

# 報道が引き起こすソーシャルメディアユーザの感情に基づくメディアの分析

小室 卓也<sup>1</sup> 風間 一洋<sup>2</sup> 吉田 光男<sup>3</sup>  
土方 嘉徳<sup>4</sup>

マスメディアやオルタナティブメディアが報道したニュースは、X (旧 Twitter) 等のソーシャルメディアで拡散され、ユーザに多様な感情を引き起こす。本研究の目的は、逆にこのユーザの感情に基づいて各メディアの特性を定量的に分析することで、メディアの報道姿勢を把握し、ニュースを中立的・俯瞰的な視点から理解できるようにすることである。本論文では、ニュース記事の URL を含む引用ポストからユーザのコメント部分のみを抽出し、T5 を用いた感情分析器で Plutchik の基本 8 感情のベクトルに変換することにより、各メディアの特性を分析する手法を提案する。メディアの感情ベクトルを Z スコアで標準化して比較し、さらに k-means 法のクラスタリング結果と照合したところ、例えば、エンタテインメント系は「喜び・期待」が強い一方、TV キー局は「怒り・恐れ・嫌悪」を強く引き起こすなどの、メディアの業種や報道姿勢による感情反応の明確な違いが確認された。

さらに、引用されたニュース記事数とその引用ポスト数には非常に強い相関があることに注目して、記事数と引用ポスト数からメディアの「影響度」も定義し、特に週刊誌などのスクープを報じるメディアで高い値になることを示した。

## 1 はじめに

マスメディアやオルタナティブメディアの報道したニュースは、X (旧 Twitter) のようなソーシャルメディア上で広く拡散され、多くのユーザに影響を与えている。それらのメディアには保守・リベラルといった報道姿勢の違いがあるだけでなく、注目を集めるためにフェイクニュースや釣り記事を意図的に拡散するメディアも存在し、実世界の予期せぬ対立や事件を招く危険性がある。このような問題を回避するためには、各メディアの報道姿勢を理解し、報道されるニュースを中立的な視点から見る必要があるが、それは容易ではない。

本研究の目的は各メディアの報道姿勢や特性を明らかにすることで報道されるニュースを中立的・俯瞰的な視点から理解できるようにすることである。ただし、メディアは同じ事件について報

道することが多いが、ニュース記事の表層的な単語の違いから報道姿勢や特性の違いを判別するには非常に高度な NLP 技術が必要となるため難しい。そこで記事を読んだユーザの意見が反映されたポスト (投稿) に着目すれば、逆にメディアの報道姿勢を推測することができると考え、本論文ではメディアの報道によって引き起こされたソーシャルメディアユーザの感情からメディアを分類、分析することでメディアごとの特徴や報道姿勢を明らかにする。実際には、まずメディアのニュース記事とその記事の URL を含んで引用したポストを収集し、ユーザの発言部分を抽出する。次に T5 を用いた感情分析器によってポスト単位の感情ベクトルを得た後にメディア単位の感情ベクトルを計算する。最後にメディア単位の感情ベクトルを用いてクラスタリングを行い分析する。

## 2 関連研究

### 2.1 感情分析に関する研究

ソーシャルメディアの分析では、喜びや怒りのような感情 (emotion) や、ポジティブやネガティブのような極性 (sentiment) を分析する手法がよく用いられる。これらの手法は、辞書ベースの手法と深層学習ベースの 2 つに大きく分けられる。

辞書ベースの手法として、ML-Ask (eMotive eLement and Expression Analysis system) [1] [2] と LIWC (Linguistic Inquiry and Word Count) がある。ML-Ask は、中村の感情分類 [3] に基づいた感情辞書と一致した数から感情を推定するシステムである。ML-Ask を用いると、対象とするテキストを 10 種類の感情 (哀, 恥, 怒, 厭, 恐, 驚, 好, 昂, 安, 喜) に分類できる。LIWC は、書き手や話し手の心理状態や感情、社会的態度を推定するテキスト分析ツールである [4]。LIWC 辞書は、複数の心理・社会・言語カテゴリーに分類され、各カテゴリーに含まれる単語の出現頻度に基づいて、文章の特徴を数値化しており、ポジティブ感情・ネガティブ感情や不安, 怒り, 悲しみなどが検出できる。日本語テキスト用に、LIWC2015 を日本語化した J-LIWC2015 が開発されている。

ただし、辞書ベースの手法は、分析結果が感情辞書の充実度合いに大きく依存するだけでなく、単語単位では判定できないような複雑な文章を正しく解析できないことがある。

深層学習ベースの手法として、福田らの BERT を用いた手法 [5] と難波らの T5 を用いた手法 [6] がある。どちらも、梶原らの日本語の感情分析データセット WRIME [7] で事前学習済み日本語モデルをファインチューニングした感情・極性分析器である。WRIME は、SNS 上のポストに Plutchik の基本 8 感情 (喜び, 悲しみ, 期待, 驚き, 怒り, 恐れ, 嫌悪, 信頼) を 4 段階 (0, 1, 2, 3), 感情極性 (ネガティブ, ニュートラル, ポジティブ) を 5 段階 (-2, -1, 0, 1, 2) で人手でラベル付けしたデータセットである。深層学習ベースの感情分析器は、辞書ベースの感情分析器よりも感情の検出率が高く、記号に加えて、複数の単語、文で表される表現も検出できるという利点がある。ただし、WRIME は怒り, 信頼の 2 つの感情のデータ数が他の感情に比べて少ないために、その 2 つの感情検出率は他の感情よりも低くなる点は考慮する必要がある。本論文では、BERT より性能が高い難波らの

