

ニュースコメントに関する対照学習を用いた ニュースの潜在的な関係性の可視化

高橋 陸[†] 牛尼 剛聰^{††}

[†] 九州大学大学院芸術工学府 〒 815-8540 福岡県福岡市南区塩原 4 丁目 9 - 1

^{††} 九州大学大学院芸術工学研究院 〒 815-8540 福岡県福岡市南区塩原 4 丁目 9 - 1

E-mail: [†]takahashi.riku.110@s.kyushu-u.ac.jp, ^{††}ushiama@design.kyushu-u.ac.jp

あらまし 近年 SNS の普及に伴い、ニュースに関する情報収集の際に SNS 投稿は重要な役割を果たすことがある。SNS を用いた情報収集では、ニュース記事のみからでは得られない情報や多様な視点からの意見を得ることができる。本論文では、SNS 上に投稿されたニュース記事とそれに対するコメントに着目し、Sentence-BERT による対照学習を用いてニュースとコメント間の関係の学習を行う。その後、学習済みモデルによりニュースに対するコメントをもとにした文章ベクトルを生成し、ニュース記事本文からでは推察することの難しいニュース間の潜在的な関連性を可視化する手法を提案する。本手法では、SNS におけるリプライ機能を利用する。リプライ関係に基づき、関連性の高い投稿から生成した文章ベクトルにおいてベクトル間の類似度が高くなるよう学習を行うことで、異なるニュースにまたがる類似コメントを明らかにし、情報収集の支援を実現する。

キーワード SNS, ニュース, ユーザ支援, 機械学習, 可視化

1 はじめに

近年、個人が手軽に情報を発信できる SNS(Social Networking Service) の普及に伴い、Twitter や Instagram など、多くの SNS で様々なトピックに関する投稿が行われている。SNS 上ではユーザがいつでも、どこでも簡単に文章を投稿できることから、多様なトピックに対する個人の意見や反応が多く存在する。また、SNS 上にはテレビ局や新聞社などの報道機関も取材などで得た情報をニュース記事として投稿しており、以上のことから、SNS は情報収集の手段として用いられることが多くなってきている。

情報収集における SNS の利用に関して、新聞通信調査会による「第 13 回メディアに関する全国世論調査」[1] によると、インターネットは情報源としての信頼度は新聞やテレビに劣る一方で、情報源として最も欠かせないメディアとして支持されている。また、NTT ドコモのモバイル社会研究所による「週 1 回以上アクセスして日常的に生活情報を得ているメディア」に関する調査[2] では、10 代 20 代においてソーシャルメディアの情報収集としての利用率はテレビや新聞による情報収集の割合を上回っており、特に若い世代を中心に SNS が情報源として広く利用されていることが分かる。

SNS を情報収集に用いることで、ユーザは多様な視点からの反応や意見を得ることが可能である。一方で、SNS 上に存在する情報はトピックごとや関連する内容に関して整理されておらず、特定のトピックに関する情報を網羅的に取得することは困難である。また、SNS 上に投稿されたニュース記事では、ニュースの詳細情報が記載されている一方で、因果関係や他の

ニュースとの関連といった情報は少ない。そのため、ニュース間の関連性の判断は、現時点ではユーザ個人の持つ知識や経験に依存している。

現在、ユーザが SNS を用いてニュースに関する情報収集を行った際に一般的に用いられている方法は、対象とするニュースに関する代表的なキーワードを用いたキーワード検索である。キーワード検索では、対象とするニュースに対してユーザが関係があると考える単語や、知りたい内容についてのキーワードをクエリとして投稿を検索する。例えば、新型コロナウイルスに関する情報を探す際には、「コロナ」や「感染状況」、「ワクチン」といった単語がクエリとして用いられることが多い。しかし、一般的に対象のニュースにおいて、そのニュースに関する SNS 投稿の全てに必ず含まれるような単語は存在しない。したがって、单一のキーワードを用いた検索では、ユーザが欲しい情報が十分に得られないことが多く、情報を網羅的に集めるにはキーワードを変えて何度も検索を行う必要がある。

SNS 上にニュース記事が投稿される場合、一般的にニュース記事本文ではニュースに関する詳細情報が述べられる。一方で、ニュース記事本文には、ニュースとなる事柄が起った経緯や他のニュースとの関係が分かる内容は少ないとから、单一のニュース記事から得られる情報から他のニュースとの関係を推察することは難しい。ここで、物価高騰に関するニュースを具体例として挙げる。物価高騰を扱ったニュース記事の本文では、現在の物価について過去のデータとの比較などが行われているが、なぜ物価高騰が起こっているのかについて十分に説明されていない。物価高騰の背景について、ウクライナにおける戦争の影響や日本の金融政策その年の気候など様々な要因が考えられる一方、こうしたトピックと物価高騰との関係について直接

