

Twitterにおける虚偽情報拡散抑制に向けた 情報拡散コミュニティ可視化手法の検討

日下部 眞太郎^{1,a)} 米澤 拓郎¹ 大越 匡¹ 鈴木 もとこ² 山田 暁² 中澤 仁³

概要：近年，SNS 上で虚偽情報の拡散が問題になっており，拡散傾向の分析や，情報の信頼性を可視化による拡散の抑止が検討されてきた．しかし，虚偽情報がどのような属性のユーザを媒介にして拡散されるのか明らかになっていないことや，既存の抑止方法の効果が限定的であるという問題がある．そこで本研究では，情報自体の信頼性の検証ではなく，虚偽情報には特有の恣意的な拡散プロセスがあるのではないかと仮説を立て，その拡散プロセスの分析結果をユーザに提供することで，ユーザが自律的に情報拡散を抑制可能とする手法を提案する．特に本稿では，仮説がどの程度妥当なものであるか，拡散者コミュニティの興味およびその時系列変化の分析手法を提案，実装し，虚偽情報を対象とした分析実験を行った．

1. はじめに

近年，Web やソーシャルネットワーキングサービス（SNS）上で，アメリカの大統領選挙や英国の欧州連合離脱を問う国民投票を発端として，虚偽情報の拡散が増大している [16]．投票行動への影響は未だ定かになっていないが，他の事例として拡散された虚偽情報が株式市場に影響を与えたり [17]，自然災害に社会的な混乱を引き起こしたとし [18][19]，実社会への影響が強く懸念されている．この問題に対処するために，虚偽情報の拡散防止策の考案や拡散プロセスの分析が今後ますます重要となる．従来のアプローチとして，報道メディアや Web，SNS 上で発信される情報の真偽を検証するファクトチェックが行われている．しかし，人手で判断するこの手法は，情報の検証に時間を要し，十分な数の情報を検証できない問題がある．また，Web や SNS の事業者自身によって，そのサービスの基盤上で共有される情報の信頼性を可視化する手法も開発されている．しかし，Facebook が実際のサービスでこの手法を適用して運用したところ，バックファイア効果により拡散防止とは逆の効果が観測された [7]．バックファイア効果とは，政治的なテーマ（イラクに大量破壊兵器があったかどうか等）について，事実を示した後でも，政治的立場によって，全く効果がなかったり，逆に誤解への確信を強めてしまう現象である [10]．よって，虚偽情報拡散の抑制には不適當であり，現在は信頼性の可視化は行われてい

はない．

このように，従来は虚偽情報の拡散防止のために情報自体の信頼性の担保を基本とした対策が行われてきたが，抑止力としての働きは低いという問題がある．そこで本研究では，情報自体の信頼性の検証ではなく，虚偽情報には特有の恣意的な拡散プロセスがあるのではないかと仮説を立て，その拡散プロセスの分析結果をユーザに提供することで，ユーザが自律的に情報拡散を抑制可能とする手法を提案する．虚偽情報の 3 割以上が政治的なデマであるという既存研究の報告 [12] を考慮すると，虚偽情報の拡散は，まずはその虚偽情報が拡散されることによって利益が得られるコミュニティから拡散される想定される．よって，どういったコミュニティで拡散されていったのか，その拡散プロセスを可視化することによって，情報の受け手が虚偽情報に特徴的に見られる拡散プロセスを理解し，情報の信頼性の判断材料になり得ると考えられる．

本稿では，上述した仮説に基づき虚偽情報の拡散傾向を明らかにするために，情報がどういった情報拡散者（リツイーター、以下 RTer）によって拡散されていったか，特にその集団（コミュニティ）としての属性および属性の時系列変化に着目する．コミュニティの属性はそのコミュニティを構成するユーザ（ここでは RTer）の様々な特徴の集合によって分析が可能であるが，本稿では上述した「虚偽情報が拡散されることによって利益が得られるコミュニティが存在」することを想定し，コミュニティの属性としてそのコミュニティの興味，という観点で分析を行うことを主な目的とする．リツイートがどういった興味を持つコミュニティで拡散されていったか，その極性状態（興味が

