

# ニュース記事へ言及するTwitter投稿の主観性と情報継承性に基づく分類

遠田哲史<sup>1</sup> 吉永直樹<sup>2</sup> 豊田正史<sup>3</sup>

近年のソーシャルメディアの普及に伴い、ニュース記事への言及を伴う投稿を通じてニュースが拡散される現象が広く見られるようになってきている。ここで問題となるのは意図的に誤情報を流布する「フェイクニュース」であるが、本物のニュース記事について書かれた投稿であっても、読者にニュースの内容に関する誤解を生じさせるような、誤解を招く表現が使われる場面もある。本論文では、COVID-19に関するニュース記事への言及を含むTwitter（現X）上の投稿を網羅的に分析し、投稿の主観性と記事の内容が投稿に継承されているか（情報継承性）に基づいた投稿内容の分類を行う。また、投稿とニュース記事に含まれる情報を利用し、投稿の分類を自動的に判定する分類器を提案する。

## 1 はじめに

近年はネット上における情報入手手段として、あるいは手に入れた情報の拡散・議論を目的として、ソーシャルメディアが活用される場面が増えている。実際に、ソーシャルメディア上でときどきまたは頻繁にニュースを入手しているアメリカ人成人は50%に及ぶという2022年7～8月に行われた調査結果があり [10]、日本においても、2021年9～10月に行われた調査で小中学生の過半数が日常的にソーシャルメディア経由でニュースを得ていると報告されるなど [17]、既存のマスメディアと同等以上に重要なニュースの入手経路となっている。これらソーシャルメディア上のニュースに関する投稿には世の中の出来事理解を促進する有益なものもあるが、誤解や意図的な誤情報に基づく投稿も多く、結果として不正確な情報が蔓延することが重大な問題となっている。とりわけ、COVID-19の感染拡大に伴い対面で情報交換をする機会が減り、オンライン上の情報への依存度が増えたことで問題が深刻化しており、誤情報が拡散し氾濫する状況（インフォデミック）により社会に多大な悪影響が及ぶ可能性が指摘されている [15]。

このような誤情報や偽情報を包括する概念としての「フェイクニュース」の問題に対し、与えられた投稿が提供する情報の信憑性を判断することでフェイクニュースを検出するファクトチェックを自動化する研究が行われている [3, 6, 14]。多くのファクトチェック研究では、ニュースの真偽の判断を行う分類器の作成を目的とし、元投稿の主張の正誤や投稿への反応の主張への賛否度合いがラベル付けされたデータセットが提供されている。

<sup>1</sup> 学生会員 東京大学大学院情報理工学系研究科

tohda@tkl.iis.u-tokyo.ac.jp

<sup>2</sup> 非会員 東京大学生産技術研究所

ynaga@iis.u-tokyo.ac.jp

<sup>3</sup> 正会員 東京大学生産技術研究所

toyoda@tkl.iis.u-tokyo.ac.jp

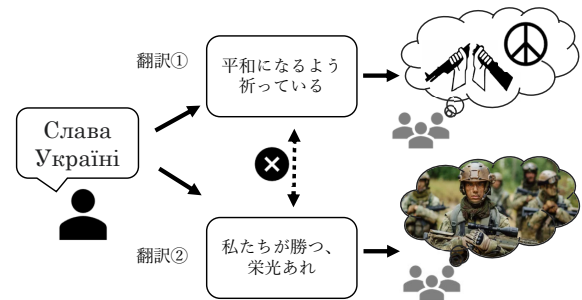


図1: 互いに異なる印象を与える、誤解を招く情報伝播の例。

しかし、ソーシャルメディア上での情報伝播過程に着目すると、ファクトチェックの対象とならない主観表現を含む投稿が多数あり、信憑性が確立された記事への直接的な言及があっても、なおニュースに関する誤解が伝わる場面がみられる。例えば話題の元となる投稿に関する間接的な言及や部分的な引用、また内容を翻訳や要約した投稿を通じて、話題の内容が改変されながら他ユーザに伝わることも多い。元となる情報源を忠実に引用しているように見せかけるなどして、元の内容と必ずしも矛盾しない恣意的な情報の取捨選択を行った場合、読者を誤解させ、誤った情報が不特定多数のユーザに拡散する状況につながる。本研究で問題とするような改変の例を図1に示す。