言及している情報は少なく、従来のキーワード検索では特定のトピックに関する情報は取得できる一方で、複数のトピックにまたがるような話題、関係性について調べる際には適していない。現在、こうしたニュース間の関係性の判断は各ユーザが持つ経験や知識に委ねられており、SNS を用いた情報収集においても解決できていない。

以上の問題点から、現状では SNS を情報源として有効に活用できているとは言い難い。

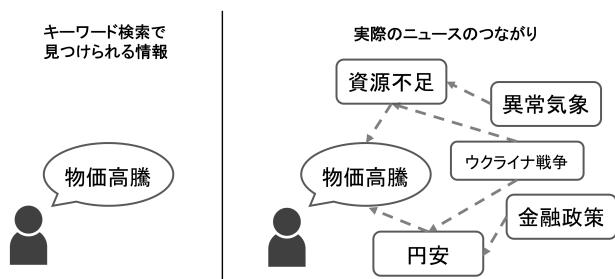


図 1 潜在的なニュースの関係性

本研究では、代表的な SNS の一つである Twitter を用いて、Twitter 上に投稿されたニュース記事とそれに対するリプライの関係に着目する。SNS 上に存在する投稿を利用し、機械学習を用いてニュースとニュースに対するコメントの関係性を学習する。その後、学習したモデルを用いてニュースやリプライの文書ベクトルを生成し、各ニュースや投稿の関係性を可視化することでユーザが情報収集を行う際の支援とすることを目的とする。

本研究の貢献は、以下の通りである。

- Sentence-BERT を用いて 2 文間の関係を学習することにより、単純な文章の類似性に寄らないニュースや SNS 投稿の反応を考慮した文書ベクトルを生成する。
- ニュース自体の内容だけでなく SNS におけるコメントを利用し、ニュースに対する反応という観点からニュースの特徴を抽出する事で、SNS による情報収集を支援する。

本論文の構成は以下のようになっている。第 2 章では、関連研究の紹介を行う。第 3 章では、提案手法について述べる。第 4 章では、提案手法の有効性を評価するための実験内容について述べ、実験結果について述べる。第 6 章では、まとめと今後の課題について述べる。

Sentence-BERT を用いて 2 文間の関係を学習することにより、単純な文章の類似性に寄らないニュースや SNS 投稿の反応を考慮した文書ベクトルを生成する。

2 関連研究

これまでにも、文章の特徴から、機械学習を用いて関連する情報を抽出する手法が提案されている。ここでは、それらの手法のうち、提案手法に関連する手法について述べる。

2.1 ニュース理解支援に関する研究

神谷 [3] らは、ウェブ上から社会問題に関する情報を抽出する手法を提案している。この手法では、対象とするウェブ上のテキストに対して、それが社会問題に関する内容であるかを判断し、関連する社会問題に分類している。この手法では、日本語 wikipedia 中の、社会問題に関する記事を教師データとして、自然言語処理の代表的な機械学習モデルの一つである BERT [4] を用いたテキスト分類モデルを構築している。この手法では、対象とするテキストが社会問題に関連するか否かを判定した後に、どの社会問題に関連するかを分類している。評価実験では、社会問題であるか否かの判定に関しては良い結果が得られなかった一方で、社会問題であると判定された記事を分類するタスクでは高い精度であった。

この手法では、社会問題との関係が明示的に示されていないテキストを、社会問題に分類するという点が本研究の提案手法と類似している。しかし、本論文の提案手法では SNS 上の投稿を学習データとして用いるため、Wikipedia に掲載されていないニュースに関しても対応可能である。

2.2 文書の特徴抽出に関する研究

大倉 [5] らは、ニュース記事に対する関連記事を判定するためのキーフレーズ抽出方法を提案している。この手法では、記事内容の特徴を捉え、かつ同一トピック内で広く使用されるフレーズを抽出するために、記事内での単語の出現頻度と類似記事における出現数からフレーズを抽出し、RNN [6] を用いて学習を行うことで関連記事を判定するためのキーフレーズを抽出している。

この手法では、文章中に現れるフレーズを用いて機械学習モデルに類似記事の学習を行っている。関連する記事を得るという点で本研究の目的を類似している。しかし、本研究では SNS 投稿を用いて間接的にニュース記事間の関係性を求めるため、記事中に含まれるフレーズに關係なくニュース記事の特徴を抽出するという点で異なる。

