

ポピュラリティ推定に基づいた SNSにおけるニュースの中立的な理解支援

川口 天佑[†] 牛尼 剛聡^{††}

[†]九州大学大学院芸術工学府 〒815-8540 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 福岡市南区塩原 4-9-1

E-mail: [†]2DS16087N@kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

あらまし 近年, SNS を通してニュース情報を取得する人々が増加している. しかし, SNS の情報環境はコミュニティによって偏りが生じておりフィルターバブル的な状態であるため, SNS 上で閲覧したニュースやそれに対する反応の中立性が偏っている場合, その偏りの影響を強く受けて自らの中立的観点を損なう可能性がある. 本研究では, この問題を解決するために, ニュースや反応の中立性を推定しユーザに提示するシステムが必要だという考えに基づいて, ニュースの中立性を推定するために有用な指標である「ポピュラリティ」を提案する. 本手法では, Twitter 上でニュースに反応したユーザにおけるニュースの話題に対する日常的な関心の強さと, それに基づく反応ユーザの分布傾向の違いに着目し, これを利用してポピュラリティを推定する.

キーワード SNS, Twitter, ニュース, フィルターバブル, 中立性, ユーザ, 反応, 興味関心

1. はじめに

近年, ニュース情報の取得において SNS を利用するユーザが, 特に若い世代を中心に増加している. これは新聞社や TV 局などの様々なマスメディアが SNS アカウントを作成し, 自社の記事のタイトルや URL の投稿を行うことでその流通を促進していることが主な要因であると考えられる.

一方で, 社会の情報化にともない, メディアリテラシーの重要性がよく取り上げられている. テレビや新聞のみではなく, SNS も一種のメディアと化した近年, 若い人々を中心とする SNS ユーザにとって, 拡散・提供されているニュース記事を読む時には, その情報に振り回されず, 主体的に読み解く力や, 必要になっている. そのためには, ニュース自身の内容が偏っている可能性や, またニュースについての他人の意見が偏っていないかを常に考慮し, 自身の中立的観点を保つことが重要であると考えられる.

しかし, SNS でニュースを閲覧する場合, ユーザの中立的な観点を保つ上である問題が存在している. それはフィルターバブル [1] と呼ばれる問題である. フィルターバブルとは, インターネットを通して情報を取得するとき, Web ブラウザや Web サービスが, ユーザの興味や嗜好を学習することで, ユーザの意図にかかわらず自分の嗜好や思想に沿った特定の傾向に偏った情報のみが提供され, 情報環境の中で孤立する現象のことである. SNS では, ユーザの意図的な選択に基づくネットワークが構築されるため, 情報環境という視点でそれを見たとき, フィルターバブル的であると十分に考えられる.

そして, このフィルターバブル下では, ニュースに対するユーザの中立的な観点を保つ上で大きく分けて2つの問題が存在すると考えられる.

第一の問題は, 中立性に偏りがあるニュースの存在である.

ニュース記事を発信するマスメディアには, それぞれの立場や見方があり, 中立的でない場合が存在する. 中立的でないニュースとフィルターバブル的な環境が組み合わさると, 大きな問題となる場合がある. SNS 上では, 自身と繋がっているユーザが記事をシェアすることでニュースが流れてくることがある. しかし, これこそがフィルターバブルであり, 自身のコミュニティに属する人々がフィルターの役割を果たし, ユーザの意思に関わらず特定の観点到偏ったニュースばかりを流通させることで, ユーザの中立的観点が歪んでいく可能性がある.

第二の問題は, ニュースに伴う他ユーザの反応 (コメント) の意見の偏りである. SNS 上でニュースに反応するユーザは, テレビ放送でのニュース番組におけるコメンテーターの役割を果たし, ニュースを読むユーザのニュース理解に影響を及ぼすことが考えられる. フィルターバブル下では配信されるニュースだけではなく, 付加されている反応が偏っている場合があると考えられ, ユーザはニュースに対する同じような傾向に偏った反応を目にし続けることで, ユーザ自身のニュースの捉え方が中立的でなくなる可能性がある.

このことから, SNS では, ニュースの中立性やニュースの反応に対する中立性を判断するために役立つ情報をユーザに示すことが重要である. そこで, 本論文では, SNS で配信されるニュースの中立的な理解を支援するために, ニュースとニュースの反応の中立性の推定手法を提案する. 提案手法では, ニュースへの反応を行ったユーザの特徴に注目する. ニュースに対して反応するユーザには大別して2種類のユーザが存在すると考えられる. 一つは日常的にそのニュースの話題に関心があるために反応を行ったユーザ, もう一つは, 日常的にそのニュースの話題への関心は低いにもかかわらず, そのニュースに対して反応を行ったユーザである.

