

Twitter のメンションを利用したニュースのコンテキスト推定と理解支援

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: [†]ikededa.sho.294@s.kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

あらまし SNS を利用してニュースを読む際には、ニュースに対するメンションを読むことで、様々なユーザのニュースへの意見を知り、ニュースの全体像を理解できる利点がある。しかし、メンションの数が多い場合には、大量のメンションすべてを確認することは困難であるために、全体像を得ることが難しい場合がある。また、SNS に不慣れなユーザにとっては、SNS で取得可能な有益な情報の取得が困難である。そこで、我々は SNS 上のメンションを利用して、ニュースの社会的位置付けや世論（コンテキスト）を自動的に抽出し、ユーザにわかりやすく提示する手法を開発することを目指す。提案手法では、SNS のメンションを利用してコンテキストとして重要なキーワードを抽出し、それに基づいてニュースに対する意見を選出する。キーワードと意見をグラフ化してユーザに提示し、ニュースの理解を支援する。

キーワード SNS, ニュース, コンテキスト推定, テキスト処理

1 はじめに

現在、インターネット上では、多くのニュース記事が配信され、ニュースサイトやアプリケーションを通じてユーザに配信されている。総務省の令和元年通信利用動向調査 [1] によると、令和元年の「情報検索（天気予報、ニュースサイト、地図・交通情報などの利用）」の利用者は、回答者 32,607 人のうち、64.2% に及ぶ。特に、20 代から 50 代では 75% 以上が利用しており、60 代でも 64.5% が利用している。また、NHK 放送文化研究所の国民生活時間調査 [2] によると、国民全体の平日のテレビ、ラジオ、新聞を含むマスメディア接触の平均時間が 2010 年の 4 時間 28 分から 2015 年には 4 時間 19 分に減少しており、ニュース閲覧の形態が変わってきていることがわかる。

インターネットニュースを利用するユーザは、自身の興味があるニュースを読む。例えば、日本における代表的なポータルサイトの一つである Yahoo! ニュース¹ では、トップページのトピックスや、スマートフォンのアプリケーションの通知によって、ユーザは興味のあるニュースを見つけ、読むことができる。しかし、同じニュースを読んだとしても、ユーザによってニュース理解の程度や、ニュースの捉え方が異なる。そのため、ニュースの理解を広げるものとして、ニュースに対する他者の意見は有益である。

近年では、情報を伝達するツールとして、SNS（ソーシャル・ネットワークワーキング・サービス）が普及している。SNS 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。令和元年では、回答者 32,607 人のうち、58.6% が SNS を利用しており、特に 13~39 歳の若年層は 75% 以上が利用している [1]。代表的な SNS として Twitter² が

あげられる。Twitter 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 の特徴として拡散性の高さがあり、他のユーザが投稿した投稿をリツイートすることによって、自分をフォローしているユーザにそのツイートを配信できる。

Twitter の普及に伴い、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、Twitter 上でニュースをニュースツイートとして配信している。例えば Yahoo! ニュースの公式アカウント (@YahooNewsTopics)³ では、Yahoo! ニュースで配信されたニュース記事を見出しや概要、その記事の URL を Twitter 上に投稿している。Twitter 上のニュースツイートは多くのユーザがリツイートすることによって拡散される。Twitter ユーザは、ニュースツイートに対し、リプライや引用リツイートで自分の意見や追加の情報を投稿できる。ユーザは、ニュースツイートに対する他者のリプライや引用リツイートを読むことにより、ニュース自体を深く理解し、そのニュースに対する世論を理解できる。実際、消費者庁の平成 28 年度の調査 [3] によると、調査対象 520 人のうち 38.8% が、ニュース等に関する情報の収集を目的に SNS を利用していると報告されている。

一方、SNS を利用してニュース閲覧を行う際の問題点として、リプライや引用リツイートといったメンションが大量に投稿されている場合、それらを全て閲覧することが困難であるため、SNS からニュースの全体像を把握することが困難な場合がある。また、SNS に不慣れなユーザにとっては、SNS で適切な操作を行うことができず、ニュースに関する関連する情報や世論の動向を理解することが困難となる場合が多い。さらには、SNS では自分と似た興味関心をもつユーザをフォローする結

1 : <https://news.yahoo.co.jp/>

2 : <https://twitter.com/>

3 : <https://twitter.com/YahooNewsTopics>

コンテキストグラフ

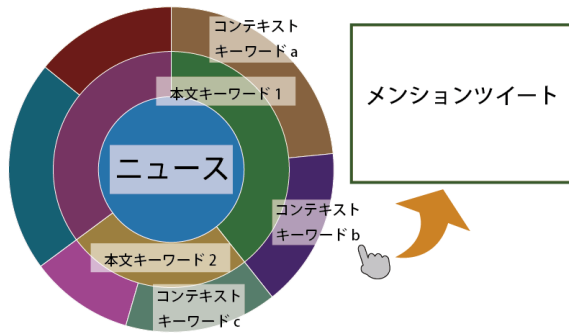


図1 コンテキストグラフの概念図

果、自分の意見と近い意見ばかりが返ってきてしまう「エコーチェンバー」と呼ばれる現象があり [4], ニュースを大局的かつ多様な視点から捉えることができずに、適切に世論を理解できない場合がある。

