

Web ニュースの主題語に着目した解説ツイートの抽出手法の提案

見塚 圭一[†] 鈴木 優^{††} 灘本 明代[†]

[†] 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

^{††} 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

E-mail: [†]s1371116@s.konan-u.ac.jp, ^{††}ysuzuki@is.naist.jp, ^{†††}nadamoto@konan-u.ac.jp

あらまし インターネット上には時々刻々と様々なニュースが流れている。しかしながらこれらニュースの中には、主題となる人物や組織の詳細や関係性がわからないなど、ニュースの背景を知らないため、そのニュース自体が理解できないことがある。一方、インターネット上にはさまざまな情報があり、その中にはニュースの解説を行っているものもある。そこで、ニュースの解説をインターネットから抽出し、提示することにより、ニュースの理解支援になると考え、本研究ではニュースの解説文の抽出・提示手法を提案する。本論文ではそのはじめの一歩として、ニュースの主題となる人物や組織をニュースの主題語と呼び、その主題語に着目した解説文を Twitter から抽出する手法の提案を行う。

キーワード Twitter, Web ニュース, 解説文抽出

1. はじめに

日々様々なニュースがインターネット上に増加している。これら増加する様々なニュースをすべて理解する事は困難である。一方、インターネットの普及によりインターネット上には様々な情報が氾濫しており、人々はインターネットから情報を取得する事が日常的になっている。そこで我々は、ニュースの解説をしている情報をインターネットから抽出しユーザに提示することは、ニュースの理解支援になると考え、本研究ではニュースの解説情報の抽出・提示手法を提案する。

ニュースの記事には、ニュースの主題となる人や企業、またそのニュースの背景等がある。本研究ではこの主題となる人や企業を主題語と呼ぶ。ニュース記事の内容を理解するためには、はじめにこの主題語を理解することが必要であると考えられる。しかしながら、ニュース記事内において主題語に関して詳しく説明しているとは限らない。

ニュースの主題語には、有名人や企業等、Wikipedia 上に記事が存在している場合がある。Wikipedia の記事は、基本的な内容であり且つ中立的な立場から解説^(注1)が行われている。一方 Twitter 上でも様々な人がそのニュースに対して解説を行っている。その中には Wikipedia に掲載していないが重要な情報や、信憑性の怪しい情報など様々な情報がある。この Wikipedia に掲載されていないが重要な情報はニュースの解説をするに当たり重要であると考えられる。そこで本研究では Twitter から主題語に関する解説を行っているツイートを抽出し、その中から Wikipedia に存在していない情報を抽出・提示する手法を提案する。

例えば、Wikipedia に掲載されている情報の例としては、人物であれば、生年月日、出演作品や、行ってきた活動などである。一方、Twitter のみに載っている情報の例としては、ニュー

スの主題語に対する中立的ではない解説や、豆知識などである。例えば「ノロウイルスなど感染性胃腸炎患者 過去 2 番目の多さに」というタイトルニュースの場合、主題語は「ノロウイルス」であり、その主題語の Wikipedia の記事には「ノロウイルス属は、ウイルスの分類上第 4 群（プラス一本鎖 RNA ウイルス）のカリシウイルス科に属している。」のような学名や、感染対策等の基本的な解説が掲載されている。一方、Twitter には「ノロウイルス今年の流行は G.2 変異株というものであるが、これは迅速キットによる検査で検出されにくい。なので、検査陰性でも実際には感染していることはよくある。」等の Wikipedia には存在していないが重要な情報がある。このような情報を抽出し、提示する。

図 1 に主題語に対する Wikipedia に掲載されていないが重要な情報の抽出、提示手順を示す。

本論文では、Wikipedia に掲載されていないが重要な情報の抽出・提示する手法の研究のはじめの一歩として、Twitter からの解説ツイートの抽出手法の提案を行う。具体的には、ニュース記事から主題語の抽出を行い、抽出した主題語を用いてツイートを収集する。そして、機能表現や文章表現に着目し、解説ツイートの抽出を行う。

