

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 6 日現在

機関番号：12608

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330362

研究課題名(和文) 外部情報を利用したソーシャルメディア上のユーザ推薦方式の研究

研究課題名(英文) Social Media User Search/Recommendation Using External Knowledge Bases

研究代表者

野呂 智哉 (NORO, TOMOYA)

東京工業大学・情報理工学(系)研究科・助教

研究者番号：80401553

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：TwitterやFacebook等のソーシャルメディア上で自分が興味を持つ話題に関連する有益な情報を発信するユーザを発見できれば、その話題に関連する情報を効率よく得られるようになる。しかし、各ユーザが発信する個々のメッセージの情報量は少なく、一般文書を対象とした従来のテキストマイニング手法のようにテキスト情報のみを利用する手法の効果は限定的である。

本研究では、ニュース記事やオントロジーを外部情報として利用することで話題関連ユーザを発見・推薦する手法を開発した。さらにこの手法を応用し、特定話題に関連するメッセージの発見手法、同じ話題に興味を持ち情報を発信するユーザクラスタの発見手法を開発した。

研究成果の概要(英文)：Finding users on social media such as Twitter and Facebook who usually post valuable messages related to a topic of interest will help us get information on the topic efficiently. However, it is not easy for us to find such users by general text mining methods, which is based on only the message text, since each message text is short.

In this research, we developed methods for finding/recommending topic-related Twitter users by using news articles and ontologies (e.g. DBpedia) as external knowledge bases. News articles are used for detecting news-topic-related hashtag communities and finding two types of influential users: content-based influential users and authority-based influential users. Ontologies are used for checking topic consistency of each user's messages and finding users who usually post messages on the same topic. In addition, we applied the methods to a method for finding topic-related messages and a method for finding hierarchical user clusters.

研究分野：自然言語処理

 キーワード：ソーシャルメディア マイクロブログ 推薦 検索 コミュニティ検出 クラスタリング オントロジー  
 ニュース記事

### 1. 研究開始当初の背景

Twitter や Facebook 等のソーシャルメディアで、自分が興味を持つ話題に関連する有益な情報を発信するユーザを発見することができれば、その話題に関連する情報を効率よく得られるようになる。しかし、ソーシャルメディア上のテキストの情報量は少なく、一般文書を対象とした従来のテキストマイニング手法のようにテキスト情報のみを利用する手法の効果は限定的である。一方、ソーシャルメディア上には、各ユーザの行動やユーザ間の関係、ニュース記事等の外部情報との関係といった、特有の情報が存在する。これらを利用することにより、効率的なユーザ検索・推薦手法の実現が期待できる。

### 2. 研究の目的

本研究チームは、注目話題の関連ニュース記事からの特徴語抽出手法や、ニュースの話題を考慮した Twitter ハッシュタグ推薦手法の開発を行ってきた。この成果を発展させ、ユーザの行動や、ニュース記事、DBpedia 等の外部情報を利用した、話題関連ユーザ検索・推薦手法を開発する。本研究課題は Twitter を対象とする。さらにこの手法を応用し、特定話題に関連するツイート検索手法、階層的クラスタ (同じ話題に興味を持ち、情報を発信するユーザ集合) の発見手法を開発する。

### 3. 研究の方法

まず、話題関連ユーザ検索の基本的手法として、ユーザの行動やユーザ間の関係に基づく手法を開発した。注目する話題を表すキーワード (話題関連語) を用意し、それによるツイート検索結果に出現するユーザ集合について、各ユーザのツイート発信、リツイート、リプライに基づく行動関係グラフと、ユーザ間のフォロー関係に基づくユーザ関係グラフを構築する。この2種類の関係グラフと各ユーザのツイート数から、各ユーザの話題関連性を計算し、順位付けを行う (図1)。