本研究では、COVID-19に関するニュース記事への直接的な言及があるTwitter（現X）上の投稿（ツイート）を対象とした分析を行い、投稿と元記事との関係を明らかにすることを目的としたアノテーションスキームを設計した上で、投稿のラベル付きデータセットの構築を行う。次に、データセットの分析を行い、投稿者による主張の含有度合い（主観性）と元のニュース記事の内容を継承している度合い（情報継承性）の軸に基づいてラベルの整理を行う。最後に、ニュース言及を含むTwitter投稿の主観性と情報継承性の分類タスクを新たに提唱し、これを解く分類器の提案、および構築したデータセットを用いた有効性の評価を行う。

## 2 関連研究

### 2.1 フェイクニュース検出・ファクトチェック

誤情報の検出を目的とした研究には、与えられた情報の信憑性を判断するファクトチェックを自動化する研究が挙げられる。ファクトチェックには3つの段階があり、それぞれ個別のタスクとして分解することが出来る [14]。最初の段階では、文中から主張を抽出し、ファクトチェックの検討に値するかどうか、つまりCheckworthyであるかどうか検討する [16]。次に主張を肯定あるいは否定する根拠となるエビデンスを（ウェブ上などから）取得する段階がある。最後に、エビデンスと主張を比較し、前者が後者を肯定ないし否定するかチェックし、主張の真偽について最終的な判断を下す [2, 12, 13]。また、主張の中の情報の新規性に着目し、情報新規性の分類タスクで学習させたモデルをファクトチェックに応用する研究も行われている [8]。

従来の研究ではニュース記事や個別の主張を含むテキストを入力とするタスク設定も多かったが、近年ではTwitter上の投

稿をファクトチェックの対象とした研究が増えている。とりわけ、RumourEval19 [5] が本研究と近い。RumourEval19では、Twitter上の投稿と、それに対する反応（リプライ）について、リプライの元投稿に対する肯定度合いをラベル付けしている。

また、COVID-19関連のファクトチェック用データセットの整備 [7]も行われている。この研究では、Twitter上の任意の投稿に対して、その内容に対応するCOVID-19についての既知の流言を取得し、対応付けされた流言に対する投稿のスタンスをラベル付けしている。このラベルを自然言語推論システムの出力に対応付けることで、投稿が流言に対して肯定的に言及しているか否かという観点からのファクトチェックを行っている。

テキスト以外のモダリティに関する誤情報のデータセットとしてはCOSMOS [1]が挙げられる。写真が適当でない説明文とともに提示されると、記事の読者が騙されてしまう現象に対抗し、画像に対する誤解を生む説明文を検出する手法を提案した。本論文で提案する、ツイートの主観性と情報継承性の分類タスクは、ニュース記事に紐づけられた投稿による誤解を防ぐことを目的とする点で、COSMOSと同類だがモダリティの異なる問題設定をしていると解釈できる。

## 2.2 含意関係認識

特定のニュースに言及したツイートの情報継承性を分類するタスクは、自然言語処理分野で研究されている含意関係認識タスクの類型だと見ることができる。含意関係認識とは、与えられたテキスト同士に意味的な含意関係が成り立つかどうかを自動的に判定する技術であり、通常は与えられた前提文が仮説文の内容を意味的に含意するかどうか判定する設定になっている。近年はより長い文章の含意関係を扱った研究も行われており、ニュース文章に関する文章単位での含意関係認識も行われている [4,11]。

