The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

ファクトチェックのための要検証記事探索の支援

Support for Searching for Articles to be Verified for Fact Checking

内山 香 *1 Kaori Uchiyama
鈴木 海渡 *1 Kaito Suzuki
田上 翼 *1 Tsubasa Tagami
塙 一晃 *1 Kazuaki Hanawa
乾 健太郎 *1*2 Kentaro Inui
小宮 篤史 *3 Atsushi Komiya
藤村 厚夫 *3 Atsuo Fujimura
町野 明徳 Akinori Machino
楊井 人文 *4 Hitofumi Yanai
山下 亮 *4 Ryo Yamashita

*1東北大学 Tohoku University
*2理化学研究所 革新知能統合研究センター RIKEN Center for Advanced Intelligence Project
*3スマートニュース株式会社 SmartNews, Inc.
*4一般社団法人日本報道検証機構 Watchdog for Accuracy inNews-reporting, Japan

False information have become a social problem and it is necessary to verify huge information such as news articles on the Internet. In this paper, we constructed a support system to efficiently verify the authenticity of news articles. This system estimates news articles that describe incorrect information based on posts that refer to them on SNS. We found that it can be expected to improve efficiency of searching for articles that require human verification by using the created system.

1. はじめに

国内外を問わず誤情報の拡散が社会的な問題となっている．アメリカの大統領選挙に関してメディアの報道やソーシャルメディア（SNS）上で投票に影響を与えうる多くの誤情報が拡散されたことや，国内でも健康・美容情報サイトにおいて人の健康に影響しうる誤った医療情報の記事が掲載されるなど，誤情報への対策が急務となっている．このような誤情報の拡散を防ぐためには政治家の言説，メディアの報道，ウェブ上のコンテンツや SNS 上の投稿など，社会に影響を及ぼしうる情報の真偽や正確性を検証するファクトチェックが必要である．

ファクトチェックの対象は政治分野だけでなく多岐に渡る．対象をニュース記事に絞っても日々配信される膨大な量のニュース記事全てに対し，人手で個々の内容を調査をすることは不可能である．そこで一般社団法人日本報道検証機構 *1 では以下の 3 つのステップによって内容を調査をするニュース記事を選別している．第 1 のステップではニュース記事に言及している SNS の投稿から (1) のような検証の必要性を示唆する情報（同機構ではこれを「端緒情報」と呼んでいる）を人手で収集する．第 2 のステップでは収集された端緒情報の内容や記事の影響度・深刻度等に基づいて，より詳細な本調査を必要とする記事（以下「要検証記事」と呼ぶ）を拾い上げる．最後に個々の要検証記事についてファクトチェックの本調査を行い，必要に応じて検証結果を記事化する．

(1) a. ○○新聞では縮小といってる。他紙と違って**不可解**ですが両方共ちゃんと裏付けしてるのか？> http://xxx

b. この記事は**誤報**では？千代田区も路上喫煙はダメで過料が科されているはずです！> http://xxx

このファクトチェックプロセスは，SNS 上の情報をある種の集合知として活用することによって要検証記事をできる限り網羅的に同定することをねらうもので，その有効性は日本報道検証機構のこれまでの実績にも示されている．ただし，膨大な SNS 情報から端緒情報を人手で探索する第 1 のステップが膨大な時間を要する作業となっており，この部分がボトルネックとなって要検証記事の探索コストを押し上げていることが深刻な問題となってきた．端緒情報の探索では，SNS の投稿を「誤報」や「デマ」などのフレーズでフィルタリングしてはいるが，それでもフィルタを通過した投稿のほとんどは (2) のようにファクトチェックには役立たない投稿である．実際に，2017 年に日本報道検証機構とファクトチェック・イニシアティブ *2，賛同メディアが連携して実施した衆議院議員総選挙ファクトチェックでは，総選挙関連の Twitter の投稿を 1 日あたり約 16,000 件人手でチェックしたが，そのうち端緒情報と認められた投稿は平均わずか 12 件であった．

