるので、風呂に水を溜めておきましょうって言ってるのを見たよ！」は、他人の発言を引用しているツイートやリツイートを利用して行動促進の内容を伝えるツイートであり、投稿主ではなく、他人による行動促進の内容を引用したツイートである。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。

4 ルールベースを用いた行動促進ツイートの抽出

ダイエットや熱中症、地震の話題を用いて収集した多数のツイートを目視で分析した結果より、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は文の末尾にその特徴が現れていることがわかった。そこで本研究では、このようにツイート内の文末で行動促進を行っている部分を「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する。ツイートの分析結果より、行動促進ツイー

表 2: 暗示的行動促進ツイートの例

ツイート
ダイエットの時の断食は、栄養を求めて必要な脂肪を燃焼させるため、健康に悪影響を与えます。
地震により家が被害を受けた時は、被害状況の写真を撮っておく。そうしなければ、証拠なしと判断され、保険適用外になる。
地震の時には備えがあるのと無いのでは、生存する可能性に天と地ほどの差があると言われている。

表 3: ルールベースに使用する行動促進フレーズのタイプ一覧

タイプ	内容
タイプ 1	ツイート内に動詞、助動詞の「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。

トに多く出現している傾向のあった 4 つの行動促進フレーズのタイプを提案する。表 3 に使用する 4 つのタイプを提示する。そして、これらいずれかの行動促進フレーズのタイプに 1 つでも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。以下 4 つの行動促進フレーズのタイプについて説明する。

[タイプ 1:] ツイート内に動詞、助動詞の「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 [24] 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、「地震の時には安全な場所に避難しよう。」といったツイートや「大地震の発生後には余震に備えよう。」といったツイートの「避難しよう」や「備えよう」といった表現である。表 4 にタイプ 1 の行動促進フレーズを含むツイートの例を示す。

[タイプ 2:] ツイート内に動詞の「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、「閉じ込められないように震災時にはドアをすぐに開ける!」といったツイートや「地震が起きたときはすぐに外には出ずに室内で待機してください!」といったツイートである。表 5 にタイプ 2 の行動促進フレーズを含むツイートの例を示す。

[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。

「逃げるといい」や「逃げた方がいい」といった表現は他人に物事を勧めたり、提案を行う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「緊急時に備えて懐中電灯を用意すると良い。」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「災害時には冷静になった方が良い。」というように「～方が良い」というタイプである。表 6 にタイプ 3 の行動促進フレーズを含むツイートの例を、表 7 にタイプ 4 の行動促進フレーズを含むツイートの例を示す。

表 4: タイプ1を満たすツイートの例

ツイート
わざわざ回数を数える必要はありませんが、30回以上嘔んでから食べる事だけ意識 <u>しましょう</u> 。ご飯の場合、よく嘔んで食べると甘く感じられるようになります。 <u>#ダイエット</u>
エナジードリンクはご存知の通りカフェインめっちゃ入ってますし飲み過ぎると腎臓にもよくないしなので依存症にはお気をつけて。持ち歩きなら <u>麦茶にしましょう!</u> ←麦茶は体を冷やす効果がありますし緑茶と違って利尿作用がないので熱中症対策にも <u>ばっちり</u> です!
土砂災害危険箇所は、 <u>がけ崩れ</u> が起きたときに人家に被害が想定されている場所です。この場所は避難対象地区に指定されており、東海地震の警戒宣言が発令された場合には、 <u>避難勧告</u> が出ます。これらの施設を前もって把握して <u>おきましょう</u> 。
※行動促進フレーズは下線で示している

表 5: タイプ2を満たすツイートの例

ツイート
のどが渇いたと思った時には熱中症の初期症状です。そうなる前に <u>まめに水分補給を心掛けて下さい</u> 。
タンパク質は、 <u>ダイエットの敵</u> じゃないですよ。しっかり食べて <u>下さい</u> 。気を付けるのは、糖質です。お米、パン、うどん、パスタなど。
遅ればせながら北海道地震関係のリストを作りました。現在70件。随時追加していきます。とりあえず官公庁・放送局・公共サービス関係優先です。 <u>情報収集にお役立てください</u> 。
※行動促進フレーズは下線で示している

5 SVMを用いた行動促進ツイートの抽出

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クラス分類のタスクであるため、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LIBSVM¹を使用する。SVMのカーネルはRBFカーネルを用いる。各種ハイパーパラメータはLIBSVMのツールを用いてグリッドサーチを行うことにより決定する。実際の話題ごとのハイパーパラメータの値は実験の章にて記述する。

本研究では、2種類のパターンの素性を用いて学習を行う。以下、それぞれのパターンの素性の説明を行う。

1. パターン1

パターン1は、ツイート中の全品詞の単語を用いて学習を行う。これは、特定の品詞に絞らず全品詞を使用することによって、出現した単語により行動促進ツイートの特徴を捉えることができる考えたためである。

¹<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 6: タイプ 3 を満たすツイートの例

ツイート
ダイエットの時には食事前に飴玉を <u>食べると良い</u> 血糖値があがり、脳の満腹中枢に働きかけ食べ過ぎないようにしてくれます。
熱中症の様な症状でしたら首の横を <u>冷やすと良い</u> ですよ！
ブレーカーの場所がわからないとか、届かないというなら、あまり無理をせず、各家電のコンセントを抜いて <u>おく</u> と <u>良い</u> ですよ！

※行動促進フレーズは下線で示している

表 7: タイプ 4 を満たすツイートの例

ツイート
ビールは毎日飲んだ方が <u>良い</u> 。500ml 缶を毎日飲むと、ダイエットに良い。ビールには、カルシウム・葉酸・ビタミンを多く含む。代謝を活発にして、循環器系を機能させる。
「暑いせいで疲労を感じる」のであれば熱中症に近づいている危険があるので水分を取った <u>ほうが</u> いいです。熱けいれんまでいくと動けなくなり自力ではどうしようもなくなるのでこの時期も油断は禁物ですよ。
地震によりガス管などの破損の可能性がある場合、電気が復旧したときに火災が発生することがありますので、ブレーカーは落としておいた方が <u>いい</u> です。

※行動促進フレーズは下線で示している

2. パターン 2

パターン 2 は、ツイートに対し形態素解析を行い得られた品詞のうち、動詞、助動詞の単語のみを対象とする。これは、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、これら 2 つの品詞のみを用いることにより、判定が可能になると考えたためである。

形態素解析器として Mecab² を使用する。パターン 1 とパターン 2 において使用する品詞の中で、活用形を持つ品詞は、活用形が行動促進であるかどうかの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の活用形で使用する。

これらのツイート内の各語を Word2Vec を用いて分散表現ベクトルに変換し、ツイート中の各単語の分散表現ベクトルの平均値を算出し、各ツイートの素性とする。

6 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。ランダムフォレストを用いる理由としては、自動でツイート内に含まれるフレーズを元に分類のルールを作成して行動促進ツイートの抽出を行うことができると考えたためである。学習には

²<http://taku910.github.io/mecab/>

表 8: ランダムフォレストにおける素性の組み合わせ一覧

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

各記号意味一覧：n=名詞，v=動詞，adj=形容詞，p=助詞，p_end=終助詞，aux=助動詞

表 9: ランダムフォレストに用いる素性の例

具体例	品詞	具体例	品詞
過ごしましょう	動詞+助動詞+助動詞	方が良い	名詞+助詞+形容詞
つけてください	動詞+助詞+動詞	なりたくない	動詞+助動詞+助動詞
必須ですよ	名詞+助詞+終助詞	乾いて欲しい	動詞+助詞+形容詞

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。

使用する素性は，表 8 における品詞の組み合わせとなっているツイート中の特定の部分を用いてツイートを特徴ベクトルに変換したものである。表 8 における品詞の組み合わせは，ツイートを分析した結果，文末付近に出現している品詞の組み合わせには規則性があったためそれらの組み合わせを抽出し，選択を行った。行動促進ツイートの特徴となる部分は文末であると考えたため，文末付近にある部分の品詞の組み合わせに着目した。ツイートの特徴ベクトルには one-hot ベクトルを用いる。ツイートに表 8 の品詞の組み合わせを持つ単語の部分が含まれていれば 1，含まれていなければ 0 とした。この素性を用いることにより，行動促進フレーズのタイプ以外の行動促進を行っている表現を抽出することができ，抽出精度が向上すると考えた。

例えば、「控えましょう」であれば、「控え（動詞）」＋「ましよ（助動詞）」＋「う（助動詞）」の組み合わせとなっているおり，表 8 の No.26 に当たるため，この「控えましょう」という部分を素性に加える。そして，ツイートの素性ベクトルの「控えましょう」の部分は 1 となる。表 9 に素性に用いる品詞の組み合わせにより得られた一部の例を示す。

