

修 士 論 文

Twitterにおける論争の分析 及びその極化の予測

Analysis of Controversies on Twitter and Prediction of Their Polarization

東京大学大学院 情報理工学系研究科
電子情報学専攻

指導教員 豊田 正史 教授
氏 名 48-176442 張 翔

提 出 日 平成31年1月31日

概要

SNS の登場により情報技術に疎い人でも簡単に意見発信や他人との意見交換が行えるようになり、多様な意見を持つ人々との議論が容易となった。しかし、「原発再稼動」や「朝鮮学校無償化」などの政治的・民族的なトピックのように、議論のトピックが論争になるようなものであった場合、そのトピックに関心を持つ人々は違う意見を避け、同じ意見の人とのみ交流しようとする極化状態が進行する傾向が指摘されている。本来議論を行うのは他人と意見交換を行い、自分にかけていた視点を獲得することが目的の一つである。しかし、極化状態が進行した議論においてはもはや人々は多様な観点を獲得することはなく、むしろ自分と同じ意見のみを繰り返し繰り返し繰り返す他者の口を通して獲得するのみである。更に問題は、他者からも自分が持つ意見と同じ意見を聞くことで、そのような、異なる視点からの反論を欠いた意見が世間一般的で世論を代表したものだとして、誤認してしまうことである。もし人々がそのような多角的な視点を欠いた意見を持つと、それをもとに世論が形成され、最終的に選挙などの形で、多角的な視点到欠した政策が現実に影響を及ぼすかもしれない、そのような社会は社会の多様性の包括力を損ねる。そこで、極化状態、および極化状態をもたらす論争はなくすることが望まれる。しかし、いったん発生した極化によって対立構造が構築された後に、その構造を崩すことは容易ではないことが、既存の論争解析研究や社会学研究によって示唆されている。そのため、すでに発生した極化を解消するだけでなく、そもそも極化を進行させないように予防するというアプローチが必要である。そこで本研究では、ある論争を招く可能性のあるトピックを特定し、極化を抑制するための第一歩として、あるトピックが論争となるのかどうかを早期に予測するタスクを提案する。本研究では、この予測タスクを多クラス識別問題として定式化し、論争候補のトピックに関連したツイートのデータセットを作り、そこからあるトピックに関心を持つユーザ達の意見自体のテキスト情報や、ユーザたちの関係性ネットワークのグラフ特徴量などを特徴量として用いて教師有り学習により識別モデルを訓練し識別を行うことでこのタスクを達成し

た．本研究の貢献としては，論争をもたらすトピックの早期予測という新しいタスクを提案した点，論争を起こしたトピックのキーワードを作成した点，そして議論参加者間ネットワークの統計量や実際の議論中の投稿テキストが予測に有効に働くことを実証した点である．

目次

第1章	はじめに	1
1.1	マイクロブログにおける論争	1
1.2	本研究の目的と貢献	2
1.3	構成	3
第2章	関連研究	4
2.1	controversial なイベントの検知	4
2.2	発生した論争の解消	5
2.3	論争度の定量化	6
2.4	その他の議論・論争に関する研究	7
第3章	論争化予測のデータセット構築と RWC の分析	10
3.1	論争化予測のためのデータセット構築	10
3.2	論争度指標 RWC の分析	14
第4章	論争化予測手法	24
4.1	タスク設定	24
4.2	分類器	26
4.3	グラフ特徴量	27
4.4	テキスト特徴量	28
4.5	ユーザ特徴量	29

第 5 章	評価実験	31
5.1	使用データ	31
5.2	評価データでの予測精度と分析	32
5.3	特徴量分析	32
5.3.1	ablation テスト	32
5.3.2	単一種の特徴量テスト	32
5.4	事例分析	35
5.4.1	成功事例	35
5.4.2	失敗事例	38
第 6 章	まとめ	41
	謝辞	43
	参考文献	45
	発表文献	50

目 次

2.1 RWCが高くなるトピックと低くなるトピックのインタラクション グラフの例	8
3.1 トピックキーワード 1046 個の RWC プロット	17
3.2 (左) トピックキーワード出現当日の全ツイートで計算した RWC の 密度分布. (右) 出現当日から 10 日間の全ツイートで計算した RWC の密度分布.	18
3.3 論争が発生していないのに論争度が高くなってしまふトピックの例. ユーザグラフがほぼ木構造になってしまっている.	19
3.4 (左) ツリー状グラフ ($ V = E + 1$) になったトピックと (右) そ れ以外のトピック, それぞれについてトピック出現当日から 10 日間 の全ツイートで計算した RWC の密度分布.	21
3.5 (左) クラスタリンクと (右) それ以外のトピック, それぞれについ てトピック出現当日から 10 日間の全ツイートで計算した RWC の密 度分布.	22
3.6 トピックキーワードの出現当日の全ツイートで計算した RWC と, 出 現当日から 10 日間の全ツイートでの RWC のプロット. ただし, ト ピックの出現当日のツイートから構築したグラフが完全にツリー状 ($ V = E + 1$) なトピックを除外している.	23

4.1	学習データと開発データ合わせたを 4387 件のトピックを $(x, y) = (RWC(1day), RWC(10days))$ でプロットしたもの. $RWC(10days) - RWC(1day)$ の大きさを 3 クラスに分けている.	25
5.1	成功トピック「憲発議」の出現当日と 10 日後のインタラクショングラフ.	37
5.2	成功トピック「新潮 45 編集部」の出現当日と 10 日後のインタラクショングラフ.	37
5.3	成功トピック「コラボ装備」の出現当日と 10 日後のインタラクショングラフ.	40

表 目 次

3.1	賛成パターンにヒットしたツイートの例	12
3.2	反対パターンにヒットしたツイートの例	13
3.3	トピックキーワードのサンプル	15
3.4	ツリーグラフ, CC が 0 となるグラフができるトピックの数	17
4.1	使用した特徴量. 特に記載のない場合はすべてスカラ値.	30
5.1	test データでの識別性能	33
5.2	開発データでの ablation test の結果の各クラスおよび全体の f1 値.	33
5.3	開発データでのグラフ特徴量を 1 種類だけ使用して学習を行った結果 (<code>linearSVC(tol = 1.0e - 5, C = 5.0).</code>)	34
5.4	予測の成功したトピックの実例	36
5.5	予測の失敗したトピックの実例	39

第1章 はじめに

1.1 マイクロブログにおける論争

スマートフォンやタブレット端末などのモバイル機器や、マイクロブログの普及により多くのユーザが身の回りの現実世界の出来事をウェブ上に投稿するようになって久しい [1,2]. 例えば, 世界中で利用されているマイクロブログサービスの一つである Twitter¹ では2018年1月現在, 1日平均5億件を超える投稿がなされている². マイクロブログは通常, 他ユーザの投稿に返信を行う機能のほか, 投稿の共有やお気に入りなどの機能が備わっており, ユーザ同士のインタラクションをより促進する仕組みとなっている [3,4].

これらの性質を備えたマイクロブログにおいては, 日々様々な現実世界の出来事についての言及が行われ, 時にはそれについて他ユーザとの意見の交換, いわゆる議論がしばしば発生する. ここで, 「アメリカ大統領選」や「原発の再稼働」のようなセンシティブで物議を醸すトピックについての議論は, ユーザ間に賛成, 反対といった明確な対立構造が生まれ, それぞれ自分の立場の正しさを主張して論じあう, いわゆる論争という状態になりやすい. このような状態に陥ると, ユーザが同じ意見を持つユーザ同士の間でのみ交流する事態を招き, 結果として極端な意見の増幅に繋がる³. この意見の極化はその解消が困難であることが報告されており [8,9], 最終的には対象のトピックにおいて適切でない攻撃的な意見や, 間違った情報が論争外にまで拡散される原因となる [5,10–12].

¹<https://twitter.com/>

²<https://blog.hootsuite.com/twitter-statistics/>

³これらは Echo Chamber [5] とも呼ばれている. また, 検索エンジンのパーソナライズによりユーザが見たくない情報が隠蔽化される Filter Bubble [6,7] とも関連している.

このような事態を避けるには、マイクロブログに出現するトピックを監視し、論争になりそうなトピックを発見しその論争化を予防する措置を取ることが必要となる。その足掛かりとして本研究では、マイクロブログ上に新しく出現したあるトピックについての議論が、将来的に論争に発展するかを予測するというタスクを設定し、教師あり学習の手法によりこれを解くことを提案する。

教師あり学習のためには分類器の学習を行うためのラベル付きデータセットが必要であるが、ある特定のトピックについてそれが将来的に論争になるか否かをラベル付けしたものは存在しない。そのため本研究ではまず既存研究で作成された、なんらかの事物に対するユーザの議論のスタンスを検出する研究 [13, 14] で得られた意見についての賛否表明パターンを用いることで、ユーザによって賛否の分かれるトピックを示す単語を自動的に収集し、論争となったトピックのデータセットを構築した。さらに、このデータセットを用いて論争がどのようなトピックで発生するか、また論争の発生の予測に必要な特徴が何であるかを確認するための分析も行なった。

トピックの論争化の予測実験では、議論の対象を示す単語として抽出したトピックが将来的に論争へと発展するかどうかを予測する問題を、そのトピックの（既存手法により計算される）論争の度合い (Random Walk Controversy; RWC) [15] が将来的に大幅に増大するか、大幅に減少するか、または大きく変化しないかを分類する3値分類問題として設定し、それを解く教師あり分類器を構築した。

実験結果の分析により、論争となるトピックの予測には、トピックの出現初期に投稿されたツイートのツイート内容、ユーザ間のインタラクショングラフの特徴量、およびそのグラフをグループに分割したときの各グループ間の分離度等の手がかりが有効であることを示した。

1.2 本研究の目的と貢献

本研究の目的は、マイクロブログ (Twitter) においてあるトピックが発生したときに、そのトピックが論争になるかを予測することによって事前に予防を可能とすることである。本研究ではあるトピックが一定期間経過したときに発生時と比べて

論争状態になっているかどうかを予測実験を行った。本研究の貢献としては、以下の3点が挙げられる。

1. マイクロブログにおけるトピックの論争化の予測という新たなタスクを提案した点
2. 論争・非論争トピックの大規模なデータセットの自動的に構築する手法を提案し、実際に構築したデータセットを分析し論争がどのようなトピックで発生するかを分析した点
3. 発生したトピックが発生時と比べて将来的に論争状態になるかどうかを予測する分類器の学習に用いる特徴量を考案し、実験によってそれらの有効性の評価を行った点

1.3 構成

本論文の次章以降の構成は以下の通りである。

- 第2章 マイクロブログから論争を検知する研究及び論争の解消・解決を行なう関連研究について述べる。
- 第3章 論争化の予測のためのデータセット構築手法を提案し、実際に構築したデータセットを論争度 (RWC) を用いて分析する。
- 第4章 論争化の予測手法について、分類器や特徴量の詳細について述べる。
- 第5章 論争化の予測実験を行った結果とその評価について述べる。
- 第6章 本研究の結論と今後の課題について述べる。

第2章 関連研究

本研究に関連する研究としては，論争になりそうなイベント（controversial なイベント）を検知するイベント検知の研究と，論争の解消・解析を行う研究が挙げられる．以下で順に紹介する．

