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

井上　大成† 　中北　雄大†† 　風間　一洋† 　吉田　光男††† 　土方　嘉徳††††

† 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
†† 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
††† 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,s236194}@wakayama-u.ac.jp, ††kazama@ingrid.org, †††mitsuo@gssm.otsuka.tsukuba.ac.jp, ††††contact@soc-research.org

あらまし　近年，ソーシャルメディア上でフェイクニュースが拡散されることでアメリカ大統領選挙の結果が左右されるなど，メディア報道とソーシャルメディアに関連する様々な問題が指摘されている．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を時系列的に関連づけながら分析する手法を提案し，ソーシャルメディア上のユーザの活動が，メディアの報道にどの程度影響されているかを検証する．実際には，各メディアが報道したニュースと，Twitter でニュースに言及したツイートのデータを取得し，いくつかのトピックに関するニュースとツイートの時系列変化をグラフにして可視化し，またニュースとツイートの時系列変化の相互相関を求めることで，メディア側の報道とユーザ側の反応に相関が存在しているかどうかを分析する．さらに，ニュースとツイートのそれぞれに変化点検出手法である ChangeFinder を用いて変化点スコアを算出し，さらに検出した変化点に基づいてニューステキストのキーフレーズを抽出した結果を可視化することで，ニュースとツイートの数の増減が，分析するトピックで起こったイベントと対応しており，検出した変化点でそのトピックからいくつかのイベントを抽出することが可能かを検証する．

キーワード　ソーシャルメディア，ニュース・マスコミ，空間・時空間・時系列データ処理

1 はじめに

Twitter のようなソーシャルメディアの普及に伴い，マスメディアにおいても公式アカウントによる情報発信やニュース記事の意見をソーシャルメディアに容易に投稿できるボタンの設置，ソーシャルメディア上の意見収集・取材などの手段で積極的に相互交流するようになった．

ただし，身近な場所で発生した自然災害や事故，事件に関しては，ソーシャルメディアのユーザによる一次情報の発信が可能であるが，経済や政治，国際情勢などの分野に関する一次情報を発信できるのは取材活動が可能なマスメディアやネットニュースなどのオルタネティブメディアに限られ，それ以外のユーザは報道されたニュースを読み，重要だと判断した場合にさらに情報拡散したり，その内容について議論することしかできない．つまり，これらの分野においては，ソーシャルメディア上のユーザの意見は，マスメディアやオルタネティブメディアの報道内容に強く影響されることになる．

さらに，この影響力を用いて，偏った視点からの報道やフェイクニュースの拡散，政府や政治家へのメディアの忖度によりニュース内容を操作することで，ユーザが真実を知る機会が奪われたり，実世界の予期せぬ深刻な対立を招くことがある．例えば，ソーシャルメディア上で意図的にフェイクニュースを拡散することでアメリカ大統領選挙の結果が左右されたり，フェイクニュースが集団リンチや殺人のような重大な事件につながった事例も存在する [1]．

本論文では，現実の事件に関するニュース報道によって，ソーシャルメディア上のユーザの意見や感情が変化すると仮定して，逆にそのユーザの行動から各メディアの報道内容の影響力や，報道姿勢の違いを推定することを目的とし，ニュースとその言及ツイートの盛り上がりにおける時間的な因果関係の有無と，その内容の違いを分析する．

2 関連研究

2.1 メディアの報道やソーシャルメディアの投稿に関する研究

メディアはその立場や主義主張から偏った報道をすることがあり，そのような報道姿勢について分析する研究が行われている．張らは，ニュース記事の書き方を「印象」という評価指標で分析し，ニュースサイトの報道傾向を視覚的に分析できる手法を提案した [2]．田中らは，公共事業に対する大手新聞社の批判的な記事について，いくつかのキーワードの掲載回数の経年変化を調べることで，その報道姿勢を分析した [3]．

また，ソーシャルメディアの投稿に注目した研究も行われている．鳥海らは新型コロナウイルス感染症が社会にどのような影響を与えたのかを明らかにするため，Twitter のデータからユーザの偏りや感情を推定し，その変化を分析した [4]．また，木村らは，東日本大震災時において，福島原発事故に関連する

情報が，テレビ報道と Twitter のそれぞれでどのように扱われたのかを分析した [5].

以上のようにメディアの報道やソーシャルメディアの投稿に関する研究は多く行われているが，報道機関またはソーシャルメディア上のユーザの一方だけに注目したものが多い．本論文では，メディアの報道とソーシャルメディアのユーザの反応の間に関係性があると仮定して，これらを関連付けて時系列的に分析する．

2.2 時系列データの可視化に関する研究

時系列分析を行う研究は様々な分野で行われており，分析を行う上でデータを可視化することは，データの特徴や性質を把握するために非常に有用な方法であるといえる．上田らはアポロ計画の会話通信記録を用いて，このような長期間に大量の発話が記録されたデータから，会話の概要などを視覚的に分析できるシステムを提案した [6]．この手法では，アポロ計画のミッション全体を表示しながら，その一部を拡大表示させたものも同時に表示することで，マクロな視点とミクロな視点を組み合わせたマルチスケールな分析が可能となっている．

本論文では，一般的な時系列分析事例における，時系列データの可視化手法にも着目し，ニュース記事とその言及ツイートを同時に可視化して，この二つを関連付けながら分析できる手法を提案する．

3 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手法

3.1 メディアの報道とソーシャルメディアの反応

ソーシャルメディアやメディア報道に関する研究では，政治や経済などのトピックが分析対象となることが多いが，その他のトピックでもソーシャルメディアのユーザの活動がメディアの報道に大きく影響されることは十分に考えられる．そこで任意のトピックを指定して，ユーザとメディア報道を時系列的に関連付けながら分析できることが望ましい．

そこで本論文では，メディアの報道とソーシャルメディアの反応を，以下の手順で時系列分析する．この分析手法の概要を図 1 に示す.

