

Twitterの反応を利用したニュースコンテキストの理解支援

池田 将† 牛尼 剛聡††

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

†† 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 4-9-1

E-mail: †ikedasho.294@s.kyushu-u.ac.jp, ††ushiana@design.kyushu-u.ac.jp

あらまし 近年の SNS の急速な普及に伴い、Twitter などの SNS を利用してニュースを閲覧するユーザが増えている。従来の新聞やテレビのニュースでは、解説者やコメンテータがニュースに対する意見を述べることで、読者や視聴者のニュース理解を助ける役割をしている。それに対して SNS では一般のユーザが、ニュースに対する多様な視点、意見をメンションとして投稿し、他のユーザはそれらを閲覧することにより、ニュースの理解を深めている。本研究では、ニュースが潜在的に有する前後関係や背景をニュースコンテキストとし、ニュースに対するメンションから、ニュースコンテキストを推定する方法を提案する。さらに、ニュースコンテキストの特徴を利用し、ユーザが閲覧するニュースに対して、前後関係やトピックが共通するニュースを抽出し、ユーザがニュース全体像を理解することを支援する。キーワード SNS, ニュース, テキスト処理, コンテキスト推定

1 はじめに

SNS(ソーシャル・ネットワーキング・サービス)は、ユーザの交友関係に基づいてコミュニケーションを行うインターネット上のサービスである。特に、代表的な SNS の一つである Twitter [1] は、ユーザはツイートと呼ばれる 140 字以内の短い文章や画像を投稿し、他のユーザがそれに対しての返信などを行うことで、ユーザ間の緩やかなコミュニケーションを促すサービスである。Twitter の利用者は、2019 年 4 月時点で約 3 億 3,300 万人いるとされる [2]。近年では、企業や公的機関など、多くの組織も Twitter 上にアカウントを開設し、情報を提供している。

このような Twitter の爆発的な普及に伴い、新聞社やテレビ局、インターネットのニュースサイトなど、多くのメディアが Twitter 上にアカウントを開設し、ニュースの概要や見出しを投稿している。ユーザは Twitter 上で見たニュースに興味を持った場合、そのツイートに記載されたリンクをクリックすることで、Web 上の記事本文を読むことができる。また、ツイートに対するリプライや引用リツイートなど、他のユーザがニュースに対して行った反応を読むことができる。

我々は、それぞれのニュースには背景や前後関係などといった文脈情報があると考えている。本稿では、これらの文脈情報を「ニュースコンテキスト」と呼ぶ。ニュースコンテキストの例を図 1 に示す。例えば、2019 年 8 月 24 日に朝日新聞によって報じられた、『秋田市に陸上イージス「もうダメ」地元自民議員が反対』というニュース [3] に対しては、「防衛」や「国防」についての背景、「国会」やそれに関連する「議員」についての情報、「秋田県民」がどう考えているか、などといったことがニュースコンテキストとして存在すると考えられる。ニュースを読むユーザは、対象とするニュースだけでなく、「ニュースコンテキスト」を知ることで、そのニュースをより深く理解でき

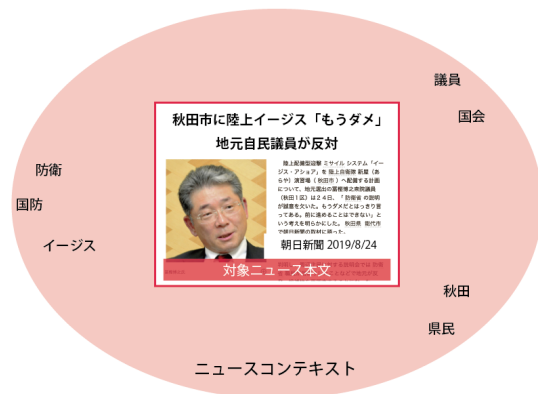


図 1 ニュースとニュースコンテキストの例

ると考えられる。

しかし、SNS ではニュースがユーザのタイムライン上に断片的に現れるため、全てのニュースを閲覧することは難しく、必要な情報を見逃してしまうことがある。また、必要なニュースがユーザのタイムラインに出現しないこともある。そうしたことから、SNS を利用したニュース閲覧において、ニュースコンテキストを理解するのは難しいと考えられる。

次に、従来のマス・コミュニケーションにおける情報伝達の仕組みについて述べる。藤江 [4] は、マス・コミュニケーションの社会的機能として、以下の 3 つをあげている。

- (1) 報道的機能
- (2) 教育的機能
- (3) 娯楽的機能

このうち、(1)の報道的機能には、マスメディアによって提供される間接的な環境について、受け手が直接的に体験できない環境を監視し、情報を提供する第一の機能の他に、その環境の全貌を受け手が十分に理解するための手助けをする分析、解説の第二の機能、環境変化に適応し、問題解決をするため、様々な

立場の専門家が意見や態度を提示し、社会の合意を助ける第三の機能があるとしている。また、(2)の教育的機能は、世代から世代へ、ある集団からある集団へ、情報や価値、社会的規範を伝達する機能であるとしている。SNS上に現れるニュース記事は、これらのうち、(1)の情報を提供する機能のみを持つものであり、ユーザの理解を手助けするシステムを構築することに意義があると考えられる。

