

外部知識のネットワーク関係を活用したニュースメディアのスタンス分類

江口 航野[†] 横山 昌平^{†, ††}

[†] 東京都立大学システムデザイン研究科情報科学域 〒191-0061 東京都日野市旭が丘 6-6

^{††} 東京大学生産技術研究所 〒153-8501 東京都目黒区駒場 4-6-1

E-mail: [†]eguchi-koya@ed.tmu.ac.jp, ^{††}shohei@tmu.ac.jp

あらまし 情報過多の、そして分極化した社会において、情報伝達メディアであるニュース記事は、配信メディアが異なると、同一トピック物事について、報じ方のスタンスに差異が生まれ、分断が生じ、正しく情報を収集できないことが増加している。そのため、ニュースメディアのスタンスを判定することが重要となるが、網羅的にニュースメディアの記事文面を取得するのは難しい上、記事のみからでは、メディア間の相対的な立ち位置を判断することは難しい。そこで、本研究では、ニュース記事以外の外部知識を活用することで、ニュース記事本文の解析を行うことなく、メディアのスタンスを判定する手法を提案する。

キーワード 議論マイニング、スタンス分類、ニュース記事、SNS、次元削減、クラスタリング

1 はじめに

2020 年に世界で消費されたデータが 59ZB[1] であり、年々世界で消費されるデータ量は膨張している。このことが象徴するように、近年、情報過多の社会の傾向が強まり、またグローバル化も合わせて、ますます複雑化・多様化の一途をたどっている。この様な社会において日々生じる社会問題も同様に複雑化しており、情報伝達メディアであるニュースの重要性が高まっている。

情報過多によって、人が処理できる情報量以上の情報が出回ることに対して、システムによる情報の推薦が行われることで効率的な情報集取を可能とした。その一方で、同質的なトピック・思想のコンテンツのみしか触れることができなくなつたことから、エコーチェンバー現象やフィルターバブルといった言葉に代表されるように、日々接する情報に偏りが生じる状況が問題となっている。情報収集の手段として代表的な存在であるニュース記事は、配信メディアが異なると、同一トピック物事について、報じ方のスタンスに差異が生まれることが多い。また、同じニュース配信メディアであっても、右寄り・左寄りのスタンスであるかは、記事のトピックによって変化していくことさえある。

アメリカのシンクタンクであるピュー研究所の調査 [2] によると、共和党と民主党の支持者の保守とリベラルのイデオロギーを数値化し、分布を比較したデータによると、調査が開始された 1994 年から、年々二つの政党のイデオロギーの分布の重なり部分の面積が減少している。つまり、保守とリベラルの間を行き来していた中道派層が減少し、分極化が進んでいるということである。この支持するイデオロギー分極化に合わせて、ニュースメディアの分極化が進んでいる [3]。

この分極化によって現実を正しく把握できなくなってしまった一つの例として、2016 年のアメリカ大統領選挙が挙げられ

る。開票前は、フォックスニュース¹は接戦を予想していた一方、ヒラリー・クリントン候補が圧勝すると CNN²やニューヨークタイムズ³など主要メディアは予測していた。蓋を開けてみると、ドナルド・特朗普候補が勝利し、主要メディアのみを見ていた民主党支持者に衝撃を与えた。

そこで、本研究では、ユーザーに対して各ニュース配信メディアの報道スタンス情報を提供することで、ユーザーが各ニュース記事より適切に情報を得る支援を行うことを目的としている。手法として、記事本文データを使用せず、記事 URL のドメイン名とツイートデータのみを使用している。これは、網羅的にメディアの記事本文を取得するためには、各メディアに合わせたシステムを設定する必要があり、また、記事本文を解析対象としたシステムは、言語ごとに合わせたシステムの構築が必要となるなど処理コストが高くなるためである。本論文では、2 章にて、先行研究および関連研究について本研究との関わりについて述べ、3 章では、提案手法を用いる際の前提知識を述べる。4 章では提案手法を述べ、5 章では実験結果を、6 章で考察を述べ、7 章で本研究のまとめと今後の展望を述べる。

2 関連研究

本章では、関連研究について述べる。本研究の提案手法において、Twitter のスタンス分類および、ニュースのスタンス分類が関連研究として挙げられる。

2.1 Twitter におけるスタンス分類

Twitter のスタンス分類の多くは、Magdy らの研究 [4] や Kutlu らの研究 [5] など教師データをともなった研究が多く、人手である程度ラベルを付けたデータを作る必要があった。

1 : <https://www.foxnews.com/>

2 : <https://www.cnn.com>

3 : <https://www.nytimes.com/>

Darwish らは、ツイートのデータから、ユーザーのスタンスを教師学習を行わず判定することを目的にしたシステムを提案している[6]。手法として、ツイートのテキスト本文、リツイート、ハッシュタグをユーザーごとのデータにまとめ、それらをベクトル化、次元削減し、その削減後のベクトルをクラスタリングすることで、教師学習を行わずしてユーザーのスタンス分類を実現している。本研究の提案手法で、ユーザーをベクトル化する際にこの研究の手法の一部を取り入れている。