手順 1　ニュース記事データとその言及ツイートデータの収集
手順 2　トピックに関連するニュース記事と言及ツイートへの絞り込み
手順 3　ニュース記事と言及ツイートの時系列分析

3.2 ニュース記事データとその言及ツイートデータの収集（手順 1）

メディアが報道したニュース記事とそれに言及したツイートを関連付けて分析するために，まずニュース記事データとその言及ツイートデータを収集する．具体的には，複数のメディアを観測対象として設定し，ニュース記事のリンクを含んだツイート・リツイートを継続的に Twitter API で収集し，さらにリンク先のニュース記事の作成日時やテキストデータも収集する．収集したデータは全文検索可能なデータベースに保存する．データベースは Elastic 社が開発しているオープンソース検索エンジンである Elasticsearch[1]を用いる．

3.3 トピックに関連するニュース記事と言及ツイートへの絞り込み（手順 2）

さらに，キーフレーズ共有性 [7] に基づいてトピックに関連するニュース記事と言及ツイートだけに絞り込む．キーフレーズとは文書中で重要となる文字列のことであり，通常は 2 単語以上の句であることが多い．キーフレーズ共有性とは，特定のトピックに関連するニュース記事が，同一のキーフレーズを共有している性質のことであり，文書の内容の類似性を用いるよりも適切に関連記事を抽出できる．

実際には，分析するトピックごとに人手で検索クエリを設定して，Elasticsearch で検索してトピックに関連するニュース記事に絞り込んでから，その記事に言及したツイートを抽出する．

3.4 ニュース記事と言及ツイートの時系列分析（手順 3）

トピックで絞り込んだニュース記事と言及ツイートに対して，以下の 4 つの手法を組み合わせて時系列分析する．

3.4.1 ニュース記事数と言及ツイート数の時系列変化の可視化

設定したトピックの各メディアのニュース記事と言及ツイートの時系列変化を可視化する．

実際には，トピックに関連するニュース記事と言及ツイートの出現回数を 1 日ごとにカウントして，ニュース数と言及ツイート数を縦軸に，その年月日を横軸にして，日付で対応づけた二つの折れ線グラフとしてプロットする．

3.4.2 ニュース記事数と言及ツイート数の相互相関の算出

ニュース記事数と言及ツイート数の時系列データの特徴を，相互相関関数を用いて分析する．

一つの時系列データに T 個のデータがあり，二つの時系列データ X，Y の t 番目のデータをそれぞれ $X(t)$，$Y(t)$，その位相ずれを Δt とした時の相互相関関数は以下の式で表される．

$$C_{XY}(\Delta t) = \sum_{t=1}^{T} X(t)Y(t+\Delta t) \tag{1}$$

分析の際は，ニュース数の時系列データとツイート数の時系列データに対し，その位相ずれ Δt を 1 日単位で増加させながら相互相関を求める．これにより，二つの時系列データの関係や周期性が存在しているかどうかを確認できる．

3.4.3 ニュース記事数と言及ツイート数の変化点検出

実世界で何らかの事件が発生した場合，さらにその事件に関する新しい事実がわかるなど状況が変化した場合には複数のメディアがニュース記事として報道し，その内容に応じて Twitter ユーザがさまざまなツイートをすると仮定して，山西らの ChangeFinder [8] を用いて，ニュース記事数または言及ツイート数の時系列変化の変化点を検出する．ChangeFinder は，AR モデル（自己回帰モデル）に 2 段階学習と忘却機能を加えた SDAR アルゴリズムを使用して，時系列データの変化点を

1：https://www.elastic.co

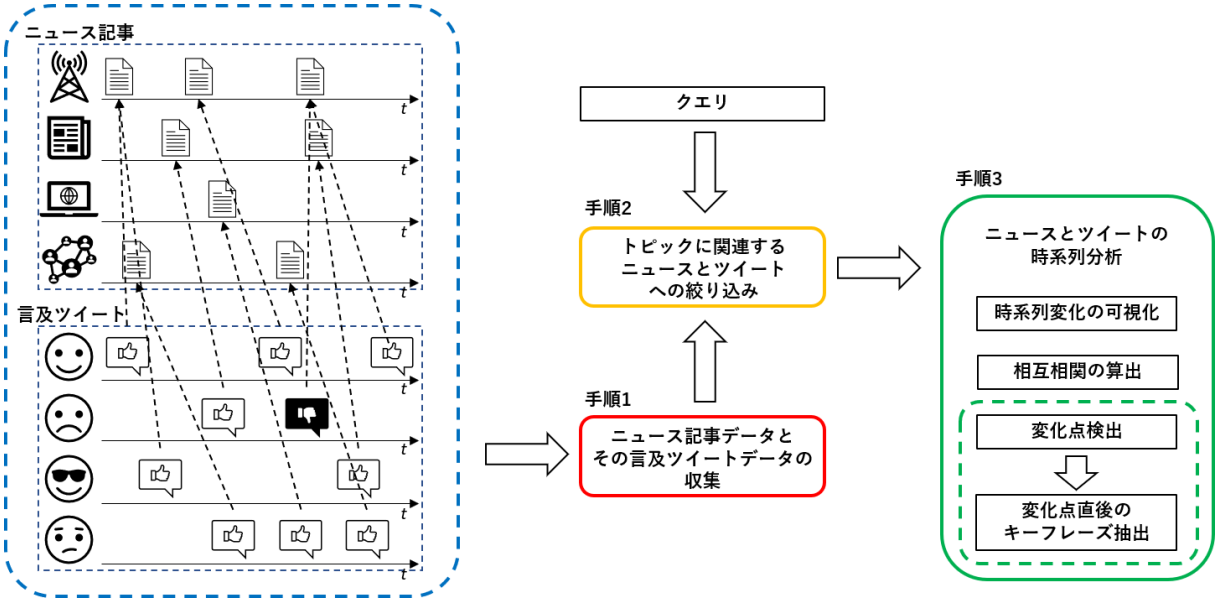


図 1　分析手法の概要図

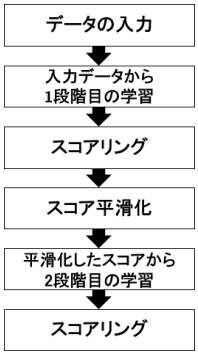


図 2　ChangeFinder の処理手順 [9]

