

Twitter 上で配信されるニュースの偏りを考慮した推薦手法

池田 将[†] 牛尼 剛聡^{††}

[†] 九州大学芸術工学部芸術情報設計学科 〒815-8540 福岡県福岡市南区塩原 4-9-1

^{††} 九州大学芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]1DS15177N@s.kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

あらまし 近年の SNS の爆発的な普及により、SNS はニュースを流通させるサービスとして重要な役割を有するようになった。しかし、SNS 上ではニュースに対する立場に基づいた無意識的なコミュニティが存在するため、そのコミュニティ間で流通するニュースや情報に差異があり、ニュースを中立的に理解できない可能性がある。本論文では、Twitter でニュースを取得しようとするユーザに対し、ニュースに対する中立的な理解を支援するために、関心はあるが配信されにくいニュースを推薦する手法を提案する。本手法では、ニュースのツイートをリツイートしたユーザを反応ユーザとし、反応ユーザの一致度に基づいてユーザが関心を示す可能性があるニュースを抽出する。反応ユーザ群のコミュニティを推定し、コミュニティごとにニュースの配信されにくさを計算し、配信されにくいニュースを推薦する。提案手法では、推薦されたニュースを見ることで、発見性が高いニュースに触れ、ユーザにニュースの中立的な理解を促すことを目指す。

キーワード SNS, ニュース, Twitter

1 はじめに

近年の SNS の爆発的な普及により、SNS がニュースを流通させるサービスとして重要な役割を果たすようになった。例えば、代表的な SNS の一つである Twitter では、多くの新聞社や通信社が自社のアカウントを開設し、ニュース記事の見出しを投稿している。ユーザは興味のあるニュースがあった場合、そのツイートから参照される Web ページの記事を読むことができる。

一方で、SNS を利用して情報を取得する場合、新聞やテレビで情報を取得する時と違い、ユーザが見たいと思う情報かどうかで取捨選択を行い、自分にとって不快な情報をフォローしなくなる傾向があるため、TV や新聞の情報を取得するときより強くバイアスがかかる可能性があるという指摘がなされている [1]。また、検索システムや SNS ではユーザの登録された情報や閲覧履歴からユーザごとに異なる情報を提示するパーソナライズ技術も利用されている [2]。これに関して、パリサー [3] はフェイスブックにおいて自身と考える異なる保守系のユーザも友達登録し、彼らのリンクやコメントを確認することで、多様な情報を集めようとしたが、自身は保守的友人のリンクと比べて革新的友人のリンクや有名アーティストの最新動画のリンクをクリックすることが多いため、保守的友人のリンクがニュースフィードに表示されないという指摘をしている。このような、インターネットにおいて、ユーザが好ましいと思う情報ばかりが選択的に提示されることで、思想的に孤立するさまをフィルターバブルといい [4]、検索システムや SNS を利用する上での問題となっている。SNS ではフィルターバブルによる無意識的なコミュニティの住み分けがあり、配信されるニュースに偏りが生じることが考えられる。このため、Web 上には膨

大なニュースや意見の中からユーザがそれぞれの立場に関わらず効率的に多様なニュース、意見を集めるのは困難である。

このような情報の偏りを防ぎ、ユーザが興味のある出来事について多様な観点からニュースや意見を閲覧できるようにするため、ユーザが興味あるトピックに関して、ユーザの目に触れにくいニュースを推薦することは重要であると考えられる。なぜなら、ユーザに対して敢えて普段閲覧しない立場や主張の記事を提示し、普段から閲覧している記事と比較させることで、ユーザがニュースを中立的に認識するよう促すことができる可能性があるからである。そこで本研究では、ニュースのツイートをリツイートしたユーザを反応ユーザとし、反応ユーザの一致度を元にユーザが関心を示す可能性が高いニュースを抽出する。そして反応ユーザ群のコミュニティを推定し、コミュニティごとにニュースの配信されにくさを計算し、配信されにくいニュースを推薦する手法を提案する。提案手法では、推薦されたニュースを見ることで、発見性が高いニュースに触れることができ、ユーザにニュースの中立的な理解を促すことを目指す。

2 関連研究

本研究では、SNS でニュースを取得する際、フィルターバブル現象により、ニュースの中立的な理解が困難となっている問題の解決を目的としているが、本研究と同様にニュースやユーザの反応の中立性に着目し、それに対する支援や提案を行うような研究はこれまでもいくつか行われている。

神瀧ら [5] は情報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、ユーザが選んだ特定の視点や情報に対する中立性を保証する情報推薦システムを提案している。片岡ら [6] は Google 検索においてユーザが気付かないうちに閲覧する情

報に偏りが生じる問題を解決するため、ユーザにパーソナライズされた検索結果とそうでない結果との差をフィルターバブルの度合いとしてユーザに提示して認知させることで、ユーザにフィルターバブルの存在を自覚させ、ユーザが情報の探索領域を広げることを促すユーザ・インターフェースを提案している。

青木ら [7] は同じトピックに関するニュースでも見るものによって意見や印象が異なる出来事についての記事は新聞社によって報道内容に違いがあることに着目し、新聞各社の報道に関する編集意図を抽出する手法を提案している。

川口ら [8] は SNS の情報環境がコミュニティにより偏ることで、SNS 上で閲覧するニュースやそれに対する反応が偏り、ユーザの中立的観点が損なわれるという問題を解決するため、ニュースに対する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である「ポピュラリティ」を抽出し、ニュースや反応の中立性を推定する手法を提案している。

切通ら [9] はユーザがニュースイベントに対して様々な視点から理解するのを支援するためのニュースアプリケーション NewsSalad を提案している。このアプリケーションでは、ニュースを意見、視点、詳しさの 3 尺度で定量化し、ユーザに対し閲覧中の記事と同じイベントでかつその差異の尺度が最も大きい記事を提示する。

