

東京大学
情報理工学系研究科 電子情報学専攻
修士論文

返信態度の分類に基づくツイート間論述関係の分析
Analyzing Statement Relations between Tweets
from Users' Communications

大和田 裕亮
Yusuke Owada

指導教員 石塚 満 教授

2013年2月

概要

東日本大震災では安否確認や被災者支援のために Twitter が活躍したが、一方で多種多様な情報が流通し、混乱を招いた。我々は、情報の信憑性や重要性を評価するには、ツイート空間の論述的な構造を解析・可視化し、情報の「裏」を取ることが大切だと考えている。本稿では、ツイートの返信および非公式 RT（以下、返信・引用と略す）に着目し、ツイート間の論述的な関係を認識する手法を提案する。具体的には、返信・引用ツイートによって、投稿者の「同意」「反論」「疑問」などの態度が表明されると考え、これらの態度を推定する分類器を教師有り学習で構築する。評価実験では、返信・引用ツイートで表明される態度の推定性能を報告する。さらに、本手法が直接的に返信・引用関係のないツイート間の論述的な関係の推定にも応用できることを示し、ツイート間の含意関係認識に基づくアプローチとの比較を行う。

Abstract

Although Twitter played an important role in supporting victims in the 2011 Tohoku earthquake and tsunami disaster, we encountered a number of situations where the vast flow of unauthorized information caused troubles. In order to assess the credibility and importance of a piece of information, we find it important to analyze the statement structure on Twitter and to understand the background of the information. In this paper, we propose a method for analyzing the statement relation between a tweet and its reply or quoting tweet. More specifically, we assume that a reply or quoting tweet expresses a statement relation (e.g., *agreement*, *rebuttal*, *question*, *other*) towards the target tweet, and build a classifier for predicting a statement relation for a given pair of tweets. The experimental results report the performance of the classifier for predicting statement relations. In addition, we demonstrate that the proposed method can be applied to analyze statement relations between tweets that have no direct reply/quoting link, and compare the proposed approach with the previous method based on the technique of textual entailment.

目次

第 1 章	序論	1
1.1	本研究の背景	1
1.2	本研究の目的	2
1.3	本研究の貢献	3
1.4	本論文の構成	3
第 2 章	信憑性判断支援のためのツイート間関係認識	5
2.1	はじめに	5
2.2	信憑性	5
2.2.1	信憑性の定義	5
2.2.2	信憑性の構成要素	6
2.2.3	人間の信憑性判断モデル	6
2.2.4	本論文における信憑性の扱い	7
2.3	Twitter 情報の信憑性判断を支援するアプリケーション	8
2.3.1	トピックに関する無秩序なまとめ	8
2.3.2	トピックに関する複数ツイート要約	8
2.3.3	命題に対する言明の分類	9
2.3.4	ツイートに対する他ツイートの関係分類	11
2.3.5	各アプリケーションのまとめ	11
2.4	まとめ	11
第 3 章	関連研究	12
3.1	Web 情報の信憑性分析	12
3.2	情報の整理と文間関係認識	13
3.3	評判分析	14
第 4 章	返信態度の分類	15
4.1	はじめに	15
4.2	各投稿形式の整理と本論文中における表記	15
4.3	返信ツイートの態度分類手法	16

4.3.1	ツイートの意図	16
4.3.2	問題設定	17
4.3.3	返信ツイートのアノテーション	19
4.3.4	分類手法	20
4.4	返信ツイートの態度分類の評価	23
4.4.1	交差検定による性能評価	23
4.4.2	各素性の有効性	25
4.4.3	考察	26
4.5	デマ検出への適用	28
4.6	まとめ	30
第 5 章	一般のツイート間関係認識	31
5.1	はじめに	31
5.2	問題設定	32
5.3	同トピックツイートペアのアノテーション	32
5.4	提案手法	33
5.5	文間関係認識に基づく分類手法	36
5.6	一般のツイート間関係分類の評価	39
5.6.1	各トピックにおける精度評価	40
5.6.2	拡張パターンの有効性	42
5.6.3	考察	42
5.6.4	論述構造の可視化	43
5.7	まとめ	45
第 6 章	結論	46
6.1	本研究のまとめ	46
6.2	今後の課題	46
	発表文献と研究活動	48
	参考文献	49
付録 A	反論表現の一覧	54

表目次

4.1	選出した 20 個のクエリ	20
4.2	データセット A におけるアノテーションの一致度	21
4.3	反論表現の例	21
4.4	返信ツイートの分類結果 (データセット B)	24
4.5	参考: 二人一致の返信ツイートの分類結果	24
4.6	各クラスで特定性の高い素性	25
5.1	一般ツイートペア分類実験の 11 トピック	33
5.2	トピック毎のアノテーション結果	34
5.3	否定表現	38
5.4	提案手法による分類	40
5.5	文間関係認識手法 1 による分類	41
5.6	文間関係認識手法 2 による分類	41
5.7	拡張パターンによる認識の精度	42
A.1	反論表現辞書	54

目次

1.1	返信・非公式リツイート，もしくは内容に基づくツイート間の論述関係 . . .	2
2.1	Prominence-Interpretation 理論 [1]	7
2.2	言論マップの分析例	10
4.1	非公式リツイートのリンク情報の復元	19
4.2	使用する特徴を変更した時の性能変化	26
4.3	反論の種類毎の再現率	27
4.4	反論・疑問の数と割合	30
5.1	直接返信のパターン	35
5.2	両方をリツイートするパターン	35
5.3	他のツイートで返信のパターン	36
5.4	片方をリツイート・もう片方に返信のパターン	36
5.5	両方を返信のパターン	36
5.6	文間関係認識例	37
5.7	文間関係認識手法 1	39
5.8	「イソジン飲むと放射能対策になる」の論述構造グラフ	44
5.9	「イソジン飲むと放射能対策になる」に関連するツイート群	44

第 1 章

序論

1.1 本研究の背景

近年，Twitter や Facebook などのソーシャルメディアが社会において大きな存在感を示している．特に，Twitter は情報発信の手軽さやリアルタイム性が魅力であり，有名人のニュース，スポーツなどの国際試合の勝利，災害の発生などの速報，アメリカ大統領選挙に代表される選挙活動，アラブの春（2010 年，2011 年）やイギリスの暴動（2011 年）など，社会に大きな影響を与えるメディアになっている．2011 年 3 月に発生した東日本大震災においても，安否確認や被災者支援のために，ソーシャルメディアが活躍した．

Twitter 上ではリアルタイムな情報交換が行われているが，誤った情報や噂も故意に，あるいは故意ではなくとも広まってしまうことがある．東日本大震災での有名な例としては，「コスモ石油の火災に伴い有害物質の雨が降る」や「地震で孤立している宮城県花山村に救助が来ず，赤ちゃんや老人が餓死している」などの誤情報の拡散が挙げられる．このような誤情報の拡散は無用な混乱を招くだけでなく，健康被害や風評被害などの 2 次的な損害をもたらす．1923 年に発生した関東大震災の時も，根拠のない風説や流言が広まったと言われているが，科学技術がこれほど進歩した 2011 年でも，流言を防げなかった．

このような反省から，Twitter 上の情報の真偽性や信憑性を判断する技術に注目が集まっている．しかしながら，情報の真偽性をコンピュータが自動的に判断するのは，技術面および実用面において困難が伴う．コンピュータが情報の真偽性を推定するには，大量の知識を使って自動推論を行う必要があるが，実用に耐えうる知識獲得や推論手法はまだ確立できていない．また，情報の真偽性は人間にも分からないことが多い．例えば，「ひまわりは土壌の放射性セシウムの除去に効果がある」という情報が間違いであることは，震災後に実際にひまわりを植えて実験するまで検証できなかった．さらに，我々は情報の信憑性と効用のトレードオフを考えて行動決定している．ある情報の信憑性が低くても，その情報を信じなかったことによるリスクが高ければ，その情報を信じて行動するのは妥当な選択と言える．このような背景から，ツイートの信憑性を計算機が直接判定する，というスタンスは必ずしも望ましいものではない．人間による信憑性判断を計算機により支援する仕組みへの要求が高まってきている．

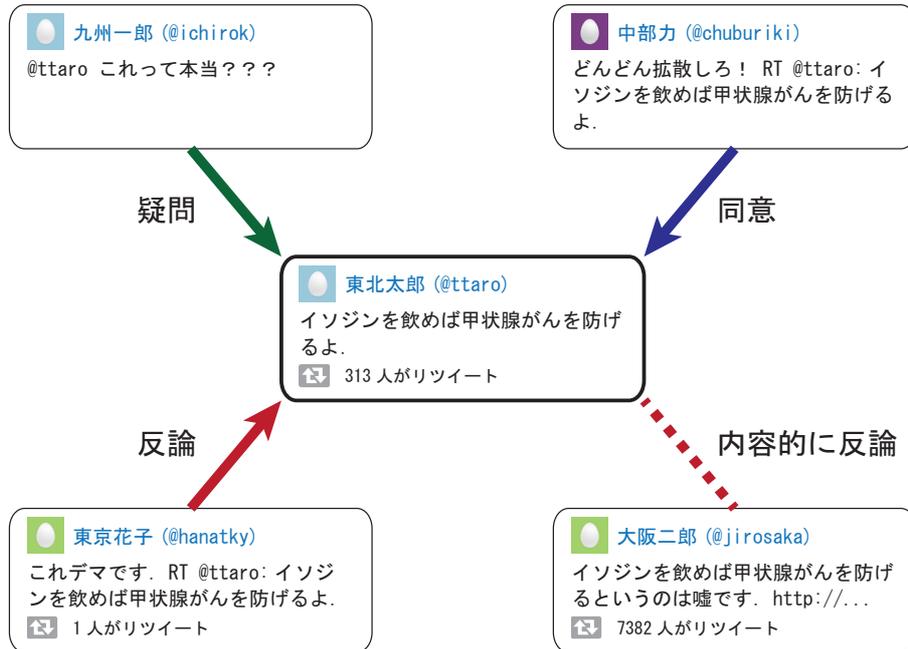


図 1.1. 返信・非公式リツイート，もしくは内容に基づくツイート間の論述関係。

1.2 本研究の目的

前節で述べたような背景を鑑み，本研究では，ツイートの信憑性を直接判断するのではなく，そのツイートの情報の「裏」を取るようなツイートを提示することで，情報の価値判断を支援することを考えている。

図 1.1 に「イソジンを飲めば甲状腺がんを防げる」という内容のツイート（中心）に対する，周囲の反応の例を示した．このツイートに対して，同意する意見，反対する意見などを提示することで，この情報の根拠や問題点，他人の判断などが明らかになる．例えば，図 1.1 左上のツイート「これって本当???'」は，中心のツイートに対して疑問を呈しており，図 1.1 左下のツイート「これデマです． RT @ttaro:イソジンを飲めば甲状腺がんを防げるよ .」は，中心のツイートに対して反論を行っている．これらのツイート間の関係情報を用いれば，中心のツイートに対して多くの反論・疑問が寄せられているため，中心のツイートの信憑性は怪しいと判断したり，右下のツイートの URL の情報を読むことで，追加情報を得ることができる．

Twitter において特徴的なのは，ツイート間に返信や非公式リツイートなどの形式を取った投稿が可能な点である．返信は，メールで返信を行うときに，返信元の内容を消去してから返信内容を書く状況に相当する．Twitter のメタデータ上では，どのツイートに対して返信を行ったのかという情報が残されている．非公式リツイートは，メールで返信を行うときに，返信元の内容を引用したままにしておく状況に相当する．元のツイートをそのままの形でフォローに送る（公式）リツイートとは異なり Twitter が提供している機能ではないが，サード

パーティ製のクライアントでサポートされており頻繁に利用されている。例えば，図 1.1 左上のツイートは中心のツイートに対する発言であること，図 1.1 左下のツイートは中心のツイートを引用したことが記されている。これに対し，図 1.1 右下のツイートは，返信や非公式リツイートの形式を取っていないため，中心のツイートを見て投稿されたものかは不明である。

本研究では，返信や非公式リツイート（以降，まとめて「返信」と表記する）の関係にあるツイートに着目し，ツイート間の論述的な関係を認識する手法を提案する。具体的には，返信ツイートによって，投稿者の「同意」「反論」「疑問」などの態度が表明されると考え，これらの態度を推定する分類器を教師有り学習で構築する。評価実験では，返信で表明される態度の推定性能を報告する。さらに，構築した返信態度の分類器を，直接的に返信関係のないツイート間の論述関係の推定に応用する手法を示す。参考のため，既存の文間関係認識器を適用して比較を行い，提案手法は深い言語処理に基づく手法と独立に使うことができ，双方を相補的に活用することでより高性能な分類が行えることを示す。

1.3 本研究の貢献

本研究の主な貢献は以下の 4 つである。

- Twitter の返信ツイートによって表明される態度を，その論述関係に基づき「同意」「反論」「疑問」「その他」の 4 クラスに分類した 2690 ツイートからなるコーパスを構築した。このコーパスは再利用可能である。
- Twitter の返信ツイートにおいて使用される反論表現を収集した。反論表現辞書は，本研究で行った実験に限らず使用可能である。
- Twitter の返信ツイートを，その論述関係に基づく上記の 4 クラスを識別する分類器を構築し，その自動認識の性能を評価した。 F_1 値は 4 クラスの Macro 平均で 0.667，反論クラスで 0.472，全体の Accuracy は 0.751 という性能を得た。また，反論表現辞書や構造的特徴の使用により性能を向上できることを示した。
- 返信態度の分類器を一般のツイート間の論述関係認識に応用する手法を提案し，具体的なトピックを用いてその有効性を評価した。提案手法の性能を既存の文間関係認識技術による手法と比較して示すと共に，二つを組み合わせることでよりツイート空間の論述構造グラフを作成し，ユーザ支援を行う構想を示した。

1.4 本論文の構成

本論文の構成は以下である。本章では，本研究の背景・目的および貢献について述べた。2 章では，そもそも信憑性とは何かについて整理した後，ユーザの信憑性判断を支援するにはどのようなアプリケーションが有効かについて述べる。3 章では本研究の関連分野を示し，代表的な研究について整理して示す。4 章では，返信ツイートで表明される態度の推定手法および評価を示し，提案手法の有効性と残された課題について述べる。5 章では，3 章で構築した返

信態度の分類器を一般のツイート間へ拡張する手法および評価を示し，文間関係認識に基づく手法との比較を行いつつ，提案手法の有効性と残された課題について述べる．6 章で結論と今後の展望を述べる．

第 2 章

信憑性判断支援のための ツイート間関係認識

2.1 はじめに

本研究の目的は、ツイート情報の「裏」を取るようなツイートを提示することで、ユーザによる情報の信憑性判断や価値判断を支援することである。本章では、まずそもそも信憑性とは何かについて述べる。その後、ユーザーによる情報の信憑性判断を支援するために必要な要素を検討し、最終的にツイート間関係認識を解くことが重要であることを述べるため、関連研究を交えながら本研究において解くべき課題について説明する。

2.2 信憑性

2.2.1 信憑性の定義

信憑性に関する研究は、1950年代にコミュニケーション学や社会心理学の分野から始まった [2]。一般に信憑性は、客観的な特性である真偽や正確さとみなされる傾向にある。しかし、これらの分野では、信憑性とは主観的な特性であり、対象とする情報だけでなく受け手によって正しさの解釈が異なると主張されている。Fogg らは、信憑性 (credibility) は believability として定義されるとした [3]。すなわち、信憑性のある人とは受け手にとって信じられる人であり、信憑性のある情報とは受け手にとって信じられる情報である。信憑性の厳密な定義は未だに明確でない [4] が、Fogg らの考えは現在のところ広く受け入れられていると言える。

信憑性について重要な性質が二つある。一つは、信憑性は知覚品質であるということである。信憑性は物、人、情報などの中に存在しているものではない。したがって、信憑性について論じる際には、信憑性の認知について論じなければならない。もう一つは、信憑性の認知は、多様な特徴を同時に評価した結果生じるということである。それらの特徴が何であるかは、信憑性研究の主要な分野の一つである。