検出する手法である．時系列データの定常性を前提とする AR モデルでは変化点を適切に検出できない時系列データに対しても，ChangeFinder では古いデータを忘却することで，変化点を検出できる．

ChangeFinder の処理手順を図 2 に示す．まず，連続値をとる時系列変数を $\{z_t : t = 1, 2, ...\}$ とし，参照する過去のデータの数を表す次数を k としたときの AR モデルは以下のように表される．

$$z_t = \sum_{i=1}^{k} \omega_i z_{t-i} + \varepsilon \tag{2}$$

ここで，$\omega_i \in R^{d\times d}(i = 1, ..., k)$ は d 次のパラメータ行列であり，$z_{t-k}^{t-1} = (z_{t-1}, z_{t-2}, ..., z_{t-k})^T \in R^{d\times k}$ と記す．ε は平均 0，共分散行列 Σ のガウス分布 $\mathcal{N}(0, \Sigma)$ に従うガウス変数である．入力される時系列データ $x_t(t = 1, 2, ...)$ を $x_t = z_t + \mu$ とすると，AR モデルの確率密度関数は以下の式で表される．

$$p(x_t \mid x_{t-k}^{t-1} : \theta) = \frac{\exp(-\frac{1}{2}(x_t - w)^T \Sigma^{-1}(x_t - w))}{(2\pi)^{d/2}|\Sigma|^{1/2}} \tag{3}$$

ここで $w = \Sigma_{i=1}^{k} \omega_i (x_{t-i} - \mu)$ であり，モデルに関するパラメータを $\theta = (\omega_1, ..., \omega_k, \mu, \Sigma)$ とする．1 段階目の確率密度学習で得られた確率密度関数の時系列データを $p_t(x)$ とすると，時刻 t のデータ x_t の外れ値スコアを以下の式で示すように対数損失で計算する．

$$Score(x_t) = -\log p_{t-1}(x_t) \tag{4}$$

さらに，スコアを平滑化するために，以下の式に示すように，幅 T の窓を移動させながら移動平均を計算する．

$$y_t = \frac{1}{T} \sum_{i=t-T+1}^{t} Score(x_i) \tag{5}$$

ここで得られた新しい時系列データ y_t に対して，2 段階目の学習として AR モデルを再び適用して得られた確率密度関数の時系列データ q_t を用いて，T' を与えられた正の整数としたとき，時刻 t における T' 平均スコアを以下で示すように対数損失で計算する．

$$Score(t) = \frac{1}{T'} \sum_{i=t-T'+1}^{t} (-\log q_{i-1}(y_i)) \tag{6}$$

これにより得られたスコアを変化点スコアとする．このように平滑化によって外れ値を取り除いてから再度学習することで，精度の高い変化点の検知を可能にする．

分析の際は，あるトピックのニュース数と言及ツイートの時系列データそれぞれに対して ChangeFinder で変化点スコアを算出し，変化点スコアの平均値の 2 倍を閾値とし，閾値を超えた日を変化点とする．

3.4.4　変化点直後のキーフレーズ抽出

得られた変化点の直後 3 日間に報道されたニュース記事からキーフレーズを抽出して，その変化点が生じた理由を分析する．

キーフレーズ抽出は Liang [10] らの手法を用いる．この手法では文書中の大局的な文脈と局所的な文脈を組み合わせてモデル化することで，文書間の長さが大きく異なるデータに対しても適切なキーフレーズの抽出が可能となっている．

文書 D を日本語形態素解析して得られるトークンを $t_1, t_2, ..., t_N$ とし，0 または一つ以上の形容詞に修飾された名詞または複合名詞（ただし，代名詞と非自立名詞は除く）をキーフレーズ候補 $KP_1, KP_2, ...KP_N$ とする．さらにトークンのベクトル表現を $H_1, H_2, ..., H_N = \mathrm{BERT}(t_1, t_2, ...t_N)$ とし，BERT [11] で事前学習した言語モデルを用いて，候補フレーズと文書全体のベクトル表現を抽出すると，候補フレーズのベクトル表現はフレーズに含まれるトークンのベクトル表現の平均 H_{KP_i} となり，文書全体のベクトル表現は以下の式で表される．

$$H_D = \mathrm{Maxpooling}(H_1, H_2, ..., H_N) \tag{7}$$

キーフレーズ候補の大局的な適合性を表す, キーフレーズ候補 KP_i と文書 D の適合度 (Phrase-Document Similarity) は，次式で計算する．

$$\mathrm{R}(H_{KP_i}) = \frac{1}{\|H_D - H_{KP_i}\|_1} \tag{8}$$

次に，キーフレーズ候補の局所的な顕著性を表す中心性 (Boundary-Aware Centrality) を求める．まず，文書をキーフレーズ候補をノードとするグラフ $G = (H_{KP_i}, e_{ij})$ として表す．また i をキーフレーズ候補の位置，n をキーフレーズ候補数として，Boundary Function を $d_b(i) = min(i, \alpha(n-i))$, 閾値を $\theta = \beta(max(e_{ij} - min(e_{ij})))$ として，ノード i の中心性を次式で計算する．また，α，β，γ はハイパーパラメータである．

$$\begin{aligned}\mathrm{C}(H_{KP_i}) &= \textstyle\sum_{d_b(i)<d_b(j)} \max(e_{ij} - \theta, 0) \\ &+\lambda \textstyle\sum_{d_b(i)\geqq d_b(j)} \max(e_{ij} - \theta, 0)\end{aligned} \tag{9}$$