本研究では, ニュースに対する反応ユーザの日常的な関心の

強さが異なり、ニュースによって日常的な関心が高い反応ユーザと低い反応ユーザの分布が異なることに注目した。この日常的関心度に基づくユーザの分布の違いが、ニュースの中立性とニュースに対する反応と中立性の相関に着目し、この日常的関心度に基づく反応ユーザの分布の傾向を示す指標をポピュラリティと呼ぶ。本論文では、ポピュラリティに基づくニュース及び反応の中立性の推定手法の提案とその有効性を検証するための実験により評価を行う。

本研究では、対象とする SNS として、代表的なマイクロブログサービスの一つである Twitter を利用する。Twitter では「フォロー」機能により、ユーザが閲覧したいアカウントを選択でき、他ユーザ間とのネットワークが構築される。さらに、Twitter ではユーザの情報を分析することでユーザやツイートを推薦する機能も備わっているため、フィルターバブル的な情報環境が存在すると考えられる。

また、新聞社や TV 局をはじめとしたマスメディアも Twitter のアカウントを取得し、自社の Web サイトのニュースページのタイトルや url を載せたツイート (以下、「記事ツイート」と呼ぶ) を投稿している。Twitter は「リツイート」と呼ばれるシェアを行う機能や、「いいね」といった共感などを示す反応を行う機能も提供しているため、一般のユーザはマスメディアの記事ツイート投稿に対して反応できる。さらに、Twitter では、ニュースの url とそれに対する自分の意見を付与したコメント付きツイートが活発に投稿されている。本研究では、このようなニュースに対する意見が付与されたツイートを明確な意志を持ったニュースへの反応だと考え、これらを「反応ツイート」とし、これを行ったユーザを「反応ユーザ」とする。

本論文では、Twitter 上でマスメディアから投稿された記事ツイートとそれに反応ツイートを行った反応ユーザのデータを収集し、これらに基づくニュースと反応の中立性推定手法の提案、及びその評価を行う。

2. 関連研究

本研究では、SNS 上に生じるフィルターバブルという問題の結果、ニュースを閲覧するユーザの中立的な観点が失われる問題を解決することを目的としているが、本研究と同様に、情報の中立性を重視し、フィルターバブル問題の解決を目的とする研究がいくつか存在する。

Hannák ら [2] は、パーソナライズの精度の上昇に伴うフィルターバブル問題に対して、Web 検索の個人化レベルや個人化に利用されるユーザ属性の強さを示す指標がないことに注目した。彼らは、この問題に対して、個人化の程度を測定する手法を提案している。Waldner ら [3] は、Twitter において、ユーザが興味のない対象を排除するのではなく、利用者の興味程度によって表示方法を変えることで、推薦システム上では排除されてしまう情報を残し、利用者にとって潜在的に重要である可能性のある情報に触れる機会を残すシステムを構築した。Sayooran ら [4] は、Facebook 上での個人化がブラックボックスである問題を解決するために、フィルターバブルのイメージに基づき、フィルタリングされている情報を泡状の UI を用い

て視覚的に提示することでシステムへのユーザの信頼性を高める手法を提案している。

神寫ら [5] は、推薦システムにおけるフィルターバブル問題を解決するため、ある特定の観点に関して中立性を高めるように推薦を行うシステムを提案している。片岡ら [6] は、ユーザが接触している情報の個人化の程度を簡単なインジケータを用いてユーザに認知させることで、ユーザが自らフィルターバブル問題を自覚して解決させる UI システムを提案している。

本研究と同様に、ニュースに焦点をあて、ニュース位に対する中立性や多角的な視点を得ることを重視した、ニュース理解の支援を目的とした研究も存在する。Park ら [7] は、ニュースメディアの欠陥として、ニュース製作過程における偏見を問題として取り上げ、それが読者に対して与える悪影響に対して NewsCube という新しいニュースサービスを提案することでその解決を試みている。NewsCube は、関心のあるニュースの出来事に対して複数の視点を自動的に作成し、迅速に提供することを目的としている。切通ら [8] も、同様の問題に対して、ユーザが閲覧しているニュースのイベントに対して様々な角度から理解することを支援するために、意見、視点、詳しさを 3 尺度において最も差異が大きいニュースを提供するニュースアプリ「NewsSalad」の開発研究を行っている。また、青木ら [9] も、同じトピックを扱うニュースでも、新聞社ごとに報道内容に違いがあることに着目し、あるトピックに対する複数の新聞社ごとの編集意図の抽出をすることでそれらの問題を解決する手法を提案している。

また、本研究では、ニュースの中立的な理解支援を行うために、ニュース自体の分析ではなく、ニュースに反応している SNS ユーザに着目した。同様に SNS やブログのユーザなどから得られる情報を利用して、ニュースの理解支援を行うことを目的とした [10]-[13] のような研究もある。

本研究では、フィルターバブルによって生じている問題の中でも特に、SNS 上のニュース自体とそのニュースに対する反応の中立性の偏りによって生じる可能性のある、ユーザが中立的な観点を歪むを解決すべき問題としている。そこで、ニュース及び反応の中立性をユーザに提示することでフィルターバブルの影響を回避し、問題の解決ができると考えた。また、ニュースに対する反応ユーザとそのユーザのニュースの話題に対する日常的な関心の強さと、ニュースによって、反応ユーザの日常的な関心度に基づく分布が異なることに着目した。これがニュースや反応の中立性と関係がある可能性に注目し、これを利用したニュース及び反応の中立性の推定手法を開発する。

3. ポピュラリティ抽出のアプローチ

SNS 上でユーザがニュースを見るとき、ニュースのタイトルや見出しだけではなく、そのニュースが得た「シェア」や「いいね」などの反応の数もそのニュースの持つ情報の一つとして取得する。ニュースに対する「シェア」や「いいね」の数が多ければ、多くの人々に閲覧された印象をユーザに対して与えるが、本当にそれが正しい印象であるとは限らない。