3 提案手法

本論文では、入力として用いる文章間の関係性を学習するため、代表的な自然言語処理モデルである BERT [4]、および BERT を文書ベクトルの生成に関して改良したモデルである Sentence-BERT [7] を用いる。BERT を用いる理由として、ニュースの理解支援のための「エアリップ」ツイート抽出 [8] におけるニュースに関する代表的なキーワードを含まない投稿の抽出において、他のモデルと比較して高い精度が得られたためである。Sentence-BERT は図 1 に示す様に、入力として与えられた二つの文章をそれぞれベクトル化し、文書ベクトルの類似性や関係性を学習することで、文章間の関係を文書ベクトルに反映するモデルである。Sentence-BERT は通常、2 つの類似した内容の文章を入力として用い、それぞれの文章のベクトルがベクトル空間上で近い位置になるよう学習を行うことで、類似文章検索などのタスクに用いられる。本研究では、入力文

としてSNSに投稿されたニュース記事の見出し部分と、それに対して行われたリプライを用いる。関係のある二つの文章のベクトルがベクトル空間上で近くなるよう学習を行うことで、ニュースとSNS投稿の関係性や投稿同士の関係性からニュースに対するSNS上の反応の特徴を抽出する。

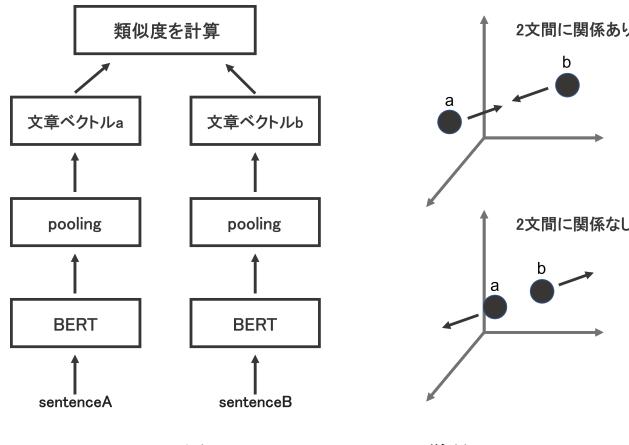


図2 Sentence-BERT の学習

本研究では、ニュースコメントを用いた機械学習を行うことで、ニュース間における潜在的な関係性を可視化する手法を提案する。本章では提案手法について、学習・ベクトル生成の手法と可視化の手法の2つに分けて紹介する。

3.1 ニュース記事およびニュースコメントを用いた文章間の関係性学習

本研究では、Sentence-BERTを用いて以下の2つのモデルの学習を行う。

- ニュース-コメントモデル
- コメント-コメントモデル

3.1.1 ニュース-コメントモデル

ニュース-コメントモデルについて、このモデルではニュース記事およびニュースコメントを入力とする。Twitter APIを用いて、報道機関により投稿されたニュース記事とそれに対するリプライを収集し、Twitterのリプライ機能で明示されたリプライ関係の有無に基づき、ラベルを付与する事でデータセットを作成する。

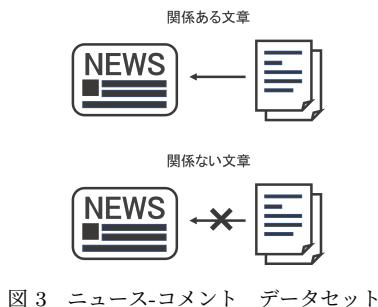


図3 ニュース-コメント データセット

モデルの学習では、リプライ関係が存在する場合は入力されたニュース記事タイトルとニュースコメントの文章ベクトルが

ベクトル空間上で近い位置に存在するよう学習を行い、リプライ関係が存在しない場合はベクトル空間上の位置が遠ざかるよう学習を行う。学習を行なったモデルによりニュースコメントの文章ベクトルを生成する事で、ニュース内容とコメント内容を考慮した文章ベクトルを生成する。

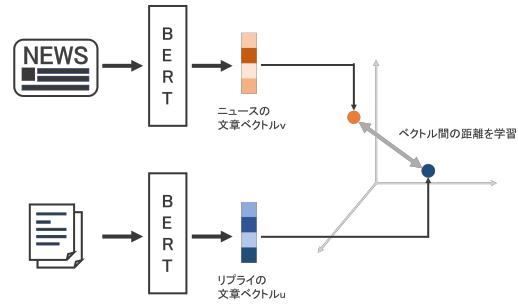


図4 ニュース-コメントモデルの学習

3.1.2 コメント-コメントモデル

次に、コメント-コメントモデルについて説明する。このモデルでは、ニュースコメント同士の関係からデータセットを作成する。学習に用いるデータセットは、Twitterのリプライ機能に基づき同一のニュースに対するコメントか否かのラベル付けを行う。

