

# ファクトチェックを必要とするニュース記事の探索の支援

田上 翼<sup>†1</sup> 浅野 広樹<sup>††2</sup>

楊井 人文<sup>‡3</sup> 山下 亮<sup>‡4</sup> 小宮 篤史<sup>††5</sup> 藤村 厚夫<sup>††6</sup> 町野 明德<sup>7</sup> 乾 健太郎<sup>††8</sup>

<sup>†</sup> 東北大学 <sup>‡</sup> 一般社団法人日本報道検証機構 <sup>††</sup> スマートニュース株式会社 <sup>‡‡</sup> 理研 AIP

{<sup>1</sup>tagami,<sup>2</sup>asano,<sup>8</sup>inui}@ecei.tohoku.ac.jp {<sup>3</sup>yanai,<sup>4</sup>rymst}@gohoo.org  
{<sup>5</sup>atsushi.komiya,<sup>6</sup>atsuo.fujimura}@smartnews.com <sup>7</sup>akinori.machino@me.com

## 1 はじめに

イギリスの欧州連合離脱是非を問う国民投票や、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) 上で多くの誤情報が拡散された。拡散された誤情報の中には、投票に影響を与えるものや社会的な混乱を招くものもあり、国内外問わず誤情報への対策が急務となっている。これに対処すべく、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、社会に影響を及ぼす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、日々配信される大量のニュース記事のすべてをファクトチェックにかけるのは不可能である。そこで、一般社団法人日本報道検証機構<sup>\*1</sup>ではまず、表1のような、ファクトチェック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 SNS から探しだし、収集した端緒情報に基づいてファクトチェックにかけるべきニュース記事を選別している。このプロセスは大きく分けて3つのステップからなる。第1のステップでは、ニュース記事に言及している SNS の投稿から端緒情報を人手で収集する。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、該当記事の中から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実施し、必要に応じて検証結果を記事化する。

このファクトチェックプロセスは、SNS 上の情報が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のステップがそれ自体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

<sup>\*1</sup> <http://wanj.or.jp/>

表1 端緒情報である投稿の例

- 今までのほとんどの調査で厚生省の飲食店全面禁煙案支持が7割だったんだけど？どっちかが嘘ついてるってことね > <http://xxx>
- ○○新聞大誤報です。水道企業団は水の卸売のみ担当ではなく末端給水も一部担っています。記事の前提の事実認識が根幹から間違えています > <http://xxx>

表2 端緒情報でない投稿の例

- あ、日本は解散総選挙なの？アメリカみたいにロシアの操作とかはなくても、流行りのデマニュース流しまくり大会はやるんですよきっと。 > <http://xxx>
- ○○○○大統領との親密さをアピール？札東でアピールの間違いだ！！ > <http://xxx>

た投稿のほとんどは表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<sup>\*2</sup>、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 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。

そこで本研究では、この端緒情報の探索を自動化し、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<sup>\*3</sup>の投稿を対象に、そこから端緒情報の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。これによって端緒情報探索の人手作業が不要になり、さらに検証必要度のランク上位の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れば本研究はひとまず成功である。本稿では、この目標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, 2, 9]、幅広いドメイン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。