ニュースが多くの反応を得ている場合でも、幅広い様々な層

のユーザから得た反応なのか、特定のコミュニティや特定の層のユーザの間で集中的に注目されたために得られた反応なのかは、反応数からだけではわからない。前者であれば問題はないが、後者の場合、そのニュースに対して特定の層からの偏った反応が集中している可能性が高い。また、そのようなニュースは中立的でない可能性がある。例えば、偏った視点のニュースは、異なる視点の層の多くのユーザから批判されたり、逆に過剰な同調を受けてその反応数が増加することが考えられる。

よって、ニュースに対する反応数だけから、ニュースと反応の中立性を推定することは難しいと考えられる。しかし、ニュースに対する反応ユーザには、そのニュースの話題に対して、日常的に関心を持っているために反応したユーザもいれば、普段はあまり関心がないが、そのニュースに対しては反応を行ったユーザもいるなど、その日常的な関心度の大きさに違いがある場合がある。

ニュースによっては、日常的関心が高いユーザが多く反応している場合もあれば、日常的関心が低いにもかかわらずユーザが多く反応している場合もあり、ニュースによって反応するユーザの特徴の分布が異なることに着目し、この違いがニュースやその反応の中立性と相関があると考えた。

本研究では、ニュースに対する反応ユーザの日常的な関心の強さを「日常的関心度」と呼び、反応ユーザの日常的関心度に基づく分布の傾向を示す指標を「ポピュラリティ」と呼ぶ。日常的関心度が低いにもかかわらず反応したユーザが多いニュースは、幅広い層のユーザに対して印象を与えたニュースでありポピュラリティが高いと考えられる。逆に、日常的関心度が高いユーザばかりに反応されたニュースは、普段から関心が高い特定の層の人々からの反応が集中しており、ポピュラリティが低いと考えられる。

本研究では、ポピュラリティの抽出手法について述べ、中立性との相関を検証する。

反応ユーザの日常的関心度の推定においては、ユーザが普段 SNS で行っている投稿内容を利用する。SNS はブログ的な機能を有しているものが多く、その投稿の内容はユーザが普段から関心を抱いているものであると十分に考えられる。よって、ニュースの話題と反応ユーザの日常的な投稿の内容を比較することで、日常的関心度の推定を行う。

以上から、本論文においてポピュラリティの抽出手法は、以下の3つの段階に分けられる。

- (1) 対象ニュースへの反応ユーザのデータの収集
- (2) 反応ユーザの日常的関心度推定
- (3) 反応ユーザの日常的関心度に基づくポピュラリティ抽出

以上を踏まえ、第4節において、具体的なポピュラリティの抽出手法について提案する。

4. 提案手法

本研究では、対象とする SNS を Twitter とし、Twitter 内で拡散されたニュース記事に対し、そのニュースに対する反応を行ったユーザのデータを利用してポピュラリティ抽出を行う。

以下に、ポピュラリティの抽出を行うための具体的な流れを説明する。なお、本研究では手法で必要となるツイートなどの Twitter 上のデータの取得のために Twitter API を利用する。

4.1. 反応ユーザのデータの収集

本研究では、対象とするニュースのポピュラリティの抽出のために、対象とするニュースに対する反応ユーザを利用する。具体的には、マスメディアが投稿したニュースのタイトルと url を含むツイートに対して、反応を行ったユーザ及び彼らのツイートを収集する。

Twitter 上では様々な反応を行うことができる。反応は大きく分けて、リツイートや「いいね」のようにアイコンに触れるだけで行えるシンプルな反応と、リプライや引用ツイートのような自分の意見の付与ができるコメント付きの反応の2種類に分けられる。本論文では、後者の反応が前者と異なり、ツイートとしての形を持つ反応であることから「反応ツイート」と呼ぶ。また今回、収集の対象として扱う反応ユーザは、リツイートや「いいね」のような、シンプルな反応を行ったユーザではなく、明確に自らの意思を伴った反応であると考えられる記事ツイートをを行ったユーザのみとした。以下、「反応ユーザ」とはこの「反応ツイート」を行ったユーザのみを指す。

反応ツイートの収集方法について、Cao ら [14]、邱 [15] らによる研究がなされている。しかし、今回は確実な反応ユーザの抽出を目指すために、記事ツイートとのリンクがあるリプライや引用ツイート、またニュース記事のサイトへのリンクがある url 付きのツイートを対象とする [16] の手法によって反応ツイートの収集を行う。

以下、具体的な3種類の反応ツイートについて説明する。

(a) 記事ツイートへのリプライ

Twitter 上では、すでに投稿されているツイートに対して、返答や意見を述べるツイートを投稿するために、リプライが利用される場合がある。リプライは Twitter ユーザに割り当てられるスクリーンネームと呼ばれる“@”から始まる文字列(例: @xxxxx)をツイート中に含んでいる。リプライの対象となるツイートは、個人の Twitter アカウントによるものだけではなく、マスメディアのアカウントから投稿されたツイートも対象になる。マスメディアの Twitter アカウントに対するリプライは、ニュース記事に対するユーザの明確な意思を伴ったツイートと考えられるため、反応ツイートとして収集の対象とする。

(b) 引用ツイート

Twitter の機能の一つとして引用ツイートという機能が存在する。これは、自分以外のユーザのツイートに対して、自分のコメントを付加した形で一つのツイートとして投稿を行うことができる機能である。