¹ 慶應義塾大学政策・メディア研究科

² 株式会社 NTT ドコモ

³ 慶應義塾大学環境情報学部

^{a)} birdman@ht.sfc.keio.ac.jp

偏っている状態)と無極性状態(様々な興味が包含される状態)の時間遷移を分析することで、虚偽情報の拡散に特徴的な傾向があることを明らかにすることを試みる。コミュニティの興味が抽出する手法としては、コミュニティを構成する個々のユーザの興味が抽出しその集合を得る方法が考えられる。Twitterにおけるユーザの属性推定手法としては、対象ユーザのツイートやプロフィール文を分析する手法[8][6][1][3]やフォロー・メンション等のユーザ間の関係を利用する手法[5][2]などがこれまで提案されてきた。一方で、これらの既存手法は対象ユーザが興味にそったプロフィール文の記述や投稿を行っていることや、コミュニティに属するユーザ間で相互フォロー関係やメンション関係があることを前提としており、多くの場合分析自体が難しいという問題が存在する。本研究ではユーザの投稿やプロフィール文、相互フォロー関係の存在に頼らない、共通フォロウィ情報を利用したユーザ集合の興味推定手法[12]を拡張し、コミュニティの興味およびその時系列変化の分析手法を提案する。本稿では特に、上述した仮説がどの程度妥当なものであるか、提案したシステムを実装し、虚偽情報を対象とした分析実験を行い、その結果を報告する。

2. 虚偽情報とその拡散性

2.1 虚偽情報の定義

SNSなどのサービスの浸透につれ誰もが容易に情報を投稿し、共有可能なインターネット環境が整ってきた。それに伴い、フェイクニュースと言われる情報が意図的、非意図的に関わらずSNS上に投稿され、急速に拡散され、人々の事実認識やそれに伴う意思決定に影響を与えたり、社会問題へと発展するケースが多く見られるようになった。フェイクニュースは、大きく以下の6つに分類される[14]。

1. 誤情報: 取り上げられた事実や事象に誤りのある情報
2. 偽情報: 取り上げられた事実や事象がそもそも存在しない情報
3. 不正確な情報: 取り上げられた事実や事象に誤りがあるとまでは言えないが、正確ではない情報
4. ミスリーディングな情報: 事実や事象に誤りがあるとまでは言えないが、見出しや表現の仕方などで誤解を生じさせかねない情報
5. 根拠のない情報: 事実や事象に誤りがあるとは言えないが、それが事実であると証明する根拠がない情報
6. 風刺や冗談: 事実や事象はそもそも存在しないか、大幅に脚色されているが、風刺や冗談であり、人を騙す意図はない情報

上記のフェイクニュースの定義の中でも、特に1. 誤情報と2. 偽情報は真実と異なる情報であり、それが広まった際の社会的影響は重大である。本稿では、フェイクニュースのうち、誤情報と偽情報をまとめて虚偽情報と呼ぶ。2016年のアメリカ大統領選では、Facebook上で、ニューヨー

ク・タイムズやCNN等の主要メディアのニュースより、フェイクニュースの方が記事のエンゲージメント数が高かったことが明らかになっている[16]。この事例から、誰でも自由に投稿が可能なインターネット環境において虚偽情報の投稿をなくすことは不可能であるため、いかにその拡散を抑制するかが、今後非常に重要になるといえる。

2.2 拡散性と抑制手法

Soroush[12]らによるTwitterの全量データ調査によると、虚偽情報は真実の情報に比べて拡散されやすいという分析結果が報告されている。SNSにおいては驚きのある内容、ネガティブな要素、新規性などが強い記事・投稿が拡散されやすい傾向にあり、虚偽情報を投稿するユーザは虚偽の情報に脚色をつけ、拡散性の強い記事を量産する傾向にある。特に記事へのアクセス数はそのまま広告収入へとつながるため、拡散されやすい虚偽情報の生産はビジネス化されている。同じく[12]によると、虚偽情報の拡散を記事・投稿のテーマ別にみると、82,600件中、「政治」27,600件、「都市伝説」16,500件、「科学・技術」12,000件、「ビジネス」11,000件、「テロ・戦争」8,000件、「エンターテインメント」6,000件、「自然災害」1,300件となっており、政治をトピックとした虚偽情報が最も拡散されやすいと報告されている。選挙期間中や重要な政治局面においては対立候補や対立政党に対するネガティブな情報が重要な戦略材料となるため、ある政党の支持者に都合の良い情報は虚偽情報であったとしても(むしろ衝撃的に脚色された虚偽情報こそ)急速に拡散されることは容易に想像できる。