そこで本研究では、SNS 上のメンションを利用して「ニュースコンテキスト」を自動的に抽出し、ユーザにわかりやすく提示することを目指す。本論文では、ニュースの社会的位置付けや世の中の意見をニュースコンテキストと呼ぶ。ニュースコンテキストは、ユーザがニュースの全体像や世論をより深く理解するのに役立つと考えられる。SNS のメンションを利用してコンテキストとして重要なキーワードを抽出し、それに基づいてニュースに対する意見を選出する。また、ニュースやそのコンテキストを簡単にかつ俯瞰的に捉えるために、キーワードと意見をグラフ化した「コンテキストグラフ」をユーザに提示し、ニュースの理解支援を行う。

従来の Twitter では、ニュースを配信するアカウントがニュースの見出しとニュース記事の URL が含まれたツイートを投稿し、それに対して多くのユーザがリプライや引用リツイートなどで意見を表明している。従来の Twitter を利用したニュース閲覧では、ユーザはニュースツイートからリンクしてニュース本文を読んだり、ニュースツイートに紐付けられたメンションを読むことで、ニュース理解を行なっている。しかし、このような閲覧手法では、ページ遷移が多く操作が煩雑になる上、メンションが大量にある場合にはそれら全てを読むことは困難であり、世論などコンテキストまで含めたニュースの全体像が掴みにくい。「コンテキストグラフ」の概念図を図1に示す。「コンテキストグラフ」では、一番内側の円にニュースタイトル、2番目の円にニュース本文で重要な「本文キーワード」、3番目の円にコンテキストとして重要な「コンテキストキーワード」を配置する。「本文キーワード」はニュースの内容を表し、「コンテキストキーワード」はニュースに対する意見や世論を表す。ニュース本文、本文キーワード、コンテキストキーワードを一覧することで、ニュースの内容以外に、そのニュースの主な論点についても全体像を把握することができる。図は円グラフを重ねたような形になっており、キーワードの重要度を直感的にわかりやすく把握できる。

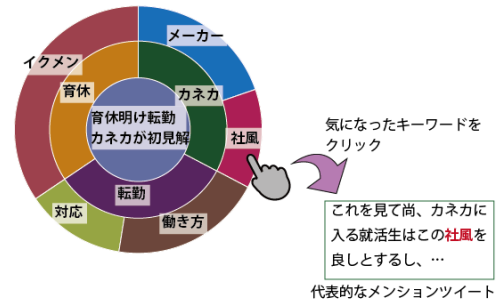
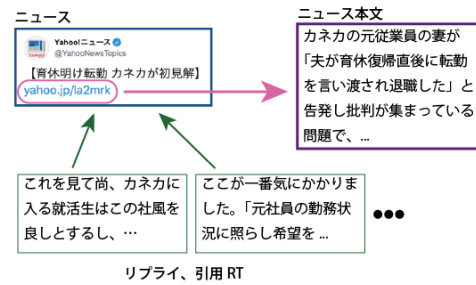


図2 コンテキストグラフの具体例

例として、2019 年 6 月 6 日に Yahoo!ニュースで配信された【育休明け転職 カネカが初見解】という記事についてのコンテキストグラフの具体例を図2に示す。この記事は、育休を取得した男性が復帰直後に転職を命じられたことを男性の妻が告発したことに関し、男性が勤めていた企業が反論をしたという記事である。コンテキストグラフでは、この記事の本文の中の重要語となる「カネカ」、「転職」、「育休」といったキーワードが本文キーワードとして配置され、「社風」、「働き方」などメンションの中で重要となった語がコンテキストキーワードとして配置される。

本稿では、コンテキストグラフ生成の手法について述べ、有効性を評価するための実験についてその結果を示し、考察する。第2章では本研究と関連する研究について、第3章では提案手法について、第4章では実験について、第5章ではまとめと今後の課題についてそれぞれ述べる。

2 関連研究

本研究では、ユーザの深いニュース理解を支援するために、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、その提示を行う手法の開発を目的とする。そこで本章では、Web ニュースでの理解支援に関する関連研究と、SNS を利用したニュース閲覧支援に関する関連研究について述べる。

2.1 Web ニュースでの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、多くの研究が行われている。Liu ら [7] 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 [8] は、ニュース記事について、そのニュースの中心的事象について述べた主題事象についての記述と、その出来事の時系列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。Saravanou ら [15] は、ランク付け学習を用いてあるニュースの重要なニュースストーリーを識別する手法を提案している。