2.2 ニュースにおけるスタンス分類

近年、フェイクニュースの検知等のタスクを目的とした研究が盛んに行われており、その研究の関連分野でもある議論マイニング(Argument mining)をテーマにした研究が多くなされている。Hanselowski らの研究[7]では、フェイクニュースの検知の第一歩として、ニュース記事のスタンスを分類することを目標として、ニュースの見出しから教師データで学習させたニューラルネットワークモデルを使い、分類する手法を提案している。これらは、1記事単位のスタンスを判定することを目的としており、また、言語的内容に依存した構成となっているため、ドイツ語での同様の研究[8]が別途行われているように、他言語のニュースを対象とする際は、その言語での教師データを別に準備する必要がある。

本研究と同様に、メディア単位のスタンスの分類の研究として、Baly らの研究[9]がある。Baly らは、各ニュースメディアの記事を収集し、記事の品詞や感情値、メディア公式のSNS上で投稿に対する「いいね」やコメント、記事のトピックや単語などの特徴量を用いて、SVM にて学習させ、分類器を生成している。また、ニュースメディアのスタンスに注目し、メディアスタンスの可視化を行った研究として、Yoshioka らの研究[10]が挙げられる。Yoshioka らは、ドナルド・トランプとヒラリー・クリントンが戦った2016年のアメリカ大統領選挙時のニュースを対象として、各メディアが出した記事の中に、出現する候補者の名前の回数を使い、メディアのバイアスの可視化を行った。これら研究は、記事中の言語的特徴や出現単語等使用しているため、各メディアに応じたクロールシステムであったり、言語に応じたシステムの構築を行う必要があるが、本研究での提案手法は、これらの処理を含んでいないため、汎用性の面で優位性がある。

また、本研究と同じように、Twitter 上のネットワークを活用して、メディア間の分極度の可視化を行った研究として、藤兼らの研究[11]が挙げられる。藤兼らの研究では、各ユーザーをユーザーノードとし、各メディアをメディアノードとして、ユーザーがメディアのリンクをツイートした場合、エッジを張ることで、2部グラフを作成した。この2部グラフを活用して、各メディア間の分極度を可視化している。本研究での提案手法2つのうち、Twitter の共起ネットワークを活用した手法(提案手法1)が似ているが、本手法が共起関係のみを用いているのに対して、藤兼らは、2部グラフから、メディア及びユーザーの共被閲覧、共閲覧関係の両方を用いて、分析を行っている。また本研究では、可視化に留まらず、各メディアのスタンスに

合わせた分類及び、精度検証までを行っている。

3 準 備

提案手法で、使用している次元削減、およびネットワークのクラスタリングについての事前知識をこの章にて述べておく。

3.1 UMAP

UMAP[12]は「Uniform Manifold Approximation and Projection」の略であり、次元削減の手法の一つである。次元削減することは、高次元のデータの計算コストを下げることに加え、データの可視化が可能となり、人間が理解できる形で表すことができるようになる。別の次元削減の手法であるt-SNE(t-Distributed Stochastic Neighbor Embedding)[13]が似た手法として挙げられるが、t-SNEはデータセットが大きくなると、性能が低下し、また処理の実行時間が指數関数的に増加してしまう。python のパッケージの scikit-learn の t-SNE での実装で、MNIST[14]のデータセット 784 次元 70000 個のデータをすべて 2 次元に次元削減するのに、1450 秒掛かるのに対して、UMAP で次元削減を行うと、87 秒で完了し、約 17 倍の速度で処理ができる。そのため、本研究では、ツイート数が非常に膨大になることから、次元削減には UMAP を用いることにした。

3.2 Louvain 法

Louvain 法[15]は、大規模ネットワークのコミュニティを検出するための手法であり、また階層的クラスタリングの一種、そしてネットワークデータに対する代表的なクラスタリング手法である。このクラスタリングは、コミュニティへのノードの割り当て精度を示す指標のモジュラリティ(Modularity)[16]が最大化するように、クラスタの結合を行う。Louvain 法が大規模データに対して使うことができ、また、処理時間も高速であるため、本研究では Louvain 法を用いることにした。

4 提案手法

本章では、ニュースメディアのスタンスをクラスタリングする2つの手法について述べる。一つは、ツイート中に出現するURLの共起ネットワークを活用した手法であり、もう一つは、ツイートのユーザー分類を活用した手法である。以下、提案手法それぞれについて個別に述べる。

4.1 提案手法 1

本手法は、データセットとなるツイートデータの取得、データの前処理、共起分析を行うことで、ニュースメディアのスタンス関係を取得するものである。概要図を図1に示す。

4.1.1 データの取得と整形

Twitterより、ニュースメディアのスタンスを調べたいトピックについてツイートを収集する。その中で、ニュースメディアの URL の記載のあるツイートを抽出する。取得したデータの各ツイートの URL から、ドメインを取得し記録していく。その後、ユーザーごとに各ユーザーがツイートしたニュースメディアの URL のドメインをまとめて、ユーザーとニュースメ