以下、2 章で関連研究について述べる。3 章で、ニュース記事から主題語を抽出する手法について述べ、4 章で Twitter から条件を用いて解説ツイートを抽出する手法について述べる。そして、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

ニュースの理解を深めるための研究は数多くされている。

馬 [1] らの研究では、情報補完の観点から、あるコンテンツに対する異なった視点のコンテンツや、より詳細に説明がなされているコンテンツの検索手法を提案している。本研究では、コンテンツに対する情報をツイートから取得し、取得した情報

(注1) : <https://ja.wikipedia.org/wiki/Wikipedia:五本の柱>

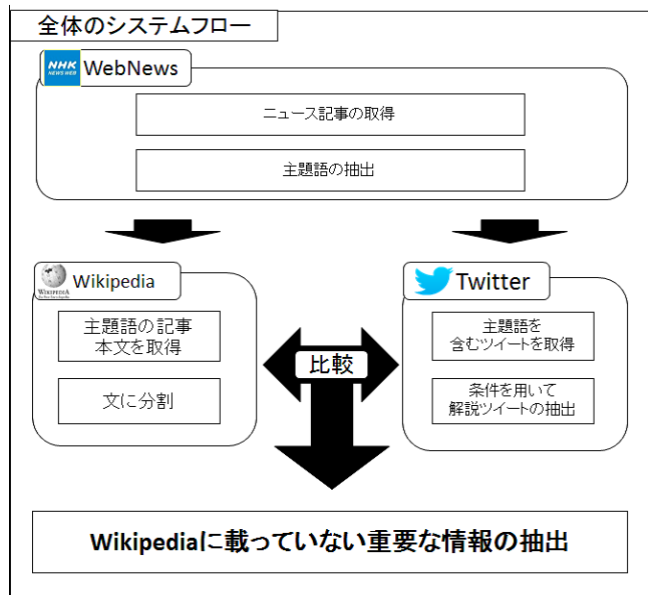


図1 提案手法のシステムフロー

から Wikipedia には載っていない重要な情報を取り出すという点で異なる。

北山 [2] らの研究では、ニュース記事内容の理解を深めるために、重要度の高い過去の関連ニュース記事を提示し、比較をすることにより、ニュース記事の理解を深めることを目的としている。本研究では、ニュース記事から主題語を抽出し、主題語に対する解説を行っているツイートを抽出し、提示することにより、ニュース記事の理解支援を行うという点で異なる。

田中ら [3] [4] の論文では、ニュース記事の理解支援のためにニュース記事中の人物や国家などの語句を抽出し Wikipedia を用いて、その語句が行った事柄等の年表を作成し提示することにより、ニュース記事の理解を深めることを目的としている。本研究では、ニュース記事から主題語を抽出し、Twitter 上から主題語に対する情報を取得し、提示するという点で異なる。

佐藤ら [5] の研究では、ニュース記事から話題を表すキーワードを取得し、取得したキーワードに関して詳細な説明をしているブログ記事を収集する手法を提案している。本研究では、ブログではなく Twitter を用いている点が異なる。

小林ら [6] は、ニュース記事から専門用語を抽出し、その専門用語に対する解説を、ソーシャルビデオ情報、ツイート情報、関連プロダクト情報といった多面的な関連情報マルチコンテンツ) から取得し、それらをユーザが利用しやすいインタフェース上に提示する手法を提案している。ツイート情報を解説に用いるという点では本研究と類似しているが、本研究はツイートから解説を行っているものを抽出しているという点で異なる。

3. ニュース記事の主題語の抽出

ニュース記事を象徴する単語はタイトルまたは本文の一文目に出現することが多いと言われている [7]。そこで、大原ら [8] はこれを参考にし、ニュース記事のタイトルに出現する名詞と第一段落に出現する名詞をそのニュース記事の主題語候補とし

て、式 (1) を用いてニュースの主題語の抽出を行っている。式 (1) において S_i は、各主題語候補の重要度を示す。本研究でも大原らの提案している式 (1) を用いてニュースの主題語を抽出する。

$$S_i = \alpha \cdot tf_{il} + \beta \cdot tf_{im} + \gamma \cdot tf_{in} \quad (1)$$