この問題に対応し虚偽情報の拡散を抑制するため、その情報の信憑性を評価するファクトチェックが行われるようになった。ファクトチェックはフェクトチェッカーと呼ばれる人々が、人手によって情報のスクリーニングから予備調査、本調査を行い、その評価結果を記事にしており、高いコストを要する。ファクトチェックのコストを軽減するため、一部のプロセスを自動化を支援する研究も行われているが、その支援は未だ限定的である[20]。一方で、ファクトチェックにより信憑性に疑いのあることが明示されたとしても、人々の拡散行動の抑制効果は低いという事実も指摘されている。例えばFacebook社は共有される記事の信憑性をファクトチェックの結果を併記することによりユーザに注意喚起を促す取り組みを行ったが、バックファイア効果[9]により期待した抑制結果が得られなかったとし、その取り組みを断念している。

2.3 研究の目的

上述したように、信頼できる第三者機関より情報自体の信憑性が評価されたとしても、虚偽情報拡散の抑制には効果が低い。従って、情報の信憑性評価とは異なった、虚偽情報の拡散抑制方法が必要である。そこで本研究では、情

報自体ではなく、その情報の拡散プロセスをユーザに提示することにより、虚偽情報の拡散の抑制を行う手法を提案する。例えば上述したように、政治的な虚偽情報はその情報が広まることによって利益が得られる特定のコミュニティによって、意図的・非意図的に関わらず情報の拡散が行われると考えられる。すなわち、情報を拡散するコミュニティの極性を分析することにより、虚偽情報の可能性が強いかどうか、判断できる可能性があると考えられる。一方で、情報がある程度拡散されてしまった後では、さらに多くのユーザによっても拡散がなされる。すなわち、コミュニティは極性状態から無極性状態へと変化することが考えられる。この時間的な遷移を分析し、極性状態から無極性状態へと移行する変化点を捉えることにより、虚偽情報に特徴的に見受けられる拡散プロセスを明らかにすることを最終的な目的とする。本稿では、この目的を達成するための最初の段階として、情報を拡散するコミュニティの属性、特にコミュニティの興味を抽出し、可視化を行う手法を提案し、虚偽情報を対象とした分析実験を行う。

3. 関連研究

3.1 ツイート解析による属性分析

リツイートを行うコミュニティの属性を分析するために、そのコミュニティを構成する個々のユーザの属性を分類する手法が考えられる。Twitter API から直接得られる定量的な情報としては、個々のユーザのフォロワー数、フォロワー数、過去のツイート数、使用言語などが挙げられる。一方、これらの基礎的なデータだけでは個々のユーザのプロファイリングとしては不十分であるため、プロフィール情報の分析や投稿内容のテキスト解析によるユーザの興味推定などの手法が提案されている。属性推定のために対象のツイート情報を用いる手法では、推定対象のユーザ自身が投稿したツイートや、付与されている URL、位置情報等を解析の対象とする。既存手法では事前に位置情報 [8][6] や支持政党 [1][3] 等、特定の属性に絞った推定を行っている。一方で、プロフィール情報は十分に書かれていなかったり、投稿されたツイートの内容が必ずしもユーザが持つ興味と合致しない場合も多いため、ユーザのプロファイル推定を行えない場合も多く存在する。

3.2 フォローメンション情報解析による属性分析

一方で、投稿ツイートの分析ではなく、ユーザ間のフォロー、メンション情報を利用するコミュニティの属性推定手法も提案されている。フォローメンション情報を用いてノード（ユーザ）間にエッジを張り、ソーシャルグラフを作る。そして、ソーシャルグラフからコミュニティを抽出し、推定対象ユーザの属性を推定している [5][2]。鳥海ら [15] は、フォローメンション情報解析の手法を用いて、虚偽情報の拡散プロセスにおいて主体になっているコミュニティ

情報の解析を試みている [15]。一方で、既存手法で分類可能な属性はその種類に制限があり、任意の虚偽情報を対象とするには制限が存在する。またフォロー、メンション情報と周辺ユーザ（1 ホップ先のユーザ）のプロフィール情報の分析を組み合わせた手法 [2] も提案されているが、対象となるコミュニティが大規模になっていくにつれ、周辺ユーザの探索に要するコストが高くなるという問題も存在する。

3.3 共通フォロウィ抽出による属性分析

本研究の提案手法に最も近い手法として、特定のユーザ群の共通フォロウィ情報を利用した手法 [13][11] が挙げられる。河野ら [13] は、実世界イベントの大衆性を分析するため、ユーザ群のフォロウィ集合の分散性に基づく評価手法を提案している。これは、後述する本研究における、コミュニティの無極性状態の抽出手法に類似する。また佐藤ら [11] は、ユーザ群の共通のフォロウィのうち上位に位置する Twitter アカウント情報を Wikipedia 情報を参照することで属性分類し、その分類結果を利用し実世界イベントの判別を提案している。これは本研究における、コミュニティの偏った極性状態の抽出手法に類似する。本研究では、コミュニティの極性・無極性状態を抽出するため、上述した2つの手法を組み合わせた手法を提案する。また、その時系列変化を観測するため、RTer の集合をスライディングタイムウィンドウ方式で分割し、分類を行う手法を提案する。これらのことから、提案手法は既存手法を組み合わせリツイートコミュニティの極性・無極性分類に適用しつつ、その時系列変化を観測するという新たな手法として位置づけられる。