本研究で比較しているニュース記事とSNS上の投稿は書き手の意見が含まれることや、前提とするニュース記事と仮説とする投稿の長さが多様であることから、既存の含意関係認識手法とは設定が異なる。そのため、既存の含意関係認識データセットで学習された判定器をそのまま適用することは難しい。一例として、比較している文章同士の長さが大幅に違うため、重複エンティティの数を数える等のヒューリスティクスが使えない等の問題がある。また、内容の情報継承性に関するラベルについても、誤解を生む可能性に着目した粒度の高いスキーマが求められている。

## 3 ニュース言及を含む投稿データの取得

本節では、ニュース記事への言及を含むTwitter投稿の取得における前処理、および高い拡散度合いが見られる投稿の選定について説明する。また、ニュース言及を含むツイートを分析し、構築したデータセットにおける投稿の選定基準の説明を行う。

### 3.1 前処理

本研究では、2022年の日本語の全量投稿<sup>\*1</sup>からCOVID-19関係の外部ニュースサイトに直接言及した単体のツイート（ソースツイート）を抽出し、これらを情報伝播上の重要性に応じて並べ替えを行い、残った上位の投稿を対象とした分析を行う。

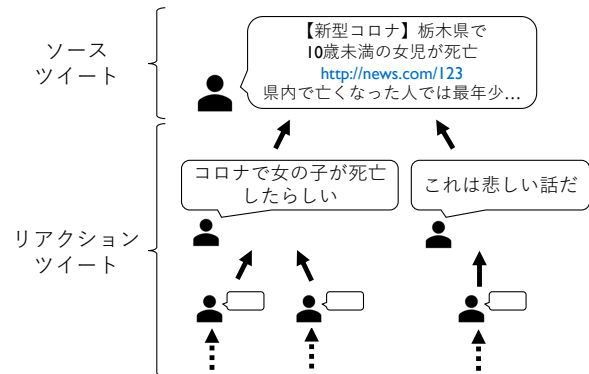


図2: リアクションツリーの例。

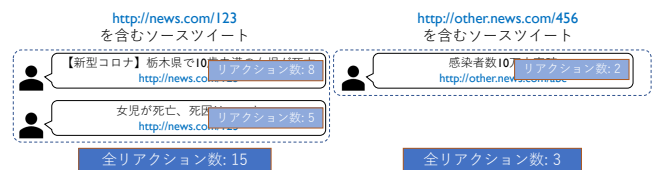


図3: リアクション数に応じた並べ替えの例。

全量ツイートを取得した後、データの前処理として以下の処理を行ってソースツイートを取得する。

- キーワードフィルター: トピックキーワード（「コロナ」「covid」「COVID」）を含まないツイートや、スパム投稿と関連性の高いキーワード（「プレゼント」等）を除外
- URLフィルター: URLが含まれないツイートを除外
- 単体ツイート抽出: リプライ・RT・QTを除外
- 非公式RT除外: リンク先のタイトルとURLのみ含むツイートを除外
- 重複ツイート除外

次に、それぞれのソースツイートの拡散度合いを判断するために、ソースツイートのリアクション数を取得する。取得対象は「リプライ」「RT」「引用RT」とし、まずソースツイートに対するリアクションから取得を開始し、これらに対するリアクション（リプライのリプライなど）をも再帰的に取得することで、ソースツイートを根ノード、リアクションを葉ノードとしたリアクションツリーを抽出する（図2）。リアクションツリーに含まれるリアクションの数を取得し、ソースツイートに対するリアクションの総数とする。なお、リアクションはソースツイートが投稿された日時より1週間以内まで取得した。

最後に、それぞれのソースツイート及びリンク先のニュース記事の拡散度合いを判断するために、ソースツイートに対するリアクション数を利用した二段階の並び替えを行う（図3）。第一段階では、各ニュース記事の拡散度合いを測るため、リンクを含むソースツイートを集め、その数およびリアクション数の総和を計算する。この値でニュース記事を降順に並べ、よりリアクション数の多いソースツイートのグループを優先的に分析・ラベル付けを行うこととした。第二段階では、各ソースツイートの拡散度合い

\*1 TwitterデータはNTTデータより提供を受けたものである。

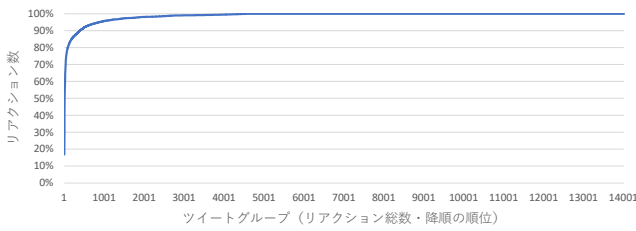


図4: 2022年5月1日のツイートグループのリアクション数.