ここでは、 l はタイトル、 m は第一段落の文章であり、 n は第二段落以降を結合した文章である。主題語候補として、タイトルと第一段落に出現する名詞を選択する。 tf_{il} , tf_{im} , tf_{in} はそれぞれ主題語候補 i の各文書 $l.m.n$ における出現頻度を表す。 α , β , γ は重みを示し、 $\alpha + \beta + \gamma = 1$ とし、適切なパラメータを代入する。ここでは、我々の実験より $\alpha = 0.9$, $\beta = 0.075$, $\gamma = 0.025$ とする。

ニュース中の主題語候補に対して S_i を求め、 S_i の値が一番高いものを、そのニュース記事の主題語とする。

4. 解説ツイートの抽出

Twitter 上には、意見や解説、雑談などが混在している。一方、Wikipedia には、解説のみが存在している。したがって、ツイートと Wikipedia を直接比較することが困難であると考えた。そこで、Wikipedia と同じ条件で比較を行うために、そのニュース記事の解説を行っているツイートを抽出する。本論文では、解説を行っているツイートを解説ツイートと呼ぶ。

4.1 機能表現を用いた解説ツイートの抽出

4.1.1 機能表現

ものごとを解説する際には、「選挙とは組織の代表を選挙権を持つ人が選出する行為のことである。」や「北海道には雪が降る日が多い。」といった文にあるような「とは」や「には」といったような表現を用いる。そのような表現に着目することにより、解説ツイートの抽出を行うことができると考え、機能表現を用いる。

機能表現とは、2 つ以上の語から構成され、全体として1つの機能的な意味を持つ表現である [9]。例えば、「僕は野球がしたい。」や「晴れているなら野球をする。」という文のように「たい」や「なら」といった、文に対して願望や順接などの意味を付与する表現である。この機能表現を用いる際には、松吉 [10] らの提案する機能表現辞書つつじを使用する。

4.1.2 解説ツイートの抽出

機能表現辞書内の意味の中で解説を行っている文に付与される意味として「話題」と「定義」が挙げられる。その表現の一部を表 1 に示す。これらの表現の入っているツイートから、「推量」「願望」「疑問」の機能表現を除いたツイートを解説ツイートとして抽出する。これは「推量」「願望」「疑問」は、筆者の考え、意見を述べる際に用いられるためである。「推量」「願望」「疑問」の表現の一部を表 2 に示す。

表 1 話題・定義の機能表現の一部
 というのは にしては に至っては かと言えば
 には とは となると となりや

表 2 推量・願望・疑問の機能表現の一部

でしょう	だろう	ちがいない	かもしれない
がいい	ねがいたい	たがる	たく
かしら	じゃないか	たっけ	

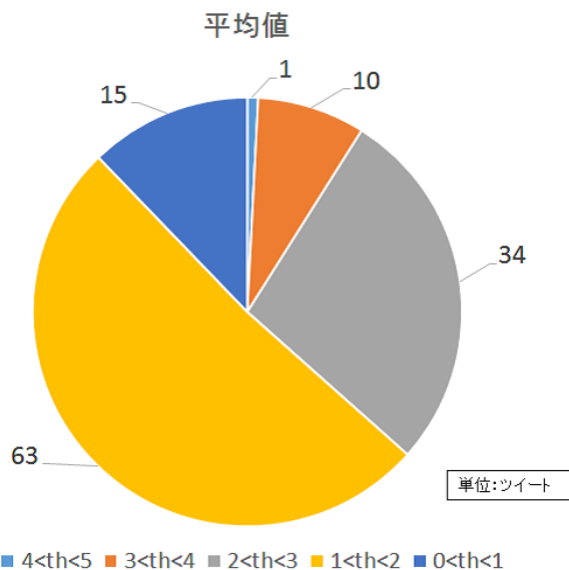


図 2 機能表現を用いた解説ツイートの抽出の評価値の割合

4.1.3 予備実験

ツイート中の機能表現に着目することにより、解説ツイートの抽出を行うことができるのかの予備実験を行った。

実験データ

大統領令をクエリとして取得したツイートから機能表現を用いて抽出したツイート 124 件を解説ツイートとして用いる。

被験者

機能表現を用いて抽出した解説ツイート 25 件あたり各々 10 人の被験者である。

実験手順

クラウドソーシングにより、実験を行った。まず、被験者は解説ツイートを見て、そのツイートが解説を行っている文であるかそうでない文であるかを 0 から 5 の 6 段階で評価する。この作業を 25 件のツイートごとに 10 人の被験者に行ってもらう。