この基本手法をベースに、外部情報を利用した話題関連ユーザ検索・推薦手法を2種類開発した。1つは、外部情報として DBpedia や YAGO といったオントロジーを利用し、発信するツイート内容が一貫しているユーザを推薦する手法である。本手法では、基本手法で上位に順位付けされたユーザについて、発信するツイートから固有表現を抽出し、オントロジーのクラス階層をもとに拡張することによって、各ユーザが発信するツイートの話題を検出する (図2)。この検出を複数期間に分けて行うことにより、内容の時間的一貫性の有無を判定する。

もう1つは、外部情報としてニュース記事を利用し、話題関連ハッシュタグで形成されるコミュニティから、内容に基づく話題関連ユーザと社会的地位に基づく話題関連ユーザを発見する手法である。本手法では、一定

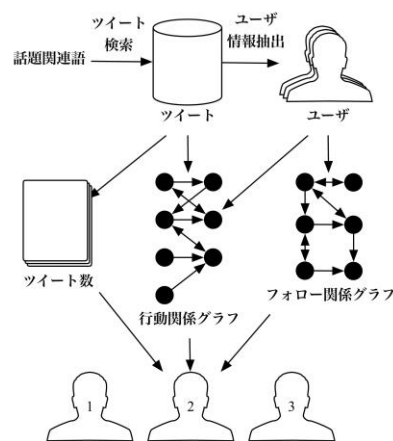


図1: ユーザ・ツイート間の関係に基づく話題関連ユーザ検索

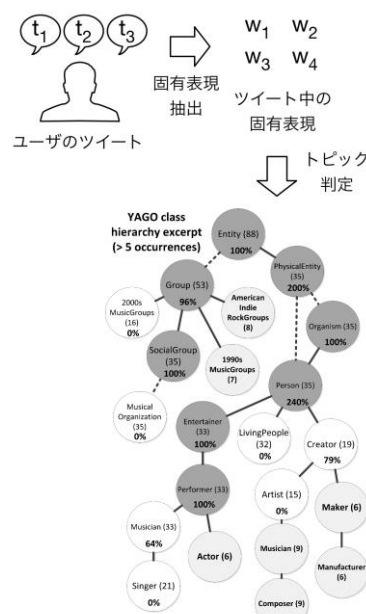


図2: オントロジーを利用した話題判定

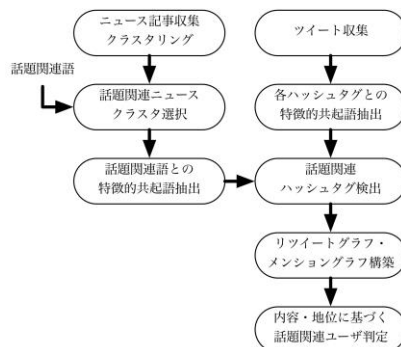


図3: ニュース関連ハッシュタグコミュニティ検出に基づく話題関連ユーザ検索

期間に発信されたニュース記事とツイートを事前に収集する。ニュース記事は内容でクラスタリングし、ツイートからは各ハッシュタグと特徴的に共起する語 (特徴的共起語) を抽出しておく。話題関連語を入力として与えると、関連するニュース記事クラスタを選

択し、そのクラスタにおける話題関連語との特徴的共起語を抽出する。ニュース記事クラスタにおける話題関連語との間の特徴的共起語と、ツイートにおける各ハッシュタグとの間の特徴的共起語から、話題関連ハッシュタグを検出する。話題関連ハッシュタグを含むツイートを発信するユーザで形成されるコミュニティから、リツイート関係とメンション関係をもとにリツイートグラフとメンショングラフの2種類のグラフを構築し、2種類の話題関連ユーザを判定する(図3)。

さらに、開発した話題関連ユーザ検索・推薦手法を応用し、注目する話題に関連するツイートを検索する手法を開発した。本手法では、話題関連語から話題関連ユーザを発見すると同時に、その話題において各ユーザが持つ力としてVoiceとImpactの2種類のスコアを定義する。話題関連ユーザが発信、リツイートしたツイートについて、それに関わったユーザのスコアをもとに話題関連度を算出し、話題関連ツイートを発見する(図4)。