マスメディアによる記事ツイートを引用した引用ツイートもその記事に対するユーザの反応が含まれているツイートと考えられるため、反応ツイートとして収集の対象とする。

(c) 記事 URL 付きツイート

(a) 及び (b) の反応は、マスメディアが発信した記事ツイートに対するユーザの反応である。一方、記事 URL 付きツイートは、それらとは少し異なる。

記事 URL 付きツイートは、そのほとんどがマスメディアのニュースサイトの記事ページに設置されている Twitter のツイートボタンによって投稿されたものである。記事ページのツイートボタンを押すと、記事タイトル及び URL が本文に自動的に含まれているツイート投稿画面が表示される。ニュースを閲覧したユーザはこの機能を利用することで、自分のフォロワーに対し、ニュースを簡単に紹介することができる。

また、この際に、自動的に含まれるタイトルと URL だけではなく、ユーザ自身の意見や感想を書き込みツイート投稿も可能である。このことから、ユーザの意見感想を含む記事 URL 付きツイートは、反応ツイートとして収集の対象とする。

以上の反応ツイートのデータは、Twitter API を通して JSON や XML などの形式で得ることができる。これには、ツイート本文だけではなく、ツイートを行ったユーザの ID なども含まれる。つまり、対象となるニュースの記事ツイートへの反応ツイートを収集することで、反応ユーザのデータを収集することができる。この反応ユーザの ID を利用して、反応ユーザが投稿したツイートを収集できる。今回、反応ユーザの日常的関心度を推定するために、反応ユーザのツイートを利用する。具体的には、ユーザ 1 人につき 200 ツイートを取得し、ユーザの日常的関心度推定に利用する。

4.2. 反応ユーザの日常的関心度推定手法

本研究では、ユーザが Twitter 上で日常的に投稿しているツイートがユーザの日常的関心を表すと考え、ニュースの内容と比較することで、反応ユーザのニュースの話題に対する日常的関心度を推定する。推定するために、まず、収集した反応ユーザのツイート 200 件を結合して 1 つの文書とし、これをユーザの日常的関心を表す文書とする。これと対象のニュースの本文との文書の類似度を求め、それをそのユーザの日常的関心度とする。

ユーザのツイートとニュース本文の文書の類似度を求めるために、文書の特徴ベクトル化する必要がある。ベクトル化のために、まず対象となる文書の形態素解析を行う。形態素解析には MeCab [17] [18] を用いる。形態素解析で得られた名詞を対象として、2 つの文書を Doc2vec [19] によりベクトル化する。

Doc2vec は、単語のベクトル化手法である Word2vec [20] を文書に拡張したものである。Doc2vec は膨大な量の文書データをコーパスとして学習することで、文書の文脈や単語の語順を考慮したベクトル化を行うことができる。本研究では、コーパスとして wikipedia の全記事を用いて、Doc2vec の学習モデルを生成する。生成された学習モデルを用いて、反応ユーザの日常的関心を表す文書 (ツイート) と対象のニュース記事本文のベクトル化を行う。

2 つの文書の特徴ベクトルの \cos 類似度を計算し、その類似度を反応ユーザの日常的関心度として利用する。対象記事 i の記事本文のベクトル \mathbf{av}_i とユーザ j の日常的関心ベクトルを \mathbf{uv}_j を用いて、対象記事 i に対する反応ユーザ j の日常的関心度 EI_{ij} を以下の式で計算する。

$$EI_{ij} = \cos(\mathbf{av}_i, \mathbf{uv}_j) = \frac{\mathbf{av}_i \cdot \mathbf{uv}_j}{|\mathbf{av}_i| |\mathbf{uv}_j|} \quad (1)$$

反応ユーザのニュースに対する日常的関心度 EI は -1 から 1 の間の値となる。

4.3. ポピュラリティの計算

第 4.2 節で求めた反応ユーザの日常的関心度を用いてニュースのポピュラリティを計算する。

図 1 は、ある 2 件のニュースを対象とし、第 4.2 節で説明した手法を用いて計算した反応ユーザ全員の日常的関心度のヒストグラムである。横軸 (階級) は、日常的関心度であり、階級幅は 0.05 である。縦軸は、階級幅中のユーザ数の相対度数である。

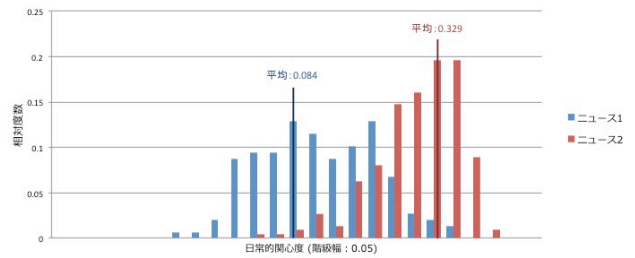


図 1 2 つのニュースに対する反応ユーザの日常的関心度に基づくヒストグラム

以下はニュース 1 及び 2 のタイトルとメディア名、報道日である。

- ニュース 1
 - 蓮舫代表が周囲に辞意を漏らす 民進関係者が明かす
 - (ライブドアニュース：2017/07/27)
- ニュース 2
 - 【蓮舫代表辞意】蓮舫氏「安倍晋三政権を追い込むだけでは…私たちが受け皿にならねば国民が不幸になる」
 - (産経ニュース：2017/07/27)