<sup>1</sup> 非会員 和歌山大学 システム工学部

s266116@wakayama-u.ac.jp

<sup>2</sup> 正会員 和歌山大学 システム工学部

kazama@wakayama-u.ac.jp

<sup>3</sup> 正会員 筑波大学 ビジネスサイエンス系

mitsuo@gssm.otsuka.tsukuba.ac.jp

<sup>4</sup> 正会員 兵庫県立大学大学院情報科学研究科

contact@soc-research.org

T5 の感情分析器を用いる。

## 2.2 ソーシャルメディアを感情分析した研究

実際にソーシャルメディアを感情分析器を用いて分析した研究が存在する。鳥海らは Twitter 上で新型コロナウイルスに関係するポストを収集し、辞書ベースの感情分析器である ML-Ask を用いてポストから感情を推定し、その変化を分析した [8]。福田らは言語モデルの 1 つである BERT を用いた感情分析器を構築し、Twitter から収集したワクチンに関するツイートから新型コロナウイルスワクチンに対する人々の感情とその要因を分析した [5]。本論文ではニュース報道を対象とするが、Web サイトやソーシャルメディアで報道されたニュース記事そのものではなく、そのニュース記事を引用して X 上に投稿されたポスト（引用ポスト）を感情分析して、その結果をニュース記事が X ユーザに与える感情と見做す。また、引用ポストにはユーザのコメント以外の情報も含まれるが、それらは除き、X ユーザ自身が直接記述したコメント部分のみを感情分析する。

## 2.3 メディアの報道姿勢やバイアスの分析に関する研究

メディアの報道姿勢や政治的バイアスの定量的な分析に関する様々な研究が存在する。Chao らは、Ad Fontes Media [9] が公開している複数の専門家によって評価されたメディアのバイアスおよび品質データを用いて、メディアの X 公式アカウントによりポストされる文からそのメディアの政治的バイアスと記事の品質を予測する LSTM モデルを構築することで、メディアの政治的バイアスと記事の品質を推定する手法を提案した [10]。Huang らは、メディアバイアスをマクロおよびミクロの両視点から分析するフレームワークを提案した [11]。マクロレベルでは、イベント言及データに基づいて各メディアのベクトル表現を生成し、報道対象の選択傾向を分析した。ミクロレベルでは、メディアごとの記事データから構築した Word2Vec モデルと心理学のセマンティック・ディファレンシャル法の考え方を用いて、対立語群への意味的近さの差に基づいて、ジェンダーや政治的傾向などのバイアスを抽出した。

上記 2 件の研究がメディアの発信コンテンツを分析対象とするのに対し、本論文と同様にソーシャルメディア上のユーザの反応に着目する研究として、久田らは、特定の政治的な話題に関するニュースへのユーザコメントにトピックモデルを適用してメディアごとのトピック推定確率を算出し、その類似度に基づくクラスタリングによりメディアの政治的バイアスを分析した [12]。本論文では、引用ポストを感情分析してメディアごとの感情ベクトルを算出して、それを用いてメディアのクラスタリングや分析を行う。

## 3 感情に基づく分析手法

### 3.1 報道とソーシャルメディアユーザの反応

メディアの報道するニュースは X のようなソーシャルメディア上で拡散され、例えば音楽やファッションなどを扱うエンタテインメント系のメディアであればユーザに喜びの感情が、スクープを報じて大衆の注目を惹くような週刊誌系のメディアであればユーザに怒りの感情を引き起こしやすいだろうと推測できる。読んだニュース記事に他の人にも知らせる価値があると判断した時

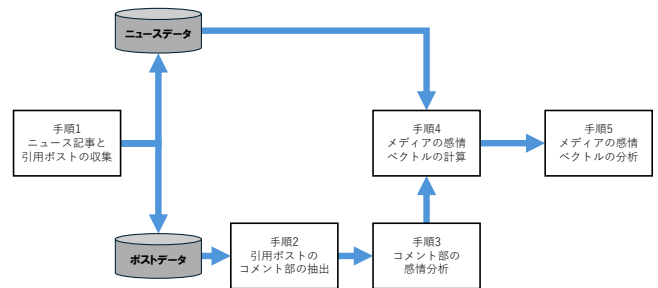


図1 提案手法の概要

には、ニュース記事の URL を含めてポストするが、ニュース記事に抱いた感情は、そのコメント部分のテキストに現れると考えられることから、感情分析器を用いて抽出する。

本論文で提案する感情に基づく分析手法の手順を以下に、概要を図1に示す。

手順1 ニュース記事データとその引用ポストデータの収集

手順2 引用ポストのコメント部の抽出

手順3 コメント部の感情分析

手順4 メディアの感情ベクトルの計算

手順5 メディアの感情ベクトルの分析

### 3.2 ニュース記事データと引用ポストデータの収集

メディアが報道したニュース記事は、X 上ではユーザがその記事を引用、すなわちその記事の URL を含むポストをすることで拡散される。そこで、ユーザのポストの中から、URL のホスト部に観測対象のメディアのホスト名を含むポスト・リポストと、その URL が示すニュース記事を継続的に収集し、データの関係や属性などを含めて全文検索可能なデータベースに保存する。データベースには Elastic 社のオープンソース検索エンジンである Elasticsearch を用いる。

### 3.3 引用ポストのコメント部の抽出

ニュース記事の引用ポストには、その URL に加えて、記事のタイトル、記事のカテゴリ、メディア名、メディアのユーザ名、ハッシュタグのような引用したニュース記事や報道したメディアに関する情報が自動または手動で挿入されていることが多い。しかし、このような情報はユーザの引用意図を反映しているわけではなく、そのまま残しておくことでノイズや誤検出の原因となる。そこで、中北らの手法 [13] を用いて、これらの情報を除去し、ユーザが自分の意思に基づいて記述したコメント部のみを抽出する。ただし、この処理を適用後にテキストが残らないポストは、ユーザのコメントが付与されていないと判断して除外する。

### 3.4 コメント部の感情分析

次に、抽出したコメント部を感情分析し、Plutchik の基本 8 感情で構成される感情ベクトルを取得する。感情分析には難波らの提案した T5 を用いた深層学習ベースの感情分析器 [6] を使用する。