本研究は、これら既存の手法と異なり、Twitter など SNS 上で配信されるニュースに対して、ユーザが中立性を保つために、ユーザ自身の目に触れていない意見を知りたいと考えた時に、ユーザの反応と趣向を用いることにより、自身の考えとは異なる意見や論点を持つニュースを提示し、ユーザに比較させることで効率的にそれを行うことを目標としており、新規性がある。

3 アプローチ

Twitter は、ユーザが「ツイート」と呼ばれる 140 字以内の短い記事を書き込み、他のユーザがそれと呼んだり、返信したりすることでコミュニケーションが生まれるインターネット上のサービスである [10]。Twitter 上で他のユーザをフォローすると、自分とフォローしたユーザのツイートがタイムライン上に表示される。Twitter 上では、他のユーザをフォローするために承認を受ける必要がなく、自分が見ていない間もタイムラインが流れていくため、「ゆるいつながり」が生まれる。また、リアルタイムにニュースなどの情報を取得する際にも利用される。

Twitter において、ニュースの取得を目的の一つと考えるユーザは、複数のニューストピックに興味を示し、リツイートやリプライなどといった反応を行う場合がある。これに関して、あるニューストピックに対して反応を行なったユーザの多くは、関連性の高い別のニューストピックに対しても反応すると考えられる。例えば、国政選挙に関するニューストピックに対して反応を行なったユーザは、各政党の主張や動き、国会での出来事に関するニューストピックにも反応を行う可能性があると考えられる。これにより、あるニューストピックに関する反応から、そのニューストピックと関連性の高い別のニューストピックを

ユーザに対して推薦できるのではないかと考えられる。

本研究では、収集したツイート集合に対してクラスタリングを行い、同じニューストピックに関するツイートをニュースクラスタとしてまとめる。ニュースクラスタに含まれるニュースツイートをリツイートしたユーザをニュースクラスタの反応ユーザとし、対象とするニュースクラスタと他のニュースクラスタの反応ユーザの一致度を計算し、一致度が高いニュースクラスタを対象とするニュースクラスタの関連ニュースクラスタとする。

一方、ユーザー一人一人が個別の意見や主張を持っているため、あるニューストピックに反応したユーザたちも意見や主張によって、ある程度のコミュニティに分かれることが考えられる。Twitter などの SNS では、ユーザがフォローしているユーザなどの情報から新たにユーザがフォローする可能性があるユーザを推薦するため、自身が支持する主張を行ったり、自身が関心を示すニュースを共有するユーザを中心にフォローすると、そうでないユーザをフォローせず、それによって自身の考えと異なる主張やニュースが Twitter 上で見えにくくなり、ユーザの Twitter 上で取得するニュースに中立性が損なわれる可能性がある。

これを防ぐため、本研究では反応ユーザのフォロー関係からコミュニティを推定し、自分の所属するコミュニティとは別のコミュニティのユーザからよく閲覧されているが、自身のコミュニティからはあまり閲覧されていないニュースを推薦する手法を提案する。本手法では関連ニュースクラスタのコミュニティへの配信されにくさを算出し、配信されにくいとされる記事をコミュニティへ推薦する。

4 提案手法

本研究では、ユーザが興味のあるトピックではあるが、ユーザに配信されにくいニュースを、ユーザの反応を利用して推薦する手法の開発を目的とする。以下に本研究の提案手法の概要を示す。

4.1 データの収集

本研究では、ユーザに興味のあるトピックではあるが、ユーザの目に触れにくいニュースをユーザの反応を用いて推薦する。そのために、ニュースに関するツイート、それに付随するニュース記事、またニュースに対するユーザの反応を取得する。Twitter 上のデータ取得には Twitter API を使用する。図 1 に具体的なデータ構成を示す。本研究では、ニュースツイート、リツイート、反応ユーザ、反応ユーザのフォローをデータとして扱う。以下にデータの詳細とその収集方法について述べる。

4.1.1 ニュースツイートおよびニュース記事本文の収集

第 1 章でも述べたとおり、Twitter では多くの新聞社や通信社が Twitter アカウントを持ち、新聞社のウェブサイトに掲載された記事へのリンクを有するツイートを投稿している。本研究では、Twitter 上でユーザに配信される可能性が高いニュースを取得するため、新聞社のアカウントから発

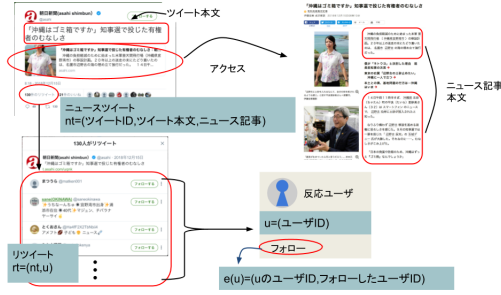


図1 データの構成

信されたツイートを利用する。また、そのようなツイートには元の記事にリンクするための url が添付されており、元の記事の本文についても収集を行う。ニュースツイートを $nt = (\text{ツイート ID}, \text{ツイート本文}, \text{ニュース記事})$ のように表現する。

4.1.2 反応ユーザの収集

Twitter には、様々な方法でユーザが反応を行うことができる。具体的には、ツイートに対するユーザの好みや関心を表す「いいね」、興味のあるツイートをフォロワーに共有する「リツイート」、ツイートに対してユーザの主張を述べる「リプライ」などがある。本研究では、コミュニティ内に流通するニュースツイートを対象にしているため、リツイートとそれを行なったユーザを反応ユーザとし、それを収集する。それぞれのユーザを $u = (u \text{ のユーザ ID})$ 、リツイートを $rt = (nt, u)$ のようにモデル化する。また、 nt のリツイートの集合を $RT(nt)$ のように表す。