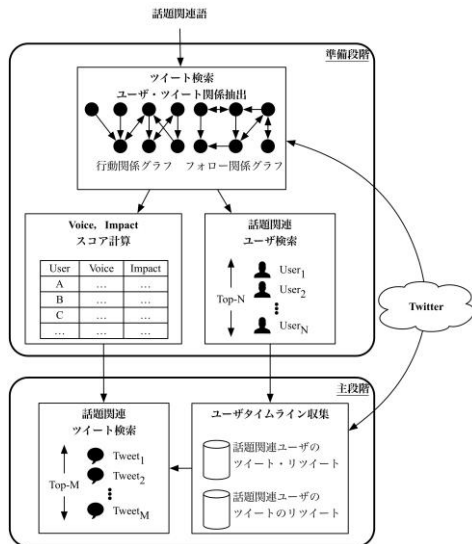


図 4: 話題関連ユーザの行動に基づく話題に関連するツイートの検索

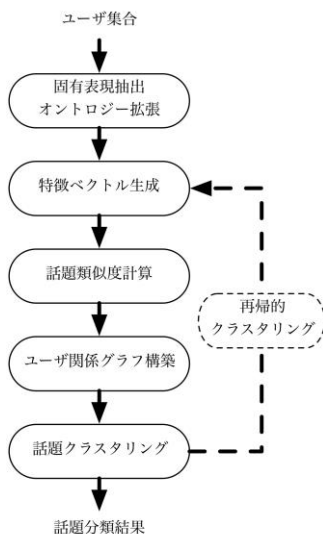


図 5: オントロジーを利用した階層的ユーザクラスタの発見

話題関連ユーザの検索と2種類のスコアの計算は準備段階で事前に行い、ツイートの話題関連性の判定と分離している。

また、オントロジーを利用し、同じ話題に興味を持ち、その話題に関連するツイートを発信するユーザのクラスタの階層的構造を発見する手法を開発した。本手法はユーザ集合を入力とし、各ユーザのツイートから、固有表現抽出とオントロジーによるクラス拡張によって、各ユーザの興味を表す特徴ベクトルを生成する。この特徴ベクトル生成において、オントロジーの上位クラスと下位クラスの重みを変化させるパラメータを導入することにより、分類の細かさを調整できる。生成した特徴ベクトルをもとにユーザ間の話題類似度を計算してユーザ関係グラフを構築し、クラスタリングにより同じ興味を持つユーザのクラスタを発見する。さらに、クラスタリング結果から1つのクラスタを選択し、それをさらに細かく分類することで階層的な構造を発見できる(図5)。

#### 4. 研究成果

個々のツイートは140字以内と短く情報が少ないため、出現単語の分布等の言語情報を利用した話題関連ユーザ検索手法の効果は限定的である。本研究では、オントロジーやニュース記事といった外部情報を利用することにより、この問題を解決する新たな話題関連ユーザ検索・推薦手法を開発した。また、開発した検索手法を応用し、話題関連ツイートの検索手法と階層的ユーザクラスタの発見手法も開発した。

基本手法の開発では、各ユーザのツイートの発信、リツイート、リプライといったTwitter上の行動に基づくユーザ・ツイート間の関係と、フォロー関係に基づくユーザ間の関係の両方を利用することにより、それぞれ単独で利用する場合より話題関連ユーザ検索性能が上がることを示した(学会発表論文④)。また、各ユーザのツイート発信のパターン、他ユーザとのやり取りのパターン等によるユーザ分類を取り入れることにより、さらに検索結果が改善することを示した(学会発表論文⑤)。

オントロジーを利用した話題関連ユーザ推薦では、DBpediaやYAGO等のオントロジーのクラス階層を利用して各ユーザが発信するツイートの内容を表現し、それを複数期間に分けて観測することにより、各ユーザの発信内容が時期によらず一貫しているかどうかを判定する。その結果、基本手法では観測期間において一時的に特定話題のツイートの発信が増加したユーザが検索結果の上位に入ってしまうことがあったが、本手法では観測期間外との間で発信内容に一貫性のないユーザを排除することができるようになった(学会発表論文⑥)。ソーシャルメディアで他ユーザをフォローする動機の一つに、自分が興味を持つ話題についてそのユーザ