2.1 controversial なイベントの検知

ウェブやマイクロブログ上で発生するイベントの検知を行う研究は数多くあるが [16–20]， controversial なイベント，つまりユーザによってそのイベントへの賛否が分かれるようなイベントを検知対象とした研究は多くない．Popescu らは Wikipedia ¹から著名人の名前の一覧を取得し，1日ごとに各有名人について言及を行なったツイートのデータセットを作り，その内のいくつかを人手で controversial であるかそうでないかをアノテーションしてラベル付きデータセットを作成した．そして，言及ツイートのテキスト特徴量や，言及ツイートをしたユーザたちのつながりから計算されるグラフ特徴量，また外部ニュースサイトで同時期に対象の有名人が言及されたニュースが存在するかという外部ニュース特徴量をデータから抽出し，所与のイベントが controversial であるかどうかを判定する識別モデルを学習し実際のツイートに適用することで controversial なイベントの検知を行なった [21]．

この研究では高い精度で controversial なイベントの識別を実現している．しかし，Wikipedia に名前が載るような著名人の人名をデータセットのシードとしているために，実際の検出時においても著名人の関わるイベントしか捉えることができない

¹<https://en.wikipedia.org/wiki/>

いという問題点がある．実際には日本が国際捕鯨委員会 (IWC) を脱退したことの是非を問う「IWC 脱退」や，不登校児童への対処法について議論を引き起こした NHK の「不登校特集」など，著名人がトピックの中心に登場しないものの，controversial となるトピックは少なくない．本研究では，あらゆるトピックに共通して現れやすい賛否表明のパタンを用いて論争が起きそうなトピックについてのツイートを自動抽出することで，トピックの偏りを抑制している．

2.2 発生した論争の解消

発生した論争そのものの解消を目指した研究がいくつか存在する．Kiran らは対立する意見を持つ2つのグループを対象に，各グループの有力者のなかから，対立する側の意見にも耳を傾ける可能性の高い有力者を選び出し，その人物らを結びつける (Twitter におけるユーザ推薦のような機能によって) ことで，対立を緩和させ，論争を解消することを目指した [22]．また Gillani らは，ユーザに自分と異なる意見にも耳を傾けさせる意識を高めるための手法を開発した．具体的には，特定のトピックに関心を持つユーザたちのつながりをグラフ化し，それをいくつかの異なる意見をもつユーザグループに分割したうえでユーザの一部にそのグラフを提示することで，ユーザ自身の情報獲得 (ユーザが属するグループ) のグラフ上におけるカバー率を自覚させることでユーザの意識を高める手法となっている．評価においては，対象ユーザのその後のフォロー傾向などを観察することで手法の妥当性を検証した [23]．Vo ら [24] は，ウェブ上でのフェイクニュースの共有により生じる誤報の拡散や議論の不健全化を防ぐための新たなアプローチを提案している．具体的には，ガーディアンと呼ばれる，ウェブ上に投稿された情報の事実確認を積極的に行っているユーザを発見し，検証が必要な情報の URL を彼らに推薦することで，事実確認への従事をより積極的なものとするを狙っており，評価実験によって推薦精度の向上が確認されている．

Kiran らや Gillani らの研究は論争の解消にあたって一定の成果を出しているものの，解決の対象となるトピックを人手で選定しなければいけないという問題が残さ

れている。Vo らのアプローチには、議論を健全化させることが可能なガーディアン①の数には限りがあるという制約が存在する。一方で本研究は人手の選定無しに将来論争が起きるトピックを探し出す手法を提案しているため、本研究によって今後の深刻な論争化が生じると予測されたトピックに対して、これらの研究のアプローチを重点的に用いることで論争の極化を抑止するという相互補完的な関係を実現できる。

一方で、Bail らは対立する 2 つのグループのそれぞれの構成員に対して、対立している側の情報を継続的に摂取してもらうことの影響を調査した結果、対立が解消されるわけではなくむしろより対立が激化する傾向にあるということを報告している [8]。

Bail らの報告は、深刻化してしまった論争への対処が困難であることを示唆している。これらの研究を通じて明らかになってきたことは、一度極化してしまった論争を解決、軽減するための有効な方策が限られているということである。このことは、論争に対する予防的なアプローチの必要性を示唆している。そこで本研究では、論争の予防のための第一歩として、将来発生しうる論争を早期段階で予測するという問題に取り組む。

2.3 論争度の定量化

本研究で対象とするウェブ上の論争を対象とし、それらの定量化を目指した研究がいくつか存在する。Adamic らは 2004 年のアメリカ大統領選挙の前 2 か月にわたって政治系ブログとその記事中にある他の政治系ブログ、関連サイトへのリンクを集めて、左派（急進、リベラル、Democrat）のブログコミュニティと右派（保守、コンサバティブ、Republican）のブログコミュニティを取得・解析し、両コミュニティの特徴やその違いを調査した。具体的には、各コミュニティのユーザらの記事へのコメント等からブログコミュニティ間の関係性をグラフ化し、そこから分析のためのグラフ統計量を算出した。その結果としては、政治に関する論争となるような事例においては「グループ間交流がグループ内交流に比べて非常に疎となる」という

特徴が見られることを報告した [25].

Kiran らは, Adamic らが報告した「グループ間交流がグループ内交流に比べて非常に疎となる」という論争事例の特徴に着目することで, マイクロブログ上のあるトピックがどの程度論争を引き起こしているかを Random Walk Controversy (RWC) という指標で定量化した. 具体的には, あるトピックに関心をもつ Twitter ユーザのインタラクションからグラフを作り, それを2つのグループへ分割し, グループ間を横断するエッジの割合をランダムウォークを用いて計算することで RWC を算出し, あるトピックがどの程度論争を引き起こしているかを定量化する手法を提案した [15]. RWC が高くなるトピックと低くなるトピックそれぞれに関するインタラクショングラフの例を Fig. 2.1 に示す. RWC の高さが2つのグループの断絶を表す指標となっていることが視覚的にも確認できる.

対して Gao らは, 議論の健全さとしての論争度を定式化することを目的として, スラッシュドット²という掲示板サイトの任意の投稿を種とする議論を木構造で表現し, 木の深さとコメント数から計算される h-index を論争度の定義とした [26].

本研究は Kiran らと同様にマイクロブログを対象としているため, 彼らの RWC をもとに対象のトピックが将来論争となるかを予測するタスクを設定した. そのためデータセット構築の際も, 論争化するトピック候補の収集後に各事例について RWC を算出し付与している. 付与した RWC を参考に各トピックを分析し, 論争化予測の手がかりとなる特徴量の考案を行っている.

2.4 その他の議論・論争に関する研究

Yoshida らは政治的な情報の拡散を, 政治への関心が薄いユーザにどの程度拡散するかという観点から情報の拡散力を評価する手法を提案し, その評価を行った [27].

Kakiuchi らは番組配信アプリの使用履歴を解析してユーザの好みの変化傾向を分析することで, Echo Chamber の発生過程にも応用可能となり得る情報獲得の変化傾向について調査した [9].

²<https://slashdot.org/>

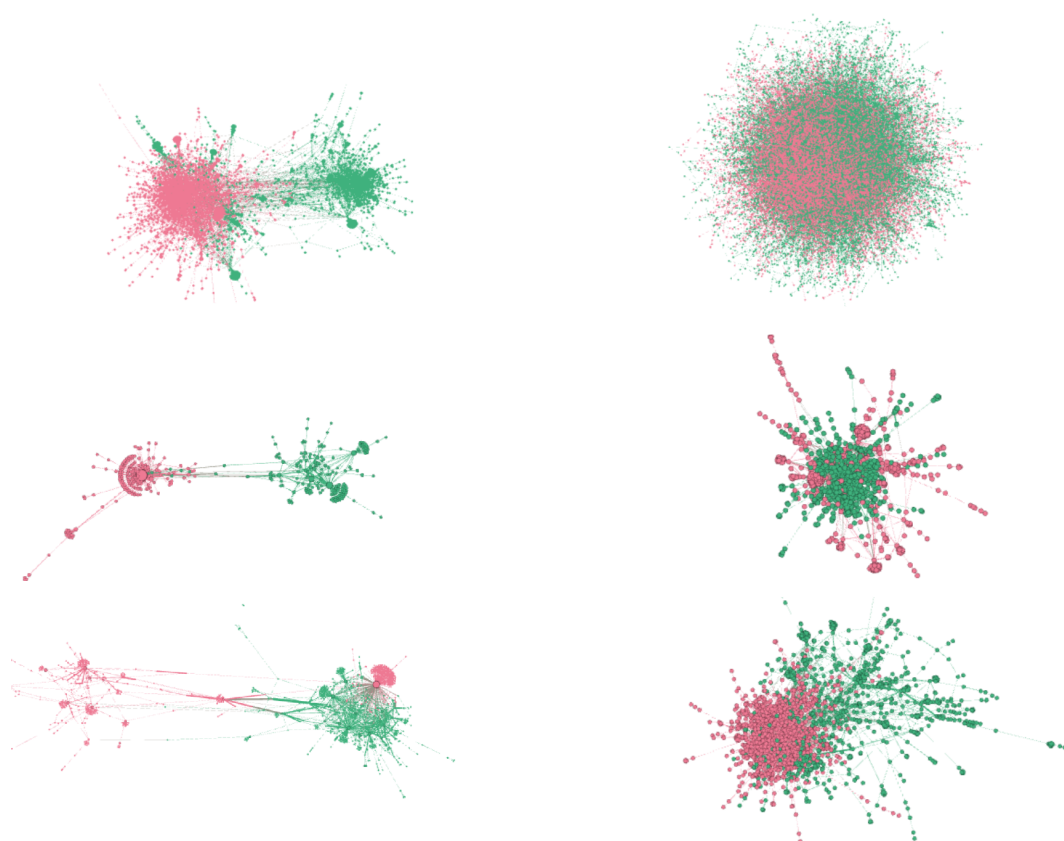


Fig. 2.1: RWC が高くなるトピック（左）と低くなるトピック（右）のインタラクショングラフの例. 左は上から順に「保育園落ちた日本死ね」、「築地移転」、「スパコン詐欺」について話題が出現した当時にツイートを行ったユーザたちのユーザグラフ. 右は「コミックマーケット開催」、「関ジャニ昴脱退」、「豊田真由子議員謝罪」についてのユーザグラフ

Benjamin らは、論争という概念に関しては長年の社会学的な知見が数多く存在するにもかかわらず、最近の計算機で論争をモデル化しようとする研究ではそれが活かされていないとし、論争に関する人文科学研究の知見を元に論争をモデル化し、さらにそのモデルを用いて既存の計算機による論争の研究を捉え直した [28].

Cheng らは、ウェブ上で行われたコメントの応酬を分析することで、ユーザに文句をつけられやすい意見がどのような性質を備えるか、またはどのようなときにユーザは文句をつけやすいかを示した [29]. 具体的には 2012 年の 12 月から約 8 ヶ月の間に CNN³ のニュースサイトにつけられたコメント約 1600 万件（約 90 万ユーザによる）の応酬を分析することで、日曜日の午後に文句をつける活動が活発化する傾向と、過去に文句を言われていたユーザほど以降もより文句を言われる傾向にある、ということを示した.

Johannes らは既存の議論研究は domain-specific なものしか存在しない点を指摘し、議論の中核となる要素は Claim（主張）であるという主張のもとで、Claim を獲得しようとする既存研究を 6 つ挙げてそれらすべての Claim の定義を確認した. その上で、それぞれのデータセット中での Claim の概念は大きく異なっているが、lexical な共通点があったと主張し、未知の domain への適用の可能性を示した [30].

本研究では、これらの知見を踏まえて、将来の論争度の予測を行うための特徴量を設計する. 特に、Cheng らの研究において、ウェブ上での議論では周囲のユーザの過去の行動が大きな影響を与えることが報告されていることから、予測対象の議論に参加しているユーザが過去に参加した議論に関する情報を特徴量として利用する.