ディアの URL ドメイン一覧が対となった対応表を作成する。

4.1.2 共起分析

ユーザーとニュースメディアの URL ドメインの対応表に基づき、共起ネットワークを作成していく。ここでの共起ネットワークとは、あるユーザーがメディア A、メディア B、メディア C の 3 つのメディアのドメインを持つ URL の記載のあるツイートをしている場合、各メディアをノード、メディア間をエッジとすると、この（メディア A ～メディア B）、（メディア A ～メディア C）、（メディア B ～メディア C）の 3 つの共起関係のネットワークができる、全ユーザーのメディアの共起関係を足し合わせたものとしている。また、エッジの太さとして、2 メディア間の共起関係の回数として表現している。共起関係を作成する際に、すべてのドメイン間に共起関係があるとするのではなく、条件以上の出現関係のみのドメインを使用して、共起関係を構築していく。その際の条件は、次の式を満たす関係である。ユーザーのツイート中に出現した URL の n 種類のドメインのうち、 i 種類目のドメインの出現回数を $domain_i$ と定義し、A($\leq n$) 番目のドメインと B($\leq n$) 番目のドメインの共起関係が成立するかは、式 1 を満たす時とした。

$$|domain_A - domain_B| < \frac{\sum_{i=0}^n domain_i}{n} \quad (1)$$

あるユーザーの使用した URL のドメイン A の出現回数が 100 回、ドメイン B が 1 回の際に、このドメイン A とドメイン B に共起関係があるとすることを防ぐことを理由に、この条件を付けて行う。この共起関係データに基づき、Louvain 法を用いてネットワーククラスタリングを行う。

4.2 提案手法 2

本手法は、ツイート中のテキスト、リツイート、ハッシュタグ

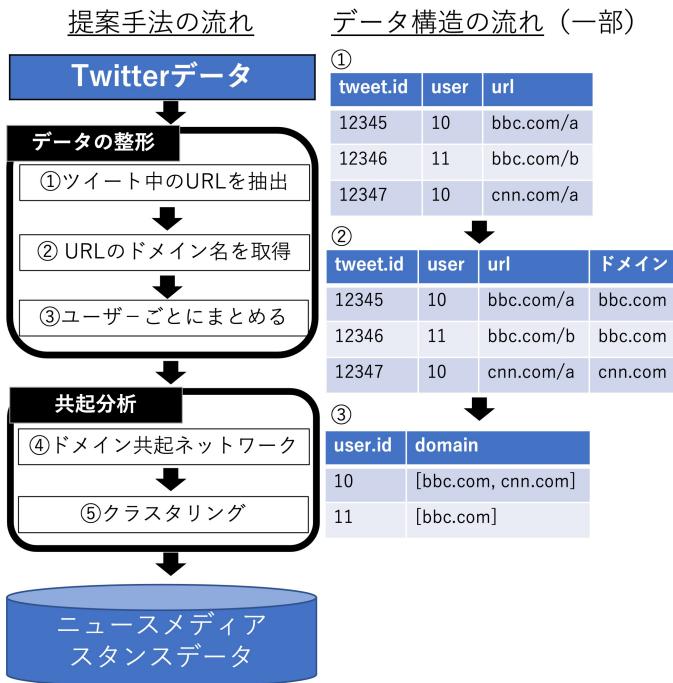


図 1 提案手法 1

グデータを特徴量としてユーザーごとにまとめ、ベクトル化することで、ユーザーの特徴量として生成する。このベクトルの次元数を削減した後、ユーザーの特徴をドメインの特徴として継承させ、ドメインごとの重心を求め、それをクラスタリングすることで、ニュースメディアのスタンス関係を取得するものである。概要図を図 2 に示す。

4.2.1 データの取得

Twitter より、ニュースメディアのスタンスを調べたいトピックについてツイートを収集する。その中で、ユーザーごとの特徴をツイートより導き出すため、リツイートも含めてツイート回数が 10 回以上のユーザーのツイートを対象とする。また、ツイートにニュースメディアの URL を含むツイート数が処理対象のツイート数の半数となるように設定する。

4.2.2 ユーザーの特徴量

ツイートごとのデータから、ユーザーごとのデータ構造に変換する。つまり、ユーザーごとに、そのユーザーがツイートした全テキスト、全リツイート、全ハッシュタグのデータを特徴量として保存する。次に、ツイートテキスト本文に対して、メンションと URL の削除を行った後、各特徴量を、カウントエンコーディングを行い、各特徴量を結合することで、ユーザーのベクトル化を行う。次に、生成したベクトルに対して、UMAP を用いて、2 次元に次元を圧縮する。次元数削減する理由として、対象とするツイート数が多くなると、「次元の呪い」によって計算に必要なメモリ量と時間が著しく増加する上、またクラスタリング時に精度が低下してしまうことを防ぐためである。また、これによって、ユーザーのスタンスを二次元空間に表現することが可能となる。