が継続的に情報を発信することを期待するということがある。本手法は、このような目的でフォローするユーザを発見する場合に効果的である。個々のツイートは140字以内と短く情報が少ないため、出現単語の分布をそのまま利用して一貫性を判定する方法は効果的ではないが、オントロジーのクラス階層を利用することによってこの問題を解決している。

ニュース関連ハッシュタグコミュニティ検出に基づく話題関連ユーザ検索は、ツイートが短く情報が少ない問題を、外部情報としてニュース記事を利用し、ニュースの話題に関連するハッシュタグを発見、利用することによって解決している(雑誌論文③)。ソーシャルメディアの利用シーンの1つに、ある話題に関連するニュース記事を読んだ後、ソーシャルメディア上でその話題に関する追加的情報を得るということがある。本手法は、このような利用シーンにおいて効果的である。例えば、“Obama”でニュース記事を検索し、その中で大統領選挙に関する話題に興味を持った場合、本手法では、大統領選挙に関連するハッシュタグを特定することによって話題関連ユーザを発見する。これにより、大統領選挙に関連するツイートを発見するのに有効なキーワードが分からなくても、話題関連ユーザを発見できるようになった。また、リツイートグラフとメンショングラフの2種類のグラフを構築することにより、発信内容に基づく話題関連ユーザと社会的地位に基づく話題関連ユーザの2種類の話題関連ユーザを区別して発見できる。例えば、大統領選挙において、自分の意見等を積極的に発信するジャーナリストや著名人は発信内容に基づく話題関連ユーザであり、関連する情報をソーシャルメディア上で得たい場合にはこのようなユーザをフォローすると良い。一方、候補者等のように自身はそれほど頻繁に発信しないが他ユーザから言及されることの多いユーザは社会的地位に基づく話題関連ユーザであり、関連する情報を追跡する場合には現実世界における彼らの言動に注目すると良い。

話題関連ユーザの行動に基づく話題関連ツイート検索では、注目する話題に関連するツイートを、誰が発信、リツイート、リプライしたかに着目することによって発見することを可能にした(雑誌論文②)。これまでに開発した話題関連ユーザ検索によって、注目する話題に関する有益な情報を発信するユーザを発見することは可能になったが、1つの話題のみを常に発信するユーザは非常に少なく、検索により得られる話題関連ユーザも、話題とは無関係のツイートを発信することがある。本手法は、話題関連ユーザが関わるツイートから、話題に関連するものを見出すものである。出現単語の分布による検索手法は、ツイートが短く情報が少ない問題において効果的ではない。また、写真や外部

Webサイトへのリンクがメインのツイートの場合はさらにツイート中の言語情報が少なくなり、オントロジーによる拡張も困難となる。本手法はこれらの問題を解決する。話題関連ユーザが多く関わり、それ以外のユーザがあまり関わらないツイートは、話題関連ツイートである可能性が高いという仮定に基づく手法であり、実験によりこの仮定が当てはまることを示した。例えば、ある大学教授は原子力の話題に関連するツイートを多く発信する原子力の話題関連ユーザである。しかし、その一方で国の財政問題や汚職等に関連するツイートも発信する。このユーザのツイートはいずれも多くのユーザにリツイートされたりリプライを受けたりするが、原子力に関連するツイートには原子力の話題関連ユーザが多く反応し、それ以外のユーザはあまり反応しない。一方、原子力以外の話題のツイートには、原子力の話題関連ユーザはあまり反応しない。