4.1.3 反応ユーザのフォローを収集

Twitter では、ユーザは興味のある他のユーザをフォローすることで、他のユーザが投稿したツイートを自身のタイムラインに配信できる。一般に公開されているアカウントから発信されている情報であれば、ユーザ同士の友人関係に関わらず閲覧することが可能だが、Twitter ユーザは基本的に自身のタイムラインから情報を取得するので、ユーザが Twitter 上で閲覧している情報を推測するには、ユーザのフォロー関係の収集が有効である。本研究ではユーザの立場や主張に基づくコミュニティに着目し分析を行うため、収集した反応ユーザがフォローしているユーザ (フォロワー) を収集する。これら反応ユーザとフォローの関係性を $e(u) = (u \text{ のユーザ ID}, \text{フォローしているユーザ ID})$ のように表現する。また、ユーザ u のフォロワーの集合を $F(u) = \{u_1, u_2, \dots\}$ のように表現する。

4.2 ニュースのクラスタリング

新聞社のアカウントでは、毎日様々なニュースが配信されているが、その中には、同じ出来事に関する異なる記事に関するツイートや、続報に関するツイートが投稿されることがある。本研究で推薦するものは同じ事柄に関する違う記事ではなく、対象に関連する記事であるため、これらを区別する必要はない。そこで、ニュースのクラスタリングを行うことを考える。

ニュースをクラスタリングするために、ニュース記事をベクトル化する。ニュースツイートの文章は非常に短いため、他の

ものと明確な差を表現するため、ニュース記事の本文をベクトル化する。日本語の文章はスペースで単語が区切られていないため、記事の本文を形態素解析し、ニュース記事から単語を抽出する。形態素解析にはオープンソースの日本語形態素解析エンジンである MeCab[11] を利用する。

ニュース記事を Doc2Vec [12] によってベクトル化する。Doc2Vec は文や文書といった任意の長さのテキストから分散表現を得る手法である。膨大な量の文書データをコーパスとして学習しモデルを作成する。作成したモデルを利用し文書の文脈や単語の語順までを考慮に入れた特徴ベクトルを作成する。Wikipedia を教師データとし、Doc2Vec でモデルを作成する。ニュース記事を作成したモデルでベクトル化する。学習する際、ニュース本文のうちストップワードとならない動詞、名詞、形容詞、形容動詞以外を除去する。

作成した記事ベクトルを対象に DBSCAN [13] でクラスタリングを行う。DBSCAN は密度準拠のクラスタリング手法であり、半径、密度をあらかじめ設定しておき、ある一点から決められた半径の中で密度を満たしていればクラスタを形成し、それを繰り返すことでクラスタを成長させていく手法である。クラスタ間の距離の閾値 Eps とクラスタを構成する最小のデータ数 $MinPts$ の 2 つを持ち、ある点 x から距離 Eps 以内にある点集合を近傍 $N_{Eps}(x)$ と定義し、以下の接続関係を満たす時、同じクラスタに分類する。

$$y \in N_{Eps}(x) \quad (1)$$

$$|N_{Eps}(x)| \geq MinPts \quad (2)$$

作成されたクラスタは、 $nc = \{nt_1, \dots, nt_i, \dots\}$ のように表現する。

4.3 関連ニュースクラスタの抽出

対象とするニュースクラスタと関連するニュースクラスタを探検し、抽出する。本研究では、ニュースクラスタ集合 $NT = \{nc_1, nc_2, \dots\}$ に対して、「対象記事 nc_1 に関心を持ったユーザ u_1 がある別の記事 nc_2 に関してどれだけ関心を持つか」ということを表す関数 $interest(u, nc_1, nc_2)$ を定義する。関数 $interest(u, nc_1, nc_2)$ は、ニュース nc_1 の反応ユーザ集合 $RT(nc_1)$ とニュース nc_2 の反応ユーザ集合 $RT(nc_2)$ の類似度を元に以下のように計算する。

$$\begin{aligned} interest(u, nc_1, nc_2) &= Jaccard(RT(nc_1), RT(nc_2)) \\ &= \frac{|RT(nc_1) \cap RT(nc_2)|}{|RT(nc_1) \cup RT(nc_2)|} \end{aligned} \quad (3)$$

本研究は、ユーザが興味を持つ記事に関して、それに関連し、かつ配信されにくいニュースを推薦することを目指しており、興味を持つニュースと関連性が高いものを抽出する。

4.4 リンクに基づくクラスタリング

対象とするニュースクラスタとそれに関連するニュースクラスタに反応したユーザをフォロー・フォロワーの関係から、コミュニティを抽出するためにクラスタリングを行う。ソーシャルグラフからコミュニティを抽出する代表的な手法には、Girvan

と Newman の手法 [14] [15] がある。この手法は分割法であり、最短経路 *betweenness* をいう指標を利用する。*betweenness* は、コミュニティ間をつなぐエッジでは高くなり、コミュニティ内をつなぐエッジでは低くなる。Girvan らは、ソーシャルネットワークを $G = (V; E)$ (V : ノード集合, E : エッジ集合)、ネットワーク G に含まれる全てのノード対の集合を U 、 U の要素のうち最短パスが $e \in E$ を通るノード対の集合を U_e とした時の e の *betweenness* を表す B_e を以下のように定義する。

$$B_e = \frac{|U_e|}{|U|} \quad (4)$$

B_e の値が高いエッジを除去していくことで、コミュニティの分割を行う。クラスタリングの結果の良さを考察するため、クラスタ構造を評価する指標 *modularity* を導入する。コミュニティ分割の良さを「コミュニティ内とコミュニティ間の辺の割合」とし、この値 *Modularity* Q を以下のように定義する。

$$Q = \sum_{i \in L} (e_{ii} - (\sum_{j \neq i} e_{ij})^2) = \sum_{i \in L} (e_{ii} - a_i^2) \quad (5)$$

