

フェイクニュース記事の誤り箇所抽出手法の提案

末吉 将也[†] 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

E-mail: †s1971054@s.konan-u.ac.jp, nadamoto@konan-u.ac.jp

あらまし メディア上にはさまざまなニュースが存在しており、そのニュースの中にはフェイクニュースと言われる誤りを含むニュースが多数ある。このようなフェイクニュースが存在することにより、本来求めていたニュースを取得することができないことやフェイクニュースが拡散されるなどの問題が存在する。その問題の解決策として、フェイクニュースの訂正を行うということが考えられる。本論文では、フェイクニュースの誤り箇所の訂正を行うためにニュースの主人公に着目し、ニュース中のその主人公を含む文に対する「どこで」、「いつ」、「何をしたか」を抽出する。そして、対象のフェイクニュースに対する正しいニュース群を抽出し、フェイクニュースの誤り箇所を抽出し、その訂正を行う。

キーワード フェイクニュース, Web ニュース

1 はじめに

インターネット上では、日々世界中で様々なニュースが発信され、その量は膨大になっている。さらに、我々が気になったことを調べるときや新規の情報を得る上でニュースは多大な役割を担っている。しかしながら、インターネット上の種々のニュースには、事実であるニュースのほかにデマや誤りを含むフェイクニュースが存在している。これらのようなフェイクニュースが存在することで、本来求めていたニュースを得ることができないほか、フェイクニュースを拡散される恐れがある。フェイクニュースが拡散されてしまうと、そのフェイクニュースを真実のニュースとして信じてしまう人が増え、様々な障害が起こる。これに対し、多くのフェイクニュースはファクトチェックが行われ、ファクトチェックサイト¹で公開されている。しかしながら、既にファクトチェックが行われたフェイクニュースや見てすぐに明らかにフェイクであると判断が可能なフェイクニュースもあれば、そうでないフェイクニュースもある。ニュースを見てそれがフェイクかどうかを常にファクトチェックすることは困難である。そこでフェイクニュースを自動で抽出する研究は様々ある [1] [2]。

また、フェイクニュースの影響は大きい。例えば、2008 年にはユナイテッド航空の親会社の倒産に関するフェイクニュースにより、ユナイテッド航空の株価が 76 パーセント下落したことがある²。また、2011 年の東日本大震災のような自然災害の時にもフェイクニュースが拡散され、これらのニュースを被災者が見ってしまうことで避難への意識や減災に向けての行動を妨げてしまうなどのことが起こっている。このように、フェイクニュースの影響は大きく、そのフェイクニュースを取り除くことや誤り箇所を訂正することがフェイクニュースの拡散防止になると考える。フェイクニュースを抽出して削除するので

はなく、誤り箇所を訂正するだけで正しいニュースに成り得るフェイクニュースも多数ある。例えば、「日本国内で消費されるトイレットペーパーの多くは中国で製造されたものを輸入しているが、新型コロナウイルスの影響で輸出入が困難になったことから日本国内ではトイレットペーパーが不足する」といったニュースでは、正しくは「日本で消費されているトイレットペーパーの多くは日本国内で製造されている」である。つまりは、「中国で製造されている」という部分が誤っているフェイクニュースである。そこで本研究では、フェイクニュースの誤りを訂正する手法の提案を行う。

本研究では、悪意の有無に関わらず誤字や脱字を含むニュースもフェイクニュースの対象とする。また、フェイクニュースの誤っている箇所を誤り箇所と呼ぶ。本論文で対象とする誤り箇所の訂正箇所は、本文中の「誰が(何が)」、「どこで」、「いつ」、「何を」、「なぜ」、「どのように」がニュース記事を作る際に重要である³。そこで我々はニュースにおいては特に「誰が(何が)」が重要であると考え、本研究では、ニュースの「誰が(何が)」を主人公とし、その主人公において関係する、「どこで」、「いつ」、「何を」、「なぜ」、「どのように」が誤っているかどうかを判定し、その修正を行う。本研究ではこの主人公に対する「どこで」、「いつ」、「何を」、「なぜ」、「どのように」をまとめて「ニュース構成要素」と呼ぶ。また、本論文では、はじめとしてこれらのニュース構成要素のうち、「どこで」、「いつ」、「何をしたか」を対象とし、フェイクニュースの誤り箇所の抽出を行い、その訂正を行う。