Twitterでは、ニュースに対する意見を発信するために、通常のツイート以外にも、ニュースや意見を共有する「リツイート」、リツイートを行いつつユーザ自身の意見も発信する「引用リツイート」、ニュースや意見に対し返信を行う「リプライ」がある。これらのうち、リプライと引用リツイートは、様々なユーザがもつ知識や意見に基づいて発信したものであり、ニュースコンテキストの推定に役立つと考えられる。

そこで、本研究では、SNS上のメンションを利用して、対象とするニュースに対してニュースコンテキストが関連するニュースを提示することで、ユーザのニュースコンテキストの理解を支援する手法を提案する。ニュースコンテキストが関連するニュースの例を図2に示す。ユーザは現在閲覧しているニュースだけでなく、提示されたニュースコンテキストが関連するニュースも合わせて読むことで、ニュースの背景や位置付け、そのニュースの全体像の理解を支援できると考えられる。

本稿では、上記を実現するための手法についての具体的な手順を述べる。また、提案手法の有効性の評価のために、行った実験について、その結果を示し、その結果を考察する。第2章では本研究と関連する研究について述べる。第3章では本研究のアプローチについて述べる。第4章では提案手法のうち、ニュースやメンションをコンピュータ上で扱うことのできる特徴ベクトルにする手法を示す。第5章ではニュース間の関連度の算出手法について述べる。第6章では実験について述べる。第7章ではまとめと今後の課題について述べる。

2 関連研究

本研究では、ユーザの深いニュース理解を促すために、他のユーザの反応を利用したニュースコンテキストの推定と、それを利用したニュース推薦を行う。そこで本章では、ニュース推薦に関する既存研究と、Webニュースでの理解支援に関する研究について述べる。

2.1 ニュース推薦に関する研究

ユーザの興味を推定し、それに適したニュースを推薦する研究はこれまでもいくつか行われており、それらのいくつかのはWebニュースを提供するWebサイトにおいて、実際に実装されている。

Liら[5]は、新しいニュースコンテンツについても効率よくパーソナライズされた推薦を行うために、バンディット・アルゴリズムを用いたニュース推薦手法を提案している。

Okuraら[6]は、閲覧履歴を入力シーケンスとしたリカレント・ニューラルネットワークを利用したニュース推薦手法を提

案している。

Yonedaら[7]は、時間減衰関数を埋め込んだクラスタベースの協調フィルタリングを行うことで、ニュースの即時性に対応したニュース推薦手法を提案している。

本研究では、ニュースが潜在的にもつコンテキストに着目し、それに基づく推薦を行うことを目標としており、これら協調フィルタリングを利用したニュースの推薦とは異なる。

2.2 ニュース理解支援に関する研究

ニュースに対して様々な視点や関連情報を提供し、ユーザのニュース理解を支援する研究として、次のようなものがある。

Rennisonら[8]は、ニュース記事間の関連性から構築されたニュース情報空間を可視化することで、ユーザがニュースについて幅広い理解を迅速に得ることができるシステムであるGalaxy of Newsを提案している。

Nadamotoら[9]は、ユーザが指定したニュースと関連した情報について、時系列順に提示することで、ユーザにニュースの進捗状況を理解させることのできるTime-based Contextualized-News Browser (T-CNB)を提案している。

Liuら[10]は、ジャーナリストは自身が属する社会集団の文化的規範と価値に影響を受けニュースを制作することから、読者に異なる社会集団から発信された記事をまとめて提示することで多様な視点を提供するLocalSavvyというパラダイムを提案している。

田中ら[11]は、ニュース記事について、そのニュースの中心的な出来事について述べた主題事象についての記述と、その出来事の経緯や位置付けを示す過去の出来事である背景・前提事象についての記述に分類し、後者を抽出する手法を提案している。

片岡ら[12]はGoogle検索においてユーザが気付かないうちに閲覧する情報に偏りが生じる問題を解決するため、ユーザにパーソナライズされた検索結果とそうでない結果との差をフィルターバブルの度合いとしてユーザに提示し、認知させることで、ユーザにフィルターバブルの存在を自覚させ、ユーザが情報の探索領域を広げることを促すユーザ・インターフェースを提案している。

切通ら[13]はユーザがニュースイベントに対して様々な視点から理解するのを支援するためのニュースアプリケーションNewsSaladを提案している。NewsSaladでは、ニュースを意見、視点、詳しさの3つの尺度で定量化し、ユーザに対し閲覧中の記事と同じイベントでかつその差異の尺度が最も大きい記事を提示する。