4. 提案手法

本章では、提案手法の概要、機能要件、設計について述べる。

4.1 概要

本研究では、虚偽情報には固有の拡散プロセスが存在するという仮説を評価するため、虚偽情報を拡散する RTer コミュニティの属性分類とその時系列変化の観測を可能とする手法を提案する。本来であれば現在進行系で拡散状態にある情報が虚偽情報であるかどうかを即座に判断しうるシステムが望ましいが、本稿では上述した仮説をまずは評価するため、予め収集したツイートに対して、解析対象となるツイートを著者らが任意に選択し、その拡散プロセスを過去に遡って可視化を行うこととする。提案手法は、解析対象となるツイートに対して、大きく分けて以下の5つのプロセスで分析される。

ステップ 1: RTer 全体集合の取得

RTer は、対象のツイートをリツイートすることで虚

偽情報を拡散させる。そこで RTer によるリツイート
を収集することで、RTer の全体集合を抽出する。

ステップ 2: 属性抽出対象となる RTer 部分集合の選択

虚偽情報が RTer 達によって拡散されるにつれて、RTer
の全体集合の構成は変化していく。この構成の変化を
抽出するために、RTer の全体集合を時系列 (1 日区切
り等任意の範囲) に分ける。

ステップ 3: RTer 部分集合の共通フォロウィ分析

全体集合から抽出した RTer の部分集合を対象にして、
共通のフォロウィを抽出する。共通のフォロウィを抽
出することで、RTer 同士の関係性を表すことができる。
共通のフォロウィとは、RTer の部分集合で構成される
ソー シャルグラフにおいて、2 人以上からフォローさ
れているユーザのことである。しかし、この方法で共
通のフォロウィを抽出すると、孫正義 (@masason) や
地震速報 (@earthquake.jp) のような、日本で多数の
のフォロワ数を持つユーザが抽出されることになる。
(孫正義のフォロワ数: 2,595,630 人、地震速報のフォ
ロワ数: 2,173,396 人) そのため、RTer 集合固有のフォ
ロウィではない可能性がある。これらのユーザは場合
によってはノイズとして扱われる対象であり、RTer
集合の固有なフォロウィかどうかを判別する必要がある。
そこで、本手法では RTer 集合のフォロウィで
あるかを判別するために、共通のフォロウィに対して
FeatureScore を割り振る。(詳細を RTer 集合の属性抽
出手法の手順で述べる)

ステップ 4: Wikipedia を参照し共通フォロウィ情報から 属性の抽出

共通のフォロウィがどのような属性を持っているかを
抽出するために、共通のフォロウィのアカウント名を
含む Wikipedeia の記事を参照する。Wikipedia は自
由に編集に参加可能なイン ターネット百科事典であ
る。Wikipedia に存在する全ての語には、属するカテ
ゴリが存在し、各カテゴリは 1 つ以上の上位カテゴリ
に属するという階層構造になっている。本手法では、
Wikipedia に存在する「百のカテゴリ」(表 1) の中か
ら「総記」「プロジェクト関 連文章」「一覧」「ウィキ
ポータル」「画像」「索引」「ウィキ ペディアのテンプ
レート」「ヘルプ」と「日本」のカテゴリを削除した。
その結果残った計 91 個のカテ ゴリを階層構造の第 1
層のカテゴリとする。カテゴリ数は 91 個だが、階層
構造の第 1 層のカテゴリを「百のカテゴリ」と呼称す
る。その第 1 層のカテゴリから 1 ステップで到達可能
なカテゴリを第 2 層、さらに第 2 層から 1 ステップで
到達可能なカテゴリを第 3 層とし、それ以降の階層
も同様に扱う。そのため、共通のフォロイに関連する
Wikipedia 上のページに割り振られている各カテゴリ
から、第 1 層カテゴリのいずれかに到達するまで幅優