³<https://edition.cnn.com/>

第3章 論争化予測のデータセット構築 と RWC の分析

本章では，論争が発生し極化が進行したトピック・しなかったトピックのそれぞれについて分析し，どのような要素が論争トピックを特徴づけるかについて考察を行う．

第1章で述べたように，トピックの論争化の予測のための教師あり分類器を構築するためにはラベル付きのデータセットが必要であるが，2019 年現在では論争を起こしたトピックを網羅したデータセットは存在しない．そのため我々は最初に Twitter から論争候補トピックと，そのトピックに言及したツイートを収集し，論争が発生したトピックとそうでないトピックへと自動的に選別することで論争が発生したかどうかのラベルを割り当て，論争候補トピックのデータセットを構築することを目指した．さらに，構築したデータセットのトピックについて 2.3 節で述べた RWC を用いてどのようなトピックが論争化するか・しないかという点に着目し分析を行うとともに，論争化の予測の手がかりとなる特徴量について考察した．

3.1 論争化予測のためのデータセット構築

本節では，論争の対象となるトピックを自動で収集する手法について説明する．ただし，本研究ではトピックは特定のキーワードで表現されるものとし，そのようなキーワードをトピックキーワードと定義し，それらを収集することをトピックの収集とした．以下ではトピックキーワードの収集方法について説明する．

ここで，収集しようとしている新しく出現したトピックに関して定義としては，

Twitter 上においてあるトピックのトピックキーワード（が含まれるツイート）が投稿されてから過去一定期間を遡った間に、同様のキーワードが一度も投稿されていないものとする。これは、すでに広く議論されているトピックを除く意図がある。また、論争の発生したトピックの定義としては、そのトピックについてのツイート集合中でトピックに賛成を示すツイートと反対を示すツイートが両方見られているものとする。これは、論争においては賛成・反対の明確な対立構造が発生するという考えに基づいている。以降では、上述した過去一定期間を過去 30 日間とみなし、その間に同じトピックキーワードが出現していなければそのトピックを新出トピックとする。

まず、論争の発生したトピックを収集するため、Twitter ユーザによって賛成または反対が示されているトピックキーワードとして、Sasaki らが作成した「賛否表明パターン」[13]を用いた。これは Twitter のユーザのツイートから、対象ユーザが賛成・反対を表明している事物を抽出するために作られたものであり、「X/大賛成です」、「X/がいい」や「X/絶対/反対」、「X/やめて」というような、人が何かに賛否を表明するときに使われるパターンが賛否それぞれに対して 100 件ずつある。このパターンにマッチするツイート文の例をマッチパターンとともに Table 3.1, Table 3.2 にそれぞれ示す。

本研究ではこのパターンを用いて、Twitter 上で賛成・反対の対象となっているキーワードを以下の手続きで集めた。

1. ツイートを正規化し、句読点に基づき文分割する
2. パターンを含む文だけを抽出する
3. 抽出した文から名詞連続を抽出する
4. 名詞連続のうち「固有名詞」と「サ変活用名詞」を両方含む連続名詞を抽出する
5. 残った連続名詞のうち、直近の 30 日間に出現していないものを抽出する

3.1 論争化予測のためのデータセット構築

Table 3.1: 賛成パターンにヒットしたツイートの例

賛成パターン	マッチしたツイート文
してくれ	都営地下鉄って言ってんだから、全部地下にしてくれよ
好評	13巻アニメイト限定セット」好評予約受付中!
してください	1日1回プレイできますので、是非プレイしてくださいね♪
好評	ご好評につき、以下の日程にて追加公演が決定いたしました!
がいい	プラトニック・ラブと称して、精神的恋愛を高尚だというのも妙だが、肉体は軽蔑しないほうがいい。
大好き	地獄の門大好き!
が必要, 必要	これからの時代の自由には「分人主義」が必要だ
おめでとうございます	遅くなりましたが、お誕生日おめでとうございます
お願いします	チャンネル登録お願いします!
好きだ	編集者時代に神保町の店(名前は失念して場所もうろ覚え)のミートドリアが好きだったことを思い出す
するべき	大事な目的は電力の安定供給再生可能エネルギーだの原発ゼロだのはあくまでも手段の選択目的を確実に達成出来る手段を選択するべき
賛成	回答者数は少ないですが、ミサイル防衛は80%以上の人が賛成。
歓迎	僕も携帯代の値下げは大歓迎だけど、値下げは県知事の権限でできるものじゃないし、県政の課題でもないですよ。
満足	テアトル新宿で公開中の傑作映画『寝ても覚めても』をより深く理解できるとともに、厳しい評論で有名な蓮實重彦氏をはじめとする濃厚な対談の数々を読むことができ、超大満足です。

Table 3.2: 反対パターンにヒットしたツイートの例

反対パターン	マッチしたツイート文
やめて しない	それをもう金輪際やめて欲しいですね。 大井町線止まってりゃ諦め付いたんですけど下手に動いて るもんだから出社しないとならないという
やめて の危険性 に反対 反対	態々画像付きで投稿しなオスのやめて コラム:世界で高まる「ドル不足」の危険性 僕がこういう意見に反対しているのは, 「そんなの嘘だ」と樹木希林が反対して、彼女の発案でセ リフを「そうでない方はそれなりに」と変えさせたと.
禁止	小学校の運動会の案内が来ていたんだけどまだ「校内での ドローンによる撮影は禁止です」の文言はなかった.
許すな 許さない	アベ改憲を許すな! 駅のエスカレーターとか電車に乗る時とか、横から割って 入ってくる奴を許さない.
は不要 いらない	責等の理由による個人の活動自粛は不要 屋内に喫煙室などいらないし不特定多数の人が集まる場所 に喫煙環境を作る必要がない..
するな は無理	大阪市はセンター求人を妨害するな! 経済的な問題もあって3人目以降は無理です中絶めちゃ多 いよ.
の問題, は嫌だ	俺は色味の問題で照明開発に邪悪な怨念をもってるけど(黄 色がかったしょんべん液晶は嫌だけどブルーライトカット してねってやつ), 世の中には有害紫外線カット日焼けマシ ン(波長が違えばダメージが無い嘘理論)とか作らされてる 人も居るし光は闇

操作4を行うのは、実際に操作3までをおこなった結果を人手で確認したところ、論争を引き起こしそうな名詞連続は「固有名詞」と「サ変活用名詞」の両方を含んでいることが多かったためである。具体的には抽出したパターンを観察したところ「おしどりマコ擁立賛成」のように「A(が)B(することに賛成)」のようなパターンが多く見られたためである。操作5を行うのは、前述したように本研究では新たに出現したトピックのみを対象とするためである。

この一連の操作1から4までを2018/8/1から2018/9/20までのツイートデータに対して行った。そして操作5において、9月分のキーワードを獲得した。元となるデータとして、著者の所属研究室で収集しているツイートデータ¹から、2018年の9月のツイート約6億件(603,048,206)を用いた。このうち、日本語を含むツイートは約1億件(114,290,698)、パターンにマッチした文は約1千万件(10,799,911)。最終的に得られたキーワードは約10万件(102,619)であった。

こうして得たトピックキーワードの中にはノイズもあることが予想されるので、各キーワードについて、そのキーワードを含むツイートで出現当日のものが50件に満たないものは除外した。最終的に、2018/9/1から2018/9/20までの20日間について、1,046件のトピックキーワードが得られた。トピックのサンプルをTable 3.3に示す。

3.2 論争度指標 RWC の分析

本節では論争化予測のベースとなる2.3節で述べたKiranらが考案したRWCを3.1節で収集したトピックに適用し、分析と考察を行う。Twitter上におけるトピックの論争度の指標であるRWCは、Twitterユーザ間のつながりをもとにしたインタラクショングラフから計算される。しかし、このグラフがツリー状であったり、クラスタリング係数などの統計量が0になるようグラフであると、そのグラフが出て

¹2011年3月から継続的に収集している。2011年の3月に30名程度の著名人を選び、その人物らにメンションを送ったり、その人物らのツイートをRTしたりしたユーザを再度ターゲットとし、データセットを拡大していった。

3.2 論争度指標 RWC の分析

Table 3.3: トピックキーワードのサンプル

controversial		non controversial	
総裁選延期	北海道節電率	軍事的緊張緩和	青葉モカ誕生祭
商業捕鯨再開案	不登校特集	おすすめオーブ	産業廃棄物扱い
ポンプ車手配	分散型発電	ソーシャルゲーム化	イベ参加中
日露平和条約締結	豊洲移転中止	サムスピ新作	裏技募集
普天間辺野古移設	北方領土放棄	広島公演2日目	3日目終了
朝鮮学校補助金廃止	代表戦中止	限定ありす	本日第9話
就活指針廃止	有害サイト認定	初回限定封入特典	にも関係
就活ルール廃止	仏海軍派遣	HUNTER 再開	15日発売
日米同盟重視	小川監督続投	横浜アリーナ追加公演	発電所停止
小泉電撃訪朝	外国人枠拡大	飲酒運転ひき逃げ	直筆サイン入
消費増税阻止	ロシア外遊	山盛りポテトフライ	画力対決
飯島内閣参与	石狩湾新港発電	ドラえもん誕生	eストア

きたトピックが論争でないのに高い RWC を示す傾向があることを事前の分析により確かめた。RWC をもとにトピックについての論争化予測の教師データを構築する際には、論争であるトピックには高い RWC が、そうでないトピックには低い RWC がついていることが望ましい。そのため、本節では RWC がグラフの形状や統計量と関係があることを示し、また、RWC が 3.1 節で収集した実際のトピックに適用した際にどのように振舞うのかを観察することで、所望の振舞いをする RWC の条件について考察する。

RWC は -1 から $+1$ までの間の実数値であり、あるトピックに関心をもつ Twitter ユーザ達から、ノードがユーザでリツイートがエッジであるようなインタラクショングラフを作り、そのグラフを 2 つのグループに分割したときの、グループ内部の交流の頻度と比較した両グループ間の交流の頻度に相当する。RWC が高ければ高いほど、そのトピックに関心をもつユーザたちは強く 2 つのグループに分割されているといえる。まず、3.1 節で収集した 1046 個のトピックについて、トピックの登場初日のみの RWC (以下、 $RWC(1)$) と、初日から 10 日間のツイートを使った RWC (以下、 $RWC(10)$) のプロットを Fig. 3.1 に、またその密度分布を Fig. 3.2 に示す。

ここで、前述した通りいくつかのトピックについて RWC 値を観察したところ、Fig. 3.3 のように、実際には論争が起きていないのに RWC が高くなってしまうケースが確認された。そのようなケースでは (i) インタラクショングラフの形状がツリー状であるか、(ii) クラスタリング係数が 0 であった。ここで、この 2 つの特徴が RWC に影響を与えるかを確認するために、(i) インタラクショングラフの形状がツリー状であるトピックとそうでないトピックとで RWC の振る舞い、また (ii) クラスタリング係数が 0 であるトピックとそうでないトピックとで RWC の振る舞いを分析する。ここでの目的は (i)(ii) と、トピックが論争であるか (RWC が高いか) との関連を観察することである。そのため筆者が 3.1 節で収集した 1,046 件のトピックキーワードから 100 件トピックをサンプリングしたものに論争であるかどうかを 2 値でアノテーション²し、このアノテーション付きのトピックリストを用いて分析を行う。ア

²アノテーションについては、対象のトピックが Twitter で出現してから話題として収束するまで (10 日間) のツイートを観察し、キーワードが指し示している人や物について、好意的な意見の