2.2.2 信憑性の構成要素

信憑性を構成する要素については様々な議論が示されてきたが、大部分の研究者に支持されている主要な要素としては、信頼性 (trustworthiness) と専門性 (expertise) の二つが挙げられる [2][3][4]。信頼性は (受け手によって認知される) 発信者の善良性や道徳性に関する要素であり、専門性は (受け手によって認知される) 発信者の知識や技量に関する要素である。これら二つの特性を評価することによって、信憑性全体を評価することができると考えられている。例えば、信憑性の高い Web 情報といった場合には、信憑性は知覚品質であることをふまえて、高い信頼性と高い専門性を持っている “と受け手が感じられる” Web 情報のことであると言える。

しかし、信憑性を構成する要素が何であるかについては、今なお議論が続いている。信頼性と専門性は二つの主要な要素ではあるものの、やや抽象的な概念である。Kapoun らは、大学生が Web 情報の信憑性を判断する際のガイドラインとして、正確さ、権威、客観性、鮮度、詳細性の五つに分類したチェックリストを提案した [5] が、ユーザのコストが高く実用性に欠けると批判的な視線を向ける研究者も存在する [6]。

2.2.3 人間の信憑性判断モデル

社会心理学における重要な概念の一つに「態度」というものがある。態度とは、ある対象についてどのように反応するか決めていた状態のことを言う。態度が変化することを態度変容と呼び、どのような場合に態度変容が引き起こされるかが重要な研究テーマとなっている。特に、態度変容を引き起こそうとするコミュニケーションのことを説得的コミュニケーションと呼ぶ。Hovland らは、説得的コミュニケーションの効果に影響する重要な要素として、メッセージの伝え手・メッセージの内容・受け手の素質・反応、の 4 つが重要な要素であると指摘した [2]。この中でも、メッセージの伝え手、つまり情報発信者に関しては、その信憑性が非常に重要な要素であるとされた。Hovland らによれば、同一のメッセージ内容でも、情報発信者が異なる場合には態度変容に異なった影響を及ぼすことが確認されている。

このように、社会心理学の分野では、説得理論の構築を目指して信憑性研究が行われてきた。特に近年では、信憑性の低いと思われる情報が氾濫する Web に注目が集まってきている。多くの研究者が、ユーザが Web 情報の信憑性をどのように判断しているのかについて、アンケートに基づく調査を行った。これらの研究では興味深い結果が得られている。まず、Web 情報は基本的に信用できているユーザは多い [7]。また、検索エンジンの検索結果が Web ページの信憑性に従ってランキングされていると考えているユーザも少なくない [8]。さらに、Web ページの信憑性に関して、情報の中身よりもデザインから大きな影響を受けることも明らかにされている [1][9]。信憑性の分析やその判断支援のための技術が要求されていることは明らかであろう。

さらに、ユーザの信憑性判断プロセスについての研究も多く行われてきた。有力なものの一

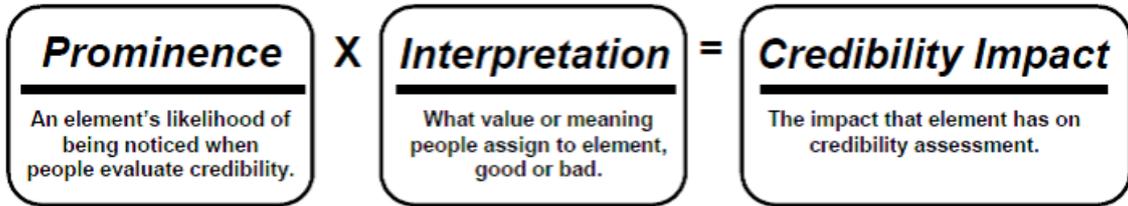


図 2.1. Prominence-Interpretation 理論 [1]

つとして、Fogg らの提案した P-I 理論がある [1]。Fogg らは、4 年間にわたり 6,500 人のユーザの協力を得て、人々が Web 情報の信憑性をどのように判断するかモデルを示した。彼らの理論によれば、Web 情報の信憑性判断に与える影響は、Prominence と Interpretation の積で表すことができるとされている (Fig.2.1)。すなわち、ユーザが Web ページにおいて信憑性に影響を与える要素に注目する確率と、見つけた要素をユーザが解釈する評価値とを掛け合わせることで、ユーザの信憑性判断に与える影響の大きさを測ることができる。例えば、フィッシングサイトを警告するツールが意外と役に立たないという研究結果があるが [10]、P-I 理論によれば、ユーザが警告に注目する確率が低いからであると説明できる。また、Hillgoss らは、信憑性判断には 3 つのレベルがあることを、ユーザ実験によって明らかにした [4]。彼らのユーザ実験からは、ユーザが信憑性判断を行う際には、信憑性の概念化 (construct)、様々なケースに適用できる経験則 (heuristics)、情報と情報源に関連付けられた特定の性質 (interaction) の 3 つのレベルの組み合わせによることが分かっている。

なお、ある情報に対し反対意見を示すことでユーザに正常な判断を促すことは心理学的にも明らかにされており [13]、信憑性判断を支援するのにもまた重要な点である。

2.2.4 本論文における信憑性の扱い

以上、主に社会心理学の分野の諸研究を主に参考とし、信憑性に関して述べた。本論文では、信憑性に関して上記と同等の立場に立つ。すなわち、信憑性とは情報の受け手によって決まる知覚品質であり、情報それ自体に絶対的に備わっているものではないと考える。Twitter 情報の信憑性と言った場合には、情報それ自体の真偽などではなく、ツイートを、あるいはツイート内の何らかの主張を信じられるかどうか、という意味合いで用いる。さらに、情報の信憑性を判断するとは、その情報を信じるかどうかを決めるということである。

本研究の目的は、ユーザが Twitter 情報の信憑性を判断するのを支援すること、すなわち Twitter 情報を信じるかどうか決める手助けをするということである。したがって、例えばある情報が書かれたツイートに対し、これはデマである、と決定して提示することを目指しているわけではない。あくまでもユーザがその情報を信じるかどうかの判断を支援するという立場である。この立場に立つ理由としては大きく二つある。一つ目は、情報がデマであるかどうか、いわば真偽の判定によりユーザの支援を行う場合、そのシステムには実現困難な高性能が要求されうるという点である。あるデマが人命に関わるようなものである場合は、システムの

提示する情報が間違っていた際には重大な実害を及ぼす可能性がある。いわゆるセーフティファーストな情報を提示することも可能であるが、それほど有用なものとはならない可能性が高い。仮に情報の真偽を 100 % の精度で判定できるシステムを構築できればよいが、それは事実上不可能である。二つ目は、真偽の明確な問題に対象を絞ると、扱える情報の範囲が狭くなる点である。例えば、「ひまわりは土壌の放射性セシウムの除去に効果がある」という情報が間違いであることは、震災後に実際にひまわりを植えて実験するまで検証できなかった。すなわち、その情報が Twitter で出回った時点では、真偽が判明していなかったということである。本研究ではこのような情報も扱いたい。実際の災害等の際に重要なのは、「今分かっていることは何か」をはっきりさせることである。真偽が明らかでないならば、そのことが分かるような、あるいは分かるための助けになるようなシステムが求められる。以上の二点から、本研究においては、「信憑性は情報それ自体ではなくユーザの受け取り方であり、機械ができるのは判断支援のみ」という立場を取る。次節では、このような立場に立ちつつ、Twitter 情報の信憑性判断を支援するアプリケーションとしてふさわしい形について検討する。

2.3 Twitter 情報の信憑性判断を支援するアプリケーション

本研究の最終目標は、Twitter 情報に対し何らかの処理を行い提示することで、ユーザの情報信憑性判断を支援することである。当然の帰結として、何らかのアプリケーションを想定する必要がある。例えば、序論の図 1.1 において示したツイートグラフはその一例である。このツイートグラフは、入力したツイートに関連するツイートを、その論述関係を伴う形で可視化する、というものである。以下、想定するアプリケーションについて、関連研究を交えながら検証する。

2.3.1 トピックに関する無秩序なまとめ

トピックとは何らかの話題を指し、抽象度が高いトピックとしては「スポーツ」「政治」などとなる。一方、抽象度の低いトピックとして表すならば「原発の再稼働に反対」などである。Twitter では、ユーザが何らかのトピックに関連するツイートを手動でまとめ、他のユーザに公開する Togetter^[11] という既存のサービスがある。Togetter は、ユーザが特定のカテゴリ（トピックに相当する）に属すると思われるツイートをまとめて表示できるものである。Togetter はそれ自体は便利で利用者の多いサービスであるが、あるユーザが適当にまとめたものが、他のユーザにとって有用なものとは限らない。まとめ方が特定の人物その他を貶める恣意的なものである、と反発するユーザも多い。信憑性判断支援には役に立たないと言ってよい。

2.3.2 トピックに関する複数ツイート要約

自然言語処理の分野では、古くから文書要約に関する研究が行われてきた。文書要約とは、「受け取った情報の最も重要な内容を簡約した形かつユーザの要求に応じる形で提示するこ

と」である。精度の高い要約を行うためには言語の意味理解を必要とするため高度な分野であるが、近年の情報量増加に伴い、文書要約の重要性はますます高まっていると言える。特に代表的なタスクとしては、重要文抽出が挙げられる。重要文抽出とは、字数などの制約条件が与えられた時に、目的関数を最大化するように文書中からいくつかの文章を選択することである。 $x = s_1, \dots, s_n$ から字数制限 B のもとで重要文の集合 $\hat{y} \subseteq x$ を選ぶタスクは、式 2.1 のように定式化される。

$$\hat{y} = \operatorname{argmax}_{y \subseteq x} F_x(y) \quad \text{s.t. } |y| \leq B \quad (2.1)$$

目的関数 $F_x(y)$ は、重要度や冗長性を考慮して決められる。

Twitter におけるトピックの要約は、複数文書からの要約に相当するタスクである。一つのツイートを文書とみなした時、あるトピックに属する多数のツイートの中から重要なツイートを選び出してユーザに提示することができれば、ユーザの情報取得にとって便利だと考えられる。代表的な研究としては Duan らの研究 [12] が挙げられる。Duan らは、時間の経過と共に変化していく Twitter におけるトピックの要約を作るため、ユーザ空間とツイート空間の重要度を相補的に利用する手法を提案した。すなわち、重要なツイートを投稿したユーザは重要度、すなわち影響力が高く、影響力の高いユーザの投稿するツイートは重要度が高いという仮定を置き、重要度が時間の経過に伴い更新されていくアルゴリズムを採用した。この重要度のランキング手法において先行研究を上回る結果を得た。

要約を見れば、あるトピックについて知りたいユーザは必要な情報を得ることができ、情報取得にかかる時間を大幅に短縮することができる。しかしながら、要約された情報を見て、ユーザが情報の少なさに不安を感じる可能性もある。要約を示すにあたっては、その要約がトピックを表す十分性を持っていることをユーザに納得させなければならない。要約は有力であるが、信憑性判断支援に役立てるには工夫が必要であると言える。

2.3.3 命題に対する言明の分類

命題とは、「A は B である」という形式での記述が可能なものであり、真偽が存在する。狭義のトピックと考えるもよい。「A は B である」という命題を与えた時に、「A は B である」と「A は B でない」の立場に分類して表示することは、ある情報に対する反対意見を示すことでもあり、前述したようにユーザに正常な判断を促すことにつながる [13]。さらに、ユーザが入力した命題に対する答えが返ってくる形式であり、ユーザは必然的に返ってきた答えに注目するため、P-I 理論に即した面も合わせ、信憑性判断支援のために非常に有効であると考えられる。

命題に対する言明の分類が直接的な形で行われているものはディベートサイトである。例えば Debate.org^{*1} では、示されたトピック（議題）に対する賛成・反対の立場を取って意見を書き込むことができる。これを第三者のユーザから見れば、対立する意見を見比べることができ、信憑性判断のために有用である。

^{*1} <http://www.debate.org/opinions/>

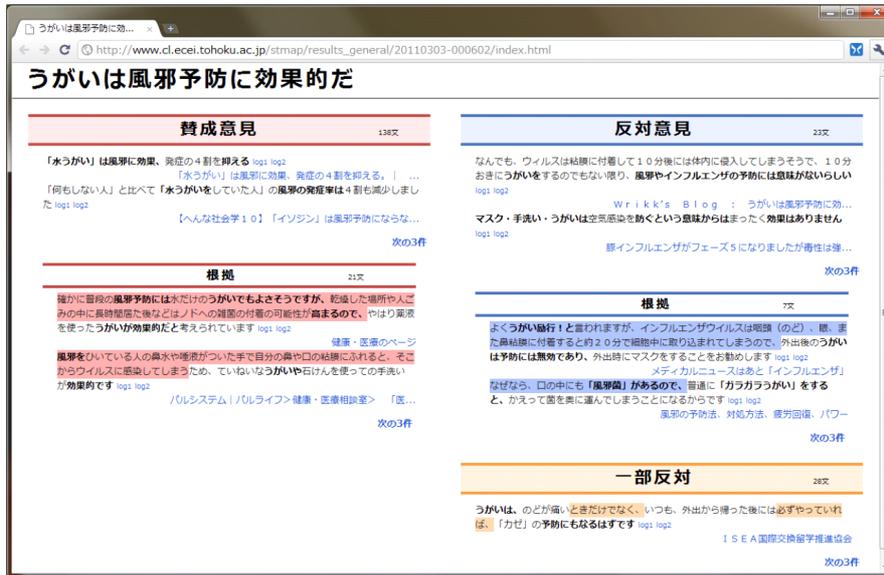


図 2.2. 言論マップの分析例 .

