メディアの報道に対するソーシャルメディアユーザの反応の時系列変化の分析手法

井上 大成† 風間 一洋†† 吉田 光男††† 土方 嘉徳††††

† 和歌山大学大学院システム工学研究科 〒 640–8510 和歌山県和歌山市栄谷 930
†† 和歌山大学システム工学部 〒 640–8510 和歌山県和歌山市栄谷 930
††† 筑波大学ビジネスサイエンス系 〒 112–0012 東京都文京区大塚 3–29–1
†††† 関西学院大学商学部 〒 662–8501 兵庫県西宮市上ケ原一番町 1–155
E-mail: †s246023@wakayama-u.ac.jp, ††kazama@ingrid.org, †††mitsuo@gssm.otsuka.tsukuba.ac.jp,
††††contact@soc-research.org

あらまし クラスタリングや LDA などで得られた文書集合の内容を推測する簡単な手法として，頻度や重要度が高い単語を抽出する方法がよく用いられる．ただし，そのような単語は抽象度が高いことが多く，より詳しい状況を知るために係り受け関係にある単語も同時に抽出する方法が用いられることがある．しかし，係り受け解析器はテキストの形式や構造に過敏であり，ツイートのような口語で記号類が多い文章や並列構造を正しく解析できないなどの問題が存在する．本論文では，ツイートのような文章に対しても，形態素列にパターンマッチングを適用することで，名詞と係り受け関係にある形態素を抽出し，文章の極性と共にわかりやすく可視化する手法を提案する．実際に，いくつかのトピックに対するメディアの報道に特に強く反応した期間のツイート群を変化点検出で取り出し，提案手法を用いることでユーザの反応の詳細と，時期による違いがわかりやすく表示できるかを検証する．

キーワード ソーシャルメディア，ニュース・マスコミ，テキストマイニング，自然言語処理応用，情報可視化

1 はじめに

Twitter（現 X）などのソーシャルメディアの普及に伴い，ユーザが投稿した大量の投稿の内容を分析する研究が盛んに行われている．内容を知るための簡単な手法として，クラスタリングや LDA などで絞り込んだ後に，出現頻度や TF-IDF，LDA のトピック出現確率などを用いて，重要な単語集合を抽出する手法がよく用いられる．しかし，これらの手法で抽出される単語は抽象度が高い傾向があり，内容を詳しく推測することは難しい．そこで，単語 n-gram や係り受け関係にある単語を表示することで，内容理解を支援する研究が行われている．特に後者の係り受け解析器はテキストの形式や構造に過敏であり，ツイートのような口語で記号類が多い文章を正しく解析できないことも多く，例えば日本語係り受け解析器としてよく用いられる CaboCha [1] には，名詞や文節の並列構造を含む文章を正しく係り受け解析できないという問題が存在する．

本論文では，重要な単語と修飾・被修飾の関係や主語・述語の関係にある係り受け関係に着目して，幾つかの品詞パターンを用意することで，ツイートのような文章に対しても，名詞と係り受け関係にある形態素を抽出し，文章の極性と共にわかりやすく可視化する手法を提案する．

実際には，メディアの報道に対するソーシャルメディアのユーザの反応を分析するため，いくつかのトピックに対するメディアの報道に特に強く反応した期間のツイート群を変化点検出で取り出し，提案手法を用いることでユーザの反応の詳細と，時期による違いがわかりやすく表示できるかを検証する．

2 関連研究

膨大なテキストデータの内容理解を支援するために，単語の組を提示する様々な研究が行われている．

単語 n-gram を用いた研究として，白田らは「シドアルジョ泥流」に関する記事や論文から抽出した 1-gram と 2-gram に LDA を用いてトピック抽出を行った [2]．また Wang らは，n-gram のトピックモデルを提案し，実際に機械学習のトップカンファレンスである NIPS のデータから 1-gram，2-gram のトピックを抽出して，その有効性を示した [3]．

係り受け関係にある単語を用いた研究として，赤尾らは LDA で抽出したトピックの特徴語に加えて，CaboCha で特徴語と係り受け関係を持つ単語を同時に抽出する手法を提案した [4]．なお，事前に係り受け解析されやすいように文章を修正すると共に，CaboCha で並列構造を正しく解析できない問題に対処するために解析結果を修正することで解析誤りを低減していた．また，福田らはコロナウイルスワクチンに対して持つ感情とその感情が現れる要因を分析するため，ツイートテキストから抽出した感情の時系列変化において，バーストが検出された係り受け関係を抽出する手法を提案した [5]．

ただし，これらの研究では，単語の組をランキングするだけであり，上位のほんの一部の結果しか表示できない．久保らは，単語の関係性を示す単語 2-gram に対して LDA を適用し，各トピック上位の 2-gram をトピックグラフとして再構築することで，元の文章で共通して使われる重要な部分とそれらに対する行動や意見を簡潔に表現する手法を提案した [6]．グラフ化す