具体的には、以下の手順で行う。

- (1) 対象のフェイクニュースの主人公を抽出する。
- (2) 対象のフェイクニュースのニュース構成要素を抽出する。
- (3) (1) で抽出した対象のフェイクニュースの主人公と同じニュースを対象のニュースサイトから抽出する。

1 : <https://infact.press>

2 : <https://jp.reuters.com/article/idJPnTK16021320080909>

3 : <http://www.at-s.com/blogs/nie/study/howto.html>

(4) (3)で抽出したニュース群から(2)で抽出したニュース構成要素の一部を含むニュースを抽出し、それを正しいニュースする。

(5) (2)と(4)のニュース構成要素から対象のフェイクニュースの誤り箇所を抽出する。

以下、2章で関連研究について述べ、3章で本研究での対象のフェイクニュースの定義を行う。4章では提案手法によるフェイクニュース記事の誤り箇所抽出について述べる。5章では提案手法での誤り箇所の抽出に対する評価を行い、結果とその考察について述べる。最後に、6章で本論文のまとめと今後の課題について述べる。

2 関連研究

フェイクニュースに対しては数多くの研究が行われている。渡邊ら[3]は、地震や台風などの出来事に対して誤情報に多く使われているキーワードを取得し、そのキーワードからツイートを収集する。さらに、収集されたツイートを誤情報の支持・拡散ツイート、誤情報の反論・訂正ツイートに分類するためのコーパスの構築し、そのコーパスを用いて教師あり学習を用いて自動分類手法の提案を行っている。中原ら[4]は、Twitterにおいてデマかどうかをユーザが判断することを支援するための基準として、非公式リツイートに付加されたコメントをリツイートしたツイートの意見として抽出し、抽出したコメントに含まれる単語によって「危険度」と「訂正ツイート」の提示を提案している。この提案手法によってデマであるツイートにて検証した結果、訂正情報として非公式リツイートに付加されたコメントを確認している。梅島ら[5]は、間違った情報(デマ)の拡散を防止するため、Twitter上におけるデマを含むツイートとデマの訂正を含むツイートの拡散に関して、ツイートの内容とそのツイートが拡散されやすいかどうかの仮説を立てている。そして、それらのツイートの内容をポジティブ、ネガティブなどの評価属性をもとに評価し、分析を行なっている。岡山ら[6]は、フェイクニュースと政治のニュースから学習を行ったフェイクニュース分類機を作成し、Yelpのレビューに対してフェイクかどうかの分析をしている。さらに、フェイクと分類されたレビューとフェイクでない分類されたレビューを比較し、フェイクなテキストを含むレビューの言語的特徴の分析を行っている。これらの研究ではフェイクニュースに対して反論や訂正ツイートの抽出や評価属性や言語的な特徴をもとに分析を行なっているのに対し、本研究ではフェイクニュースに対して間違い部分を抽出することを行なっている点が異なっている。

QUら[7]は、デマの特徴や発生の原因を分析、考察しデータの前処理と機械学習によりデマ情報を検知する分類機を作成し、デマ検知アルゴリズムを開発した。山本ら[8]は、ウェブ検索プロセスにおいて検索結果の信憑性判断への注意喚起を目的とする。新しいタイプのクエリ推薦手法を提案している。これらの研究ではデマ情報やデマの発生原因の特徴やウェブ検索の特徴を用いているのに対し、本研究では新聞記事の構成を用いている点が異なる。

3 フェイクニュースの定義

フェイクニュースは様々な研究によって種々定義されている。Lazerら[9]は「フェイクニュースとはニュースメディアのコンテンツを形式的には模倣するが、組織的な考えや意図は模倣しない物である」と定義している。Zhangら[10]は「フェイクニュースとは主にインターネット上で公開・配信され、金融、政治、その他の利益のために意図的に読者を誤解させ、騙し、誘い込む為のあらゆる種類の情報やストーリーのこと」と定義している。本研究を行うにあたり、我々はファクトチェックの推進や普及に取り組む団体であるFIJ⁴によりニュースを以下のようにレーティングされたものを参考に、本研究でのフェイクニュースを定義する。