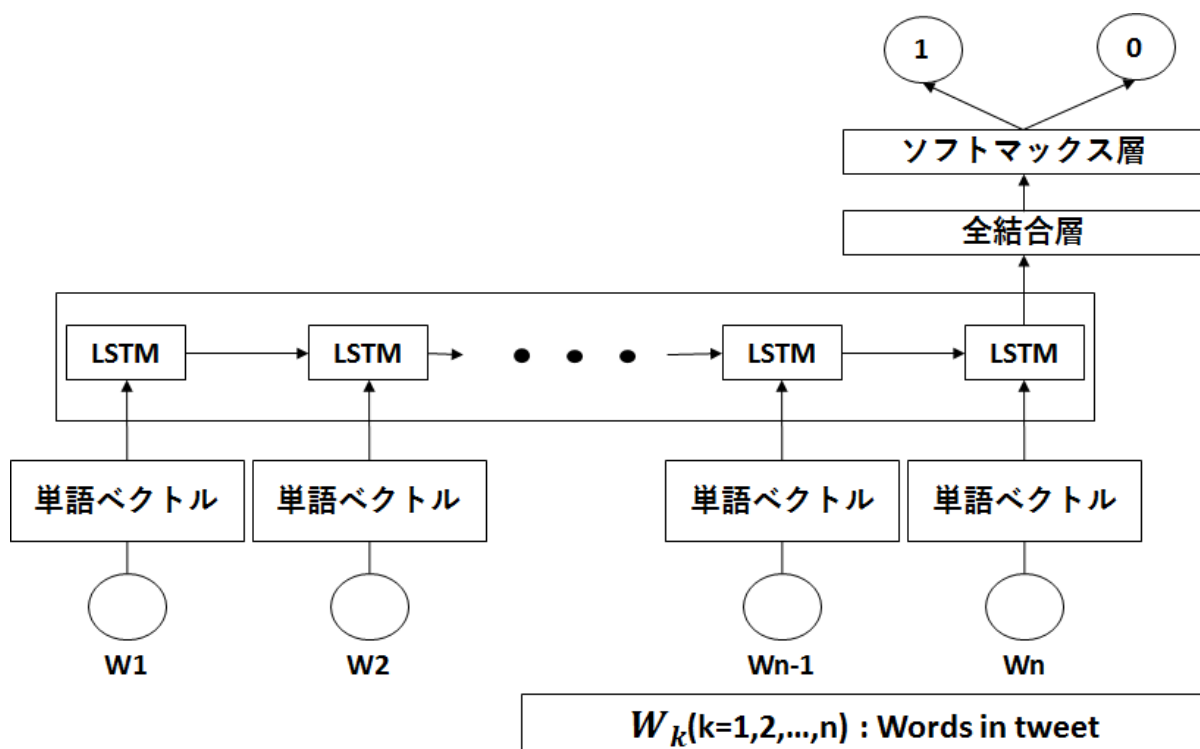


図 2: LSTM のネットワークの構造

7 LSTM を用いた行動促進ツイートの抽出

5 章にて Word2Vec を用いてツイートの分散表現ベクトルを獲得し、学習を行った。しかしながら、SVM では、出現単語の分散表現ベクトルの平均値を使用しているため、ツイート中の単語の出現順序を考慮することができない。そこで、単語の順序を考慮することのできる LSTM を用いて行動促進ツイートの抽出を行う。LSTM を用いる理由は、行動促進ツイートの判別時に文の単語の順序を考慮することにより、他の手法では考慮することのできない表現や、ルールベースでは誤って判定されていた行動促進フレーズが含まれているが、行動促進ツイートではないツイートを正しく判定することができると考えたためである。LSTM の実装には、Python の機械学習ライブラリである Chainer³ を用いる。

素性は、ツイート内に出現する全単語を Word2Vec を用いて分散表現ベクトルに変換したものを使用する。学習済み Word2Vec のモデルとして SVM に使用したものと同一ものを用いる。使用する単語は、ツイートに対して形態素解析を行い、得られた全品詞の単語を用いる。全品詞を用いる理由としては、文の単語の流れを考慮することにより、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。また、使用したモデルに存在していない未知語に関しては、今回は 0 ベクトルを使用した前処理として使用するデータ内から、url、Twitter のユーザ名を取り除く処理を行う。

使用する LSTM のネットワークの構造を図 2 に示す。LSTM の各種ハイパーパラメータは、隠れ層の数は 1、ユニット数は 300、バッチサイズは 500、エポック数は 300、学習率は 0.001、

³<https://chainer.org/>

옵ティマイザーは Adam を利用する. LSTM モデルは 1 層で構成され, LSTM のユニットの数はツイートの単語数とする. 300 次元の特徴ベクトルは, LSTM のユニットへの入力であり, LSTM のユニットの出力は次のユニットの入力となっている. 最後の LSTM のユニットの出力は全結合層への入力となっている. この全結合層は, 2 つの値を出力し, その値を用いて, ソフトマックス層を用いて行動促進部分を含むか否かを判別する.

8 評価実験

提案したどの手法が適している手法であるのかを確認するための以下の 2 つの実験を行った.

1. 実験 1

平常時の話題を用いて提案した行動促進ツイートの抽出手法を用いて行動促進ツイートの抽出を行い, どの手法が適しているのかを確認する. 平常時の話題として, 「ダイエット」, 「熱中症」の 2 つの話題を使用する.

2. 実験 2

災害時の行動促進ツイートから, 提案手法を用いて行動促進ツイートの抽出を行い, 災害時の行動促進ツイートの抽出に適した手法を確認する. 災害時の話題として, 「地震」の話題を使用する. さらに, 平常時の行動促進ツイートと災害時の行動促進ツイートの比較, 分析し, 災害時の行動促進ツイートの特徴抽出を行う.

8.1 実験 1

実験 1 では, まず, 平常時の話題を用いて提案した行動促進ツイートの抽出手法を用いて行動促進ツイートの抽出を行い, どの手法が適しているのかを確認する. 平常時において, Twitter 上には様々な情報が散見され, その中には行動を提案していたり, 促しているものが存在している. そこで, まずは多様な情報の中から行動促進ツイートの抽出を行うのに適した手法がどの手法であるのかの検証を行う. 平常時の話題として, 「ダイエット」, 「熱中症」の 2 つの話題を用いる. この 2 つの話題を選択した理由は健康に関する話題では, 人に有益な情報を伝えたり, 経験したことを人に勧めるなど, 他人に影響を与えるようなツイートが多いと考えたためである.

8.1.1 実験データ

SVM, ランダムフォレスト, LSTM の学習, 評価時に使用する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ウドソーシングを利用した. 具体的には, ダイエット, 熱中症の 2 つの話題に関して, それぞれ「ダイエット」, 「熱中症」のクエリを用いて, ランダムに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 人の被験者に判定をしてもらい, 3 人以上の判定がどちらかに偏った場合にツイートを正例または負例に分類し, タグ付けを行った. そして, 表 10 に示すツイート数を正解付きデータとして決定した. これらのデータを実験データとして学習と評価を行う.

表 10: 実験 1 における話題ごとの実験データの数

話題	教師データ			テストデータ			総数		
	正例	負例	合計	正例	負例	合計	正例	負例	合計
ダイエット	1,696	3,195	4,891	416	807	1,223	2,112	4,002	6,114
熱中症	1,850	2,046	3,896	465	509	974	2,315	2,555	4,870

表 11: 実験 1 における話題ごとの SVM のパラメータ

Topic	Cost Parameter C	Kernel Parameter g
ダイエット	8.0	0.3125
熱中症	8.0	0.03125

8.1.2 実験方法

評価時のテストデータとして、話題ごとにランダムに実験データから 2 割のツイートを抽出する。それらをテストデータとして各手法の評価に利用する。残りのツイートを教師データとして利用する。表 10 にそれぞれの話題の教師データ数、テストデータ数を示す。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、それぞれの話題にて、ルールベースと機械学習の 3 つの手法を用いて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。ランダムフォレスト、LSTM の各種パラメータに関しては、第 6 章、第 7 章で説明したものをを用いる。実験 1 にて SVM の学習に用いる話題ごとの各種ハイパーパラメータは表 11 に示す。ここでは、パターン 2 の素性を用いて抽出を行う。SVM と LSTM に用いる 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²を利用する。

8.1.3 結果と考察

図 3, 表 12, 表 13 に各話題における 4 つの提案手法の適合率、再現率、F 値、10 交差検定の結果を示す。まず、各手法の考察を行う。

ルールベース

ルールベースにおいては、各話題において適合率が低く、再現率が高い結果となった。これは、ルールベースにおいて、行動促進フレーズがツイート内に含まれていれば、行動促進ツイートとして抽出を行っているため、行動促進フレーズを含むが行動促進を行っていないツイートを多数抽出したと考えられる。しかしながら、再現率が高いため、ある程度行動促進ツイートを抽出できていることがわかる。

SVM

SVM では、ダイエットの話題において、適合率、再現率共にある程度高い結果となった。これは、行動促進ツイートの特徴が動詞、助動詞に出現していたためであると考えられる。一方、