さらに，文書の最初と最後に重要な情報があるとと仮定して，キーフレーズの文書中の出現位置に応じて重みを設定するために，p_1 を各キーフレーズ候補が出現した最初の位置として位置バイアスの重み $p(KP_i) = 1/p_1$ を求めて，次式のように Softmax 関数で正規化する．

$$\hat{p}(KP_i) = \frac{\exp(p(KP_i))}{\sum_{k=1}^{n} \exp(p(KP_i))} \tag{10}$$

ノード i の中心性 Boundary-Aware Centrality は，次式で求める．

$$\hat{\mathrm{C}}(H_{KP_i}) = \hat{p}(KP_i) \cdot \mathrm{C}(H_{KP_i}) \tag{11}$$

最後に，大局的な適合度 $\mathrm{R}(H_{KP_i})$ と局所的な顕著性 $\hat{\mathrm{C}}(H_{KP_i})$ を掛けてスコアを計算し，上位 k 個のキーフレーズ候補をキーフレーズとして抽出する．

$$\mathrm{S}(H_{KP_i}) = \mathrm{R}(H_{KP_i}) \cdot \hat{\mathrm{C}}(H_{KP_i}) \tag{12}$$

4 分　　析

4.1 データセット

ニュースに言及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 9 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 3,531,139 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，統合型メタ検索エンジン Ceek.jp News [12] 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 2,787,416 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，言及数上位 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 4 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」,「円安」,「上島竜兵」,「ロシア or ウクライナ」を Elasticsearch のクエリとして検索して，ニュース記事を抽出し，さらにそれらのニュースに言及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．

（3） **円安**

ロシアによるウクライナ侵攻や，日米金融政策の方向性の違いなどから急速に進行している円安に関する報道．2020 年 10 月には対ドルの円相場が 150 円台になり，32 年ぶりに円安水準を更新した．

（4） **上島竜兵**

2022 年 5 月 11 日にお笑いタレントの上島竜兵が亡くなった事に関する報道．

4.3 ニュース記事数と言及ツイート数の時系列変化の分析

各トピックに対する 1 日のニュース記事数とその言及ツイート数の時系列変化を，前者を上側に青色で，後者を下側に橙色でプロットして図3 に示す．縦軸はニュース記事数または言及ツイート数を，横軸はその投稿日を表す．

トピック「コロナウイルス」に関しては，図3(a) に示すように，ニュース記事数・言及ツイート数共に 1 月末と 7 月末に大きく増加している．これはコロナ禍における第 6 波と第 7 波の時期と一致しており，それらの実世界のイベントがニュース報道と Twitter 上の発言に影響したと考えられる．しかし，全体的な傾向としての値の増減とは別に，1 日ごとの値の変動にも着目すると，平日に値が高くなり，休日に値が低くなるという一週間単位の周期性が確認できる．これは，例えば新聞には休刊日や異なる内容の日曜版を発行する，各自治体の新規陽性者数の発表などの周期的なパターンが存在し，それが影響を与えていると推測できる．

トピック「ロシアによるウクライナ侵攻」に関しては，図3(b) に示すように，ニュース記事数と言及ツイート数共

にロシアが侵攻を開始した 2 月 24 日から値が急増し，その後一旦ピークに達した後はゆるやかに値が減少しているが，言及ツイート数の方の減少度合いが大きい．また，侵攻前の 15 日から 18 日には，ニュース記事数だけに値の増加が確認できる．これらの事柄から，ニュース報道が Twitter 上の発言に影響を与えていることが推測できるが，実際に侵攻されていない時点や，侵攻を開始してしばらく経過した時点では，Twitter ユーザの反応が鈍くなっていることが考えられる．また，このトピックに関しても，一週間の周期性が確認できる．

トピック「円安」に関しては，図 3(c) に示すように，ロシアのウクライナ侵攻が始まり，円安が始まった 3 月末からニュース・ツイート共に値が増加している他，1998 年以来，約 24 年ぶりに 135 円台まで下落した 6 月ごろや，円安がさらに進行して 150 円台に突入した 10 月ごろにもニュースとツイートの盛り上がりが確認できる．

トピック「上島竜兵」に関しては，図 3(d) に示すように，訃報が報じられた 5 月 11 日にニュースとツイートの値が大きくなっているが，その後すぐに値は減少している．このような報道では，Twitter ユーザの意見はほぼ同じで，異論はほとんど起こらないために，ユーザの間でそのトピックに関する議論が拡大・継続するようなことは起こらないからだと考えられる．

4.4 ニュースデータとツイートデータの相互相関

次に，各トピックのニュース記事数とその言及ツイート数の相互相関をプロットして図 4 に示す．横軸はニュース記事数とその言及ツイート数の一日単位のずれ（ラグ）を，縦軸はそのラグの時の相互相関函数の値を示す．

どのグラフでもラグ 0，つまり同日におけるニュース記事数と言及ツイート数の相関がもっとも高く，ニュース報道に対するユーザの言及にはほぼ遅延がないことがわかる．

また，トピック「コロナウイルス」とトピック「ロシアによるウクライナ侵攻」，トピック「円安」に関しては，それぞれ図 4(a) と図 4(b)，図 4(c) に示すように，ラグ 7 とラグ-7 での相関がラグ 0 の次に高い値を示しており，1 週間周期で相関の盛り上がりが確認できる．ただし，トピック「コロナウイルス」の方がトピック「ロシアによるウクライナ侵攻」よりも周期性が明確である．これは，メディア側に存在する周期性以外にも，地方自治体や保健所の体制に影響されて新規陽性者数の増減には一週間単位の周期性があることが知られており，それらが複合して影響したからと考えられる．また，トピック「円安」はトピック「コロナウイルス」よりも周期性に乱れが見られる．これは為替はより複雑に変動するために，それが影響を与えていると勘がられる．