## 2 要検証記事の探索を支援する手法

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。手順は端緒情報の判定と記事の検証必要度

<sup>\*2</sup> 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://fj.info/>

<sup>\*3</sup> <https://twitter.com/>

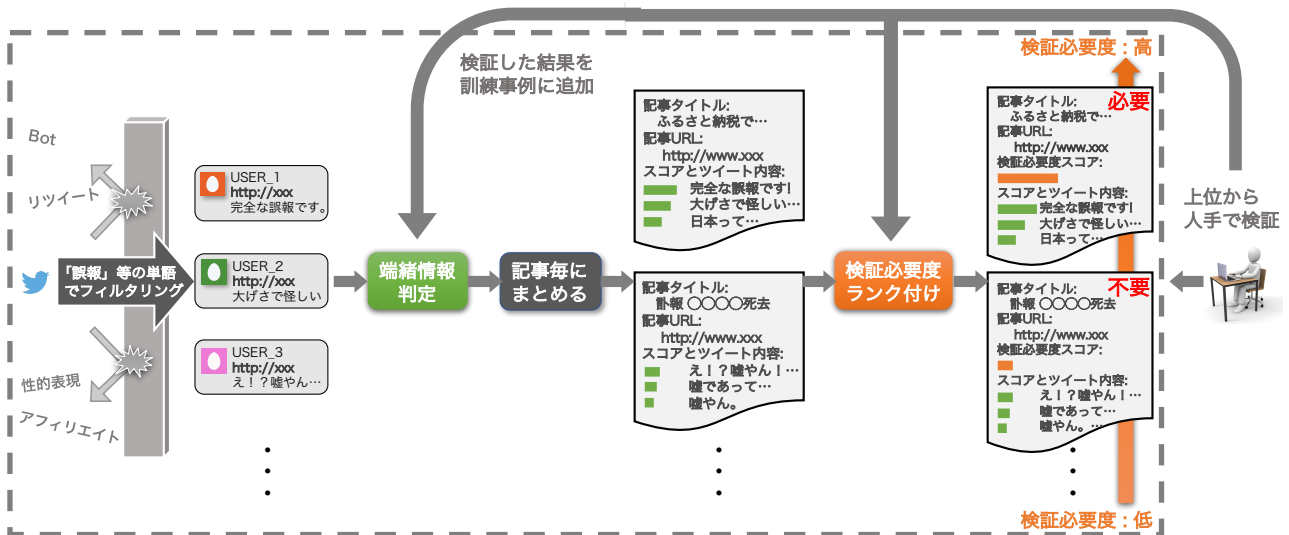


図1 ファクトチェック支援システムの構成

表3 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						

表4 記事タイトルを含むツイートの例

- ○○でデマを流す男、○○。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: ○○○○新聞 <http://xxx>

のランク付けの2つに分けられる。本節ではこの2つにステップについて説明を行う。

### 2.1 端緒情報の判定

端緒情報を抽出するために、Twitter上の投稿から日本報道検証機構が実際に用いている表3のフレーズと、記事などのURLを含むものを収集する。またニュース記事のURLを含むツイートには表4のような、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、URLから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。本論文ではこの判定を行う分類器を構築し、その性能を評価する。

### 2.2 記事の検証必要度ランク付け

端緒情報が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。本論文ではこ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する。

### 2.3 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検証し、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、訓練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 データセットの作成

本論文では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 ツイート単位データセットの作成

データセットを作成するために、朝日新聞デジタル<sup>4</sup>、産経ニュース<sup>5</sup>、日経電子版<sup>6</sup>、毎日新聞のニュース・情報サイト<sup>7</sup>、YOMIURI ONLINE<sup>8</sup>のいずれかの媒体の記事URLを含むツイートを20,000件抽出し、人手で各ツイートが端緒情報であるか判定した。正例の数は606件であり全体の3.03%であった。また負例は正例でなく、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でないツイートのうち2,851件を用いた。

### 3.2 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

<sup>4</sup> <https://www.asahi.com/>

<sup>5</sup> <http://www.sankei.com/>

<sup>6</sup> <https://www.nikkei.com/>

<sup>7</sup> <https://mainichi.jp/>

<sup>8</sup> <http://www.yomiuri.co.jp/>

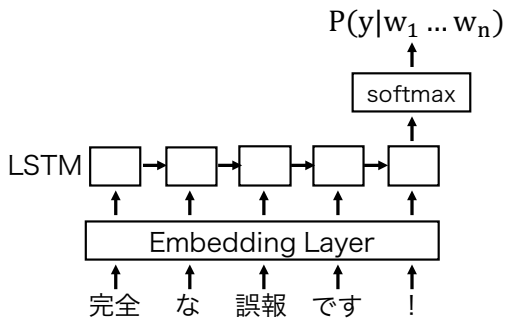


図2 LSTMを用いた分類器の概略図

も含まれている記事を正例とした。ツイート単位データセットと同様の条件で 20,000 件の記事 (50,333 ツイート) を抽出し、人手で判定した。正例である記事は 564 件であり全体の 2.82% であった。また負例は正例でない記事のうち 1,271 件を用いた。

## 4 実験

第3節で作成したデータセット用いて、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。

### 4.1 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つの手法で構築した。形態素解析には、形態素解析器 MeCab-0.996<sup>\*9</sup>と、辞書として mecab-ipadic-neologd<sup>\*10</sup>を使用した。すべてのモデルの実装には、深層学習ライブラリ Keras<sup>\*11</sup>を使用した。

- LR (N-grams) 素性に Bag-of-Words と Bag-of-Bigrams を用いたロジスティック回帰モデル。
- FFNN (N-grams) 素性に Bag-of-Words と Bag-of-Bigrams を用いた順伝播型ニューラルネットワークモデル。隠れ層は 2 層であり、活性化関数は ReLU[7]、出力層は softmax で、損失関数には交差エントロピーを使用した。またこの他に Dropout[8] と Batch Normalization[4] を用い、最適化には Adam[5] を使用した。
- LR (AE) ツイートに含まれる各単語の分散表現の平均を素性としたロジスティック回帰モデル。単語の分散表現は、表3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[6] を用いて学習した。次元数は 300 とした。
- FFNN (AE) ツイートに含まれる各単語の分散表現の平均を素性とした順伝播型ニューラルネットワークモデル。隠れ層は 3 層であり、それ以外は FFNN (N-grams) と同じ条件で学習した。
- LSTM ツイートに含まれる各単語の分散表現を入

<sup>\*9</sup> <https://taku910.github.io/mecab/>

<sup>\*10</sup> <https://github.com/neologd/mecab-ipadic-neologd>

<sup>\*11</sup> <https://keras.io/>

表5 端緒情報判定の性能

手法	適合率	再現率	F 値
LR (N-grams)	<b>0.75</b>	0.57	0.65
FFNN (N-grams)	0.72	0.59	0.65
LR (AE)	0.53	<b>0.75</b>	0.62
FFNN (AE)	0.70	0.57	0.62
LSTM	0.74	0.64	<b>0.69</b>