また、ニュース理解のための新たな可視化手法やアプリケーションに関する研究として、次のようなものがある。Rennison ら [5] 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 [6] 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、ユーザがニュースの進捗状況を理解できる Time-based Contextualized-News Browser (T-CNB) を提案している。片岡ら [9] 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・インタフェースを提案している。切通ら [11] 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。

本研究はこれらの研究と異なり、SNS から抽出した意見、コンテキストをユーザに俯瞰視させることで、ユーザのニュース理解を促すことを目的としている。

2.2 SNS を利用したニュース閲覧支援に関する研究

SNS におけるニュース消費を分析し、ユーザのニュース理解や議論を促す研究に、次のようなものがある。Becatti ら [14] は選挙期間中のニュース消費において、異なる意見をもつコミュニティに対しても影響があるバイラルなコンテンツを特定するため、2018 年のイタリアの総選挙期間中のツイートを分析している。この研究では、認証アカウントと非認証アカウントの2層の無向リツイートネットワークを利用してコミュニティを推定し、ユーザとツイートの2層の有向ネットワークを用いてバイラルコンテンツを推定している。Barbera ら [10] は、実際の Twitter のデータを利用し、ニュースイベントでのエコーチェンバーの状態や推移を分析している。

また、SNS において、ユーザにニュースに関する情報を提供し、理解を促す研究として以下のようなものがある。Phuvipadawat ら [12] は、Twitter における速報 (Breaking News) を収集し、それらをグループ化、ランク付け、追跡を行う手法を提案してい

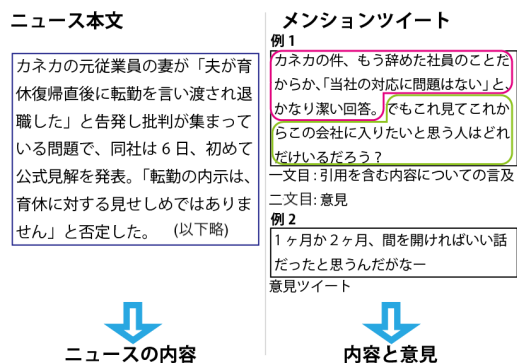


図3 ニュース本文とメンションツイートの分析

る。また、その手法を Hotstream というアプリケーションとして実装している。Mizuka ら [13] は、ニュース記事に対するコメントのうち信憑性が不明瞭なものに対して警告をつけるための第一歩として、ルールベースの手法と SVM を組み合わせて、ニュース記事に対するコメントツイートを抽出する手法を提案している。Vo ら [16] は、SNS でフェイクニュースが氾濫している問題の解決のために、ガーディアンと呼ばれる SNS でのファクトチェック活動を行い、それらを拡散するユーザに対して、刺激を与えることで正しい情報が拡散されるようにするファクトチェック URL 推薦モデルを提案している。

本研究では、ニュースの持つ世論や意見といったコンテキストを提供するために SNS を用いており、これらの研究とは異なる。

3 提案手法

本研究では、Twitter の本文とメンションから、ニュースの内容を示すものとして本文キーワードを、ニュースのコンテキストを示すものとしてコンテキストキーワードを抽出する。本文キーワードはニュース本文の重要語として、コンテキストキーワードはメンションの重要語からニュース本文の重要語を比較して抽出する。この理由を 2019 年 6 月 6 日に Yahoo! ニュースで配信された【育休明け転勤 カネカが初見解】という記事を例に示す。図3にニュース本文とメンションツイートの分析を示す。ニュース本文は、出来事について事実を記述しているため、そこから抽出されるキーワードもニュースの内容を示すものと考えられる。メンションツイートは、例1のようにニュース本文への言及とユーザ自身の意見を両方含むものや、例2のように意見のみが述べられたものがあり、そこから抽出されるキーワードは、内容を示すものと意見を示すものの両方が含まれると考えられる。このため、メンションキーワードの単語重みとニュース本文キーワードの単語重みを比較することで、コンテキストキーワードを判別できると考えられる。

本研究の流れを図4に示す。本研究ではまず、ニュース本文キーワードとメンションキーワードをそれぞれ抽出する処理を行う。抽出されたニュース本文キーワードとメンションキーワードを比較し、重要度を算出することで、コンテキストキーワードを抽出する。

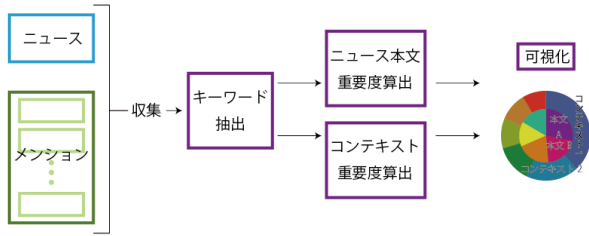


図4 本研究の流れ

3.1 ニュース本文キーワードの抽出

本研究では、キーワードを抽出する手法として、tf-idf 法 [17] を利用する。tf-idf 法は文書中の単語の重要度を計算する手法である。tf-idf 法での重みは、tf 値と idf 値の積である。tf 値はある単語がある文書内に現れる頻度であり、idf 値はある単語が現れる文書数の逆数である。これにより、文書内で重要で、かつ、その文書で特徴的である単語を重要視する。