ることで多くの単語の組を簡潔に表示できるが，一覧性に劣るなどの欠点も存在する．本論文では修飾関係を踏まえて，その関係が存在する量や文書の極性も同時に可視化し，メディアの報道に対するユーザの反応の詳細を簡潔に表現する手法を提案する．

3 提案手法

3.1 提案手法の概要

提案手法は任意のテキスト集合に対して適用可能であり，基本的には，LDA の各トピックの上位の名詞など，任意の名詞集合を指定して，その名詞集合の係り受け関係を分析できる．本論文では，TF-IDF の値が高い名詞集合を，入力に用いた文書集合において特に重要で注目すべきとみなし，実際にメディアのニュース報道に対するソーシャルメディアのユーザの反応を可視化することで，反応の詳細と時期による違いがわかりやすく表示できるかを検証する．

具体的には，井上らの手法 [7] を用いてメディアのニュース報道に対する引用ツイート数の時系列変化に変化点検出を適用し，変化点直後の 3 日間の引用ツイートを特に反応が強く表れている入力データとして，次の手順で処理する．

（1） ユーザの発言部分の抽出
（2） パターンマッチングによる係り受け関係の抽出
（3） 重要単語の抽出
（4） 発言の極性の推定
（5） 係り受け関係の可視化

また，引用ツイートで変化点がなくニュースでのみ変化点があるような場合でも，何らかの事件が起こって報道が活発に行われており，そのような場合でもユーザがどのような感情を抱き，どのような意見をしていたのか分析するため，ニュース記事数の時系列変化に対しても変化点検出を適用し，引用ツイート数の変化点直後 3 日間と重ならない期間の引用ツイートも入力データに加える．

3.2 ユーザの発言部分の抽出

本論文で用いる引用ツイートは，ニュース記事の URL に加えて，ニュースのタイトルやメディア名なども含まれることが多いが，そのような部分はユーザの意見を反映しておらず，また自動的・慣例的に挿入されるために頻度も比較的高く，重要な単語や発言の極性を検出する上でノイズとなることが多い．そこで中北らの手法 [8] を用いて，ユーザの発言部分だけを抽出する．

3.3 パターンマッチングによる係り受け関係の抽出

次に，名詞の詳細を知ることができるように，引用ツイートの発言部分から，名詞と係り受け関係にある形容詞または動詞を抽出する．このような場合は，係り受け解析を用いる方法が一般的であるが，形態素解析よりも入力テキストの形式や構造に過敏であり，特にツイートのような口語で記号類が多い文章はうまく処理できないことも多い．

そこで，提案手法では名詞との係り受け関係だけに注目する

表 1 使用パターンと実例

	使用パターン	実例
1	形容詞／名詞	軽い／症状
2	名詞／助詞／形容詞	感染者／が／多い
3	名詞／助詞／動詞	攻撃／が／始まる
4	名詞／助詞／名詞（サ変接続）／動詞	侵攻／を／非難／する

という特徴を利用して，MeCab で日本語形態素解析して得られる形態素列にパターンマッチングを適用することで，入力データに対して比較的頑健で高速な係り受け関係抽出を実現する．

使用するパターンは，表 1 に示すように，修飾語と被修飾語の 1 パターンと主語と述語の 3 パターンの計 4 通りである．

この結果，上の 2 パターンで名詞と形容詞の組を，下の 2 パターンで名詞と動詞（サ変接続の名詞を含む）の組としてまとめて，ツイートと対応づける．ただし，抽出した形容詞や動詞の中には，ユーザの反応の詳細を知る上で有益ではない単語も含まれるため，4.4.1 節で述べるようにストップワードリストを作成して処理対象から除外する．

3.4 注目する名詞集合の抽出

抽出した係り受け関係から注目すべき重要単語を含んでいるものに限定するため，TF-IDF を適用して，注目する名詞集合を抽出する．

TF-IDF は情報探索やテキストマイニングなどの分野で利用される，文書中に出現した単語がどのくらい特徴的であるかを識別するための指標である [9]．全文書数 D に含まれる全単語数 K のうち，i 番目のある単語 w_i の TF-IDF の値 $TF\text{-}IDF(w_i)$ は，ある文書に単語が出現する度合いを表す TF（Term Frequency）の値 $TF(w_i)$ と，単語が文書全体に出現する度合いの逆数である IDF（Inverse Document Frequency）の値 $IDF(w_i)$ の積で与えられる．本論文では，ツイートを文書として扱う．

$$TF(w_i) = \frac{n(w_i)}{\Sigma_{k=1}^{K} n(w_k)} \tag{1}$$

$$IDF(w_i) = \log \frac{D}{d(w_i)} \tag{2}$$

$$TF\text{-}IDF(w_i) = TF(w_i) \times IDF(w_i) \tag{3}$$

ここで $n(w_i)$ は単語 w_i の出現頻度を表し，$d(w_i)$ は単語 w_i を含む文書数を表す．

実際には，ツイートの発言部分を MeCab で日本語形態素解析し，一般・自立・固有名詞・サ変接続・形容動詞語幹である名詞だけを抽出し，TF-IDF の値が上位 20 件の名詞集合を抽出する．