先探索を再帰的に行う。第 1 層カテゴリを発見した場
合、発見された百のカテゴリに対してポイント (カテ
ゴリポイント) を 1Point 割り振る。また、第 1 層カテ
ゴリを同じ距離で複数発見した場合、発見した各カテ
ゴリに対し 1Point を割り振る。最後に、1Point も割
り振られていなかったカテ ゴリに関して 0Point を割
り振る。その結果、ユーザを 91 次元の特徴ベクトル
で分類することができる。また、ユーザに割り振られ
うるカテゴリポイントの総数は、Wikipedia の各記事
のタグに依存するため、全てのユーザを同じ尺度で表
すため、特徴ベクトルの正規化を行う。正規化後の特
徴ベクトルを、共通のフォロイの属性として用いる。
このカテゴリ分類法を、共通のフォロウィとして抽出
されたアカウントの上位 10 件に対して適応する。

表 1: 百のカテゴリ一覧

社会, 政治, 経済, 産業, 交通
教育, 歴史, 福祉, 医療, 健康
環境, 市民活動, 平和, 軍事, 大学
芸術, 文化, 言語, 宗教, 遊び
趣味, 伝統芸能, 文学, 音楽, 美術
映画, 演劇, アニメ, 漫画
イラストレーション, スポーツ
ゲーム, 賭博, ファッション, 食文化
建築, マスメディア, 芸能
大陸, アジア, アフリカ, オセアニア
北アメリカ, 南アメリカ, 北海道
東北, 関東, 中部, 近畿, 九州
沖縄, 自然, 宇宙, 元素, 気象
災害, 海洋, 生物, 植物, 動物
鉱物, 学問, 哲学, 論理学, 言語学
心理学, 文学, 宗教学, 政治学, 経営学
社会学, 教育学, 数学, 物理学, 化学
生物学, 人類学, 地球科学, 医学, 薬学
歯学, 農学, 工学, 技術, コンピュータ
ネットワーク, エレクトロニクス
バイオテクノロジー

ステップ 5: 結果を統合可視化

ステップ 2 からステップ 4 で抽出した部分集合ごとの
カテゴリポイントの割り振りを統合して、時系列でど
のようにカテゴリポイントが変化するかを可視化す
る。可視化の手法は 2 つ設定した。1 つ目は、カテ
ゴリの大小をひと目で把握するために WordCloud 形式
の可視化である。2 つ目は、時系列でのカテゴリポ
イントの特徴量の変化量を棒グラフで可視化する方法で
ある。

上述したうち、特にステップ 3 と 4 に関しては佐藤ら [11]
の手法を応用した手法となっており、その詳細を次節で説
明する。

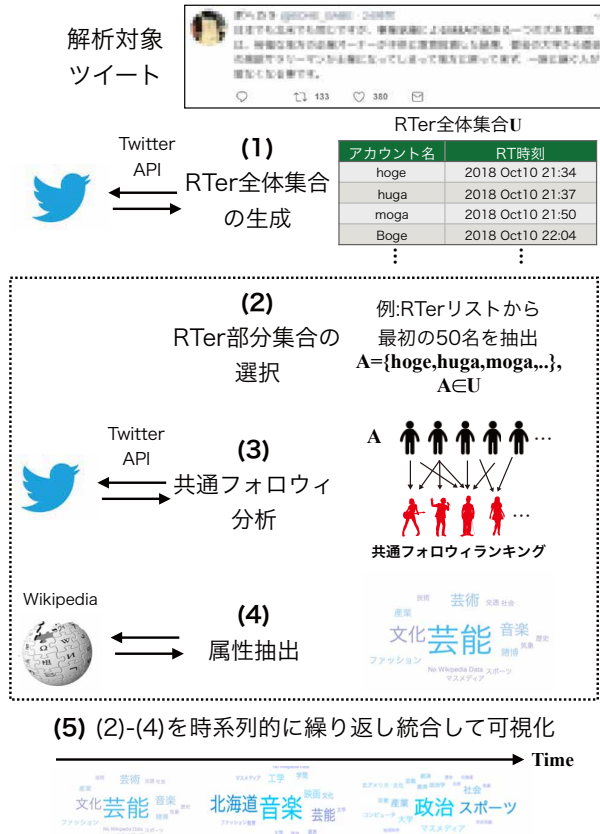


図 1: 手法の概要

4.2 RTer 集合の属性抽出手法

提案手法では、RTer 集合の極性を抽出するため、その特徴を分析する。佐藤らの提案手法を空間的なユーザ集合ではなく時間的に変化する RTer の集合に対して応用し、共通フォロウィに含まれる著名なユーザをノイズとして除去する手法を考慮し、以下の手順で分析を行う。

手順 1

RTer 部分集合に含まれるユーザのフォロウィのうち、2 人以上のユーザからフォローされているフォロウィの集合を $F = \{f_0, f_1, \dots, f_n\}$ とする。