リアクション数%	順位
50	8
60	17
70	32
80	85
90	388
95	864
100	4535

表1: 2022年5月1日のツイートグループのリアクション数と順位.

を測るため、同様にリアクション数の多さでソースツイートを並べ替えた。また、最終的なデータセット構築の際には、ニュースサイト（NHK、共同通信、Yahoo!ニュース）へのリンクを含まないものを除外するフィルターを設けた。

### 3.2 ニュース言及を含むTwitter投稿の分析

ニュース言及を含むツイートの分析を行うにあたり、上記手法で2022年5月1日に投稿されたニュース記事へのリンクを含むツイート23,508件を取得した。これらのツイートをリンク別にグループ分けした結果が図4である。ツイートグループに含まれるリアクションの大半は、上位の限られた数のニュース記事に関するものであることがわかる。具体的な数値は表1に示すが、上位8件のニュース記事に関する投稿が、当日に行われたツイートに対するリアクションの半数を占めている。リアクション数上位100件のニュース記事を扱うことで、当日に行われたCOVID-19関連ニュースにまつわるツイッター上のリアクションの8割以上を網羅できることがわかる。

図5に上位4ツイートグループ内のソースツイートにおけるリアクション数の分布を示す。同じ記事のリンクを含む投稿同士でもリアクション数に大きな差異がみられ、特に上位数ツイートにリアクションの大多数が集中していることが見受けられる。

これらの結果より、リアクション数の多いツイートグループを選び、その中でもリアクション数の多いソースツイートを選ぶことで、その日に行われたTwitter上の情報伝播の大きな部分を占める重要性の高いツイートを得られることが分かる。

## 4 ニュース言及を含むTwitter投稿データセット

本研究で構築した、ニュース記事への言及を含むTwitter投稿データセットの概要を説明する。まず、前節で説明した手法に従い、2022年10月1日から4日にかけて投稿されたCOVID-19関連ニ

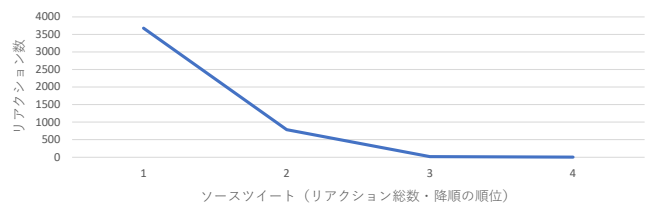
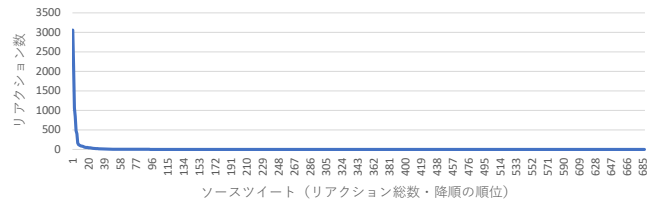
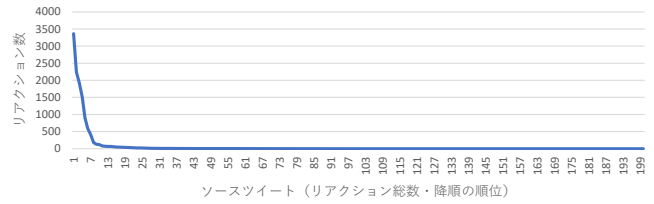


図5: 2022年5月1日の上位4ツイートグループ内のソースツイートのリアクション数.