3.5 発言の極性の推定

さらに，ユーザがどのような感情を持ってニュース報道に言及したかに関する情報も利用するために，引用ツイートから極性を抽出する．

極性の抽出には，Nanba らの T5 を使った感情分析器 [10] を使用した．これは約 100GB の日本語コーパスで事前学習した T5 モデルを，梶原らの主観と客観の感情分析データセット

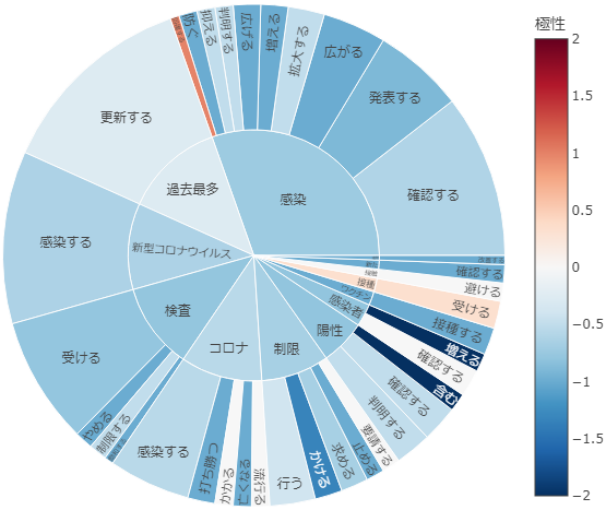


図 1　二重ドーナツグラフ可視化の例

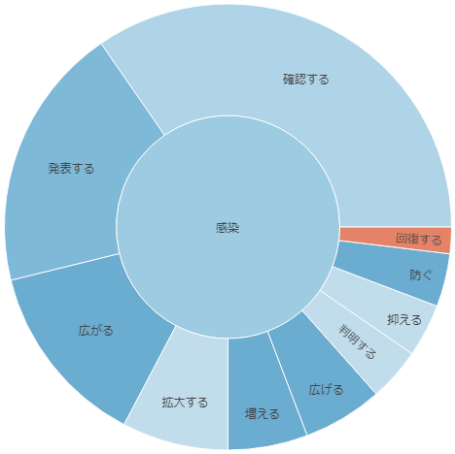


図 2 「感染」に注目して拡大表示した例

WRIME [11][1]を用いてファインチューニングした感情分析器である．WRIME は，日本語の感情分析の研究のために，SNS 上の投稿に対して，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．このような深層学習ベースのアプローチは，学習用データセットがあれば比較的容易に感情分析器を構築でき，検出率も比較的高い利点がある．

実際には，ツイートの発言部分に感情分類器を適用し，その極性の値（-2∼2）を抽出して，引用ツイートと対応づける．

3.6　可　視　化

本論文では，得られた係り受け関係をそのままランキングするのではなく，名詞と係り受け関係にある形容詞や動詞の関係を，二重ドーナツグラフを用いて可視化する．なお，形容詞は名詞の性質や状態を表し，動詞は名詞の動作や作用を表すと考えて，実際には表 1 の各パターンでそれぞれ出現頻度上位 N 件の係り受け関係を求め，パターン 1 とパターン 2 を形容詞に，パターン 3 とパターン 4 を動詞にまとめて可視化する．動詞に関して，$N = 20$ で二重ドーナツグラフとして可視化した例を，図 1 に示す．内側のドーナツグラフには注目する名詞を，外側のドーナツグラフにはそれと係り受け関係にある形容詞または動詞を，それらの出現頻度に応じた面積で表示する．

これだけでは単語の区別が難しいことと，表示された単語がどのような意図で用いられていたかを示すために，極性の値を用いて着色する．まず，各形容詞や動詞に関して，それらが出現した引用ツイートの極性値の平均を計算する．次に，名詞に関して，それらと関係がある形容詞または動詞の値の重み付き平均を計算する．最後に，形容詞・動詞および名詞を，ポジティブ（2）からネガティブ（-2）まで赤から青に変化するように着色する．

本可視化システムは，Python と Web アプリケーションのフレームワークである Streamlit を用いて Web ブラウザ上で対話的に操作できるシステムとして実装した．トピックと対象期間を指定すれば図 1 のような二重ドーナツグラフ可視化が可能であり，さらに内側の名詞から，詳しく調べたい単語をクリックすることで図 2 に示すように，その名詞だけに限定して拡大表示することもできる．