⁴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 12: 実験 1 におけるダイエットの話題の抽出結果

手法	Precision	Recall	F-measure	10-cross validation
Rule Base	0.4532	0.8726	0.5965	—
SVM	0.8407	0.7740	0.8060	0.8773
Random Forest	0.8597	0.5745	0.6888	0.834
LSTM	0.8537	0.7716	0.8106	0.8627

表 13: 実験 1 における熱中症の抽出結果

手法	Precision	Recall	F-measure	10-cross validation
Rule Base	0.6557	0.7742	0.7101	—
SVM	0.6534	0.7054	0.6784	0.8773
Random Forest	0.8202	0.7849	0.8021	0.8340
LSTM	0.8004	0.8538	0.8262	0.8627

熱中症に関しては、適合率が低い結果となった。これは、動詞、助動詞を用いた行動促進フレーズの内、意志形を含む表現と同様の表現が投稿主の意志を示している場合があり、それらが教師データの負例に多く出現したため、動詞、助動詞だけでは正確に判定できなかったと考えられる。

ランダムフォレスト

ランダムフォレストにおいては、ダイエットの話題では、再現率が低い結果となった。これは、負例にも出現する文末表現がテストデータにおける行動促進ツイート内に多く出現していたため、行動促進ツイートが抽出できていないと考えられる。一方、熱中症の話題では、適合率、再現率、ともに高い値となった。これは、ダイエットの話題に比べ、行動促進を行う際の表現が決まっており、教師データにおける負例の表現があまりテストデータ内の行動促進ツイートに出現していなかったためであると考えられる。

LSTM

LSTMにおいては、全話題において、最も F 値が高い結果となった。これは、単語の順序を考慮することにより、行動促進フレーズが他人への行動促進を示しているのか、ユーザ自身の行動を示しているのかを判別できたためであると考えられる。また、単語の順序を考慮することで、他の手法では獲得することのできなかった表現を学習することができ、正しく判定を行っていると考えられる。

また、2つの話題のそれぞれの値のグラフより、LSTM が最も良い結果となっていることがわかる。このことから、LSTM により、ある程度単語の順序を考慮することができ、行動促進ツイートの抽出を行えていることがわかる。そのため、LSTM が平常時の行動促進ツイートの抽出に適した手法であると考えられる。以下の総評にて具体的な行動促進ツイートの抽出結果を述べる。

総評

手法毎の考察

ダイエットの話題を例に各手法による行動促進ツイートの抽出結果を表 14 に示す。表 14 の No.1, No.2 においては、ルールベース以外の手法にて、正しく行動促進ツイートの抽出を行うことができている。このことから、機械学習を用いることにより、“大切です”や“OK ですよ”

表 14: 実験 1 におけるダイエットに関する提案手法を用いた行動促進ツイートの抽出結果

No.	Correct	Rule	SVM	RF	LSTM	Tweet
1	○	×	○	○	○	【ゆっくり食べるとダイエット効果】食事は時間をかけて食べると、約 15-20 分後には満腹中枢に指令が届き、満腹感を感じられるようになるので食べる量が減らせます。また嘔む回数が増えると、細くなった食材からの栄養吸収が高まる為、痩せやすい体作りにも大切です。
2	○	×	○	○	○	ダイエットに役立つと言われるクエン酸を充分にとると、脂肪や乳酸が効率よくエネルギーに分解されます。最近話題のもろみ酢には、このクエン酸が豊富に含まれていますよ。酸っぱいのが苦手な方は薄めて飲んでも OK ですよ。
3	○	○	×	×	○	ダイエットの最大の敵である停滞期の突破として、最も効果的なのが筋トレと高タンパクな食生活の組み合わせ。絶食はマジで身体壊すことになるのでやめた方がいいです。僕は一度壊したし、即効でリバウンドしました。停滞期で困ったらとりあえず筋トレしてプロテイン飲んでけば間違いありませんよ。
4	○	○	×	×	○	酸っぱいのが気になるなら酢を一回沸騰させれば酸味少し軽減するよ。ダイエットしてるなら砂糖はやめたほうがいいよ。食べ方の順番わかんないけど、酢の物たべてから最後に炭水化物がいいよ。
5	○	×	×	×	○	ダイエット中でしたら糖質の多い根菜類はお控えくださいませ。芋や人参の根菜、またタマネギなどは糖分を多く含んでおりますので、野菜は葉もの中心に摂取することをオススメ致します。
6	×	○	○	×	×	苦手な食べ物がダイエットに良いので食べるように頑張ろう。
7	×	○	○	×	×	毎日もやし（をラーメンにトッピングして）食べるダイエット始めようかな。
8	○	×	×	×	×	朝起きたらグーンと背伸び！コレはお腹空いたときにやるのもダイエットに効果的です。

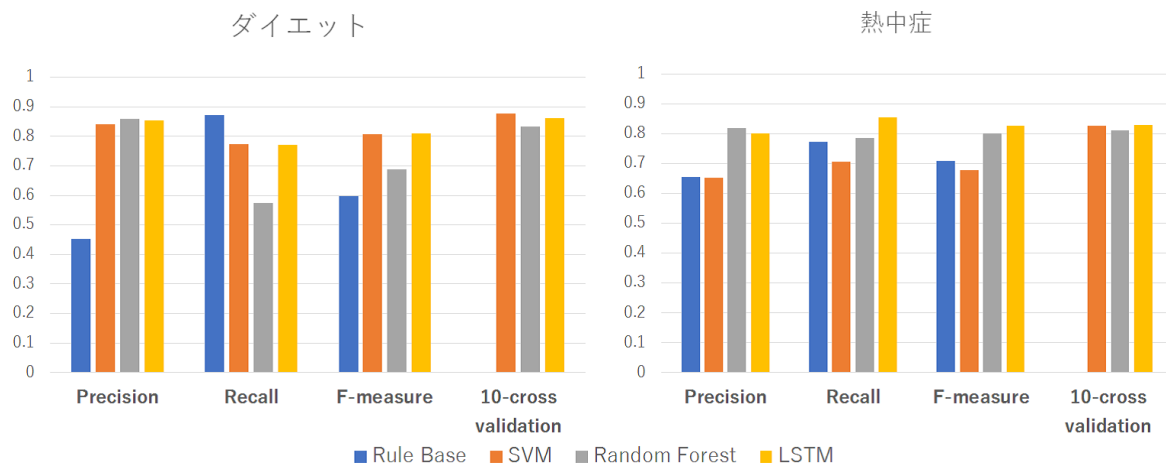


図 3: 実験 1 における 2 つの話題の実験結果

といった、行動促進フレーズを含まないが、行動促進を行っている表現を含む行動促進ツイートを抽出することができることがわかる。ルールベースでは、行動促進フレーズを含むかどうかのみを考慮して行動促進ツイートの抽出を行っているため、行動促進フレーズが含まれていなければ、行動促進ツイートを抽出することができない。一方、機械学習を用いた手法では、ツイート内に含まれる行動促進フレーズ以外の表現を考慮することにより、行動促進ツイートの抽出を正しく行えていると考えられる。

表 14 の No.3, No.4 においては、ルールベース, LSTM においては、ツイート内の行動促進を行っている部分を考慮して判定を行うことができている。一方, SVM とランダムフォレストは誤った判定をしている。SVM の判定が誤っている理由としては、ツイート内に“方が良い”という動詞や助動詞を用いていない行動促進フレーズが含まれているが、SVM は動詞, 助動詞の特徴を考慮して分類している。そのため, SVM では、行動促進フレーズに動詞, 助動詞が含まれていないツイートには対応できないため、抽出ができなかったと考えられる。ランダムフォレストにより抽出ができている理由としては、1 つのツイート内に複数の文末表現が出現していた場合、それらの文末表現が教師データの負例に多く出現している等、誤って判定をしているためであると考えられる。

LSTM を用いることにより、表 14 の No.5 のように、他の 3 つの手法では抽出することのできない行動促進ツイートを抽出することができた。No.5 の例では、“オススメ致します”といった他の手法では考慮することのできない表現を含む行動促進ツイートを抽出することができた。これは、LSTM ではある程度単語の順序を考慮することができ、教師データに出現している行動促進ツイートにおける、行動促進に用いる様々な表現を考慮することができたため、他の手法では抽出できなかったものを抽出することができたと考えられる。

また、No.6, No.7 のようにルールベース, SVM では誤って行動促進であると抽出しているが、LSTM では正しく判定をすることができている。これは、ある程度単語の順序を考慮することで、LSTM では、ツイートの主体が誰であるのかをある程度考慮することができ、行動促進フレーズを含むが行動促進を行っていないツイートを正しく判定できているためであると考えられる。ルールベース, SVM では、行動促進フレーズが誰の行動を示しているのかを考慮せずに抽出しているため、誤って抽出を行っている。ランダムフォレストにおいては、負例に多