### 3.5 メディアの感情ベクトルの計算

次に、引用ポストの感情ベクトルから、各メディアの感情ベクトルを計算する。なお、メディアは、機械的に処理できるように、

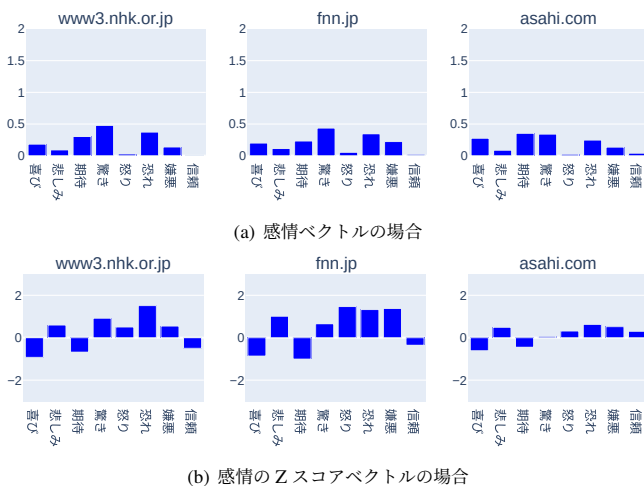


図2 メディアの感情特徴の可視化例

ニュース記事の URL のホスト部で区別する。さらに、メディア名は、ホスト名の FQDN の先頭に“www.”がある場合にはそれを取り除いた文字列を用いる。

メディアの感情ベクトルは次の手順で計算する。

1. ある記事に対するユーザのすべてのポストの感情ベクトルの平均値を計算し、それをユーザのその記事に対する感情ベクトルとする。この理由は、特定のユーザが同じ記事を複数回引用することがあるので、特定のユーザの偏りを排除し、引用したすべてのユーザの影響を均等に評価するためである。
2. ある記事を引用してポストしたすべてのユーザのその記事に対する感情ベクトルの平均値を計算し、それをその記事の感情ベクトルとする。
3. あるメディアのすべての記事の感情ベクトルの平均値を計算し、それをそのメディアの感情ベクトルとする。

### 3.6 メディアの感情ベクトルの分析

得られた 8 次元のメディアの感情ベクトルは、次に述べる 3 種類の方法で分析する。

#### 3.6.1 メディアの感情特徴の可視化

メディアの感情ベクトルをそのまま可視化しても、ソーシャルメディアの発言における感情の偏りや教師データ由来の感情分析器の感情の検出能力の偏りが存在するために、メディアごとの特徴が分かりにくい。そこで、あるメディアがメディア全体に対して、相対的にどのような感情が突出しているかを示すために、各感情ごとに Z スコアを計算した感情の Z スコアベクトルを作成し、それを可視化する。標準化された値である Z スコアを用いることで上記の偏りによる影響は軽減される。実際に一部のメディアの感情特徴を、感情ベクトルをそのまま用いて可視化した場合と Z スコアベクトルを可視化した場合の例を、図 2 に示す。

#### 3.6.2 メディアの感情ベクトルによるクラスタリング

感情ベクトルが類似しているメディア群を把握するために、メディアの感情ベクトルに基づいてクラスタリングする。なお、ニュース記事数が非常に少ないメディアは限られた記事やユーザ

の影響が相対的に大きくなるためか、感情ベクトルの L2 ノルムが大きくなりクラスタリング結果に悪影響を与えるために、閾値を設定して除外した。クラスタリングには k-means 法を用いて、k の値はメディアの感情特徴の可視化結果を手手で確認しながら、類似した感情特徴のメディアが同一グループになるように決定する。

#### 3.6.3 メディアの感情ベクトルの分布の可視化

メディアの感情ベクトルがどのように分布しているかを把握するために、次元削減手法の 1 つである UMAP [14] を用いて 2 次元に可視化する。なお、メディアはクラスタリング結果に基づいて、同一クラスターが同じ色になるように彩色する。

## 4 分析

### 4.1 データセット

ニュース記事のデータセットとして、統合型メタ検索エンジン Cseek.jp で扱われているニュース記事のうち、2022 年 1 月 1 日から 2022 年 3 月 31 日までの 1,139,350 件のニュース記事を利用した。

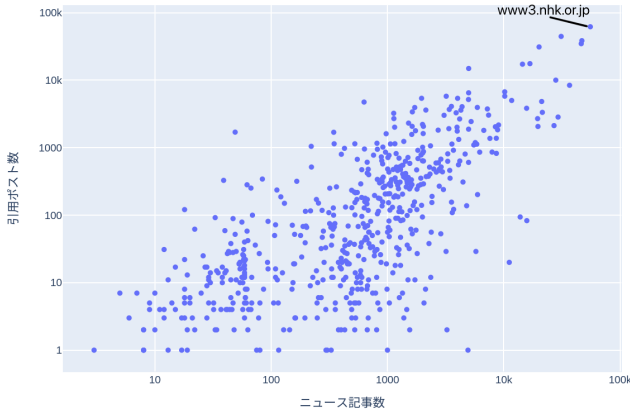
ニュースを引用したポストのデータセットとして、X API で収集した 2022 年 1 月 1 日から 2022 年 11 月 30 日までのポストのうち、ニュース記事の URL が含まれる 993,358 件のポストを使用した。このデータセットに含まれていたメディア数は 566 個であった。

### 4.2 メディアの統計分析

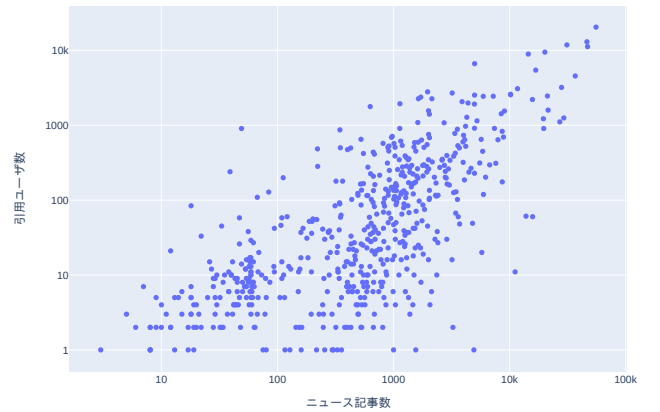
まず、メディアのニュース記事数と引用ポスト数、引用ユーザ数の関係について分析した。

各メディアのニュース記事数と引用ポスト数およびメディアのニュース記事数と引用ユーザ数を両対数プロットした結果を、図 3(a) と図 3(b) に示す。横軸はメディアのニュース記事数、縦軸はメディアの引用ポスト数または引用ユーザ数であり、プロットされた各点はメディアを表す。ニュース記事数と引用ポスト数の相関係数は 0.80、ニュース記事数と引用ユーザ数の相関係数は 0.81 で、どちらも強い相関があった。引用ポスト数に加えて引用ユーザ数を用いた理由は、ユーザがニュース記事を複数回引用ポストすることが考えられるので、その影響を除外するためである。一番ニュース記事数が多い右上のメディアは NHK であり、全体的に相関は高いが、同じニュース記事数でも引用ポストされる回数や人数はかなり異なることがわかる。