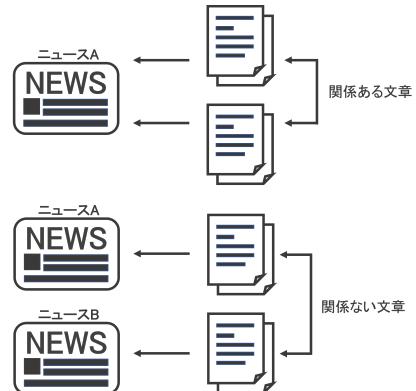


図5 コメント-コメント データセット

モデルの学習では、2つのニュースコメントを入力とし、入力が同一のニュースに対するコメントである場合はその文章ベクトルがベクトル空間上で近くなり、異なるニュースに対するコメントの場合はベクトル空間上の位置が遠くなるように学習を行う。このモデルでは、コメント同士の関連性の学習からニュース内容に依存しないニュースコメント間の類似性を考慮した文章ベクトルを生成することを目的とする。

3.2 潜在的なニュース間の関係の可視化

各モデルにより生成したニュースコメントの文章ベクトルを用いる事で、ニュース記事のみでは得ることの難しい、潜在的なニュース間の関係を可視化する。本研究では、BERTまたはSentence-BERTにより生成された768次元の文章ベクトルに対し、t-SNE[9]を用いる事で2次元に変換する。t-SNEを用いることで、Sentence-BERTにより生成された768次元の文

章ベクトルを2次元に変換する際に、各ベクトル間の距離の情報を保つつつ次元圧縮が可能である。その後、2次元に変換したニュースコメントの文章ベクトルとニュースコメント本文からデータベースを作成し、このデータベースを基にニュース記事ごとに色分けされた散布図を作成する。散布図上からユーザが気になるデータ点を選択することで、各ニュースにおける代表的なコメントやニュース間の境界に位置するコメントを参照する。

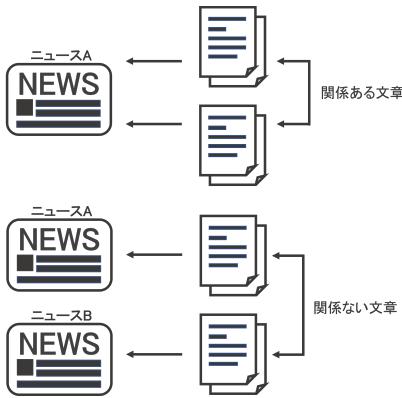


図 6 コメント-コメント データセット

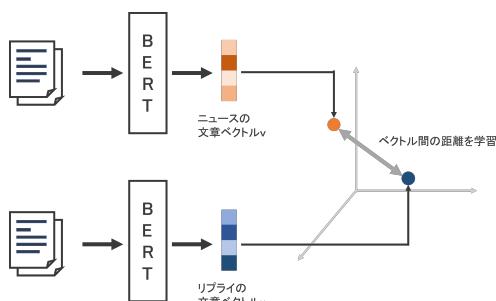


図 7 コメント-コメントモデルの学習

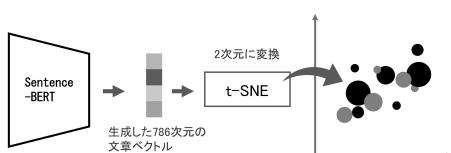


図 8 t-SNE による文章ベクトルの変換

4 評価実験

4.1 実験方法

本実験では提案手法を用いて、ニュース-コメント、コメント-コメント、BERT の3つのモデルに関して、同一のデータセットにより学習を行ない、ニュースコメントの文章ベクトルを生成した。各モデルに対する評価として、ニュースコメントの文章ベクトルと学習に用いた各ニュース記事タイトルの文章ベクトルの類似度を計算し、類似度に対して閾値を設定すること

で、ニュースコメントとニュース記事間の関係の有無の判定の精度を調べた。また、提案手法によるニュース内容とコメント内容を考慮した文章ベクトルの生成について、ニュースコメントの文章ベクトルから各ニュース記事との類似度を求め、類似度の Top_k 内に正解のニュースが含まれるかどうかを調べる事で評価を行なった。最後に、Sentence-BERT によって生成された 768 次元の文章ベクトルに対し、ニュースコメントが行われたニュースごとに DBSCAN を行った。距離 8 以内、クラスタを構成する最小データ数を 30 として DBSCAN を行うことで、各ニュースに対するコメントに含まれるニュース理解の参考にならないようなノイズコメントの除去を行った。その後、t-SNE を用いて次元削減を行うことで2次元に変換し、散布図に可視化することで各ニュース間の関係について分析した。