表 15: 実験 1 における 2 つの話題の LSTM を用いた抽出結果

話題	Precision	Recall	F-measure	10-cross validation
ダイエット	0.8537	0.7716	0.8106	0.8627
熱中症	0.8004	0.8538	0.8262	0.8289

く出現した表現がツイート内に存在しているため、正しく判定できたと考えられる。このように、LSTM を用いることでしか抽出できないツイートや、LSTM を用いることにより行動促進ではないツイートであると正しく判定することができている。さらに、他の手法では抽出できなかった行動促進ツイートに関しても抽出を行うことができている。

しかし、No.8 の例のように、“効果的です”といった行動促進を行っているような表現が含まれていたとしても、LSTM では抽出できていない例があった。これは、ツイートが短い場合、単語の順序により、行動促進となる文脈であるのかということ考慮できていないことが原因であると考えられる。

話題毎の考察

表 15 に話題ごとの LSTM の適合率、再現率、F 値、10 交差検定の精度を示す。2 つの話題を比較すると、ダイエット、熱中症の話題では、適合率、再現率が共に高い値となった。これは、この 2 つの話題の行動促進ツイートを分析した結果、自分自身が経験したことや、知っている有益な情報を他人に伝えようとする傾向がある為、行動促進フレーズが教師データの正例に多く出現していたと考えられる。そこで、それぞれの話題のツイートを分析すると、ダイエットの話題に関しては、ツイートを閲覧したユーザーに対するアドバイスを行っているようなツイートが多く見受けられた。熱中症の話題に関しては、他人に対しての注意喚起を行っているようなツイートが数多く見受けられた。“食べよう”や“対策しよう”といった行動促進フレーズを含むが、行動促進をしていないツイートが多く存在している。しかしながら、ツイートの文字数が少ない場合、明確にツイートの主体を考慮していないため、行動促進フレーズを含むが、行動促進を行っていないツイートの判定が困難となっている場合があると考えられる。そのため、ツイートの主体を考慮し、学習を行うことにより、ツイート内の行動促進フレーズが他人への行動促進を行っているのか、投稿主の行動を示しているのかを判定でき、適合率が向上すると考えられる。

以上の結果より、話題により結果が異なるが、平常時の話題において LSTM を用いた行動促進ツイートの抽出手法が最も良い結果となった。このことから、単語の順序が行動促進を行っているかに影響を与えているということがわかる。また、LSTM によって、他の手法では対応できなかった行動促進ツイートを抽出することができている。そのため、単語の順序を考慮することのできる LSTM が行動促進ツイートの抽出に適した手法であるといえる。

8.2 実験 2

実験 2 では、災害時の行動促進ツイートから、提案手法を用いて行動促進ツイートの抽出を行い、災害時の行動促進ツイートの抽出に適した手法を確認する。平常時の話題のツイートと比べると、災害時のツイートは緊急度が高く、情報が直接命に関わる場合もある。そのため、災害時の話題において、Twitter 上から行動促進ツイートを抽出することは重要であるといえる。このことから、災害時の話題を用いて行動促進ツイートの抽出を行う。本研究では、災害時の

話題として、地震の話題を用いる。さらに、平常時の行動促進ツイートと災害時の行動促進ツイートの内容にはそれぞれ特徴があると考えられることから、それぞれの話題の比較、分析を行う。

8.2.1 実験データ

実験データとして、災害時のツイートをアンケートの被験者に提示し、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ウドソーシングを利用した。具体的には、北海道胆振東部地震発生後に「地震」のクエリを用いてランダムに収集した10,000件のツイートをを用いる。それらのツイートに対し、アンケートにより行動促進ツイートであるか、そうでないかの判定を行う。判定の方法は、1ツイートあたり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、全被験者が行動促進でないと答えたツイートを負例とする。その結果、正例2,405件、負例5,696件の合計8,101件のツイートが正解付きデータとして判定された。これらの正解付きデータのうち、正例、負例ともに同程度にするために、負例として判定された5,696件のツイートからランダムに抽出した2,405件のツイートを実験データの負例データとして用いる。使用する実験データは正例、負例ともに2,405件で合計4,810件である。ダイエット、熱中症の話題を平常時の話題として用い、地震のツイートのデータと同様に、実験1で用いたそれぞれの話題のデータに対し、正例、負例の数を同程度にする処理を行い、実験データとして利用する。LSTMの各種パラメータに関しては、第7章で説明したものをを用いる。実験2にてSVMの学習に用いる各話題ごとの各種ハイパーパラメータを表16に示す。実験2において用いるSVMとLSTMにて利用する分散表現の獲得手法はWord2Vecを改良した手法であるfastText⁴を用いる。fastTextの学習済みモデルデータとして、国立国語研究所のNWJC2Vec[25]を用いる。

8.2.2 実験方法

学習時の各話題の教師データ、テストデータとして表17に示す数のツイートをそれぞれの手法の評価に用いる。全体のデータから正例、負例それぞれ2割のツイートをランダムに抽出したものをテストデータとして用い、残りのデータを教師データとして用いる。これらの教師データを用いてSVMの2パターンの素性を用いた手法とLSTMを用いた手法の計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災害時の行動促進ツイートの抽出手法として決定する。さらに、災害時の話題だけでなく、平常時の話題にも同様の手法を用いて行動促進ツイートの抽出を行い、それぞれの状況における行動促進ツイートの分析を行う。

8.2.3 結果と考察

図4、表18、表19、表20に各話題における3つの提案手法の適合率、再現率、F値、10交差検定の結果を示す。

⁴<https://fasttext.cc/>

表 16: 実験 2 における話題ごとの SVM のパラメータ

話題	パターン	コストパラメータ C	カーネルパラメータ g
地震	1	8.0	0.03125
	2	8.0	0.03125
ダイエット	1	2.0	0.125
	2	2.0	0.03125
熱中症	1	2.0	0.03125
	2	2.0	0.03125

表 17: 実験 2 における話題ごとの実験データの数

話題	総数			教師データ			テストデータ		
	正例	負例	合計 1	正例	負例	合計	正例	負例	合計
地震	2,405	2,405	4,810	1,935	1,913	3,848	470	492	962
ダイエット	2,112	2,112	4,224	1,691	1,688	3,379	421	424	845
熱中症	2,314	2,202	4,516	1,849	1,763	2,314	465	439	904

まず、ルールベース、SVM、LSTM の 3 つの手法の各々の考察を行う。

各手法ごとの考察

ルールベース

ルールベースの結果は、地震の話題においても適合率が低く、再現率が高い結果となった。再現率が高い理由としては、災害時において行動促進を行う際には、「してください」といった命令形を含む行動促進表現が多く用いられており、この表現は他人に行動を促す際に多く用いられているため、正しく抽出することができたと考える。また、再現率が高い事より、ルールベースは行動促進ツイートがある程度正確に抽出できたといえる。しかしながら、ルールベースのみでは抽出することができていない行動促進ツイートが存在している。そのことから、更なるルールの追加、または機械学習を用いた行動促進ツイートの抽出が必要であることがわかる。

また、ルールベースを用いて抽出を行った際に誤判定の結果のツイートの例を表 21 に示す。タイプ 1 においては、「着よう」や「でよう」といったツイート内の行動促進フレーズが個人の意志や行動を示している場合が存在していた。タイプ 3 においては、物事に対する願望や希望といった内容が誤って抽出されていた。タイプ 4 に関しては、「方がいいのでは？」という表現が多く見受けられ、他人に疑問を投げかける場合に同様の表現が見受けられた。

SVM

まず、SVM の 2 つのパターンに関して考察を行う。

パターン 1 : 全品詞を素性に使用

パターン 2 : 動詞と助動詞のみを素性に使用

図 5 に SVM に使用したパターン 1、パターン 2 それぞれのツイートの分散表現ベクトルを次元圧縮を行い、2次元上にプロットしたものを示す。どちらもある程度正例、負例毎にまとまっているため、SVM を用いて行動促進ツイートであるかを判定することができたと考えられる。そして、パターン 1 とパターン 2 の精度を比較すると、適合率はパターン 2 の方が高いものの、再現率、F 値はパターン 1 のほうが優れていることがわかる。このことから、使用した品詞にお