- 正確

事実の誤りはなく、重要な要素が欠けていない。

- ほぼ正確

一部は不正確だが、主要な部分・根幹に誤りはない。

- 誤り

全て、もしくは根幹部分に事実の誤りが存在する。

- 虚偽

全て、もしくは根幹部分に事実の誤りが存在し、事実でないこと知りながら伝えた疑いが濃厚である。

- 不正確

正確な部分と不正確な部分が混じっていて、全体として正確性が欠如している。

- ミスリード

一見事実とは異なることは言っていないが、釣り見出しや重要な事実の欠如などにより、誤解の余地が大きい。

- 根拠不明

誤りと証明できないが、証拠や根拠がないまたは、非常に乏しい。

- 判定留保

真偽を勝目することが困難。誤りの可能性が強くはないが、誤りでないとも否定もできない。

本研究では上記レーティングの内、ほぼ正確と誤り、不正確を対象とする。例えば、「三笥は2014年にFIFAワールドカップに出場した」や「FIFAワールドカップは2022年にフランスで開催された」などのニュースは全体として正確性が欠如されているため「不正確」に分類される。

つまりは本論文では、捏造されたニュースや真実のニュースを改ざんしているフェイクニュースに書き換えたもの、単に不正確なニュースなども含めて、悪意の有無に関わらず、誤字や脱字のような間違いを含むニュース全てをフェイクニュースと定義する。

4 フェイクニュースの誤り箇所の抽出手法

まず初めに、フェイクニュースの誤り箇所の訂正を行うため

4: <https://fij.info/introduction/guideline>

に誤り箇所抽出を行う。図1にフェイクニュース抽出のイメージ図を示す。

4.1 対象のフェイクニュースの主人公の決定方法

本研究ではニュースの根幹部分の抽出を行うため、まず初めにニュースの主人公を抽出する。ここで、ニュースの主語は人だけでなく、物や動物等である場合があるため、主人公は人だけでなく、名詞も対象とする。

主人公の抽出手順を以下に示す。

(1) タイトルに人物名がある場合、その人物名をそのニュースの「主人公候補」とする。

(2) タイトルに人名がない場合、本文中で(名詞)+「は」または(名詞)+「が」が出てきた時、その名詞を「主人公候補」とする。

(3) 「主人公候補」がない場合、タイトルにある名詞を「主人公候補」とする。

(4) 「主人公候補」が1人、1つの場合、その「主人公候補」を「主人公」とする。

(5) 「主人公候補」が複数ある場合、そのニュースの第1段落で最も出現数の多い人物名を「主人公」とする。

4.2 対象のフェイクニュースのニュース構成要素の抽出方法

対象のフェイクニュースに対して決定した「主人公」を含む文のみを抽出する。抽出した文それぞれに対してMecab⁵にて形態素解析を行い、ニュース構成要素をそれぞれの文から抽出する。ここで、フェイクニュースの主人公を M とする。ニュース構成要素の抽出手法を以下に示す。

- いつ (When)

「いつ (D_i)」は時間を表しているため、(数)+「年」や(数)+「月」が出てきた時、その(数)+「年」や(数)+「月」を「いつ」として抽出する。

- どこで (Where)

「どこで (P_i)」は場所を表しているため、「東京」や「大阪」などの地域を表す固有名詞や「国会」などの一般名詞+「で」や固有名詞または一般名詞+「に」が出てきた時、その固有名詞または一般名詞を「どこで」として抽出する。

- 何をしたか (What)