トピック「上島竜兵」に関しては，図 4(d) に示すように，ラグの減少に伴い値が急減し，周期性も見られない．これは，ニュース記事数と言及ツイート数の時系列変化の結果と一致する．

4.5 ニュース記事数と言及ツイート数の変化点の分析

4.3 節で示した，一日ごとのニュース記事数とその言及ツイート数の時系列変化について，それぞれ ChangeFinder を用いて，次数 $k = 1$，窓サイズ $T = 5$，SDAR アルゴリズムの忘却パラメータ $r = 0.02$ に設定して算出した変化点スコアをプロットして，図 5 に示す．横軸は投稿日，第 1 縦軸はその日のニュース記事数または言及ツイート数，第 2 縦軸はその日の変化点スコアを表す．

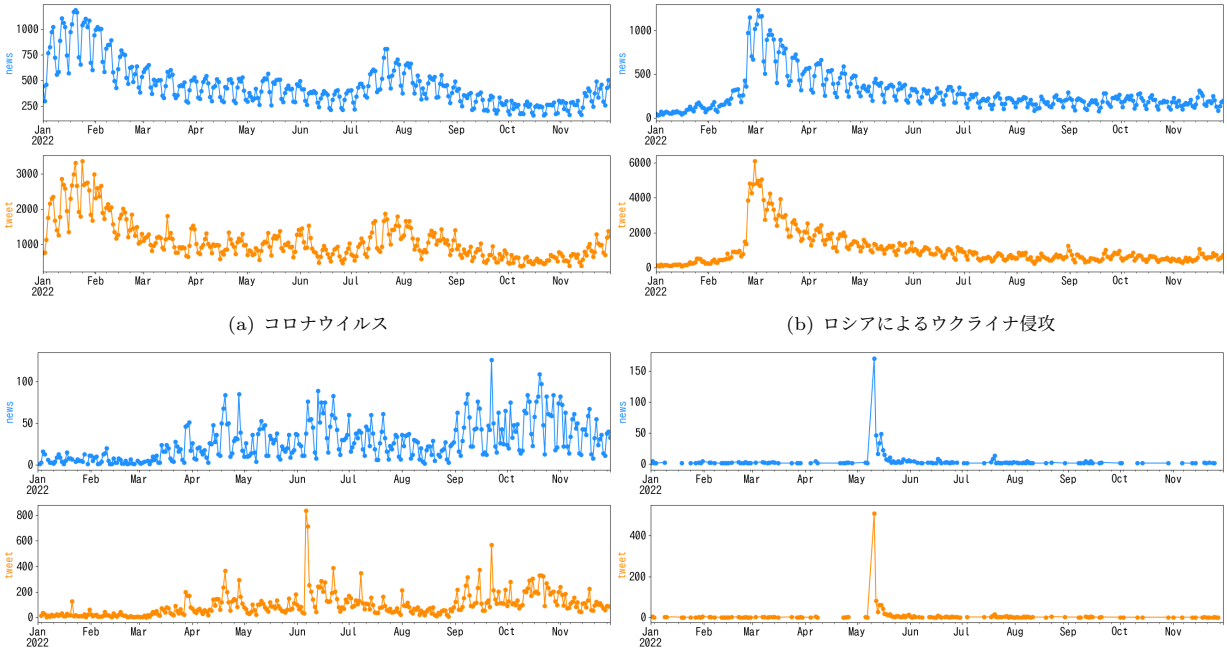
トピック「コロナウイルス」に関しては，図 5(a) に示すように，ニュース記事数では二つの変化点を検出している．一つ目の変化点として検出した 1 月 4 日では，一日の新規陽性者数が約 3 か月ぶりに 1000 人を上回るなど，第 6 波で感染が激しく流行し始めた時期であった．二つ目の変化点である 7 月 22 日では全国の感染確認数が 3 日連続で過去最多となるなど、こちらでは第 7 波で感染が激しく流行し始めた時期に変化点を検出したことが考えられる．

言及ツイート数では，1 月 5 日に変化点を検出し，ニュース記事で検出した変化点から一日遅れているものの，ニュース記事数と言及ツイート数共に新規陽性者数が大きく増加したタイミングで変化点スコアが高くなっていることが確認できる．また，他の変化点は検出されず，トピックに対する関心・姿勢がメディア側と Twitter ユーザ側で異なっている可能性があると考えられる．

トピック「ロシアによるウクライナ侵攻」に関しては，図 5(b) に示すように，ニュース記事数で一つ目の変化点として検出した 2 月 18 日では，ロシアがアメリカに回答した文書の内容を公開し，ウクライナに軍事侵攻する意図を否定しながらも，ウクライナの NATO 加盟を警告するなど，外交による解決を試みながらも，ロシアとウクライナ間で急速に緊張が高まりだした時期であった．二つ目の変化点である 2 月 24 日では，ロシアがウクライナに侵攻を開始した時期であり，ニュース記事数では侵攻前の緊張状態と侵攻開始の時期で二つの変化点を検出したことが考えられる．

言及ツイート数では，2 月 23 日に変化点を検出しているが，言及ツイート数の時系列変化において，実際に言及ツイート数が大きく増加したのは 2 月 24 日であることから，この変化点はニュースにおける一つ目の変化点を遅れて検出したのではなく，侵攻が開始する前に変化点を検出したと推測でき，ChangeFinder ではユーザの反応から実世界で大きなイベントが起こる前に，それを予測することが可能であると考えられる．