この式において、 L は全ネットワークのコミュニティの集合、 e_{ii} はネットワーク全体の辺の本数に対するコミュニティ $i \in L$ 内部のノード同士をつなぐエッジ数の割合、 e_{ij} はネットワーク全体の辺の本数に対するコミュニティ $i \in L$ からコミュニティ $j \in L$ に繋がっている割合、 a_i は e_{ii} はネットワーク全体の辺の本数に対するコミュニティ $i \in L$ から出ている辺の割合である。

また Newman は、Girvan と Newman の手法を改良し、全クラスタに対し、2 つのクラスタを結合した場合の *Modularity* である Q の増減 ΔQ を計算し、 Q の増加を最も大きくする 2 つのクラスタを結合していく手法を提案している [16]。この処理は Q が最大になるまでコミュニティの結合を繰り返す。 ΔQ は以下のように計算できる。

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad (6)$$

これらの手法では、グラフ内のノード、エッジの量が増えると計算量が膨大になるという問題がある。そこで、Clauset、Newman、Moore は貪欲法を利用して *Modularity* を計算する手法を提案している [17]。

本研究では、反応ユーザをノード、反応ユーザ間の相互フォローをエッジとしてソーシャルグラフを作成し、Clauset らの手法を利用してコミュニティ抽出を行う。ユーザ u の所属するコミュニティ (ユーザ集合) を $C(u) = \{u_1, u_2, \dots\}$ とする。

4.5 コミュニティへ配信されにくいニュースの抽出

ユーザ u_2 を介してニュース nc_1 に関心を持った利用者 u_1 に対して、ニュース nc_2 がどの程度配信されやすいかを $dist(u_1, nc_2)$ とする。ユーザ u_1 への配信されにくさは、 u_1 のタイムラインに nc_1 を流したユーザ u_2 の配信されにくさとして考える。また、同じコミュニティに属するユーザへのあるニュースの配信されにくさは同じであると見なし、コミュニティへのニュースの配信されにくさをユーザへの配信されにくさとする。コミュ

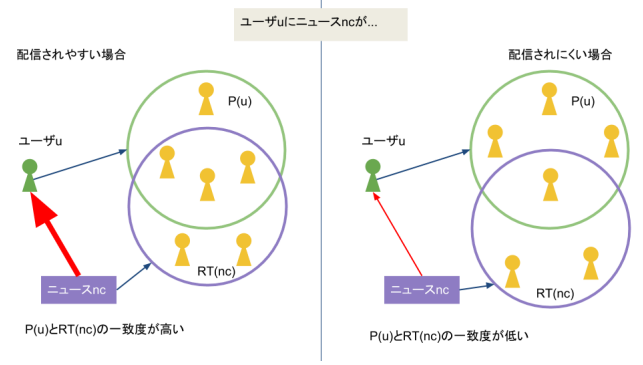


図 2 ユーザへのニュースの配信されにくさ

ニティ内でニュース nc_2 をリツイートした人の割合が低い場合、そのニュースはコミュニティ内であまり流通していないと考えられ、ニュース nc_2 はコミュニティにとって配信されにくいニュースであるということが考えられる。ユーザ u のフォロワーの所属するコミュニティに属するユーザの集合を $P(u) = \bigcup_{x \in F(u)} C(x)$ とし、 $dist(u, nc)$ は以下のように計算する。

$$dist(u, nc) = \frac{|P(u) \cap RT(nc)|}{|P(u)|} \quad (7)$$

この $dist(u, nc)$ の値が低いほど、ユーザ u にとって配信されにくいニュースであるとする (図 2)。

4.6 ユーザへ推薦するニュースの決定

ユーザが興味のあるトピックではあるが、ユーザに配信されにくいニュースを推薦するため、式 (4.1) と式 (4.5) を利用し、どのニュースをユーザに推薦するか決定する。ユーザ u とニュース nc_2 の適合度を $suitable(u, nc_2)$ とし、式 (4.1) と式 (4.5) を利用し、以下のように定義する。

$$suitable(u, nc_1, nc_2) = \frac{interest(u, nc_1, nc_2)}{dist(u, nc_2)} \quad (8)$$

ユーザ u には $suitable(u, nc_1, nc_2)$ が高い記事を推薦する。

5 実験

提案手法の有用性を確認するため、収集したニュース記事のクラスタリングと、ニュースクラスタの関連記事を抽出する実験を行なった。

5.1 実験手法

5.1.1 データセットの準備

対象とする新聞社の Twitter アカウントとして、毎日新聞 (@mainichi)、産経新聞ニュース (@Sankei.news)、朝日新聞 (@asahi) を利用した。2018 年の 9/16 から 12/16 までにこれらのアカウントから投稿されたニュースのツイートをすべて収集した。ニュースツイートは合計 20503 件収集できた。さらに、ツイートに掲載されている URL から新聞社の Web サイト [18] [19] [20] 上で配信されているニュース記事本文も収集した。適切に取得できなかったニュース記事を除くと、最終的に

17098 件のニュースを収集できた。また、収集した新聞記事のツイートをリツイートしたユーザの ID を収集した。収集された ID は 152589 個となった。