また、比較手法として BERT モデルを用いた。このモデルでは、リプライ関係に基づきニュースコメントを入力、ニュース記事を正解ラベルとして分類タスクを行い、その後、学習を行なった BERT モデルの最終層の cls トークンを用いて文章ベクトルを生成した。

データセットとして、本研究では代表的な SNS の一つである Twitter を用いた。Twitter API を利用して、投稿されたニュース記事 16 件と各ニュースに対するニュースコメントの取得を行なった。図 3, 5 に示すように、取得したニュース記事とコメントについて、Twitter のリプライ機能で関係が明示されたニュース記事とニュースコメントには関連性が存在するとし、リプライとして関連が明示されていない投稿間には関連性が存在しないとしてラベル付けを行い、データセットを作成した。データセットの正例と負例の数は 1:1 とし、合計 14282 件のデータセットを作成した。

表 1 データセットに用いたニュース一覧

ニュース記事見出し
岸田首相は当初は積極的に取材に応じる姿勢を示していたが…
先進 7 カ国 7 は 49 日オンラインで首脳会議を開きロシア産石油…
日経平均株価は先週末の終値より 684 円安い 2 万 6319 円 34 銭で…
新型コロナウイルスの水際対策について緩和すべきと思う人…
財務省は 10 日税収で将来返済する必要がある国の借金長期債…
新型コロナウイルスについて東京都は 8 日新たに 4711 人の感…
東京都医師会は 101 日午後に行われた定例会見で段階的に感染…
小池氏、今度は「神宮外苑再開発」で樹齢 100 年の樹木を約 1000…
ロシアの国連代表部に勤務してきた露外交官、ボリス…
近年、「体育座り」は子供の体に負担がかかるとの指摘があり…
マイナンバーカードと健康保険証の機能を併せ持つ「マイナ保険…
国民審査 “在外の日本人が投票できないのは違憲” 最高裁判決
東京都は、一戸建て住宅を含む都内の新築建物に、太陽光パネル…
「他意はなく、不用意な発言だった。心からおわびしたい」と…
性行為伴う AV 禁止する法制定を別途検討 立憲民主党が方針
【誤送金問題】「約 4299 万円を法的に確保した」阿武町長が会見…

4.2 ニュース-コメント

注目するニュース記事に関して、リプライ機能で関係が明示されているニュースコメントはニュース記事との関連性が存在

するとし、異なるニュース記事に対するコメントとには関連性が存在しないとして、Sentence-BERT の入力として用いるデータセットを作成した。このデータセットを用いてニュース記事とニュースコメントの関係について学習を行い、学習したモデルを用いてニュースコメントとニュース記事タイトルの文書ベクトルを生成した。その後、ニュースコメントと各ニュース記事タイトルとの類似度を計算し、2 文間の関係性の有無の判定とニュースコメントのニュース記事への分類を行った。最後に t-SNE を用いて文書ベクトルを 2 次元にすることで可視化を行った。

4.3 コメント-コメント

同一のニュース記事に対するコメント間には関係があるとし、異なるニュースに対するコメント間には関係が存在しないとしてデータセットを作成した。作成したデータセットを用いて各ニュースに対する Twitter 上の反応に関してモデルの学習を行い、学習したモデルを用いてニュースコメントとニュース記事タイトルの文書ベクトルを生成した。その後、ニュースコメントと各ニュース記事タイトルとの類似度を計算し、2 文間の関係性有無の判定とニュースコメントの分類を行い、t-SNE を用いて文書ベクトルを 2 次元にすることで可視化を行った。

4.4 結 果

学習したモデルがニュースコメントの関係性を正しく学習できているかを調べるために、ニュースコメントの文書ベクトルを用いて類似度を計算した。ニュースコメント間の関係性の判定では、求めた 2 文間の類似度に対して閾値を設定することにより 2 文間の関係の有無を判定した。図 9 より、ランダムに選ばれた 2 つのニュースコメントが同じニュースに属するものであるか同課の判定では、ニュース記事のタイトルとニュースコメントの関係を学習したモデルが最も精度が高く、ニュースコメント同士の関係を学習したモデルにおいても一定の精度が得られたことが分かる。

次に、ニュースコメントからそのコメントが属するニュースの予測を行った。ニュースコメントから生成した文書ベクトルに対して各ニュースのタイトルから生成した文書ベクトルとの類似度を求め、その類似度 top k に正解のニュースが存在するかを調査した。図 10 を見ると、ニュース-コメントの学習を行ったモデルでは、ニュースコメントと正解ニュースとの類似度が高く、予測できていることが分かる。一方で、コメント同士の学習を行ったモデルでは正確な予測は難しいことが分かった。また、表 2 に示す BERT によるニュースコメントのニュース記事への分類精度と比較すると、提案手法によるニュースコメントからコメントが属するニュースの予測の精度は低い結果となった。