本論文では、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。

$$tf(w_i, d_j) = \frac{f(w_i, d_j)}{\sum_{w_k \in d_j} f(w_k, d_j)} \quad (1)$$

また、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。

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

tfidf 値は tf 値と idf 値の積なので、以下のように定義される。

$$tfidf(w_i, d_j) = tf(w_i, d_j) \cdot idf(w_i) \quad (3)$$

重みの計算は非自立語や接尾詞を除いた名詞の単語を対象に行う。

3.2 メンションキーワードの抽出

メンションキーワードはニュースに対するメンションの重要語である。メンションキーワードを抽出する手法としては、tf-idf 法を改良した手法を用いる。まず、tf にあたる部分は、対象とするニュースに対するメンションのうち、目的の単語が使われたメンションの割合とする。理由としては、ある特定のユーザが同じ単語を何度も使うことで、本来重要でない単語が重視されてしまうことを防ぐためである。次に、idf にあたる部分は、Wikipedia の idf を用いる。この理由として、ある特定のニュースが大きく盛り上がり、そのニュースに関する報道や、そのニュースに対するメンションが増えた場合、本来一般的でない単語が一般的なものとされてしまうことを防ぐためである。

あるニュース n_j のメンションに出現する単語 w_i の重み $weight(w_i, n_j)$ は、メンションを m 、ニュース n_j に対する全メンション集合 M_{n_j} 、単語 w_i を含むメンションを $m(w_i)$ 、単語 w_i の Wikipedia での idf を $wiki_idf(w_i)$ とすると、次のように定義される。

$$weight(w_i, n_j) = \frac{|m|m = m(w_i), m \in M_{n_j}|}{|M_{n_j}|} \cdot wiki_idf(w_i) \quad (4)$$

重みの計算は非自立語や接尾詞を除いた名詞の単語を対象に行う。

3.3 コンテキストキーワードの抽出

コンテキストキーワードは、ニュース本文キーワードとメンションキーワードを使って抽出する。コンテキストキーワードの重みは、メンションキーワードとしての重みからニュース本文キーワードとしての重みを引くことで計算する。

まず、それぞれのキーワードの重みの分散を等しくするため、ニュース本文キーワードの重みとメンションキーワードの重みをそれぞれ標準化する。標準化後の重み集合 W_{std} を、標準化前の重み集合 W 、 W の標本平均 \bar{W} 、 W の標本標準偏差 s_W から以下のように求める。

$$W_{std} = \frac{W - \bar{W}}{s_W} \quad (5)$$

また、標準化された重みを 0~1 の間に収める処理を行う。処理後の重み W' は、 W_{std} の最大値 $W_{std,max}$ と W_{std} の最小値 $W_{std,min}$ から、以下のように求められる。

$$W' = \frac{W_{std}}{W_{std,max} - W_{std,min}} - \frac{W_{std,min}}{W_{std,max} - W_{std,min}} \quad (6)$$

(5),(6)の処理を行い更新されたニュース n_j での単語 w_i のニュース本文重みを $nw(w_i, n_j)$ 、メンション重みを $mw(w_i, n_j)$ とする。

ニュース n_j での単語 w_i のコンテキスト重み $cw(w_i, n_j)$ を、以下のように求める。

$$cw(w_i, n_j) = mw(w_i, n_j) - nw(w_i, n_j) \quad (7)$$

$cw(w_i, n_j)$ の高いものをコンテキストキーワードとして抽出する。

3.4 可視化

抽出した本文キーワード、コンテキストキーワードを利用してコンテキストグラフを生成する。一番内側の円にニュース本文、内側から 2 番目の円にニュース本文キーワード、一番外側の円にコンテキストキーワードを配置する。円グラフの各単語の面積は、その単語の重みによって決められる。各キーワードをクリックするとそのキーワードが含まれた代表的なメンションツイートが表示される。

4 実験

本章では、実際に収集したデータを利用して、提案手法を実装し、有用性を評価した。本章では、評価のために行った実験について述べる。

4.1 データの準備

実験に利用するニュースは Yahoo!ニュース (@YahooNewsTopics) 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の収集を行った。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。また、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 30 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 2020 年 11 月 1 日から同年 11 月 15 日にかけて行い、上記の条件を満たす記事 660 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。

4.2 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 など不要な情報を消す
- 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情報を消す
- 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 MeCab⁴を利用した。