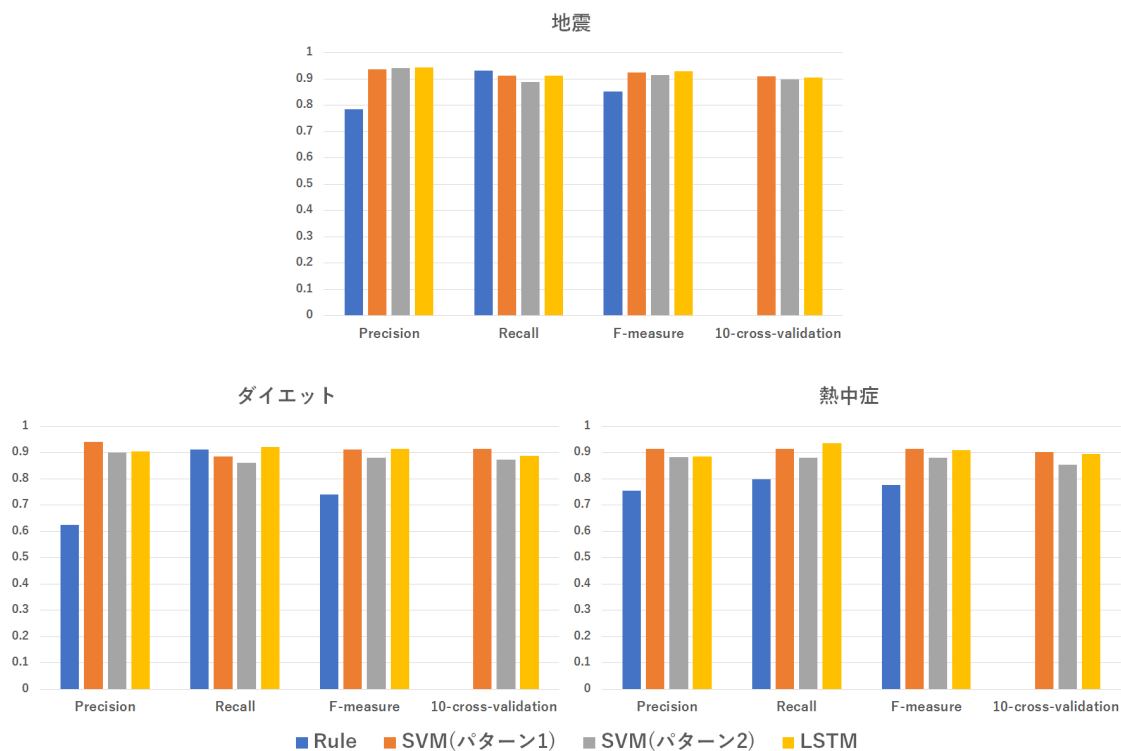


図 4: 実験 2 における 3 つの話題の実験結果

表 18: 実験 2 における地震の抽出結果

手法	Precision	Recall	F-measure	10-cross validation
Rule	0.7835	0.9319	0.8513	—
SVM(パターン 1)	0.9367	0.9128	0.9246	0.9111
SVM(パターン 2)	0.9414	0.8894	0.9147	0.8986
LSTM	0.9449	0.9128	0.9286	0.9062

いて、動詞、助動詞以外の品詞においても行動促進を示しているものが存在しているといえる。そのため、SVMを用いた行動促進ツイートの抽出には全品詞を使用するパターン 1 を用いる。

次にパターン 1 を用いた SVM について考察する。パターン 1 を用いた SVM の結果は、他の手法と比べて適合率、再現率共に高い値となっている。これは、単語の出現の組み合わせを考慮することにより、行動促進ツイートの抽出を行うことができていると考えられる。しかしながら、文字数が多いツイートに関しては、行動促進フレーズが含まれていたとしても、正しく判定できていない場合が存在していた。これは、ツイートの分散表現を求める際に、出現単語の分散表現ベクトルの平均値を用いているため、一部の単語が行動促進をしている単語であっても、単語数が多いツイートの場合、その特徴が埋もれているからと考えられる。

LSTM

LSTM の結果は、他の手法と比べて F 値が高い結果となった。これは、LSTM を用いることで、ある程度単語の順序を考慮することができ、新たな行動促進の表現を抽出できたためと考

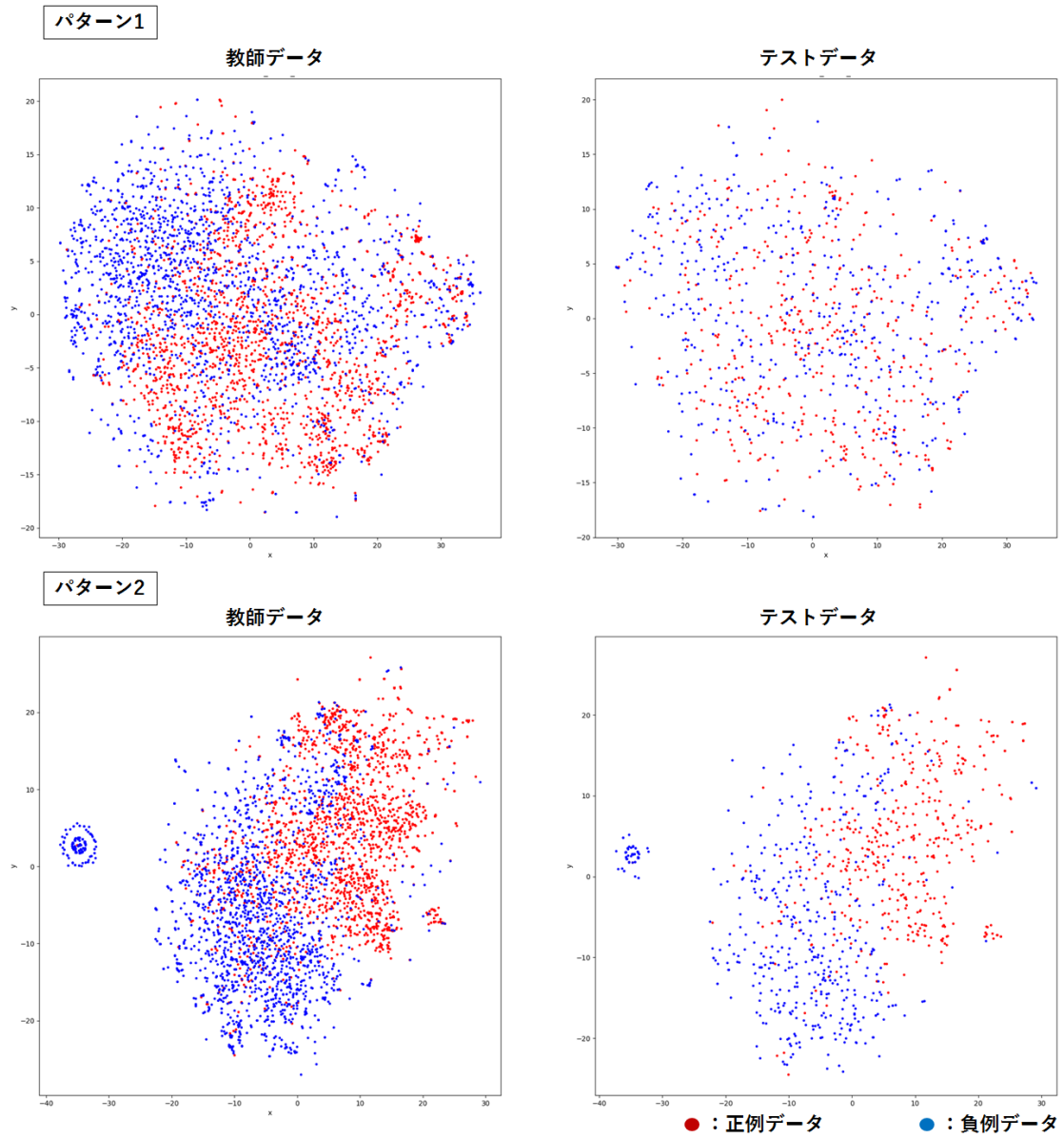


図 5: 実験 2 における地震の話題のツイートのプロット結果

表 19: 実験 2 におけるダイエットの抽出結果

手法	Precision	Recall	F-measure	10-cross validation
Rule	0.6244	0.9121	0.7413	—
SVM(パターン 1)	0.9395	0.8860	0.9120	0.9129
SVM(パターン 2)	0.9005	0.8599	0.8797	0.8736
LSTM	0.9044	0.9216	0.9129	0.8885

表 20: 実験 2 における熱中症の抽出結果

手法	Precision	Recall	F-measure	10-cross validation
Rule	0.7556	0.7978	0.7762	—
SVM(パターン 1)	0.9140	0.9140	0.9140	0.9014
SVM(パターン 2)	0.8815	0.8796	0.8805	0.8540
LSTM	0.8841	0.9355	0.9091	0.8938

えられる。特に、SVMと比較すると、文字数の多いツイートでも、重要な部分を考慮し抽出を行うことができた。しかしながら、行動促進フレーズ以外の表現では、抽出を行えないツイートが存在していた。これは、教師データを追加し、行動促進フレーズのパターンを増やすことが必要であると考えられる。

また、SVMとLSTMはほぼ同じ抽出精度となっており、ルールベースを用いるよりも機械学習を用いた方が、より正確に行動促進ツイートの抽出が出来ることがわかる。そこで、SVMとLSTMの実際のテストデータにおける抽出結果を比較する。