(2) a. 本当に信じられない。**嘘**であって欲しい。言葉が見つからないけどご冥福をお祈りします。! > http://xxx

b. ○○○○大統領との親密さをアピール？札束でアピールの**間違い**だ !! > http://xxx

そこで本研究では，この端緒情報の収集を自動化し，人手による要検証記事探索作業を技術的に支援する仕組みの構築を考える．具体的には，代表的な SNS である Twitter*3 の投稿を対象に，そこから端緒情報である可能性が高い投稿を自動抽出し，それらの情報に基づいてニュース記事を検証必要度の観点からランキングする．これによって検証必要度が高い記事をチェックするだけで要検証記事を網羅的に収集できるようになると考えられる．本稿では，要検証記事の探索支援に向けた研究開発の途中経過を報告する．なお我々が調べた限りでは，インターネット上の誤情報を検出する研究は行われているものの [1, 4]，Twitter の投稿から端緒情報を抽出し，要検証記事を収集する研究は本研究が初である．

連絡先: 田上 翼，東北大学，宮城県仙台市青葉区荒巻字青葉 6-6-05，022-795-7091，tagami@ecei.tohoku.ac.jp

*1 http://wanj.or.jp/

*2 ファクトチェック活動の支援を目的として 2017 年 6 月に設立され，2018 年 1 月に NPO 法人を取得． http://fij.info/

*3 https://twitter.com/

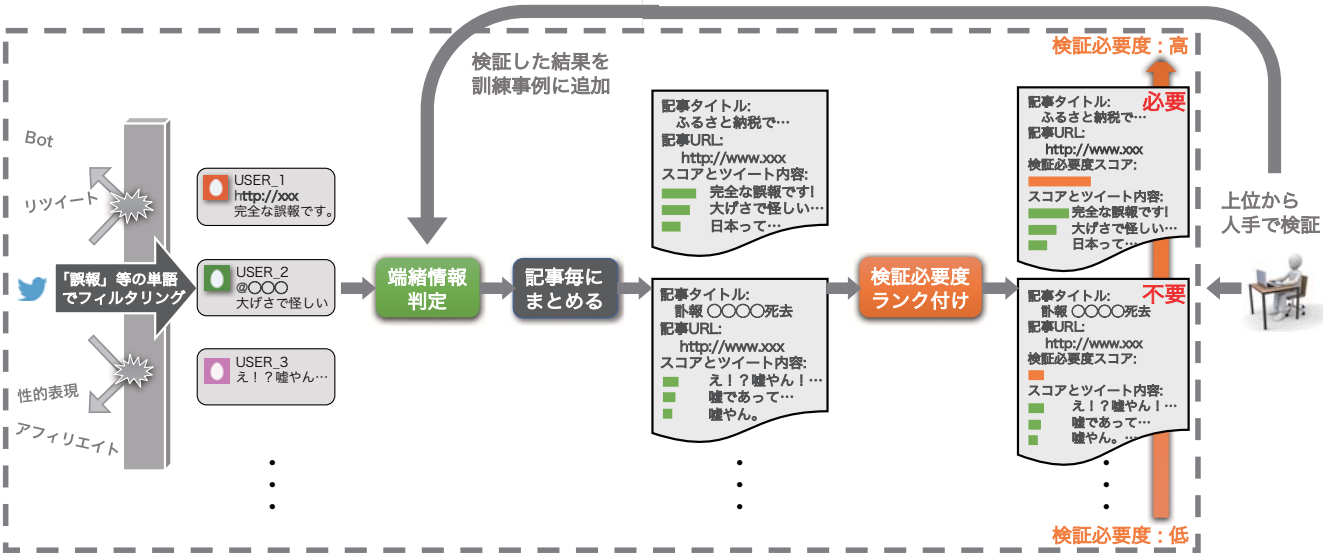

図 1: ファクトチェック支援システムの構成

2. 要検証記事の探索

要検証記事を探索するために用いる手法の概略を図 1 に示す．手順はツイートの収集，端緒情報の判定，記事の検証必要度のランク付け，人手による検証の 4 つに分けられる．