オントロジーを利用した階層的ユーザクラスタの発見では、話題の粒度(分類の細かさ)の調整と、階層的なクラスタ構造の発見が可能になった(雑誌論文①)。例えば、スポーツ全般の話題に関連するツイートを発信するユーザを1つにまとめるか、野球やサッカー、テニス等の話題をそれぞれ区別するかを、パラメータの調整によって変更できる。また、スポーツ全般の話題に関連するユーザのクラスタを発見した後に、それをさらに細かいクラスタに分割できる。ユーザ集合が与えられたとき、その中にどのような話題に興味を持つユーザがいるかを知りたい場合に、本手法は効果的である。出現単語の分布そのものではなく意味を考慮したクラスタリング手法としてLDAに基づく手法が知られているが、それよりも本手法はクラスタリング性能が高く、個々のツイートが短く情報が少ない問題を解決することを示した。

本研究課題は、日本語と英語のツイートを利用して行った。言語処理のためのツールやオントロジーは各言語に合わせたものが必要であるが、本研究課題の成果の本質部分については言語に依存せず、他言語にも適用可能である。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

- ① Kristian Slabbekoorn, Tomoya Noro, and Takehiro Tokuda. “Ontology-Assisted Discovery of Hierarchical Topic Clusters on the Social Web”. *Journal of Web Engineering*, Vol.15, No.5&6, Rinton Press, pp.361-396, 2016. 査読有  
<http://www.rintonpress.com/journals/jweonline.html#v15n56>

- ② Tomoya Noro, and Takehiro Tokuda. "Searching for Relevant Tweets Based on Topic-Related User Activities". *Journal of Web Engineering*, Vol.15, No.3&4, Rinton Press, pp.249-276, 2016. 査読有  
<http://www.rintonpress.com/journals/jweonline.html#v15n34>
- ③ Feng Xiao, Tomoya Noro, and Takehiro Tokuda. "Finding News-Topic Oriented Influential Twitter Users Based on Topic Related Hashtag Community Detection". *Journal of Web Engineering*, Vol.13, No.5&6, Rinton Press, pp.405-429, 2014. 査読有  
<http://www.rintonpress.com/journals/jweonline.html#v13n56>

[学会発表] (計3件)

- ④ Tomoya Noro, and Takehiro Tokuda. "Effectiveness of Incorporating Follow Relation into Searching for Twitter Users to Follow". *Web Engineering, Lecture Notes in Computer Science*, Vol.8541, Springer, pp.420-429, 2014. 査読有  
 DOI: 10.1007/978-3-319-08245-5\_27  
[http://link.springer.com/chapter/10.1007/978-3-319-08245-5\\_27](http://link.springer.com/chapter/10.1007/978-3-319-08245-5_27)  
 The 14th International Conference on Web Engineering, July 1-4, 2014, Toulouse, France.
- ⑤ Tomoya Noro, Atsushi Mizuoka, and Takehiro Tokuda. "Towards Finding Good Twitter Users to Follow Based on User Classification". *Information Modelling and Knowledge Bases XXVI, Frontiers in Artificial Intelligence and Applications*, Vol.272, IOS Press, pp.343-354, 2014. 査読有  
 DOI: 10.3233/978-1-61499-472-5-343  
<http://ebooks.iospress.nl/volumearticle/38510>  
 The 24th International Conference on Information Modelling and Knowledge Bases, June 3-6, 2014, Kiel, Germany.
- ⑥ Kristian Slabbekoorn, Tomoya Noro, and Takehiro Tokuda. "Towards Twitter User Recommendation Based on User Relations and Taxonomical Analysis". *Information Modelling and Knowledge Bases XXV, Frontiers in Artificial Intelligence and Applications*, Vol.260, IOS Press, pp.115-132, 2014. 査読有  
 DOI: 10.3233/978-1-61499-361-2-115  
<http://ebooks.iospress.nl/volumearticle/35694>  
 The 23rd European-Japanese Conference on Information Modelling and

Knowledge Bases, June 3-7, 2013, Nara, Japan.

## 6. 研究組織

### (1) 研究代表者

野呂 智哉 (NORO TOMOYA)

東京工業大学・大学院情報理工学研究科・助教

研究者番号：80410553

### (2) 研究分担者

徳田 雄洋 (TOKUDA TAKEHIRO)

東京工業大学・大学院情報理工学研究科・教授

研究者番号：30111644