4.1.4 結果・考察

全体のツイートごとの平均値の割合を図 2 に示す。正解データは、評価値の平均が 3.0 以上または、評価値の平均が 2.5 以上かつ分散値が 1 以下のものとする。正解データの数 19 個であり、適合率は 0.153 となった。

図 2 からわかるように、0 から 2 の間の評価値のツイートの数が多くなっている。これらのツイート表現には、顔文字や、感嘆符、笑いを示す「w」等が含まれていることが多い傾向にある。また、クエリを「大統領令」としてツイートを取得しているが、評価値の高いものの中には、「大統領令」と異なる語がツイート中の主題となってしまう場合がある。

これらのことから、機能表現を用いた抽出手法では「大統領令は無効のはず！」や「米国大統領令は報道機関で翻訳しない

表 3 解説ツイートの条件一覧

番号	条件
(1)	主題語+「は」または「が」、主題語+格助詞+「は」の形が含まれる。
(2)	文末に終止形、過去形の表現が含まれる。
(3)	〇〇一、〇〇～を含まない。(〇:ひらがな)
(4)	!, ?, (笑), ww を含まない。
(5)	ひらがなの小文字を含まない。
(6)	連続のひらがなを含まない。
(7)	一人称の名詞を含まない。
(8)	顔文字を含まない。
(9)	文字数が 30 文字以上。
(10)	文末に思考を表現する動詞を含まない。

で米国での弁護士資格を持った人が正確に翻訳して欲しい。」といった意見や願望を含むツイートが抽出されてしまう。また、主題語を含むツイートをを用いて抽出しているが、主題語に対する解説を行っていないツイートも抽出されてしまう。そこで、機能表現のみだと解説ツイートを抽出するのは困難であると考え、さらに文章表現に着目し、解説ツイートの抽出を行う。

4.2 文章表現を用いた解説ツイートの抽出

文章表現に着目して解説ツイートの抽出を行うに当たり、予備実験の結果を分析し、以下の 10 個の条件を抽出した。これらの条件すべてを満足したツイートを解説ツイートとする。

4.2.1 文章表現を用いた解説ツイート抽出の仮説

予備実験の結果から、得られた 10 の解説ツイートの条件を表 3 と以下に示す。

条件 1: 主題語+「は」または「が」、主題語+格助詞+「は」の形が含まれる。

物事を解説する際には、必ず解説対象と格助詞の「は」、または、解説対象と「は」の間に格助詞が連続して登場するケースが多い。例えば、「ノロウイルスには、次亜塩素酸ナトリウムが有効である。」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感染予防ができる」といった文である。

条件 2: 文末に終止形、過去形の表現が含まれる。

解説を行う際に、文末が終止形、もしくは過去の形であれば、解説の対象の動作が確実に行われること、終了したことであるということの判断が可能であるためである。例えば、「僕は免許を取った。」や「僕は免許を取る。」といったツイートは、文末が基本形または過去の形になっている。一方、「僕は免許をとりたい」、「僕が免許を取ればな」といったような文末が基本形以外で終わっている場合は願望や推定などを表す語が含まれ、主観的な意味になることが多い。

条件 3: 〇〇一、〇〇～を含まない。(〇:ひらがな)

ひらがなの後に「一」や「～」といった表現が続く場合は、主観的な表現に多く、解説をしている文書には出現していない。例えば、「明日は雨かなー?」といったツイートや「今日は何しようかなあ～」といったツイートである。