表 3、表 4 はニュース-コメントモデル、コメント-コメントモデルそれぞれにおいて、各ニュースが他のどのニュースとの関連性が高いかを示したものである。各ニュースに対するニュースコメントから文書ベクトルを生成し、ニュースごとのニュースコメントの平均ベクトルから類似度を比較することで、

類似度が最も高いニュースを関連ニュースとした。この結果から、ニュース-コメントモデルではニュース内容が類似しているニュースが類似度が高くなっている、ニュース内容を考慮したニュースコメントのベクトル生成が出来ていると考えられる。一方で、コメント-コメントモデルではニュース-コメントモデルと比較するとニュース内容の関連性を正しく学習できているとは言えない。これは、コメント-コメントモデルではモデルの学習はコメント同士で行っており、ニュース内容が含まれないためであると考えられる。

続いて、入力に用いた文章間の関係を学習したモデルによる文書ベクトルを散布図に可視化した図について確認する。Sentence-BERT 及び BERT による文書ベクトルは 768 次元で表現されており、これを散布図にプロットするにあたって t-SNE を用いて 2 次元にベクトルを変換した。また、リプライとニュースの分類を行った BERT による文書ベクトルは最終層の CLS トークン部分を用いた。

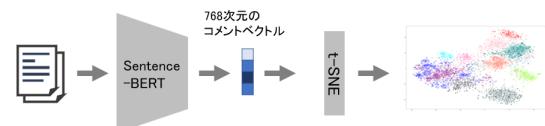


図 9 文章ベクトルの可視化の流れ (DBSCAN なし)

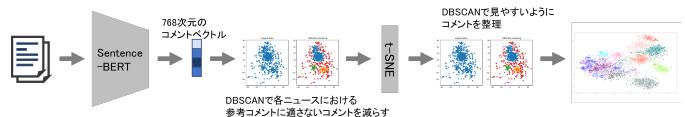


図 10 文章ベクトルの可視化の流れ (DBSCAN 適用)

図 11~13 は各ニュースに対するコメントの文書ベクトルを、t-SNE を用いて 2 次元に変換し散布図にしたものである。図 11 から、ニュース記事のタイトルとそれに対するコメントの関係を学習したモデルでは、内容の類似したニュースに対するコメントの文書ベクトルが近くに現れており、投稿の文書ベクトルにニュースの内容を含んでいるといえる。一方で、図 12 を見ると、ニュースコメント同士の関係を学習したモデルによる文書ベクトルでは、ニュースの内容は考慮されておらず、ニュースに対する純粋な SNS 上の反応の類似性を表している。図 13 の BERT による文書ベクトルでは、Sentence-BERT を用いた場合と比べ、ニュースごとに SNS 投稿の文書ベクトルの分布が固まっていることが分かる一方、各ニュース記事間の接面がはっきりせず独立していることから、BERT を用いた分類タスクによる文書ベクトルの生成では、ニュース間の関係やコメントの類似性は考慮されておらず、本論文の目的には合わない結果となった。

4.5 考察

実験結果より、提案手法によりニュースの内容を考慮した文書ベクトルとニュースに対するSNS上の反応を特徴を表す文書ベクトルを生成できたと考える。一方で、リプライの文書ベクトルからニュースを予測するタスクでは、リプライ同士の学習を行ったモデルでは精度が低い結果となった。これは、リプライにはニュースの内容が含まれているとは限らず、生成した文書ベクトルにはニュースの内容が含まれていないためだと考えられる。また、表3, 4に示したように、ニュースの予測ではなく関係性の強弱を求める場合では2つのモデルで似た結果となっていることから、それぞれのモデルにおいてニュースごとの特徴を学習できていると考えられる。