4　評価・分析

4.1　データセット

ニュースに引用したツイートデータセットとして，Twitter API で取得した 2022 年 1 月 1 日から 2022 年 11 月 30 日までのツイートのうち，ニュース記事の URL が含まれる 4,491,671 件のツイートを利用した．

また，ニュースデータセットとして，ニュースポータルサイト Ceek.jp News[2]で扱われているニュース記事のうち，その URL がツイートデータセット内のツイートに含まれているニュース記事 3,466,472 件を利用した．

4.2　分析対象トピック

データセットの期間内における，引用数上位 10 件のニュース記事から，時系列的な特性が異なる以下の 2 個のトピックを人手で選定して，それぞれ「コロナウイルス」，「ロシア or ウクライナ」を Elasticsearch のクエリとして検索して，ニュース記事と，さらにそれらのニュースを引用しているツイートを抽出した．

（1）　コロナウイルス

2019 年に発生してから世界中で感染が確認されているコロナウイルスに関する報道．感染者が急激に増えた時期として，2022 年 2 月ごろをピークとする第 6 波，7 月ごろをピークとする第 7 波がデータセットの期間内に存在している．

（2）　ロシアによるウクライナ侵攻

2022 年 2 月 24 日にロシアが開始したウクライナへの軍事侵攻に関する報道．一週間前の 2022 年 2 月 17 日ごろには米当局の情報で侵攻の可能性があることが知らされており，侵攻前

1：https://github.com/ids-cv/wrime

2：http://news.ceek.jp/

表 2　CaboCha とパターンマッチング法の対応

	CaboCha	パターンマッチング法
1	形容詞／名詞	名詞／助詞／形容詞
2	名詞／形容詞	名詞／助詞／形容詞
3	名詞／動詞	名詞／助詞／動詞 名詞／助詞／名詞（サ変接続）／動詞

表 3　係り受け抽出性能

	適合率	再現率	F 値
パターンマッチング法	0.891	0.475	0.619
CaboCha	0.738	0.870	0.799

日の 23 日にはウクライナ全土に非常事態宣言が出された．

4.3　係り受け抽出性能の評価

パターンマッチング法はツイートのような厳密に解析しにくいテキストに対する頑健性と処理速度を重視している．しかし，当然離れた係り受け関係は検出されないため，ある程度の検出漏れがあるはずである．そこで係り受け解析器である CaboCha と比較し，提案手法がどの程度係り受け関係を抽出できているかを評価した．

評価用テキストデータとしては，トピック「コロナウイルス」の 1 つ目の変化点である 1 月 5 日直後 3 日間に報道されたニュース記事の文字数が 30 文字以上の引用ツイートからランダムに 100 件を抽出し，抽出率を向上させるための前処理・後処理は行わずに，URL とハッシュタグだけを除去して使用した．正解データは，人手で表 1 に示す品詞の係り受けパターンを判定して作成した．なお，評価の際の CaboCha とパターンマッチング法で対象とする係り受け関係の対応を表 2 に示す．

まず，評価用テキストデータを CaboCha とパターンマッチング法で 100 回処理した時間を 10 回計測して平均値を求めた．パターンマッチング法は 9.71 秒に対して CaboCha は 44.03 秒で，22%の時間で処理できることがわかった．

パターンマッチング法と CaboCha の解析結果を適合率，再現率，F 値で評価した結果を表 3 に示す．この結果を見ると，CaboCha の方が再現率，F 値共に高く，予想通りパターンマッチング法は検出漏れがあった．しかし，適合率に関してはパターンマッチング法の方が高かった．この原因を調べると，ツイートでは助詞や句点の脱落が頻繁に出現し，そのような場合は CaboCha は解析誤りを引き起こすが，パターンマッチング法ではそもそも抽出対象にはならないからであった．

これらの結果から，パターンマッチング法は，ツイートのテキスト分析のように，大量のテキストデータを対象に統計処理を行う場合のように，多少の検出漏れは構わないが，解析誤りによるノイズは低減させたい用途には向いていると思われる．

4.4　ソーシャルメディアユーザの反応の分析

次に，2 つのトピックに対してソーシャルメディアユーザの反応を提案手法を用いて可視化し，その結果を分析した．また，可視化する出現頻度上位の係り受け関係の件数は $N = 20$ とした．

表 4　ストップワードリスト

形容詞	ない，やすい
動詞	する，いる，れる，なる，ある，てる，できる（出来る），思う，られる，いう（言う），くる，せる，見る，やる，出る，考える，知る

4.4.1　ストップワードの選定

ツイートデータセットを MeCab で形態素解析して，形容詞と動詞の出現頻度の上位 20 件の単語を目視で確認し，そのうち内容を補足するための情報量が少なく有益とは思わない単語をストップワードとした．ストップワードリストを表 4 に示す．