本研究では、ニュース本文内に直接的に現れないような潜在的な背景や前後関係も考慮するため、SNSでのユーザの反応を利用して、ニュースのコンテキストを推定する。

3 アプローチ

本研究では、ニュースの潜在的な前後関係や背景であるニュースコンテキストを利用して、ユーザのニュース理解を深めるこ

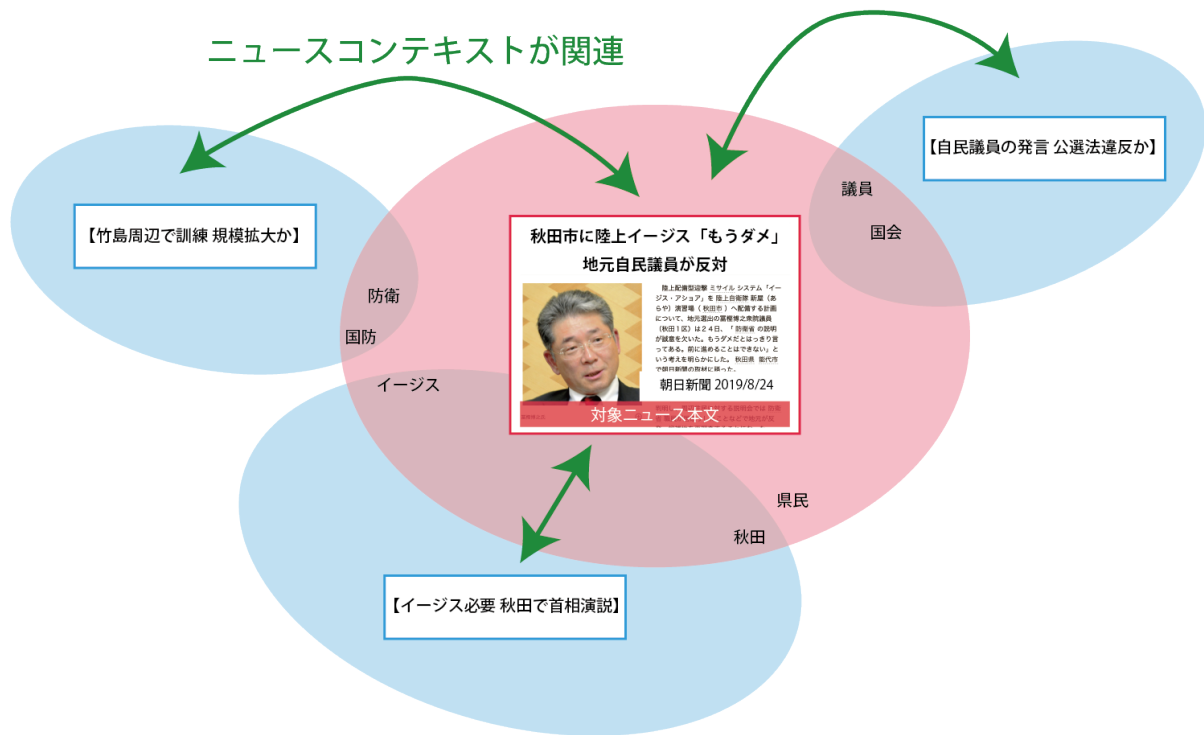


図2 ニュースコンテキストが関連するニュースの例

とを促す関連ニュースを抽出する。

ユーザに配信されるニュースは、主として伝える出来事と、その背景に関する記述から構成される。しかし、それぞれのニュースについての背景に関する記述は短くまとめられており、その出来事に関しては、記述されているより膨大な背景が存在することが多い。本研究では、ニュースに関する多様な背景や前後関係であるニュースコンテキストを知ることが、ニュースの全体像の理解を促進すると考える。

本研究では、ニュースコンテキストを推定する方法として、SNSでのメンションを利用する。本研究ではメンションとして、Twitterでの引用リツイートとリプライを利用する。これらを利用する理由は、SNSユーザはニュースに対して、ユーザが理解しているニュースコンテキストに基づいて意見を発信するためである。様々なユーザが多様な意見をメンションとして発信しているため、様々な視点を考慮したニュースコンテキストの推定が期待できる。

本研究でのアプローチの概念図を図3に示す。本研究で開発するシステムでは、ユーザが閲覧しているニュースに対して、ニュースコンテキストが関連するニュース記事を、ニュースデータベースの中から抽出し、ユーザに関連ニュースとして提示する。ニュース間のニュースコンテキストの関連度を算出する手法の概念図を図4に示す。ニュース間のニュースコンテキストの関連度を算出する手法では、主な処理として以下の2つがある。(1) ニュース本文やニュースに対するメンションからニュースコンテキストを考慮した特徴量を作成する処理を4章にて述べる。(2) 算出した特徴量を利用して、ニュース間の関連度を算出する処理を5章にて述べる。

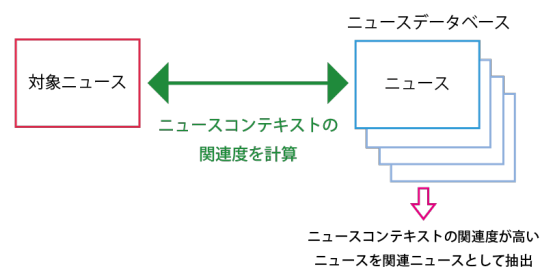


図3 アプローチの概念図

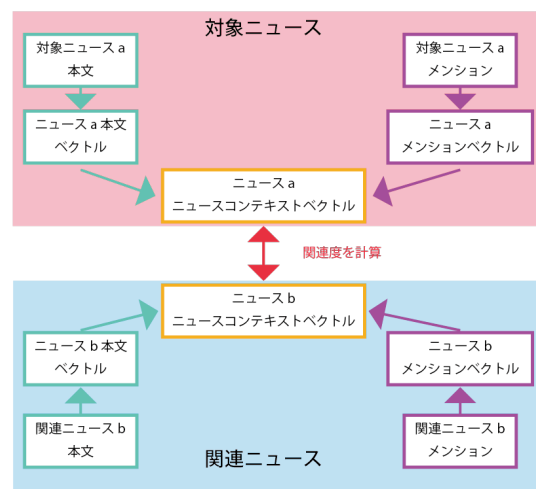


図4 ニュースコンテキスト関連度の算出

4 ニュースコンテキストを考慮した特徴量の作成

本研究では、個々のニュース本文や反応を、BoWを利用し

てモデル化する．それぞれのニュースとメンションの文書中に含まれる単語について，ニュースコンテキストとしての重要なキーワードの重みを重くすることで，ニュースコンテキストを考慮する．