ュースの言及を含むツイートを網羅的に取得した。これらをグループに分けたのち、各日リアクション数上位200のツイートグループの中でそれぞれ上位5ツイートまでを取得し、得られた914ツイートすべてに対してアノテーション作業を行った。これらの作業は、論文著者の1名が全て実施した。

### 4.1 アノテーションスキーム

取得したツイートについて、ツイートとリンク先のニュース記事の関係を細やかに表すラベルを設け、特にツイート内でリンク先のニュース記事の内容がどの程度忠実に表現されているか明らかにすることを念頭にアノテーション用のスキームを構築した。本スキームでは、ツイートに含まれる内容を表すカテゴリを設け、それぞれ元のニュースと比較した時の改変度合いに応じた数値をラベルとして付与する。

カテゴリとしては、次の5種類を用意した。

**言い換え・要約:** リンク先のニュース記事に含まれる情報を言い換えたり、記事の要約を行っている

**翻訳:** ニュース記事の言語とツイートの言語が異なり、翻訳された情報が含まれている

**外部情報の参照:** リンク先のニュース記事に含まれていない情報が含まれている

**投稿者の意見:** ツイートの投稿者の意見や主張が含まれている

	タイトル	言い換え・要約	外部情報の参照	投稿者の意見	ツイート数	全体に占める割合 (%)
	✓	-	-	-	228	24.9
	-	✓	-	-	33	3.6
	-	-	✓	-	1	0.1
	-	-	-	✓	62	6.8
	✓	✓	-	-	97	10.6
	✓	-	✓	-	19	2.1
	✓	-	-	✓	285	31.2
	-	✓	✓	-	2	0.2
	-	✓	-	✓	34	3.7
	-	-	✓	✓	32	3.5
	✓	✓	✓	-	3	0.3
	✓	✓	-	✓	51	5.6
	✓	-	✓	✓	60	6.6
	-	✓	✓	✓	5	0.5
	✓	✓	✓	✓	2	0.2
合計					914	-

表2: データセットのカテゴリラベルの組み合わせ別ツイート数.

カテゴリ	ツイート数	全体に占める割合 (%)
言い換え・要約	227	24.8
翻訳	0	0
外部情報の参照	124	13.6
投稿者の意見	531	58.1
タイトル	745	81.5
合計	914	-

表3: データセットのカテゴリ別ツイート数.

改変度合い	ツイート数	カテゴリに占める割合 (%)
改変なし	105	46.3
誤解の生じない改変あり	100	44.1
誤解の生じうる改変あり	10	4.4
完全な改変あり	12	5.3
合計	227	-

表4: 「言い換え・要約」カテゴリの改変度合い別ツイート数.

タイトル: 記事のタイトルが含まれている

それぞれのカテゴリについて、情報改変度合いを0-3の数値で表す。ここでの数値の意味は次の通りとなる。

- 0 (改変なし): 記事に含まれる情報は完全に保持されている。
- 1 (誤解の生じない改変あり): ニュース記事に含まれる情報について、表現上の改変は行われているが、その内容は十分に継承されている。読者がニュース記事を読んだときに得る理解と、ツイートを read したときに得る理解に大きな差はない。
- 2 (誤解を生じうる改変あり): ニュース記事に含まれる情報について、誤解を生じうる程度の改変が行われている。読者がニュース記事を読んだときに得る理解と、ツイートを read したときに得る理解に差がある。
- 3 (完全な改変あり): ツイートに含まれる情報は完全に間違っている。

アノテーション作業には、ツイートの本文及びリンク先のニュース記事のタイトルと本文が提示される。まず、5カテゴリのそれぞれについて、そのカテゴリに該当する要素がツイートに含まれるかどうかを判断する。次に、該当したカテゴリについて、情報改変の度合いをラベルとして付与する。なお、単一のツイートが複数のカテゴリに該当することもある（記事内容を要約し、