4.4.2　トピック「コロナウイルス」

ニュース記事数と引用ツイート数の時系列変化から変化点検出を行った結果，トピック（1）ではニュース側で 1 月 4 日と 7 月 22 日に，ツイート側で 1 月 5 日に変化点を検出した．そこで引用ツイート数の変化点である 1 月 5 日に加えて，ニュース記事数でのみ検出した 7 月 22 日を分析対象とした．提案手法を適用した結果を図 3 に示す．左側は形容詞の可視化結果，右側は動詞の可視化結果である．

図 3(a) は，1 つ目の変化点である 1 月 5 日直後の可視化結果である．名詞・形容詞では，第 6 波の感染拡大時期であったことから，「接種・酷い」や「接種・重い」などワクチン接種後の副作用を示唆する表現や「米軍・甘い」や「米軍・悪い」など米軍基地におけるクラスターの発生を示唆する表現が多かった．また，「オミクロン・強い」などオミクロン株の感染が注目されている時期でもあったと思われる．名詞・動詞では，「感染・確認する」や「新型コロナウイルス・感染する」などコロナウイルスの感染を示唆する表現が多く抽出された．極性に関してはどちらも負の極性を持つ文章がほとんどであり，正の極性を持つ単語も一部あるが，よく見ると必ずしも正とは言えなかった．

図 3(b) は，2 つ目の変化点である 7 月 22 日直後の可視化結果である．名詞・形容詞では，第 7 波の感染拡大時期を反映して，「コロナ・憎い」や「コロナ・恐ろしい」など，コロナウイルスに対する感情を示唆する表現や，「症状・軽い」や「症状・重い」など感染後の症状を示唆する表現が多かった．特に後者に関しては，「重い」，「しんどい」に対して，「軽い」，「無い」という逆の意味の単語が同時に抽出されており，提案手法ではそれらの割合をある程度推定することも可能だと思われる．名詞・動詞では，1 つ目の変化点と同様に「感染・確認する」など，コロナウイルスの感染に関する表現が多いが，「感染・拡大する」や「過去最多・更新する」，「感染者・増える」など，感染者数の増加に関するさまざまな表現が使われている．極性に関しては，この期間もネガティブなものが多い．

4.4.3　トピック「ロシアによるウクライナ侵攻」

トピック（2）ではニュース側で 2 月 18 日と 2 月 24 日に，ツイート側で 2 月 23 日に変化点を検出した．そこで 2 月 23 日に加えて 2 月 18 日を分析対象とした．提案手法を適用した結果を図 4 に示す．左側は形容詞の可視化結果，右側は動詞の可視化結果である．

図 4(a) は，1 つ目の変化点である 2 月 18 日直後の可視化結

果である．

名詞・形容詞では，ロシアのウクライナ侵攻の直前なので，「批判・強い」や「侵攻・無い」のような戦争を否定する表現や，「ロシア・怖い」，「プーチン・強い」，「ウクライナ・弱い」などのそれぞれの情勢を表す表現が同時に抽出されている．名詞・動詞では，「ロシア・侵攻する」や「攻撃・受ける」，「侵攻・決断する」などロシアによるウクライナ侵攻を示唆する表現が多かった．極性に関しては，どちらもネガティブなものが多いが，「外交・粘り強い」など一縷の希望を示唆する表現も抽出されている．

図 4(b) は，2 つ目の変化点である 2 月 23 日直後の可視化結果である．名詞・形容詞では，開戦直後ということもあり，制裁という名詞が多く使われ，それに対して「厳しい」，「強い」，「甘い」，「軽い」，「重い」などの対極の意味の形容詞が組み合わされた表現が抽出され，それらの割合から厳しい制裁を望むユーザが多かったことが推測できる．名詞・動詞では「戦争・始まる」や「侵攻・受ける」，「攻撃・受ける」などのロシアの侵攻開始を示唆する表現の出現に加えて，「戦争・反対する」，「ロシア・非難する」などロシアを批判する表現も多かった．極性に関しては，負の極性を持つ表現がほとんどである．なお，「良い・戦争」という表現が正の極性として抽出されているが，ツイートデータを調べたところ，確かにそのような用語が使用されていたが，それは否定的な文脈，または否定表現を伴っていて，提案手法では極性が反転する場合には対応できていないことがわかった．

4.5 考　　察

提案手法の妥当性について考察する．

まず，提案手法は実際に名詞と係り受け関係にある形容詞や動詞を単純にリスト表示すると最大で 80 行になる内容を，コンパクトに表示できることがわかった．全体の可視化結果では小さく表示されてしまう場合もあるが，特定の単語に限定すれば詳細に見ることができるために，全体的な俯瞰と詳細な分析を両立できていた．