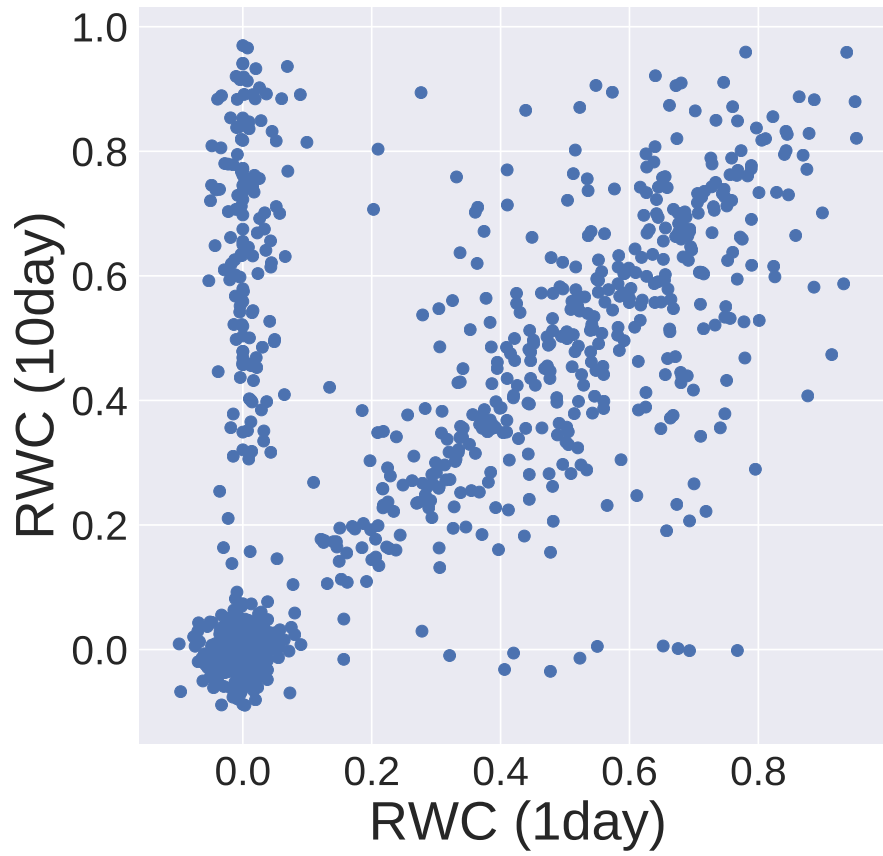


Fig. 3.1: トピックキーワード 1046 個の RWC プロット

Table 3.4: ツリーグラフ, CC が 0 となるグラフができるトピックの数

	$CC = 0$	$CC > 0$
tree graph	479	0
non tree graph	194	373

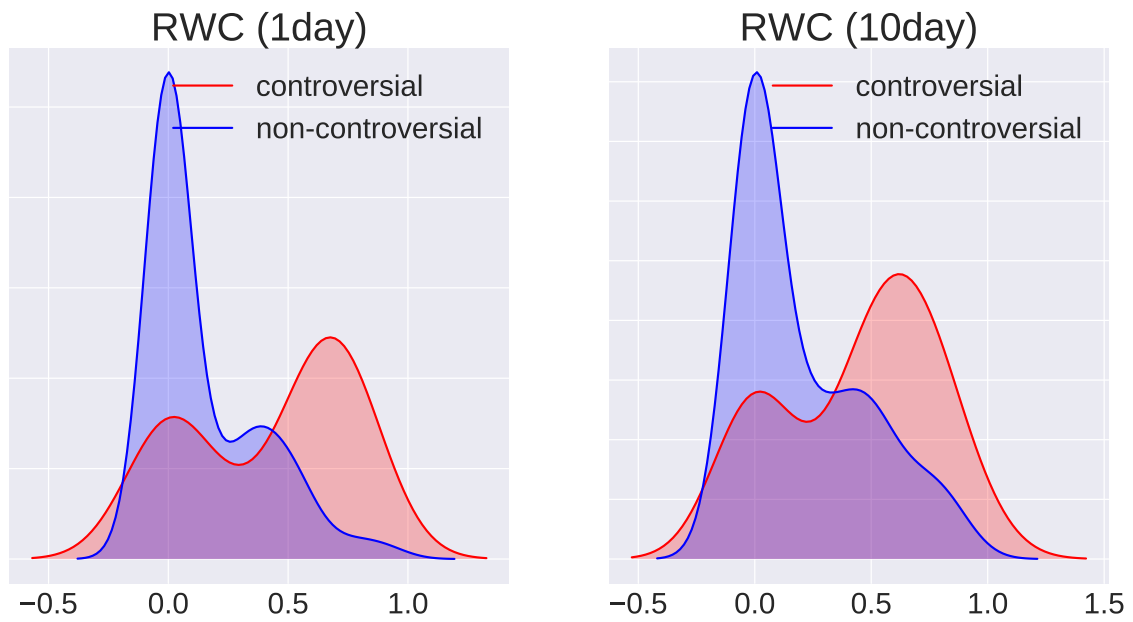


Fig. 3.2: (左) トピックキーワード出現当日の全ツイートで計算した RWC の密度分布. (右) 出現当日から 10 日間の全ツイートで計算した RWC の密度分布.

3.2 論争度指標 RWC の分析

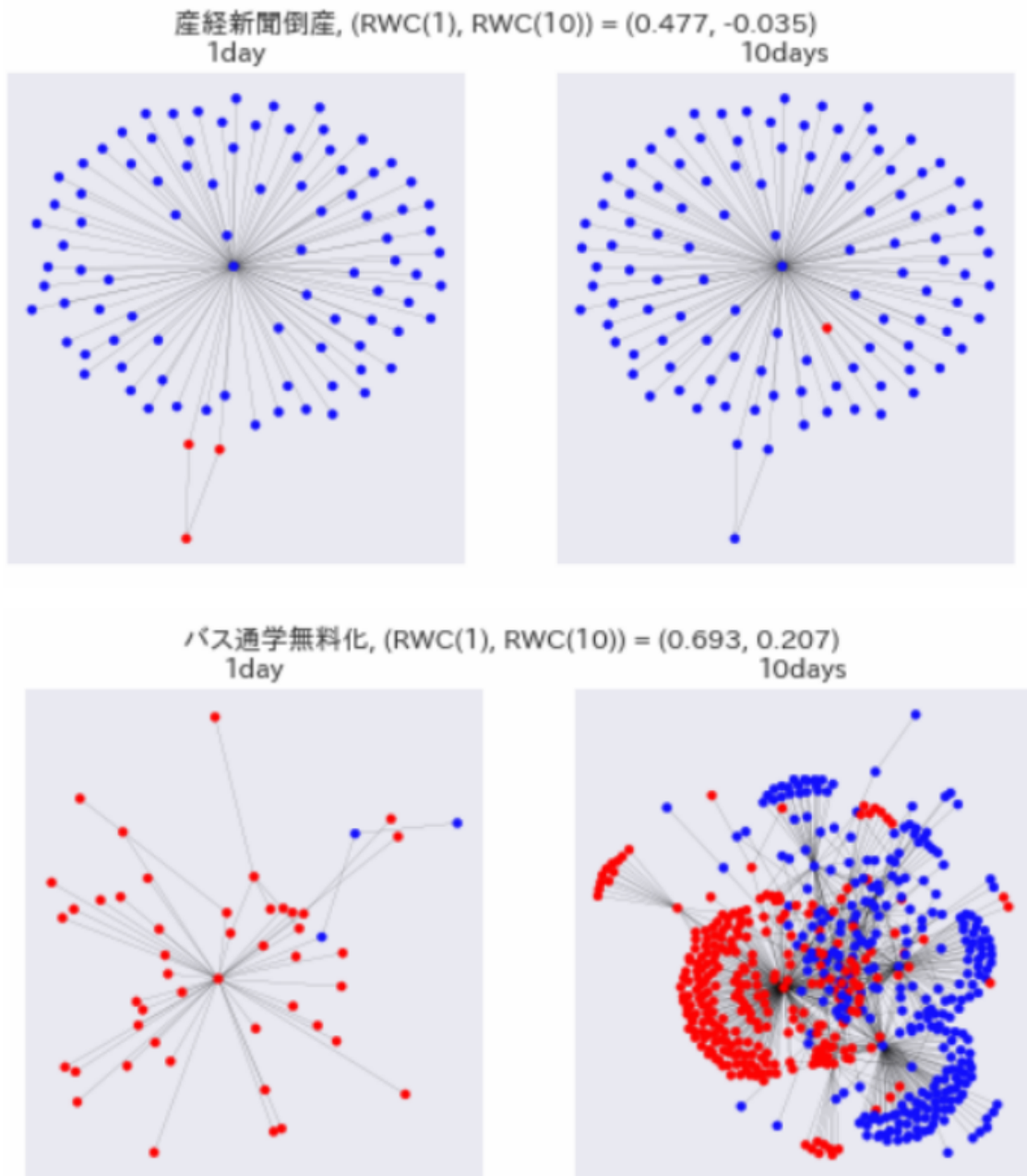


Fig. 3.3: 論争が発生していないのに論争度が高くなってしまいうトピックの例。ユーザグラフがほぼ木構造になってしまっている。

ノテーションの結果，論争であるトピックは 20 件，論争でないトピックは 80 件であった。

それぞれの結果を Fig. 3.4, Fig. 3.5 に示す．また，Table 3.4 にケース (i)(ii) となるトピックの数も示す．Fig. 3.4 から，ケース (i)(ii) のどちらにおいても該当するトピックが論争であるかどうかによらず RWC が 0 に集中していることがわかる．よって，予測の対象トピックは，そのトピックから作られるグラフがツリー状でもなくクラスタリング係数も 0 ではないようなものが望ましい．しかし，ツリーグラフを除去した場合には 45% のトピックが残るのに対し，クラスタリング係数が 0 であるグラフのトピックを除去すると，Table 3.4 から全体の約 35% ほどしか残らずサンプル数が著しく低下してしまう．そのため，予測の対象とするトピックはトピック登場初日のインタラクショングラフがツリー状でないものだけを取り扱うこととして，クラスタリング係数に関しては制限をかけないものとする．

最後に，Fig. 3.6 に，トピックの登場初日のツイートから構築したインタラクショングラフがツリー状であるトピックを除外した RWC(1) と RWC(10) のプロットを示す．当初の予測通り，概ね $y = x$ 状に分布している．また，RWC が大きく増加しているトピックは多いが，大きく減少しているトピックはほとんど無い．

ツイートと否定的な意見のツイートとが両方確認できれば 1 を，そうでなければ 0 をアノテーションした．

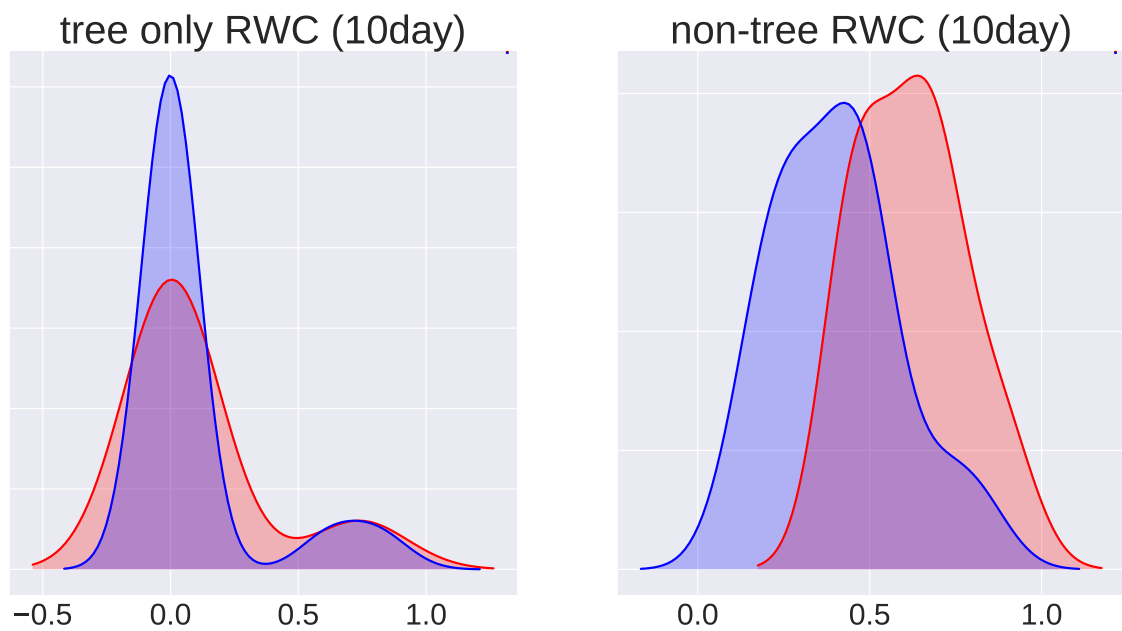


Fig. 3.4: (左) ツリー状グラフ ($|V| = |E| + 1$) になったトピックと (右) それ以外のトピック, それぞれについてトピック出現当日から 10 日間の全ツイートで計算した RWC の密度分布.