4.3 キーワード抽出結果

以下にキーワード抽出結果の例を示す。ニュース a⁵についてのニュース本文キーワードの抽出結果を表 1、メンションキーワードの抽出結果を表 2、コンテキストキーワード抽出結果を表 3 に示す。

【ニュース a】

【反対多数 大阪市は存続へ】大阪市を廃止して四つの特別区に再編する大阪都構想の是非を問う住民投票が 1 日行われ、約 1 万 7 千票の僅差（きんさ）で反対多数となった。前回 2015 年に続く否決だ。大阪維新の会代表の松井一郎市長は 23 年 4 月の任期満了で政界を引退すると表明した。当日有権者数は 220 万 5730 人、投票率は前回は 4・48 ポイント下回る 62・35%だった。… (以下略)
2020/11/1

表 1 ニュース a のニュース本文キーワード

キーワード	重み
特別区	1.0000
大阪市	0.7222
構想	0.6791
維新	0.5975
否決	0.5975
住民投票	0.5293

表 2 ニュース a のメンションキーワード

キーワード	重み
大阪	1.0000
大阪市	0.7993
維新	0.7434
反対	0.6124
構想	0.5598
大阪都構想	0.5316

表 3 ニュース a のコンテキストキーワード

キーワード	重み
大阪	0.9373
大阪市民	0.4717
大阪都構想	0.4498
反対	0.3864
松井	0.3722
イソジン	0.3666

4.4 キーワードの評価実験

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、データセットの中からランダムで選んだニュース 10 件について、メンションキーワード 30 語を被験者に提示し、評価実験を行なった。実験で使ったニュースの見出しとその配信日を以下の表 4 に示す。

表 4 ニュースの見出しとタイトル

ニュース番号	ニュース見出し	配信日
1	加州バイデン氏勝利確実 CNN	11/4
2	陣営活発も トランプ氏は沈黙	11/5
3	米大統領選 きょう勝者確定か	11/6
4	Zeebra 中林美和と離婚	11/2
5	全国の感染者 2 日連続で最多	11/13
6	イベント緩和、当面延期へ	11/12
7	反対多数 大阪市は存続へ	11/1
8	参院選広島 落選側も県議に金	11/10
9	9 割にワクチン効果 識者早計	11/10
10	トランプ氏 選挙結果否定せず	11/9

コンテキストキーワードはメンションキーワードに含まれるので、メンションキーワードとコンテキストキーワードの比較を行うことができる。九州大学に所属する学生男女 7 人に対して、以下の質問に 5 段階で回答してもらった。

(1) そのキーワードは、そのニュース自体の内容を表すものだと思いますか？

(2) そのキーワードは、そのニュースに対する世の中の人の意見を表すものだと思いますか？

(3) そのキーワードをみて、そのキーワードに関する意見(ツイート)を読みたいと思いますか？

4.4.1 結 果

被験者の評価値の平均と提案手法での推定値(重み)の相関を見ることで、提案手法の有効性を検証した。以下に評価実験の結果を示す。

4 : 形態素解析エンジン MeCab, <http://taku910.github.io/mecab/>

5 : <https://news.yahoo.co.jp/pickup/6375312>, ニュース本文は削除済み

質問1「そのキーワードは、そのニュース自体の内容を表すものだと思いますか？」のニュースごとのメンションキーワードとコンテキストキーワードの評価値との相関値を表5に示す。この質問の結果を見ると、メンションキーワードについて、多くのニュースで弱い相関を示している。一方で、コンテキストキーワードでは、ほとんどのニュースで相関を示していない。コンテキストキーワードでは、ニュース本文に出るキーワードの重みが低くなっているため、この結果は妥当であると考えられる。

表5 質問1の結果

ニュース番号	メンション	コンテキスト
1	0.4418	0.3166
2	0.4110	0.0412
3	0.6468	0.2836
4	0.5339	-0.0105
5	0.6071	0.2105
6	0.4959	0.4437
7	0.3908	-0.0322
8	0.3073	-0.1871
9	0.4763	-0.3071
10	0.1556	-0.1846

質問2「そのキーワードは、そのニュースに対する世の中の人の意見を表すものだと思いますか？」のニュースごとのメンションキーワードとコンテキストキーワードの評価値との相関値を表6に示す。この質問の結果を見ると、全てのニュースにおいてメンションでもコンテキストでも相関を示していない。この理由として、コンテキストキーワードにおいて、ニュース本文語の重みが低くなりきっていないことなどが考えられる。また、メンション数の少ないニュースでは、意味のない単語が上位に来ていることも理由として考えられる。

表6 質問2の結果

ニュース番号	メンション	コンテキスト
1	0.1759	0.0617
2	0.0739	-0.0089
3	0.1749	0.0829
4	0.0985	0.0331
5	-0.0248	0.0966
6	0.3308	0.3392
7	0.0068	0.1517
8	-0.1732	-0.2105
9	-0.0277	-0.1635
10	-0.2036	-0.3092