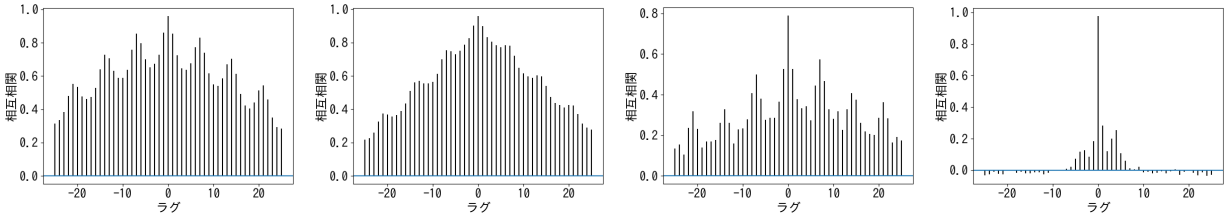
トピック「円安」に関しては，図 5(c) に示すように，まずニュース記事数で 5 つの変化点を検出した．一つ目の変化点として検出した 4 月 20 日では，一時 1 ドル=129 円台となり，約 20 年ぶりの円安水準を更新し，日銀が国債を無制限に買い入れる措置を取り始めるなど，円安の進行と，その対策を取り出した時期であった．二つ目の変化点として検出した 6 月 16 日では，5 月の貿易収支が過去最大の赤字と発表され，三つ目の変化点として検出した 9 月 8 日では，経常収支が過去最少の黒字と発表されるなど，これらは円安の影響を受けた報道が行われた時期であった．四つ目と五つ目の変化点として検出した 9 月 23 日と 10 月 22 日では，どちらも前日に政府の為替介入が行われており，どの変化点も円安の相場に関するニュース記事に加えて，政府の対策措置などその影響を受けた事柄が報じられ



(a) コロナウイルス　(b) ロシアによるウクライナ侵攻

(c) 円安　(d) 上島竜兵

図 3　ニュース記事数と言及ツイート数の時系列変化



(a) コロナウイルス　(b) ロシアによるウクライナ侵攻　(c) 円安　(d) 上島竜兵

図 4　ニュース記事数と言及ツイート数の相互相関

た時期を変化点として検出された考えられる．

言及ツイート数では，ニュース記事数と同様に 4 月 20 日に一つ目の変化点を検出したが，二つ目に 6 月 6 日を変化点として検出した. この日は日銀総裁が講演で「家計の値上げ許容度が高まってきている」との見解を示したと報道するニュース記事があり，このような波紋を呼ぶニュース記事に反応したユーザが多くいたことが推測できる．このトピックでは，ニュース記事数と言及ツイート数でそれぞれ検出した変化点に大きく異なる点があり，盛んに報道されないようなイベントでも Twitter ユーザが意見の対立や議論を呼ぶようなニュース記事であれば反応するという性質が大きく影響したと考えられる．

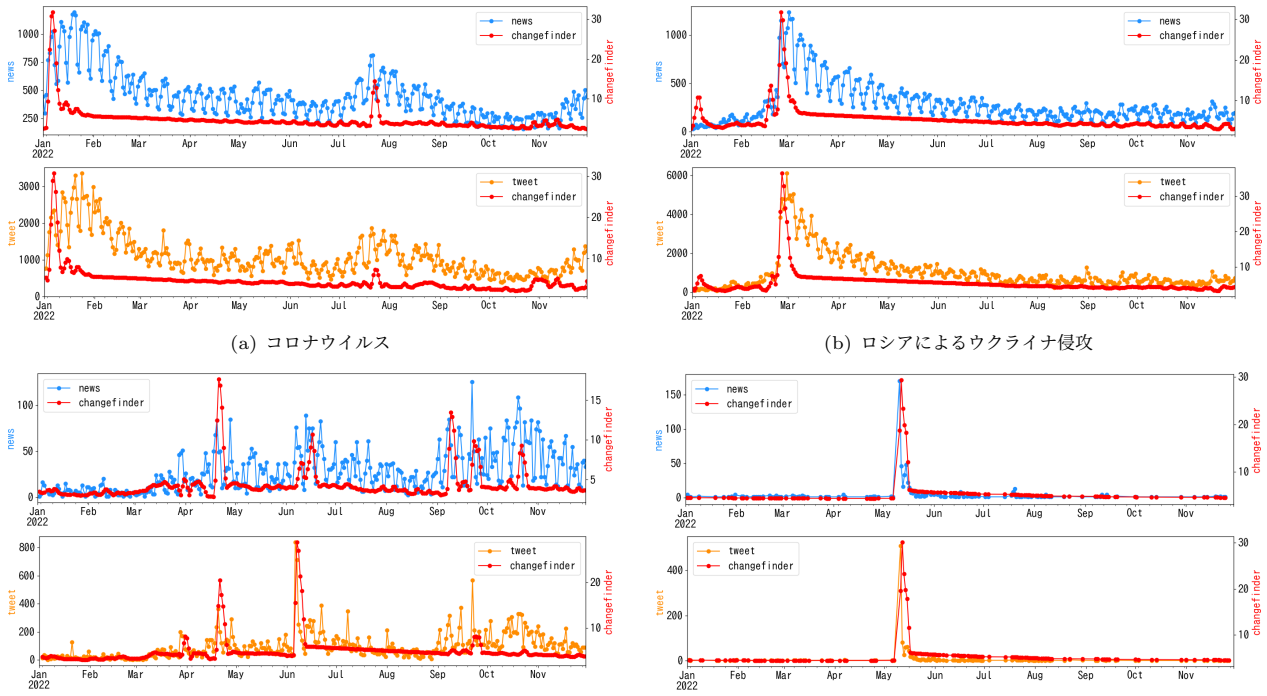
トピック「上島竜兵」に関しては，図 5(d) に示すように，ニュース記事数と言及ツイート数どちらでも訃報があった 5 月 11 日に変化点を検出している．このトピックに関しても，メディアとユーザに大きな影響を与えたと考えられるイベントが起こったその日に変化点スコアが高くなっており，またニュース側とユーザ側でトピックに対する姿勢・状況は似通っていると考えられる．