「何をしたか (W_i)」は「何を」という目的語と「したか・するか」という述語の連語である。まずは、「したか・するか」に着目し、動作を表す動詞を抽出する。しかし、「笑う」や「泣く」などは目的語がなくともある程度の意味は理解できる。しかしながら、「食べる」や「読む」などは目的語がないと文を完全に理解することが困難である。そのため、(名詞)+「を」+(動詞)や、(名詞)+(動詞)の形になっているものを「何をしたか」として抽出する。以上のような形になっていない場合には、動詞のみを「何をしたか」として抽出する。1つのフェイクニュースは主人公 M 、いつ (fD_i)、どこで (fP_i)、何をしたか (fW_i) から構成される。1つのフェイクニュースに主人公 M は1つであるが、いつ (fD_i)、どこで (fP_i)、何をしたか (fW_i) は1つ

とは限らない。これを $\{M, fD_i, fP_i, fW_i\}$ で示す。 f はフェイクニュース f を示し、 i は1つのフェイクニュースにおけるニュース構成要素の数である。

具体的な例を表1に示す。

4.3 対象のフェイクニュースに対する正しいニュースの取得

正しいニュースとして、毎日新聞⁶と朝日新聞⁷を使用する。これらの新聞社のニュース記事には、まれにフェイクニュースが紛れ込んでしまっている場合がある。しかしながら、本論文ではフェイクニュースかどうかの判定は研究の対象としていないため、毎日新聞と朝日新聞のニュース記事は全て真実が書かれているという想定で研究を行っている。決定したフェイクニュースのニュース構成要素より対象のフェイクニュースに対する正しいニュースの取得を行う。以下に正しいニュースの取得手順を述べる。

(1) フェイクニュースの主人公 M を含むニュースを毎日新聞と朝日新聞から抽出する。抽出したニュースを正しいニュース候補と呼ぶ。

(2) 正しいニュース候補から、ニュース構成要素 $\{M, SD_{jk}, SP_{jk}, SW_{jk}\}$ を全て抽出する。この時、 S は正しいニュース候補 S を示し、 j は S の数を示す。また、 k は j に含まれるニュース構成要素の数を示す。

(3) $\{M, fD_i, fP_i, fW_i\}$ と $\{M, SD_{jk}, SP_{jk}, SW_{jk}\}$ を比較し、 M 以外の要素が1つ以上あっている場合は、その S を正しいニュースとする。

例えば、「三笥は2014年にFIFAワールドカップに出場した。」というフェイクニュースがあった場合、 M が「三笥」で fD_1 が「2014年」で fP_1 が「FIFAワールドカップ」で、 fW_1 が「出場した」となる。これに対し「三笥は2022年にFIFAワールドカップに出場した。」というニュースは M が「三笥」で SD_{11} が「2022年」で SP_{11} が「FIFAワールドカップ」で、 SW_{11} が「出場した」となる。つまりは、フェイクニュースは三笥、2014年、FIFAワールドカップ、出場したとなり、正しいニュース候補は三笥、2022年、FIFAワールドカップ、出場したとなる。これにより、この S は正しいニュースであることがわかる。表2にフェイクニュースとそれに対する正しいニュースの例を示す。

4.4 対象のフェイクニュースの誤り箇所抽出

4.3節にて抽出した対象のフェイクニュースのニュース構成要素と正しいニュースのニュース構成要素を比較することにより、対象のフェイクニュースの誤り箇所を抽出するとともに、正しい値も決定する。以下に比較結果について述べる。