条件 4: !, ?, (笑), ww を含まない。

解説をしているツイートは客観的な表現であると考えられるため、主観的なツイートに多く見受けられる!や?等の記号や

表4 思考動詞一覧

思う	わかる	考える	知る	感ずる
感じる	願う	期待する	予想する	

表5 使用した主題語一覧

トランプ大統領	大統領令	TPP
メキシコ	天降り	尖閣諸島 横綱

表現を含んでいない。例えば、「明日の天気は晴れだ！うれしい！」といったツイートや、「今日は雨が降るのではないだろうか？」といったツイートである。このように、感嘆符や疑問符といったものは主観的な表現に多い。

条件5： ひらがなの小文字を含まない。

「あ、い、う、え、お、わ」といった通常使われないひらがなの小文字は、口語的な表現であるため、解説をしているツイートには含まれない。例えば、「つかれたあ」といった表現や「ええ…」といったような表現である。

条件6： 連続のひらがなを含まない。

話し言葉でのみ用いられる表現は解説を行っている文には現れていない。例えば、「うわああああ」といった表現や、「おおおおお!!!」といった表現である。

条件7： 一人称の名詞を含まない。

「私」や「僕」といった一人称の名詞は主観的な表現に多く見られ、解説を行っている文には現れていない。例えば、「私はノロウイルスにはかからない。」といったツイートや、「僕は、インフルエンザではないけど、ノロウイルスにかかった。」といったツイートである。

条件8： 顔文字を含まない。

顔文字を含む文書は、主観的な文書が多いため、解説を行っているツイートには出現しない。例えば、「腹痛すぎて寝れない(´д`)」といったツイートや、「ノロウイルスは嫌だ。(´ω`)」といったツイートである。

条件9： 文字数が30文字以上。

文字数が少ない文書は、解説を行っている文書とはいえない。例えば、「安倍首相とトランプ大統領がゴルフ。友情が築ければ良いですね。」といったツイートや「これでもトランプ大統領は支持されてるといふから怖い」というツイートである。このようなツイートは情報が少なく、解説であるとはいえない。

条件10： 文末に思考を表現する動詞を含まない。

解説の中には主観的な内容が含まれていることが少ないため、主観的な内容が書かれているツイートに多く含まれる表4中の動詞が含まれるか含まれないかを条件とする。これにより、客観的な内容のツイートであるかの判定が可能である。例えば、「明日は晴れると思う。」といった表現や、「テストでいい点が取れると期待する。」といったように表4にある動詞が用いられている文は主観的な表現であるといえる。

尚、表3条件10における動詞を思考動詞と呼び、思考動詞の一覧を表4に示す。

上記すべての条件を満たすものを解説ツイートとし、主題語を含むツイート群から抽出を行う。

5. 実験

本論文では、ニュース記事から主題語を抽出し、主題語を含むツイートから解説ツイートを抽出する手法の提案を行った。

機能表現の有用性の実験は、4.1.3節の予備実験にて行ったため、提案手法の有用性を図るために以下の2つの実験を行った。

- (1) 文章表現のみを用いた解説ツイートの抽出の実験
- (2) 機能表現と文章表現を用いた解説ツイートの抽出の実験

5.1 文章表現のみを用いた解説ツイートの抽出の実験

文章表現のみを用いた解説ツイート抽出のための10の条件の有用性を図るために文章表現のみを用いた解説ツイートの抽出の実験を行った。

5.1.1 実験条件

実験データ

抽出したニュースの主題語を用いて、主題語を含むツイートを収集し、解説ツイートを抽出する。主題語は無作為に抽出した7つの主題語である。ツイートは各々の主題語に対して、100件のツイート、合計700件のツイートである。使用した主題語を表5に示す。

被験者

抽出した解説ツイート25件あたり10人の被験者である。

5.1.2 実験手順

具体的な実験手順を以下に示す。実験にはクラウドソーシングを用いた。

- (1) 無作為に抽出した主題語を含むツイート群から文章表現を用いて解説ツイートを抽出。
- (2) 抽出したツイートを被験者に提示し、解説ツイートであるかを0-5の6段階で評価。
- (3) 各主題語に対し、(1)から(2)の処理を行う。