2.1 ツイートの収集

対象のニュースサイトで掲載された記事へ言及しているツイートを以下の 3 つの方法を用いて収集した．また収集対象は 2017 年 1 月から 6 月のツイートとした．

URL を含むツイート

ツイートが言及している対象の記事を明確にするため，対象のニュース記事の URL を含むツイートを収集した．

リプライツイート

Twitter では特定のツイートに対する返信をリプライという．いくつかのニュースサイトでは Twitter アカウントを所有しており，記事を配信すると同時にその記事についてツイートをして周知をすることがある（以下ではこのツイートを「ニュースツイート」と呼ぶ）．このツイートに対するリプライツイートは記事に対して言及していると考えられるため，ニュースツイートに対するリプライツイートを収集した．

リツイート直後のツイート

Twitter では他のユーザのツイートを共有するリツイートと呼ばれる仕組みがある．図 2 に示すように，あるユーザがリツイートした直後にそのツイートに関連する内容のツイートをすることがある．そこでニュースツイートをリツイートしたユーザのリツイート直後のツイートを収集した．収集したツイートには記事に関連していないものも含まれるため，1016 件のツイートを対象にリツイートとの関係の有無を調査したところ，リツイートとその直後にしたツイートまでの時間に図 3 に示す関係があることがわかった．図 3 より記事に関係する内容のツイートは時間が経つごとに減少することがわかる．また 240 秒を超えた付近で関係のあるツイートの割合が大きく下がるため，本研究では 240 秒以内に投稿されたツイートのみを収集した．

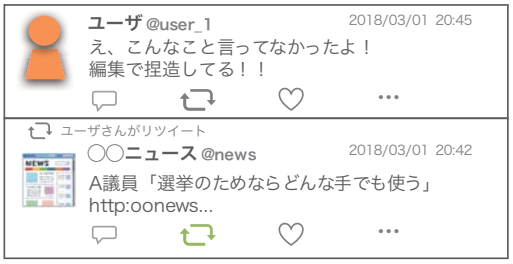

図 2: リツイート直後のツイートの例

2.2 端緒情報の判定

端緒情報を抽出するために，2.1 節で収集したツイートから日本報道検証機構が実際に用いている表 1 のフレーズでフィルタリングした．それぞれの手法で収集されたツイートのフィルタリング後の数は URL を含むツイートが 468,188 件，リプライツイートが 17,232 件，リツイート直後のツイートが 18,429 件であった．また，ニュース記事について言及しているツイートには，(3) のように記事のタイトルや本文の一部が引用されることが多い．解析の対象としたいのは記事のタイトルや本文を除いたコメント部分であるため，言及している記事のタイトルと本文を獲得しツイートとの重複部分を除去する．この他にツイート中の URL とハッシュタグ，記号および英数字を除去する．またリツイートや Bot によるツイートなどは除外する．次に，収集した各ツイートを 3. 章で述べる端緒情報の分類器で判定し，端緒情報である確率によりスコア付けする．本研究ではこの判定を行う分類器を作成したデータセットを用いて構築し，その性能を評価した．

(3) 〇〇でデマを流す男、〇〇。2018 年の ACL 鹿島はストレートインだろうが。＞サッカー次期代表監督、外国人に限定しないで: 〇〇〇〇新聞 http://xxx

2.3 記事の検証必要度ランク付け

端緒情報か判定されたツイートを記事ごとにまとめ，各ツイートのスコアを用いてそれぞれの記事の検証必要度をランク付けすることで，要検証記事の探索を支援する．本研究ではこ