それに対する意見を述べるツイートなど）ため、該当するカテゴリ全てについて情報改変度合いのラベルを付与することとする。

#### 4.2 データセットの特徴

表2にカテゴリラベルの組み合わせ別ツイート数を、表3にカテゴリ別のツイートの分布を示す。カテゴリ別ツイート数を見ると、最も多くツイートが該当したカテゴリがタイトルであり、次いで投稿者の意見、言い換え・要約が続く。タイトルの多さについては、多くのニュースサイトではSNSへ投稿するためのボタンが設けられており、これを押すと自動的にニュース記事へのリンクとタイトルがツイート本文に挿入される機能があるためだと考えられる。また、カテゴリラベルの組み合わせに着目すると、投稿者の意見とタイトルの組み合わせが最も多く見られた。これについては、ニュースに言及したうえで自分の意見を述べる形の投稿が多く見られることを示している。

表4に「言い換え・要約」カテゴリに属するツイートの情報改変度合いの分布を示す。言い換え・要約を含む投稿のうち、記事内容の改変が全く見られないものは半分以下となっていることがわかる。また、10%弱の投稿には、少なくとも誤解を生じうる程度の改変がみられることが示されている。これは、アノテーションが行われたツイートが全てのツイートの集合から無作為に選ばれたのではなく、よりリアクション数の多いものが選ばれていることに着目すると、Twitter上における情報改変を含む情報伝播の

ラベル	基準
客観的	「投稿者の意見」ラベルが含まれていない
混在	「投稿者の意見」ラベルに加え、他のラベルを含む
主観的	「投稿者の意見」ラベルのみ含まれている

表5: ツイートの主観性ラベルの付与基準.

システム	主観性			情報継承性		
	P	R	$F_1$	P	R	$F_1$
最多クラス法	0.17	0.33	0.22	0.28	0.33	0.30
提案手法	<b>0.66</b>	<b>0.65</b>	<b>0.64</b>	<b>0.47</b>	<b>0.45</b>	<b>0.45</b>

表8: 提案手法 (入力:  $F^{tw}$ ) の性能評価.

ラベル	基準
高	「言い換え・要約」かつ 「誤解を生じない改変」未満を含む
中	「外部情報の参照」を含む
低	「言い換え・要約」かつ 「誤解を生じる改変」以上を含む

表6: ツイートの情報継承性ラベルの付与基準.

ラベル	ツイート数
客観的	155
混在	184
主観的	347

(a) 主観性

ラベル	ツイート数
高	708
中	122
低	22

(b) 情報継承性

表7: ツイートの主観性・情報継承性ラベルの分布.

多さを強く印象付ける結果となっている。

このデータセットを後述のタスクに適した形にするため、ツイートの主観性と情報継承性の軸に合わせたラベルの整理を行う。ラベルの付与基準は表5・表6にて示し、それぞれのラベルの分布は表7に示す。なお、主観性ラベル付与の際には「タイトル」カテゴリを無視（「タイトル」カテゴリのみ含むツイートを除外）し、情報継承性ラベル付与の際には「投稿者の意見」カテゴリのみ含むツイートを除外した。前処理で除外されたツイートがそれぞれ異なるため、主観性および情報継承性ラベルの付与されたツイートの数はそれぞれ異なる。

## 5 ニュース言及を含むTwitter投稿の主観性と情報継承性の分類

構築したデータセットを利用し、ニュース記事への言及を含む任意のツイートについて、投稿内容の主観性とニュース記事の内容に対する情報継承性を分類するタスクを新たに設定する。このタスクを解く事前学習済みの大規模言語モデルを利用した分類手法を提案し、基本的な性能評価を行う。

### 5.1 提案手法の概要

提案手法は事前学習済み言語モデルRoBERTa [9] とロジスティック回帰を組み合わせたものである。入力はツイート本文 ( $F^{tw}$ ) およびリンク先記事のタイトル ( $F^t$ )、記事の本文 ( $F^b$ ) とし、それぞれ独立したRoBERTaエンコーダの入力とする。次に、各エンコーダの[CLS]トークンの最終隠れ層を全て連結したベクトルを作る。最後に、このベクトルを入力、ツイートの正ラベルを出力とした多項ロジスティック回帰モデルを学習する。