近年は、特に日本において、命題に対する言明を自動で分類し可視化する研究が行われている。代表的なものとしては、WISDOM[14] や言論マップ [15]*2 が挙げられる。これらの研究は、構文解析などの深い言語処理を行うことで、ある命題に対する肯定・否定を Web ページ上から探索・分類し、ユーザーの意思決定を支援することを目指している。例えば、「うがいは風邪予防に効果的だ」という命題を与えた時に、効果的であるという立場と効果的ではないという立場の意見を分類して表示する。WISDOM は、それに加えてその意見を発信した人の属性の統計情報なども合わせて示し、どのような人がどのような意見を持っているのかをユーザに知らせる工夫をしている。一方言論マップは、その意見の根拠を提示したり、一部賛成・一部反対など微妙な立場の意見の分類を行っており、論理関係の解析に重きを置いたシステムである (図 2.2)。

しかしながら、これらのシステムを Twitter で実行するには二つの問題がある。一つ目は、Twitter は、Web ページ上の情報のようなしっかりとした文章形式で書かれたツイートは少ないため、深い言語処理を行う難しさが大幅に増すことである。Twitter には 1 ツイートに 140 字以内という制限があり、その名前が表すように投稿は「つぶやき」の類が多く書式もまばらであるため、Web ページのテキストと同等に扱うのは困難である。二つ目は、真偽という形で分類できる命題を与えなければいけないことである。時には命題化の難しい問題に対する情報が欲しいこともある。

*2 言論マッププロジェクト:<http://www.cl.ecei.tohoku.ac.jp/stmap/>

2.3.4 ツイートに対する他ツイートの関係分類

入力としてあるツイートを与えた時に、それらと何らかの論述関係を持つツイートを抽出して表示することになる。序論の図 1.1 で示したツイート間の論述関係グラフがこれに該当する。このようなアプリケーションのメリットとしては、命題化を行わずとも使用できることや、グラフ化して表示するという観点から必然的に要約的要素も含むことが挙げられる。抽出できたものの精度が高ければ、ユーザに対するシステムの説得力としても高いものになる。また、ツイートに対する関係を用いることから、返信や引用といった特殊な関係の存在するツイートに対しては、抽出については難しくないため分類のみを行えばよいという利点もある。一方デメリットとしては、ツイート間の関係認識は、少なくとも命題に対する言明の分類以上の難しさがある。ツイートは単文とは限らないことや、その書式がまばらであること、そもそも賛成・反対のような形で分類できるとは限らないことから、少なくとも文間関係認識以上には難しい。また、直接返信や引用の関係にあるツイートが多いケースはむしろ珍しく、それらの関係にないツイートとの関係を認識できなければ、あまりよいグラフができあがらないということがある。

2.3.5 各アプリケーションのまとめ

以上の内容をふまえて、想定するアプリケーションについての見解をまとめると以下である。

- 入力としてはトピックかツイートであり、ツイートのほうが生成の手間はかからない。
- 出力としては要約も有効な手段であるが、情報が偏る可能性は否定できない。関係を表すに留めるほうが無難である。
- トピックに対する関係の分類とツイートに対する関係の分類では、言語处理的には後者のほうが難しい。
- ツイートに対する関係の分類では、返信や引用などの構造的な関係を利用できる場合がある。

2.4 まとめ

前節までで、信憑性とは何かについて述べ、ユーザの信憑性判断を支援するためのアプリケーションとしてどのようなものが想定されるかについてまとめた。本研究では、2.3.5 節で列挙した点のうち、特に最後の点に注目する。すなわち、ツイートに対する関係の分類を言語处理的に解こうとした場合はトピックに対する関係の分類より難しいが、Twitter 上の構造を利用することでの改善が望めると考える。本研究では、ツイート間の論述関係を、命題への変換などの特殊な過程を通さずに直接行うことを考え、それを返信などの構造を利用することにより実現する手法を提案する。

第 3 章

関連研究

本研究の関連分野は多岐に渡るが、特に関連の深い分野としては、情報の信憑性分析と、文間関係認識、評判分析などの分野が挙げられる。本節ではこれらの研究分野について、Twitter を対象に行われているものを中心に整理する。

3.1 Web 情報の信憑性分析

Web 情報の信憑性を判断する研究はこれまでも多く行われてきているが、近年は Twitter を対象としたものも多い。Twitter 上に流れるデマを見つけ出す研究はその典型である。Castillo ら [16] は、手動で信憑性のラベル付けをしたデータを学習させ、ツイートやアカウントの特徴を用いて Twitter 上の情報の信憑性を自動で判別する分類器を訓練し、高い性能を得た。彼らは流行しているトピックをニュースとその他に分類し、ニュースが正しいか間違っているかを分類している。Qazvinian らは、教師あり学習を用いて Twitter 上に流れる「噂」を発見する研究を行った [17]。彼らは噂を誤報とデマの二種に分け、誤報は単なる誤り情報であり、デマは意図的な誤り情報であると定義した。特徴としてテキスト内容の特徴・ネットワークの特徴・Twitter 特有の情報の大まかに 3 種に分け、さらにそれらを細かく分類し、計 7 種の特徴グループを作った。各特徴グループについて、噂かそうでないかの分類器をナイーブベイズを用いて構築し、最終的に各グループの和を取ることで噂かそうでないかを判定するという手法を取り、用意したベースライン手法を上回る性能を得た。また、ソーシャルネットワークの関係をグラフ化し、解析する研究も盛んである。Twitter のグラフ関係を利用したマイニングを行っているものとして、Guputa らの研究 [18] が挙げられる。Gupta らは、Castillo らと同様の分類器を用いた手法と、グラフ上の最適化の組み合わせにより、Twitter 上で観測されたイベントの信憑性の推定を行った。彼らは信憑性を、イベントが起こった確からしさとして扱っている。アカウント・ツイート・イベントに分類器の結果に基づく信憑性の初期値を与え、それらをノードとするネットワークを構成し、リンク関係を用いて信憑性の値を更新していくことで、最終的にイベントの真偽を判定する。これらの研究は、いずれもデマかそうでないかという真偽判定の問題を扱っている。しかしながら、情報の真偽を判断する事はそもそも難しく、真偽が存在しない情報に対応できないという問題点がある。

一方，ユーザの信憑性判断を支援するという立場の研究も多い．信憑性を自動で判断する手法はそもそも事実性の明確な情報にしか適用しづらいが，真偽の不明確な情報に対しても信憑性の検討に対する要求は存在するため，このような立場は有力である．例えば，検索エンジンにおける信憑性判断支援を目的とした研究として，山本らの研究 [19] や Schwatz らの研究 [20] が存在する．山本らは，ユーザの信憑性判断モデルの推定により検索結果の再ランキングを行う手法を提案した．また，Schwatz らは，信憑性判断に役立つと思われる情報を検索結果のページに表示し，その有効性を評価した．他にも，ある情報に対し反対意見を示すことでユーザに正常な判断を促す [13] ことに着目し，ユーザが Web ページを閲覧中に疑わしいと思ったセンテンスを選択すると，反証候補を Web から検索し提示する DisputeFinder[21] などの研究がある．

本研究は，情報信憑性分析の研究としての立場は後者に属する．すなわち，直接情報の真偽性の判断を行うのではなく，価値判断を行いたい情報の周辺を整理し提示することで，情報の受け手側の行動決定を支援することを目指す．

3.2 情報の整理と文間関係認識

多様な情報を整理する際にまず考えられるのは，情報の要約を行うことである．要約とは，情報の中から重要な情報を過不足なく抜き出してまとめることを指し，情報単位としては文や文書が主であるが，近年では Twitter の要約も行われている．Duan らは，重要なユーザーは重要なツイートをするという仮説に基く Twitter 上の話題（トピックと呼ぶ）に関するツイート群の要約を行い，先行研究を上回る要約性能を得た [12]．しかしながら，信憑性分析支援を行うという我々の立場では，重要なツイートを抽出するだけでは不十分であり，ツイート間の関係を明らかにする必要がある．

本研究に最も近いものは，水野らの言論マップ [15] である．言論マップは，Web 上の文章を対象とし，与えられたクエリと検索対象文の間の意味的關係，すなわち同意するか対立するかに基づいて意見を分類している．この文間関係認識は，近年盛んに行われている含意関係認識 [22] を対立や根拠などの関係に拡張したもので，根拠関係などのより広範な分類を扱うことにより，言論構造の把握を目指している．このような文間関係認識を，本研究で目的とするツイート間の論述関係の推定に適用するには，いくつかの問題がある．1 つ目は，ツイート同士が共に単文でなければ，文間関係認識は直接適用できないということである．ツイートが複数の文から成る場合には，そのツイートの談話構造を解析し，ツイートの意味するところをまとめなければならないが，このような解析は容易ではない．もう 1 つは，ツイートの書式の乱雑さである．水野らの研究では，整った文に対しても対立関係を認識するのは容易ではないという結果が出ており，字数が制限されただけの文体も多いツイートに対して認識を行うのは一層困難となる．ツイート間の関係認識を行うには，言語処理のみでは限界があると考えられる．本研究は，ネットワーク構造に基づくアプローチを提示し，深い言語処理と組み合わせることにより高精度なツイート間の関係分類を行える可能性を示すものである．

3.3 評判分析

本研究は、返信・引用ツイートによって、投稿者の「同意」「反論」「疑問」などの態度が表明されると仮定しているため、評判分析に関する研究とも関わりが深い。評判分析を本研究に当てはめて述べるとするならば、評判の対象は返信元・引用先のユーザー、もしくはその主張である。評判分析に関する研究の中で近年特に盛んなのは、極性分類と呼ばれるタスクである。評価極性による意見の分類を行う研究は、Turney の研究 [23] 以降、数多く存在する。極性としては良い/悪いという評価、あるいは正の感情/負の感情が用いられる。Speriosu らはグラフ上での値の更新を利用したツイートの感情分類を提案している [24]。グラフを構成するノードとしてはユーザやツイートの他に、単語 N グラムや分類器の学習に使用した感情表現などを加えている。ツイート及び分類器の学習に用いた素性をシードとして各ラベルの確率重みを初期値とし、それらのラベル（ポジティブとネガティブの2値）をグラフを通して拡散させることで先行研究を上回る結果を得て、ネットワーク構造の利用の有効性を示しているが、ユーザーのフォロー関係の有効性については示せていない。

元々は製品に対するレビューなどを大雑把に分類するのが目的であったが、近年では、評価対象をより明確にした極性分類が盛んである。例えば、「Windows7 は Vista よりもずっと良い」という文から、Windows7 に対する良い/正の感情と Vista に対する悪い/負の感情を区別して抽出したいという要求がある。代表的なものとしては、WISDOM[14] が挙げられる。WISDOM では、ある評価対象による極性にに基づき意見を分類し、ユーザーの意思決定を支援することを目指している。Jiang らは、評価対象との構文関係に関するルールを素性に用いた上で、リツイートなどの関連ツイートを利用することにより、評価対象を明確にした極性分類の性能を向上させた [25]。Hassan らの研究 [26] は、目的こそ違うものの本研究に近いタスクを解いている。彼らは、ディベートサイトにおける各議論トピックへの参加者のグループ分けを行った。まずユーザ間のポストを positive か negative かに分類し、同じ意見を持つ者同士は positive なやり取りが多く、違う意見を持つ者同士は negative なやり取りが多いという仮定に基づき、最終的に同じ意見を持つ者同士をまとめたグループを出力し、高い精度を得た。これは、評価対象を返信先のユーザーとした評判分析を利用し、投稿間の関係分類を行っていると言える。ただし、本研究は多様な意図の存在するツイートデータを対象としているおり、ディベートサイトでの投稿のように、相手の投稿への同意や反論を仮定することはできない。「疑問」の認識を目指す点も異なる。

第4章

返信態度の分類

4.1 はじめに

2章では、ユーザーの情報信憑性判断を支援するためには、ツイート間の論述関係を認識することが重要であることを述べた。本論では、ツイート間の論述関係を認識するための手法を、2段階に分けて述べる。まず本章では、返信・引用という特殊な構造に着目し、それらにより表明される「同意」「反論」「疑問」などの態度を教師あり学習により分類する手法を示すと共に、その性能を評価する。次章では、構築した分類器を利用し、一般のツイート間関係の認識に役立てる手法を示し、実際に抽出できたツイート間の論述関係を提示する。

4.2 各投稿形式の整理と本論文中における表記

本研究における「返信」の概念を明確にするため、Twitterにおけるいくつかの投稿形式を整理する。

返信 @で始まる投稿であり、特定のツイートに対する返信として投稿すること。Twitterの提供する機能であり、関係が記録されている。

公式リツイート ある投稿をそのままの形でフォロワーに拡散すること。Twitterの提供する機能である。

非公式リツイート ある投稿を必要に応じて編集しつつ、自身のコメントを付加して投稿すること。Twitterの提供する機能ではないが、使用者が多い。なお、自身のコメントを付加していないものについても非公式リツイートとして扱う。

本論文では、特に断りのない限り、「返信」と「非公式リツイート」の二つをまとめて「返信」と表記する。分けて扱う必要がある場合においては、前者を「返信」、後者を「引用」と表記する。また、「公式リツイート」については「リツイート」と表記する。

4.3 返信ツイートの態度分類手法

本研究では、返信・非公式リツイート（以下引用）の態度分類を行うことで、ツイート間の論述構造を解析する手法を提案する。返信や引用では、返信先・引用元の主張に対する態度や、ユーザーそのものに対する態度が表明される。これらの態度を分類することで、ツイート内容やユーザー間の関係を用いてツイート空間の構造を解析することができる。本節では返信・引用で表明される態度分類の手法について述べる。

4.3.1 ツイートの意図

返信は、あるユーザー A から他のあるユーザー B に対し発信するツイートである。したがって、ツイート中には A から B に対する何らかの態度が含まれると考えられる。一方引用は、あるユーザー A が他のあるユーザー B のツイートの全体もしくは一部を引用し、文を追加して発信するツイートである。この中にも、A から B に対する何らかの態度が含まれると考えられる。返信・引用ツイートに含まれる意図は様々である。相手の主張に同調、あるいは反発するツイートもあれば、相手や周囲に疑問を投げかけるツイート、引用して情報を補足するツイートなどもある。他愛もない会話や、時には相手の主張ではなく相手そのものに対する敵意のむき出しなどもある。多くのツイートを観察した結果、返信・引用に多く見られる意図は以下のようなものが見られた。

主張の支持 相手のツイートで主張されている内容を支持するもの。

主張の否定 相手のツイートで主張されている内容を否定し、あるいは反発するもの。

感謝 相手に感謝の意を表明するもの。

崇拜 相手に対し、崇拜の意を表明するもの。周囲へ知らしめようとするケースが多いため、引用に多く見られる。

強い注意 デマである可能性を呼び掛けるなど、相手に強い注意を促すもの。

質問 相手や他のフォロワーに対する質問。

回答 質問に対する回答。

懐疑 情報や情報源に対する疑いの気持ちを示すもの。

補足 情報の追加提供を主な目的とするもの。部分的な支持・否定である場合も多い。

会話 上記のどの意図も特に含まれていない、他愛もない会話。

無関係 返信先・引用元の内容とは無関係の別話題に対するツイート。

ツイート空間の論述構造を分析するためには、これらの多様なツイートを認識する必要がある。しかしながら、上で挙げたクラスを全部識別するのは、第一に非常に困難である。第二に、ここまでの細かいクラス分けは本研究の目的とそぐわない。すなわち、ユーザーの情報信憑性判断を支援するには情報が過剰である。実際にこれらのクラスを識別することができたとしても、それをユーザーに分かりやすい形で提供するのが困難である。したがって本研究で

は、より目的に沿った形でのツイートの分類を目指す。

4.3.2 問題設定

返信には様々な意図のツイートが存在する。相手の発言に同調、あるいは反発するツイートもあれば、相手や周囲に疑問を投げかけるツイート、引用して情報を補足するツイートなどもある。ツイート空間の論述構造を分析するためには、これらの多様なツイートを、その投稿の意図に従いいくつかのグループに分類する必要がある。そこで本研究では、返信で表明される態度を、「同意」「反論」「疑問」「その他」の4クラスに分類するタスクを考える。

同意 主張の支持や感情的な同調（感謝や崇拜も含む）など、返信先のツイートに対して明確な同意の意図が感じられるもの。

反論 主張の否定や感情的な反発など、明確な反論の意図が感じられるもの。発言者に対し強く注意を促すようなものも含める。

疑問 返信先に対して情報を要求しているが、明確な反論とは言えないもの。情報源を要求するようなものや、引用部に対する疑問の吐露なども含まれる。

その他 上記のどれにも分類できないもの。

また、具体的なツイート例とその分類を示す。

1. デマです。みんな冷静になろう。@XXX コスモ石油が否定 「火災で有害物質降る」のメール連鎖 <http://...> 同意
2. @XXX 火災で壊滅した、コスモ石油のコンビナートにあった科学燃料の詳細を記者会見で発表して報道しろ！「危険というのはデマです。」なんて情報で納得するわけないだろ！（返信先：... コスモ石油の爆発で有害物質の雨が降る件はデマ。広げてしまった方はツイート削除の上、訂正を / コスモ石油が否定...）反論
3. そうであったとしても、雨カップとかは持ってた方が良いよね。QT @XXX: ... 千葉県、近隣圏に在住の方に有害物質が雨などと一緒に飛散するという虚偽のチェーンメールが送られています。... その他
4. 硫黄分の多い原油が燃えると酸性雨につながる可能性があるんで、それに尾ひれはひれがついたものと推測します。 @XXX ... 千葉県、近隣圏に在住の方に有害物質が雨などと一緒に飛散するという虚偽のチェーンメールが その他
5. @XXX JFE ケミカル等含めた火災による被害が無いと言うことでよろしいですか？（返信先：... コスモ石油の爆発で有害物質の雨が降る件はデマ。広げてしまった方はツイート削除の上、訂正を / コスモ石油が否定...）疑問
6. @XXX NHKのニュースでは今のところ有毒物質が発生することはないと言っていますが、あなたのツイートのソースは何ですか？（返信先：... コスモ石油の爆発により有害物質が雲などに付着し、雨などといっしょに降るので... コピペとかして皆さんに知らせてください！！）疑問