4.2.3 スタンス分類

ユーザーのツイート内容は、そのユーザーのスタンスを継承しており、またそのツイート（リツイートも含む）に URL として出現するニュースメディアのスタンスも、ユーザーのスタンスを継承しているとして、ニュースメディアのベクトルを、そのユーザーベクトルから継承する。これは、ユーザーは、同じ思想のユーザーのツイートをリツイートする傾向が強いという結果 [17] を発展させ、その中に出現するニュースメディアも同じ思想である傾向があるとしたためである。よって、ユーザーのベクトル情報から、各 URL が 2 次元空間にプロットされたベクトル空間が生成される。このベクトル情報に基づき、各 URL のドメインごとの重心を求め、その重心をニュースメディアのベクトルであるとする。ただし、重心を取る際に、50 件に満たないドメインは、対象外として処理する。最後に、K-means 法を用いて、クラスタリングを行う。

5 実験

本章では、2 つの提案手法を用いた実験について述べる。

5.1 データセット

ツイートデータセットは、表 1 の条件のツイートデータ⁴を

4: Twitter データは NTT データから提供を受けたものである

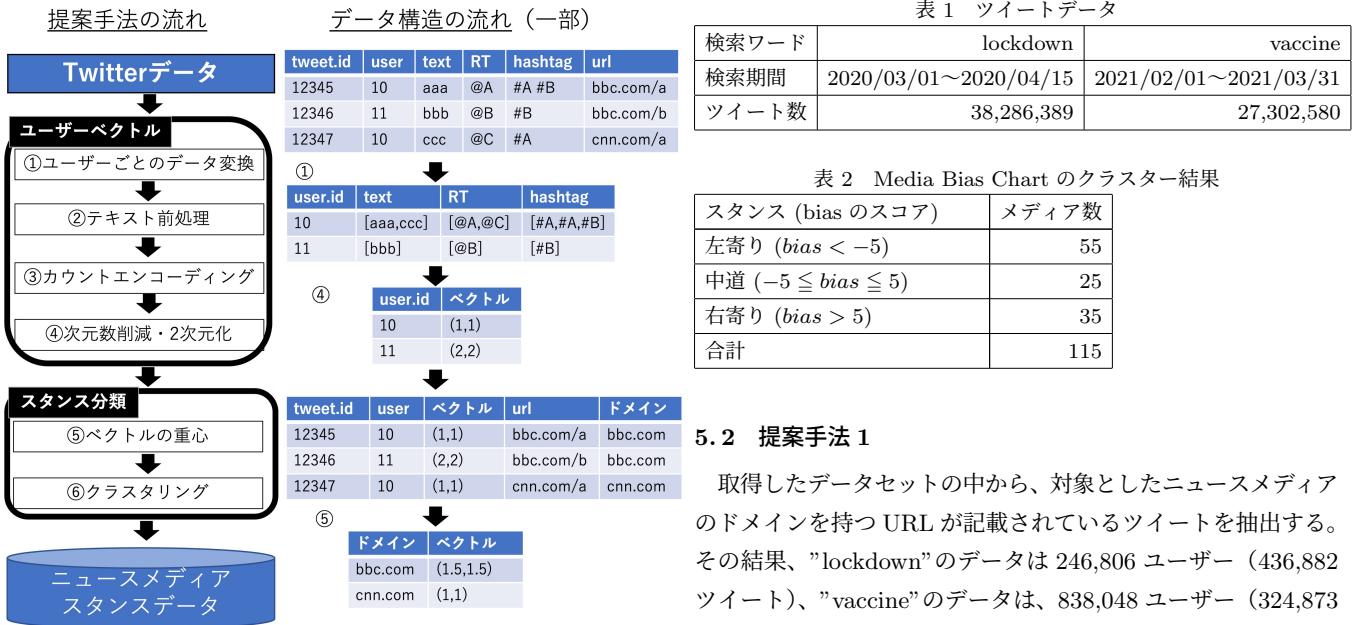


図 2 提案手法 2

収集した。対象とするニュースメディアは、アメリカのニュースメディアを中心に、主張の政治的偏りおよび内容の信頼性を図示している Media Bias Chart⁵に掲載しているニュースメディアとした。また、正解クラスタとしては、Media Bias Chart の Bias の指標を基に、左寄り、中道、右寄りの 3 つのクラスターに分けた。その結果、表 2 の通りのニュースメディア数となつた。また、Bias Score 横軸に、Reliability Score を縦軸にして可視化させると、図 3 のようになった。

5 : <https://adfontesmedia.com/>

| 検索ワード | lockdown | vaccine |
|-------|-----------------------|-----------------------|
| 検索期間 | 2020/03/01～2020/04/15 | 2021/02/01～2021/03/31 |
| ツイート数 | 38,286,389 | 27,302,580 |