質問3「そのキーワードをみて、そのキーワードに関する意見(ツイート)を読みたいと思いますか？」のニュースごとのメンションキーワードとコンテキストキーワードの評価値との相関値を表7に示す。この質問でも、ほとんどのニュースで、相関が示せていない。

表7 質問3の結果

ニュース番号	メンション	コンテキスト
1	0.4181	0.3320
2	0.1852	-0.0627
3	0.1780	0.0286
4	0.0613	0.0899
5	0.1351	0.1671
6	0.3441	0.3424
7	-0.1003	0.1955
8	-0.0507	-0.2210
9	0.3336	-0.2650
10	-0.2093	-0.3137

4.4.2 考察

質問2の「そのキーワードは、そのニュースに対する世の中の人の意見を表すものだと思いますか？」について、メンションキーワードとコンテキストキーワードの間で差を示す事ができなかった理由として、まず、コンテキスト重みではニュース本文の中に登場する単語の重みを十分に下げる事ができなかったことが考えられる。キーワードとしてニュース本文に現れる単語の略称や同じ意味として理解される単語が存在しており、それらについての処理がうまくいかなかったという事が考えられる。質問1では、メンションキーワードについて評価値と重みの弱い相関を示しており、ニュース本文に現れるキーワードは内容を示しているため、コンテキストキーワードでは、メンションキーワードからニュース本文に現れる語を取り除いたものとするのが良いのではないかと考えられる。

キーワードのみでは意味を理解するのが難しい事も考えられる。特に対象となったニュースについての知識が少ないと、世論としてのキーワードなのか、ノイズなのか、判別がつかないのでは無いかと考えられる。キーワードとキーワードについて言及されたメンションツイートを同時に提示する事で、世論を理解できるか検証する必要がある。

4.5 コンテキストグラフの評価実験

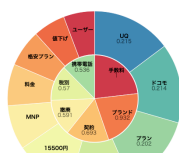
本研究では、ユーザがニュースのコンテキスト理解を支援するために、コンテキストグラフというニュース閲覧インタフェースを提案している。そこで本項では、実際にコンテキストグラフを実装し、被験者実験を行う。

本研究では、コンテキストグラフというニュース閲覧インタフェースをWebアプリケーションとして実装した。その具体例を図5に示す。

提案インタフェースは3つの要素から構成される。左からニュース本文ブロック、コンテキストグラフブロック、メンションブロックである。ニュース本文ブロックでは、そのニュースの本文が掲載されている。コンテキストグラフブロックでは、コンテキストグラフが掲載されている。内側の円グラフが本文キーワードであり、外側のドーナツグラフがコンテキストキーワードである。キーワード抽出実験を行なった結果、コンテキストキーワードにおいて、本文中の語が多く見られることがあったため、今回はメンションキーワードから本文中に表れた

ニュース本文

携帯電話大手のKDDI（au）は9日、主力ブランドと格安ブランドとの間で契約を乗り換える際の手数料を廃止し以降に適用すると発表した。ソフトバンクも一部の店舗での手続きを廃止して対応する。これは最大1万5500円（税別）が必要だった。ソフトバンクは業者からの予定で、具体的な時期は改めて公表する。 撤廃するのには、ブランド移行の手続きをした場合の契約更新手数料や、電話番号を変えに前の携帯電話会社へ乗り換える「番号ポータビリティ」（持ち運び）制度の手続き料、税金、同様に廃止や予定が報じられている。定額契約の途中で乗り換える場合の契約解除料も無料にする。



キーワード: ドコモ

YahooNewsTopics 結果、ドコモが撤廃か
YahooNewsTopics ドコモの発表にauもソフトバンクもついてこれないじゃないか こんなクソみたいな対応してるでやめて欲しいな
ドコモにきてほしいなよ、こんなじゃ
https://t.co/rU61Uz0eiu
YahooNewsTopics [auとSB 格安ブランド移行無料] タイトルを見て、auのサブブランドとソフトバンクのサブブランド間の移行が無料になるのかと期待してびっくりした。ドコモのdhamoに対抗するたにこれからは競合していくのか
auソフトバンクはそれじゃ勝負できないのか。ドコモに軍配が上がる。 https://t.co/rU61Uz0eiu
YahooNewsTopics それだけでドコモに負けるよ
ワイモバやUQへの移行を促さずしたところで

図5 コンテキストグラフの具体例⁶

語を取り除くことでコンテキストキーワードとした。メンションブロックでは、ユーザが気になったコンテキストキーワードを選ぶことで、それに関するメンションツイートを読むことができるようになっていく。

提案手法であるコンテキストグラフのコンテキスト理解の有効性を確認するため、ワードクラウドを用いたインタフェース(図6)、実際のTwitter UIを模したインタフェース(図7)と比較する。