また，ニュースコンテキストをコンピュータ上で扱えるようにするため，ニュース本文，メンションを特徴ベクトルにする．

4.1 ニュースコンテキストの抽出

ニュースコンテキストとして重要になる単語に重みをつける方法として，以下の3つの手法を検討する．

4.1.1 idf 法

Yoneda ら [7] は，ニュース記事の特徴ベクトル \mathbf{a} を，ニュース記事に含まれる単語の分散表現を利用し，ニュース記事 a に含まれる単語集合 W_a ，単語 w_i の特徴ベクトル \mathbf{w}_i として，次のように定義している．

$$\mathbf{a} := \frac{\sum_{w_i \in W_a} idf(w_i) \mathbf{w}_i}{\|\sum_{w_i \in W_a} idf(w_i) \mathbf{w}_i\|} \in \mathbb{R}^d \quad (1)$$

これに倣い，本研究でも idf を語の重みとする方法を検討する．idf (Inverse Document Frequency) は，文書中の語の重要度を評価する代表的な手法である tf-idf 法 [14] で利用される特徴量の一つであり，全文書に対してある単語が現れる文書数の割合の逆数である．この特徴量では，出現している文書の割合が低い単語に対して値が大きくなる．

本論文では，単語 w_i の idf 値を，全文書数を N ，単語 w_i の出現した文書数を $df(w_i)$ とした時，以下のように定義する．

$$idf(w_i) = \log \frac{N + 1}{df(w_i) + 1} + 1 \quad (2)$$

例として，以下の【ニュース 1】のメンションにおける，idf 値の高い単語の例を，表 1 に示す．

【ニュース 1】 2019/8/24

【イージス 秋田の自民議員反対】陸上イージスを秋田市に配備する計画について、地元選出で自民の富樫博之衆院議員が「防衛省の説明が誠意を欠いた。もうダメだとはつきり言ってある、前に進…

表 1 【ニュース 1】のメンションにおける idf 値が高い語の例

単語	idf 値
切取り	7.73221
新屋	7.73221
兵力	7.73221
中泉	7.73221
線分	7.73221

4.1.2 tf-idf 法

tf-idf 法は，先述の idf に文書内の単語の出現頻度である tf(Term Frequency) をかけた tf-idf を重みとすることで，文書内での出現頻度も考慮した特徴を抽出する．ある文書 d_j に出現する w_i の tf 値を，文書 d_j 中に出現する w_i の回数 $f(w_i, d_j)$ を利用し，以下のように定義する．

$$tf(w_i, d_j) = \frac{f(w_i, d_j)}{\sum_{w_k \in d_j} f(w_k, d_j)} \quad (3)$$

ある文書 d_j に出現する w_i の tf-idf 値は，(3) の $tf(w_i, d_j)$ ，(2) の $idf(w_i)$ を利用し，以下のように定義する．

$$tfidf(w_i, d_j) = tf(w_i, d_j) \cdot idf(w_i) \quad (4)$$

例として，先述の【ニュース 1】のメンションにおける，tf-idf 値の高い単語の例を，表 2 に示す．

表 2 【ニュース 1】のメンションにおける tfidf 値が高い語の例

単語	tfidf 値
秋田	0.64930
配備	0.30485
イージス	0.22834
議員	0.22033
防衛	0.19659

4.1.3 メンションにおいて，tfidf 法を改良した方法

ニュース本文においてあまり出現しないが，メンションにおいて多く出現する語は，そのニュースのコンテキストとして重要なキーワード，注目されている視点であると考えられる．そこで，メンションにおいて，メンションの tf とニュース本文の tf の差を利用することを検討する．本稿ではこの重みを tfdbd-idf(Term Frequency Difference between Documents - Inverse Document Frequency) とし，メンション文書 m_a における w_i の tfdbd-idf 値を，ニュース本文 n_a を利用して以下のように定義する．

$$tfdbdidf(w_i, m_a, n_a) = (tf(w_i, m_a) - tf(w_i, n_a))idf(w_i) \quad (5)$$

この手法を用いる際，ニュース本文の重みは tf-idf を利用する．

例として，先述の【ニュース 1】のメンションにおける，tfdbd-idf 値の高い単語の例を，表 3 に示す．

表 3 【ニュース 1】における tfdbd-idf 値が高い語の例

単語	tfdbd-idf 値
議員	0.11760
国会	0.06591
県民	0.06129
国防	0.05810
役目	0.05565

4.2 ニュースコンテキスト特徴量の算出

上記のような重みを利用して，ニュース本文の特徴ベクトルとメンション特徴のベクトルを求める．単語ベクトルは Word2Vec [15] を利用する．

Word2Vec は，ニューラルネットワークを利用して単語の分散表現を得る手法である．Word2Vec では，文章中のある単語とその周辺にある単語を入力し，その単語の次にどの単語が出現するかという確率を出力するモデルについて，学習を繰り返