表 2 Media Bias Chart のクラスター結果

| スタンス (bias のスコア) | メディア数 |
|------------------------------|-------|
| 左寄り ($bias < -5$) | 55 |
| 中道 ($-5 \leq bias \leq 5$) | 25 |
| 右寄り ($bias > 5$) | 35 |
| 合計 | 115 |

5.2 提案手法 1

取得したデータセットの中から、対象としたニュースメディアのドメインを持つ URL が記載されているツイートを抽出する。その結果、”lockdown”のデータは 246,806 ユーザー（436,882 ツイート）、”vaccine”のデータは、838,048 ユーザー（324,873 ツイート）をを解析対象とした。

このツイートデータに対して、4.1 で述べたように、データの整形の処理を行い、共起ネットワーク図を作成した（図 4）。このネットワークに基づき、python の igraph のパッケージ [18] を使用して、Louvain 法でのクラスタリングを行った。

5.3 提案手法 2

提案手法 1 同様に、取得したデータセットの中から対象としたニュースメディアのドメインを持つ URL が記載されているツイートを抽出し、ツイート回数が多い順にユーザーを抽出し、これらユーザーを対象とした。また、対象ドメインを持つツイート以外にも、対象ユーザーとしたユーザーのツイートも解析対象として追加する。これによって、その他のツイート情報から、より正確にユーザーの Twitter 上のネットワーク上で