まず、表 22 に SVM によって行動促進ツイートを抽出できていない例を示す。文字数の多い No.1, No.2 のツイートのように行動促進フレーズが含まれていても正しく判定できていない場合があった。これは、ツイートの特徴ベクトルを取得する際に、ツイート内の単語の分散表現ベクトルの平均を使用しているため、単語数が多い場合行動促進をしている単語が出現していたとしても、特徴が抽出できていないためであると考えられる。一方、No.3, No.4 のように文字数が少ないツイートでは、行動促進フレーズが含まれているツイートは抽出できているが、行動促進フレーズが含まれないツイートは正しく判定することができていなかった。これは、単語数が少ない場合、行動促進フレーズ以外の行動促進を行っている単語の特徴が現れず、判別することができていないと考えられる。これに対し、ツイートの文字数が少ない場合は、「しましよう」といった行動促進の特徴が現れやすいため、行動促進フレーズを含んでいるツイートは正しく判定することができた。

次に、表 23 に LSTM の行動促進ツイートの抽出例を示す。No.1 の例のように、行動促進フレーズが含まれないが、行動促進ツイートであるものが抽出することができなかった。これは、教師データ内に出現している同様の表現が少なかったためと考えられる。また、No.2 のように、SVM では特徴が埋もれてしまって抽出ができなかった行動促進ツイートも LSTM を用いることにより抽出することができた。これは、LSTM を用いることで、単語の順序を考慮でき、かつツイート内の行動促進をしている部分を学習できたためと考えられる。一方、No.3, No.4 の例のように、ツイート内に行動促進フレーズが含まれていたとしても抽出できない例が存在していた。ツイートが短い場合、単語の流れを考慮できず、抽出が困難になっていると考えられる。このことから、LSTM が単語の順序を考慮することにより文字数の多いツイートにも対応

表 21: 実験 2 におけるルールベースによる誤判定の例

タイプ	ツイート
1	何か沖縄で大きな地震が起こるかもとの眩きが多いね。救助隊に発見されても困らないように寝るときは <u>衣着よう</u> …。屋内裸族 <u>学校でよう</u> としたら地震きてびびった
3	田舎にいた頃、地震で3日以上停電してたけど、あれは冬だったからまだマシなだけで、暑さは和らいだと言えどまだまだ暑いじゃん。早く復旧 <u>する</u> といいなあ…。 東京の強風なんて屁のレベルってぐらい大阪の被害がすごすぎてツイッターみながら震えてた:(´ω`):そして今日は北海道の地震…停電続いてるみたいだけど、早く復旧 <u>する</u> といいね(´;ω;`)ゆきちゃんが無事であることはツイートで把握していた…w
4	原発事故は悲惨ですが目の前の危機をどう乗り切るか。原発は稼働していてもいなくても地震によるリスクはさほど変わりません。それなら稼働させて電力の安定供給に <u>使った方がいい</u> のでは？ <u>マンション借りる</u> けど、最近の地震とかの影響もあって、川の近くはやめといた方がいいのかな？

※行動促進フレーズは下線で示している

することができ、行動促進ツイートの抽出が出来ていることがわかる。また、文字数の少ないツイートに対応するためには、文字数の少ない行動促進ツイートの教師データを増やすことにより、抽出が可能になるのではないかと考えられる。

また、表 22、表 23 に示すツイートのように、SVM や LSTM の機械学習の結果より、行動促進ツイートのルールでは設定していない行動促進をしている表現が抽出できることがわかった。例えば、「お願いします」や「注意が必要です」という表現である。しかしながら、このような表現は行動促進でないツイートにも出現している。これは、適合率が低下する原因になるため、単純にルールに追加することはできない。さらに、機械学習を用いることにより行動促進ツイートの抽出ができた理由としては、ツイート内のほかの単語を考慮することができたためであると考えられる。以上のことより、機械学習の手法がルールベースよりも優れているといえる。

どの話題においても、LSTM の精度がよく、単語数の多いツイートにおいても行動促進ツイートの抽出ができています。そのため、LSTM が災害時の話題において行動促進ツイートの抽出に適した手法であると考えられる。

総評

以下の総評にて地震の具体的な行動促進ツイートの抽出結果を述べる。

手法毎の考察

災害時の話題である地震を例に各手法による行動促進ツイートの抽出結果を表 24 に示す。表 24 の No.1, No.2 は、ルールベース以外による行動促進の表現を機械学習を用いることにより抽出することができた例である。SVM では、全ての品詞を用いることにより、行動促進の特徴が抽出できている。LSTM では、教師データに含まれる行動促進の表現を学習することができ、行動促進を行っている部分を考慮して抽出することができている。これらのことより、機械学習を用いることにより、行動促進ツイートを抽出できることがわかる。

表 22: 実験 2 における SVM の誤抽出の例

No.	Correct	Rule	SVM	LSTM	Tweet	文字数
1	○	○	×	○	【休館情報】ガラス工芸館は地震の影響で、煙突かられんがが一部剥がれ落ちています。屋根の上にもれんがが乗っている状態で、引き続きれんが落下の危険があるため、当面休館となります。建物のまわりに近づかないようにお願い致します。	131
2	○	○	×	○	私も幼少の時に阪神大震災、成人してから東大日本地震経験してライフラインが麻痺したときの不便さやストレスを感じた。自分は大丈夫って思っても回りがそうなる自分もつられて影響してしまうからね。何が言いたいかというと、災害時用バック用意しておくの良いよ	121
3	○	×	×	○	本日（914）15 時台の電力需要は、震災前日比で「11.1 %減」でした。もう一段節電のご協力をお願いします。#北海道みんなで節電#北海道#地震#meti_saigai	84
4	○	×	×	×	胆振東部地震で震度が大きい地点では土砂崩れにも注意を 02:51:37 震度 4 北海道早来追分#地震	47

表 24 の No.3, No.4 は, SVM のみ行動促進ツイートと判定できていない例である。表より行動促進フレーズが含まれていても, SVM では抽出することができないものが存在していることがわかる。これは, SVM ではツイートの単語が含まれているか否かのみで判定を行っているからである。それに対し, LSTM ではツイート内の単語の順序を考慮しているため, 行動促進を行っている部分を特定し抽出出来ていると考えられる,

LSTM を用いることにより, 表 24 の No.5, No.6 のように, ルールベースと SVM では抽出することのできない行動促進ツイートを抽出することができた。No.5 の例では, 「注意」といった他の手法では考慮することのできない表現を含む行動促進ツイートを抽出することができた。また, No.6 の例では, ツイート内の引用を行っている部分に行動を促している表現が出現している。この場合, LSTM では, 行動促進をしていると考えられる「注意が必要です」といった表現を考慮し判定することができている。これに対し SVM には, 全ツイート内の単語の平均の値をツイートの分散表現として用いているため, 抽出することができていると考えられる。これは, 単語の順序を考慮することにより, 行動促進フレーズの出現場所や, 教師データに出現している様々な行動促進のフレーズ表現を考慮することができたため, 他の手法では抽出できなかった行動促進ツイートを抽出することができたと考えられる。

以上の結果より, LSTM が災害時の行動促進ツイートの抽出に適しているといえる。

表 23: 実験 2 における LSTM の抽出例

No.	Correct	Rule	SVM	LSTM	Tweet	文字数
1	○	×	×	×	つまり、北海道で起きた地震の時の状況が西日本全域に存在しているということを伝えるものでもあったのです。本当のメカニズム云々は別にしても、以前から言っているように西日本においては足元に、直下型の地震に備える必要があるのです。恐れるのではなく、正しく備えるのです。	130
2	○	○	×	○	iPhone4S は緊急地震速報がデフォルトで OFF になっています。利用したい方は ON にしましょう。「設定をオンにするとバッテリーの持ち時間が短くなる場合があります。」と記載がありますが、au 版は、BC-SMS の仕様上、オン設定で短くなることはありません。_a6m	131
3	○	○	○	×	エレベーター内で地震に襲われたら、階のボタンを全部押してください。そうすることで、どこかの階で停止してくれる可能性が高くなる。	63
4	○	○	○	×	地震→停電のときは復電時に火出るからブレーカー落とそう！	28

平常時と災害時の手法比較

表 18, 表 19, 表 20 より, 各手法における平常時と災害時における行動促進ツイートの抽出結果を比較する. ルールベースにおいては, 地震の話題において, 平常時の話題に比べると適合率が高い結果となっている. これは, 緊急時には自分自身の意志を述べるツイートではなく, 他人への行動を促進している内容が多く存在しているためと考える. 特に, 行動促進に用いる表現は「してください」といった命令形を用いたものが多く存在していたため適合率が高くなったと考えられる.