すことで、中間層のパラメータを調整する。単語を入力した時の中間層の値がその単語の分散表現を表す特徴ベクトルとなる。

ニュース本文の特徴ベクトルとメンションの特徴ベクトルは、含まれる各単語の単語ベクトルの重み付き平均として算出する。例として、ニュース a のニュース本文 n_a の特徴ベクトル \mathbf{n}_a を、 n_a の BoW である W_{n_a} , W_{n_a} の中に出現する単語 w_i , その単語の Word2Vec ベクトルを \mathbf{w}_i として、以下のように定義する。

$$\mathbf{n}_a := \frac{\sum_{w_i \in W_{n_a}} \text{weight}(w_i, n_a) \mathbf{w}_i}{\|\sum_{w_i \in W_{n_a}} \text{weight}(w_i, n_a)\|} \quad (6)$$

重みとして idf を利用する場合には、重みの引数は w_i のみとなる。

また、ニュース a のメンション m_a の特徴ベクトル \mathbf{m}_a を、 m_a の BoW である W_{m_a} , W_{m_a} の中に出現する単語 w_i , その単語の Word2Vec ベクトルを \mathbf{w}_i として、以下のように定義する。

$$\mathbf{m}_a := \frac{\sum_{w_i \in W_{m_a}} \text{weight}(w_i, m_a) \mathbf{w}_i}{\|\sum_{w_i \in W_{m_a}} \text{weight}(w_i, m_a)\|} \quad (7)$$

こちらも重みとして idf を利用する場合には、重みの引数は w_i のみとなる。また、重みとして tfdbd-idf を利用する際は、重みの引数として w_i, m_a, n_a の 3 つが必要となる。

5 ニュース間の関連度の算出

前章において、ニュース本文の特徴ベクトルとメンション特徴のベクトルを算出する方法を示した。本章においては、これらの特徴ベクトルを利用して、ニュース間の関連度を求める方法について述べる。

ニュース a について、ニュースコンテキストのベクトル \mathbf{N}_a を、ニュース本文の特徴ベクトル \mathbf{n}_a , メンションの特徴ベクトル \mathbf{m}_a を利用して、以下のように定義する。

$$\mathbf{N}_a = \alpha \mathbf{m}_a + (1 - \alpha) \mathbf{n}_a \quad (8)$$

上記の式において、 α はニュース本文とメンションの特徴ベクトルの重要度の割合を決めるパラメータである。

また、ニュース間の関連度は特徴ベクトルの類似度として求める。類似度にはコサイン類似度を利用する。例えば、ニュース a とニュース b の関連度は以下ようになる。

$$\cos(\mathbf{N}_a, \mathbf{N}_b) = \frac{\mathbf{N}_a \cdot \mathbf{N}_b}{\|\mathbf{N}_a\| \|\mathbf{N}_b\|} = \frac{\sum_{i=1}^{|V|} N_{a,i} N_{b,i}}{\sqrt{\sum_{i=1}^{|V|} N_{a,i}^2} \sqrt{\sum_{i=1}^{|V|} N_{b,i}^2}} \quad (9)$$

上記の式において、 V は次元数である。

以上の手法を用いて、元のニュースとニュースコンテキストが関連するニュースを抽出する。評価実験を行い、適切にニュースコンテキストが関連するニュースを抽出できるようなパラメータを決定する。

6 実 験

本章では、Web から収集したデータを利用した評価実験によって、提案手法の有用性を評価する。

6.1 データの準備

評価実験のために利用したデータセットについて述べる。Yahoo!ニュース (@YahooNewsTopics) [16] で配信されているニュースを対象として、TwitterAPI を利用しニュースツイートを収集した。Yahoo!ニュースを対象とした理由として、新聞社、通信社など様々なメディアからニュースを提供され、それを配信しているため、多様な視点からのニュースが配信されるとともに、集められる反応にも主張の偏りが無いと考えたためである。

次に、Yahoo!ニュースのニュースツイートに対して行われたリプライ、引用リツイートを集めた。本稿ではこれらを反応ツイートと呼ぶ。

さらに、ニュース記事の本文は、Tweet につけられている URL を元にスクレイピングにより収集した。本研究では、反応を利用してニュースの特徴をはかるため、そのニュースに対する反応がある程度必要である。そこで今回は、リプライと引用リツイートが 50 回以上なされているニュースを対象とした。

データ収集は 2019 年 6 月 10 日から同年 8 月 30 日にかけて行い、上記の条件を満たす記事 1677 件とそれらのメンションをデータセットとして構築した。

6.2 前 処 理

形態素解析の前に、テキスト本文には以下の前処理を行った。

- URL など不要な文字列を消す
- 全角英数記号を半角に直す

またメンションには以下の前処理を行った。

- URL, スクリーンネームなど不要な文字列を消す
- 絵文字などを消す
- 全角英数記号を半角に直す

テキスト本文、メンション共に単語ごとに分割するため形態素解析を行った。形態素解析には、オープンソースの日本語形態素解析エンジンである MeCab [17] を利用した。

6.3 ニュース本文、メンションのベクトル化

単語のベクトル化を行うための Word2Vec モデルを準備した。Yoneda ら [7] は Word2Vec モデルを学習する際のコーパスとして、膨大なニュース記事を利用している。しかし本研究では、十分な量のニュースコーパスが用意できなかった上、メンションなどにニュースコーパス内に現れない重要語が含まれることが考えられたことから、日本語版 Wikipedia 全記事 (2019 年 7 月 19 日現在) をコーパスとして Word2Vec モデルの学習を行った。ベクトルの次元は 300、窓サイズは 15 で、出現回数が 20 回未満の単語を除いて学習した。