5.1.2 ニュースクラスタの形成

収集したニュース記事本文に対し MeCab を使って形態素解析を行い、文書ごとの単語集合を作成した。gensim [21] の Doc2Vec の実装を使用し、日本語版 Wikipedia の全記事を使い、PV-DM のモデルを学習した [22]。日本語版 Wikipedia 全記事の取得は Wikimedia Downloads [23] から行い、出現頻度が 7 未満の単語を含む文を削除した上で、モデルを学習した。この際、ベクトルの次元は 300 とし、イテレーション回数は 55 とした。学習したモデルを使い、ニュース記事のベクトル化を行った。

DBSCAN を用い、ニュースツイートのクラスタリングを行った。DBSCAN の半径を決定するため、半径を 1 から 0.2 ずつ大きくしていき DBSCAN のテストを行い、クラスタがいくつできるか、例外が何個存在するか、一つのクラスタが平均何個の記事で構成されているか調べた。

一つのクラスタに存在する記事の個数の平均が指定した密度よりも大きくなりすぎると、半径を大きくしすぎたためにクラスタの精度が低くなってしまうことが予想される。そのため、一つのクラスタに存在する記事の個数の平均が指定した密度よりも大きくなりすぎず、クラスタも多く形成できるパラメータとして半径を 11、密度を 2 とし、クラスタリングを行った。その結果クラスタが 1,447 個、クラスタに含まれない例外は 13,424 個だった。記事本文の収集あるいはベクトル化がうまく行えなかったニュースツイートが存在し、それに起因すると思われるほとんど意味を為さないクラスタが 1 つ存在し、そのクラスタに含まれるニュースツイートも例外として扱うことにしたため、最終的にクラスタが 1,446 個、例外が 13,940 個になった。クラスタに属されない例外とされた記事は今回は一つの記事で一つのクラスタを構成するものとして考える。

5.1.3 関連ニュースクラスタの抽出

提案手法を元に個々のニュースクラスタに関して他のニュースクラスタとの関連度を算出し、それを元に関連ニュースクラスタの抽出を行なった。他のニュースクラスタとの関連度を算出し、関連度の高いもの上位 30 件を関連ニュースクラスタとして抽出した。表 2 はニュースクラスタ 1,425 の関連ニュースクラスタのうち、反応ユーザの類似度が高い上位 5 件の関連ニュースクラスタである。

5.1.4 ユーザ集合のリンクに基づくコミュニティ抽出

提案手法を元に、収集した反応ユーザ集合に対してコミュニティ抽出を行なった。また、抽出されたコミュニティがどの程度妥当であるかを確認するために、評価実験を行なった。対象とするニュースクラスタの抽出評価を行うユーザに実際に配信されたニュースクラスタを実験の対象とするニュースクラスタとして抽出した。ユーザに配信されたかどうかは、評価を行うユーザがフォローしているユーザが、ニュースクラスタをリツイートしているかで判断した。コミュニティの抽出対象とするニュースクラスタと、それぞれのニュースクラスタの関連



図 3 被験者に提示したユーザリストとタイムラインの例

ニュースクラスタ 30 件に反応したユーザのフォローユーザを TwitterAPI で収集し、ソーシャルグラフの作成、コミュニティ抽出を行なった。ソーシャルグラフの作成、コミュニティ抽出には、NetworkX [24] を利用した。評価実験の手法得られたコミュニティのうち、ユーザへ推薦されたニュースの評価を行うユーザにニュースツイートを配信したユーザが所属しているコミュニティについて、その妥当性を確認した。ここでの妥当性とは、「コミュニティに所属しているユーザ同士が同じような出来事に興味をもったり、近い意見を持つか」と定義する。

評価実験では、対象ユーザとユーザが所属するコミュニティからランダムに選ばれた 10 人を合わせた 11 人を提案手法のユーザ集合、対象ユーザとニュースツイートに反応したユーザからランダムに選ばれた 10 人を合わせた 11 人をベースラインのユーザ集合とする。被験者に対し、「それぞれのリストに含まれているユーザについて、提示されているユーザのプロフィールや添付されている url からアクセスできるタイムラインをおおよそ 100 件程度見た上で、同じような出来事に興味があったり、近い意見を持っていたりするユーザのまとまりになっていると感じますか？」という質問をし、以下の 5 段階の尺度で提案手法とベースラインそれぞれに回答してもらった。

- (1) 全くまとまりがない
- (2) あまりまとまりがない
- (3) 少しまとまりがある
- (4) まとまりがある
- (5) とてもまとまりがある

またユーザには図 3 のようにリストに含まれているユーザのプロフィールを提示し、url からリストに含まれているユーザのタイムラインにアクセスしてもらった。

5.1.5 コミュニティへのニュースの配信されにくさの推定

提案手法に基づき、関連ニュースクラスタ 30 件のコミュニティへの配信されにくさを推定した。評価を行うユーザに対象となるニュースを流通させたユーザが所属するコミュニティへの配信されにくさを算出し、それを評価を行うユーザへの配信されにくさとした。

5.1.6 ユーザへ推薦するニュースの決定

算出した配信されにくさを元にユーザに推薦するニュースの決定を行う。ニュース nt が配信されたユーザ u に対して、 $suitable(u, nc_1, nc_2)$ の値が高いものから順に 5 つを推薦した。評価実験の手法被験者の過去のタイムラインに実際に現れた