次に、引用数 0 のニュース記事も集計対象に含まれるので、引用がないニュース記事を除外して分析した。各メディアのニュース記事数と引用ポスト数およびメディアのニュース記事数と引用ユーザ数を両対数プロットした結果を、図 4(a) と図 4(b) に示す。横軸はメディアのニュース記事数、縦軸はメディアの引用ポスト数または引用ユーザ数である。ニュース記事数と引用ポスト数の相関係数は 0.98、ニュース記事数と引用ユーザ数の相関は 0.96 と、どちらも非常に強い相関があった。対数変換を行うと右上がりの直線状の分布となるため、引用される記事に絞った場合はニュース記事数と引用ポスト数の間にはべき乗則に従う関係があると推測する。引用数 0 のニュース記事を除外することで相関係数が大きく向上し、より線形の分布に近づく理由として、速報や

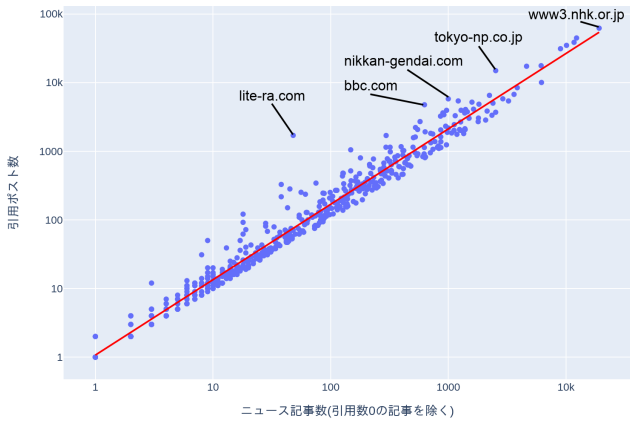


(a) ニュース記事数と引用ポスト数

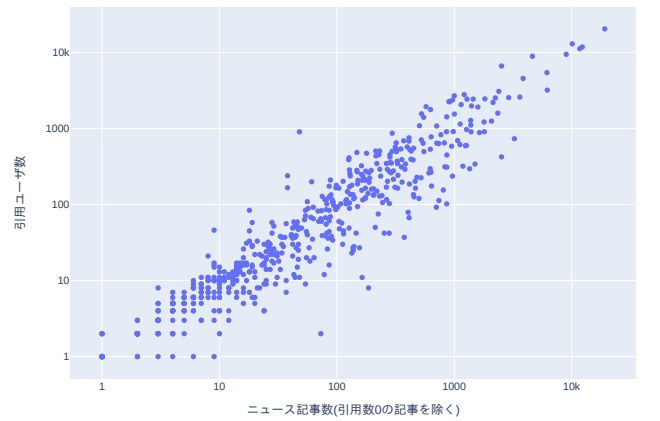


(b) ニュース記事数と引用ユーザ数

図3 メディアのニュース記事数と引用ポスト数・引用ユーザ数の関係



(a) ニュース記事数と引用ポスト数



(b) ニュース記事数と引用ユーザ数

図4 メディアのニュース記事数と引用ポスト数・引用ユーザ数の関係 (引用数0を除いた場合)

転載などのニュース記事の種類やトピックは引用されるかどうかに関与するが、引用されるニュース記事に絞ると大きな差はないからだと推測する。また図4(a)が図4(b)よりも線形に近い理由として、ユーザ単位に集約しなければその傾向が顕著になるからだと推測する。

そこで、図4(a)に最小二乗法で回帰直線を求めてプロットした。この結果を見ると、少数のメディアは回帰直線からかなり上方に存在し、同じ記事数でもより多くの引用数を獲得していることがわかる。そこで、メディアの影響度  $E(i)$  を、メディア  $i$  の引用数を  $c_i$ 、回帰式から求めた引用数を  $c'_i$  として、次のように定義する。

$$E(i) = \text{sgn}(\log c_i - \log c'_i)(\log c_i - \log c'_i)^2 \quad (1)$$

残差を強調するために2乗する。さらに、符号関数  $\text{sgn}(\dots)$  を用いて、回帰直線の上なら影響度が正、回帰直線の下なら負となるようにする。

影響度の上位15件のメディアの影響度と記事数、引用数、クラスター（後から述べる）を表1に示す。ただし、記事数が非常に少ないメディアはユーザに与える影響が小さいので、記事数が

10未満のメディアは除外した。1, 2, 3, 8, 10, 11位はWebニュースメディア、4, 14, 15位は新聞社、5, 7, 12, 13位は雑誌社、6位はテレビ局、9位はWebポータルサイトであり、複数の種類のメディアが混在していた。

特に、ニュース系メディアとして *lite-ra.com* (リテラ)、*bbc.com* (BBC)、*wezz-y.com* (wezzy)、*nikkan-gendai.com* (日刊ゲンダイ)、*facta.co.jp* (FACTA)、*tokyo-np.co.jp* (東京新聞) のようにスクープやスキャンダルを報じるメディアが、趣味やエンタテインメント系として *numero.jp* (Numero TOKYO)、*tetsudo-shimbun.com* (鉄道新聞)、*cinra.net* (CINRA) が含まれており、このようなメディアの記事は一般的な新聞社やTV局などのマスメディアよりも多くの引用を生み、ソーシャルメディアのユーザに与える影響力が高くなると考えられる。

#### 4.3 メディアの感情特徴の分析

ニュース記事に対してユーザが持つ感情が、メディアによってどのように異なるかについて分析した。メディアの感情のZスコアベクトルの可視化結果を、ニュース記事数の上位から60件取り出して、左から右、上から下に並べて図5に示す。横軸はPlutchikの基本8感情、縦軸は感情のZスコアの値である。この