また，今回は同一トピックで 2 つの時点を分析対象としたが，提案手法ではそれらの違いがわかりやすく，また名詞も形容詞や動詞と組み合わせることでその文脈を推定したり，同じ意味を持つ組み合わせの判断が容易になることがわかった．

さらに，同一の名詞に対して意味が反対の形容詞が同時に抽出される現象も確認でき，その場合はそれらの割合から意見の多様性と多数派の判断ができる可能性も示された．

極性に関しては，ポジティブな単語とネガティブな単語の区別が容易であったが，一部に妥当と思われない極性もあった．これは提案手法は，それらの係り受けが出現したツイート全体の極性を集計して用いているために齟齬が生じることがあることと，否定的な文脈または否定表現が伴っていても，それを反映できていないことがわかった．なお，今回用いたトピックの可視化結果ではほとんどがネガティブなものばかりであった．これが選択したトピックによるものか，ソーシャルメディアではネガティブな発言が多い傾向にあるのか，システム的なバイアスなのかは不明であり，今後多彩なトピックに関して分析を行う必要がある．

5 おわりに

本論文では，いくつかのトピックに関するメディアのニュース記事と，それに言及したツイートデータに対して提案手法を用いて，メディアの報道に特に強く反応した期間のツイート群を変化点検出で取り出し，ユーザの反応の詳細と，時期による違いを分析した．その結果，提案手法では多くの情報をコンパクトにわかりやすく表示できること，同一トピックでも異なる時期の違いが判別しやすいこと，ある事柄に対する多彩な意見と多数派・少数派の識別ができる可能性が示された．ただし，ツイートの極性を統計処理して単語に適用する手法に関しては，一部の単語で正しい極性が得られない問題も確認した．

今後の課題としては，より多くのトピックで分析を行うことで，トピック単位ではなく，ソーシャルメディア全体の性質や傾向について調査すると共に，可視化システムについて，抽出元のツイートテキストも表示することで，1 つの単語に着目した時により詳細な分析ができるように改良し，可視化結果の評価を行う．

謝　　辞

本研究は JSPS 科研費 21H03557 の助成を受けた．

文　　献

[1] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. 情報処理学会論文誌, Vol. 42, No. 6, pp. 1834–1842, 2002.

[2] Yussanti Nur Fajrina, Yukari Shirota, and Riri Fitri Sari. Topic extraction analysis for sidoardjo mudflow impacts. 学習院大学 経済論集, Vol. 53, No. 3, pp. 101–114, 2016.

[3] Xuerui Wang, Andrew MSCallum, and Xing Wei. Topical N-grams: Phrase and topic discovery, with an application to information retrieval. In *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM '07)*, pp. 697–702, 2007.

[4] 赤尾貴仁, 風間一洋. 係り受け関係を用いた LDA のトピックの内容理解の支援. WebDB 夏のワークショップ 2022, B3-1, 2022.

[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.

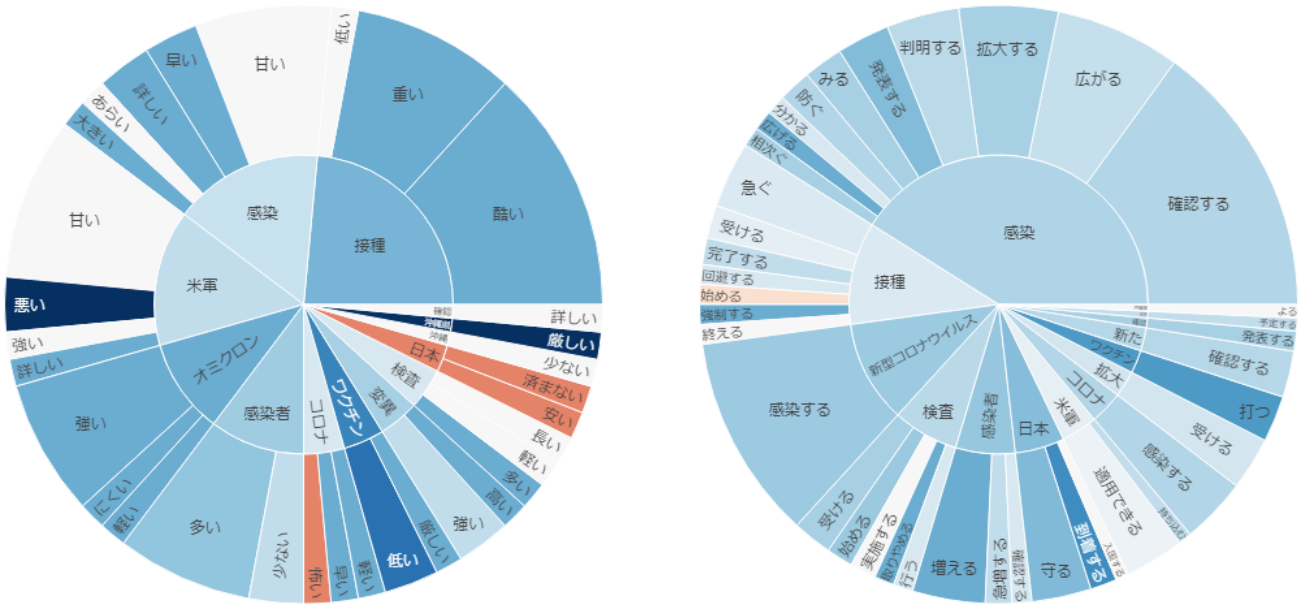
[6] 久保侑哉, 風間一洋, 榊剛史. 単語 2-gram に基づくマルチトピックグラフの可視化. 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2018), B1-5, 2018.