表 1 ニュースクラスタの例

クラスタ ID	ツイート	新聞社	投稿日時
1425	岩屋防衛相、辺野古移設は「国民のため」 https://t.co/W8qelWWJzM →岩屋毅防衛相は、米軍普天間飛行場の名護市辺野古への移設について「日米同盟のためではない。日本国民のためだ」と述べた。	産経	2018/12/15 15:57:06
	辺野古移設「国民のため」 岩屋防衛相、抑止力強化 https://t.co/ZYX7lDzQhc	毎日	2018/12/15 21:15:06
1445	卓球・張本、シングルス 最年少V グランドファイナル https://t.co/YGxFn18Fuu	朝日	2018/12/16 20:20:04
	張本智和選手が初めてグランドファイナルを制しました。15歳172日での優勝は男女シングルスで大会史上最年少です。 https://t.co/bYSrFZ3med #卓球	朝日	2018/12/16 20:44:24
	卓球・張本、シングルス 最年少V 伊藤・早田組も優勝 https://t.co/YGxFn18Fuu	朝日	2018/12/16 21:52:03
1446	【速報】16日午後8時半ごろ、札幌市豊平区平岸の飲食店で爆発があった模様 https://t.co/G178usq2A9	朝日	2018/12/16 21:11:21
	札幌・豊平区の飲食店で爆発 建物倒壊・けが人の情報も https://t.co/tIvaPfsKZW	朝日	2018/12/16 21:22:04
	札幌の飲食店付近で爆発 建物倒壊・けが人の情報も https://t.co/tIvaPfsKZW	朝日	2018/12/16 21:56:05
	札幌の飲食店付近で爆発 建物倒壊・けが人数の複数人の情報も https://t.co/tIvaPfsKZW	朝日	2018/12/16 22:00:04
	札幌の飲食店付近で爆発 建物倒壊、20人以上が搬送 https://t.co/tIvaPfsKZW	朝日	2018/12/16 22:32:05

表 2 ニュースクラスタ 1425 の関連ニュースクラスタ (上位 5 件)

クラスタ ID	ツイート	新聞社	投稿日時	反応ユーザの類似度
14520	菅義偉官房長官、文大統領を牽制 徴用工判決「尊重」発言で https://t.co/L4QSLGGplv →「韓国政府に対し（判決による）国際法違反の状態を是正することを含め、適切な措置を講じることを強く求めているところだ」と述べた。	産経	2018/12/14 18:58:58	0.1
14797	松江市沖で無人の木造船が漂流しているのが発見されました。船首にはハングルが書かれていました。 https://t.co/2PX8ucr1b4	産経	2018/12/16 14:23:04	0.0997
13932	迎撃ミサイルの実験成功 日米共同開発 https://t.co/kTNPDkMhy →防衛装備庁は、日米が共同開発した改良型迎撃ミサイル「SM3ブロック2A」を使った発射実験を米国防総省がハワイ沖で日本時間11日に実施し、迎… https://t.co/SDjcop13dO	産経	2018/12/12 07:35:57	0.0966
14439	岩屋防衛相、辺野古の埋め立てに着手したと発表 https://t.co/psQPbtCZFX →岩屋防衛相「22年前に約束した普天飛行場の返還を今度こそ実現したい」「普天間飛行場の危険性を除去し、返還を実現するということも沖縄の民意だ」	産経	2018/12/14 12:08:30	0.0864
14207	ファウウェイなどを調達しないよう民間企業に要請する方針だとの一部報道について、菅官房長官は「現段階で予定はない」と述べました。一方、政府調達からは事実上排除します。 https://t.co/fxfVGEwHQit	産経	2018/12/13 13:49:04	0.0848

ニュースツイートを配信されたニュースツイートとし、それに対して $suitable(u, nc_1, nc_2)$ が高いものを被験者に提示し、評価実験を行なった。被験者一人当たり 5 件の配信されたニュースについて、推薦されたニュースの評価を行なってもらった。提案手法の他に、関連ニュースクラスタとして抽出された 30 件のうちからランダムで 5 件をベースラインとした。

質問 1 として配信されたツイートがユーザの興味とどの程度合致するか調べるため、「「タイムラインに現れたニュース」に関して、あなたはどの程度関心を持っているか」を 5 段階で評価してもらった。

質問 2 として推薦された提案手法とベースラインのそれぞれのニュースに対し、被験者が「どの程度関心を持っているか」を 5 段階で評価してもらった。

質問 3 として推薦された提案手法とベースラインのそれぞれのニュースに対し、「そのニュースが「タイムラインに現れたニュース」や、それに関する出来事をより中立的に (広い視点で) 理解するためにどの程度重要なニュースであると思うか」を 5 段階で評価してもらった。

5.2 実験結果

5.2.1 ニュースクラスタの形成

作られたニュースクラスタ 1,446 個のうち 100 個を無作為に抽出し、ニュースクラスタに含まれるニュースツイートが同じニューストピックについて扱ったものであるかを確認した。また、異なる新聞社によって配信されたニュース同士のクラスタが何個あるかを確認した。

結果として 100 件全てで同一のニューストピックについて扱ったニュースツイートがまとめられていることを確認した。また、異なる新聞社によって配信されたニュースどうしのクラスタが 5 個あることも確認した。

5.2.2 コミュニティの妥当性

九州大学芸術工学部の学生男女 7 人を被験者として、評価実験を行った。評価実験で得られた提案手法とベースラインの評価値の平均を図 4 に示す。

提案手法の評価値の平均は 2.637、ベースラインの評価値の平均は 2.648 であった。また、得られたデータに対して t 検定を行ったところ、 $p=0.965$ であった。考察以上の実験結果より、