表 1 影響度上位のメディア

	メディア名	影響度	引用数	記事数	クラスタ
1	lite-ra.com	1.833	1700	48	7
2	news.yahoo.co.jp	0.563	328	38	2
3	careerhack.en-japan.com	0.454	121	18	0
4	globe.asahi.com	0.367	1047	148	1
5	nikkei-science.com	0.365	282	45	1
6	bbc.com	0.329	4745	627	4
7	numero.jp	0.326	217	38	4
8	tetsudo-shimbun.com	0.308	92	18	2
9	bengo4.com	0.235	1693	296	4
10	wezz-y.com	0.204	252	56	2
11	cinra.net	0.199	343	75	3
12	nikkan-gendai.com	0.192	5787	999	4
13	facta.co.jp	0.179	72	19	4
14	tokyo-np.co.jp	0.165	14921	2527	4
15	webronza.asahi.com	0.158	799	179	1

結果から、メディアによってユーザがニュース記事に抱く感情の傾向が大きく異なることがわかる。例えば、テレビ局や新聞社のメディアである fnn.jp (フジニュースネットワーク)、sankei.com (産経新聞)、yomiuri.co.jp (読売新聞)、news-tv-asahi.co.jp (テレビ朝日)、news24.jp (日本ニュースネットワーク)、news.tbs.co.jp (ジャパン・ニュース・ネットワーク) のニュース記事に対しては、ユーザは怒り・恐れ・嫌悪の感情を他よりも顕著に持つことがわかる。これに対して、趣味やエンタテインメント系のメディアである natalie.mu (ナタリー)、hobby.dengeki.com (電撃ホビー)、ascii.jp (ASCII)、fashion-press.net (FASHION PRESS)、spice.eplus.jp (SPICE)、crank-in.net (クランクイン)、realsound.jp (Real Sound) のニュース記事に対しては、ユーザは喜び・期待の感情を他よりも顕著に持つことがわかる。また、公共放送である www3.nhk.or.jp (NHK) のニュース記事に関しては、ユーザは驚き・恐れ・嫌悪の感情は顕著に持つが、怒り・嫌悪を持つことは相対的に少なく、民間放送とは異なる傾向を示すことがわかる。ただし、同じ業種であっても、すでに述べた感情特徴を明確に示すとは限らず、メディアの報道姿勢に強く影響されると考えられる。

#### 4.4 クラスタの分析

類似する感情特徴を持つメディア群の存在可能性を確認できたので、メディアを感情ベクトルに基づいてクラスターリングし、各クラスタの特徴を分析した。

ただし、すべてのメディアを分析対象にした場合には、ニュース記事数が少ないメディアがクラスタ分割に悪影響を与えることがわかった。この理由を分析した結果、ニュース記事数の少ないメディアは少数のニュース記事だけから感情ベクトルが計算されるために、感情ベクトルの各要素が大きくなってしまうことが原因であった。メディアのニュース記事数と感情ベクトルの L2 ノルムの関係を図 6 に示す。横軸はメディアのニュース記事数、縦軸はメディアの感情ベクトルの L2 ノルムである。

図 6 に示されるように、記事数の非常に少ないメディアでは感情ベクトルの L2 ノルムが大きくなる傾向があり、メディアの感情ベクトルとするには妥当ではないと考えた。そこで、ニュース記事数が 10 以上のメディアだけを対象に k-means 法でクラスターリングした。クラスタ数  $k$  は 9 とした。どのようにクラスタ

に分割されたかを調べるために、まず感情ベクトルに基づいて UMAP でメディアを 2 次元に配置し、クラスタ別に着色して可視化した。この可視化結果を図 7 に示す。この結果から、特徴空間においてクラスタが必ずしも明確に分離してはいないことがわかる。

次に各クラスタがどのような感情特徴を持つかを分析した。クラスタの中心の感情ベクトルから Z スコアベクトルに変換して可視化した結果を、図 8 に示す。可視化結果の一番左側のクラスタ 5 は喜び・期待の感情が強く、一番右側のクラスタ 7 は悲しみ・驚き・怒り・恐れ・嫌悪が強く、可視化結果の右と左で大きく感情特徴が違うことがわかる。さらに、喜びの感情が強いクラスタ 0 はクラスタ 5 の右隣に配置され、期待の感情が強いクラスタ 2 はクラスタ 5 の右下に配置されていた。また、驚きの感情が強いクラスタ 8 はクラスタ 7 の左上に、悲しみ・驚き・怒り・恐れが強いクラスタ 4 はクラスタ 7 の左下に配置され、感情的特徴の違いと連続性を反映した配置になっていることがわかる。

次に各クラスタにどのようなメディアが含まれるかを分析した。各クラスタのニュース記事数上位 15 件のメディアを、降順に配置した結果を表 2 に示す。まず可視化結果の左側に注目すると、クラスタ 5, 0, 2 はエンタテインメント系のメディアであり、さまざまな分野を報道しているメディアが混在していた。ただし、クラスタ 0 に隣接しているクラスタ 3 もエンタテインメント系であるが、その上位は sponichi.co.jp (スポニチ)、tokyo-sports.co.jp (東スポ)、hochi.news (スポーツ報知) などのスポーツ紙で占められており、メディアの傾向が少し変わっていた。ただし、クラスタの感情の Z スコアベクトルを考慮すると、ソーシャルメディアユーザに与える感情的な影響は小さいと考えられる。

可視化結果の右側に注目すると、一番右側のクラスタ 7 の上位は news.tv-asahi.co.jp (テレビ朝日)、news.tbs.co.jp (TBS)、news24.jp (日本テレビ)、ytv.co.jp (読売テレビ) などのテレビのキー局・準キー局で占められていた。クラスタの感情の Z スコアベクトルで悲しみ・驚き・怒り・恐れ・嫌悪などの感情が顕著だったことも考慮すると、このクラスタのメディアはソーシャルメディアユーザに与える影響が極めて高いと推測できる。その左下のクラスタ 4 の上位は www3.nhk.or.jp (NHK) のような公共放送と、asahi.com (朝日新聞)、mainichi.jp (毎日新聞)、sankei.com (産経新聞)、yomiuri.co.jp (読売新聞) などの全国紙の新聞社と、jiji.com (時事通信社) などの通信社などが、右上のクラスタ 8 の上位は東海テレビや名古屋テレビ、テレビ宮崎などのローカル局と山陽新聞や徳島新聞などの地方紙によって占められていた。さらに、クラスタ 4 に隣接するクラスタ 1 の上位は nikkei.com (日本経済新聞)、sankeibiz.jp (SankeiBiz)、toyokeizai.net (東洋経済)、diamond.jp (ダイヤモンド・オンライン) のような経済専門紙が多く、同じ新聞社でも傾向が異なることがわかる。