表2 BERTによるリプライのニュース分類

	val acc	test acc
BERT	0.734375	0.7058

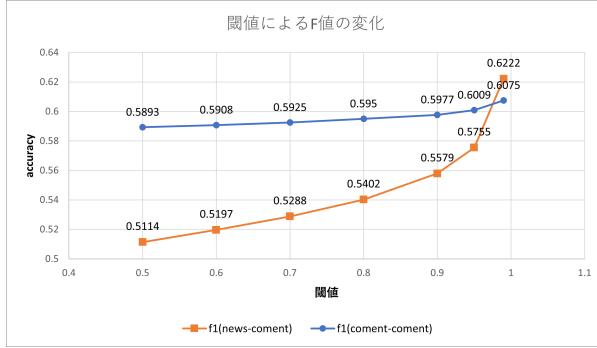


図11 二つの文章間の類似度による関係性有無の判定

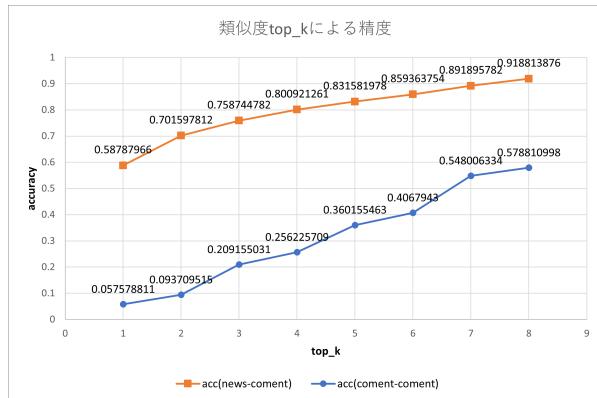


図12 リプライとニュースの予測精度

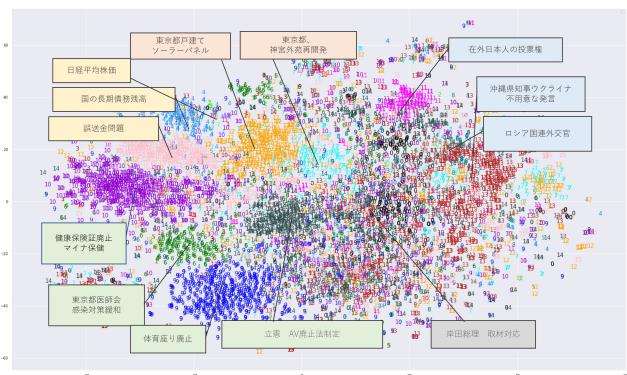


図13 文章ベクトルの可視化（ニュース-リプライ）

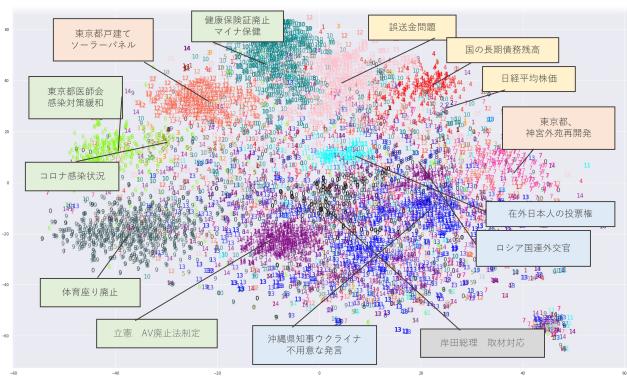


図14 文章ベクトルの可視化（リプライ-リプライ）

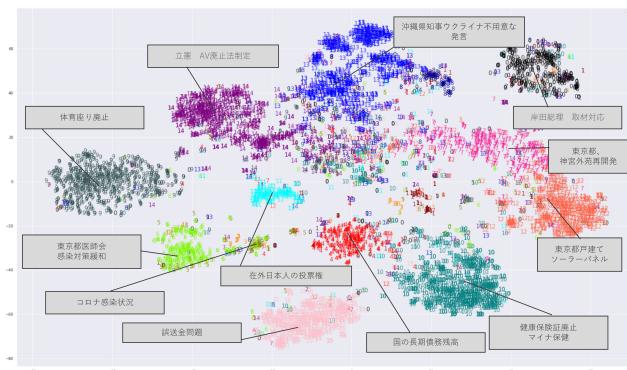


図15 文章ベクトルの可視化（BERT）

表 3 類似度の最も高いニュースの例（ニュース-リプライ）

ニュース	類似度の最も高いニュース
岸田首相は当初は積極的に取材に応じる姿勢を示していたが…	ロシアの国連代表部に勤務してきた露外交官、ボリス…
先進 7カ国 7は 49 日オンラインで首脳会議を開きロシア産石油…	財務省は 10 日税収で将来返済する必要がある国の借金長期債…
日経平均株価は先週末の終値より 684 円安い 2 万 6319 円 34 銭…	財務省は 10 日税収で将来返済する必要がある国の借金長期債…
新型コロナウイルスの水際対策について緩和すべきと思う人…	東京都は、一戸建て住宅を含む都内の新築建物に、太陽光パネ…
新型コロナウイルスについて東京都は 8 日新たに 4711 人の感…	東京都医師会は 101 日午後に行われた定例会見で段階的に感染…
小池氏、今度は「神宮外苑再開発」で樹齢 100 年の樹木を…	ロシアの国連代表部に勤務してきた露外交官、ボリス…
近年、「体育座り」は子供の体に負担がかかるとの指摘があり…	新型コロナウイルスの水際対策について緩和すべきと思う人…
マイナンバーカードと健康保険証の機能を併せ持つ「マイナ保険…	【誤送金問題】「約 4299 万円を法的に確保した」阿武町長が会…
国民審査 “在外の日本人が投票できないのは違憲” 最高裁判決	小池氏、今度は「神宮外苑再開発」で樹齢 100 年の樹木を…
「他意はなく、不用意な発言だった。心からおわびしたい」と…	性行為伴う A V 禁止する法制定を別途検討 立憲民主党が方針