ニュース本文、メンションをベクトル化する際には、文書中の名詞のみを利用した。

6.4 ニュース間の関連度の算出

【ニュース 1】について、idf 法で重み付けしたベクトルで選択された関連ニュースを表 4, tfidf 法で重み付けしたベクトルで選択された関連ニュースを表 5, tfdbdidf 法で重み付けしたべ

質問1

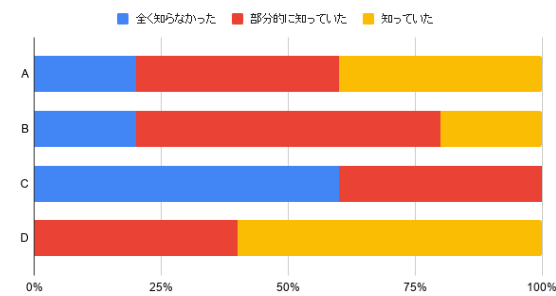


図5 質問1の回答結果

クトルで選択された関連ニュースを表6に示す。

関連ニュースとして抽出したニュースの多くが、【ニュース1】と同じ防衛、外交関連のニュースとなっている。

6.5 評価実験

本研究の手法において、ユーザがニュースコンテキストを知るために有効なパラメータを求めるため、評価対象のニュース(「対象ニュース」と呼ぶ)と、抽出した関連ニュース10件を被験者に提示し、評価実験を行なった。被験者に対して、以下の質問に回答してもらった。

- i (対象ニュース、関連ニュースについて、) そのニュースを知っていたか
- ii (対象ニュースと関連ニュースを比較して、) 同じ話題に関するニュースだと思うか
- iii (対象ニュースと関連ニュースを比較して、) 背景や原因に関連性があると思うか

質問iは3段階で、質問ii、質問iiiは5段階で回答してもらった。実験に使用したニュースを表7に示す。今回の評価実験は、九州大学大学院芸術工学府、および九州大学芸術工学部の学生男女5人に回答してもらった。また、関連ニュースは、重み付け3手法、関連度4手法の合計12手法で算出された12種の類似度の合計に基づき、関連が高いと考えられるものを各10件を選び、それを評価してもらった。

6.6 結果

各対象ニュースに対する、質問iの回答の結果を図5に示す。

各ニュースの質問iiの評価値と質問iiiの評価値の間の相関係数を表8に示す。

ニュースコンテキストベクトルをidf手法で重み付けし、パラメータを変化させた際の、各対象ニュースと関連ニュース間の類似度と、質問iiの評価値との相関係数を表9に、質問iiiの評価値との相関係数を表10に示す。

ニュースAとニュースCについて、パラメータ、つまりニュースコンテキストベクトルでのメンションベクトルの割合が高くなるにつれて、idfでのニュース間の類似度と質問iiの評価値間の相関、およびidfでのニュース間の類似度と質問iiiの評価値間の相関において、より強い相関を示している。これは、ニュースコンテキストを推定する際に、ニュース本文だけ

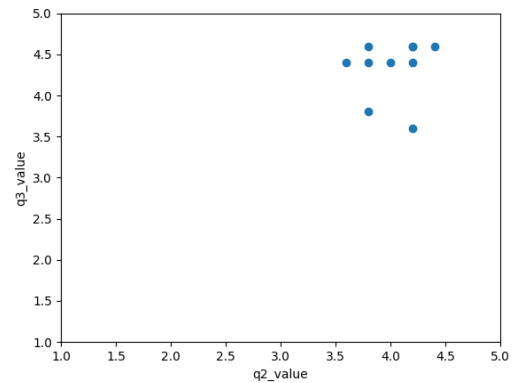


図6 ニュースDの質問iiの評価値と質問iiiの評価値の散布図

でなく、メンションを利用することが有効である可能性を示していると考えられる。ただし、これら2つの対象ニュースはそれぞれ、質問iiの評価値と質問iiiの評価値の間の相関が強く出ており、元のニュースと直接的な関連は少ないが、背景が関連しているニュースを選ぶことまではできていない。

図6にニュースDの質問iiと質問iiiの評価値の散布図を示す。ニュースDにおいては、ニュース間のベクトルの類似度と、質問iiiの評価値との間に負の相関を示しているが、図6のように関連ニュースとして抽出したニュースが直接的、背景的に強く関連するものに偏っており、様々な関連度合を考慮した十分な評価を行うことができていないと言える。

また、tfidf法、tfdbd-idf法でベクトル間類似度との相関を求めた場合、idf法と比べ、ニュースコンテキストベクトルのパラメータ α を変化させていったときの相関に大きな変化が見られなかった。

6.7 考察

本研究で抽出を行いたいと考えている、対象ニュースと直接的にはあまり関連していないものの、ニュースコンテキストが強く関連するニュースの例として、以下の【ニュースB】に対する【ニュースB'-5】があげられる。

【ニュースB】 2019/8/27

【年金財政検証30年後に2割減】厚労省は27日、年金の長期見通しを試算する財政検証結果を公表。標準的なケースで約30年後に年金は実質的に2割近く目減りするが、給付水準は50%を維持し、経済成長が見込めれば制度は持続可能とした。