図 1 に示すように、ニュース 1 とニュース 2 に対する反応ユーザの日常的関心度に基づく分布は異なる。ヒストグラムの山の形状を見たとき、ニュース 2 の山はニュース 1 に比べ右寄りであり、ニュース 2 には日常的関心度が高い反応ユーザが多い。一方、ニュース 1 は、このニュースの話題に対して日常的関心度が低いにも関わらず、反応を行ったユーザがニュース 2 に比べて多い。つまり、ニュース 1 はポピュラリティが高いニュースであり、ニュース 2 はポピュラリティが低いニュースである。

本手法では、ニュースの反応ユーザの日常的関心度に基づく分布の特徴を示すために平均値を用いる。ニュース i に反応ユーザが n_i 人存在するとき、そのニュースに対するポピュラリティ P_i は、ニュース i に対する反応ユーザ j の日常的関心度 EI_{ij} に基づき、以下の式で計算する。

$$P_i = -\overline{EI}_i = -\frac{1}{n_i} \sum_{j=1}^{n_i} EI_{ij} \quad (2)$$

例として示した2つのニュースは、対象としている話題は同じであるが、読者が感じるニュース自体の中立性と反応の中立性に違いが生じる可能性がある。そして、その違いと反応ユーザの分布に関係があると考えられる。このことから、このポピュラリティとニュース及び反応の中立性には関連性が存在すると考え、その検証実験を次の第5節にて行う。

5. 実 験

今回、抽出したポピュラリティが、リツイートやいいね、またプライなどの反応数とは相関のない、独立した値であることを検証するための予備実験をおこなった。

また、ポピュラリティとニュースの中立性と反応の中立性との相関を検証する実験を行った。以下にその詳細を述べる。

5.1. データセット

今回の実験の対象としたニュース記事および反応ユーザ、また反応ユーザの日常的関心度推定で使用する Doc2vec の学習モデル作成に用いたデータの詳細を以下に述べる。

- ニュース記事：321 件
 - － 2017年7月20日から27日の間に Twitter 上でマスメディアのアカウントにより投稿されたニュース記事
 - － ニュース取得対象としたマスメディア (Twitter アカウント名：スクリーンネーム)
- (a) 朝日新聞 (asahi shinbun)：@asahi
- (b) 産経ニュース：@Sankei_news
- (c) NHK ニュース：@nhk_news
- (d) Yahoo!ニュース 速報や地震情報も：@YahooNewsTopics
- (e) ライブドアニュース：@livedoornews
 - － これらのニュース記事のうち、記事ツイートが Twitter 上で100件以上の反応ツイートを得たもの
- 反応ユーザ：93854 人
 - － 対象とするニュース記事に対して反応を行った Twitter ユーザ
 - － 各ユーザの日常的関心度を求めるためのツイート 200 件
- Doc2vec の学習モデル
 - － 学習対象：Wikipedia 全記事 1066856 件 (2016/12/02 時点)
 - － Wikipedia の記事一つを文書とし、本文を形態素解析、名詞のみを抽出
 - － 次元数を 100 に設定し、学習を行いモデルを作成

5.2. 実験 1 - ポピュラリティの独立性の検証

今回、ニュースの反応ユーザの日常的関心度に基づいて抽出を行ったポピュラリティが、Twitter 上で得ているリツイートの数や、「いいね」の数、また反応ツイートの数とは関係のない、独立した値であることを検証するための実験を行った。

5.2.1. 実験 1 - 実験内容

データセットのニュース記事 321 件全てに対して、提案手法に基づくポピュラリティの計算を行い、この値が、ニュースが Twitter 上で得たリツイート数、いいね数、反応ツイート数、以上3種類の反応数に影響されていない独立した値であること

を検証するために比較を行う。

5.2.2. 実験 1 - 実験結果

以下の図 2, 3, 4 は、横軸を提案手法に基づき算出したポピュラリティとし、縦軸をそれぞれリツイート数、いいね数、反応ツイート数として、ニュース 321 件をプロットした図である。なお、縦軸は3種類の反応数いずれの場合も対数軸を使用している。

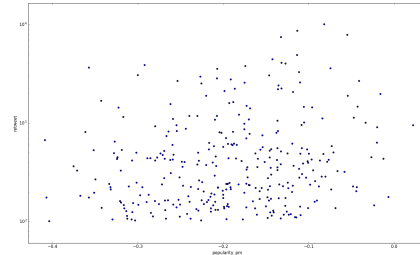


図 2 ポピュラリティとリツイート数によるプロット図

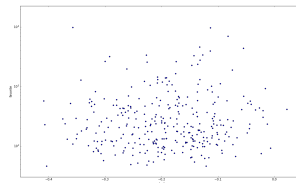


図 3 ポピュラリティといいね数によるプロット図

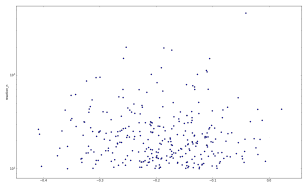


図 4 ポピュラリティと反応ツイート数によるプロット図

5.2.3. 実験 1 - 考察

図 2, 3, 4 から、ポピュラリティと3種類それぞれの反応数の間にはどれもほぼ相関がなく、今回、提案手法に基づき算出したポピュラリティが反応数から独立した値であることがわかる。つまり、反応数にかかわらず、ニュースに対する反応ユーザの日常的関心度に基づく分布は、ニュースごとに異なることがこの結果から示されたといえる。