7. @XXX ありがとうございます！ 同意
8. 天皇陛下！万歳！\ (T-T) / RT @XXX: 天皇陛下は、宮内庁が京都御所に避難してほしいと要望したのを... お断りになったそうじゃないか。... 私は心から尊敬します。
同意
9. (怒) RT @XXX: ... 東京まで健康被害が現れるような Pu は飛んでこない... RT @YYY: もしプルトニウムが漏れてたら東京から逃げますか?... 反論
10. 馬鹿左翼、煽るな。RT @XXX: ... 日本政府は事故の重大性をまったく認識していない。今すぐに多国籍軍を総動員して封じ込めないとチェルノブイリ以上の被害が出る 反論

これらの4クラスを設計した意図について、具体例を参照しつつ述べる。ツイート空間の整理に必須なのが、対立構造の抽出である。そのために「同意」と「反論」の2クラスを設定する。「同意」のクラスによって結びつけられたツイート群は、何らかの主張やユーザーに対して同様の態度を表明し、「反論」のクラスによって結びつけられたツイート群は、何らかの主張やユーザーに対して異なる態度を表明する。この2クラスを設定することで、ツイート空間を、同じ態度を表明するツイートクラスタの集合として整理できる。なお、「同意」や「反論」については態度が明確なもののみを分類する。前述したように、これら2クラスがツイート間の論述構造を解析する上で中心的な役割を果たすため、曖昧な態度のツイートをこれらの2クラスに分類しないように注意する必要がある。ツイート例(1)(2)(3)は、「コスモ石油の爆発で有害物質の雨が降るとするのはデマ」というツイートに対する返信である。(1)はその内容を受け入れているので「同意」、(2)は納得しないと主張しているので「反論」に分類する。(3)は、デマであることを受け入れてはいるが、完全に安心はしていないことが読み取れる明確な同意とまでは言えないため、「その他」に分類する。また、(4)は内容を受け入れてはいるが、返信先の情報への同意よりも自分の考えを表明するためのツイートと見られるため、このようないわば補足的な意図を持つツイートも「その他」に分類する。

「同意」と「反論」の2クラスが対立構造を持つのに対し、「疑問」と「その他」の2クラスは、話者の立場としては中立と言える。もしツイートAがツイート空間全体において重要なツイートであり、ツイートBがツイートAと「その他」の関係にある場合は、ツイートBは論述構造を明らかにする上で重要とは考えられない。さらに「疑問」を加える。情報の信憑性を判断するにあたり、「疑問」を持つユーザーの存在自体を示すことも重要であるし、疑問に対する回答は有益な情報を与えるケースが多い。そのため、論述構造を整理する上で重要な関係と考え、「疑問」クラスを設定した。ツイート例(5)は純粹な疑問、いわば質問の例である。(6)は情報源を求める、明確な反論とまではいかない懐疑的なツイートであり、「疑問」に分類するものとする。

なお、(7)(8)(9)(10)のような感情的な同調あるいは反発のツイートは、内容(引用の場合は付加部分)には有用な情報は含まれていないが、誰が誰に同調、あるいは反発したかというユーザー同士の関係を推測することは、論述構造分析の助けになると考えられる。したがって本研究においては、感情に基づくと考えられるツイートについても、「同意」や「反論」として同等に扱い分類を行なっている。

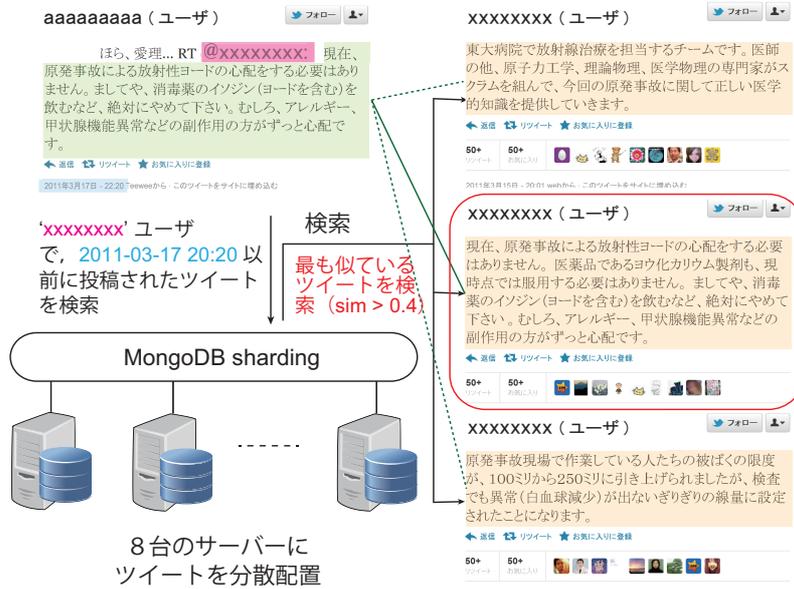


図 4.1. 非公式リツイートのリンク情報の復元 .

4.3.3 返信ツイートのアノテーション

返信ツイートの分類を行うにあたり、正解データを準備する．本論文で行う実験で使用するデータは、ホットリンク社より提供された、2011 年の 3 月 9 日から 4 月 4 日までのツイートデータ（以下 hotto コーパス）である*1．hotto コーパスには、#tsunami や#jishin など震災に関連するハッシュタグまたはキーワードが含まれるツイートと、そのツイートを投稿したユーザーやそのプロフィール情報などが収録されている．収集対象ユーザー数は約 100 万人、ツイート数は約 2 億 1 千万ツイートである．また、各ツイートには公式リツイート（以下 RT）や返信の関係が含まれており、これらはそのまま利用した．

ただし、非公式リツイートの引用元については記録されていないので、以下の方法で復元した（図 4.1 参照）．ツイートの “[RQ]T username:” というパターンが出てくるとき、そのパターンに続く部分（図 4.1 左上で緑色の網掛けの部分）をどこかのツイートから引用したと考える．そして、ユーザー “username” のツイートの中で、投稿日時が引用ツイートよりも前のものをすべて検索する．これらのツイート群に対し、引用部のテキストとの類似度を文字トライグラムのオーバーラップ係数で計算し、これが最大になるものを引用元と推定する．なおオーバーラップ係数は、文字トライグラムの集合 X と Y に対し、以下の式で定義される（シン普森係数とも呼ばれる）．

$$\text{オーバーラップ係数} = \frac{|X \cap Y|}{\min(|X|, |Y|)}$$

*1 東日本大震災ワークショップで提供されたツイートデータを利用したかったが、返信の情報が含まれていなかったため、断念した．

表 4.1. 選出した 20 個のクエリ

イソジン	コスモ石油	トルコ	花山村	関西電力
教科書 福島	空中投下	埼玉 水道水	東大 入学	双葉病院
煮沸 ヨウ素	谷垣	自衛隊 支援物資	ホウ酸	天皇陛下
茨城 県知事	医師会	メルトダウン 爆発	農水省	築地市場

図 4.1 では，“XXXXXXXX” のツイートを検索した結果，右側真ん中のツイートの類似度が最も高かったため，左上のツイートの引用元と推定している．ただし，オーバーラップ係数の最大値が 0.4 未満である場合は引用元が見つからなかったとする．

返信関係のアノテーションには，震災の際に出回ったデマ一覧^{*2} を参考に，20 個のクエリ（トピック）を選んだ．デマ一覧を利用したのは，東日本大震災の時に情報の信憑性で問題が上がったことと，反論関係が比較的多い割合で含まれると考えたからである．クエリの選出基準は，そのクエリを含むツイート集合に対し，十分な返信の数が存在することである．具体的には，後で述べるように，返信・引用が各 100 個ずつ以上取得できるようなクエリを選んだ．選んだクエリを表 4.1 に示す．各クエリについて hotto コーパスを検索し，公式リツイート（以下 RT）が多い順に並べる．上位のツイートから順に，返信ツイートを取得する．なお，あるユーザー間で相互の返信が続いている場合には，反論の応酬であるなど重要な論述構造が存在する可能性があり，有用なデータとなるため，返信ツイートが続いている限り全て取得している．このようにして，まず各クエリに関するツイートから返信・引用をそれぞれ 100 個ずつ，計 200 ツイートを集める．さらに，20 個のクエリに対して同様の作業を行い，重複するツイートがあった場合には一つを残して削除し，クエリ毎の返信ツイートを補充した．以上により，計 4000 個の返信ツイート（以降データセット A）を準備した．

データセット A に対し，三人のアノテーター（以下 a・b・c）の手により，前項で述べた方針に従い 4 クラスのラベル付けを行った．アノテーションの一致度合いを表 4.2 に示す．4000 ツイートのうち，3 人のラベルが一致したツイートは 2690 ツイートであった．ラベルの一致度を評価するため，ペアワイズに Cohen のカッパ係数を計算した．

$$\frac{p_0 - p_c}{1 - p_c}$$

ただし p_0 は対応クラスの出現数の一致率， p_c は対応クラスの偶然の一致率である．ab 間が 0.660，ac 間が 0.621，bc 間が 0.641 となり，十分な一致と見なせる．3 人のアノテーションが一致した 2690 ツイートを集めたものをデータセット B とする．

4.3.4 分類手法

返信の態度を教師有り学習で分類する手法について述べる．学習に用いる素性は大まかに 3 種類に分かれ，ツイート内容に関する素性，ツイート間の素性，ユーザ間の素性である．

*2 「東北関東大震災に関するデマまとめ」のまとめ <http://togetter.com/li/112556>

表 4.2. データセット A におけるアノテーションの一致度

	A	B	C	三人一致
同意	2052	2008	2284	1651
反論	416	434	496	263
疑問	418	432	369	282
その他	1114	1126	851	494
合計	4000	4000	4000	2690

表 4.3. 反論表現の例

デマを否定する表現	その他の表現
RT を取り消してください	あなたの発想が軽すぎる
このような事実はありません	お前にそれがわかってんのか
テキトーな情報流すな	余計なこと書くな

ツイート内容の素性は、単語ユニグラム、単語バイグラム、URL 数、ハッシュタグ数、デマ否定単語の有無、反論表現との一致度である。これらの素性では、返信ツイートであればツイート全文、引用ツイートであれば追加部分（ツイート全体から引用箇所を除いたもの）を対象とし、本文中のアカウント名や URL、ハッシュタグを正規表現により取り除く。取り除いたアカウント・URL・ハッシュタグに対しては、それぞれに対し 0 個・1 個・2 個以上を含むという 3 値の素性を作る。この設定は、例えば URL を 1 個も含まない場合と 1 個以上含む場合には大きな差があり、また 1 個だけ含む場合と多数を含む場合にも大きな差があると考えられる一方で、3 個か 4 個かには大きな差はないという考えに基づく。不要箇所を取り除いた後の本文に対して、MeCab[27]^{*3} による形態素解析を行い、単語ユニグラムと単語バイグラムを抽出する。なお、原形が存在する場合には原形を、存在しない場合には表層形を採用した。

デマ否定単語の有無とは、ツイート本文がデマの否定を示す単語を含むかどうかを表すもので、デマ否定単語は「デマ」「ガセ」「誤報」「虚報」「削除」「訂正」の 6 つとした。これらの単語は必ずしもデマを否定していることを示すとは限らないが、返信で用いられる場合では、デマの否定に用いられるケースが多く、反論の抽出に有効と考えられる。反論表現との一致度とは、あらかじめ作成した反論表現辞書を用い、反論に固有の表現をどの程度含んでいるかを表す。反論表現辞書は、データセット A の中から反論を表すと思われる表現を手で抽出して作成する。それらの表現にはデマを否定する表現も含まれるが、相手に対する感情的な反発を表す表現なども含まれる。表 4.3 に反論表現の例を示す。

なお、この二種類の表現は厳密に区別できるわけではない。例えば、表 4.3 のその他の表現例の「余計なこと書くな」については、「余計なこと」が指す内容がデマの可能性もある。実

^{*3} <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

際にはこれら二種類を分けて扱う必要はないが、反論表現をたくさん集めることが、デマによる否定以外の反論を認識する上で有効である。また、一見すると反論を表す表現であるが、反論表現辞書に含めるべきでないものも存在する。例えば、「それはデマです」は反論表現として挙げた一方で、「デマです」「デマらしいです」は反論表現にはふさわしくないことがある。例えば、以下のようなケースが存在する。

デマです、注意！ RT @xxx: (デマを否定する情報) RT @yyy: (デマ情報)
 デマらしいです RT @xxx: (デマを否定する情報) RT @yyy: (デマ情報)

この場合、引用元との間には共にデマを否定しているという“同意”の関係が成り立っているため、反論表現としてはふさわしくない。このように、伝聞形で使われそうな表現などは反論表現辞書から除外している。実際に収集した100個の反論表現は付録に載せた。

以上のように作成した反論表現辞書を用い、反論表現との一致度を求め、素性として使用する。まず、反論表現辞書中の各反論表現に対して、その単語バイグラムをどの程度含むかを求める。例えば、反論表現辞書が「それ/は/デマ/です」「これ/デマ/ね」の二つからなる場合を考える。「デマ/です」を含むツイートがあった場合、「それ/は/デマ/です」の3個のバイグラムのうち1個を含むので1/3、「これ/デマ/ね」の2個のバイグラムは1個も含まないので0となる。これらのうちの最大値（この場合は1/3）を、反論表現との一致度とする。実際には反論表現は100個あるので、それら100個に対して計算を行い、最大値を反論表現との一致度としている。定式化すると以下である。N個の反論表現を $r_i (i = 1, 2, \dots, N)$ とし、 r_i の単語バイグラムを $b_{ik} (k = 1, 2, \dots, w_i, w_i$ は r_i 中のバイグラムの個数)とする。あるツイートの単語バイグラムが $b_{tk} (k = 1, 2, \dots, w, w$ はツイート t 中のバイグラムの個数)で表される時、

$$\text{反論表現との一致度} = \max \left\{ \frac{|b_{ik} \cap b_{tk}|}{|r_i|} \right\}$$

である。

返信先のツイートとの間に関する素性では、ツイートのタイプ、空文、単語バイグラムのコサイン類似度を使用する。ツイートのタイプとは、返信か引用かを表す。空文とは、引用でありながらも自身のコメントを付加していないものに発火する素性である。単語バイグラムのコサイン類似度とは、返信ツイートの単語バイグラム（引用ツイートの場合は、ツイート全体から引用箇所を除いた部分）と、返信先ツイートの単語バイグラムのコサイン類似度である。コサイン類似度は、バイグラム集合XとYに対し

$$\frac{|X \cap Y|}{\sqrt{|X||Y|}}$$

で表される。

返信先のユーザとの間に関する素性として、返信の回数、返信の方向性、RTの回数、RTの方向性、共通のツイートをRTした回数、共通のURLをツイートに含んだ回数、共通のツイートに返信・引用した回数を利用する。返信の回数は、hottoコーパスの中から、返信している2ユーザー間で返信が行われた回数である。方向性とは、2ユーザー間の返信の履歴

を調べ、双方向のやりとりがある、返信相手への返信のみがある、返信がない、のいずれかを指す。RT の回数や方向性についても同様である。これらの素性はユーザー間の関係から返信による態度を推測するために使われる。例えば、RT が相互でたくさん行われている場合、新たな発言も「同意」である可能性が高い。返信が一方向的なものであった場合、それらの発言は「反論」である可能性がある。また、共通のツイートを RT した回数、共通の URL をツイートに含んだ回数、共通のツイートに返信した回数の3つは、第三者を媒介して表れる素性である。例えば、よく共通のツイートを RT しているユーザー同士での返信があった場合は、「同意」の可能性が高い。

以上、各素性の設計について述べたが、本研究において特に重要と考えているのは、反論表現辞書との一致度および構造的特徴の利用である。ここでいう構造的特徴とは、返信か引用かや、返信先のユーザとの間の関係についての各素性であり、これらは Twitter 上の返信構造により生じる素性である。また、反論表現辞書は、前述したように返信で用いられる固有表現を元に作成しており、これらもまた Twitter 上の返信構造により生じる素性と言える。次節の評価では、返信構造により生じるこれらの素性がどの程度性能に寄与するのかについても示す。