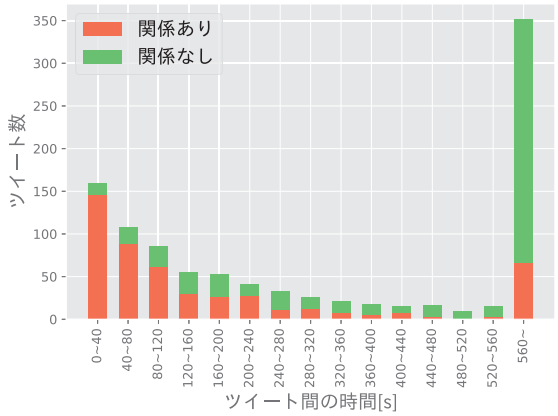

図 3: ツイート間の時間ごとのリツイートとその直後のツイートの関係の有無

表 1: フィルタリングに用いるフレーズ一覧

怪しい	偽	偽装	違和感	嘘	噂	改竄	ガセ	疑い	疑惑	記事について
誤り	誤解	誤認	誤報	誤用	根拠	信用しない	信用できない	ステマ	デマ	とんでもない
捏造	パクリ	否定	フェイク	不可解	偏向	変造	マスゴミ	間違い	ミスリード	無根

の記事の検証必要度のランク付けによる要検証記事の分類性能も評価した．

2.4 人手による検証

実際にファクトチェックを行う時にはシステムの出力を人手でチェックし，要検証記事であるかを判定する．このときシステムが判断した要検証記事と端緒情報を人手で修正することにより，学習のトレーニングデータを拡充することができる．この仕組みにより端緒情報の判定と記事の検証必要度をランク付けするモデルの性能改善が期待される．

3. 実験

人手でラベル付けを行ったデータセットを用いてツイートが端緒情報であるか判定する分類器を構築し，その性能を評価した．また構築した分類器を用いて記事の検証必要度をランク付けし，その性能についても評価した．

3.1 データセット

ツイートの端緒情報判定に用いるためのツイート単位データセットと，ニュース記事の要検証記事分類に用いるための記事単位データセットの 2 つを人手で作成した．

3.1.1 ツイート単位データセットの作成

2.1 節で述べた 3 つの方法で収集した各ツイートを人手で端緒情報であるか判定した．これらのツイートうち正例のツイートの数は 1036 件で URL を含むツイートで 50,000 件中の 606 件（1.21%），リプライツイートで 5173 件中の 423 件（8.17%），リツイート直後のツイートで 508 件中の 7 件（1.37%）であった．また負例については正例でないツイートのうち bot によるツイートや，記事の本文やタイトルを引用しているだけのツイート，リンク切れなどの要因で記事の本文やタイトルとの重複部分を自動で除去できないツイートを人手で取り除いたものを使用した．負例のツイートの数は 6739 件でその内訳は URL を含むツイートで 50,000 件中の 2914 件（5.82%），リプライツイートで 5173 件中の 3468 件（67.0%），リツイート直後のツイートで 508 件中の 357 件（70.3%）であった．

表 2: 端緒情報判定の性能

モデル	適合率	再現率	F 値
ロジスティック回帰	0.66	0.58	0.62
LinearSVM	0.66	0.58	0.62
決定木	0.49	0.48	0.48
ランダムフォレスト	0.67	0.39	0.49
LSTM	0.62	0.55	0.59

3.1.2 記事単位データセットの作成

ツイートが言及している記事 URL ごとにまとめたツイートの集合に，端緒情報であるツイートが 1 件でも含まれている記事を正例とし，ツイート単位データセットと同様に人手で判定した．正例である記事は 564 記事であり，また負例は正例でない記事のうち 1,271 記事を用いた．

3.2 手法

ツイートが端緒情報であるかを判定する分類器を，以下の 5 つのモデルで構築した．形態素解析には，形態素解析器 MeCab-0.996[*4] と，辞書として mecab-ipadic-neologd[*5] を使用した．モデルの実装には機械学習ライブラリ scikit-learn[*6] と深層学習ライブラリ Keras[*7] を使用した．