- $\{M, fD_i, fP_i, fW_i\}$ と $\{M, SD_{jk}, SP_{jk}, SW_{jk}\}$ が全て同じ場合

ニュース構成要素が全て同じ場合は S もフェイクニュースである。

- $\{M, fD_i, fP_i, fW_i\}$ と $\{M, SD_{jk}, SP_{jk}, SW_{jk}\}$ が全て

5 : <https://taku910.github.io/mecab/>

6 : <https://mainichi.jp>

7 : <https://www.asahi.com>

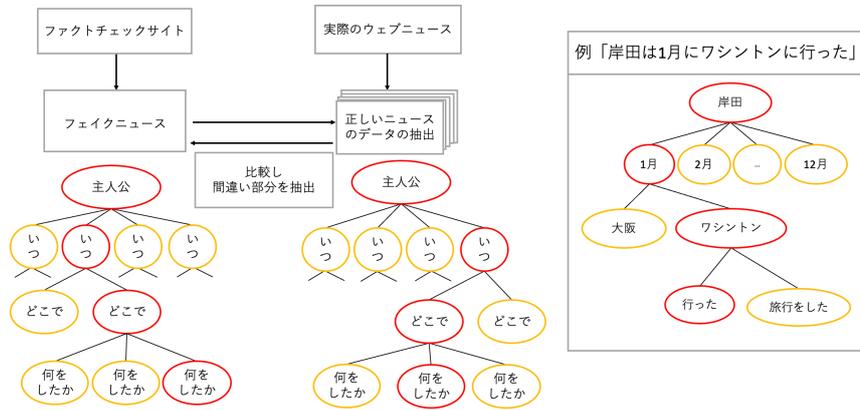


図1 フェイクニュース抽出のイメージ図

表1 ニュース構成要素

例文	主人公	どこで	いつ	何をしたらか
大谷翔平は2000年にWBCで先発投手を務めた。	大谷翔平	WBC	2000年	先発投手を務めた
ダルビッシュ有はInstagramで引退発表をした。	ダルビッシュ有	Instagram	0	引退発表をした
錦織圭は2008年にプロデビューをした。	錦織圭	0	2008年	プロデビューをした
大坂なおみは監督を辞任した。	大坂なおみ	0	0	監督を辞任した

表2 誤り箇所抽出

	フェイクニュース	フェイクニュースに対する正しいニュース
(1)	FIFAワールドカップは2022年にフランスで開催された。	FIFAワールドカップは2022年にカタールで開催された。
(2)	侍ジャパンは2015年に東京ドームでキューバと対戦した。	侍ジャパンは2017年に東京ドームでキューバと対戦した。
(3)	ダルビッシュ有は東京オリンピックに出場した。	ダルビッシュ有はWBCに出場した。
(4)	WBC第5回大会は2020年に開催された。	WBC第5回大会は2023年に開催された。
(5)	大谷翔平は2020年にホームランダービーに出場した。	大谷翔平は2021年にホームランダービーに出場した。

異なる場合

フェイクニュースとトピックが異なるニュースである。

• $\{M, fD_i, fP_i, fW_i\}$ と $\{M, SD_{jk}, SP_{jk}, SW_{jk}\}$ が1つ以上異なる場合

その異なっている要素が誤り箇所であり、その要素の S の要素の値が誤り箇所の正しい値である。

このようにしてフェイクニュースの誤り箇所を抽出する。

5 実験

主人公の抽出手法とニュース構成要素の抽出手法、および誤り箇所の抽出手法の有用性を示す3つの評価実験を行った。

5.1 実験1：主人公の抽出実験

実験の目的として、抽出した主人公が実際にそのニュースの主人公であるかの実験を行う。

実験方法

2022年12月から2023年2月までの毎日新聞よりニュース記事をランダムに100件を対象に、各々の主人公を抽出し、適合率を求める。正解データは人手により求める。

結果と考察

適合率は0.93であり、ほとんどのニュースにおいて主人公が抽出できていることがわかった。表3に実験に用いたデータの一部と抽出した主人公を示す。主人公を誤って抽出した悪い例は

表3の(4)(5)(6)である。(4)の場合はあるグループについてのニュースであるが、タイトルで一部メンバーを説明しているため、その人が主人公となった。(5)の場合は、ある人が別の人について語る際に主人公が入れ変わってしまった。これは、語る人についての説明や「ダルビッシュ有は言及した」のように、語る側の人の名前がタイトルにも第1段落にも多く出てくるためであり、今後の課題である。(6)の場合は、主人公に対してのあだ名などの別名の取得ができなかった。これは、タイトルでは主人公の本名またはあだ名のどちらかしか書いておらず、ニュース本文にて本名とあだ名どちらも書いている場合があるためである。これは今後の課題である。