5.3. 実験 2 - ポピュラリティと中立性との相関の検証

第4節で提案、抽出をおこなったポピュラリティと、ニュースの中立性と反応の中立性との間に相関があるかを検証するために、以下の実験を行った。

5.3.1. 実験 2 - 実験内容

被験者にアンケートを行い、対象のニュース記事の中立性と反応の中立性に関する印象を評価してもらった。これを正解データとして、提案する尺度の関連を調べた。

対象とするニュースは、データセットに含まれるニュース 321 件から抽出した 100 件である。抽出方法は、データセットのニュース 321 件全てに対して提案手法によるポピュラリティ推定を行った上で、推定ポピュラリティ値の順に並べたものを 100 段階に分割し、各段階からランダムに1つの記事を選択した。

アンケートは、クラウドワークス上で 100 名に依頼した。依頼者の内訳は、10代から60代以上の男女であった。

アンケートでは、ニュース記事を読んでもらい、そのニュースに対しての印象を尋ねる以下の2つの質問を行った。

- (1) このニュースに対する意見は、対立する複数の立場が存在すると思いますか？(はい/いいえ)
- (2) このニュースは「中立的」であると思いますか？(1～6の6段階評価)

ニュースに対する反応ツイートのうち、投稿された時間が早いツイート上位 15 20 件の反応を被験者に対して提示し、その反応の印象について以下の3つの質問を行い、6段階で回答してもらった。

- (3) これらの反応は「中立的なものが多い」と思いますか？
- (4) これらの反応は「多様性」があると思いますか？
- (5) これらの反応に対して全体的に共感できますか？

5.3.2. 実験2 - 実験結果

アンケートにより得た5つの評価値を、今回、個別の値としてそれぞれ一つづつ提案手法によるポピュラリティと比較しその相関の検証を行った。

質問(1)の回答は「はい/いいえ」の2択であるため、解答100件中の「はい」の割合を計算し、その値を質問1の評価結果とした。また、質問(2)～(5)は6段階の評価値であるため、解答100件の評価値の平均をそれぞれ計算し、それを質問に対する評価結果とした。

各質問の評価結果と提案手法によるポピュラリティの相関を検証するために、相関係数を求めた。相関係数には「ピアソンの相関係数」または「スピアマンの順位相関係数」を利用した。ピアソンの相関係数を用いるためには、相関を調べたい2種類のデータがそれぞれ正規分布に従っている必要がある。そのため、Shapiro-Wilk (シャピロ-ウィルク) 検定を用いて正規性の検定を行った。以下の表1にその結果を示す。question_x が質問xの評価結果、popularity_pm が提案手法に基づくポピュラリティの結果である。なお、有意水準5%にて検定を行った。Wはそのデータの検定統計量、p value はp値である。

表1 実験2: Shapiro-Wilk 検定の結果

	W (Test statistic)	p value
question_1	0.95	0.0011
question_2	0.98	0.057
question_3	0.96	0.0047
question_4	0.98	0.21
question_5	0.98	0.28
popularity_pm	0.99	0.58

Shapiro-Wilk 検定は、帰無仮説を「データが正規分布に従っている」、対立仮説を「データが正規分布に従っていない」と設定し、その結果のp値によって対象としたデータが正規分布に従っているかを検定できる。有意水準5%の時、得られたp値が0.05未満であれば、データが正規分布に従っているという帰無仮説が棄却されてしまう。p値が0.05よりも大きいならば、帰無仮説は棄却されず、そのデータは正規分布に従っていると判断できる。

今回、2, 4, 5及び提案手法に基づくポピュラリティは正規分布に従っていると判断されたが、質問1と質問3の評価結果はp値が有意水準0.05を超えず、帰無仮説が棄却され、この2

つは正規分布に従っていないと判断された。そこで、正規分布に従っていない場合でも用いることが出来るスピアマンの順位相関係数をすべての組み合わせに対して求め、正規分布に従っている質問2, 4, 5の評価結果とポピュラリティの組み合わせに関してはピアソンの相関係数による値をさらに求めた。

各質問の評価結果と提案手法に基づくポピュラリティ間の相関係数の値r及びp値を以下の表2に示す。

表2 各質問結果とポピュラリティ間の相関係数

	Spearman's rank r		Pearson's r	
	r	p value	r	p value
Q1 - popularity	-0.482	3.89E-07	-	-
Q2 - popularity	0.562	1.19E-09	0.572	4.67E-10
Q3 - popularity	0.525	2.12E-08	-	-
Q4 - popularity	0.102	0.313	0.106	0.293
Q5 - popularity	0.410	2.23E-05	0.389	6.35E-05

また、図5～9は縦軸を質問の評価結果とし、横軸を提案手法に基づくポピュラリティとしてニュースをプロットした図である。グラフ中に表示されているマーカー(点)はニュースを示している。なお、正解データおよび提案手法に基づくポピュラリティの値はともに平均0、分散1となるように正規化をおこなっている。