4.4 返信ツイートの態度分類の評価

本節では、前節で述べた返信の態度を分類する手法について性能を評価する。

4.4.1 交差検定による性能評価

前節でアノテーション済のデータセット B に対し、前述した4クラスに識別する多クラス分類を行う。本実験では、分類器として最大エントロピーモデルを用いる。分類器の実装として、Classias^{*4}の `pegasos.logistic` (L2 正則化ロジスティック回帰) を使用した。実験では、2クエリずつ計10個のデータに分割し、10分割交差検定を行った。表4.4に、クラス毎の精度、再現率、 F_1 値、全体の正答率を10分割のマイクロ平均を示した。 F_1 値の定義は以下の通りである。

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

各クラス毎の性能について見ると、「同意」と「疑問」に対してある程度高い分類性能が得られた一方で、「反論」と「その他」の分類性能は低くなっている。「その他」については、他の3クラスと異なりツイートの意図が広範囲にわたるため、共通の素性を得にくくなっているためと考えられる。素性のクラスに対する特定性の指標として、以下の数式を導入する。素性 f におけるクラス c_0 の特定性は、各クラス c の学習結果の重み w_{fc} に対して

$$c_0 \text{の特定性} = \frac{\exp(w_{fc_0})}{\exp(\sum_{c \neq c_0} w_{fc})} \quad (4.1)$$

^{*4} Classias のホームページ: <http://www.chokkan.org/software/classias/>

表 4.4. 返信ツイートの分類結果 (データセット B)

	<i>Precision</i>	<i>Recall</i>	F_1 値	<i>Accuracy</i>
同意	0.855(1389/1624)	0.841(1389/1651)	0.848	
反論	0.488(120/246)	0.456(120/263)	0.472	
疑問	0.766(239/312)	0.848(239/282)	0.805	
その他	0.537(273/508)	0.553(273/494)	0.545	
全体			0.667(Macro 平均)	0.751(2021/2690)

表 4.5. 参考：二人一致の返信ツイートの分類結果

	<i>Precision</i>	<i>Recall</i>	F_1 値	<i>Accuracy</i>
同意	0.778(1684/2166)	0.787(1404/1651)	0.782	
反論	0.483(172/356)	0.412(172/418)	0.444	
疑問	0.636(292/459)	0.755(292/387)	0.690	
その他	0.520(475/913)	0.501(475/949)	0.510	
全体			0.607(Macro 平均)	0.674(2623/3894)

で表す。データセット B に含まれる全てのデータで学習した場合に、「その他」を除く各クラス毎に特定性が上位の素性 5 個ずつを表 4.6 に示す。

「同意」と「疑問」の 2 クラスについては、それぞれを表すような特徴が比較的うまく抽出されていることが分かる。「同意」については、「！」・URL が 0 個である・感謝を表す表現などが有効な素性である。これらは、ツイートを観察したところ明らかに同意に多く見られ、組み込みやすい素性である。また、バイグラムのコサイン類似度の素性も上位に来ていることから、元のツイートと同じような内容を繰り返す場合にも同意である可能性が高いと考えられる。「疑問」については、疑問符や問いかけの表現が上位に来ている。これらもまた、ツイートを観察したところ明らかに疑問に多く見られ、組み込みやすい素性である。一方、「反論」については、反論表現との一致度が有用であるが、それらを除くとあまり反論に特有の単語とは思えない特徴が多く並んでおり、反論を示唆する表現は、n-gram などの単純な抽出法ではなかなか見つからないことが分かる。反論の認識に比較的有効であった素性としては、「返信相手への返信のみがある」、すなわち返信が一方通行であるという特徴が挙げられる。なお、反論に関してもう少し特定性の低いものを見ていくと、「お前」などの表現が見つかる。しかしながら、このような蔑称の類が登場する事例はさほど多くなかったので、反論に関する学習が十分に行えなかったと考えられる。

反論に関する学習を十分に行うための直接的なアプローチとしては、学習データを増やすことである。表 4.4 にはアノテーションが三人で一致したデータセット B の分類結果を示した

表 4.6. 各クラスで特定性の高い素性

同意		反論		疑問	
素性	特定性	素性	特定性	素性	特定性
!	0.257	(反論表現一致度)	0.642	?	3.874
(URL: 0 個)	0.246	(デマ否定単語あり)	0.229	か	0.213
(バイグラムの cos 類似度)	0.176	よ	0.166	です, か	0.162
ありがとう	0.154	デマ	0.156	?	0.129
ね	0.153	怒る	0.131	う, か	0.12
です, ね	0.138	そう, です	0.126	の, です	0.119
感謝	0.103	すぎる	0.120	は	0.115
すごい	0.100	だけ	0.115	の, ?	0.113
た。	0.097	あなた	0.113	本当	0.111
た	0.093	やる	0.111	真偽	0.102

が、比較のため、アノテーションが二人で一致したデータでの分類結果を表 4.5 に示す。全クラスで Precision・Recall・ F_1 値が低下していることが分かる。二人一致のデータは、三人一致のデータに比べて、アノテーターの判断を迷わせるツイートが多く含まれており、学習の精度を下げることとなってしまっている。学習データを増やすにあたっては、多くの人間が分類を試みてもクラスが一致するような、綺麗なデータを増やさなければならない。本研究では、再利用可能なコーパスとしてデータセット B を構築したが、綺麗な学習データを増やすことで性能向上を図ることはできるだろう。しかしながら、学習データを増やすには手間や時間がかかる上に、本研究で提案した 4 クラスの分類基準は絶対のものではない。例えば対象依存の評判分析を高精度なルールと共に組み込むなど、他の角度からの検討が必要である。本論文における以降での実験等は、全てデータセット B により構築された分類器を用いることとする。

4.4.2 各素性の有効性

本手法において特徴的なのは、反論に特有な表現を持つ事例を分類するための反論表現辞書と、そのような表現を持たない事例を分類するための構造的特徴を用いたことである。そこで、これらの有効性を評価するべく、以下の比較を行った。単語ユニグラム・単語バイグラム・元ツイートとのコサイン類似度・URL 数、ハッシュタグ数のみを特徴として使用する場合をベースライン (Base) とする。ベースラインにデマ表現と反論表現辞書との一致度を加えた場合を反論表現あり (+Con-Exp)、構造的特徴 (返信か引用か・共通の RT や URL 引用回数・相互の RT や返信回数・相互の RT や返信の方向性) を加えた場合を構造的特徴あり (+Structure)、反論表現辞書と構造的特徴を加えた場合を全使用 (+All) とする。これらの 4 つの場合において、それぞれ分類器を構築し、性能を比較したのが図 4.2 である。本研究で

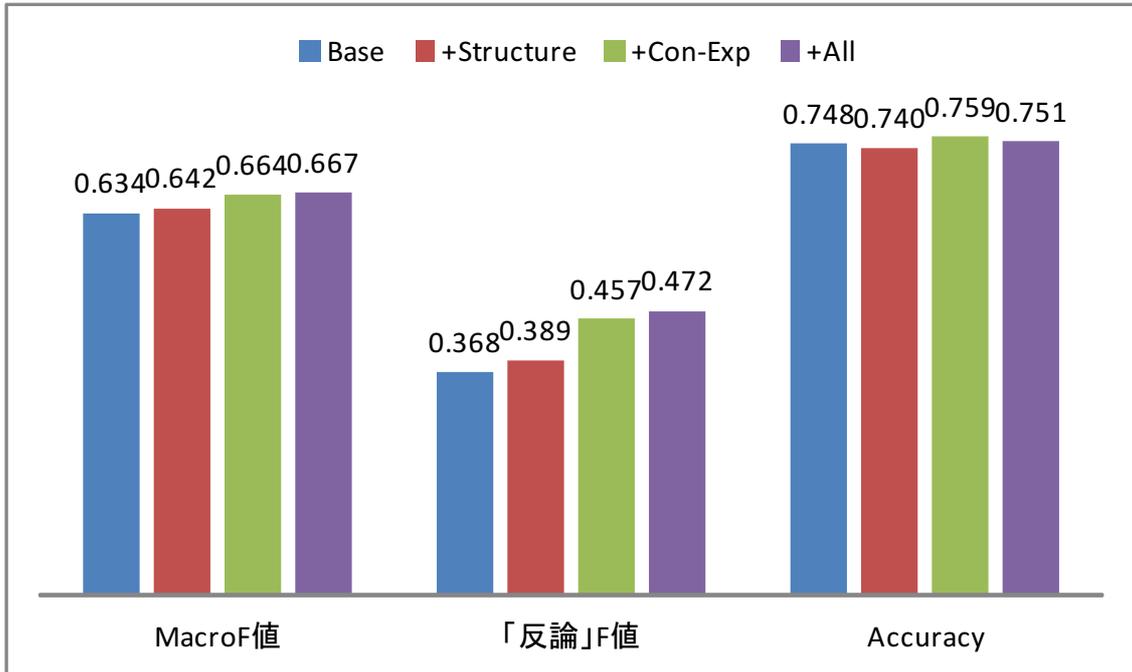


図 4.2. 使用する特徴を変更した時の性能変化

は、「反論」の認識が重要であると考え、マクロ F_1 値、「反論」クラスの F_1 値、正答率の3指標で性能を比較した。

まず、Base に対する +Con-Exp、および +Structure に対する +All の性能から、反論表現辞書は反論クラスの識別を中心として大きく寄与することが分かる。反論表現辞書を充実させることで、分類器の性能をさらに改善することができると考えられる。次に、Base に対する +Structure および +Con-Exp に対する +All の性能について、それぞれ「反論」 F_1 値が向上していることから、構造的特徴は反論クラスの識別性能を多少上げるのに貢献していると言える。一方で他クラスの識別の失敗例が増えることにより全体の Accuracy はやや低下しており、今後の検討が必要である。例えば、返信の回数・方向性については、分類対象のツイートに関連するようにトピックや時系列を限定することで、より正確にツイート間の関係に繋がる特徴となる可能性がある。

4.4.3 考察

本研究の目的であるツイート間の論述構造解析では、「反論」を高精度で識別することが重要である。そこで、反論の識別に失敗した事例を調査・分析した。ツイートで反論を行うパターンは、大きく3種類に分けられる。1つ目は、発言者への反論であり、2つ目は返信先の内容への反論である。そして3つ目は、発言者と返信先の内容の両方に対する、いわば複合的な反論である。これら3種類の反論の例を以下に示す。

1. @xxx 偉い学者さんなら、もっともらしい事言ってないで国に話し通すなり、非難勧告

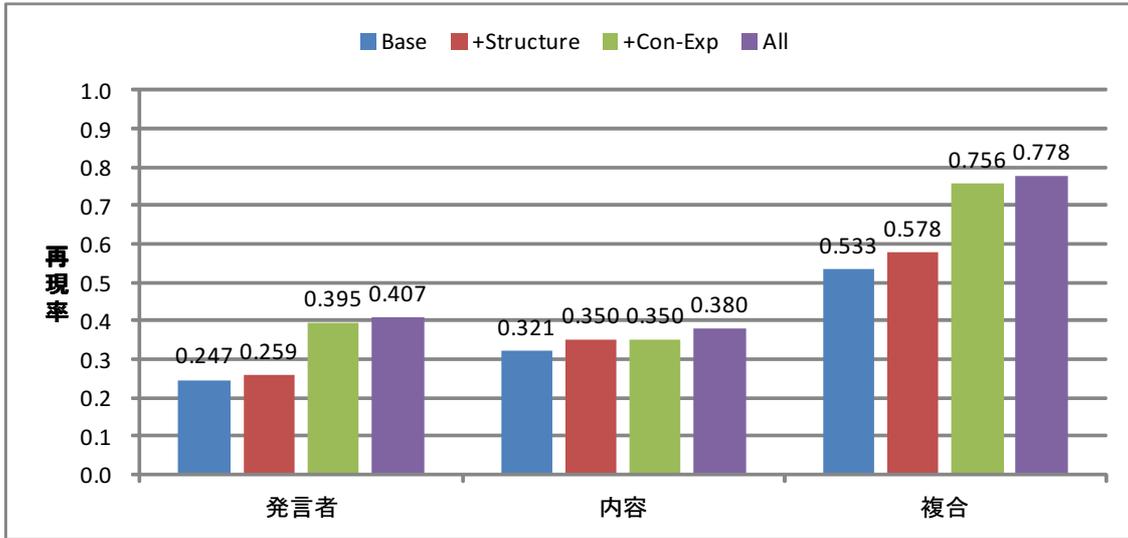


図 4.3. 反論の種類毎の再現率

- するなり、どうにかしろよ。なんにも出来ないなら不用意に被災者の不安煽る様な事言うな。発言者への反論
- 私が言うのも変ですが水で冷やしている限りメルトダウンはしません。問題が水が送れるか否かです。RT @xxx 水で冷やして「炉心熔融・メルトダウン」を停止できるという原子炉の専門家はいない。内容への反論
 - @xxx 防護ケース（格納容器のことか？）が万が一破裂しても压力容器があるので、即炉心が外界に露出ということではない。誤解を生む RT は控えて、RT 元は吟味していただきたい。複合的な反論

データセット B に含まれる 263 個の反論のうち、発言者への反論が 81 個、内容への反論が 137 個、複合的な反論が 45 個であった。図 4.2 における各特徴を使用した時毎の再現率を図 4.3 に示す。

発言者への反論については、返信先の内容に依らない表現が使われることが多いため、反論表現の抽出が有効である。上記(1)の例では、「不安煽る様な事言うな」という文は内容の影響を受けずに、反論を表すために広く用いられる。本研究では、集めた返信ツイートの中からこのような表現を集め、反論表現辞書とした。反論表現を持つという特徴を使用することで、分類の精度を上げることができる。次に、内容への反論については、返信先の内容によりツイートに含まれる表現は千差万別であり、内容を理解しなければうまく分類できない。この例では、返信先が「水で冷やしてもメルトダウンを防げない」という内容であることを理解し、「水で冷やしている限りメルトダウンしない」という主張が反論関係にあることを認識する必要があるが、このような事例を提案手法で分類することは難しい。

このような問題を解くためのアプローチとしては、返信先の主張の対象が「水で冷やしてもメルトダウンを防げない」であることを認識し、それに対し、返信の「水で冷やしている限り

「メルトダウンしない」という主張が対立関係にあることを認識しなければならない。これは、対象依存の感情分析 [25] や言論マップ [15] の矛盾認識で取り扱う事項である。しかしながら、Twitter では字数制限のある上にくだけた文体が多いため、評価極性の対象を同定するのは容易ではないため、本稿ではそのようなアプローチを取っていない。図 4.3 からは、構造的特徴の利用により、内容への反論の識別数が若干向上していることが読みとれる。構造的特徴の導入による解決できた事例としては、例えば以下のようなものがある。

- 教科書問題は救済措置が出たそうです！ RT @xxx: 子どもには何も落ち度はないのに？ RT @yyy 娘の学校に福島原発から避難してきた子が転校してきた。でも、新学期になってもその子には、教科書が配布されない。「転校ではなく、避難だから」。
- 東大 HP にて、大丈夫だそうです。 RT @xxx: RT @yyy: 前期の東大で合格した親戚が宮城で被災。入学に必要な書類の提出期限は 14 日の消印。状況が状況だけに延期してもらえないかと電話しても「規則だからダメ。#jishin #jisin

これらのツイートには使えそうな反論表現が含まれていないため、構造的特徴を中心として他の手段に頼るしかない。しかしながら、現状では“偶然”素性が発火して分類できているにすぎず、学習パラメータを少し変更するだけで分類できなくなる可能性も高く、改善は今後の課題である。現状では対象依存の評判分析からは成果が得られていないが、ディベートサイトを対象とした先行研究 [26] ではこのアプローチが最も大きな成果を挙げている。Twitter は、ディベートサイトと異なり賛成・反対が前提とされておらず、また書式も整っていないことが多いため、そのまま適用するのは難しいが、より深い言語処理との組合せが必要とされている。