実験では、抽出した主題語に対する解説ツイートを被験者に提示し、そのツイートが解説を行っていれば1-5の5段階で評価、解説を行っていなければ0と評価してもらう。

5.1.3 結果と考察

10人の被験者による評価値の平均を用いて、実験の結果を示す。平均値、分散値により結果の評価を行う。平均値、分散値の計算の際に、被験者の回答の合計値を算出し10人の被験者の中で最も合計値が高い被験者と低い被験者は削除した8人の被験者の平均値、分散値を用いる。これは、被験者ごとの値のばらつきを少なくするためである。

正解データは評価値の平均が3.0以上または評価値が平均2.5以上かつ分散値が1.0以下のものとする。正解データの数は、194個であり、適合率は0.277となった。4.1節で行った機能語を用いた解説ツイートの抽出の予備実験では、適合率が0.153であったため、10の条件を用いることにより、適合率を上げることができた。

全体のツイートごとの平均値の割合を図3に示す。結果としては、評価値の平均が0から2となったものが約60%となった。一方、評価値が3.0を超えたものは数が少ないという結果になった。また、各話題ごとの評価値の分布は図4に示す。こ

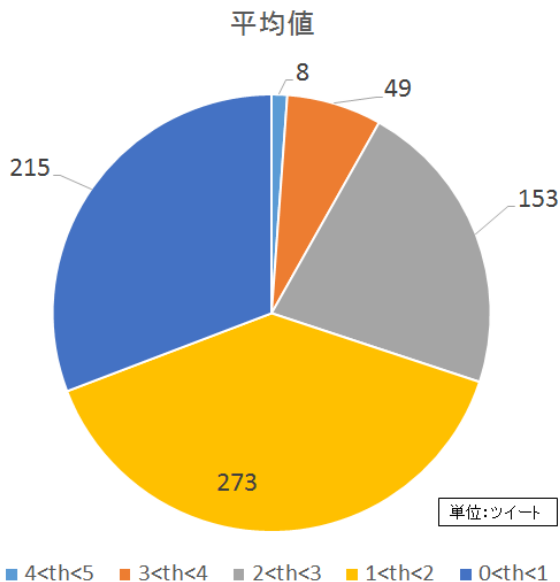


図3 文章表現を用いた解説ツイートの抽出の評価値の割合

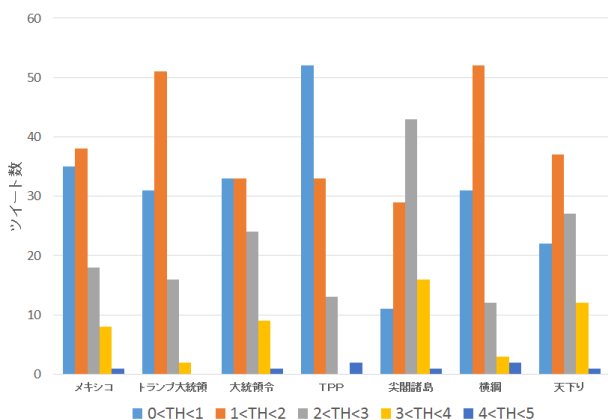


図4 各話題ごとの評価値の平均の分布

のことからわかるように、各話題に対し、数は少ないが解説ツイートの抽出は行うことができていることがわかる。抽出できた解説ツイートの一部を表6に示す。

以下に各条件についての考察を述べる。

条件1： 主題語+「は」または「が」、主題語+格助詞+「は」の形が含まれる。

条件1を用いることにより主題語に対する解説ツイートを抽出できた。例えば、主題語が「横綱」の場合、「横綱とは、元々は格付けではなくて、土俵入りの時に付けるあの白い綱のことを指していた。」といったツイートや「日本人横綱は19年ぶりなんだ。」といったようなツイートである。

条件2： 文末に終止形、過去形の表現が含まれる。

文末に終止形、過去形の表現がある場合、「税金の無駄遣いになる。」や「世界中で広く行われている。」といったように、実際に物事が起きている状態や、既に物事が発生した状態を示す解説ツイートを抽出することができた。