4.6　変化点直後のキーフレーズの分析

前節で示した変化点スコアから，変化点として決定した日を含む 3 日間のニューステキストについて，出現数上位 20 個のキーフレーズと，その出現数をプロットして図 6 から図 9 に示す．縦軸は抽出したキーフレーズを，横軸はその出現数を表す．また，一つのトピックで複数の変化点が検出された場合は，そのうち一つの変換点の観測期間でしか出現しなかったキーフレーズに下線を引いている．

トピック「コロナウイルス」に関しては，図 6 に示すように，二つのグラフのどちらにも「感染」や「新た」などのキーフレーズが確認でき，どちらの観測期間でも新規感染者数を発表するニュースが多かったことがわかる．また，図 6(a) でのみ確認されたキーフレーズには,「変異株」や「オミクロン株」などがあり，オミクロン株が流行しニュースで多く取り上げられた期間であったことが考えられる．図 6(b) でのみ確認されたキーフレーズの中には「過去最多」などがあり，これは前節でも述べた全国の感染確認数が 7 月 23 日時点で 4 日連続の過去最多となっていることから出現したキーフレーズと考えられる．

トピック「ロシアによるウクライナ侵攻」に関しては，図 7 に示すように，二つのグラフのどちらにも「ロシア軍」や「ウクライナ侵攻」などのキーフレーズが確認でき，どちらの観測期間でもロシア軍のウクライナ侵攻に関するニュースが多かったことがわかる．また，図 7(a) でのみ確認されたキーフレーズに



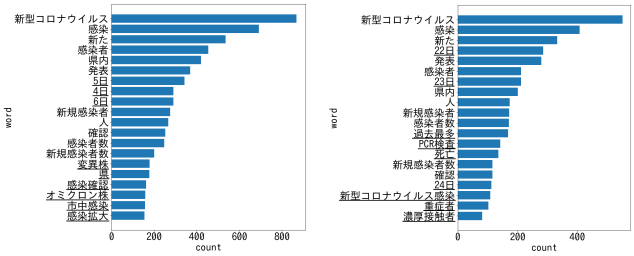
(a) コロナウイルス　(b) ロシアによるウクライナ侵攻

(c) 円安　(d) 上島竜兵

図 5　ニュース記事数と言及ツイート数の変化点

は「緊張」「ウクライナ東部」などがあり，この期間にはロシア軍がウクライナ東部で軍事演習を行ったことに対して，ロシアとウクライナ間で急速に緊張が高まりだしたとするニュースが多く報道されたと考えられる.「ドーピング問題」「選手」などのキーフレーズもこの観測期間でのみ出現しているが，これは冬季オリンピックでのロシア選手のドーピング問題に関するニュースが報道されていたためと考えられる．図 7(b) でのみ確認されたキーフレーズには「首都キエフ」「軍事侵攻」などがあり，24 日から首都キエフを含むウクライナ各地で砲撃や空襲が行われ，この期間では軍事侵攻が実際に開始されたことに関するニュースが多く行われたと考えられる.「制裁」もこの観測期間でのみ出現しており，これは侵攻に対する経済制裁など，他国の反応に関する内容も報道されていたことが影響していると考えられる．

トピック「円安」に関しては，図 8 に示すように，どのグラフでも「円」「円相場」「日銀」などのキーフレーズがあり，トピック全体と関連する単語が含まれていることが確認できる．ここからこのトピックで一つのグラフでのみ出現した単語について確認していく．図 8(a) では「上昇」「高騰」などのキーフレーズが出現しており，円安が進行していること自体を報道しているニュースが多かったことが考えられる．図 8(b) では「金融政策決定会合」「金融緩和策」などのキーフレーズが出現しており，円安の進行自体を報道していた前の期間と異なり，この期間では円安に対する政府の対応に関する報道が多く行われるようになったと考えられる．図 8(c) では「iPhone」「価格」などのキーフレーズが出現しており，この期間でも円安の進行自体が取り上げられているのではなく，円安による iPhone の値上げに関するニュースが多かったことが考えられる．図 8(d) で



(a) 1 月 4 日〜1 月 6 日　(b) 7 月 22 日〜7 月 24 日

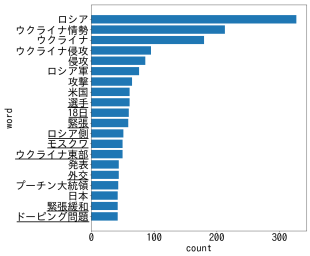
図 6　変化点直後のキーフレーズ（コロナウイルス）

は「円買い介入」「緩和」，図 8(e) では「為替介入」などのキーフレーズがあり，これは 9 月 22 日と 10 月 21 日に 2 度行われた政府の為替介入について報道しているニュースが多かったことが考えられる．

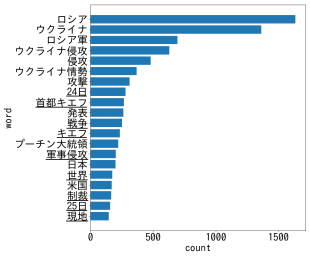
最後にトピック「上島竜兵」に関しては，図 9 に示すように，キーフレーズの抽出は一回しか行われなかったが,「上島竜兵」「死去」「相談窓口」などのキーフレーズが出現している．ニュース記事には死因について伏せたものが多いが，キーフレーズから自殺を図ったことが推測でき，ニュース記事のテキストには直接書かれていないような内容もキーフレーズの抽出によって推測できることがわかる．