1. ロジスティック回帰
2. LinearSVM
3. 決定木
4. ランダムフォレスト
5. LSTM[2]

モデル 1 から 4 については素性に Bag-of-Words と Bag-of-Bigrams を使用した．また，前処理として助詞を全て取り除き，また形態素解析の結果から基本形の数を数え，ストップワードとして頻度の高いの単語を利用した．モデル 5 については，単語の分散表現は，表 1 のフレーズを含む約 450 万件の日本語ツイートを抽出し，word2vec[3] を用いて学習した．次元数は 300 とした．

3.3 実験 1：端緒情報判定

上述の 5 つのモデルを用いて端緒情報の分類器を構築し，それぞれの手法の適合率と再現率，F 値を評価するために，5 分割交差検証を行なった．またそれぞれの手法について，今後データを増やすことで分類器の性能向上を期待できるかを確認するために，トレーニングデータのサイズを 1/8，1/4，1/2 と変化させた場合の F 値の変化についても計測した．

表 2 に各手法で構築した端緒情報分類器の適合率と再現率，F 値を示し，図 4 にトレーニングデータのサイズを変化させた時の学習曲線を示す．表 2 より，ロジスティック回帰と LinearSVM が F 値で最も高いスコアを示した．LSTM 以外の 4 つの手法では素性として Bag-of-Trigtams も用いる場合や，Bag-of-Words のみを用いた場合でも同様の実験を行なったが，Bag-of-Words と Bag-of-Bigrams を用いた場合が最も優れたスコアを示した．次に図 4 より，各手法でトレーニングデータのサイズを大きくしていくと，F 値は緩やかに上昇しており，まだ飽和は生じていないことがわかる．このことから実際にシステムを利用してトレーニングデータを拡充することにより，いずれの手法でも端緒情報分類器の性能向上が期待できる．

*4 https://taku910.github.io/mecab/
*5 https://github.com/neologd/mecab-ipadic-neologd
*6 http://scikit-learn.org/stable/
*7 https://keras.io/

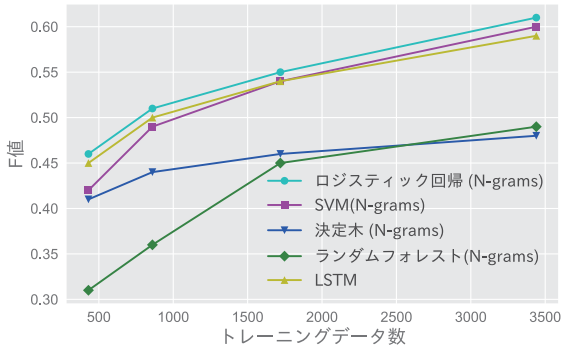

図 4: トレーニングデータサイズを変えた時の F 値

表 3: 端緒情報判定の誤り例：+1 は端緒情報であり，-1 は端緒情報でないことを表す．

ツイート例	予測	正解
(1) ○○の「県民の意向」は、聞き飽きた … 何処の県民？河北とか四川とか、○○省の間違いだろ	+1	-1
(2) 金の掛かる万博、本当にやりたいんだろうか。東京みたいにミスリードするな。	+1	-1
(3) 遼寧は普通に 30 ノット出してたし格別遅くはないぞ。嘘はよくないな。	-1	+1

いずれの手法でも正しく判定できなかったツイートを表 3 に示す．表中の (1) と (2) はいずれの手法でも誤って端緒情報であると判断したツイートの例であり，(3) はいずれの手法でも誤って端緒情報でないと判断したツイートの例である．例 (1) は単純に皮肉を言っているものであり，例 (2) はミスリードを指摘しているが，指摘の対象が記事ではなく行政などを指すものである．このようなツイートを正しく判定するには，ツイートが言及する記事などの外部知識を用いて，皮肉の検出や指示対象の特定などの深い処理が必要であると考えられる．次に，例 (3) のツイートに含まれる，「嘘」という単語はフィルタリングに用いた表 1 の他の単語に比べて日常的にツイートに使われやすく，データセットの負例中の約 24%のツイートに含まれている．このため「嘘」を含むツイートは単純な手法では誤って端緒情報でないと判断されやすいのだと考えられる．