最後に、複合的な反論については、基本的な特徴のみで比較的良好に識別できており、反論表現の利用によりさらに性能が向上している。これは、複合的な反論の中でも発言者への反論部分から特有の表現をとらえているためである。また、基本的な特徴のみでよく識別できている理由としては、複合的な反論はある程度の長さがあり、またしっかりとした文体で書かれていることが多く、有効な素性が抽出されやすいためと考えられる。

4.5 デマ検出への適用

返信態度の分類を直接的に用いた、デマ検出への応用を検討する。デマ検出は Twitter 上から機械的にデマを検出するものであり、多くの先行研究が存在する。関連研究の項でも述べたように、本研究の取る立場はあくまでもユーザーの信憑性判断を支援するという立場であり、デマ検出とは立場を異にするが、本章で行った返信態度の分類のデマ検出への応用性について検討する。

反論や疑問を多く集めるツイートは多くのユーザに受け入れられていない、すなわち信憑性が低いツイートだと考えられる。そこで以下の仮説 X を設定する。

仮説 X : 反論や疑問を多く集めたツイートはデマである

すなわち、反論や疑問が一定数、あるいは一定の割合存在するツイートを見つけることで、デ

マの可能性が高いことを示唆することができるかと仮定する．この仮説を検証するため，以下の実験を行った．まず，震災時に Twitter 上で出回ったデマとして以下の 11 個のトピックを選ぶ．

1. イソジンを飲むと放射能対策になる
2. コスモ石油の火災で有害物質の雨が降る
3. 花山村が孤立している
4. 関西電力が節電を呼びかけている
5. 福島から避難した子供に教科書が配布されない
6. 救援物資の空中投下は法律で禁止されている
7. 埼玉の水道水に異物が混入し危険
8. 東大が入学を取り消し
9. 双葉病院の医師が患者を置き去りにした
10. 谷垣総裁が入閣要請を断った
11. 茨城県知事が災害派遣要請をしていない

これらのトピックは命題型，すなわち「...である」の形式で表すことができる．各トピックについて，リツイート数上位のツイートを集める．それらのツイートに対し，返信態度の分類器を用いて抽出される反論・疑問の数と割合を調べる．リツイート数が多いツイートを集めるのは，デマは多くリツイートされることによって拡散されるためである [16]．ただし，これらのトピックを持つツイートを集めるのは容易ではない．本研究では，ツイートのトピックを決定する手法については範囲外としている．したがって，各トピックに対応する収集用のクエリを用いて，リツイート数が上位 30 件のツイートを集めた後，各トピックに該当するかどうかを判断し，該当するものツイートだけを残す．トピックに該当するかどうかとは，例えば (1) の「イソジンを飲むと放射能対策になる」のトピックについては，イソジンを飲むと放射能対策になるかどうか，についてツイート内で触れられているものを該当するツイートとみなす．この際，放射能対策になると言っているかならないと言っているかはどちらでもよく，また放射能対策になるかよく分からない，といった内容のものでもよい．ただし，イソジンでうがいをすると放射能対策になる，のようなツイートは別トピックと考え除外する．全トピックについてこの作業を行った結果，各トピックについて 16~30 個のツイートが残った．これらのツイートに対し返信しているツイートを集め，構築した分類器を適用する．抽出した反論および疑問を，命題の真偽毎に分類して集計する．たとえばトピック (1) に関しては，イソジンを飲むことで放射能対策になる，という立場のツイートと，放射能対策にならないという立場のツイートをそれぞれ集め，各グループ内で反論および疑問の数と割合を集計する．全トピックについての集計結果を図 4.4 に示す．ただし，凡例の incredible はデマである側（トピック (1) であれば，イソジンを飲んでも放射能対策にならないという立場）を指し，credible はデマでない側を指す．

図 4.4 から分かるように，反論・疑問の数や割合が多いと分類器が出力したかどうかだけでは，デマを検知するのは難しいと見られるが，反論・疑問の数や割合が極端に多いツイート集

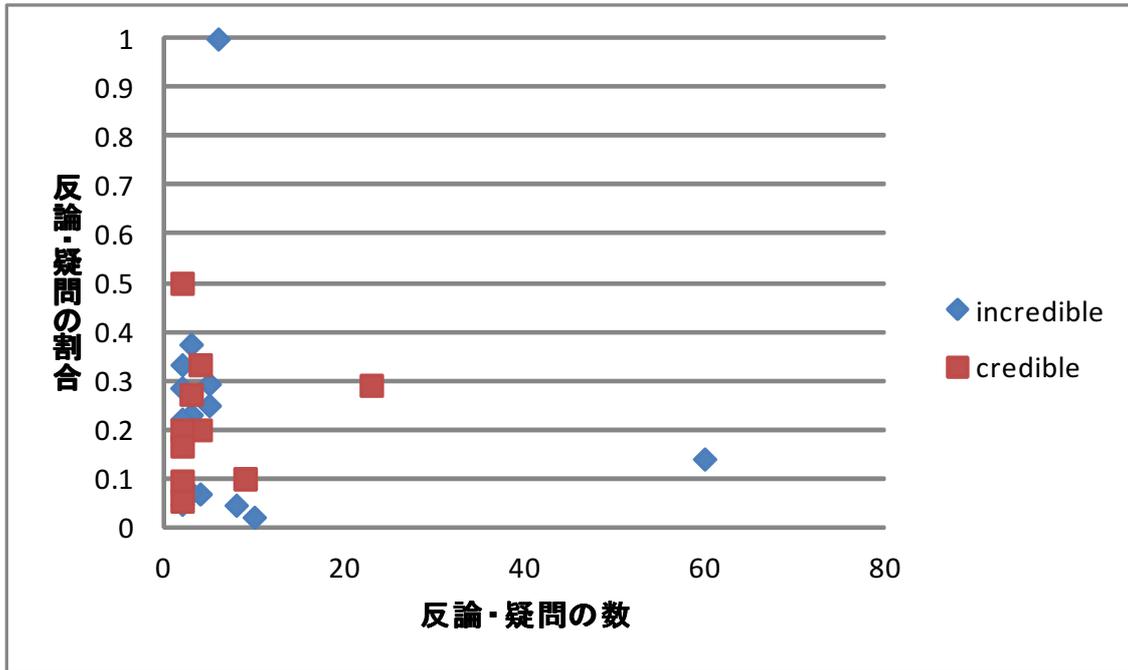


図 4.4. 反論・疑問の数と割合

合は、やはりデマ側になっていることが分かる。したがって、分類器の精度を上げることで、直接デマ検知に繋げることも可能と考えられるが、現状の性能では困難である。

4.6 まとめ

本章では、返信態度の分類について述べた。具体的には、返信態度を「同意」「反論」「疑問」「その他」の4クラスに分類する分類器を教師あり学習を用いて構築した。それに伴い、コーパス(データセットB)や反論表現辞書などの資源を作ることができた。分類性能は、全体の反論の F_1 値で 0.472, 4 クラスの F_1 値の Macro 平均で 0.667, Accuracy で 0.751 であった。構築した返信態度の分類器の直接的なデマ検知への適用に関しても検討したが、現状の性能では難しく、ユーザ支援には別の利用法を考えなければならない。

第 5 章

一般のツイート間関係認識

5.1 はじめに

本節では、返信に限らない全てのツイート間関係認識を行うことを考える。その際、前節までで述べた返信の態度分類を、直接的に返信関係のない一般のツイート間の論述関係分析に応用する手法について述べる。一般に、返信のツイートが全ツイートに占める割合は非常に少ないため、前章で作成した分類器をそのまま適用するのは難しい。特に、反論表現辞書の有効性は低いと考えられる。反論表現辞書に含まれる表現は、返信先への態度を表す際に利用されることの多い表現であり、返信になっていない一般のツイートに含まれる可能性は低い。

返信の関係のない一般のツイート間に対して 4 クラスの分類を行うには、その言語的な内容を利用することが考えられる。例えば、ツイート間の反論関係は、厳密に同じものではないものの、含意関係認識課題 [22] における文間の矛盾関係 (RTE3 [28]) に相当する。文間の矛盾関係の認識に取り組んだ Marneffe ら [29] は、RTE3 のテストデータに対して精度 22.95%、再現率 19.44% を達成した。また、日本語を対象とした場合、NTCIR9-RITE [30] および NTCIR10-RITE2 においては、矛盾関係を含む含意認識課題である MC (multi class) タスクが取り扱われている。MC サブタスクは、二つのテキストペアを与えた時に、含意関係あり (順方向・逆方向・双方向) と含意関係なし (矛盾・独立) の 5 クラスに分類する問題である。前章で示した 4 クラスと正確に対応するわけではないが、含意見解ありは「同意」と、矛盾は「反論」と、独立は「その他」と近いものと考えてよい。RITE2 のフォーマルランにおける矛盾関係認識性能は、精度 52.17%、再現率 19.67% が最高性能であった*1。これらの認識性能が示す通り、言語的な内容に基づく文間の矛盾関係の認識は容易ではない。

そこで本章では、返信態度の分類をツイート間関係分析に利用する手法について述べる。具体的には、ツイートペアが与えられた時に、両ツイートおよび両ユーザー、および関連するツイート・ユーザーに拡張したネットワークを作り、ネットワーク内に存在する返信関係を分類器で解く。その結果から、元のツイートペアの間関係を推測する。また、含意関係認識をカスタマイズしてツイート間関係認識を行えるようにし、提案手法との性能比較を行う。な

*1 <http://www.cl.ecei.tohoku.ac.jp/rite2/doku.php?id=wiki:results>

お、本章の内容は東北大学の乾・岡崎研究室^{*2}と共同で行われており、含意関係認識をカスタマイズしたツイート間関係認識分類器については、かの研究室の水野氏より提供を受けた。

5.2 問題設定

一般のツイートの意図は広範囲に渡る。意見の主張に始まり、ニュースなどの拡散、さらに他愛もないつぶやきや独り言の類も多い。これら全てのツイートを整理することはできない。したがって、整理が完了した際の有用性を考慮し維持しながら範囲を狭めることが必要である。そこで本稿では、ある特定の“トピック”に関するツイート群を集め、それらを整理するタスクを考える。トピックとは、そのツイートが属する何らかの話題である。ツイートにトピックを与えれば、同じトピックに属するツイート間には何らかの関係を付与することが可能となる。ツイートのトピックを決定する手法については本稿の手法の対象外であり、人間の手で行うものとする。

本節では、同じトピックに属するツイートのペアが与えられた場合に、関係を分類するという問題を解く。関係としては、返信の関係分類の際と同じく、「同意」・「反論」・「疑問」・「その他」の四つとする。

同意 同様の主張や感情を示すなど、明確な同意の意図が感じられるもの。

反論 対立する主張や感情を示すなど、明確な反論の意図が感じられるもの。

疑問 情報・情報源の要求や、疑問の吐露など。

その他 上記のどれにも分類できないもの。

これらの分類基準は、返信の場合と似ているが全く同じにはならない。前述したように、ツイートペアの関係性は、トピックに関する主張同士の関係性としてとらえる。返信の関係が存在しないツイートペアにおいては、お互いに対する単純な賛成や反対、感情的な同調や反発などは表れにくい。また、お互いに対する直接的な質問もまず存在しないため、「疑問」はかなり出現頻度が少ないと予想されるが、存在した場合には重要であると考え、クラスは設定しておく。このようにして、返信の関係にないツイートペアに対しても、表向きは同様のクラスを考えることとする。なお、ツイートペアのトピックが明らかに異なる場合についても、この 4 クラスに分類しようとするれば「その他」のラベルを付けることは可能である。しかし、「その他」のツイートは現状ではツイート空間の整理の際には用いないため、ツイートペアのトピックが明らかに異なる事例は最初から除外している。

5.3 同トピックツイートペアのアノテーション

上記のタスクを行うにあたり、正解データを準備する。データは hotto コーパスより集める。まず、4.5 節でも使用した 11 個のトピックを用意する。これらのトピックは返信の分類

^{*2} <http://www.cl.ecei.tohoku.ac.jp/>

表 5.1. 一般ツイートペア分類実験の 11 トピック

トピック	収集用クエリ	対応ツイート数
イソジンを飲むと放射能対策になる	イソジン	27
コスモ石油の火災で有害物質の雨が降る	コスモ石油	26
花山村が孤立している	花山村	29
関西電力が節電を呼びかけている	関西電力	21
福島から避難した子供に教科書が配布されない	教科書 福島	22
救援物資の空中投下は法律で禁止されている	空中投下	17
埼玉の水道水に異物が混入し危険	埼玉 水道水	16
東大が入学を取り消し	東大 入学	19
双葉病院の医師が患者を置き去りにした	双葉病院	19
谷垣総裁が入閣要請を断った	谷垣 入閣要請	30
茨城県知事が災害派遣要請をしていない	茨城 県知事 災害派遣要請	30

のために集めたときのものの中から設定した．対応するクエリを用いて検索されたツイートのうち，RT 数が上位のものから 30 件ずつ集める．RT 数上位のものから順に集めるのは，それらのツイートは多くのユーザーの目に触れるものであり，論述構造を分析・整理する意義が大きいためである．クエリは，該当トピックをうまく集められそうなものを選んでいますが，中にはトピックに関連するツイートだけを集めるようなクエリの調整が難しく，実際にトピックに対応するツイートの数は 30 未満になるものもある．トピック・クエリ・トピックに対応するツイート数を表 5.1 に示す．

これらの 11 トピックに対し，ペアワイズな全てのツイートの組み合わせを考える．例えば，「イソジンを飲むと放射能対策になる」のトピックに関しては，27 個のツイートが存在するので，その組み合わせは全部で 351 個ある．これらの組み合わせについて，前述した 4 クラスのラベル付けを行う．ラベル付けは，アノテーション経験の豊富な一人のアノテーターの手により行った．アノテーションの結果を表 5.2 に示す．「同意」の割合が高く，予想した通り「疑問」の割合は非常に少ないことが分かる．また，このアノテーション済のデータを，以降データセット C と呼ぶ．

5.4 提案手法

同じトピックの中でのツイートペアの関係分類を，返信関係分類器を用いて行う手法について述べる．大まかな方針としては，直接の返信の関係が見当たらない場合，ツイートやユーザーからなるネットワークを拡張し，返信の関係を探す．前節で述べた反論表現辞書などを用いることで，返信関係にないツイートペアに比べ格段に関係が推測しやすい．そして，返信分

表 5.2. トピック毎のアノテーション結果

トピック	同意	反論	疑問	その他	トピック毎の合計
イソジンを飲むと放射能対策になる	206	130	1	14	351
コスモ石油の火災で有害物質の雨が降る	254	68	0	3	325
花山村が孤立している	291	115	0	0	406
関西電力が節電を呼びかけている	86	101	0	23	210
福島から避難した子供に教科書が配布されない	133	66	32	0	231
救援物資の空中投下は法律で禁止されている	66	70	0	0	136
埼玉の水道水に異物が混入し危険	61	45	0	14	120
東大が入学を取り消し	87	81	3	0	171
双葉病院の医師が患者を置き去りにした	111	28	0	32	171
谷垣総裁が入閣要請を断った	299	116	9	11	435
茨城県知事が災害派遣要請をしていない	325	86	24	0	435
全トピックの合計	1919	906	69	97	2991