また, LSTM の結果に着目すると, 地震の話題においては適合率が 0.94 と高い値となっている. 一方, 平常時の話題では, 適合率は地震の話題と比べると, 低い値となっている. これは, 平常時の行動促進ツイートは, 災害時の行動促進ツイートに比べ, 行動促進を行う際の表現が複雑であり, 同じ表現でも行動促進を行っているもので行っていないものがあるためと考えられる. このことに対応するためには, 行動促進の表現の主体を考慮し, 誰に対する行動促進を行っているかを考慮し学習を行う必要がある. 一方, どの話題においても LSTM が良い結果となっていることから, 災害時, 平常時どちらの話題においても LSTM が行動促進ツイートの抽出に適している手法であるといえる.

表 24: 実験 2 における地震に関する提案手法を用いた行動促進ツイートの抽出結果

No.	Correct	Rule	SVM	LSTM	Tweet
1	○	×	○	○	【休館情報】ガラス工芸館は地震の影響で、煙突かられんがが一部剥がれ落ちています。屋根の上にもれんがが乗っている状態で、引き続きれんが落下の危険があるため、当面休館となります。建物のまわりに近づかないようお願い致します。
2	○	×	○	○	とにかく今は余震に備えて。東日本大震災、熊本地震を教訓に。診察取り止めの病院多いので、できればお薬手帳と処方された薬を持っての避難を。血栓症注意。サランラップ、除菌ウェットティッシュ、ビニール袋は色々役に立つよ。#避難
3	○	○	×	○	先程停電が復旧しました。ポンプも動き出して水道も復旧。冷蔵庫は開けないようにしていたので、この時間まで冷凍庫の氷も解けていませんでした。だが油断はしない。また地震が来たり、いきなり大量に電気使った停電になったりの可能性。まずは水を貯めよう。
4	○	○	×	○	私も幼少の時に阪神大震災、成人してから東日本大地震経験してライフラインが麻痺したときの不便さやストレスを感じた。自分は大丈夫って思っても回りがそうだと自分もつられて影響してしまうからね。何が言いたいかというと、災害時用バック用意しておくといいよ
5	○	×	×	○	【気象庁】特に2~3日程度は規模の大きな地震の発生が多いので注意
6	○	×	×	○	《今まで沈降をしていた札幌および苫小牧は沈降から大きく隆起に転じました。注意が必要です》(8月22日) 東京大学名誉教授で JESEA (地震科学探査機構) 会長の村井俊治さん。村井さんの警告をよく確認しておきたい。

表 25: 各話題における行動促進ツイート

話題	Tweet
地震	北海道のみなさん、今日は大気の状態が不安定になって、夕方にかけて局地的に強い雨や落雷、突風の恐れがあるそうです。また、7日の夕方から8日の午前中は、さらに広い範囲で雨が降るようです、台風直後の地震で、地盤が緩んだりしていると思いますのでお気を付けください！
	北海道の皆さん大丈夫ですか？台風の次は地震って…家の中がグチャグチャになってる人も多いと思います。まずは深呼吸して落ち着いて！自分の身の周りの安全を確保し、停電地域は復旧まではブレーカーを落としといてね！本震がこのあと来る可能性が高いから本当に警戒しててください。
ダイエット	老廃物たっぷりの食材老廃物をたっぷりと含んだ食材は、食品添加物、着色料、動物性たんぱく質です。これらを取りすぎると老廃物がどんどん溜まって疲れがとれにくく、お肌もガサガサ、ダイエットも上手くいかないという案循環を引き起こします。出来るだけ無添加の食材を取りましょう！
	炭水化物ダイエットはアルコールは厳禁？ 炭水化物ダイエットは、炭水化物の摂取量を抑えなくてはなりません。そのため、糖分が含まれているアルコールは飲まないほうがいいです。
熱中症	原発事故は悲惨ですが目の前の危機をどう乗り切るか。原発は稼働していてもいなくても地震によるリスクはさほど変わりま熱中症対策の豆知識として、重度の熱中症になりかけた場合はとにかく体に水をかけるなどして物理的に体温を下げましょう。重度の場合は水分補給よりもまずそれ。首、脇あたりを重点的にするとなお良し。ねねねーさんが命がけで体験した知識だからいざという時はちゃんと使うんだぞ
	おはようございます(*^v^*)!今日は久々の雨の予報ですね。草木にとっては恵みの雨ですが、皆様は傘をお忘れなく！また湿度と温度が上がると熱中症の危険性もアップします。こまめな水分補給もしてくださいね！

平常時と災害時の行動促進ツイートの分析

表 25 に話題ごとの行動促進ツイートの例を示す。

平常時の話題であるダイエット、熱中症に関しては、それぞれの話題で特徴が異なっていた。まず、ダイエットの話題に関して述べる。ダイエットの特徴は、自分自身の意志で行うという点である。そのため、行動促進の内容としては、「するといいよ」、「した方が良いよ」といったような提案を行うような緩やかな行動促進を行っているツイートが存在している。それに対し「してください」といった命令形を含むようなものは少ない傾向にあった。

一方、熱中症の特徴は、病気になるように対策を自発的に行うものである。そのため、ダイエット同様に熱中症への対策に関して提案を行うものが存在している。また、実際に熱中症になったユーザによる危険な状態や、対策などの注意喚起を行うような行動促進が多く見受けられた。経験から情報を伝えることにより、注意喚起を行うツイートが多い傾向にあった。

平常時の話題において共通する内容として提案を行うような緩やかな行動促進を行っているツイートが多く見受けられていた。これは、情報の重要度がそれほど高くないものが平常時に

多く存在しているためであると考えられる。

災害時の話題である「地震」の話題のツイートは、災害時の情報が直接命に関わる可能性があるため、「しましょう」や「してください」といった直接的に行動を促す表現が多く見受けられた。また、テレビやラジオなどで聞いた情報だけでなく、他の Twitter のユーザから聞いた情報や、リツイート (RT) されたツイート内にある情報を他の人に伝えるようなツイートも多く見られた。このように、災害時には情報を強く伝えて行動を促そうといったツイートや、危険であることを伝え、してはいけない行動を促す注意喚起などの行動促進の表現が多い傾向にあるといえる。しかしながら、同じような情報であったとしても、情報に差が生じ、何が正しいのかわからない状況に陥ってしまう危険性があると考えられる。

以上のことより、災害時と平常時の行動促進の違いとしては以下の2点が挙げられる。

1. 災害時と平常時では、情報の重要度が異なり、災害時では情報が直接命に関わる可能性があることから、強く行動を促す傾向がある。
2. 平常時の話題においては、行動の提案など、強く行動を促す表現は少ない傾向がある。

9 まとめと今後の課題

本研究では、Twitter 上に存在する行動促進ツイートを抽出するための複数の手法を提案し、それぞれの手法に関して抽出精度の比較実験を行った。使用した話題は、平常時の話題として「ダイエット」、「熱中症」の2つと、災害時の話題として「地震」の話題の合計3つの話題を用いた。平常時の話題として「ダイエット」と「熱中症」の話題を選択した理由としては、ユーザが経験した情報や他人から聞いた有益な情報を他人に伝えようとする傾向があり、行動を促すツイートが多く存在すると考えたためである。災害時の話題として「地震」を選択した理由としては、近年地震が多発しており、地震に関するツイートが多く存在していると考えたためである。また、地震発生時には情報が直接命に関わることが多く、特にその中でも行動を促しているツイートは重要であり、そのツイートを抽出することが必要であると考えたためである。具体的には、Twitter からの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、平常時の話題、災害時の話題それぞれの話題の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どの手法であるのか、どの話題においても有効な手法であるのかを実験を行い、検証した。その結果、LSTM を用いた行動促進ツイートの抽出手法が最も適していることが確認できた。

今後の課題は以下の通りである。

- 現在対応できていない行動促進の表現への対応
本研究において、ルールベースでは抽出できていないツイートが存在しており、機械学習を用いることによりそのようなツイートを抽出することができた。しかしながら、機械学習では抽出することができていないツイートも存在していた。これらを抽出する為には、教師データを追加し、教師データの行動促進の表現の多様性を増加させることが必要であると考えられる。
- 異なる話題における行動促進ツイートの抽出実験
本研究において、平常時の話題として「ダイエット」、「熱中症」の2つの話題、災害時の

話題として「地震」の1つの話題の計3つの話題を用いて実験を行った。しかしながら、それぞれの話題においてLSTMの結果が最も良い結果となったが、それ以外の話題に関しても同様の実験を行い、適した手法を確認する必要があると考えられる。具体的には、平常時の話題である「インフルエンザ」や「ノロウイルス」、災害時の話題である「大雨」や「大雪」といった話題に関して実験を行う予定である。

- 提示する行動促進ツイートの選択

本研究では、Twitter上の行動促進ツイートの抽出手法を提案している。次の研究の手順として、行動促進ツイートを提示する必要があると考えられる。しかしながら、単純に抽出された行動促進ツイートを提示するだけでは、誤った情報をユーザに提示することや、不必要にユーザの不安をあおってしまう可能性があると考えられる。そこで、Tweet内に存在している情報を元に提示する行動促進ツイートの判定が必要であると考えられる。具体的には、ツイート内の行動促進を行っている根拠を抽出し、その根拠に含まれている情報を元に提示するべきであるのかを判別する手法を考案する必要があると考えている。