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を用いて，その時系列変化を分析することで，ニュースとソーシャルメディアのユーザの反応の間には，トピックによって多少の差はあるものの，強い関係性があると同時に，ほぼ遅延がないことがわかった．

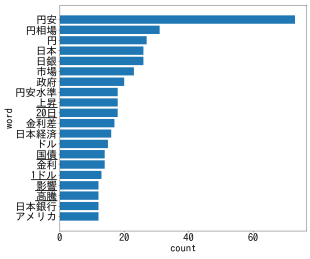


(a) 2 月 18 日〜2 月 20 日

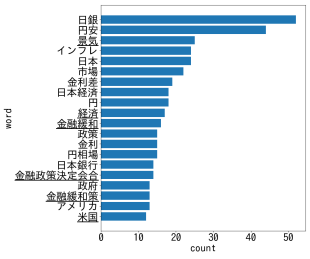


(b) 2 月 24 日〜2 月 26 日

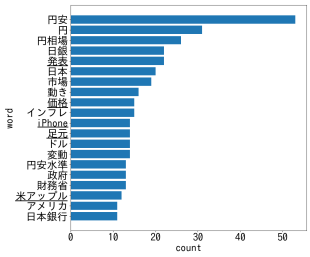
図 7　変化点直後のキーフレーズ（ロシアによるウクライナ侵攻）



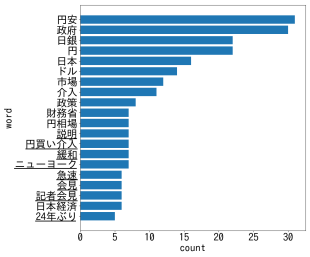
(a) 4 月 20 日〜4 月 22 日



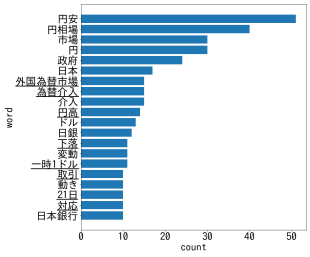
(b) 6 月 16 日〜6 月 18 日



(c) 9 月 08 日〜9 月 10 日

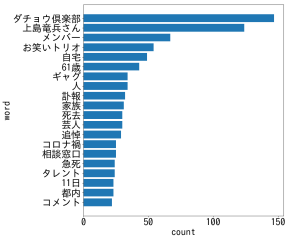


(d) 9 月 23 日〜9 月 25 日



(e) 10 月 22 日〜10 月 24 日

図 8　変化点直後のキーフレーズ（円安）



(a) 5 月 11 日〜5 月 13 日

図 9　変化点直後のキーフレーズ（上島竜兵）

また，ChangeFinder で検出した変化点は，対象トピックで起こった現実世界のイベントにほぼ対応していることがわかった．また，メディアのニュース報道や，それに伴う Twitter 上の言及には一週間の周期性が存在するが，ChangeFinder はそのような周期性の影響を受けにくいことも確認できた．さらに，変化点直後の期間のキーフレーズ抽出を行った結果，変化点が生じた原因の把握に役立つことも確認できた．

今後の課題として，まず，本論文ではキーフレーズ抽出を用いた分析はニュース記事のテキストのみに対して行ったが，さらに言及ツイートに対しても同様の分析を行うとともに，メディアとユーザに違いが存在するかどうかと，存在する場合はその違いが生まれる原因を明らかにする．さらに，複数の分析手法を統合して，言及ツイート数の時系列変化から大きな影響を与えるニュースを自動発見し継続して監視できるようなフレームワークの構築を試みる予定である．

謝　　辞

本研究は JSPS 科研費 21H03557 の助成を受けた．

文　　献

[1] Akemi T. Chatfield, Christopher G. Reddick, and K. P. Choi. Online media use of false news to frame the 2016 Trump presidential campaign. In *Proceedings of the 18th Annual International Conference on Digital Government Research (dg.o'17)*, p. 213–222. Association for Computing Machinery, 2017.

[2] 張建偉, 河合由起子, 熊本忠彦, 白石優旗, 田中克己．多様な印象に基づくニュースサイト報道傾向分析システム．知能と情報, Vol. 25, No. 1, pp. 568–582, 2013.

[3] 田中皓介, 神田佑亮, 藤井聡．公共政策に関する大手新聞社報道についての時系列分析．土木学会論文集 D3, Vol. 69, No. 5, pp. I_373–I_379, 2013.

[4] 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.

[5] 木村浩, 佐藤亮佑, 芝田雄吾, 鳥海不二夫, 榊剛史, 風間一洋, 福田健介．福島第一原子力発電所事故に関するテレビ報道とツイッター情報との関連性．日本原子力学会年会・大会予稿集, 2012.

[6] 上田叶, 脇田建．アポロ計画の長期会話記録を用いたマルチスケールな視覚的分析．第 36 回人工知能学会全国大会 (JSAI2022), 1M5-OS-20c-03, 2022.

[7] 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フレーズ抽出．言語処理学会第 24 回年次大会, pp. 1251–1254, 2018.

[8] Yamanishi Kenji and Takeuchi Jun-ichi. A unifying framework for detecting outliers and change points from time series. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 4, pp. 482–492, 2006.

[9] 山村翔, 熊谷充敏, 神谷和憲, 倉上弘．変化点検知を用いた新種スキャンの早期発見手法の検討．コンピュータセキュリティシンポジウム 2017 (CSS2017), 2017.

[10] Liang Xinnian, Wu Shuangzhi, Li Mu, and Li Zhoujun. Unsupervised keyphrase extraction by jointly modeling local and global context. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 155–164, 2021.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, pp. 4171–4186, 2019.

[12] Ceek.jp news - 最新ニュース検索．`http://news.ceek.jp/`. (Accessed on 04/27/2022).