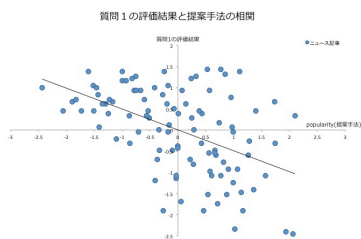


図5 質問1の評価結果とポピュラリティのプロット図

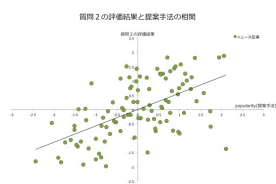


図6 質問2の評価結果とポピュラリティのプロット図

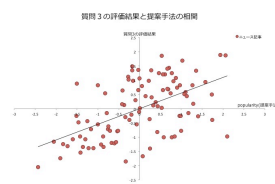


図7 質問3の評価結果とポピュラリティのプロット図

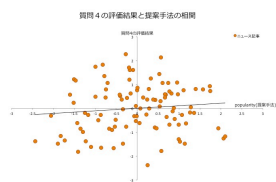


図8 質問4の評価結果とポピュラリティのプロット図

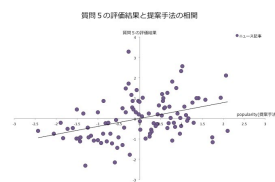


図9 質問5の評価結果とポピュラリティのプロット図

5.3.3. 実験2 - 考察

質問(1)は、設問のニュース本文を被験者に提示し、読んでもらった上で、そのニュースに対して、対立するような複数の立場の意見が存在すると思うかを「はい/いいえ」の2択で答えてもらう質問であり、「はい」の割合を評価結果の値としている。つまり、この値が高いほど、そのニュースには対立する複数の立場の意見が存在する。質問(1)への評価結果と提案手法によるポピュラリティはスピアマンの相関係数が $-0.482(p < 0.01)$ であり、有意な中程度の負の相関があるといえる。つまり、ポピュラリティが低いほど、対立するような複数の立場の意見は存在しない可能性がある。対立するような意見が存在するだろうと予見されるニュースは、ニュース自身の中立性に対して注意を払うべきであり、また、その反応に対しても片方の立場のみから偏った影響を受けないように気をつける必要があると考えられる。よって、ポピュラリティはユーザがニュースやその反応を見る際に、中立性について注意を払うための指標として有効であるといえる。

質問(2)は、設問のニュース本文を被験者に読んでもらった上で、そのニュースが「中立的」かどうかについて6段階の評価を行ってもらう質問であり、その評価値の平均を評価結果の値としている。つまり、この値が高いほどそのニュースは中立的であることを示している。質問(2)への評価結果と提案手法によるポピュラリティはピアソンの相関係数が $0.562(p < 0.01)$ であり、有意な中程度の正の相関があるといえる。つまり、ポピュラリティが高いほどそのニュースは中立性が高いということであり、本論文で提案したポピュラリティはニュースの中立性を示す指標として有効であることが示された。

質問(3)は反応の中立性を全体的に評価してもらう質問であり、(2)同様、その評価値の平均を評価結果の値としている。つまり、この値が高いほどニュースに対する反応は中立的なものが多く、逆に低ければ中立的な反応が少ない可能性が高い。質問(3)への評価結果と提案手法によるポピュラリティはスピアマンの相関係数が $0.525(p < 0.01)$ であり、有意な正の相関があるといえる。中立的な反応が多ければ、フィルターバブルによる反応の偏りをあまり気にせず済むが、もしも中立的な反応が少ないならば、反応を目にする際に注意を払う必要がある。ポピュラリティはニュースに対する反応の中立性を示す指標として有効であり、ユーザが反応の中立性を判断する際の助けとなる可能性がある。

質問(4)は反応の多様性を評価してもらう質問であり、その評価値の平均を評価結果の値としている。反応に多様性があれば、ユーザは一つの反応だけではなく、複数の反応を見て参考にすることでニュースを多角的に捉え、中立的な観点を維持しやすくなる可能性がある。しかし、今回は質問(4)への評価結果と提案手法によるポピュラリティはピアソンの相関係数は、有意な相関を認めることができなかった。ポピュラリティは、反応ユーザの日常的関心度の分布に基づくものであるが、その分布の違いによって反応の多様性を測ることは難しいとわかった。

質問(5)は反応全体に対しての共感度を評価してもらうものであり、その評価値の平均を評価結果の値としている。この評

価値が高い反応であるならば、それは全体の意見、世論に近い意見だと言えるので、フィルターバブルの影響を気にしすぎる必要はない。しかし、共感できない反応が多いならば、それはフィルターバブルにより世間とずれた偏った影響を受けてしまう可能性があるため、中立的観点を維持する上で問題となる。この質問への評価結果と提案手法によるポピュラリティは、ピアソンの相関係数が $0.389(p < 0.01)$ 、スピアマンの相関係数であれば $0.410(p < 0.01)$ であり、やや弱めではあるが、有意な正の相関があるといえる。強い相関ではないが、ユーザが反応に目を通す際に、自分以外の人々の共感度を知ることができれば、自身の中立的観点を意識するための指標として役立つため、有用である可能性がある。

これら5つの質問の評価結果との比較により、本研究で提案を行ったポピュラリティはニュース及びその反応の中立性に対して相関をもつことが示された。また、ポピュラリティは反応ユーザの日常的関心度の分布の傾向に基づく値であるが、この実験結果から、ニュースや反応の中立性は閲覧したユーザに対して影響を与え、中立性の違いにより、反応するユーザの傾向にも違いがでていると考えられる。そのため、反応ユーザの特徴を分析して得られたポピュラリティから逆にニュースや反応の中立性を推定することができたと思われる。以上より、ポピュラリティはSNS上でニュースを閲覧する際にその中立性を示す値として有効であると考えられる。