手順 2

集合 F に含まれるそれぞれのフォロウィの RTerScore と BasicScore を計算する。

集合 F の任意の要素 f_x が RTer 集合からフォローされている数を $RTerFollow(f_x)$ とすると、 $RTerScore(f_x)$ は以下のように定義される。

$$RTerScore(f_x) = \frac{RTerFollow(f_x)}{\sum_{i=0}^n RTerFollow(f_i)}$$

また、集合 F の任意の要素 f_x が Twitter でフォローされている数を $BasicFollow(f_x)$ とすると、 $BasicScore(f_x)$ は以下のように定義される。

$$BasicScore(f_x) = \frac{BasicFollow(f_x)}{\sum_{i=0}^n BasicFollow(f_i)}$$

手順 3

集合 F に含まれるそれぞれのフォロウィの FeatureScore を計算する。FeatureScore(f_x) は以下の式で求められる。

$$FeatureScore(f_x) = RTerScore(f_x) - BasicScore(f_x)$$

FeatureScore の値が高いほど、その共通のフォロウィは RTer 集合に特徴的なフォロウィの可能性があり、FeatureScore の値が低いほど、その共通のフォロウィは RTer 集合に関係のないフォロウィの可能性が高いと判断できる。本研究では、この FeatureScore を各共通のフォロウィに割り振り、降順にソートを行い、上位 10 名を佐藤ら [11] の Wikipedia 参照による属性推定手法で分析する。

4.3 システム設計と実装

本研究では本手法を適用し、RTer のコミュニティの変遷を時系列的に可視化するシステムの設計、実装を行った。本システムは、コミュニティの変遷を可視化することで、RTer 属性のカテゴリポイントの変化を、質的かつ量的に分析することができる分析者向けの解析ツールである。同解析ツールは以下の機能を有する。

- RTer 全体集合の取得機能
- RTer 部分集合の共通フォロウィ解析機能
- RTer 部分集合の属性解析機能
- 解析結果表示機能

システム構成図を図 2 に示す。以下、各機能の設計・実装の詳細について述べる。

4.3.1 RTer 全体集合の収集機能

本機能では、虚偽ツイートのリツイートを取得することで、RTer を特定する。虚偽ツイートの投稿内容を検索のキーワードとして設定し、そのリツイートを検索する。検索のために、Twitter の REST API を用いた。Twitter API は Twitter 上で投稿された過去のツイートやユーザを検索条件を指定して取得できる仕組みである。過去一週間以前の虚偽ツイートのリツイートを検索するために、REST API に含まれている Premium Search API を用いた。Premium Search API では過去一週間以前の全期間のツイートを検索することができる。この API を用いて、ツイートを検索するエンドポイントである tweets/search をリクエストして得られた Status オブジェクトと User オブジェクトを取得する。Status・User オブジェクトからは、RTer が虚偽ツイートをリツイートした際の日時や ID、RTer のユーザ ID やテキストを取得できる。これらの RTer 固有の情報を集約して RTer の全体集合のリストを得る。

4.3.2 RTer 部分集合の共通フォロウィ解析機能

本機能では、RTer 全体集合から RTer 部分集合を抽出し、その共通のフォロウィ上位 10 人を取得する。まず、RTer 全体集合のリストにある各 RTer の生成日時（虚偽ツイートのリツイート日時）を用いて、任意の時系列で RTer の部分集合を抽出する。次に、各 RTer のフォロウィ情報を REST API のエンドポイントである friends.ids と users/lookup から取得する。そして、取得したフォロウィ情報を用いて、本手法で述べたフォロウィの各フレンドの FeatureScore を算出し、各 RTer 部分集合の共通フォロウィ上位 10 人を抽出する。

4.3.3 RTer 部分集合の属性解析機能

本機能では、RTer 部分集合の共通フォロウィ上位 10 人を Wikipedia のカテゴリ分類にかけて、コミュニティの属性を推定する。まず、RTer 部分集合の共通フォロウィ 10 人の名前を Wikipedia のデータベースで検索し、該当する記事に含まれるカテゴリリストを取得する。取得したカテゴリを起点にして、親カテゴリ（百のカテゴリ）に到達するまで幅優先探索のアルゴリズムでカテゴリの階層を遡る。探索の結果割り振られたポイントを正規化する。

4.3.4 解析結果表示機能

本機能では、RTer 部分集合の属性解析機能によって得られた結果を WordCloud 形式や、テキスト形式で出力する。この可視化結果を元に時系列によるコミュニティ変遷の傾向を分析することができる。