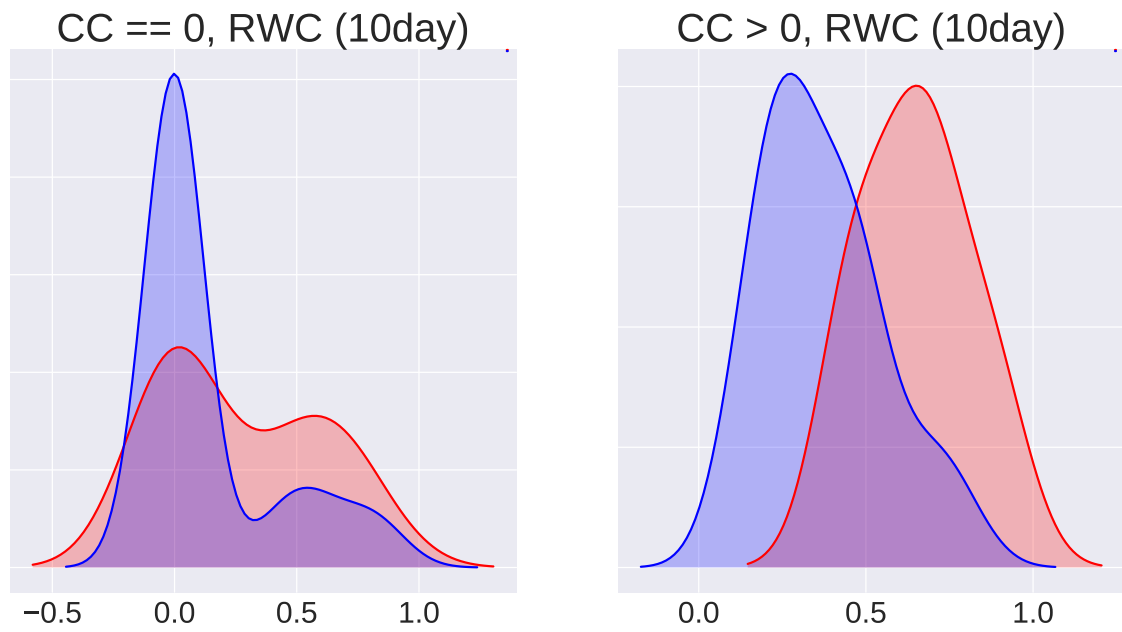


Fig. 3.5: (左) クラスタリンクと (右) それ以外のトピック, それぞれについてトピック出現当日から 10 日間の全ツイートで計算した RWC の密度分布.

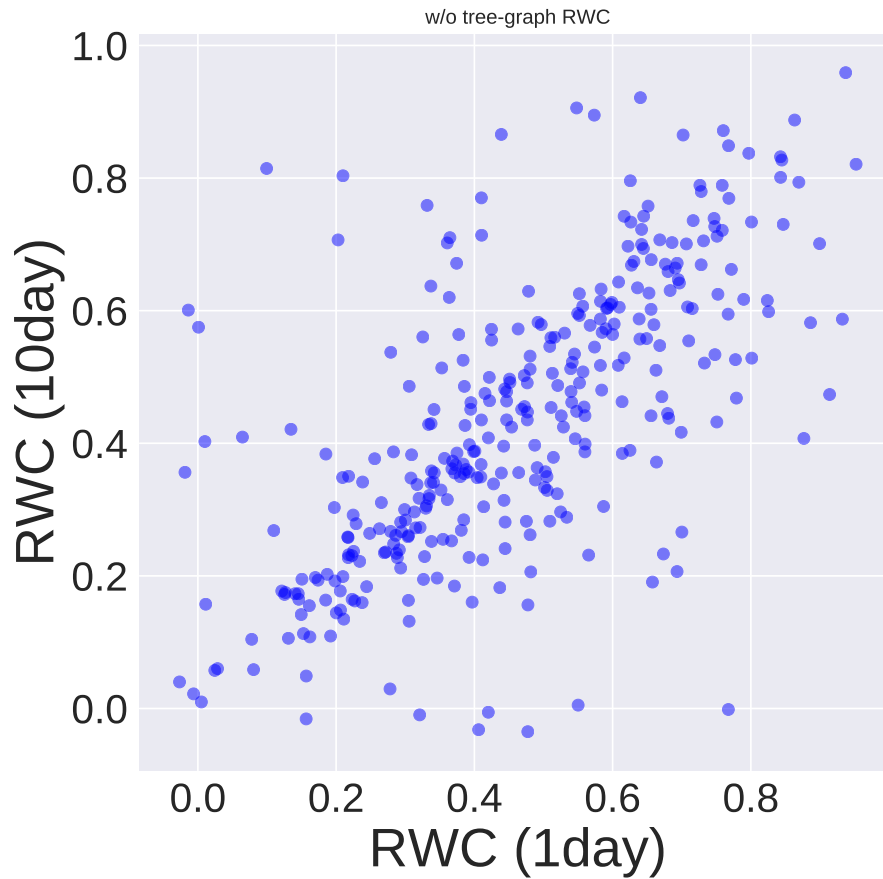


Fig. 3.6: トピックキーワードの出現当日の全ツイートで計算した RWC と，出現当日から 10 日間の全ツイートでの RWC のプロット．ただし，トピックの出現当日のツイートから構築したグラフが完全にツリー状 ($|V| = |E| + 1$) なトピックを除外している．

第4章 論争化予測手法

本章では本研究で提案する論争化トピックの予測手法に関して述べる．考案した特徴量を実際に用いて分類器を学習することで，対象の事例（トピック）が将来論争化するか否かの予測を行う．具体的にはマイクロブログに新しく出現したトピックの将来の RWC が出現当日と比較して上昇しているか，下落しているか，または大きく変化しないかの3値に予測する．

4.1 タスク設定

Fig. 3.6 からわかるように，大多数のトピックは直線 $y = x$ の近傍に分布し，いくつかのトピックが大きく $y = x$ から離れて同図の左上や右下に分布している．そこで，本研究では Fig. 4.1 のように，所与のトピックが (a) 直線 $y = x$ の近傍に来るか，(b) 図の左上の領域に来るか，(c) 図の右下の領域に来るかを予測するタスクを設定する．このタスクを十分な精度で解ければ，あるトピックが論争になるかどうかや，あるトピックで論争が発生しているときそれが10日後に収束しているのかどうか予測することが可能となる．

本研究ではこのタスクを3.1節で作成したデータセットを用いた分類器によって解くことを目的とする．特徴量に関しては，3章における分析の結果を踏まえて選定を行う．

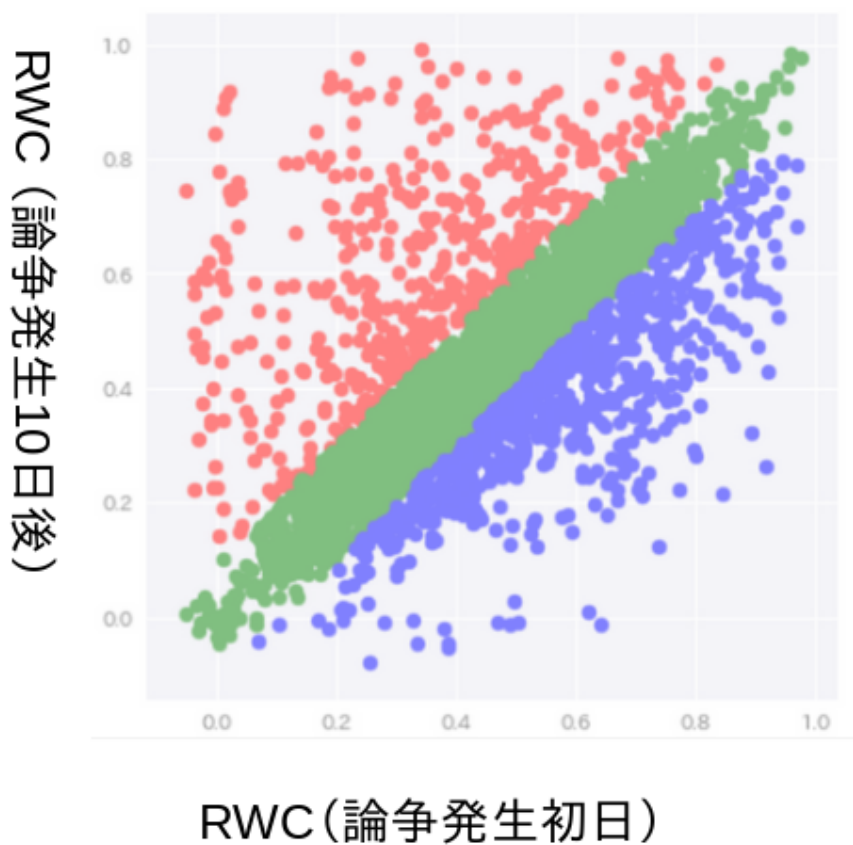


Fig. 4.1: 学習データと開発データ合わせたを 4387 件のトピックを $(x, y) = (\text{RWC}(1\text{day}), \text{RWC}(10\text{days}))$ でプロットしたもの. $\text{RWC}(10\text{days}) - \text{RWC}(1\text{day})$ の大きさに 3 クラスに分けている.