ニュース本文

携帯電話大手のKDDI（au）は9日、主力ブランドと格安ブランドとの間で契約を乗り換える際の手数料を廃止し以降に適用すると発表した。ソフトバンクも一部の店舗での手続きを廃止して対応する。これは最大1万5500円（税別）が必要だった。ソフトバンクは業者からの予定で、具体的な時期は改めて公表する。 撤廃するのには、ブランド移行の手続きをした場合の契約更新手数料や、電話番号を変えに前の携帯電話会社へ乗り換える「番号ポータビリティ」（持ち運び）制度の手続き料、税金、同様に廃止や予定が報じられている。定額契約の途中で乗り換える場合の契約解除料も無料にする。



キーワード

検索
YahooNewsTopics 結果、ドコモが撤廃か
YahooNewsTopics ドコモの発表にauもソフトバンクもついてこれないじゃないか こんなクソみたいな対応してるでやめて欲しいな
ドコモにきてほしいなよ、こんなじゃ
https://t.co/rU61Uz0eiu
YahooNewsTopics [auとSB 格安ブランド移行無料] タイトルを見て、auのサブブランドとソフトバンクのサブブランド間の移行が無料になるのかと期待してびっくりした。ドコモのdhamoに対抗するたにこれからは競合していくのか
auソフトバンクはそれじゃ勝負できないのか。ドコモに軍配が上がる。 https://t.co/rU61Uz0eiu
YahooNewsTopics それだけでドコモに負けるよ
ワイモバやUQへの移行を促さずしたところで

図6 ワードクラウドの例⁶図7 Twitter UIを模したインタフェースの例⁶

2020年12月7日から同年12月26日の間にYahoo!ニュースで掲載され、Twitterにおいて見出しが配信されたニュースのうち、リプライと引用リツイートが合わせて50件以上あったものを対象に、570件のニュースのデータセットを構築した。今回のインタフェースの評価実験では、データセットの中から無作為に選んだニュース12件を使って行なった。

九州大学に所属する学生男女15人に評価実験を行なってもらった。被験者を5人ずつの3グループに分け、それぞれの

グループに異なる順でインタフェースを提示して評価してもらった。

被験者に対し、各ニュースを3分間提示し、各ニュース毎に以下の質問に5段階で回答してもらった。各ニュースは、3分間経過すると、自動的に質問ページにジャンプするようにした。各ニュース毎に行なった質問を、それぞれ質問1-1、質問1-2とする。

- (1) ニュースの内容を理解できたと思いますか？(質問1-1)
- (2) ニュースに対する世論を理解できたと思いますか？(質問1-2)

また、全てのニュースを提示した後に、それぞれのインタフェースの使いやすさについて、以下の質問に5段階で回答してもらった。インタフェースの使いやすさについての質問を、それぞれ質問2-1、質問2-2とする。

- (1) ニュースを読むためのアプリケーションとして、使ってみたいと思いますか？(質問2-1)
- (2) ニュースアプリケーションとして、使いやすかったですか？(質問2-2)

4.5.1 結果

質問1-1の各インタフェースの平均値を図8に、質問1-2の各インタフェースの平均値を図9に示す。質問1-2に関して、提案手法が他のインタフェースと比較して有意な差を示していることがわかる。よって、提案手法は、ニュースコンテキストを理解するのに有効であると言える。

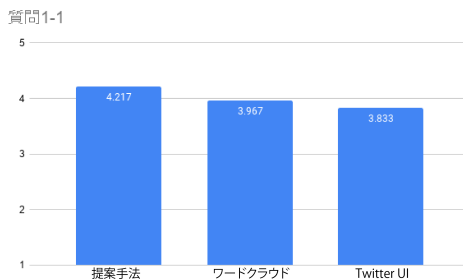


図8 質問1-1に対する各インタフェースの評価の平均値

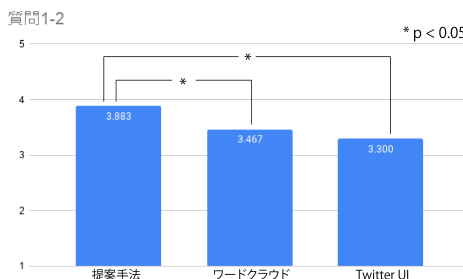


図9 質問1-2に対する各インタフェースの評価の平均値

質問2-1「ニュースを読むためのアプリケーションとして、使ってみたいと思いますか？」の評価値の平均値を図10に、質問2-2「ニュースアプリケーションとして、使いやすかったですか？」の評価値の平均値を図11に示す。質問2-1の結果