【ニュースB'-5】 2019/6/18

【働く世代の負担 日本が世界一】国連は、65歳以上の人口に対する25～64歳の人口比率を示す「潜在扶養率」が、2019年には日本が世界最低の1.8を記録したとの統計を発表。少子高…

ニュースBに対するニュースB'-5の評価値は、質問iiが2.6、

表4 【ニュース1】のidf法での関連ニュース

対象ニュースのベクトル	関連ニュースのベクトル	関連ニュース	関連度	日付
$\alpha = 0$ (ニュース)	$\alpha = 0$ (ニュース)	【イージス必要 秋田で首相演説】	0.8118	2019/7/13
		【自民 金融庁報告書の撤回要求】	0.7849	2019/6/11
		【N 国 丸山穂高氏に入党要請】	0.7847	2019/7/25
$\alpha = 0$ (ニュース)	$\alpha = 1$ (メンション)	【イージス必要 秋田で首相演説】	0.7690	2019/7/13
		【首相 9 月に内閣改造を明言】	0.7609	2019/8/26
		【首相 内閣改造「大幅」で検討】	0.7559	2019/8/17
$\alpha = 1$ (メンション)	$\alpha = 0$ (ニュース)	【日韓関係悪化 歴史的転換点か】	0.8323	2019/7/17
		【イージス必要 秋田で首相演説】	0.8295	2019/7/13
		【自分に遺憾?首相発言に違和感】	0.8123	2019/7/4
$\alpha = 1$ (メンション)	$\alpha = 1$ (メンション)	【イージス必要 秋田で首相演説】	0.9398	2019/7/13
		【破棄 ミサイル情報の質低下か】	0.8989	2019/8/23
		【安倍氏 改憲議論が国民の審判】	0.8956	2019/7/22

表5 【ニュース1】のtfidf法での関連ニュース

対象ニュースのベクトル	関連ニュースのベクトル	関連ニュース	関連度	日付
$\alpha = 0$ (ニュース)	$\alpha = 0$ (ニュース)	【イージス必要 秋田で首相演説】	0.7952	2019/7/13
		【政府 竹島訓練「中止求める」】	0.6818	2019/8/25
		【海外視察に 1850 万円 熊本市】	0.6605	2019/8/29
$\alpha = 0$ (ニュース)	$\alpha = 1$ (メンション)	【イージス必要 秋田で首相演説】	0.7711	2019/7/13
		【離島の警備 警察が専門部隊】	0.6950	2019/8/29
		【防衛相 有志連合に派遣考えず ³ 】	0.6861	2019/7/16
$\alpha = 1$ (メンション)	$\alpha = 0$ (ニュース)	【イージス必要 秋田で首相演説】	0.8546	2019/7/13
		【政府 竹島訓練「中止求める」】	0.7222	2019/8/25
		【首相 輸出規制の正当性を主張】	0.7120	2019/7/7
$\alpha = 1$ (メンション)	$\alpha = 1$ (メンション)	【イージス必要 秋田で首相演説】	0.9077	2019/7/13
		【防衛相 有志連合に派遣考えず ³ 】	0.7862	2019/7/16
		【離島の警備 警察が専門部隊】	0.7695	2019/8/29

表6 【ニュース1】のtfbdfidf法での関連ニュース

対象ニュースのベクトル	関連ニュースのベクトル	関連ニュース	関連度	日付
$\alpha = 0$ (ニュース)	$\alpha = 0$ (ニュース)	【イージス必要 秋田で首相演説】	0.7952	2019/7/13
		【政府 竹島訓練「中止求める」】	0.6818	2019/8/25
		【海外視察に 1850 万円 熊本市】	0.6605	2019/8/29
$\alpha = 0$ (ニュース)	$\alpha = 1$ (メンション)	【イージス必要 秋田で首相演説】	0.7332	2019/7/13
		【離島の警備 警察が専門部隊】	0.6993	2019/8/29
		【政府 竹島訓練「中止求める」】	0.6878	2019/8/25
$\alpha = 1$ (メンション)	$\alpha = 0$ (ニュース)	【イージス必要 秋田で首相演説】	0.8464	2019/7/13
		【首相 輸出規制の正当性を主張】	0.7674	2019/7/7
		【選挙を頑張れば予算 発言波紋】	0.7546	2019/7/1
$\alpha = 1$ (メンション)	$\alpha = 1$ (メンション)	【イージス必要 秋田で首相演説】	0.8841	2019/7/13
		【選挙を頑張れば予算 発言波紋】	0.8094	2019/7/1
		【政府 竹島訓練「中止求める」】	0.8034	2019/8/25

表7 評価する対象ニュース

対象ニュース	ニュースのタイトル	配信日
ニュース A	【内定辞退予測購入 20 社超公表】	2019/8/25
ニュース B	【年金財政検証 30 年後に 2 割減】	2019/8/27
ニュース C	【イージス 秋田の自民議員反対】	2019/8/24
ニュース D	【韓国首相 関係修復へ対話を】	2019/8/28

表8 質問 ii の評価値と質問 iii の評価値の間の相関係数

対象ニュース	質問 ii 評価値平均	質問 iii 評価値平均	相関係数
A	2.74	3.04	0.862
B	3.12	3.68	0.667
C	2.84	2.98	0.961
D	4.02	4.34	0.112