4.2 分類器

本研究では前述した将来の論争化の予測についての3値分類問題を解く教師あり分類器として Support Vector Machine (SVM) [31] を採用する．以下で SVM について説明する．

以下のように分類器の学習のための事例となる素性ベクトル \mathbf{x} と，その事例が正例か負例かを表す2値の正解ラベル t を用意する． t は +1 のとき正例，-1 のとき負例とする．式 (3.1) は，分類の際に用いる判別関数で，正例と負例とを分ける識別平面となる． \mathbf{x} が正例のときに正，負例のときに負となる．

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (4.1)$$

一般に，分類問題では取りうる識別平面が無数に存在することとなるが，SVM は識別平面から最も距離が近い事例をサポートベクトルとし，サポートベクトルまでの距離（マージン）を最大化するようなパラメータ \mathbf{w}, b を決定する．これはマージン最大化と呼ばれており，具体的にはラグランジュ乗数 $a_n (a_n \geq 0)$ を導入してラグランジュ関数式 (4.2) を最大化する \mathbf{a} を求め，求めた \mathbf{a} を用いて式 (4.3) と式 (4.4) によりパラメータ \mathbf{w}, b を決定する．これはラグランジュの未定乗数法と呼ばれている．

$$L(\mathbf{a}) = \sum_{n=1}^N a_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N a_n a_m t_n t_m (\phi(\mathbf{x}_n) \cdot \phi(\mathbf{x}_m)) \quad (4.2)$$

$$\begin{aligned} \text{s.t. } \sum_{n=1}^N a_n t_n &= 0 \\ a_n &\geq 0 \end{aligned}$$

$$\mathbf{w} = \sum_{n=1}^N a_n t_n \phi(\mathbf{x}_n) \quad (4.3)$$

$$b = \frac{1}{|X_S|} \sum_{\mathbf{x}_n \in X_S} \left(t_n - \sum_{\mathbf{x}_m \in X_S} (\mathbf{x}_n, \mathbf{x}_m) \right) \quad (4.4)$$

$\phi(\cdot)$ は \mathbf{x} を高次元空間へと写像するカーネル関数である． N は学習に用いた事例の数を表し， X_S はサポートベクトルの集合を表している．予測の際は学習したパラメータ \mathbf{w} , b とテスト事例 \mathbf{x}' を式 (4.1) へと代入し，値の正負を調べることによりテスト事例を分類する．SVM は 2 値分類器であるが，多値分類の際は one-versus-rest 法か one-versus-one 法を使う．複数のクラス C_1, \dots, C_n に対して，one-versus-rest 法はそれぞれの C_i について C_i か C_i 以外かのどちらであるかを判別する分類器を作成する．one-versus-one 法はすべての C_i と C_j の組み合わせについてどちらであるかを判別する分類器を作成する．

4.3 グラフ特徴量

本研究で予測モデルの学習に使う特徴量を Table 4.1 に示す．所与のトピックのインタラクショングラフ構造を用いた特徴量としては，まずノード数やエッジ数，クラスタリング係数や平均次数などの単純なグラフ統計量が挙げられる．他にも，単純なグラフ統計量のほか，議論に参加しているユーザたちのもともとの関係性をグラフ構造から獲得することを考え，議論に参加しているユーザのもともとの RT 関係からグラフを構築し，そのグラフの統計量も用いた．これは，例えばもとから中の悪い二人のユーザがある話題でも別々のグループに入り，対立を起こすというような状況のように，もともとの人間関係が論争を引き起こしているような可能性を考慮したためである．具体的なユーザの関係性グラフは，次のように得た．

- 議論のインタラクショングラフを得る
- インタラクショングラフ中の各ノード（Twitter ユーザ）のユーザ ID のリストを獲得する
- トピックの登場した日から過去 30 日間におこなわれた全部のエッジ（リツイート・メンション・クオート）のうち，エッジの端の少なくとも一方のユーザの ID が，先に獲得した ID リストに存在するものだけを収集する

ある時期にあるトピックのインタラクショングラフを得たあと、そのグラフの全ノード（Twitter ユーザ）の RT やメンションを過去 1 ヶ月間に遡って取得してまた別のインタラクショングラフを作りその別のグラフから、ノードだけを元のインタラクショングラフと一致するように部分グラフを導出することで、あるトピックの議論に参加しているユーザたちのもともとの人間関係を表現したグラフを取得できると考えた。このように導出した部分グラフを以下、リレーショングラフと呼ぶ。

4.4 テキスト特徴量

あるトピックに言及したツイートの集合があるとき、それらのツイート中に特定の語句が出現すると、そのトピックはその後論争になる可能性が高くなる（例: 「安倍晋三」、「憲法改正」、「移民受け入れ」）。こういった語句の特徴を捉えるために、対象のトピックに関するツイート集合の単語ベクトルを平均し、分類器の特徴量として用いる。

単語ベクトルとしては、別のコーパスを用いて単語の分散表現である Skip-gram [32] を事前に訓練し得た訓練済みベクトルを用いる。Skip-gram は文中の単語が与えられた時、対象となる単語の周辺単語を予測する言語モデルである。式を挙げて具体的に説明する。文中のある単語 w_I が与えられた時、その次の単語 w_O が出現する確率は以下の式で求められる。

$$p(w_O|w_I) = \frac{\exp(\mathbf{u}_{w_O}^T \cdot \mathbf{v}_{w_I})}{\sum_{w_v \in V} \exp(\mathbf{u}_{w_v}^T \cdot \mathbf{v}_{w_I})} \quad (4.5)$$

V は語彙数で、 $\mathbf{u}_w, \mathbf{v}_w$ が学習される単語ベクトルである。Skip-gram はこの確率式を拡張し、前後窓幅 C 個分の単語の出現を以下の同時確率の式で予測する。

$$p(w_{O,1}, w_{O,2}, \dots, w_{O,C}|w_I) = \prod_{c=1}^C \frac{\exp(\mathbf{u}_{w_{O,c}}^T \cdot \mathbf{v}_{w_I})}{\sum_{w_v \in V} \exp(\mathbf{u}_{w_v}^T \cdot \mathbf{v}_{w_I})} \quad (4.6)$$

学習の際は、以下の式で同時確率を最大化するよう確率的勾配降下法によりパラメータを調整する。

$$\operatorname{argmax}_{\mathbf{W}, \mathbf{W}'} p(w_{O,1}, w_{O,2}, \dots, w_{O,C}|w_I) \quad (4.7)$$

W, W' はそれぞれ u_w, v_w を列ベクトルとして並べた行列である．最終的には，学習後に得られた v_{w_I} を単語ベクトルとして用いる．

ツイートからの語の抽出方法としては，対象のツイートをすべて形態素解析し，名詞・動詞・形容詞・助詞と判断された単語のみを抽出する．ただし，RT によって同一の語句が複数回出現することが考えられるため，RT や引用されたツイートは複数回出現していても 1 回のみの出現として扱う．また，論争予測の分類器の訓練の際は，データセット中で出現頻度が 5 回以上の単語のみを対象とする．

4.5 ユーザ特徴量

ユーザの性質を捉えることは議論が論争になるかどうかの予測に役に立つ．例えば，論争そのものが好きなユーザもいるかもしれないし，論争を引き起こしやすい政治関係のトピックを好むユーザもいる．そこで，あるトピックに参加しているユーザがこれまでに参加してきた議論に論争になったものが多いかどうかを考慮する特徴を分類器に加える．具体的には，あるユーザが参加している議論をデータセット中で全て列挙し，それらの議論の RWC の和や平均をそのユーザの論争参加率とみなし特徴量として使用する．値としては，それぞれ 1 次元のスカラーとなる．

Table 4.1: 使用した特徴量. 特に記載のない場合はすべてスカラ値.

	使用特徴量
グラフ特徴量 (過去 30 日のインタラクショングラフに についてもこれらの特徴量を求める)	ノード数 エッジ数 クラスタリング係数 平均次数 RWC
テキスト特徴量	単語ベクトルの平均 (200 次元)
ツイート特徴量	議論参加者の過去に参加した 議論の RWC の総和・平均

第5章 評価実験

本章では提案した特徴量の有効性を，実際に提案した特徴量を用いた場合の予測精度から定量的に評価する．また，予測の成功した例・失敗した例を取り上げて分析することで，提案した特徴量の捉えられている点・捉えられていない点についての定性的な評価も行う．

5.1 使用データ

本実験で利用するツイートデータセットは，筆者の所属研究室で収集しているツイートデータのうち，2017年10月から2018年10月までに投稿されたものを利用した．これらのツイートデータセットを利用して3.1節で述べたものと同様の収集方法で出現当日そのキーワードを含むツイートが50件以上存在したものをトピックキーワードとして選択し，4753件のトピックを収集した．そして収集したこれらのトピックを学習データ，開発データ，評価データの順で時系列順になるよう，連続的な期間毎にデータを分割した．この理由としては，仮に時系列をシャッフルしたデータによって学習された分類器を利用した場合には未来のツイートデータを利用した予測となってしまうことになるため，これを避けるためである．具体的な分割方法としては，以下のように選んだ．

学習データ 2017年10月～2018年7月に出現したトピック: 3399件

開発データ 2017年8月～2018年9月に出現したトピック: 948件

評価データ 2018 年 10 月のトピック: 406 件

5.2 評価データでの予測精度と分析

Table 5.1 にテストデータでの予測精度を示す。ベースラインとしては, UP, DOWN, REMAIN の全クラスを一様な確率で予測する識別器を使用した。結果として, DOWN クラスや REMAIN クラスでの予測精度は他の特徴量のほうが大きくなったが, クラスサンプル比率を無視するマクロ平均においては提案した特徴量を用いたモデルが最良の結果を出した。これにより, 提案した特徴量の有用性が示された。

5.3 特徴量分析

本研究で提案した各特徴量が将来の論争度の予測にどの程度有効かを検証するために 2 種類の分析を行った。本節ではその結果を報告する。

5.3.1 ablation テスト

各特徴量の寄与率を測るため, ablation テストを行った。具体的には, 各特徴量を除いたモデルを学習し, その予測精度の増減を調べた。特徴量除去によって精度が減少した特徴量は予測に有効な特徴量であると言える。5.2 章にてテキスト特徴量のみの予測精度, ユーザグラフ特徴量のみの予測精度は計算済みであり, またテキスト特徴量として今は 200 次元の埋め込みベクトルのみを使用しているため, ここではユーザグラフ特徴量の各項目について, その寄与率を計算する。結果を Table 5.2 に示す。

5.3.2 単一種の特徴量テスト

Ablation テストはある特徴量を除いた際の予測精度の増減を検証するため, 特徴量単体が予測にどの程度貢献するかは明らかではない。そこで, 各特徴量のみを用

Table 5.1: test データでの識別性能

model	baseline(uniform)	w2V only	Graph only	w2V+Graph
UP	0.165	0.199	0.308	0.316
DOWN	0.258	0.268	0.512	0.508
REMAIN	0.480	0.521	0.385	0.444
macro	0.301	0.329	0.402	0.423
micro	0.337	0.384	0.404	0.433

Table 5.2: 開発データでの ablation test の結果の各クラスおよび全体の f1 値. 分類器は linearSVC($tol = 1.0e - 5$, $C = 5.0$) を使用. 「ablation なし」と比較して「グラフ特徴量 (初日)」を使用しなかった場合に予測精度が最も下がった (太字).

ablation	UP	DOWN	REMAIN	macro	micro
ablation なし	0.309	0.459	0.459	0.409	0.423
- グラフ特徴量 (初日)	0.282	0.444	0.321	0.349	0.348
- グラフ特徴量 (初日)	0.316	0.444	0.395	0.385	0.389
- RWC (初日)	0.261	0.346	0.462	0.356	0.384
- RWC (過去)	0.297	0.455	0.413	0.388	0.395
- 過去議論 RWC	0.307	0.450	0.422	0.393	0.402

Table 5.3: 開発データでのグラフ特徴量を 1 種類だけ使用して学習を行った結果
($\text{linearSVC}(\text{tol} = 1.0e - 5, C = 5.0)$.)