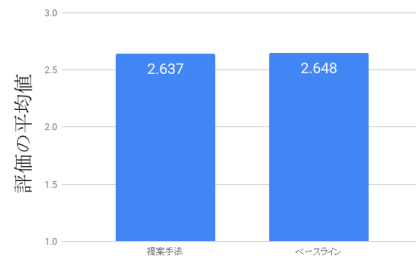


図 4 コミュニティの評価

提案手法のコミュニティがベースラインに対して適切なコミュニティであるということを示すことができなかった。適切なコミュニティが抽出できなかった原因として、ユーザ同士の関係についてフォロー関係のみを用いたことが考えられる。同じような出来事に興味を持っていたり、近い意見を持っていたりするユーザ同士でコミュニティを抽出するには、ユーザの意見をコミュニティに反映するためのユーザの反応などを用いる必要があると考えられる。

5.2.3 ユーザへ推薦するニュースの決定

九州大学芸術工学部の学生を中心に男女 6 人に対し、評価実験を行なった。なお、この評価実験では被験者ごとにパーソナライズ化されたデータを元の実験を行なっているため、被験者によって回答しているニュースが異なる。

質問 1 の「「タイムラインに現れたニュース」に関して、あなたはどの程度関心を持っているか」について、それぞれのニュースに対するユーザの評価の結果を図 5 に示す。

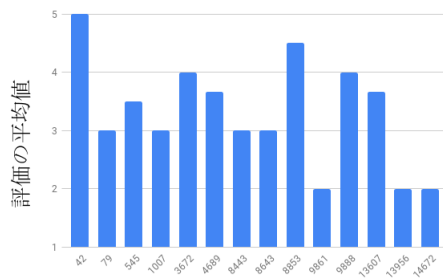


図 5 質問 1 の結果

質問 2 の推薦された提案手法とベースラインのそれぞれのニュースに対し、被験者が「どの程度関心を持っているか」について、評価実験で得られた提案手法とベースラインの評価値の平均を図 6 に示す。

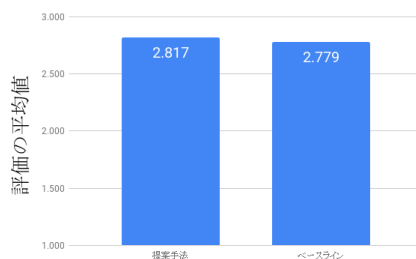


図 6 質問 2 へのユーザの評価

提案手法の評価値の平均は 2.817、ベースラインの評価値の平均は 2.779 であった。また、得られたデータに対して t 検定を行ったところ、 $p=0.767$ であった。

質問 3 の推薦された提案手法とベースラインのそれぞれのニュースに対し、被験者が「そのニュースが「タイムラインに現れたニュース」や、それに関する出来事をより中立的に (広い視点で) 理解するためにどの程度重要なニュースであるか」について、評価実験で得られた提案手法とベースラインの評価値の平均を図 7 に示す。

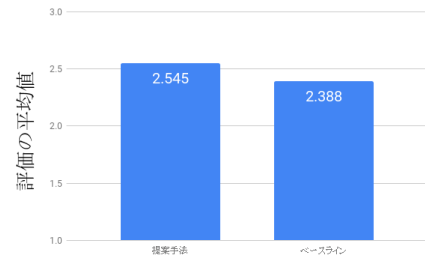


図 7 質問 3 へのユーザの評価

提案手法の評価値の平均は 2.545、ベースラインの評価値の平均は 2.388 であった。また、得られたデータに対して t 検定を行ったところ、 $p=0.137$ であった。

5.2.4 考察

質問 2 の実験結果より、ベースラインと比較して、提案手法が被験者に対して、関心のあるニュースを推薦することができたとは言えなかった。また、質問 3 の実験結果より、ベースラインと比較して、提案手法が「タイムラインに現れたニュース」や、それに関する出来事をより中立的に理解するためのニュースを推薦できているとは言えなかった。これらの原因として、コミュニティの分割が適切に行われていないことが考えられる。今回の実験で、ユーザ間をつなぐ関係として、フォローのみを用いたが、ニュースに対するユーザの意見などに関するコミュニティを抽出する場合、この関係が適切でなかったことが考えられる。そのため、引用リツイートやプライなど、ユーザの主張をより直接的に捉えられる反応を利用して、コミュニティ抽出を行うことで、より適切な推薦を行うことができると考えられる。また、提案手法の指標を改善することによってより適切な結果を推薦するように改良する必要がある。

5.3 追加実験

実験で有用な結果が得られなかったため、追加実験を行なった。12 月 14 日の「普天間基地の移設に伴い、政府が名護市辺野古沿岸部に土砂の投入を開始した」というニュースについて、朝日、産経、毎日のニュースツイートの反応ユーザを取得し、それぞれのユーザの自己紹介文 (description) を基準にコミュニティ分類を行い、それぞれのコミュニティへ推薦するニュースの推定を行った。コミュニティはそれぞれのユーザの政治への興味やニュースへの立場を基準に 5 つに分類した。推薦するニュースは、元のニュースが配信された時間以降のものを対象とした。表 3 に、辺野古への米軍基地移設に反対の立場のユー

表 3 辺野古への米軍基地移設に反対の立場のユーザに推薦するニュース

ツイート	新聞社	投稿日時
日本遺族会、安倍首相の靖国参拝を求める https://t.co/6aP16Gwwj1 →会長の水落敏栄参院議員「戦後世代が8割以上を占める今日、遺族会が戦争の悲惨さ、平和の尊さを伝えていかなければならない」 https://t.co/XHjLpBxhQx	産経	2018/12/14
岩屋防衛相、辺野古移設は「国民のため」 https://t.co/W8qelWWJzM →岩屋防衛相は、米軍普天間飛行場の名護市辺野古への移設について「日米同盟のためではない。日本国民のためだ」と述べた。	産経	2018/12/15
辺野古移設「官製デマ」 政府に深まる玉城県政への不信 https://t.co/Y4N5RLH2BZ →玉城デニー知事は政府との対話を重視する姿勢は崩していないが、政府は県による情報操作に神経をとがらせている。	産経	2018/12/14
速報】挺身隊の控訴審で三菱重工業に賠償命令 韓国 https://t.co/S9bbbHle4q	産経	2018/12/14
渋谷ハロウィーンの軽トラ横転、2人に罰金10万円 https://t.co/fQ5ukAaPN0 →20代の山梨県富士吉田市の男性と、東京都世田谷区の男性の2人を略式起訴。東京簡裁は同日、2人にそれぞれ罰金10万円の略式命令を出した。	産経	2018/12/14