[7] 井上大成, 中北雄大, 風間一洋, 吉田光男, 土方嘉徳. メディアの報道とソーシャルメディアの反応の時系列分析. 第 15 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 2023), 5b-5-1, 2023.

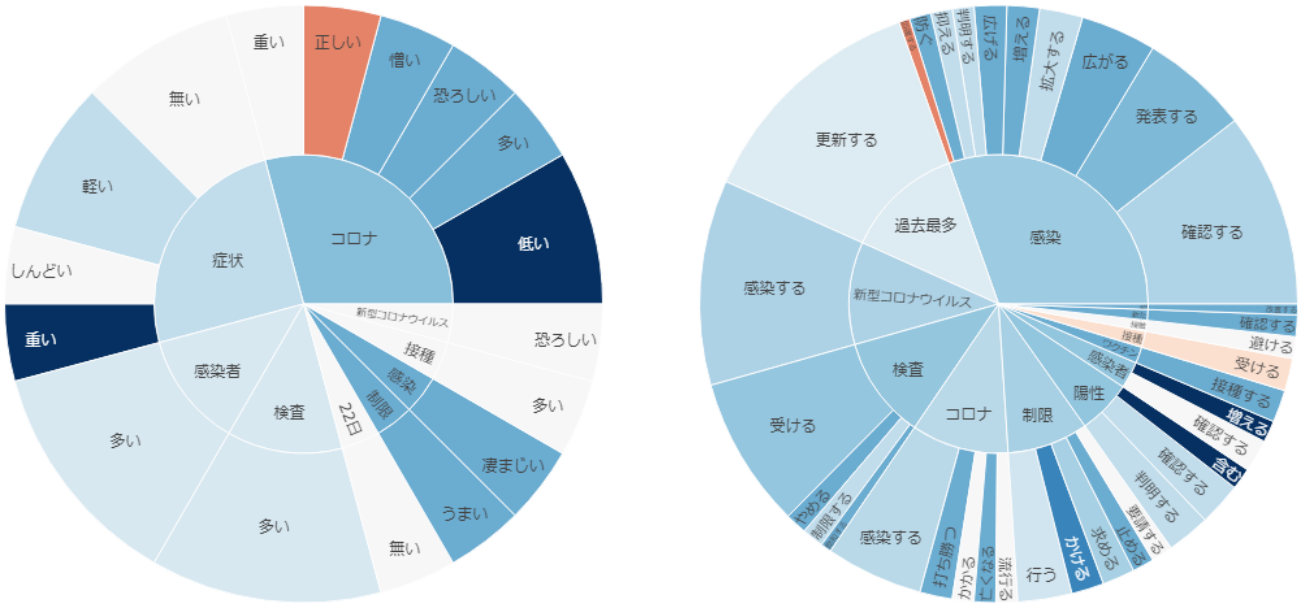
[8] 中北雄大, 風間一洋, 吉田光男, 土方嘉徳. 感情とトピックに注目したメディアの報道姿勢の分析. 第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2022), B24-2, 2022.

[9] Karen Spärck Jones. A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation*, Vol. 28, No. 1, pp. 11–21, 1972.