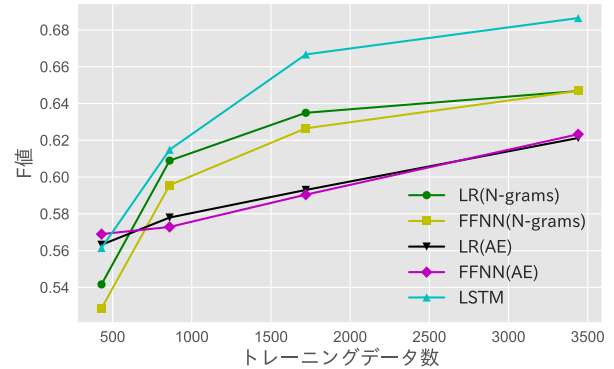


図3 トレーニングデータサイズを変えた時の F 値

力とする LSTM[3] モデル。上記の LR (AE) および FFNN (AE) では文脈の情報が失われてしまうので、文脈の情報も考慮するためにツイート中の単語の分散表現を、図2のように LSTM に入力する分類器を構築した。損失関数には二乗誤差を使用した。

### 4.2 実験1：端緒情報判定

上述の5つの手法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 値を計測した。またそれぞれの手法について、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5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3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5より、単語の分散表現と長期の文脈を学習できる LSTM が、F 値で最も高いスコアを示した。また1単語の文脈を考慮できる、LR (N-grams) と FFNN (N-grams) が文脈を考慮しない LR (AE) と FFNN (AE) よりも高いスコアを示した。このことより端緒情報の判定において、長期の文脈情報が有効であるといえる。また LR (N-grams) と FFNN (N-grams), LR (AE) と FFNN (AE) の比較より、端緒情報の判定では同じ素性を用いた場合に、ロジスティック回帰とニューラルネットワークで F 値での差異はほとんど生じないことがわかる。

次に図3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 値が緩やかに上昇していることがわかる。この結果から学習するデータが多いほど、端緒情報の性能向上が期待できることがわかる。

いずれの手法でも正しく判定できなかったツイート

表 6 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らうか。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

表 7 要検証記事の分類性能

手法	適合率	再現率	F 値
LSTM	0.66	0.64	0.65

を表 6 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る。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。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 3 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 27% 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。このようなツイートを正しく判定するには，ツイートだけでなく記事のデータも用いて，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。

#### 4.3 実験 2：検証必要度に基づく記事のランク付け

第 4.2 節で用いた端緒情報分類器を利用し，第 3 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平均をとった。端緒情報分類器には，実験 1 で最も性能が良かった LSTM を用いた。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，Precision@K と Recall@K による評価も行った。

表 7 に適合率と再現率，F 値を示し，図 4 に Precision@K と Recall@K を示す。図 4 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，検証必要性スコア上位の 1/4 の記事を見れば 6 割，1/2 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。

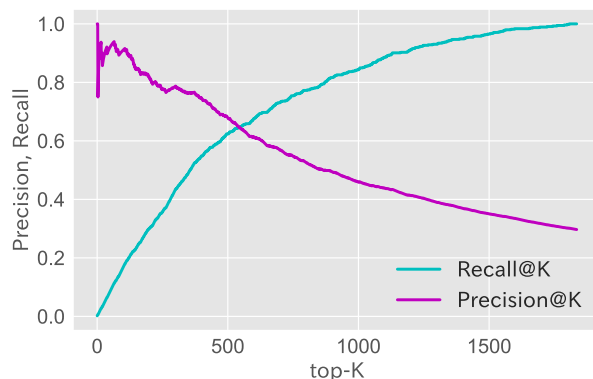


図 4 検証必要度のランク付け性能

## 5 おわりに

本論文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仕組みの構築に取り組んだ。実験ではファクトチェックの必要性を示唆している端緒情報の分類器を構築した。この分類器を用いて端緒情報の判定と記事の検証必要度をランク付けすることで，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 の支援を受けたものである。

## 参考文献

- [1] Mihai Dusmanu, Elena Cabrio, and Serena Villata. Argument Mining on Twitter: Arguments, Facts and Sources. *Emnlp*, pp. 2307–2312, 2017.
- [2] 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.
- [3] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [4] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. pp. 448–456, 2015.
- [5] Diederik Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.
- [6] Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.
- [7] Vinod Nair and Geoffrey E Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on Machine Learning*, No. 3, pp. 807–814, 2010.
- [8] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, Vol. 15, pp. 1929–1958, 2014.
- [9] 鍋島啓太，渡邊研斗，水野淳太，岡崎直観，乾健太郎. 訂正パターンに基づく誤情報の収集と拡散状況の分析. 自然言語処理 = *Journal of natural language processing*, Vol. 20, No. 3, pp. 461–484, jun 2013.