- 行動促進ツイートの提示方法の提案

本研究では、最終目標として、行動促進ツイートに対しアラートを出すことを目的としている。そのことから、ユーザに対し、どのように行動促進ツイートを提示するのが重要である。行動促進ツイートは実際にはそれほど多く存在している状態ではないため、行動促進ツイートがタイムラインに現れた時に目立たせる必要があると考えられる。具体的には、Twitterのタイムラインに行動促進ツイートが現れた際に、ツイートの背景色を黄や赤などの目立つ色に変更し、後々にまとめて行動を促していたツイートを閲覧することができるようなシステムを作成したいと考えている。

謝辞

本研究を進めるにあたり、大変多くの方々にお世話になりました。ここに感謝の意を示します。まず、研究において、さらに学生生活において多くの面において、約3年半もの長きに渡り熱く、心のこもった議論していただきました恩師、灘本先生に厚く御礼申し上げます。この研究室でなければここまで研究を続けることができていないのではないかと思います。また、精神的にも大きく成長することのできた約3年半の研究生活であったと感じております。そして論文にまとめるにあたり、有益な御助言とご教示を賜りました甲南大学小出武先生、甲南大学永田亮先生に心より謝意を申し上げます。また、研究の内容や、学会発表、論文執筆の際に丁寧なご指導をして下さりました奈良先端科学技術大学院大学鈴木優先生に深く感謝を申し上げます。そして、研究に対して度々議論に参加してくださいました先輩方、実験の際に快く被験者を引き受けてくださった後輩たちの皆にも厚く御礼申し上げます。最後に、大学院への進学を快く受け入れてくださり、また、金銭的な支援だけでなく不自由なく生活をさせていただきました、父・見塚喜市、母・見塚久美子に深く感謝致します。

研究業績

国際会議

- Keiichi Mizuka, Yu Suzuki, Akiyo Nadamoto, “Extraction of Commentary tweets about news articles”, International Conference on Information Integration and Web-based Applications & Services(iiWAS2017), pp. 188-192, 2017.
- Keiichi Mizuka, Yu Suzuki, Akiyo Nadamoto, “Extracting Japanese Behavioral Facilitation Tweet in Disaster Situations”, The 20th International Conference on Information Integration and Web-based Applications & Services(iiWAS2018), pp. 302-305, 2018.
- Keiichi Mizuka, Yu Suzuki, Akiyo Nadamoto, “A Behavioral Facilitation Tweet Detection Method”, BIGCOMP '19: Proceedings of the 2019 International Conference on Big Data and Smart Computing (BigComp), (to appear).

国内会議

- 見塚圭一, 鈴木優, 灘本 明代, “Twitter 上のニュースに対する解説ツイート抽出手法の提案”, 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017), 信学技報, Vol. 117, No. 212, DE2017-19, pp.111-116, 2017 年 9 月
- 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), B5-2, 7 pages, 2018 年 3 月
- 見塚圭一, 鈴木優, 灘本 明代, “Twitter 上の行動促進ツイート抽出手法の比較検討”, 第 11 回 Web とデータベースに関するフォーラム (WebDB Forum 2018), 信学技報, vol. 118, no. 213, DE2018-17, pp. 49-54, 2018 年 9 月
- 見塚圭一, 米田吉希, 鈴木優, 灘本明代, “災害時の行動促進ツイート抽出手法,” 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019), (To appear)
- 米田吉希, 見塚圭一, 鈴木優, 灘本明代, “行動促進ツイートからの根拠部分抽出手法,” 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019), (To appear)

ポスター発表

- 見塚圭一, 鈴木優, 灘本 明代, “Twitter 上のニュースに対する解説ツイート抽出手法の提案”, 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017), 信学技報, Vol. 117, No. 212, DE2017-19, pp.111-116, 2017 年 9 月
- 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), B5-2, 7 pages, 2018 年 3 月
- 見塚圭一, 鈴木優, 灘本 明代, “Twitter 上の行動促進ツイート抽出手法の比較検討”, 第 11 回 Web とデータベースに関するフォーラム (WebDB Forum 2018), 信学技報, vol. 118, no. 213, DE2018-17, pp. 49-54, 2018 年 9 月

- 見塚圭一, 米田吉希, 鈴木優, 灘本明代, “災害時の行動促進ツイート抽出手法,” 第11回データ工学と情報マネジメントに関するフォーラム (DEIM2019), (To appear)

研究業績-表彰

学生プレゼンテーション賞

- 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), B5-2, 7 pages, 2018年3月

学生奨励賞

- 見塚圭一, 鈴木優, 灘本明代, “Twitter上の行動促進ツイート抽出手法の比較検討”, 第11回Webとデータベースに関するフォーラム (WebDB Forum 2018), 信学技報, vol. 118, no. 213, DE2018-17, pp. 49-54, 2018年9月

参考文献

- [1] Akiyo Nadamoto, Mai Miyabe and Eiji Aramaki, “Analysis of Microblog Rumors and Correction Texts for Disaster Situation” Proc. of International Conference on Information Integration and Web-based Applications & Services, pp. 44-52, 2013.
- [2] P. Jin, L. Mu, L. Zheng, J. Zhao, and L. Yue., “News feature extraction for events on social network platforms.” in *WWW '17 Companion Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp.69–78.
- [3] 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.
- [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] 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.
- [6] 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.
- [7] A. Ritter, Mausam, O. Etzioni, and S. Clark. Open domain event extraction from twitter. In *KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1104–1112, 2012.
- [8] T. Sellam and O. Alonso. Raimond: Quantitative data extraction from twitter to describe events. In *International Conference on Web Engineering ICWE 2015: Engineering the Web in the Big Data Era*, pages 251–268, 2015.
- [9] A.-M. Popescu and M. Pennacchiotti. Detecting controversial events from twitter. In *CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1873–1876, 2010.
- [10] A.-M. Popescu, M. Pennacchiotti, and D. Paranjpe. Extracting events and event descriptions from twitter. In *WWW '11 Proceedings of the 20th international conference companion on World wide web*, pages 105–106, 2011.
- [11] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *GIS '09 Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 42–51, 2009.
- [12] 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.

- [13] R. M. Tripathy, A. Bagchi, and S. Mehta. A study of rumor control strategies on social networks. In *CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1817–1820, 2010.
- [14] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: identifying misinformation in microblogs. In *EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599, 2011.
- [15] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer. Hoaxy: A platform for tracking online misinformation. In *WWW '16 Companion Proceedings of the 25th International Conference Companion on World Wide Web*, pages 745–750, 2016.
- [16] M. A. Cameron and B. Robinson. Emergency situation awareness from twitter for crisis management. In *WWW '12 Companion Proceedings of the 21st International Conference on World Wide Web*, pages 695–698, 2012.
- [17] K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, and S. Ghosh, “Extractingsituational information from microblogs during disaster events: a classification-summarization approach,” in *CIKM '15 Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 583–592.
- [18] 榑 剛史, 原 久美子, 吉田 光男, 鳥海 不二夫, 篠田 孝祐, 栗原 聡, 風間 一洋, 野田 五十樹, “災害情報基盤構築に向けたテキストデータからの地理情報抽出システム”. 人工知能学会全国大会予稿集, JSAI2014, pp1H2NFC02a4-1H2NFC02a4, 2014.
- [19] 坂巻 英一, 亀井 悦子, 鍛冶 伸裕, 吉永 直樹, 喜連川 優, “Twitter 上のつぶやきに関するテキストマイニングの事例研究-大規模災害発生時の被災地における現状把握への応用-, ” 日本経営工学会論文誌/65 巻 (2014-2015) 1 号 pp39-50, 2014.
- [20] A. Silva, H. Valiatinn, S. Guimaraes, and W. M. Jr. From individual behavior to influence networks: A case study on twitter. In *WebMedia 2011 Proceedings of the 17th Brazilian Symposium on Multimedia and the Web on Brazilian Symposium on Multimedia and the Web - Volume 1*, page 18, 2011.
- [21] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *SIGIR '12 Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 545–554, 2012.
- [22] A. Mogadala and V. Varma. Twitter user behavior understanding with mood transition prediction. In *DUBMMSM '12 Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media*, pages 31–34, 2012.
- [23] Y. Shuhei, K. Wakabayashi, T. Satoh, and N. K. Y Nozaki. Twitter user growth analysis based on diversities in posting activities. *International journal of web information systems*, 13:370–386, 2017.
- [24] 加藤 重広, 日本語文法入門ハンドブック, 研究社, 2006.

- [25] M. Asahara, “NWJC2Vec: Word embedding dataset from ‘NINJAL Web Japanese Corpus’,” *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication*, vol. 24, no. 2, pp. 7–25, Feb. 2018.