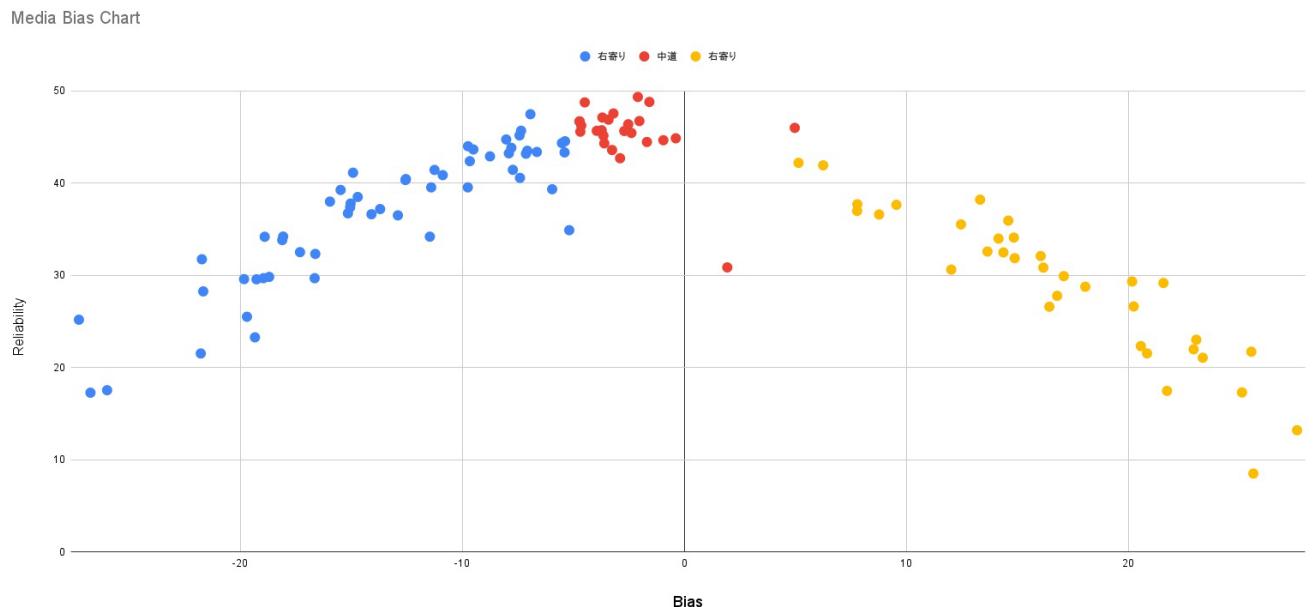


図 3 Media Bias Chart の正解クラスター

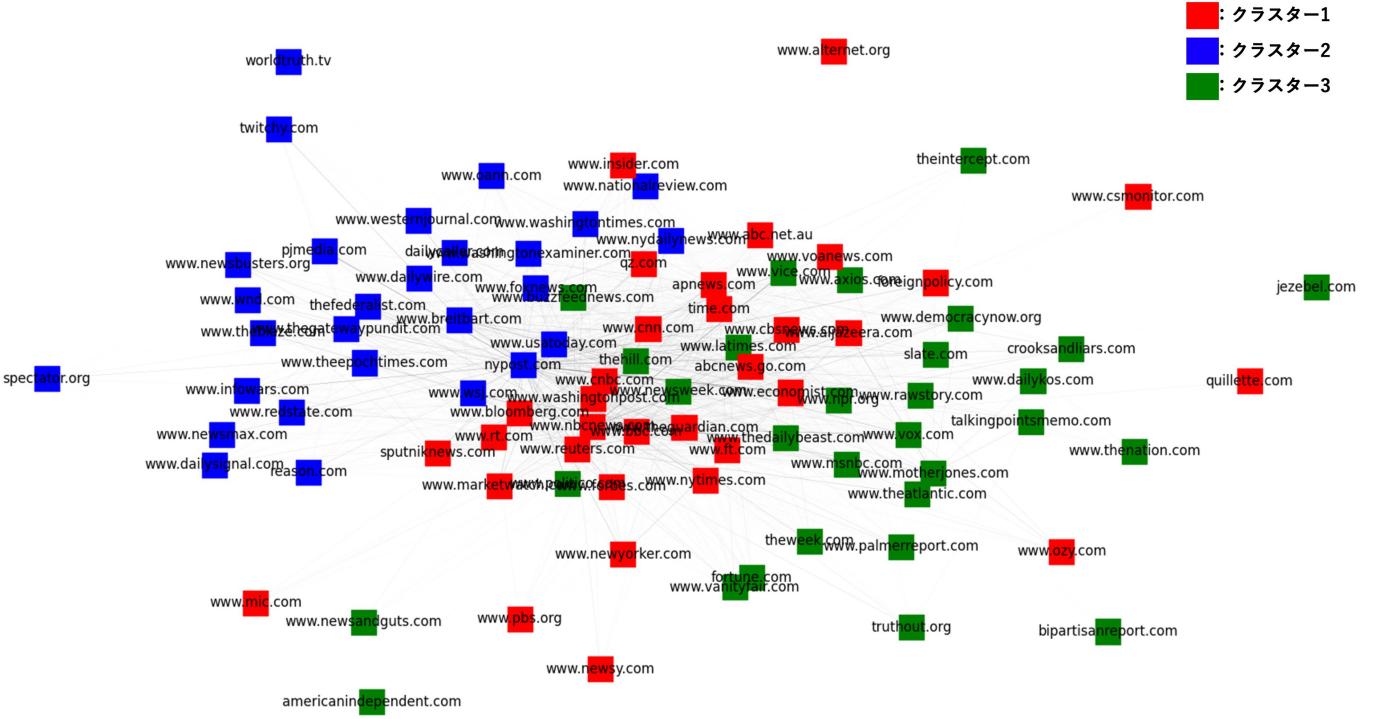


図4 提案手法1_クラスタリング結果(使用データ:lockdown)

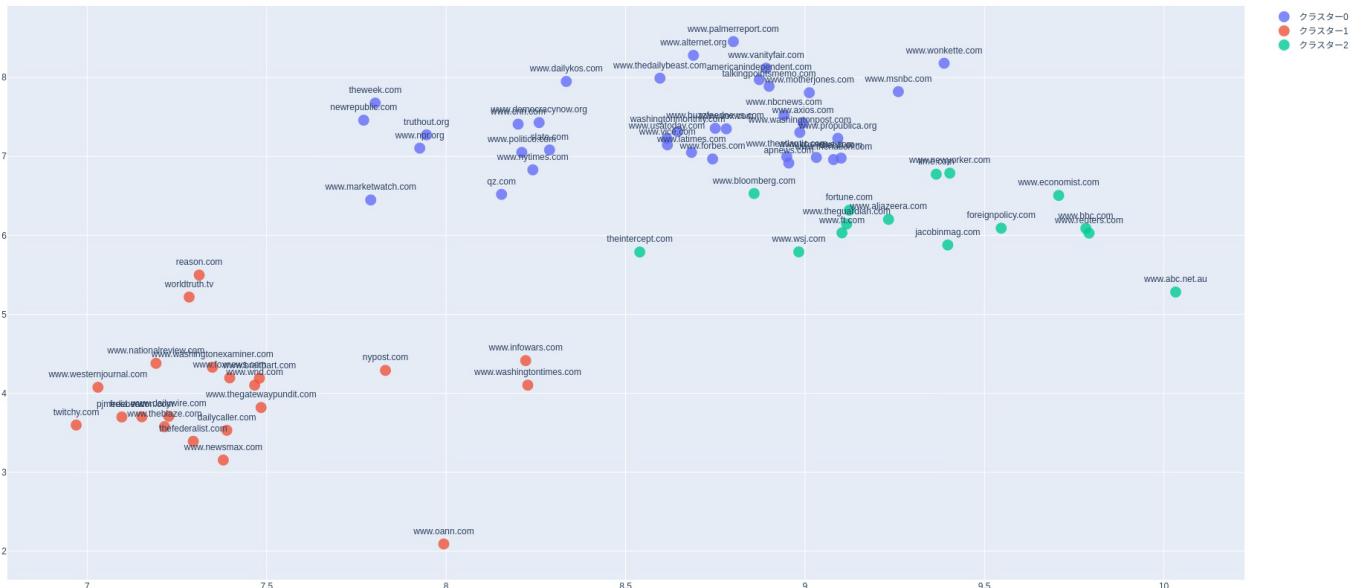


図5 提案手法2_クラスタリング結果(使用データ:vaccine、ユーザー数:25,000)