さらに、クラスタ 1 とクラスタ 2 の間のクラスタ 6 は nikkansports.com (日刊スポーツ) のようなスポーツ紙と hokkaido-np.co.jp (北海道新聞)、chunichi.co.jp (中日新聞)、kyoto-np.co.jp (京都新聞)、nishinippon.co.jp (西日本新聞) などの地方紙によって占められており、中間の特徴を示すことがわかる。クラスタの感情の Z スコアベクトルから、このクラスタのメディアの与える



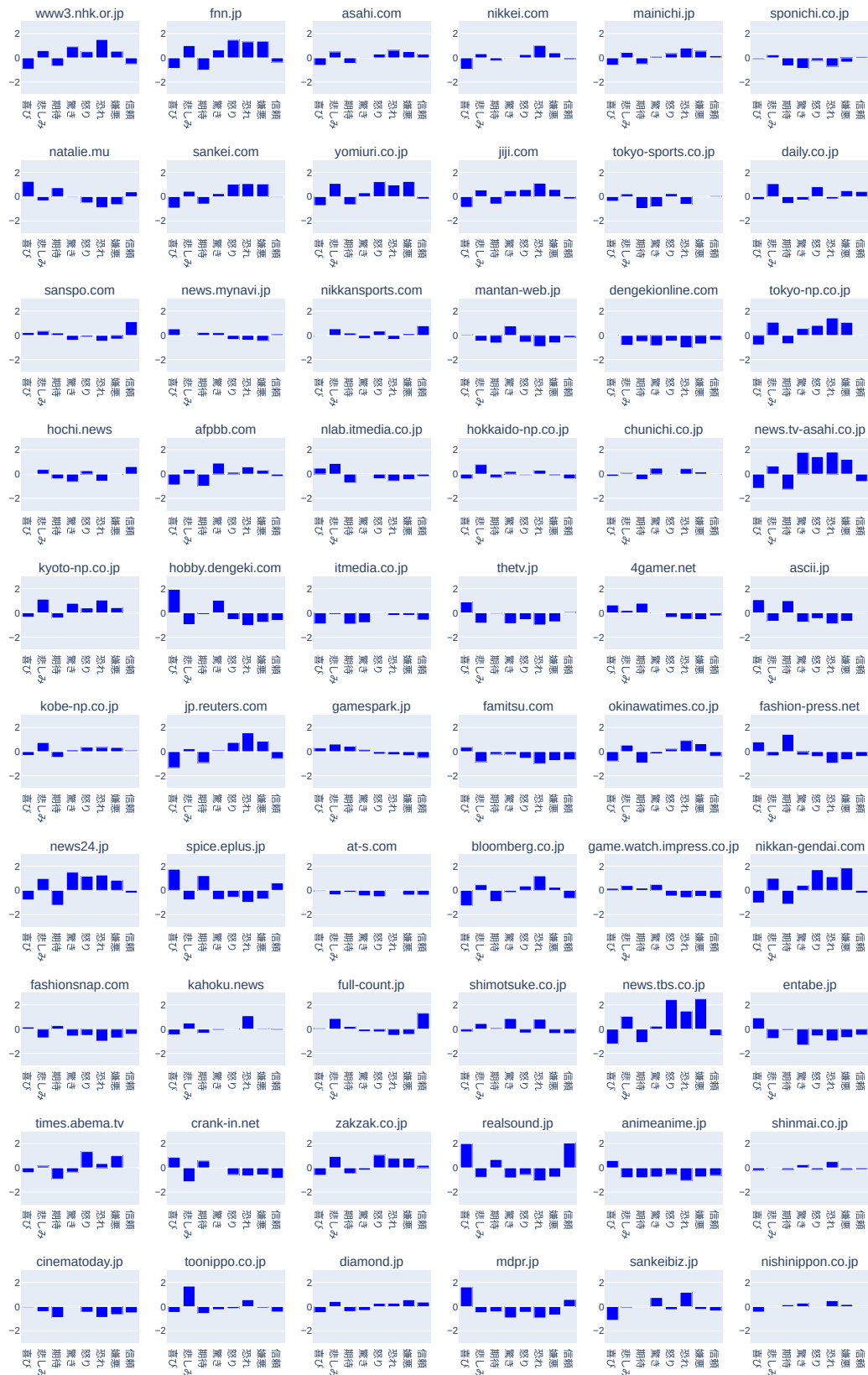


図5 メディアの感情のZスコアベクトルの可視化結果

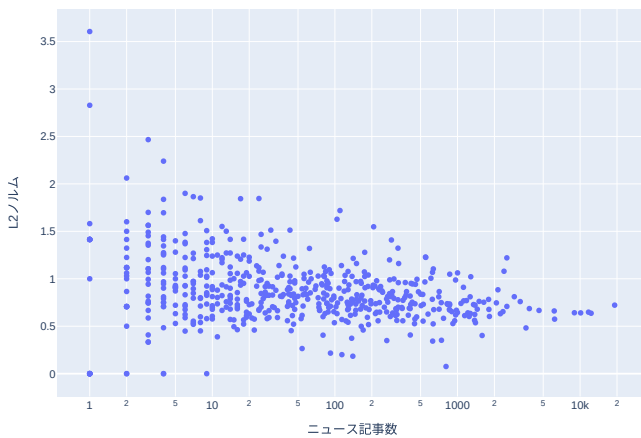


図6 ニュース記事数と感情ベクトルの L2 ノルムの関係

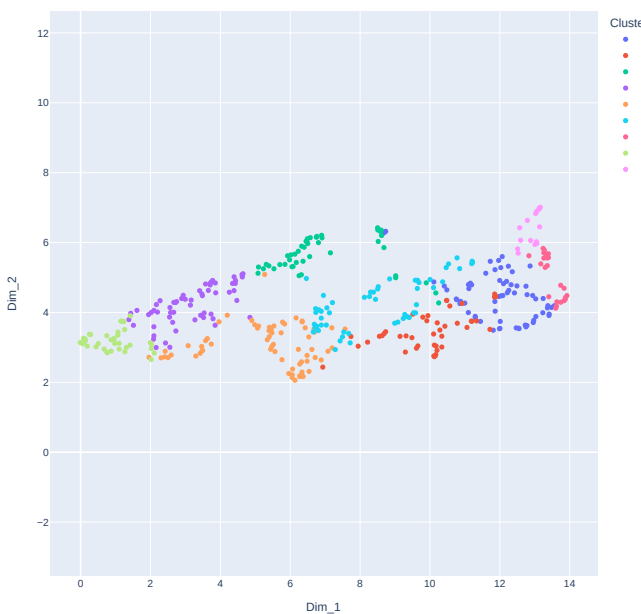


図7 感情ベクトルによるメディアの2次元可視化結果

感情的な影響は、クラスタ4よりも明らかに小さい。

以上を考慮すると、可視化結果のクラスタ間に明確なギャップが存在しないことからわかるように、業種分類が同じメディアでも複数のクラスタに分散する傾向はあるものの、クラスタによって業種分類の傾向が異なり、特に今回の可視化結果の右側にはマスメディアが、可視化結果の左側にはオルタナティブメディアが配置される傾向があった。なお、ネガティブな感情的な影響力に関しては、特に一番右側のクラスタ7が強かった。すなわち、ソーシャルメディア上でテレビのキー局が全国紙よりも強い感情的影響力を持つことになるが、その理由として、テレビでは報道番組の一部を動画として公開することが多く、そのような情報伝達方法が原因であることが考えられる。

#### 4.5 影響度とクラスタの関係の考察

表1を見ると、影響度上位のメディアのクラスタは幅広く分布していた。このことから特定のクラスタが強い影響力を持つとい

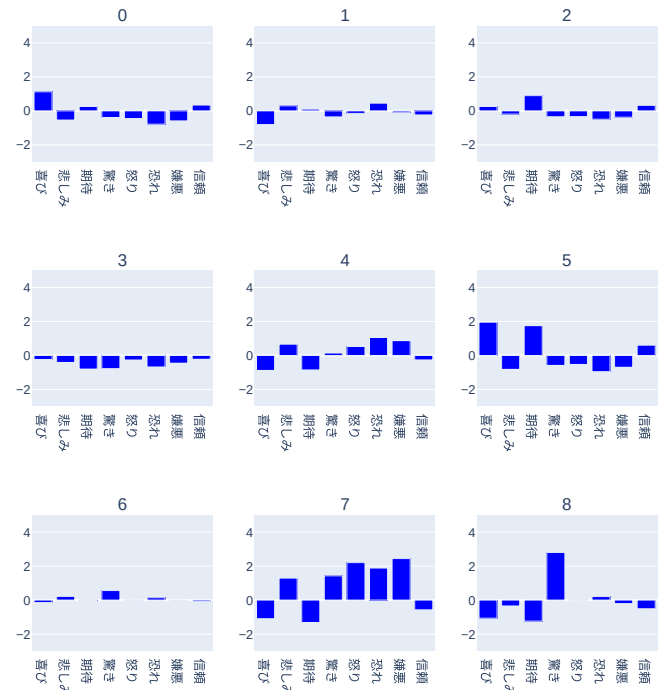