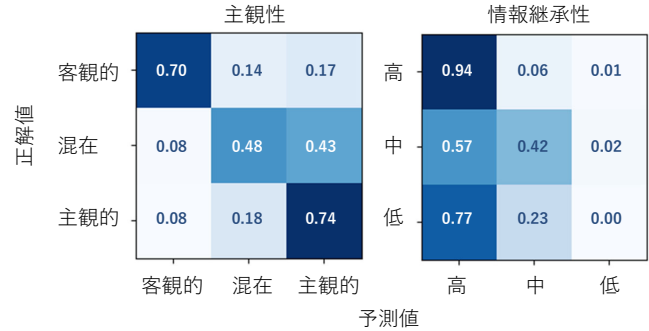


図6: 提案手法 (入力:  $F^{tw}$ ) の混同行列.

実験では、主観性と情報継承性の分類を独立したタスクとして扱う。ラベル付けされた訓練データについて、それぞれ学習を行った後、各タスクのテストデータに対する推論結果を得る。

### 5.2 実験

前節で述べた分類器を日本語 Wikipedia および Web テキストから事前学習した日本語RoBERTaモデル<sup>\*2</sup>を用いて実装した。前処理としてJUMAN++<sup>\*3</sup>を用いた分かち書きを行い、 $F^{tw}$ 、 $F^t$ 、 $F^b$ それぞれの入力には別途エンコーダを用意する。なお、用いた言語モデルの最大入力長である128トークンに入力文長を合わせるため、パディングおよび超過分の切り捨てを適宜行う。多項ロジスティック回帰モデルの実装はscikit-learnのデフォルト<sup>\*4</sup>に準じるが、重みの更新回数の上限は3000に設定する。

データは、本研究で構築した主観性・情報継承性のラベルでそれぞれアノテーションされたデータセットを利用し、それぞれのタスクについて独立で10分割交差検証を行う。なお、同一の記事に言及する投稿が複数存在することを鑑みて、訓練データとテストデータへの各分割において、テストデータに含まれる全てのニュースへのリンクを取得し、訓練事例にこのリンクが含まれる投稿がある場合は全て除去した。

評価尺度としてはマクロ平均のPrecision(適合率)、Recall(再現率)、 $F_1$ 値を用いる。最初の実験として、提案手法の基本的な性能評価を行う。ベースラインは学習データ内の最多クラスを常に選ぶ最多クラス法とし、提案手法の入力は $F^{tw}$ のみとする。

表8に実験結果を示す。提案手法は全ての評価尺度で多数決法を大幅に上回っていることが確認できる。また、図6に示すモデルの予測結果の混同行列から、主観性・情報継承性ラベル共に一定の性能が確認できる。ただし、情報継承性については正解値を

<sup>\*2</sup> <https://huggingface.co/nlp-waseda/roberta-large-japanese>

<sup>\*3</sup> <https://github.com/ku-nlp/jumanpp>

<sup>\*4</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

入力	Precision	Recall	$F_1$
$F^{tw}$	<b>0.66</b>	<b>0.65</b>	<b>0.64</b>
$F^{tw} \& F^{ti}$	0.64	0.62	0.62
$F^{tw} \& F^{ti} \& F^b$	0.64	0.62	0.62

表9: 入力情報追加時の比較 (主観性).

入力	Precision	Recall	$F_1$
$F^{tw}$	0.47	0.45	0.45
$F^{tw} \& F^{ti}$	0.45	0.43	0.43
$F^{tw} \& F^{ti} \& F^b$	<b>0.48</b>	<b>0.46</b>	<b>0.46</b>

表10: 入力情報追加時の比較 (情報継承性).