の立ち位置を判定することを目指している。4.2で述べたように、このツイートデータからユーザーのベクトルを取得する。カウントエンコーディングおよび、k-means法クラスタリングは、pythonのScikit-Learnのパッケージ[19]を用いて、エンコーディングとクラスタリングを行った。次元数の削減については、pythonのUMAPのパッケージ[20]を用いてUMAPでの次元削減を実装した。クラスタリング結果は、図5となった。

5.4 検証と結果

実験結果は、表3の通りである。検証には、純度(Purity)を用いた。これは、クラスター内に含まれる正解ラベルのうち、

最もクラスター内に多くのインスタンス数を持つものの割合の重み付き和で、式2で与えられる。

$$Purity = \frac{1}{N} \sum_{r=1}^k \max_i (C_r \cap A_i) \quad (2)$$

この式で、 k はクラスター数、 N をニュースメディア数として、 r 番目の生成したクラスターを C_r 、 i 番目の真のクラスターを A_i として示している。

6 考察

提案手法1と提案手法2の両方とも、各メディア記事の内容

表 3 実験結果

| 使用データ | 手法 | ユーザー数 | ツイート数 | 純度 (Purity) |
|----------|--------|---------|---------|--------------|
| lockdown | 提案手法 1 | 246,806 | 436,882 | 0.747 |
| | | 10,000 | 111,397 | 0.690 |
| | | 15,000 | 166,793 | 0.685 |
| | 提案手法 2 | 20,000 | 219,846 | 0.703 |
| | | 25,000 | 276,425 | 0.661 |
| | | 30,000 | 333,262 | 0.729 |
| | | 324,873 | 838,048 | 0.805 |
| vaccine | 提案手法 1 | 10,000 | 124,651 | 0.826 |
| | | 15,000 | 181,783 | 0.789 |
| | | 20,000 | 237,139 | 0.775 |
| | 提案手法 2 | 25,000 | 291,553 | 0.836 |
| | | 30,000 | 343,879 | 0.760 |

を取得することなく、Twitter 上のユーザーのネットワーク関係より、どちらのデータでも各手法で純度のスコア最大 70%以上を、手法 2 の”vaccine”のデータでは最大 83.6%を達成した。この結果から、本手法が分類を行う上で十分な精度であることが確かめられた。また、Twitter 上で形成される思想的スタンスのクラスターと、ニュースメディアの報道スタンスに同質性が存在し、本手法がこの同質性を反映していると確認できた。

提案手法 1 のクラスタリング結果の図 4 を見てみると、右側に散布している青色のクラスター 2 が右寄りメディアのクラスターであり、中央に散布している赤色のクラスター 1 は中道のメディアのクラスターであり、右側に散布している緑色のクラスター 3 は、左寄りのメディアのクラスターであることがわかる。同様に、提案手法 2 のクラスタリング結果の図 5 を見てみると、左下に散布している赤色のクラスター 1 は、右寄りメディアのクラスターであり、中央に散布しているクラスター 2 は、中道メディアのクラスターであり、右側に散布している青色のクラスター 0 は、左寄りのメディアのクラスターであることが分かる。

ここから、提案手法 1 および提案手法 2 の両方とも、中道メディアのクラスターを挟んで左右寄りのメディアのクラスターが位置するという各クラスターの相対的な位置が同じであることが観測でき、また、これは正解データとして与えた Media Bias Chart でのクラスターの位置関係(図 3) とも同じであることが確認出来る。特に、提案手法 2 のクラスタリング結果の左寄りメディアのクラスター 0 と、中道メディアのクラスター 2 が近く、また、中道メディアのクラスター 2 と右寄りメディアのクラスター 1 の境界が明確になっている点などが、Media Bias Chart と非常に類似していることが観測できた。

また、”lockdown”のツイートデータよりも”vaccine”のツイートデータの純度が約 10%ほど高い結果となった。この結果から、本手法がトピックに基づいたスタンスを分類していると言えるのではないかと考える。”lockdown”のツイートは、新型コロナウイルスが流行したころで、外出制限を設けるロックダウン政策についてコメントしているツイートである。民主党支持者の多い州 (blue states) は、共和党支持者の多い州 (red states) よりも積極的な感染防止措置の対策を行っているとい

う記事 [21] がある一方で、Chilton らの研究 [22] によると、支持政党に関わらずその当時は国家的危機として、右寄り・左寄りのスタンス関わらず感染防止のための私権制限を容認する人の割合が高いという調査もあるなど、これまで一般的な右寄り・左寄りのスタンスにあまり固定されなかった。

その一方、”vaccine”のツイートは、新型コロナウイルスのワクチン接種に対するコメントのツイートである。ワクチン接種については、左寄りメディアは、ワクチン接種を促す報道をする一方、右寄りのメディアは懐疑的な情報を織り交ぜながら報道する [23] など、これまでの伝統的なメディアスタンスに沿った論調となっている。この点を踏まえると、トピックに応じたメディアのスタンス情報を取得できていると考えられる。