表 4 類似度の最も高いニュースの例（リプライ-リプライ）

ニュース	類似度の最も高いニュース
岸田首相は当初は積極的に取材に応じる姿勢を示していたが…	「他意はなく、不用意な発言だった。心からおわびしたい」と…
先進 7カ国 7は 49 日オンラインで首脳会議を開きロシア産石油…	日経平均株価は先週末の終値より 684 円安い 2 万 6319 円 34 銭…
新型コロナウイルスの水際対策について緩和すべきと思う人…	マイナンバーカードと健康保険証の機能を併せ持つ「マイナ保険…
財務省は 10 日税収で将来返済する必要がある国の借金長期債…	東京都は、一戸建て住宅を含む都内の新築建物に、太陽光パネ…
新型コロナウイルスについて東京都は 8 日新たに 4711 人の感…	東京都医師会は 101 日午後に行われた定例会見で段階的に感染…
小池氏、今度は「神宮外苑再開発」で樹齢 100 年の樹木を…	ロシアの国連代表部に勤務してきた露外交官、ボリス…
近年、「体育座り」は子供の体に負担がかかるとの指摘があり…	新型コロナウイルスについて東京都は 8 日新たに 4711 人の感…
マイナンバーカードと健康保険証の機能を併せ持つ「マイナ保険…	【誤送金問題】「約 4299 万円を法的に確保した」阿武町長が会…
国民審査 “在外の日本人が投票できないのは違憲” 最高裁判決	小池氏、今度は「神宮外苑再開発」で樹齢 100 年の樹木を…
「他意はなく、不用意な発言だった。心からおわびしたい」と…	性行為伴う A V 禁止する法制定を別途検討 立憲民主党が方針
【誤送金問題】「約 4299 万円を法的に確保した」阿武町長が見…	財務省は 10 日税収で将来返済する必要がある国の借金長期債…

5 おわりに

本論文では、ニュース記事の内容を考慮した文書ベクトルの生成と、ニュースに対する SNS 上の反応の特徴を埋め込んだ文書ベクトルを生成する手法を提案した。実験の結果から、Sentence-BERT を用いて、ニュースと SNS 投稿の関係性を学習することで、SNS 上に投稿されたニュース記事の特徴を、SNS 投稿を用いて間接的に抽出可能であることが示された。今後の課題として、生成された文書ベクトルから、各ニュース同士の境界における SNS 投稿の内容の確認や、代表的な SNS 投稿の抽出を行うことで、より良いニュースの理解支援の実現を目指す。また、Sentence-BERT を用いた手法では、入力された文章のベクトルから、類似度を計算することで関連性の有無の判定を行う。そのため、学習に用いていないニュースに対しても対応できる可能性があると考えられる。今後は、より大規模な学習を行った際の、未知のニュースに対する文書ベクトルの生成と分類精度などの調査を行いたいと考えている。

向調査

- [3] 神谷晃, 長谷川徳賢, 白松俊. 市民協働のための web 記事上の社会問題の自動タグ付けと関連事例抽出手法. 人工知能学会全国大会論文集 JSAI2020.
- [4] Devlin Jacob, Chang Ming-Wei, Lee Kenton, Toutanova Kristina. Bert: Pre-training of deep bidirectional transformers for language understanding. In *arXiv preprint arXiv:1810.04805*, 2018.
- [5] 大倉俊平, 小野真吾. 関連記事判定のためのニュース記事キーワード抽出.
- [6] Williams Ronald J, Zipser David. A learning algorithm for continually running fully recurrent neural networks. In *Neural computation*, 第 1 卷, pp. 270–280. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info …, 1989.
- [7] Nils Reimers, Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks.
- [8] 高橋陸, 牛尾剛聰. ニュースの理解支援のための「エアリブ」ツイート抽出. In *DEIM Forum 2022 B24-1*, 2022.
- [9] Laurens van der Maaten, Geoffrey Hinton. Visualizing data using t-sne. In *Journal of Machine Learning Research* 9(2605), pp. 2579–2605, 2008.

謝 辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

文 献

- [1] 新聞通信調査会. 第 13 回メディアに関する全国世論調査.
- [2] NTT ドコモモバイル社会研究所. 2021 年一般向けモバイル動