5.2 実験2：ニュース構成要素の実験

本実験の目的は、ニュース記事の「主人公」からニュース構成要素がそれぞれ正しく抽出できているかの実験を行う。

5.2.1 ニュース構成要素「いつ」の実験

ニュース記事の「主人公」からニュース構成要素である「いつ」が正しく抽出できているかの実験を行う。

実験方法

2022年12月から2023年2月までの毎日新聞よりニュース記事をランダムに抽出した100件のニュース記事に対して、抽出したニュース構成要素の「いつ」の適合率を求める。また、本実験で用いた主人公は全て正しい主人公を使用した。正解

表3 主人公の抽出実験の結果

ニュース番号	ニュース概要	抽出した主人公	正解
(1)	永岡桂子は8日の衆院予算委員会で、督促や過料を怠っていたと認めた。	永岡桂子	永岡桂子
(2)	立憲民主党が衆院憲法審に欠席の方針を説明した。	立憲民主党	立憲民主党
(3)	高梨沙羅がインスタグラムに心境つづる。	高梨沙羅	高梨沙羅
(4)	日本サッカー協会は、なでしこジャパンに熊谷紗希らを選出した。	熊谷紗希	日本サッカー協会
(5)	ダルビッシュ有が栗山監督について言及した。	ダルビッシュ有	栗山監督
(6)	東谷義和はアラブ首長国連邦などに滞在し、一度も登院していない。	東谷義和	東谷義和 ガーシー

データは人手により決定した。

結果と考察

適合率は0.94であり、高い適合率となった。表4に抽出したニュースと主人公、「いつ」の結果を示す。「いつ」が抽出できなかった例として表4の(5)(6)があり、「10年ぶり」や「18年越し」といった相対的な年数や日数を正しく抽出できなかった。これは今後の課題である。

5.2.2 ニュース構成要素「どこで」の実験

ニュース記事の「主人公」からニュース構成要素である「どこで」が正しく抽出できているかの実験を行う。

実験方法

2022年12月から2023年2月までの毎日新聞よりニュース記事をランダムに抽出した100件のニュース記事に対して、抽出したニュース構成要素の「どこで」の適合率を求める。また、本実験で用いた主人公は全て正しい主人公を使用した。正解データは人手により決定した。

結果と考察

適合率は0.90であり、高い適合率となった。表5に抽出したニュースと主人公、「どこで」の結果を示す。「どこで」が抽出できなかった例として表5の(5)(6)(7)がある。本研究では、「Instagramで」や「Twitterに」などのSNSに投稿されたものも「どこで」として抽出している。そのため、「どこで」を(名詞)+「に」としているため、本来取得したい場所や試合名とは関係のないものを取得することがある。これは今後の課題である。

5.2.3 ニュース構成要素「何をしたか」の実験

ニュース記事の「主人公」からニュース構成要素である「何をしたか」が正しく抽出できているかの実験を行う。

実験方法

2022年12月から2023年2月までの毎日新聞よりニュース記事をランダムに抽出した100件のニュース記事に対して、抽出したニュース構成要素の「何をしたか」の適合率を求める。また、本実験で用いた主人公は全て正しい主人公を使用した。正解データは人手により決定した。

結果と考察

適合率は0.90であり、高い適合率となった。表6に抽出したニュースと主人公、「何をしたか」の結果を示す。「何をしたか」が抽出できなかった例として表6の(4)(5)がある。(4)は「日本には存在する」という意味での「日本にはいる」という述語であるが、「日本に入国する」といった「はいる」のみを「何を

したか」として抽出した。また、(5)では「子育てがしやすい」といった述語が「子育てをする」のみを「何をしたか」として抽出し、正しく意味を伝えられないことがある。これらは今後の課題である。

5.3 実験3：誤り箇所抽出実験

提案したフェイクニュースの誤り箇所の抽出手法の有用性を示す実験を行う。

実験方法

フェイクニュースとして、InFactよりファクトチェック済のフェイクニュース94件を抽出する。このフェイクニュースは新型コロナウイルスや参議院選など様々なトピックからなる。正しいニュースは2019年7月1日から2022年12月31日までの毎日新聞と朝日新聞のデータを使用する。各々94件のフェイクニュースから提案手法を用いて誤り箇所を抽出する。そして、正解データと比較して適合率を求める。正解データは人手により、各々フェイクニュースの誤り箇所を決定する。