5. 実験

本章では、提案手法による属性推定の動作実験の結果を示す。動作実験では、実際のツイートサンプル（図 3）を用いて属性推定を行った。対象サンプルは、ファクトチェック機関の FIJ が主催したプロジェクト『FactCheck 沖縄県知事選 2018』で誤情報として判定されたツイートである [21]。対象サンプルは 2018 年 9 月 13 日にツイートされており、サンプリング時点（2018 年 11 月 7 日 18 時 35 分）で 231 リツイートされていた。リツイートした RTer の共通の友人と Wiki による参照を経て抽出された結果を図 4 に示す。

図 4 は初期のリツイートから 1 日毎に、2018 年 9 月 13 日から 2018 年 10 月 29 日まで期間のコミュニティ変遷を可視化した結果である。今回の可視化では、サンプルツイートがツイートされてから 1 日ごとに範囲を分け、10 人以上の RTer が存在する日の属性推定の結果を示した。また、属性の特徴量として振り分けられたカテゴリポイントを、WordCloud 形式と棒グラフ形式で可視化している。

これらのデータを基に、コミュニティの変遷の傾向を質的に述べる。コミュニティの変遷は T1(9 月 13 日 9 月 16 日)、T2(9 月 17 日 9 月 20 日)、T3(9 月 21 日 10 月 22 日)の 3 つに分けられる。T1 の期間では、政治の特徴ベクト

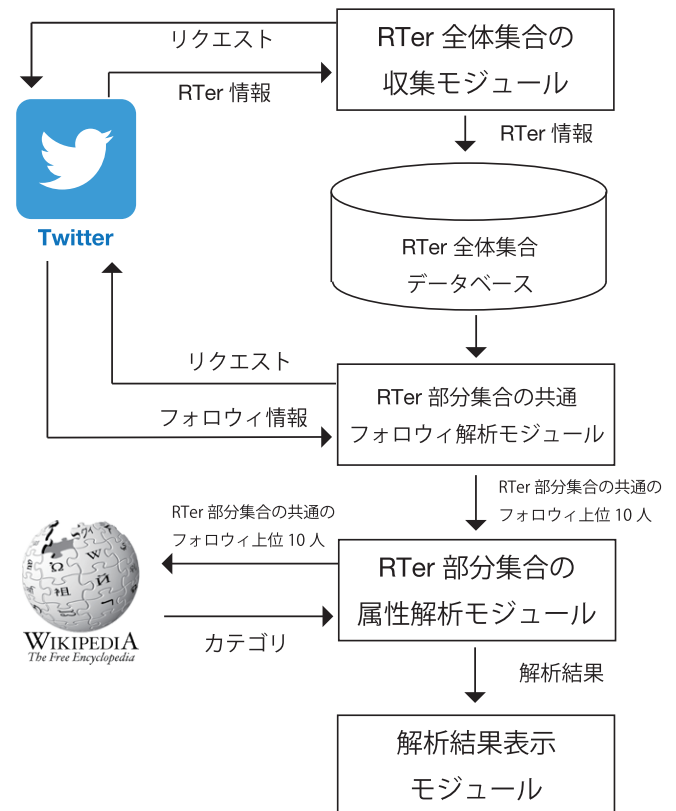


図 2: システム構成図

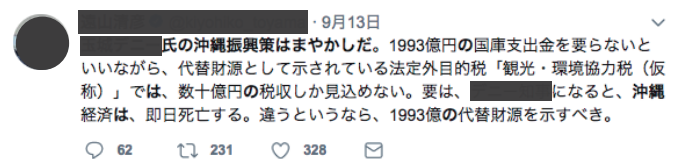


図 3: ツイートサンプル

ルが非常に高く出ており、コミュニティは政治カテゴリに偏って推移していることが分かる。T2 では、T1 で支配的であった政治以外のカテゴリが増え、属性の割合がより分散的に推移するようになった。T3 では、コミュニティの変遷が、緩やかに極性的な構成と分散的な構成を交互に繰り返しながら推移していることが分かる。これらの可視化結果から、コミュニティの変遷が質的に分析可能であることがいえる。

6. おわりに

本稿では、Twitter 上でユーザが情報の信頼度を判断するために、情報拡散者の属性傾向を可視化する手法を提案した。また、その動作実験を行い、属性推定結果のサンプルを示した。今後の課題として、極性的なコミュニティから分散的なコミュニティへの移行タイミングの傾向や属性を構成する特徴量の変化を量的に検証するため、より多くのサンプルに基づいて属性推定を行う予定である。