条件3-8 条件3-8に関しては、くだけた文章に含まれる表

現である。例えば、表7に示すツイートのようなものである。条件3-8を削除することにより、このようなくだけた表現を取り除くことができた。

条件9 文字数が30文字以上。

文字数が30文字以下の抽出結果の評価値の平均を見てみると、表8のようになった。このように、文字数が少ないツイートは情報が少なく、多くの人に低い評価をつけられる傾向がある。したがって、文字数が30文字以上のものが解説ツイートとして選ばれる事が多いといえる。

条件10 文末に思考を表現する動詞を含まない。

個人の考えを含むツイートには、思考を表現する動詞が登場した。例えば、「ものすごい勢いで署名されていく大統領令が及ぼす影響を、政治経験のないトランプも周りも知らないのではないかと思う。」といったツイートや「トランプ大統領の入国禁止令。米企業が難民を助ける策を発表していることは良いと思う。」というツイートである。条件10を用いることにより、このようなツイートを省くことができた。

これらのことから、条件1-10は解説ツイートの抽出には有用であるといえる。しかし、適合率が0.277と非常に低い結果となってしまった。これは、解説ツイートを抽出する際に、動詞の後ろに現れる機能表現により付与される願望や疑問の意味を取り除くことができなかつたからであると考えられる。

5.2 機能表現と文章表現を用いた解説ツイートの抽出の実験

本論文では機能表現と文章表現を用いてTwitterからニュースの主題に対する解説ツイートを抽出する手法を提案した。この我々の提案する手法の有用性を示すことを目的として、機能表現と文章表現を用いて解説ツイートを抽出する実験を行った。実験条件は5.1.1節と同様である。

5.2.1 実験手順

具体的な実験手順を以下に示す。

- (1) 5.1節で評価されたツイートに対し機能表現を用いて、「推定」「願望」「疑問」の意味を持つツイートを抽出。
 - (2) 抽出したツイートの評価値を確認。
 - (3) 各主題語に対し、(1)から(2)の処理を行う。
- 実験では、抽出されたツイートの確認を行う。

5.2.2 結果と考察

抽出された結果を表9に示す。実験に用いた700ツイートから表9のようなツイートを抽出した結果、107ツイートを抽出することができた。適合率は0.317である。4.1.3節の予備実験での機能表現のみを用いた解説ツイートの抽出の適合率が0.153であり、5.1節の文章表現のみを用いた解説ツイートの抽出の適合率が0.277であることより、機能表現と文章表現両方を用いることにより、適合率を上げることができた。これにより、機能表現と文章表現両方を使用してTwitterから解説ツイートの抽出を行うことは有用であるといえる。

6. まとめと今後の課題

本研究では、ニュース記事の理解支援のために、ニュースの主題語を抽出し、主題語に対する解説を行っているツイートを

表 6 解説ツイートの一部

- ・横綱は、大勢いる力士の中で唯一注連縄を付けることを許された存在、つまり神の依り代ということ。神様なんだから、そりゃあ簡単にはなれんわなあ。因みに横綱とは、元々は格付けではなくて、土俵入りの時に付けるあの白い綱のことを指していた。
- ・マティス米国防長官が3日の安倍晋三首相との会談で、沖縄県の尖閣諸島が米国の対日防衛義務を定めた日米安全保障条約第5条の適用対象になると明言したことが分かった。政府高官が明らかにした。
- ・大統領令は大統領が議会の承認を待たずに政府や軍に直接発令する行政命令。法律と同等の効力を持つ。議会は反対する法律を作って対抗でき、また最高裁判所も違憲判断を出すことができる。トランプの入国制限に対し世界中から非難が。違憲訴訟の動きも。トランプが発言の一部修正の声明。

表 7 くれた表現のツイート

ツイート

- ・グローバル人材企業の会社は軒並みトランプ氏の大統領令に反対意見なのは当然だよなー
- ・今日のお昼のニュースはトランプ大統領令づくしだった！
- ・大統領令訴えられるてwww
- ・(´-`)oO (TBS はアメリカの大統領令に対抗する必要があるのか…
- ・なんで大統領令が悪いものみたいになってるの？w
- ・アメリカのトランプ大統領、私は好きになれない。