問わず「高」の予測値が多く出ており、正解クラスあたりのデータ数が大幅に異なる不均衡データの問題に十分に対処できていない。さらに、情報継承性の正ラベルが「低」であるツイートを正しく分類できていないが、これはツイートのみを利用して情報継承性の判断を下すことの困難さを反映していると考えられる。

表9, 表10に  $F^{tw} \cdot F^{ti} \cdot F^b$  と性質の異なる文章を入力に追加したモデルの分類性能を示す。主観性の分類ではツイート本文のみを利用した場合の性能が最も良いが、これは主観性がツイート本文の情報のみで判断可能であることを示唆している。また、情報継承性については全ての入力を使った場合に最も高い性能を示しており、ツイートと言及されているニュース記事の文章同士を比較し、含意関係を捉えることの必要性が強調されている。

## 6 おわりに

本研究では、ニュース記事へ言及するTwitter上の投稿の分析を通じて投稿内容と元記事との関係を明らかにし、これを整理した上で「主観性」「情報継承性」の軸に沿ったラベル付きデータセットの構築を行った。また、このデータセットを使った分類問題を解く手法を提案し、性能評価を行った。

今後の研究方針として、Twitter投稿の主観性と情報継承性の分類タスクにおける提案手法を改良し、モデルの性能向上に向けた取り組みを行うこと、また、日本語の投稿が英語の記事を参照している場合など、言語横断的な情報伝播現象を捉えたデータセットの拡張を行うことを考えている。

## 謝辞

本研究は、三菱総合研究所が内閣官房の委託を受けて推進している「ポストコロナ時代の実現に向けた主要技術の実証・導入に向けた調査研究業務」の一環として、実施したものです。本研究で構築したデータセット、および使用した分類手法は、JST CREST JPMJCR19A4, および、JSPS科研費JP21H03445の支援を受けて開発したものです。

## 参考文献

[1] Shivangi Aneja, Chris Bregler, and Matthias Nießner. COSMOS: Catching out-of-context misinformation with self-supervised learning, 2021.

- [2] Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4685–4697, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [3] Terry Flew, Christina Spurgeon, Anna Daniel, and Adam Swift. The promise of computational journalism. *Journalism Practice*, 6(2):157–171, 2012.
- [4] Tirthankar Ghosal, Amitra Salam, Swati Tiwari, Asif Ekbal, and Pushpak Bhattacharyya. TAP-DLND 1.0 : A corpus for document level novelty detection. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).
- [5] Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, Minneapolis, Minnesota, USA, June 2019. Association for Computational Linguistics.
- [6] Lucas Graves. Understanding the promise and limits of automated fact-checking. *Reuters Institute for the Study of Journalism*, 2018.
- [7] Tamanna Hossain, Robert L. Logan IV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. COVIDLies: Detecting COVID-19 misinformation on social media. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online, December 2020. Association for Computational Linguistics.
- [8] Rina Kumari, Nischal Ashok, Tirthankar Ghosal, and Asif Ekbal. What the fake? probing misinformation detection standing on the shoulder of novelty and emotion. *Information Processing & Management*, 59(1):102740, 2022.
- [9] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.
- [10] Pew Research Center. Social media and news fact sheet, 2023.
- [11] Tanik Saikh, Tirthankar Ghosal, Asif Ekbal, and Pushpak Bhattacharyya. Document level novelty detection: Textual entailment lends a helping hand. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, pages 131–140, Kolkata, India, December 2017. NLP Association of India.
- [12] Gautam Kishore Shahi and Durgesh Nandini. FakeCovid - a multilingual cross-domain fact check news dataset for covid-19. *ArXiv*, abs/2006.11343, 2020.
- [13] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. FEVER: a large-scale dataset for fact extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [14] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA, June 2014. Association for Computational Linguistics.
- [15] World Health Organization. Infodemic, 2023.
- [16] Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. Mining dual emotion for fake news detection. In *Proceedings of the Web Conference 2021*, pages 3465–3476, 2021.
- [17] 読売新聞教育ネットワーク。中3の75%「snsでニュース知る」——読売・電通総研共同調査, 2022.