質問 iii が 4.6 であった。一方、このニュースのような、直接的関連は低いものの、ニュースコンテキスト関連が高いニュース

は、今回の手法では十分に抽出できなかった。

表 9 idf での類似度と質問 2 の評価値間の相関係数

パラメータ α	ニュース A	ニュース B	ニュース C	ニュース D
0.0	-0.080	0.181	0.117	0.074
0.25	0.200	0.229	0.245	0.102
0.5	0.747	0.295	0.514	0.217
0.75	0.823	0.351	0.819	0.401
1.0	0.804	0.429	0.846	0.318

表 10 idf での類似度と質問 3 の評価値間の相関係数

パラメータ α	ニュース A	ニュース B	ニュース C	ニュース D
0.0	-0.150	0.412	0.154	-0.534
0.25	0.031	0.489	0.279	-0.517
0.5	0.506	0.556	0.529	-0.537
0.75	0.642	0.526	0.782	-0.490
1.0	0.647	0.548	0.770	-0.414

7 終わりに

本研究では、ユーザのニュースコンテキストの理解を支援するために、ニュース本文とメンションからニュース間の関連づけを行う手法を提案した。結果として、元のニュースに対し、内容として関連するニュースを抽出することができた。一方で、直接的ではないものの、ニュースコンテキストが関連し、ユーザの知識を増強したり、ニュース理解を支援するような、特徴的なニュースを十分に抽出できたとは言えない。原因として、実験用のニュースを収集する期間が短く、十分な量のニュースで実験を行えていないことや、提案手法を利用して抽出したニュース本文やメンションのベクトルの類似度が高い関連ニュースに絞って評価実験を行なったことが考えられる。

今後の課題として、評価するニュースの幅を広げるため、様々な検証を行なっていきたいと考えている。具体的には、ニュース本文ベクトルの類似度とメンションベクトルの類似度を元に、比較ニュースをマッピングし、ニュースコンテキスト関連が深いニュースが持つ類似度の特徴を調査する。また、Word2Vec モデル作成の際のコーパスを Wikipedia のみでなく、実際のニュース記事を利用したり、実験に利用するニュースを増やすなどの工夫を行う予定である。

文 献

- [1] Twitter, <https://twitter.com/>
- [2] “【2019 年 11 月更新】人気 SNS の国内&世界のユーザー数まとめ (Facebook, Twitter, Instagram, LINE)”, *comnico*, 2019 年 11 月 7 日更新, <https://blog.comnico.jp/we-love-social/sns-users>, 2019 年 12 月 31 日閲覧
- [3] 朝日新聞デジタル, “秋田市に陸上イージス「もうダメ」 地元自民議員が反対”, 2019 年 8 月 24 日更新, <https://www.asahi.com/articles/ASM8S6RGWM8SUTFK00R.html>, 2019 年 12 月 26 日閲覧
- [4] 藤江俊彦, “はじめてのマスコミ論”, 同友館, 2006
- [5] Lihong Li, Wei Chu, John Langford, Robert E. Schapire, “A Contextual-Bandit Approach to Personalized News Article Recommendation”, Presented at the Nineteenth International Conference on World Wide Web (WWW 2010), Raleigh, NC, USA, 2010
- [6] Shumpei Okura, Yukihiro Tagami, Shingo Ono, Akira Tajima, “Embedding-based News Recommendation for Millions of Users”, *KDD’ 17*, August 13 - 17, 2017
- [7] Takeshi Yoneda, Shunsuke Kozawa, Keisuke Osone, Yukinori Koide, Yusuke Abe, Yoshifumi Seki, “Algorithms and System Architecture for Immediate Personalized News Recommendations”, *WI ’ 19*, October 14 - 17, 2019, Thessaloniki, Greece
- [8] Earl Rennison, “Galaxy of News An Approach to Visualizing and Understanding Expansive News Landscapes”, In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pp. 3 - 12. ACM, 1994.
- [9] Akiyo Nadamoto, Katsumi Tanaka, “Time-based Contextualized-News Browser (T-CNB)”, In *Proceedings of the 13th international World Wide Web conference on Alternate track papers posters*, pp. 458 - 459. ACM, 2004.
- [10] Jianhui Liu, Larry Birnbaum, “LocalSavvy : Aggregating Local Points of View about News Issues”, *Proc. 1st International Workshop on Location and the Web*, 2008, 33-40, 2008
- [11] 田中祥太郎, ヤフトアダム, 田中 克己, “ニュース記事の理解のための背景・前提事象の抽出と分析”, *DEIM Forum 2015 B4-2*, 2015
- [12] 片岡 雅裕, 橋山 智訓, 田野 俊一, “情報推薦システムにおいて閲覧する情報の偏りを気付かせる UI の設計”, *31st Fuzzy System Symposium*, 2015.9
- [13] 切通 恵介, 楠見 孝, 堀江伸太郎, 馬 強, “多様性指向のニュースアプリの開発とその有用性評価”, *DEIM Forum 2016*, 2016.3
- [14] Karen Spärck Jones, “A statistical interpretation of term specificity and its application in retrieval”, *Journal of Documentation*, 28, 11-21, 1972
- [15] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space”, *Proceedings of the International Conference on Learning Representations*, 2013
- [16] Yahoo! ニュース, <https://news.yahoo.co.jp/>
- [17] 形態素解析エンジン MeCab, <http://taku910.github.io/mecab/>