使用特徴量	UP	DOWN	REMAIN	macro	micro
すべて	0.309	0.459	0.459	0.409	0.423
only グラフ特徴量 (初日)	0.249	0.333	0.426	0.336	0.356
only グラフ特徴量 (初日)	0.228	0.305	0.234	0.255	0.251
only RWC (初日)	0.248	0.442	0.078	0.256	0.248
only RWC (過去)	0.200	0.291	0.180	0.224	0.225
only 過去議論 RWC	0.207	0.308	0.133	0.216	0.209

いた際の予測性能についても検証を行った．Fig. 5.3 に単一種類の特徴量のみを使用した場合の識別性能の表を示す．Ablation テストの結果と同様に，初日のグラフ構造がもっとも寄与していることが分かった．

5.4 事例分析

本節では，提案した手法によって，クラスを正確に予測できたトピックの分析を行うことで，提案手法の能力と限界点について考察する．

5.4.1 成功事例

クラス予測に成功したトピックについて，実際にどのような議論が Twitter で行われていたかを分析した．そのようなトピックのうち，実際に論争が発生している，あるいは論争が収束していることが確認された事例を Table 5.4 に示す．

論争の解消傾向が実際に確認されたトピックの一例として，「憲発議」の出現当日および 10 日後における RWC の値とインタラクショングラフの可視化結果を Fig. 5.1 に示す．本トピックに関する Twitter での議論を分析したところ，トピックが出現した当日には，安倍議員の支持層の議論と石破議員の支持層との議論に極化が生じていることが観測された．実際に，インタラクショングラフにおける両支持層の間にはエッジが少なく，RWC の値も 0.877 と高いことが同図より分かる．本トピックの出現から 10 日が経過した後では，RWC の値は 0.407 へと減少した．同図右のインタラクショングラフからも分かるように，この時点ではどちらの派閥のツイートも RT する人が増えており，極化が多少解消されていることが実際に確認できた．提案手法は，このようなトピックに対して，正しいクラス DOWN を予測することができた．

また，Fig. 5.2 には論争に発展すると予測し実際に論争にまで繋がった事例「新潮 45 編集部」の出現当日および 10 日後におけるインタラクショングラフを示す．このトピックに関する Twitter での議論を分析したところ，トピックが出現した当日

Table 5.4: 予測の成功したトピックの実例

自民総裁選	ネトウヨ体験
人種差別撤廃委員会	衆議院通過
新潮 45 休刊	福島市議聴取
獣医学部構想	築地市場先行破壊
オスプレイ暫定配備	拉致被害者全員救出
副次的文化系合唱祭	原田真人監督
東日本大震災復興	日活再始動
講演会中止	歌番組出演
人種差別撤廃委員会	衆議院通過
主演男優賞受賞	ジェジュン特集
しづか監督	強化拡張パック
麻生大臣支持	築地女将さん会
天体観測展プレゼント	東京カメラ部共催
新潮 45 休刊	福島市議聴取
ドクターヘリ緊急救命	脳内解放
サンダース大学付属	自衛隊ヘリ墜落
スプレイ暫定配備	SHOCK 公演 1600 回
東地区優勝	総合視聴率
ペルソナ 4 発売	エイル降臨
ドールえっち	沖波進水日
レトロコンシューマー愛好会	依田啓示
一休そうじゅん	散華行ブルース
オーマイジャンプ	復幸男
企画展参加	城崎広告

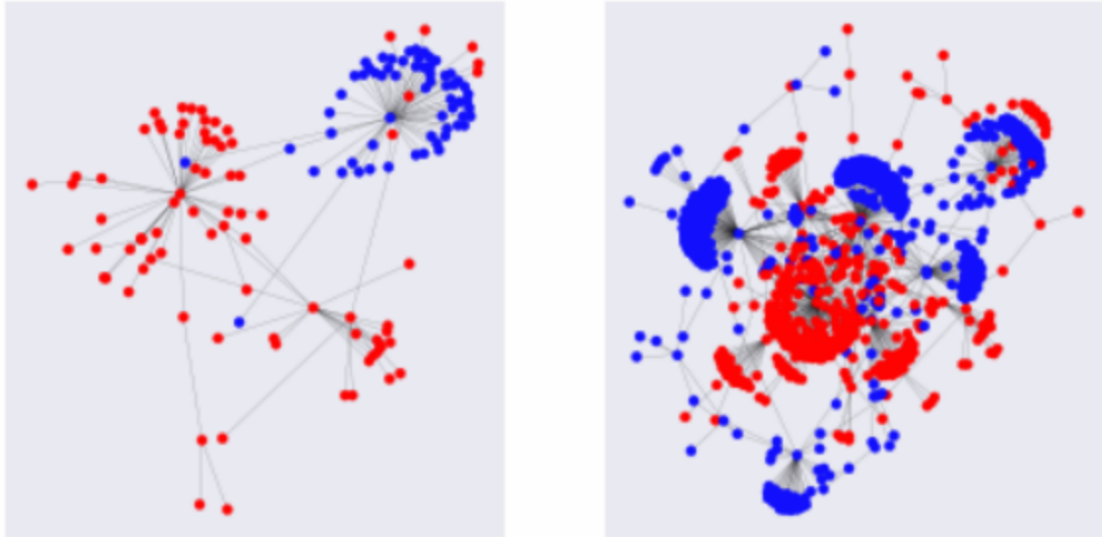


Fig. 5.1: 成功トピック「憲発議」の出現当日（左）と10日後（右）のインタラクショングラフ．RWC値は， $(RWC(1), RWC(10)) = (0.877, 0.407)$ ．

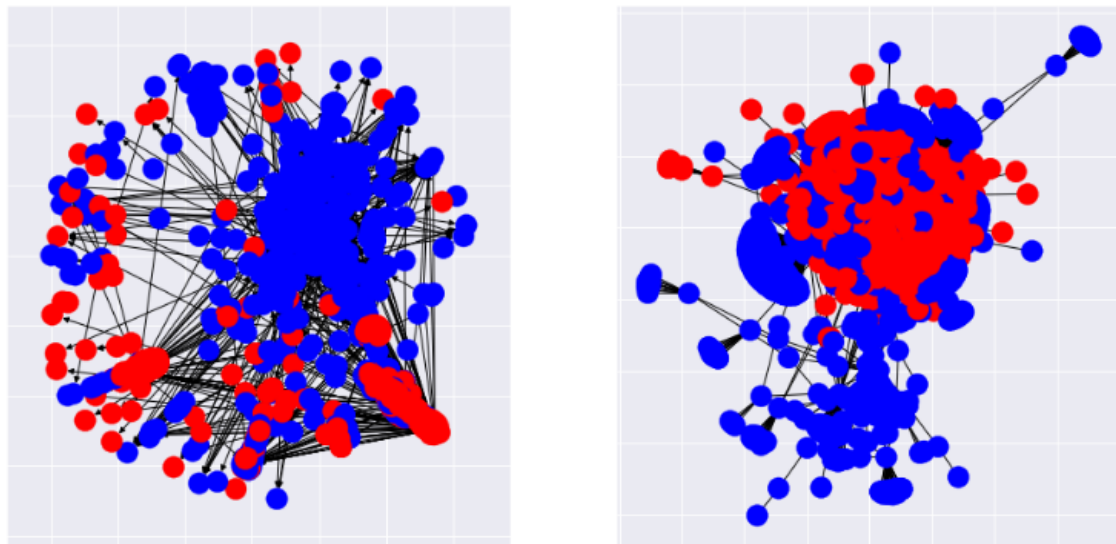


Fig. 5.2: 成功トピック「新潮45編集部」の出現当日（左）と10日後（右）のインタラクショングラフ．RWC値は， $(RWC(1), RWC(10)) = (0.080, 0.382)$ ．

には、新潮 45 という週刊誌への批判意見がほとんどであった一方、徐々に賛成意見が増えていた。これは、新潮社の公式アカウントが新潮 45 に対する批判ツイートを連日リツイートしたことに対する賞賛の意見が増えたことに起因している。

5.4.2 失敗事例

提案手法によって正しいクラスを予測できたものの、Twitter での実際の議論を確認してみると、UP クラスであるにも関わらず論争が発生していない事例や、DOWN クラスであるにも関わらず論争が収束していない事例が存在した。その事例を Table 5.5 に示す。

提案手法によって正しいクラス DOWN が予測されたものの、実際には論争の収束が確認されなかったトピック「コラボ装備」について、出現当日および 10 日後における RWC の値とインタラクショングラフの可視化結果を Fig. 5.3 に示す。同図のインタラクショングラフを見ると、確かにトピック出現から 10 日までの間で論争が激化しているように見える。しかし、実際の議論の内容を分析してみたところ、複数のソーシャルゲームがコラボ装備と呼ばれる機能¹を同時期にリリースしており、それぞれのゲームによってプレイヤー層が異なるため、インタラクションに分離が生じていただけであり、論争は生じていなかった。この結果は、事前に選定したトピックの中に複数の異なるトピックが混在していると、予測の結果に悪影響を及ぼすことを示唆している。将来の論争度のより正確な予測のための課題として、均一な粒度のトピックを同定することが考えられる。また、本研究では平均次数に基づきツリー型のグラフを事前に省いていたが、本分析の結果から、信頼性の高い予測の実現のためには、その閾値を適切に選択する必要があると考えられる。

¹あるゲーム内において別のゲームのキャラクターの装備が利用可能になること。

Table 5.5: 予測の失敗したトピックの実例

歌番組出演	強化拡張パック
ドクターヘリ緊急救命	サンダース大学付属
SHOCK 公演 1600 回	散華行ブルース
城崎広告	渡辺麻友卒業
ドール登録日	ベストジーニスト投票
山 P 入所	スクフェス感謝祭
づエレ合同	スクフェス感謝
国観客動員数 200 万人突破	小梅ちゃん合同
濃縮鶏煮干つけめん	白坂小梅合同誌
石かりの日	ルルアタック
アムロック	轟雷起動
SIR メジャーリリース	Enjoy 更新
重要美術品指定日	女子高生収容
女子高生収容所	大野智ソロアルバム熱望
女性専用車両反対街宣	咲実写イベント
日本再始動	ことり卒業
聖刻レギオン	プロ強行採決反対
潤くん入所	轟雷起動日
大野智ソロアルバム熱望	3 分チャレンジ

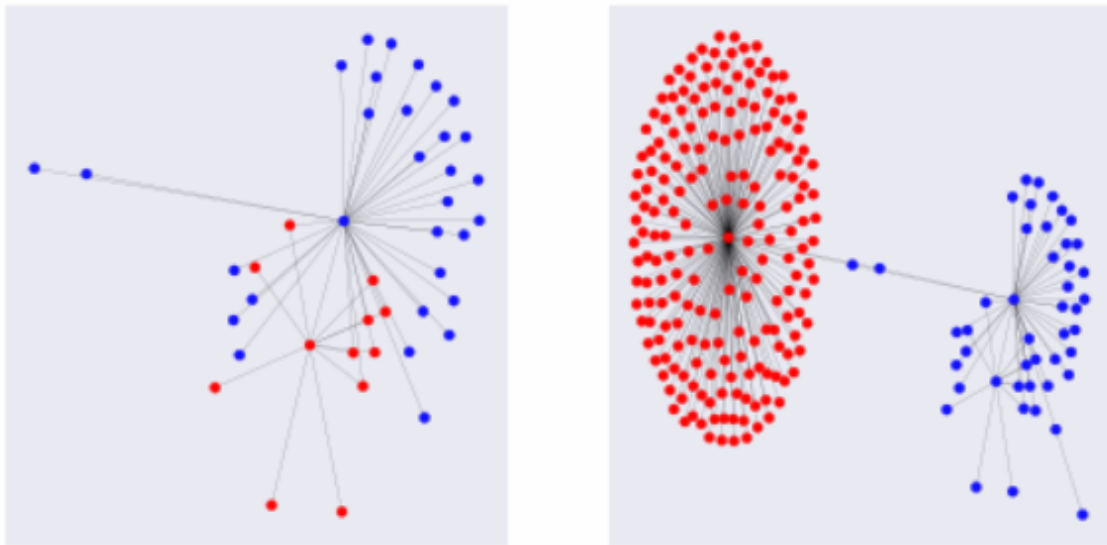


Fig. 5.3: 成功トピック「コラボ装備」の出現当日（左）と10日後（右）のインタラクショングラフ．RWC 値は， $(\text{RWC}(1), \text{RWC}(10)) = (0.099, 0.814)$ ．