図8 クラスタの感情のZスコアベクトルの可視化結果

うわけではないと推測する。ただし、例外的なクラスタは6と8である。この2つのクラスタは地方の放送局や、地方紙の新聞社が多く属しており、キー局や全国紙が多いクラスタよりもユーザーに与える影響が小さいと推測する。なお、影響度11位のメディアが感情のZスコアベクトルの低いクラスタ3に属しているが、感情ベクトルの大きさというより、その分布の類似性でクラスタリングされるため、同一クラスタでも影響度の大きさはさまざまであると考えられる。

そこで、ニュース報道の分野に絞った場合に、どのようなメディアが影響度上位に来やすいかについて分析する。具体的には、クラスタ4, 6, 7, 8から影響度順上位10件のメディアを抽出して、表3に示す。ニュース報道関係で上位を占めるメディアは、リベラルまたは中立的なマスメディアと、雑誌社またはオンラインニュースサイトの2種類に大きく分かれる。前者は、bbc.com (BBC), tokyo-np.co.jp (東京新聞)であり、社会の問題を取材して世の中に広く知らせることを目的として報道しており、それゆえにスクープで注目されることも多い。bengo4.com (弁護士ドットコム)も、社会の問題を取り上げて弁護士の法的な観点からニュース報道するので、こちらに近いと考えられる。後者は、lite-ra.com (リテラ), nikkkan-gendai.com (日刊ゲンダイ), facta.co.jp (FACTA), president.jp (PRESIDENT Online), jisin.jp (女性自身), smart-flash.jp (Smart FLASH), dot.asahi.com (AERA.dot)であり、新聞社や放送局とは異なる報道姿勢を持っており、スクープだけでなくスキャンダルも報道することがある。つまり、影響度が高くなるメディアは単にニュース報道するだけでなく、スクープを連発したり、スキャンダルを報じるなど、一般的なメディアとは異なる点があると考えられる。