3.4 実験 2：検証必要度に基づく記事のランク付け

3.3 節で用いた 5 つのモデルによる端緒情報分類器を利用し，3.1 節で作成したデータセット上で記事の検証必要度をランク付けする実験を行った．各記事の検証必要度のスコアは記事毎にまとめられた各ツイートの端緒情報分類器によるスコアのうち最も高い値を利用した．評価尺度には，要検証記事の分類性能を評価するために適合率と再現率，F 値を用いた．また，十分な量の要検証記事を獲得するには上位何件の記事を検証すれば良いかを調べるために Recall@K による評価も行った．

表 4 に適合率と再現率，F 値を示し，図 5 に Recall@K を示す．図 5 より，正例と負例が 3:7 の割合である記事単位データセットにおいて，端緒情報分類器としてロジスティック回帰と linearSVM と LSTM を用いた場合では，検証必要性スコア上位の 26%の記事を見れば 6 割，45%を見れば 8 割の正例を獲得できることがわかる．このように今回作成したデータセット上では有効なことは確かめられたが，実際は要検証記事は全体の数%であるため今後はデータセットを拡充し実際の条件に近づけたときの性能を調べる必要があると考えられる．また本研究では各記事の検証必要度のスコアを，記事毎にまとめられた各ツイートの端緒情報分類器によるスコアのうち最も高いものを利用したが，この方法では記事に対するツイート数などを考慮しておらず，評価指標を検討し直す必要があると考えられる．

表 4: 要検証記事の分類性能

モデル	適合率	再現率	F 値
ロジスティック回帰	0.73	0.62	0.67
LinearSVM	0.74	0.63	0.68
決定木	0.62	0.34	0.44
ランダムフォレスト	0.81	0.34	0.48
LSTM	0.64	0.75	0.69

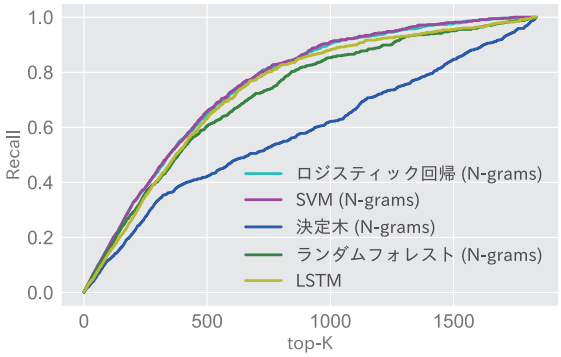

図 5: 検証必要度のランク付け性能

4. おわりに

本研究ではファクトチェックを支援するために，要検証記事の探索を支援するシステムの構築に取り組んだ．システム中では端緒情報の判定と記事の検証必要度をランク付けするために，端緒情報の分類器を構築した．この分類器を用いることでデータセット上においては要検証記事の探索作業の効率化を期待できることが確かめられた．今後は実際にこの仕組みによるシステムをファクトチェック業務に利用し，端緒情報と要検証記事のデータセットの拡充や学習のアルゴリズムの改良を進め，要検証記事の検出性能の向上を目指す．また収集されたツイートの言語的特徴の分析や実用上で生じる課題の検討に取り組む予定である．

謝辞

本研究の一部は NTT コミュニケーション科学基礎研究所，JSPS 科研費 15H01702，東北大学 Step-QI スクールの支援を受けた．

参考文献

[1] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1803–1812. ACM, 2017.

[2] Sepp Hochreiter and J Urgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

[3] Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[4] 宮部真衣, 梅島彩奈, 灘本明代, 荒牧英治. 流言情報クラウド: 人間の発信した訂正情報の抽出による流言収集.