[10] Hidetsugu Nanba, Katsuya Yamamoto, Satoshi Fukuda, Hi-



(a) 1 月 5 日〜 1 月 7 日

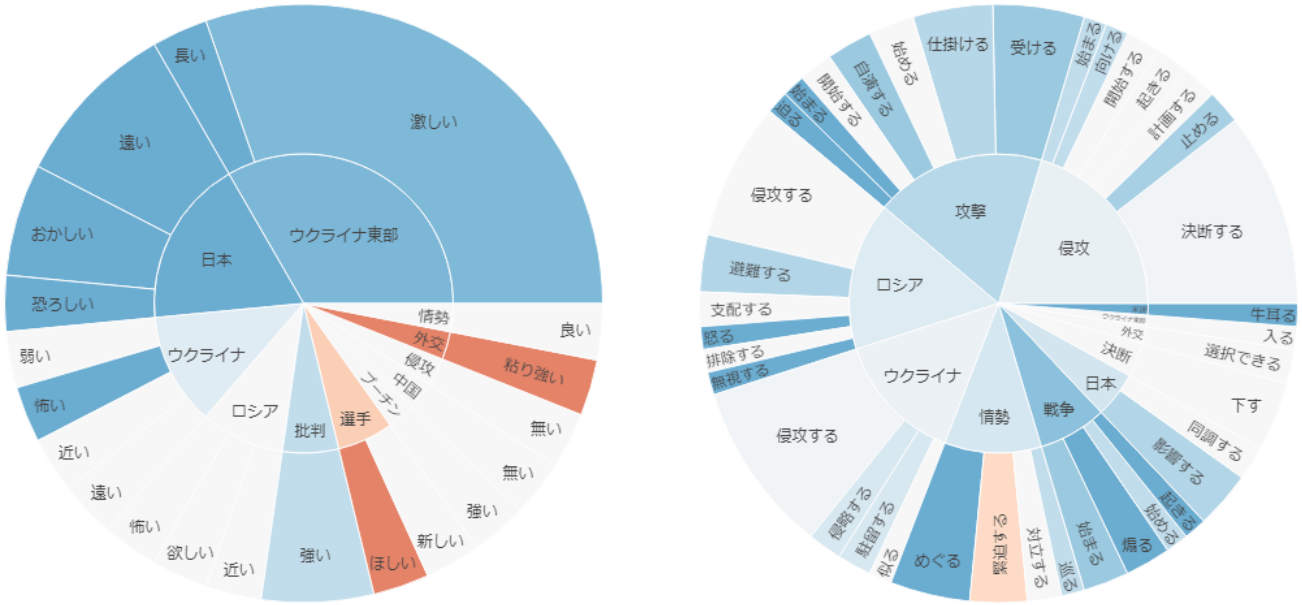


(b) 7 月 22 日〜 7 月 24 日

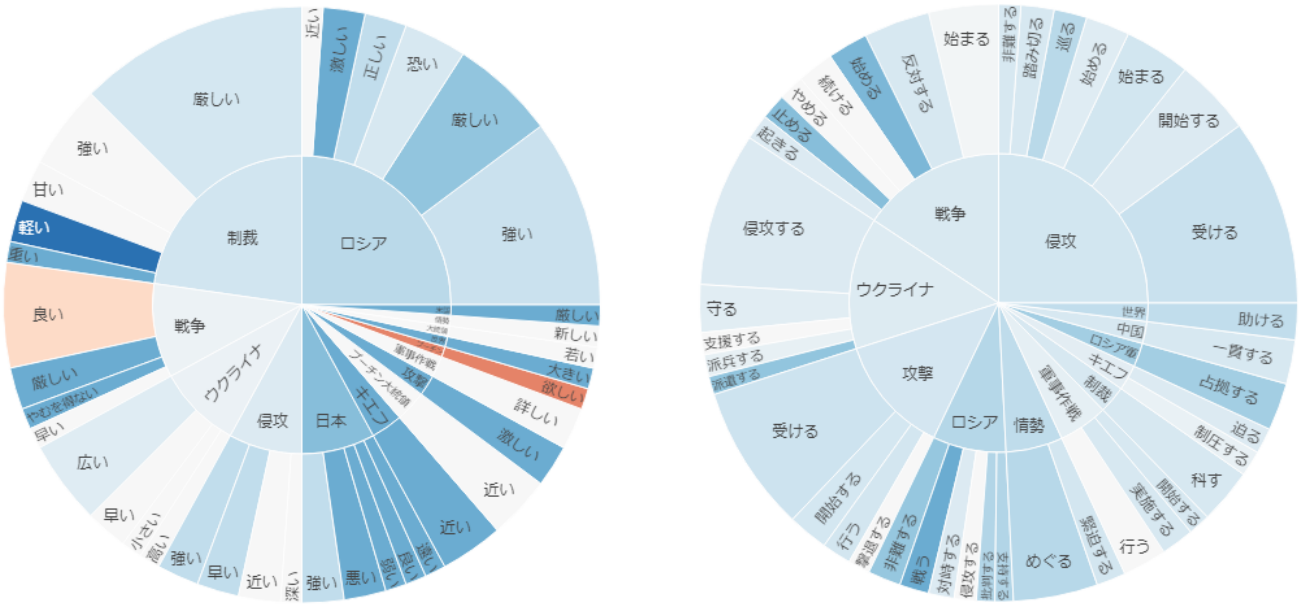
図 3　トピック「コロナウイルス」の可視化結果

roko 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.

[11] Tomoyuki Kajiwara, 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.



(a) 2 月 18 日〜 2 月 20 日



(b) 2 月 23 日〜 2 月 25 日

図 4　トピック「ロシアによるウクライナ侵攻」の可視化結果