6. おわりに

我々は、近年、ニュースを取得する際に大きな影響を有するSNSにおいて、フィルターバブル的な環境から生じるユーザのニュースに対する中立的観点を保持するためにニュースの中立性と反応の中立性の推定が必要であると考えた。そこで、本論文では、「ポピュラリティ」と呼ばれる指標を提案し、それがニュースの中立性と反応の中立性の推定に有効であるかの検証を行った。

我々は、「ポピュラリティ」の抽出のために、SNS上に存在するニュースへの反応ユーザを、ニュースの話題への日常的関心度に基づいて特徴づけることができると考えた。そして、ニュースがSNS上での反応ユーザらの日常的関心度に基づくポピュラリティ抽出手法を提案した。ポピュラリティの推定は、反応ユーザの収集と、反応ユーザの日常的関心度推定、ニュースのポピュラリティ計算の3段階で行い、各段階の具体的手法について述べた。

また、提案する「ポピュラリティ」がニュースの中立性と反応の中立性の推定に有効であるかを検証するため、被験者アンケートに基づく検証実験を行った。5種類の質問を被験者アンケートで行い、うち4種類の評価結果とポピュラリティの間に相関があると示すことができ、本研究における提案手法が有効であると示した。

今後は、このポピュラリティを用いてニュースとその反応の中立性をSNSユーザに提示することで実際にフィルターバブル的な環境下においても、ユーザの中立的な観点を維持したニュースの理解支援を行うことが出来るシステムの実装を目

指す。

また、今回提案を行ったポピュラリティではニュースや反応の全体的な印象が中立的であるかそうでないかのみしか示すことができず、中立的ではない場合、どういう方向に偏っているかわからないため、ニュースの偏りの方向の分析や、また、反応1つ単位でもその中立性を測ることのできる手法が必要となると考えられる。

また、ニュースのみではなく、他のコンテンツに対しても、このポピュラリティを利用できる可能性についても検討する予定である。

文 献

- [1] E. Pariser, “The Filter Bubble: What The Internet Is Hiding From You”, Penguin UK, 2016.3.1 (イーライ・パリサー, 井口耕二 (訳), “フィルターバブル——インターネットが隠していること”, 早川書房, 2016.3.9)
- [2] A.Hannák, P.Sapiezynski, A.M.Kakhki, B.Krishnamurthy, D.Lazer, A.Mislove, C.Wilson, “Measuring Personalization of Web Search”, Proceedings of the 22nd international conference on World Wide Web, 2013
- [3] W.Waldner, J.Vassileva, “Emphasize, don’t filter!: displaying recommendations in Twitter timelines”, Proceedings of the 8th ACM Conference on Recommender systems, 2014.
- [4] N.Sayooran, J.Vassileva, “Providing awareness, explanation and control of personalized filtering in a social networking site”, Information Systems Frontiers, 2015
- [5] T.Kamishima, S.Akaho, H.Asoh, J.Sakuma, “情報中立推薦システム”, 人工知能学会全国大会論文集 26, 2012
- [6] 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
- [7] S. Park, S. Kang, S. C. J. Song, “Newscube: delivering multiple aspects of news to mitigate media bias”, CHI 2009, 2009
- [8] 切通恵介, 楠見考, 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
- [9] 青木伸也, 湯本高行, 角谷和俊, 新居学, 高橋豊, “論点に対する極性に注目したニュース記事からの編集意図の抽出手法”, Vol.2009-DBS-149 No.16, 2009.11.21
- [10] 見塚圭一, 鈴木優, 灘本明代, “Web ニュースの主題語に着目した解説ツイートの抽出手法の提案”, DEIM Forum 2017, 2017
- [11] 若井祐樹, 熊本忠彦, 灘本明代, “ニュースに対するつぶやきの感情分析”, 情報処理学会関西支部 支部大会, 2012.09.14
- [12] 田中翔也, 安藤一秋, “SNS での注目度と教育的知識に基づいた小学生のための Web ニュースランキング手法”, 第 14 回情報科学技術フォーラム (FIT2015), 2015
- [13] 佐藤大輔, 中川博之, 田原康之, 大須賀昭彦, “閲覧中のニュース記事に対するブログ記事から主張を抽出して提示するシステムの提”, 電子情報通信学会論文誌 2011/11 Vol. J94-D No.11,2011.11
- [14] X.Cao, K.Chen, R.Long, G.Zheng, and Y.Yu, “News comments generation via mining microblogs,”, In WWW 2012, 2012
- [15] 邱起仁, 樋山淳雄, “ニュース記事に関連するツイート収集手法の提案とその評価”, Vol.2014-HCI-157 No.57, 2014.3.15
- [16] 川口天佑, 牛尾剛聡, “Twitter ユーザの反応を利用した立場の異なるニュースの発見手法”, DEIM Forum 2016, 2016.3
- [17] 形態素解析エンジン MeCab, <http://mecab.sourceforge.net/>
- [18] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto, “Applying conditional random fields to japanese morphological analysis”, Proceedings of EMNLP 2004, 2004
- [19] Quoc Le, Tomas Mikolov, “Distributed representations of sentences and documents”, In Proceedings of the 31th In-

ternational Conference on Machine Learning(ICML 2014), 2014

- [20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, Jeff Dean, “Distributed representations of words and phrases and their compositionality”, Advances in Neural Information Processing Systems 26(NIPS 2013), 2013