結果と考察

適合率が0.70であり、あまり高くなかった。結果の一部を表7に示す。誤り箇所を正しく抽出できない例として、表7の(3)(4)があり、(3)の場合は、正しいニュースが取得できなかったため、誤り箇所の抽出や修正ができなかった。(4)の場合は、フェイクニュースの主人公としてファミリーネームのみを抽出している。そのため、同姓でニュース構成要素のうち1つでも同じ場合には、関係のない人でも正しいニュースとして取得され誤り箇所が抽出される。これは今後の課題である。

6 まとめと今後の課題

本論文では、ニュース記事の重要な語である主人公に着目し、その主人公に対する、「どこで」、「いつ」、「何をしたか」で構成されるニュース構成要素により、フェイクニュースの誤り箇所の抽出方法の提案を行った。具体的には、まず初めにフェイクニュースの「主人公」の抽出手法を提案して、主人公を決定した。そして、その主人公に対してニュース構成要素の取得手法を提案した。最後にニュース構成要素を用いて他のニュースと比較することにより、対象のフェイクニュースの誤り箇所の抽出手法を提案した。評価実験を行い、これら3つの提案手法の有用性を示した。本論文では、「なぜ」と「どのように」の2つの要素の抽出手法の提案を行っていなかった。これを今後の課題とする。また、主人公決定時に、対象の文に主語が

表 4 ニュース構成要素「いつ」の実験結果

番号	ニュース概要	主人公	いつ
(1)	ダルビッシュ有は 2 月 9 日にサンディエゴ・パドレスと 6 年間総額 1 億 800 万ドルで契約延長に合意した	ダルビッシュ有	2 月 9 日
(2)	大谷翔平は 1 月 6 日に日本代表入りが正式に決まった。	大谷翔平	1 月 6 日
(3)	錦織選手は 2022 年 1 月に股関節を手術し、2022 年 10 月には右足首を捻挫、公式戦への復帰に向けリハビリを続けています。	錦織圭 錦織圭	2022 年 1 月 2022 年 10 月
(4)	2 月 20 日に WTA 世界ランキングが発表され、大坂なおみは前週から変わらず 66 位をキープした。	大坂なおみ	2 月 20 日
(5)	バンクーバー五輪代表の織田信成は 10 年ぶりに全国大会に出場した。	織田信成	10 年
(6)	石井一成は 18 年越しの夢を果たすためにチャレンジャーとして K-1 に臨む。	石井一成	18 年

表 5 ニュース構成要素「どこで」の実験結果

番号	ニュース概要	主人公	どこで
(2)	京都府舞鶴市で元市議の鴨田秋津氏が現職の多々見良三氏ら 3 人を破って初当選した。	鴨田秋津氏	京都府舞鶴市
(3)	車いすテニスの第一人者の国枝慎吾さんが 7 日、東京都で記者会見をした。	国枝慎吾	東京都
(4)	1 月に開催されたテニスの全豪オープン車いすの部男子シングルスで準優勝した小田凱人が 6 日、愛知県大口町で結果を報告した。	小田凱人 小田凱人	全豪オープン車いすの部男子シングルス 愛知県大口町
(5)	高梨沙羅はインスタグラムを更新し、チームメイトに感謝を表現した。	高梨沙羅	チームメイト
(6)	政府は 8 日、重要設備を導入する際に事前審査の指針案を公表した。	政府	際
(7)	石川祐希は真摯にバレーボールに向き合っていきたいと語った。	石川祐希 石川祐希	真摯 バレーボール

表 6 ニュース構成要素「何をしたか」の実験結果

番号	ニュース概要	主人公	何をしたか
(1)	ダルビッシュ有は真ん中高めのツーシームをスタンドまで運んだ村上を絶賛した	ダルビッシュ有	村上を絶賛した
(2)	ロサンゼルス・エンゼルスの大谷翔平が MLB ネットワークの選手ランキングで 2 年連続の 1 位に輝いた。	大谷翔平	1 位に輝いた
(3)	男子テニスの錦織圭が、エントリーしていた 2 月のデルレービーチ・オープンの出場を取りやめた。	錦織圭 錦織圭	エントリーしていた 出場を取りやめた
(4)	13 頭のジャイアントパンダが日本にはいる。	ジャイアントパンダ	はいる
(5)	福井県は子育てしやすい県に選ばれた	福井県 福井県	選ばれた 子育てする