表 4 辺野古への米軍基地移設に賛成の立場のユーザに推薦するニュース

ツイート	新聞社	投稿日時
辺野古への土砂「民意、意図的に逆で」 立憲・枝野氏 https://t.co/v1LFy0sTQn	朝日	2018/12/15
9月の沖縄県知事選では「辺野古反対」の玉城デニー氏が大勝。それでも辺野古沿岸の埋め立てが強行されました。「沖縄はゴミ箱か」。有権者にむなしさがこみ上げます。 https://t.co/Ab1chKwAy2 https://t.co/0MMBbdoS1H	朝日	2018/12/15
玉城知事「胸をかきむしられる」 辺野古の土砂投入視察 https://t.co/t5S1AiAFPz	朝日	2018/12/15
「沖縄はゴミ箱ですか」知事選で投じた有権者のむなしさ https://t.co/c19seP3gEU	朝日	2018/12/14
土砂投入強行、知事「県民の怒り、ますます燃え上がる」 https://t.co/UmTLzPvP1x	朝日	2018/12/14

に推薦するニュース上位5件を、表4に、辺野古への米軍基地移設に賛成の立場のユーザに推薦するニュース上位5件を示す。

6 ま と め

本研究では、Twitterを利用してニュースを取得する際、フィルターバブル問題により、ニュースを中立的に理解することが困難になっているという問題を解決するため、ユーザが興味のあるトピックに関して、ユーザの目に触れにくいニュースを推薦する手法を提案した。またそれに対し評価実験を行い、提案手法が有用であるか調べた。結果、提案手法が問題を解決するために有用な手法であると示すことができなかった。原因としては、ユーザのコミュニティ抽出が適切に行われなかったことなどが考えられる。今後はより適切なコミュニティ抽出を行い、ユーザにとって有用な推薦が行われる手法の開発を目指す。

文 献

- [1] “20.2.4 情報の偏りと信憑性”，はいばーワックブック，https://hwb.ecc.u-tokyo.ac.jp/current/information/technology_and_society/demerit/bias_and_credibility/，2019.1.20 閲覧
- [2] “パーゾナライズ (パーソナライゼーション) とは”，IT用語辞典 e-Words，2018.12.11 更新，<http://e-words.jp/w/%E3%83%91%E3%83%BC%E3%82%BD%E3%83%8A%E3%83%A9%E3%82%A4%E3%82%BA.html>，2019.1.20 閲覧
- [3] イーライ・バリサー，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
- [4] “フィルターバブルとは”，コトバンク，<https://kotobank.jp/word/%E3%83%95%E3%82%A3%E3%83%AB%E3%82%BF%E3%83%BC%E3%83%90%E3%83%96%E3%83%AB-1813752>，2019.1.20 閲覧
- [5] 神島 敏弘，赤穂 昭太郎，麻生 英樹，佐久間 淳，“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] 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方法”，Vol.2009-DBS-149 No.16, 2009.11.21
- [8] 川口 天佑，牛尼 剛聡，“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”，DEIM Forum 2018, 2018.3
- [9] 切通 恵介，楠見 孝，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, 2016.3
- [10] “ツイッターとは”，コトバンク，<https://kotobank.jp/word/Twitter-13825>，2019.1.28 閲覧
- [11] 形態素解析エンジン MeCab，<http://taku910.github.io/mecab/>
- [12] Quoc Le, Tomas Miklov, “Distributed Representations of Sentences and Documents”, Proceedings of the 31st International Conference on Machine Learning, 2014
- [13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, “A Density-Based Algorithm for Discovering Clusters”, KDD-96, 1996
- [14] M. Girvan, M. E. J. Newman, “Community structure in social and biological networks”, Proc. Natl. Acad. Sci. USA 99, 7821-7826, 2002
- [15] M. Girvan, M. E. J. Newman, “Finding and evaluating community structure in networks”, Preprint condmat/0308217, 2003
- [16] M. E. J. Newman, “Fast algorithm for detecting community structure in networks”, Phys. Rev. E 69, 2004
- [17] Aaron Clauset, M. E. J. Newman, Christopher Moore, “Finding community structure in very large networks”, Physical Review E 70(6), 2004
- [18] 毎日新聞のニュース・情報サイト，<https://mainichi.jp/>，2019.1.10 閲覧
- [19] 産経ニュース，<https://www.sankei.com/>，2019.1.10 閲覧
- [20] 朝日新聞デジタル，<https://www.asahi.com/>，2019.1.10 閲覧
- [21] gensim: Topic modelling for humans，<https://radimrehurek.com/gensim/>，2019.1.6 閲覧
- [22] Tadao Yamaoka, “doc2vec で Wikipedia を学習する”，Tadao Yamaoka の日記，2017.4.29 更新，<http://tadaoyamaoka.hatenablog.com/entry/2017/04/29/122128>，2018.11.17 閲覧
- [23] Index of /jawiki/latest/，<https://dumps.wikimedia.org/jawiki/latest/>，2018.11.17 閲覧
- [24] NetworkX，<https://networkx.github.io/>，2019.1.21 閲覧