類器の出力の結果を元のツイートペア間の関係の推測に用いることとする。以下、ツイートペア a と b の関係を求めるために、直接返信の関係になっている場合を含め、ネットワークの拡張により返信分類器を適用する 4 通りのパターンについて図で説明し、そのパターンにおける各クラスのスコアの式を示す。ネットワークの拡張は、ツイートやユーザを周囲に広げることにより様々な経路で元のツイートペア a と b を結び、その経路内の返信関係に分類器を適用し、各クラスのスコアを得るものである。なお、図のリンクについている名称は関係を表している。リンク関係には「author」「RT」「返信」の三種類がある。「author」はツイートとそのツイートの投稿者の間に、「RT」はツイートとそのツイートをリツイートしたユーザ、返信はツイートとそのツイートに返信しているユーザの間に成り立つ関係である。また、各パターン p で出力されるクラス c のスコアを $Score_{p,c}$ とすると、最終的に出力される ab 間の関係クラス c_{ab} は、

$$c_{ab} = \operatorname{argmax}_c \sum_p Score_{p,c}$$

となる。

1. 直接返信の関係になっている場合 (図 5.1)。返信に分類器を適用し、ロジスティック回帰により算出される各クラス c の確率値 q_c を用いて、

$$Score_{Direct,c} = q_c$$

となる。

2. 両ツイートをリツイートしているユーザがいる場合 (図 5.2)。本論文では、リツイート

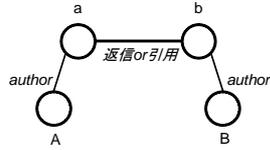


図 5.1. 直接返信のパターン

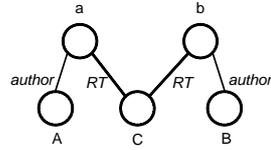


図 5.2. 両方をリツイートするパターン

は同意であるという仮定を置くことにする．しかしながら，反論の意図を持ってリツイートしている場合もあるため，片方をリツイートしたユーザ数に対し，両方をリツイートしたユーザ数の割合が大きいときほど同意とみなしやすと考え， a をリツイートしたユーザ集合と b をリツイートしたユーザ集合の間の Jaccard 係数

$$Jaccard(a, b) = \frac{|RT_a \cap RT_b|}{|RT_a \cup RT_b|}$$

を同意のスコアとする．ただし， RT_a と RT_b はそれぞれ a と b をリツイートしたユーザ集合である．なお， $Jaccard(a, b)$ が小さい場合は同意と判断する根拠が少ないと考え，同意スコアを出力しない．具体的には 0.01 未満であれば 0 とする．

$$Score_{BothRT-c} = \begin{cases} Jaccard(a, b) & \text{if } c = \text{同意 and } Jaccard(a, b) \geq 0.01 \\ 0 & \text{if } c \neq \text{同意 or } Jaccard(a, b) < 0.01 \end{cases}$$

である．

- 片方のツイートの投稿者が，もう片方のツイートに対し他のツイートで返信している場合（図 5.3）．このケースでは， a - b 間の関係を a 1- b 間が表していると考え，スコアには，ロジスティック回帰により求められた各クラス c の確率値 q_c を用いて

$$Score_{OtherRep-c} = q_c$$

とする．

- 片方をリツイートしてもう片方に返信しているユーザがいる場合を考える（図 5.4）．このケースでは，リツイートは同意であるという仮定に基づき， a - b 間の関係を c 1- b 間が表していると考え，同様にロジスティック回帰により求められた各クラス c の確率値 q_c を用いて

$$Score_{RTandRep-c} = q_c$$

とする．

以上のようなルールは，要求に応じて拡張が可能である．例えば， a をリツイートしたユーザ C と b をリツイートしたユーザ D の間に返信関係があった場合に，分類器によるその関係の出力を用いることができる．ただし，ルールを増やすほど精度は下がると考えられる．例えば，両方のツイートに返信しているユーザがいる場合（図 5.5）などが考えられるが，このケースは返信分類器を適用すべき箇所が 2 箇所存在する．それぞれによって得られた結果からスコアを算出することは可能ではある．例えば，二つの返信に返信分類器を適用した結果，片方が

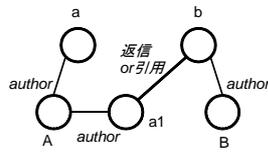


図 5.3. 他のツイートで返信のパターン

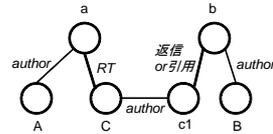


図 5.4. 片方をリツイート・もう片方に返信のパターン

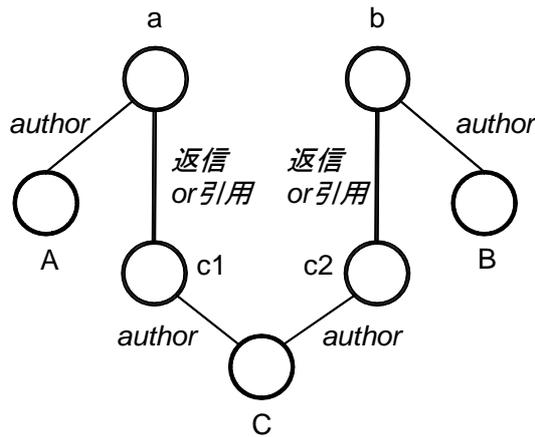


図 5.5. 両方を返信のパターン

同意でもう片方が反論と出力された場合， ab 間の関係としては反論と出力されるべきである．したがって，同意と出力されたほうについて，同意と反論のスコアを入れ替え，もう片方のスコアとの和を取るなどが考えられる．しかしながら，現状の分類器では，適用箇所が 2 箇所以上に増えた場合に精度を保つのは困難であると考えられる．そのため本論文では使用しなかったが，今後の検討課題である．

5.5 文間関係認識に基づく分類手法

本節では，2 文を与えたときにその間の意味的关系を返す文間関係認識器を用いてツイート間の関係を分類する手法について述べる．なお，この文間関係認識器は，前述した通り東北大学の乾・岡崎研究室の水野氏より提供された．そのため，詳細については本論文に必要な範囲のみに留める．文間関係認識の部分課題に位置づけられる含意関係認識は，前述の通りワークショップが開催されるなど広く研究されている．含意関係認識課題で取り扱われている関係の種類を以下に示す．

含意 一方の文 (T) を真としたとき，他方の文 (H) も真であると判断できる

T 川端康成は「雪国」などの作品でノーベル文学賞を受賞した

H 川端康成「雪国」の著者である

矛盾 2 つの文が同時に成立し得ない

- T* モーツァルトはザルツブルグで生まれた
H モーツァルトはウィーンで生まれた
 その他 上記以外
T 川端康成は小説家である
H 川端康成は大阪で生まれた

本稿で取り扱う 4 つの関係とは、含意が同意に、矛盾が反論に対応するが、同意は厳密な含意ではなく、反論は矛盾しているとは限らない。そこで、水野ら [15] の同意、対立関係を対象とした文間関係認識器を用いる。彼らの定義する同意、対立関係は、本稿で定義する同意、反論関係とほぼ対応する。疑問関係は取り扱えないため、同意、反論、その他の 3 種類の関係に分類する。

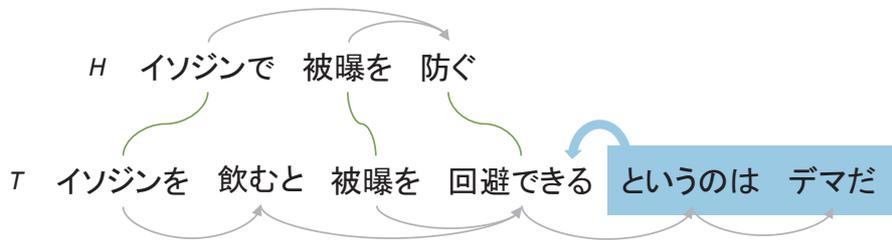


図 5.6. 文間関係認識例

まず、図 5.6 を用いて、彼らの手法を簡潔に述べる。次に、本課題のために変更した点について述べる。

言語解析 入力となる 2 文それぞれに対して、形態素解析 [27]，係り受け解析 [31]，述語項構造解析 [32]，拡張モダリティ解析 [33]，評価極性判定を行う。図 5.6 では、係り受け構造を文の上下の矢印で示す。

仮説対応部分の同定 テキスト中には、仮説の内容と関連の低い情報も含まれている。そこで、まず、テキスト中で仮説文と内容的に対応する部分を同定する。次に、対応する部分の関係に基づいて文間の関係を同定する。これにより、後段の関係分類では、対応する名詞間や述語間の関係を考慮することで、文間の関係を同定することができる。

テキスト中で仮説に対応する部分の同定は、文節アライメント、局所構造アライメント、文節アライメントの選択という 3 段階で行われる。以下にそれぞれの手続きを示す。

文節アライメント 文節中に含まれる単語間の類似・関連性に基づいて文節間に文節アライメントを付与する。文節中の内容語が類似しているとき、それを含む文節単位でアライメントされる。図 5.6 において、「イソジンで (*H*)」と「イソジンを (*T*)」は、イソジンという内容語が共通することからアライメントされる。

形態素の類似・関連性は、その表層に基づいて判断されるだけでなく、日本語 WordNet [34] や動詞含意関係データベース [35] を用いて、意味的な類似度に基づいて判断される。図 5.6 において、「防ぐ (*H*)」と「回避できる (*T*)」は、意味的に類似しているためアライメントされる。

表 5.3. 否定表現

誤り	ねつ造	誇張
嘘	誤報	デマ
今のところない	ソースがありません	事実はありません
根も葉もない噂	根拠がない	ニュースはない

局所構造アライメント 文節アライメントによってアライメントされた文節の中には、文全体の意味を考えるとアライメントすべきではない場合がある。例えば、「イソジン飲んで被曝を防ぐ」と「ワカメの味噌汁を飲むと良い」という2文を考える。2文間で「飲む」が共通しているが、その対象は異なっているため、アライメントすべきではない。

この問題に対して、文中の依存構造および述語項構造を対応付けるのが局所構造アライメントである。図 5.6 において、文節アライメントされる「被曝を (H) -^{*3} 被曝を (T)」と、「防ぐ (H) - 回避できる (T)」に着目する。これらは H 側、 T 側のいずれにも依存構造が存在し、この構造間にアライメントを付与するのが局所構造アライメントである。「イソジンで (H) - イソジンを (T)」と「防ぐ (H) - 回避できる (T)」は、 H 側には依存構造が存在するが、 T 側には直接の依存関係は存在しない。しかしながら、「飲むと (T)」を介して依存構造の繋がりが存在する。本研究では水野らと同様に、4 文節まで介しても良いという上限を設けた。

文節アライメントの選択 文節アライメントされた文節対のうち、局所構造アライメントもされている文節対のみを選択する。

関係分類 アライメント結果を入力として、まず、同意および対立と、その他との分類を行い、次に、同意と対立の分類を行う。1 つめの分類は、仮説側の文節が全て他方に対応付けられたかによって判断される。2 つめの分類について、対応付けられた述語が、否定の関係にあるか、評価極性が異なるか、反義の関係にある場合、対立に分類し、いずれにも当てはまらないものは同意に分類される。

本研究で対象としている、震災時のツイッターデータには、水野らに対立関係分類の手がかりとしている述語間の否定、評価極性、反義だけでなく、「～というのはデマです」や「～という事実はありません」といった文の後方での否定・反論が少なくない。そこで、このような表現を否定表現と呼び、対応付けられた述語よりも後方に否定表現が現われた場合、同意ならば対立に、対立ならば同意に関係を反転させる工程を付け加えた。表 5.3 に利用した否定表現を示す。これらは、評価データ以外のツイートから収集した。表 4.3 に示される反論表現とは、一般ツイート中の表現である点異なるが、重なる表現も存在する。図 5.6 において、 T 側の「回避できる」は、その後方の「～というのはデマだ」という表現によって否定されている。

*3 「-」で前後の単語が文節アライメントされることを示す

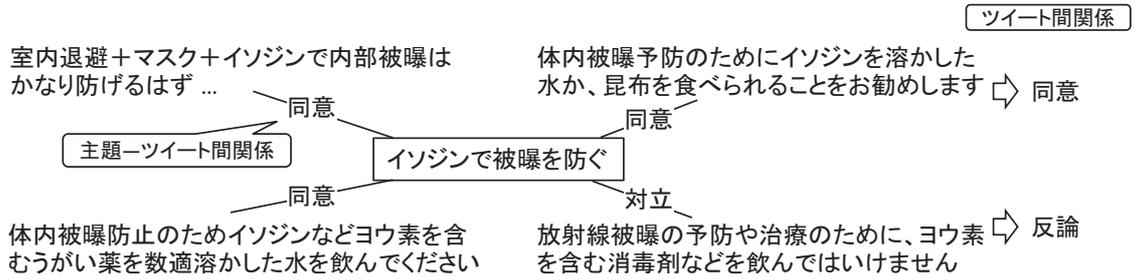


図 5.7. 文間関係認識手法 1

彼らの文間関係認識器は、仮説が単文（一つの述語といくつかの項からなる文）であることを前提としている．しかし、ツイートに含まれる文は、単文であることは少ない．そこで、トピックごとに単文を用意し、単文とツイート間の関係を求めることで、ツイート間の関係を求める．図 5.7 に、この主題文を用いてツイート間の関係を求める手法を示す．まず、トピックごとにそのツイート集合を代表する主題となる文を作文する．例えば、「イソジン」というトピックに対して、関連するツイート集合から、「イソジンで被曝を防ぐ」という主題文を入手で作文する．次に、関係を求めたい 2 つのツイートに対して、ツイートと主題文との間の関係を求める．その際、ツイートは、まず文分割され、次にトピック中の内容語（トピック単語）が全て含まれる文のみが選択される．そのような文が複数含まれている場合は、一番後ろの一文を選択する．図上部の例はいずれも同意で、図下部の例は、一方が同意で、他方が対立に分類されている．最後に、求められた 2 つの関係から、ツイート間の関係を求める．主題とツイートとの間の関係から導かれるツイート間の関係の組み合わせを、以下に示す．

1. 2 つの関係のうち 1 つでも「その他」⇒「その他」
2. 2 つの関係がいずれも「同意」⇒「同意」
3. 2 つの関係がいずれも「対立」⇒「同意」
4. 2 つの関係が異なる、すなわち「同意・対立」または、「対立・同意」⇒「反論」

以降、本手法を「文間関係認識手法 1」と呼ぶ．

ツイート間の関係を直接同定する手法も評価する．まず、文間関係認識手法 1 と同様に、文分割およびトピック単語を含む文選択を行う．次に、関係を求めたい任意の 2 ツイートについて、選択された 2 文間の関係を求める．同定された文間関係は、同意はそのまま同意関係に対応し、対立は反論関係に対応する．以降、本手法を「文間関係認識手法 2」と呼ぶ．

5.6 一般のツイート間関係分類の評価

本節では、前節で述べた一般のツイート間の関係認識手法の評価を行う．本研究の目的は、図 1.1 のように、ツイート間の論述関係を可視化し、ユーザに情報の「裏」を提示し、情報の価値判断を支援することである．誤分類事例が多いと、正しく情報の裏を取ることができないため、高い精度を実現することが重要であり、その上で再現率を上げていく戦略をとる．本論

文では，論述関係の可視化を実現するために必要な程度の数のツイートに対して，正しく関係分類が行えたかに着目して評価する．

5.6.1 各トピックにおける精度評価

正解の関係ラベルを付与したデータセット C の各ツイートペアについて，関係分類実験を行った．提案手法については，図 5.1～図 5.4 に示すいずれのパターンにも当てはまらない場合は，関係を出力しない．

表 5.4・表 5.5・表 5.6 にトピック毎の分類結果と全トピックのマクロ平均を示す．トピック毎の分類結果は，*Precision*，*Recall*， F_1 値を同意・反論それぞれについて示している．また，紙面の都合上，トピックは省略した形で書いているが，実体は表 5.1 に示している通り命題型のトピックである．

表 5.4. 提案手法による分類

トピック	P (同)	P (反)	R (同)	R (反)	F_1 (同)	F_1 (反)
イソジン	0.600(6/10)	0.750(3/4)	0.029(6/206)	0.023(3/130)	0.056	0.045
コスモ石油	0.850(17/20)	0.000(0/3)	0.067(17/254)	0.000(0/68)	0.124	0.000
花山村	0.600(9/15)	0.833(5/6)	0.031(9/291)	0.043(5/115)	0.059	0.083
関西電力	0.615(16/26)	0.000(0/1)	0.186(16/86)	0.000(0/101)	0.286	0.000
教科書	0.722(13/18)	0.000(0/1)	0.098(13/133)	0.000(0/66)	0.172	0.000
空中投下	0.571(12/21)	0.667(2/3)	0.182(12/66)	0.029(2/70)	0.276	0.055
水道水	0.600(3/5)	N/A	0.049(3/61)	0.000(0/45)	0.091	0.000
東大	0.400(2/5)	N/A	0.023(2/87)	0.000(0/81)	0.043	0.000
双葉病院	0.750(6/8)	N/A	0.054(6/111)	0.000(0/28)	0.101	0.000
谷垣総裁	0.650(104/160)	N/A	0.348(104/299)	0.000(0/116)	0.453	0.000
茨城県知事	0.900(9/10)	N/A	0.028(9/325)	0.000(0/86)	0.054	0.000
合計	0.660	0.375	0.099	0.009	0.156	0.017