表 8 30 文字以下のツイート

ツイート

- ・TPP がどうしても PPAP にしか聞こえない。
- ・大統領令が周りの人に影響が及ぶと流石に無関心ではられない
- ・TPP はぶっちゃけアメリカにはあんまり利益なかったよね
- ・米国はヒラリーが勝っても TPP は離脱してた。
- ・大統領令がこわい。これがしばらく続くのか。
- ・ちょっと感動しちまった。やっぱ横綱は日本人じゃなきゃ。

表 9 推定・願望・疑問の意味を持つツイート

ツイート

- ・数々のアメリカ大統領令が各国の最悪な結末を迎えなければいい
- ・TPP には賛成でも二国間の貿易協定には消極的な人は多いだろう
- ・行政にろくな利権がなければ、天下りは問題ないし天下りも起きないだろう。
- ・日本とか中国に対する課税は内需拡大になるかも (可能性低い) しれないけど、メキシコは壁の費用に当てられるとかトランプの頭はわけがわからないよ
- ・壁って進撃の巨人みたいなことかと思ったけど違ったな。まあメキシコが悪いんだろうね

抽出する手法を提案した。解説ツイートの抽出には文章表現と機能表現を用いた。

今後の課題は、適合率が低いため、条件を追加し、精度をあげることである。また、解説というあいまいなものを抽出しようとしているため、機械学習を用いた抽出を行うことも思案している。抽出した解説ツイートがニュースの理解につながる内容になっているのかという実験を行うことも必要である。これは、解説ツイートを抽出したが、抽出したツイートが理解支援につながるかどうかというものはわからないため、実験の必要があるといえる。そして、抽出した解説ツイートの中から、Wikipedia には載っていない重要な情報を抽出することである。

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] 馬強, 田中 克己, “補完情報の検索に基づくコンテンツ統合”. 情報処理学会研究報告データベースシステム (DBS), 2004, 72(2004-DBS-134),337 - 343,2004-07-14.
- [2] 北山 大輔, 角谷 和俊, “ニュースアーカイブのためのコンテンツ構成順序を用いた比較ニュース検索”. 電子情報通信学会 第 18 回データ工学ワークショップ (DEWS2007), A9-4, 2007.
- [3] 田中 祥太郎, ヤフトアダム, 田中 克己, “事象と記事の主題・背景関係に基づくニュース記事の理解支援”. 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2016), C1-3, 2016.
- [4] 田中 祥太郎, ヤフトアダム, 田中 克己, “ニュース記事の理解支援のための背景知識抽出と補完”. 研究報告情報基礎とアクセス技術 (IFAT), 2002,104(2002-NL-152),145 - 152, 2002-11-12.
- [5] 佐藤 由紀, 横本 大輔, 牧田 健作, 宇津呂武仁, 福原 知宏, “ニュース記事中の話題に関連するブログ記事の収集手法”. 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2011), A6-3, 2011.
- [6] 小林 透, 柴田 和樹, 中山 僚, “マルチコンテンツによる「ニュースの言葉」自動生成システムの研究”. マルチメディア、分散協調とモバイル (DICOMO2014) シンポジウム, 2014, 85 - 92,2014-07-02.
- [7] 池田 大介, 藤木 稔明, 奥村 学, “blog とニュース記事の自動対応付け”. 言語処理学会第 11 回年次大会論文集, pp,1030-1033,2005.
- [8] 大原 正章, 真下 遼, 灘本 明代, “Web ニュースからの観点抽出手法の提案”. 研究報告データベースシステム (DBS), 2015-DBS-162,27,1 - 6,2015-11-19.
- [9] 宇津呂武仁, 松吉俊, 土屋雅稔, 鈴木 敬文, 島内 蘭, “自然言語処理における日本語機能表現の解析”, 語彙・辞書研究会 第 38 回発表会 資料集, pp. 1-8, November 2010.
- [10] 松吉 俊, 佐藤 理史, 宇津呂 武仁 “日本語機能表現辞書の編纂”, 自然言語処理, Vol. 14, No. 5, pp. 123146,2007.