7 おわりに

本研究では、Twitter 上のユーザーのネットワークに注目することで、ニュースメディアの報道スタンスを、各メディアの記事を参照することなく分類することを目指した。ツイート中に出現する URL の共起ネットワークを活用した手法と、ツイートのユーザー分類を活用した手法で、Media Bias Chart に掲載されているメディアを対象にして検証を行った。検証の結果、両手法で最大で 80%以上の純度のスコアを記録することができた。加えて、これまで伝統的なメディアの報道スタンスとは異なった新型コロナウイルスへの対策のロックダウン政策についてデータ (“lockdown” データ) と、これまで伝統的なメディアのスタンス通りの報道を行ったワクチン接種に対する政策についてのデータ (“vaccine” データ) を比較することで、トピックに合わせたスタンス分類が行えていることを示せた。

次なる目標は、新出メディアに本手法が対応できることの確認である。本研究の提案手法は言語や特定のメディアに依存していないため、新たなメディアの報道スタンスの判定ができると考えられる。Twitter ユーザーがそのニュースメディアの URL をツイートする限り、新たなメディアが登場したとしても、報道スタンスを判定することが可能ではないかと考えている。そのため、新たなメディアに対するスタンス分類の検証を今後の課題として更なる研究を行っていきたい。

謝 辞

本研究の一部は JSPS 科研費 19K11982 の助成を受けたものである。

文 献

- [1] D. Reinsel and J. Rydning, “Worldwide global storage forecast, 2020–2024: Continuing to store more in the core,” IDC: The Premier Global Market Intelligence Company, International Data Corporation (IDC), 2020.

- [2] P. R. Center, "Political polarization, 1994-2017," *Pew Research Center*, 2017.
- [3] M. Jurkowitz, A. Mitchell, E. Shearer, and M. Walker, "Us media polarization and the 2020 election: A nation divided," *Pew Research Center*, vol. 24, 2020.
- [4] W. Magdy, K. Darwish, N. Abokhodair, A. Rahimi, and T. Baldwin, "# Isisnotislam or# deportall-muslims? predicting unspoken views," in *Proceedings of the 8th ACM Conference on Web Science*, 2016, pp. 95–106.
- [5] M. Kutlu, K. Darwish, and T. Elsayed, "Devam vs. tamam: 2018 turkish elections," *arXiv preprint arXiv:1807.06655*, 2018.
- [6] K. Darwish, P. Stefanov, M. Aupetit, and P. Nakov, "Unsupervised user stance detection on twitter," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 141–152.
- [7] A. Hanselowski, A. PVS, B. Schiller, *et al.*, "A retrospective analysis of the fake news challenge stance detection task," *arXiv preprint arXiv:1806.05180*, 2018.
- [8] L. Mascarell, T. Ruzsics, C. Schneebeli, *et al.*, "Stance detection in german news articles," in *Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER)*, 2021, pp. 66–77.
- [9] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, "Predicting factuality of reporting and bias of news media sources," *arXiv preprint arXiv:1810.01765*, 2018.
- [10] M. Yoshioka, M. Jang, J. Allan, and N. Kando, "Visualizing polarity-based stances of news websites," *NewsIR@ ECIR*, vol. 2079, pp. 6–8, 2018.
- [11] 藤兼由生, 風間一洋, 吉田光男, and 土方嘉徳, "グラフの分極度によるマスメディアとtwitterユーザの意見の偏りの分析," in 第 13 回データ工学と情報マネジメントに関するフォーラム (DEIM2021) , I25-4, 2021.
- [12] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *ArXiv e-prints*, Feb. 2018. arXiv: 1802.03426 [stat.ML].
- [13] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [14] Y. LeCun, "The mnist database of handwritten digits," <http://yann.lecun.com/exdb/mnist/>, 1998.
- [15] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, P10008, 2008.
- [16] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.
- [17] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. 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.
- [18] G. Csardi, T. Nepusz, *et al.*, "The igraph software package for complex network research," *InterJournal, complex systems*, vol. 1695, no. 5, pp. 1–9, 2006.
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [20] L. McInnes, J. Healy, N. Saul, and L. Grossberger, "Umap: Uniform manifold approximation and projection," *The Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.
- [21] R. Brownstein, "Red and blue america aren't experiencing the same pandemic," *The Atlantic*, Mar. 2020. [Online]. Available: <https://www.theatlantic.com/politics/archive/2020/03/how-republicans-and-democrats-think-about-coronavirus/608395/>.
- [22] A. Chilton, K. L. Cope, C. Crabtree, and M. Versteeg, "The normative force of higher-order law: Evidence from six countries during the covid-19 pandemic," Available at SSRN 3591270, 2020.
- [23] D. Bauder, "Conservative media offers mixed messages on covid-19 vaccine," *AP News*, Jul. 2021. [Online]. Available: <https://apnews.com/article/joe-biden-health-government-and-politics-arts-and-entertainment-media-74f36fead348661aeba170422381a972>.