表 2 各クラスターの主要メディア

クラスター 0	クラスター 1	クラスター 2
hobby.dengeki.com natalie.mu mdpr.jp oricon.co.jp ascii.jp thetv.jp buzzfeed.com news.dwango.jp otakei.otakuma.net moviecollection.jp maidonnews.jp animatetimes.com astage-ent.com entabe.jp joshi-spa.jp	nikkei.com sankeibiz.jp kanaloco.jp toyokeizai.net diamond.jp kahoku.news businessinsider.jp nhk.or.jp watch.impress.co.jp toonippo.co.jp newsswitch.jp business.nikkei.com k-tai.watch.impress.co.jp response.jp webronza.asahi.com	news.mynavi.jp sanspo.com fashionsnap.com fashion-press.net 4gamer.net gamespark.jp full-count.jp japan.cnet.com sportiva.shueisha.co.jp withnews.jp logmi.jp minyu-net.com dime.jp nikkan-spa.jp av.watch.impress.co.jp
クラスター 3	クラスター 4	クラスター 5
sponichi.co.jp tokyo-sports.co.jp nlab.itmedia.co.jp hoichi.news famitsu.com itmedia.co.jp daily.co.jp cinematoday.jp townnews.co.jp animeanime.jp akiba-pc.watch.impress.co.jp saga-s.co.jp dengekionline.com sanin-chuo.co.jp tv-tokyo.co.jp	www3.nhk.or.jp asahi.com mainichi.jp sankei.com jiji.com yomiuri.co.jp fnn.jp tokyo-np.co.jp afpb.com jp.reuters.com bloomberg.co.jp okinawatimes.co.jp zakzak.co.jp chibanippo.co.jp nikkan-gendai.com	spice.eplus.jp realsound.jp barks.jp thefirsttimes.jp walkerplus.com crea.bunshun.jp mi-mollet.com wpb.shueisha.co.jp okmusic.jp hamakei.com lisani.jp enterstage.jp chichibu.keizai.biz gs.dengeki.com numero.jp
クラスター 6	クラスター 7	クラスター 8
nikkansports.com hokkaido-np.co.jp chunichi.co.jp kyoto-np.co.jp trafficnews.jp nishinippon.co.jp kobe-np.co.jp kuruma-news.jp huffingtonpost.jp shinmai.co.jp mantan-web.jp news-postseven.com japan.focustaiwan.tw at-s.com dot.asahi.com	news.tv-asahi.co.jp news.tbs.co.jp news24.jp ytv.co.jp chosunonline.com hbc.co.jp webun.jp mbs.jp nnp.co.jp newsdigest.jp asahi.co.jp stv.jp 47news.jp hicbc.com lite-ra.com	tokai-tv.com nagoyatv.com umk.co.jp sanyonews.jp topics.or.jp japanese.cri.cn kyt-tv.com kts-tv.co.jp tosonline.jp logi-today.com viet-jo.com agara.co.jp rkc-kochi.co.jp teny.co.jp tdb.co.jp

表 3 クラスター 4, 6, 7, 8 の影響度上位のメディア

	メディア名	影響度	引用数	記事数	クラスター
1	lite-ra.com	1.833	1700	48	7
2	bbc.com	0.329	4745	627	4
3	bengo4.com	0.235	1693	296	4
4	nikkan-gendai.com	0.192	5787	999	4
5	facta.co.jp	0.179	72	19	4
6	tokyo-np.co.jp	0.165	14921	2527	4
7	president.jp	0.137	2700	575	4
8	jisin.jp	0.115	479	127	4
9	dot.asahi.com	0.107	2213	526	6
10	smart-flash.jp	0.091	435	126	4

以上のことから、影響度は特定の分野のメディアを判別することはできないが、本論文で行ったようなユーザの感情ベクトルに基づくクラスタリングのような何らかの報道姿勢やメディア種別と組み合わせ分析した場合には、その分野において特に影響力の高いメディアを発見するために役に立つだけでなく、その影響力の高さの理由の分析も可能になると思われる。

## 5 おわりに

本論文では、メディアのニュース報道がソーシャルメディアのユーザに与える感情的影響の分析方法を提案した。まず、引用されたニュース記事数とその引用ポスト数には非常に強い相関があることに注目して、メディアがユーザに与える影響の度合いである影響度を定義した。さらに、メディアの報道でユーザが抱く感情に基づくクラスタリング手法を提案し、メディアの報道姿勢によって異なるクラスターに分かれることを示した。さらに、それらを組み合わせることで、ある特定の分野でニュース報道の影響力があるメディアを抽出できることを示した。

## 謝辞

本研究は JSPS 科研費 23K21727 と 23K28194 の助成を受けた。

## 参考文献

- [1] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka, and Kenji Araki. A System for Affect Analysis of Utterances in Japanese Supported with Web Mining. 知能と情報, Vol. 21, No. 2, pp. 194–213, 2009.
- [2] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, Kenji Araki, and Fumito Masui. ML-Ask: Open source affect analysis software for textual input in Japanese. *Journal of Open Research Software*, Vol. 5, No. 1, p. 16, 2017.
- [3] 中村明. 感情表現辞典. 東京堂出版, 1993.
- [4] Yla R. Tausczik and James W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, Vol. 29, No. 1, pp. 24–54, 2010.
- [5] 福田悟志, 難波英嗣, 庄司裕子. コロナ禍におけるワクチンに対する人々の感情変化とその要因の分析. 知能と情報, Vol. 34, No. 3, pp. 592–600, 2022.
- [6] Hidetsugu Nanba, Katsuya Yamamoto, Satoshi Fukuda, Hiroko Shoji, Masayoshi Tanishita, Yasushi Kyutoku, and Mitsuru Yamashina. Modeling the social acceptability of technologies using Twitter data. In *Proceedings of the 2023 IEEE Conference on Systems, Man, and Cybernetics (IEEE SMC 2023)*, 2023.
- [7] Tomoyuki Kajiwar, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. WRIME: A new dataset for emotional intensity estimation with subjective and objective annotations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2021)*, pp. 2095–2104, 2021.
- [8] 鳥海不二夫, 榊剛史, 吉田光男. ソーシャルメディアを用いた新型コロナウイルス禍における感情変化の分析. 人工知能学会論文誌, Vol. 35, No. 4, pp. F-K45\_1–7, 2020.
- [9] Ad Fontes Media. Interactive media bias chart. <https://adfontesmedia.com/interactive-media-bias-chart/> [アクセス日: 2025 年 11 月 3 日].
- [10] Zehan Chao, Denali Molitor, Deanna Needell, and Mason A. Porter. Inference of media bias and content quality using natural-language processing. arXiv:2212.00237 [physics.soc-ph], 2022.
- [11] Hong Huang, Hua Zhu, Wenshi Liu, Hua Gao, Hai Jin, and Bang Liu. Uncovering the essence of diverse media biases from the semantic embedding space. *Humanities and Social Sciences Communications*, Vol. 11, No. 1, p. 656, 2024.
- [12] 久田祥平, 村山太一, 矢田峻太郎, 若宮翔子, 荒牧英治. SNS コメントを用いたニュースメディアバイアスの分析. 第 35 回人工知能学会全国大会論文集 (JSAI 2021), 1D2-OS-3a-04, 2021.
- [13] 中北雄大, 風間一洋, 吉田光男, 土方嘉徳. 感情とトピックに注目したメディアの報道姿勢の分析. 第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2022), B24-2, 2022.
- [14] Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 [stat.ML], 2020.