表 7 誤り箇所抽出実験の結果

番号	誤り箇所を含む文	フェイクニュースのニュース構成要素	正しいニュースのニュース構成要素	誤り箇所
(1)	羽田雄一郎は都内の会合で、体調不良を訴えて倒れる	{ 羽田雄一郎, 0, 会合, 倒れる }	{ 羽田雄一郎, 0, 車中, 倒れる }	会合
(2)	TikTok が他アプリで入力中のテキストを読み取っている。	{ TikTok, -, 他, アプリ, テキストを読み取っている }	{ TikTok, -, 使用中, クリップボードを読み取っている }	他, アプリ フェイクニュース
(3)	国交省は、2005 年に北海道に中国人 500 万人移住計画を発表した。	{ 国交省, 2005 年, 北海道, 中国人 500 万人移住計画を発表した }	{ 0, 0, 0, 0 }	1 から作成された フェイクニュース
(4)	ジョーンズがテキサス州で不正投票事件により逮捕された。	{ ジョーンズ, 0, テキサス州, 逮捕された。 }	{ ジョーンズ, 0, フロリダ州, 逮捕された。 }	テキサス州

抜けている場合を考慮していないため、このように主語が抜けている場合の対応をする。さらに同義語にも対応する予定である。また本論文では、フェイクニュースか否かが分かっていないニュースを対象としていないが、今後これらのニュースに対しても対応してゆく予定である。

7 謝 辞

論文の一部は JSPS 科研費 19H04218, 20K12085, 及び私学助成金 (大学間連携研究助成金) 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] 古川 凌也, 伊藤 大貴, 高田 雄太, 熊谷 裕志, 神園 雅紀, 白石 義明, 森井 昌克, “ SNS ユーザの自己紹介文にあらわれる偏りに着目したフェイクニュース検知 ” In: IEICE Conferences Archives.

- The Institute of Electronics, Information and Communication Engineers, 2021.
- [2] 柳 裕太, 田原 康之, 大須賀 昭彦, 清 雄一, “ 画像付きフェイクニュースとジョークニュースの検出・分類に向けた機械学習モデルの検討 ” 研究報告知能システム (ICS), 2019-ICS-193, pp.1-8, 2019.
- [3] 渡邊 研斗, 鍋島 啓太, 岡崎 直観, 乾 健太郎, “ Twitter 上での誤情報と訂正情報の自動分類. ” 言語処理学会第 19 回年次大会, 2013.
- [4] 中原 英美, 富永 一成, 牛尼 剛聡 “ リット構造を用いたデマ拡散防止支援手法 ” 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), pp. 1-7, 2012.
- [5] 梅島 彩奈, 宮部 真衣, 荒牧 英治, 灘本 明代, “ 災害時 Twitter におけるデマとデマ訂正 RT の傾向. ” 研究報告情報基礎とアクセス技術, Vol. 2011, No.4, pp. 1-6, 2011.
- [6] 岡山 光平, 石川 博, 廣田 雅春, “ フェイクニュース分類機を用いた口コミサイトのレビューの分析. ” 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), pp. 1-5, 2018.
- [7] Qu Lijing, 成凱 “ 機械学習を用いたデマ検知アルゴリズムの評価 ” 情報処理学会研究報告 (IPSJ SIG Technical Report)

pp. 1-6 , 2021.

- [8] 山本祐輔, 田中克己 “反証センテンスの提示による信憑性指向のウェブ検索支援” 情報処理学会論文誌 データベース Vol.6 No.2 , pp. 42-50,2013
- [9] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. ”The science of fake news.” Science 359, 6380 (2018), pp. 1094-1096.
- [10] XichenZhang and AliAGhorbani, ”An overview of online fake news : Characterization,detection,and discussion.”, Information Processing & Management Volume 57, Issue 2 (2020).