図 4: 時系列による可視化

謝辞

本研究の一部は、国立研究開発法人情報通信研究機構によって支援頂いた。

参考文献

- [1] S. Esparza , M. O'Mahony and B. Smyth, "Cat-Stream: Categorising Tweets for User Profiling and Stream Filtering", Proc. of the IUI'13, pp. 25-36, 2013.
- [2] W. Zhou , H. Jin and Yan Liu, "Community Discovery and Profiling with Social Messages", Proc. of the KDD'12, pp.430-438, 2011.
- [3] M.Pennacchiotti, A.Popescu, "Democrats,Republicans and Starbucks Afficionados: User Classification in Twitter", Proc. of the KDD'11, pp.430-438, 2011.
- [4] Z. Xu, Rong Lu, L. Xiang and Q. Yang, "Discovering User Interest on Twitter with a Modified Author-Topic Model", Proc. of the WI-IAT'11, pp.422-429, 2011.
- [5] J. McGee and J. Caverlee and Z. Cheng, "Location Prediction in Social Media Based on Tie Strength", Proc. of

- the ICWSM'13, 2013.
- [6] H. Chang, D. Lee, M. Eltaher and J. Lee, “@Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage”, Proc. of the ASONAM'12, pp. 111-118, 2013.
 - [7] Pennycook, Gordon, Rand and David G, The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings, 2017.
 - [8] B. Hecht, L. Hong, B. Suh and E. Chi, “Tweets from Justin Bieber’s Heart: The Dynamics of the “Location” Field in User Profiles”, Proc. of the CHI'11, pp. 237-246, 2011.
 - [9] 上里和也, 浅井洋樹, 奥野峻弥, 山名早人, Twitter ユーザを対象とした属性推定の精度向上-周辺ユーザの属性補完を利用して-, DEIM Forum 2015, 2015.
 - [10] Brendan Nyhan, When Corrections Fail: The persistence of political misperceptions, Political Behavior, Issue2, pp 303-330, Vol32, 2010.
 - [11] Soroush Vosoughi, Deb Roy, Sinan Aral, The spread of true and false news online, Science 09 Mar 2018:Vol. 359, Issue 6380, pp. 1146-1151, 2018.
 - [12] 佐藤翔野, 米澤拓郎, 河野慎, 中澤仁, 川崎仁嗣, 太田賢, 稲村浩, 徳田英幸, ソーシャルネットワークにおける 共通の友人に着目した実世界イベント分類手法, 情報処理学会論文誌, No.10, 2222-2235, Vol.57, 2016.
 - [13] 河野慎, 米澤拓郎, 中澤仁, 川崎仁嗣, 太田賢, 稲村浩, 徳田英幸, ソーシャルネットワークにおけるフォロー集合分析に基づく実世界イベント分類手法, 情報処理学会論文誌, No.1, pp.72-82, Vol.56, 2015.
 - [14] BuzzFeed, Fake news. It’s complicated, <https://firstdraftnews.org/fake-news-complicated/>, 2018/11 月 1 日取得.
 - [15] 鳥海不二夫, 榊剛史, バースト現象におけるトピック分析, 情報処理学会論文誌, Vol.58, No.6, 1287-1299, June. 2017.
 - [16] BuzzFeedNews, フェイクニュースが民主主義を壊す facebook が助長したその実態. <https://www.buzzfeed.com/jp/sakimizoroki/fake-news-on-sns-and-democracy>, 2018 年 11 月 1 日取得.
 - [17] K. Rapoza, “Can ‘fake news’ impact the stock market?”, Forbes, 26 February 2017, www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/, 2018 年 11 月 4 日取得.
 - [18] M. Mendoza, B. Poblete, C. Castillo, “Twitter under crisis: Can we trust what we RT?”, in Proceedings of the First Workshop on Social Media Analytics (Association for Computing Machinery, ACM, 2010), pp.71-79, 2010.
 - [19] A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, “Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy”, in Proceedings of the 22nd International Conference on World Wide Web (ACM, 2010), pp. 729-736, 2010.
 - [20] 田上翼, 浅野広樹, 楊井人文, 山下亮, 小宮篤史, 藤村厚夫, 町野明徳, 乾健太郎, ファクトチェックを必要とするニュース記事の探索の支援, 言語処理学会 第 24 回年次大会, 2018.
 - [21] FIJ, 玉城デニー氏は国庫支出金を要らないといっているのは事実ではない, <http://fij.info/archives/1804>, 2018 年, 2018 年 11 月 8 日取得.