第6章 まとめ

本論文では、人々が同じ意見を持つ人同士の間でのみ交流する事態を避けるため、あるトピックの議論が論争になるかを予測するタスクを設定し、また論争度の変化を予測するにあたってデータセットの整備を行った。具体的には、議論の対象を示す単語の集合として抽出したトピックが将来的に論争へと発展するかどうか予測する問題を、そのトピックの（既存手法により計算される）論争度が将来的に元の論争度の定数倍を超えるかどうかの識別問題として定式化した。その上でこの問題を議論の初期の段階における、ツイートの内容、ユーザ間のインタラクショングラフの特徴量、およびそのグラフを分割したときの分離度等の手がかりを用いた分類器で解くことを試みる。本論文では、既存のスタンス検知研究で得られた賛否表明ボタンを用いてユーザによって賛否の分かれるトピックを収集することで論争トピックのデータセットを構築し、また指標が正常にふるまわないケースを解析してそのようなケースを除去することで、予測用のデータセットの整備を行った。

今後の課題としては、論争になると予測されたトピックに対する予防手法の提案が挙げられる。この手法として、簡単には対立が発生している2つのグループが存在するとき、対立しているグループの情報を通知することで相互理解を図り論争を予防する、という手法が考えられる。しかし Bail らが報告したように単純に対立している側の情報を通知するだけではむしろ対立が深まる可能性がある [8]。また、池上らが紹介する Sidanius らの研究によると、異なる思想の持ち主同士が対立する理由の根底には異なる思想を持つ人々へのステレオタイプが理由であるという [33, 34]。すなわち、異なる思想を持つ人々すべてを深く理解するには人間の情報処理能力は不足しているため、身近な関わりのある人以外を類型化することで、情報を取捨選

択しようとする機能が人間に備わっており，その結果ステレオタイプが醸成され，醸成されたステレオタイプが極化を招くということである．

これに対して2つの観点からの対策が考えられる．ひとつは，人は身近な人物であればステレオタイプを抱きにくいという点に着目し，特定のトピックにおいては対立する意見を持つものの，それ以外のトピックにおいてはむしろ同じ意見を持つ人物同士をリコメンドなどにより結びつけることが考えられる．これは，同じ池上らが紹介している，異なる意見を持つグループでもグループを横断して積極的に交流を行う個体があれば，その2つのグループで徐々に融和が見られる，という集団間友情現象を狙ったものである．

もうひとつの方向性として，ステレオタイプに基づいて判断する機能が情報の取捨選択が必要であるため生じた点に着目し，その取捨選択を代行することが考えられる．現代においてステレオタイプを正すような情報は容易に入手可能であるが，そもそも世の中は多くの情報が溢れており，その中から人が持つステレオタイプを正すような情報の獲得が容易ではないために，論争が発生しているという考察に基づく．この方向性に沿った研究として，例えば Umemoto らは人が知りたいトピックに対する多角的な情報を容易に獲得できる ScentBar という検索インタフェースを提案し，ユーザが知りたいと思ったトピックに対する容易に多面的な情報を獲得できたことを報告している [35]．この手法をそのまましても先述したような異なる意見グループに対するステレオタイプから，忌避感を示されるかもしれないが，この2つの方向性を融合させることで，論争の無い世界へ一歩近づけるかもしれない．

謝辞

この研究を行うにあたって、豊田正史教授、吉永直樹准教授、小宮山純平助教、そして梅本和俊特任助教には多くの助言をしていただきました。豊田正史教授には研究の進め方や実験プログラムの基礎、またそもそも研究者としてのあり方など、今後の人生においての基盤となる技術や哲学を教わりました。吉永直樹准教授にはプログラムの実装の際の心構えや研究のストーリーを決める際の定石などを教わりました。お二人共甘い研究計画を投げかけたらバツサリ切られるので、自然と研究計画の練り方が身についたと思います。小宮山純平助教には機械学習の基礎的な質問をすることが多く、そのたびに初歩的な質問にもかかわらず丁寧に教えてもらい、申し訳ないくらいです。また研究がうまく行かない不安を相談したら学生に近い視点から答えてくださることが多く、落ち込んでいるところに復帰の手助けになることが多かったです。梅本和俊特任助教には学振や中間発表の資料など、なんども研究資料を見ていただきました。先の二人と同じく、スキのある研究計画を見せるとすべてのスキを指摘してくださるので、指摘された点を治していくたびに計画書の質が上がっていくことを感じました。ありがとうございます。そして、このような素晴らしい指導環境を作ってくださった喜連川優教授に感謝いたします。ラボの先輩である佐藤文一さん・石渡祥之佑さん・金洪善さん・佐藤翔悦さん・赤崎智さん・澤田頌子さん・陳鍵さん・大原康平さんらには、辛いときに研究の面でもそれ以外の面でも助言や相談をして頂き、実りのある研究生生活を行うことが出来ました。ラボの後輩である三條嵩明君・別所祐太郎君・福田展和君・大葉大輔君・土屋潤一郎君・蔦侑磨君・杉山普君・左天池君とは僕の研究についてまっさらな視点からの意見をもらえて、研究の意義などを見つめなおすいい機会をもらえました。また日夜

遅くまで研究に励む後輩の姿を見て、身が引き締まる思いでした。そして同期である保田和彦君・遠田哲史君・根石将人君・佐久間仁君・清水洸希君・羅博明君に感謝します。みんなでご飯食べにいたりダラダラ話したりした時間が最高の財産です。2年間みんなとともに走ってこれたことが誇らしいと思います。保田君，遠田君，羅君のこれからの活躍を祈っています。社会に出ても頑張ってください。そして根石君，佐久間君，清水君はこれからも研究やったりダラダラしたりしましょう。他にも，研究室の事務作業などでお世話になった秘書さんにも感謝したいです。最後に，遠くから支えてくれた家族に感謝します。

平成 31 年 1 月 31 日

参考文献

- [1] W Glynn Mangold and David J Faulds. Social media: The new hybrid element of the promotion mix. *Business horizons*, Vol. 52, No. 4, pp. 357–365, 2009.
- [2] Jan H Kietzmann, Kristopher Hermkens, Ian P McCarthy, and Bruno S Silvestre. Social media? get serious! understanding the functional building blocks of social media. *Business horizons*, Vol. 54, No. 3, pp. 241–251, 2011.
- [3] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pp. 591–600. AcM, 2010.
- [4] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 56–65. ACM, 2007.
- [5] Kathleen Hall Jamieson and Joseph N Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.
- [6] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.
- [7] Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

-
- [8] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, Vol. 115, No. 37, pp. 9216–9221, 2018.
- [9] Kota Kakiuchi, Fujio Toriumi, Masanori Takano, Kazuya Wada, and Ichiro Fukuda. Influence of selective exposure to viewing contents diversity. *CoRR*, Vol. abs/1807.08744, , 2018.
- [10] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David AL Levy, and Rasmus Kleis Nielsen. Reuters institute digital news report 2017. 2017.
- [11] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 211–36, 2017.
- [12] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, Vol. 359, No. 6380, pp. 1146–1151, 2018.
- [13] Akira Sasaki, Kazuaki Hanawa, Naoaki Okazaki, and Kentaro Inui. Other topics you may also agree or disagree: Modeling inter-topic preferences using tweets and matrix factorization. In *ACL*, 2017.
- [14] Akira Sasaki, Kazuaki Hanawa, Naoaki Okazaki, and Kentaro Inui. Predicting stances from social media posts using factorization machines. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3381–3390, 2018.
- [15] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. *Trans. Soc. Comput.*, Vol. 1, No. 1, pp. 3:1–3:27, January 2018.

-
- [16] Qiankun Zhao, Prasenjit Mitra, and Bi Chen. Temporal and information flow based event detection from social text streams. In *AAAI*, Vol. 7, pp. 1501–1506, 2007.
 - [17] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pp. 851–860. ACM, 2010.
 - [18] Jianshu Weng and Bu-Sung Lee. Event detection in twitter. *ICWSM*, Vol. 11, pp. 401–408, 2011.
 - [19] Charu C Aggarwal and Karthik Subbian. Event detection in social streams. In *Proceedings of the 2012 SIAM international conference on data mining*, pp. 624–635. SIAM, 2012.
 - [20] Xiangmin Zhou and Lei Chen. Event detection over twitter social media streams. *The VLDB Journal—The International Journal on Very Large Data Bases*, Vol. 23, No. 3, pp. 381–400, 2014.
 - [21] Ana-Maria Popescu and Marco Pennacchiotti. Detecting controversial events from twitter. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 1873–1876. ACM, 2010.
 - [22] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Reducing controversy by connecting opposing views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pp. 81–90, New York, NY, USA, 2017. ACM.
 - [23] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. Me, my echo chamber, and i: Introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pp. 823–831,

- Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.
- [24] Nguyen Vo and Kyumin Lee. The rise of guardians: Fact-checking url recommendation to combat fake news. In *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '18, pp. 275–284, New York, NY, USA, 2018. ACM.
- [25] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43. ACM, 2005.
- [26] Huiji Gao, Jalal Mahmud, Jilin Chen, Jeffrey Nichols, and Michelle X Zhou. Modeling user attitude toward controversial topics in online social media. In *ICWSM*, 2014.
- [27] Mitsuo Yoshida and Fujio Toriumi. Information diffusion power of political party twitter accounts during japan’s 2017 election. In Steffen Staab, Olessia Koltsova, and Dmitry I. Ignatov, editors, *Social Informatics*, pp. 334–342, Cham, 2018. Springer International Publishing.
- [28] Benjamin Timmermans, Tobias Kuhn, Kaspar Beelen, and Lora Aroyo. Computational controversy. In Giovanni Luca Ciampaglia, Afra Mashhadi, and Taha Yasseri, editors, *Social Informatics*, pp. 288–300, Cham, 2017. Springer International Publishing.
- [29] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pp. 1217–1230, New York, NY, USA, 2017. ACM.

-
- [30] Johannes Daxenberger, Steffen Eger, Ivan Habernal, Christian Stab, and Iryna Gurevych. What is the essence of a claim? cross-domain claim identification. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2055–2066. Association for Computational Linguistics, 2017.
- [31] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, Vol. 20, No. 3, pp. 273–297, 1995.
- [32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.
- [33] 池上知子. 差別・偏見研究の変遷と新たな展開. 教育心理学年報, Vol. 53, pp. 133–146, 2014.
- [34] Jim Sidanius and F Pratto. Social dominance: An intergroup theory of social oppression and hierarchy. *Cambridge: England. Cambridge University Press. Smith, SJ, Axelton, AM, & Saucier, DA (2009). The effects of contact on sexual prejudice: A meta-analysis. Sex Roles*, Vol. 61, No. 3-4, pp. 178–191, 1999.
- [35] Kazutoshi Umemoto, Takehiro Yamamoto, and Katsumi Tanaka. Scentbar: A query suggestion interface visualizing the amount of missed relevant information for intrinsically diverse search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 405–414. ACM, 2016.

発表文献

査読なし国内会議

1. 張翔, 豊田正史, 吉永直樹, Twitter における論争の分析及びその極化の予測
第11回データ工学と情報マネジメントに関するフォーラム (DEIM 2019), 福岡, 2019.