を見ると、提案手法が他のインタフェースと比較して有意に高い評価を受けた。質問 2-2 の結果を見ると、提案手法はワードクラウドに対して有意に高い評価を受けた。

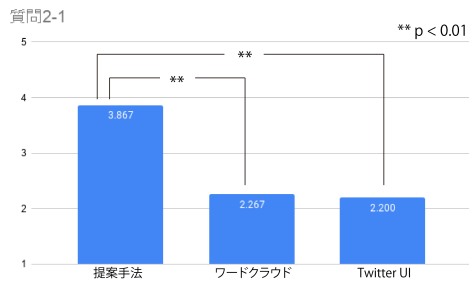


図 10 質問 2-1 に対する各インタフェースの評価の平均値

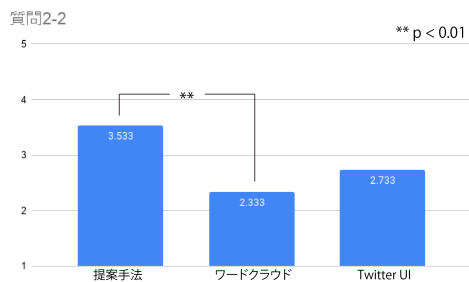


図 11 質問 2-2 に対する各インタフェースの評価の平均値

4.5.2 考察

被験者からは、キーワードについてまとめられたインタフェースの方が民意の理解がしやすかった、そのニュースで連想しなかったキーワードが提示されると、メンションを積極的に読む気になったなどの感想があり、コンテキストグラフがニュースのコンテキスト理解に役立つ手法であったと言える。また、Twitter のメンションを利用したため、批判的な意見が目立ったなどの感想もあり、メンションを投稿したユーザの情報も見るができるようにすると、よりコンテキスト理解を促すのではないかと考えられる。

5 まとめ

本研究では、SNS を利用してニュースを読む際に、ニュースの全体像を俯瞰的に理解したり、世論を偏りなく把握するのが難しいといった問題を解決するため、ニュースコンテキストの理解を促すキーワードを抽出し、ユーザが理解しやすいように可視化して提示する手法を提案した。この手法では、キーワードの抽出手法について、十分な結果が得られなかったが、コンテキストグラフを提示する手法においては、既存のニュース閲覧と比較して有効であるという結果が得られた。このことから、ユーザがニュースコンテキストを理解するには、関連するニュースや、キーワードを提示するという手法でなく、キーワードの重要度や、キーワードについて言及した SNS ユーザの意見も合わせて提示する手法が有効だということがわかった。

今後の課題としては、キーワードの抽出手法に関して改良を行うことや、リアルタイムに動作するシステムとして実装すること、システムを利用したユーザのニュース閲覧履歴から評価を行うことがある。

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] 総務省, 令和元年通信利用動向調査報告書 (世帯編), <https://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR201900.001.pdf>, 2020/5/29 公表
- [2] NHK 放送文化研究所, 国民生活時間調査, <https://www.nhk.or.jp/bunken/yoron-jikan/>
- [3] 消費者庁, 平成 28 年度 消費者庁委託調査事業 インターネット消費者トラブルに関する総合的な調査研究 報告書 SNS に関するアンケート結果 (受託者: 三菱 UFJ リサーチコンサルティング株式会社), https://www.caa.go.jp/policies/policy/consumer_policy/caution/internet/pdf/adjustments_index_1.170111.0002.pdf
- [4] 総務省, 情報通信白書令和元年度版, <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd114210.html>
- [5] 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.
- [6] Akiyo Nadamoto, Katsumi Tanaka, “Time-based Contextualized-News Browser (T-CNB)”, In Proc. of the 13th international World Wide Web conference on Alternate track papers posters, pp. 458 - 459. ACM, 2004.
- [7] 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
- [8] 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
- [9] 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
- [10] Pablo Barbera, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau, “Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?”, Psychological science, vol. 26, no. 10, pp. 1531 - 1542, 2015.
- [11] 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
- [12] Swit Phuvipadawat, Tsuyoshi Murata, “Breaking News Detection and Tracking in Twitter”, In WI-IAT '10 Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, volume 03, pages 120-123, 2010
- [13] Keiichi Mizuka, Yu Suzuki, Akiyo Nadamoto, “Extraction of Commentary tweets about news articles”, iiWAS '17, 2017.12
- [14] Carolina Becatti, Guido Calderelli, Renaud Lambiotte, Fabio Saracco, “Extracting significant signal of news consumption from social networks: the case of Twitter in Italian political elections”, Palgrave Communications 5, 2019
- [15] Antonia Saravanou, Giorgio Stefanoni, Edgar Meij, “Identifying Notable News Stories”, Proc. of The 42nd European Conference on Information Retrieval 2020, 2020
- [16] Nguyen Vo, Kyumin Lee, “The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News”, SIGIR '18, July 8 - 12, 2018
- [17] Karen Spärck Jones, “A statistical interpretation of term specificity and its application in retrieval”, Journal of Documentation, 28, 11-21, 1972