「反論」の分類精度を見ると，提案手法は文間関係認識 1 を大きく，また文間関係認識 2 をわずかに上回っており，反論の識別に有効である．しかしながら，再現率に関しては，提案手法は文間関係認識手法に比べて低い値となっており，その結果 F_1 値も低くなっている．この理由として，文間関係認識手法は言語処理に基づく手法であり，似たツイートに対しては同じ結果を出力することができる．一方，提案手法はネットワーク拡張と返信分類器の適用からなる手法であり，似たツイートがあったとしてもネットワークを拡張した場合に同等の関係を見つけれらるとは限らないため，文間関係認識手法のように同じ結果を出力できるとは限らない．本研究では特別に処理を行っていないが，いずれ目指す最終的な出力を想定すれば，クラ

表 5.5. 文間関係認識手法 1 による分類

トピック	$P(\text{同})$	$P(\text{反})$	$R(\text{同})$	$R(\text{反})$	$F_1(\text{同})$	$F_1(\text{反})$
イソジン	0.581(18/31)	0.343(12/35)	0.087(18/206)	0.092(12/130)	0.152	0.145
コスモ石油	1.000(21/21)	N/A	0.083(21/254)	0.000(0/68)	0.153	0.000
花山村	1.000(254/254)	0.000(0/1)	0.873(254/291)	0.000(0/115)	0.932	0.000
関西電力	1.000(3/3)	N/A	0.035(3/86)	0.000(0/101)	0.067	0.000
教科書	N/A	N/A	0.000(0/133)	0.000(0/66)	0.000	0.000
空中投下	0.500(2/4)	0.000(0/8)	0.030(2/66)	0.000(0/70)	0.057	0.000
水道水	1.000(1/1)	N/A	0.016(1/61)	0.000(0/45)	0.032	0.000
東大	N/A	N/A	0.000(0/87)	0.000(0/81)	0.000	0.000
双葉病院	0.900(18/20)	0.250(1/4)	0.162(18/111)	0.036(1/28)	0.275	0.063
谷垣総裁	N/A	N/A	0.000(0/299)	0.000(0/116)	0.000	0.000
茨城県知事	0.858(187/218)	0.148(25/189)	0.575(187/325)	0.291(25/86)	0.689	0.196
合計	0.855	0.148	0.169	0.038	0.214	0.037

表 5.6. 文間関係認識手法 2 による分類

トピック	$P(\text{同})$	$P(\text{反})$	$R(\text{同})$	$R(\text{反})$	$F_1(\text{同})$	$F_1(\text{反})$
イソジン	1.000(5/5)	0.250(2/8)	0.024(5/206)	0.015(2/130)	0.047	0.029
コスモ石油	1.000(1/1)	0.000(0/5)	0.004(1/254)	0.000(0/68)	0.008	0.000
花山村	1.000(236/236)	1.000(24/24)	0.811(236/291)	0.209(24/115)	0.896	0.345
関西電力	1.000(4/4)	N/A	0.047(4/86)	0.000(0/101)	0.089	0.000
教科書	1.000(1/1)	N/A	0.008(1/133)	0.000(0/66)	0.015	0.000
空中投下	1.000(55/55)	1.000(1/1)	0.833(55/66)	0.014(1/70)	0.909	0.028
水道水	1.000(1/1)	0.000(0/1)	0.016(1/61)	0.000(0/45)	0.032	0.000
東大	1.000(3/3)	0.000(0/5)	0.034(3/87)	0.000(0/81)	0.067	0.000
双葉病院	1.000(45/45)	0.000(0/8)	0.405(45/111)	0.000(0/28)	0.577	0.000
谷垣総裁	1.000(2/2)	0.000(0/1)	0.007(2/299)	0.000(0/116)	0.013	0.000
茨城県知事	0.989(222/229)	1.000(2/2)	0.683(222/325)	0.023(2/86)	0.808	0.045
合計	0.999	0.361	0.261	0.024	0.315	0.041

表 5.7. 拡張パターンによる認識の精度

	同意	反論	合計
パターン 1	12/17(0.706)	1/2(0.500)	13/19(0.684)
パターン 2	97/154(0.630)	9/16(0.563)	106/170(0.624)
合計	109/171(0.637)	10/18(0.556)	119/189(0.630)

スタリングなどの処置が妥当である。

ネットワーク拡張手法は、簡単なパターンの追加だけで、分類された事例は少ないものの、直接の返信の関係にないツイート間の関係分類を行うことができた。文間関係認識に基づく手法について、一部のトピックについては、多くの事例を高精度で分類できている。いずれの手法でも分類された事例数の少なかったトピックとして、「東大入学取り消し」、「埼玉の水道水が危険」などがある。

5.6.2 拡張パターンの有効性

ネットワーク拡張手法のポイントは、同じ態度を仮定できる箇所をネットワーク上で見つけ、返信の分類器を適用することであった。該当するパターンとしては図 5.3 と図??である。まず、これらのパターンによって同意もしくは反論と認識されたツイートペアについて、その精度を表 5.7 に示す。

パターン間の精度に大きな差異は見られなかったが、パターン 2 のほうが抽出できる数は圧倒的に多いことが分かる。また、反論について 0.5 を超える精度が得られている。同じ態度を仮定し返信分類器を適用することで、現状の分類器の性能でも反論をある程度当てることができる。ただし、同意と合わせた合計正解率は 0.630 であり、これは全部同意と判定した時のチャンスレートとほぼ等しいため、十分とは言えない。

5.6.3 考察

ネットワーク拡張手法の問題点は、ほとんどのトピックについて、分類された事例の数が少ないことである。この問題の解決策として、パターンの拡張、データの増加が挙げられる。前者は、ネットワークをより広範囲にわたって探索して、関係を探してくるようによって、分類可能事例を増やす。ただし、ノイズの多いパターンを採用することで精度は下がる可能性が高い。例えば、RT は同意であるという仮定は常に成り立つわけではない。また、元のツイート間のパスが長くなるほど、そのリンク関係の信用度も下がる。本研究で目指すツイート空間の整理は、大量のツイートをまとめるよりは、少数でも重要なツイートを高精度で分類、整理することが有効であると考えているので、ルールの拡張は最適点を見つける問題に帰着することになる。一方で、後者の対策である、分類対象となるデータを増やすことは有効であると考えられる。データを増やすことで、パターンを拡張することなく、新しいリンクが

見つかる可能性は高い。

提案手法が文間関係認識手法よりも有効だった「救援物資の空中投下」について、その理由を述べる。提案手法により正しく分類できた 2 つのツイートを示す。

- a 【デマ 3 3】 「日本では物資の空中投下が認められていない」 そんなことはありません。
- b なんと驚いた情報です！日本では物資の空中投下が認められていないんだそう！とっくに自衛隊が孤立被災者に実施してると思ってた。これでは本当に孤立者が死んでしまう。救出前にヘリで食糧を落として何が悪いんだろう。わたしは今これを知り怒りで全身が震えてます。みなさんリツイートをお願い！

ツイート a が、括弧による引用に対して否定する構造となっているため、文間関係認識手法が対応できなかった。この事例は、図 5.4 のパターンによって分類された。パターン中の c1 として使われたツイート例としては、以下のようなものがある。

- @XXX 【お願いします】問題のツイート削除してください！このままでは本当に物資の空中投下が出来ないと誤解されます。物資が遅れている、投下より安全な輸送を優先している、これは事実です。しかしツイッター上では議論がずれてしまってる現状をご理解ください！
- 始まってます。http://bit.ly/fVC21T 現地ですと情報が錯綜して大変だと思いますが応援しています。RT @mikajohn: なんと驚いた情報です！日本では物資の空中投下が認められていないんだそう！とっくに自衛隊が孤立被災者に実施してると思ってた。これでは本当に孤

これらは、a をリツイートしたユーザが b に対して反論した例であり、それが正しく分類できたことで、a, b 間の関係を正しく分類することができた。比較的信用できる情報を RT しつつ、デマと思われるツイートに反論するのは、このパターンで抽出できる典型的なパターンと見られる。

5.6.4 論述構造の可視化

今回の実験で得た結果を元に、「イソジンを飲むと放射能対策になる」のトピックに関するツイート空間の論述構造グラフを作成した。作成の手順は以下である。まず、「イソジンを飲むと放射能対策になる」に対応する 354 個のツイートペアに対し、提案手法による関係分類を試み、「反論」か「疑問」と識別したツイートペアについてはそれらの関係で結ぶ。関係が結ばれたツイートペアを全て集めたものを、グラフに載せるツイートの集合とする。次に、そのツイート集合内に生じる全てのツイート間のペアのうち、すでに「反論」か「疑問」の関係がついたペア以外については、1 つ以上の手法で「同意」と識別されたものがあれば、「同意」関係で結ぶ。以上の流れで論述構造グラフを作成することで、「反論」や「疑問」といった重要な関係を抽出しつつ、無闇にツイート数を増やすことのないグラフを得ることができる。

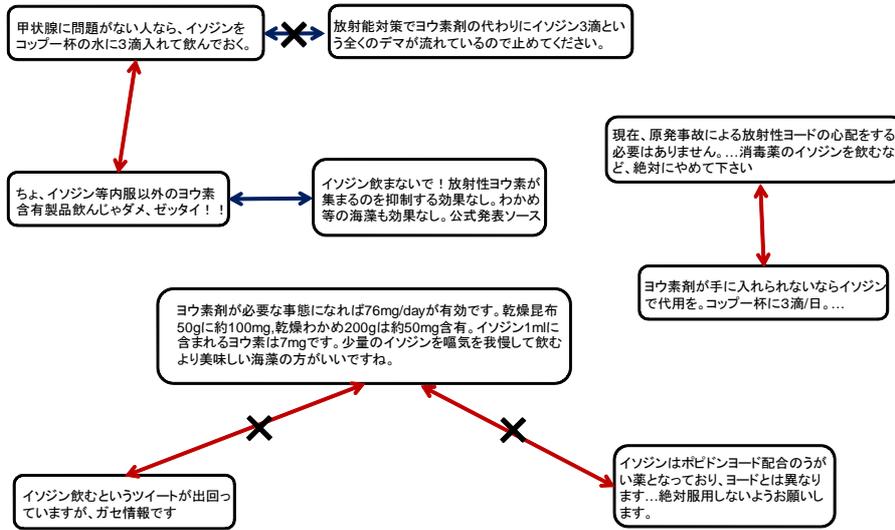


図 5.8. 「ヨウ素を飲むと放射能対策になる」の論述構造グラフ。

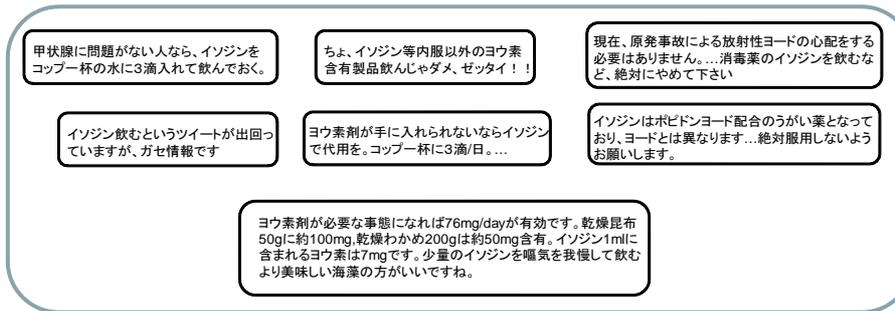


図 5.9. 「ヨウ素を飲むと放射能対策になる」に関連するツイート群

今回の実験結果を元に作成できたグラフを図 5.8 に示す。

グラフ中の赤線は「反論」、青線は「同意」と分類された関係を表している。ただし、これらの分類結果が誤っていた関係には×印を付与してある。この実験では、「疑問」を識別することはできなかった。

現状の課題として、クラス分類を行った関係に対して誤りが多いことが挙げられる。正しく分類されている関係のみを見ることができれば、ユーザーは十分な情報を得ることができるが、間違って分類されている関係が混在していると、情報を取得するにあたり混乱を招くこととなる。より実用性のある論述構造を得るには、分類の精度を高めなければならない。また、「疑問」の関係にあるツイートや URL を含むツイートを取得し、論述構造グラフに組み込むことも目指している。

なお、現状の精度では、無理にグラフ化しても分かりやすい論述構造を得るのは難しい。「反論」や「疑問」の関係が抽出されたツイート群を、単に関連のあるツイートとして並べたほうが有用な可能性も否定できない。参考として、そのような想定の下での出力を図 5.9 に示す。

これは第 2.3.2 節で述べたトピックに関する要約に近い形である．要約としての詳細な評価は省くが，トピックに関する情報はバランスよく得られており，ユーザの信憑性支援判断に有用であると言ってよい．

5.7 まとめ

本節では，返信態度の分類器を一般のツイート間関係認識へ拡張する手法を提案し，その評価を行った．ネットワークの拡張による手法はある程度の精度が得られたものの，再現率については低い．「疑問」の関係にあるツイートを抽出するためには，リツイート数が上位のものを探索するだけでは不十分であることも確かめられた．次に，論述構造マップのサンプルについて，その有用性を述べた．現状では誤った識別が多く，また十分な整理も行われていないため，ユーザにとって利用しやすいものとは言えない．一方で，単に反論と識別されたツイートペアを並べるだけのやり方も示した．トピックの要約とも言えるこの形も，ユーザの信憑性判断支援にはある程度十分な情報量が得られていることを確認した．

第 6 章

結論

6.1 本研究のまとめ

本研究では、震災時など多様な情報が Twitter 上で氾濫するような状況を想定し、ツイート空間の論述構造を解析する試みを行った。多岐に渡るツイートを整理することは容易ではないが、我々はまず返信で表明される態度に着目し、教師あり学習による 4 クラスへの分類を行った。反論表現辞書や構造的特徴を用いることで性能を向上できることを示し、特に重要な論述関係である「反論」の識別性能の F 値で 0.472、4 クラス全体の Accuracy で 0.751 という性能を得た。

続いて、一般のツイート間の論述関係を分類する手法として、ネットワークの拡張による提案手法と、文間関係認識による手法を比較した。前者は、あるツイート間の関係を間接的に表すと思われるツイートやユーザーの関係を探索し、必要に応じて返信の分類器を適用することで、元のツイートペア間の関係を推測する手法である。それに対し後者は、ツイート間の関係を直接言語処理で解く手法であり、従来研究の含意関係認識器を本稿のタスク用にカスタマイズして使用した。各手法にはそれぞれ得意なタスクがあり、2 つの手法は相補的に作用することが期待される。また、提案手法により論述構造グラフのサンプルを示した。

6.2 今後の課題

本研究は、最終的には実際のアプリケーションによるユーザの支援を目標としている。そのための今後の課題としては、まず返信の態度分類器の性能をさらに高める必要がある。具体的には、反論表現辞書の拡張および、より有効な構造的特徴の開発が必要となる。さらに、より有用な論述構造グラフ作成に向けては、ネットワークの拡張による手法が重要である。具体的には、より有用なパターンの検討・実装とそれに基づく広範なネットワークの探索、およびそのためのデータ整備を行う。また、深い言語処理との組み合わせを進めていくことも必要である。

実際にユーザ支援に役立つアプリケーションを開発するためには、ユーザに見せるための工夫、すなわち論述構造グラフの作成アルゴリズムと提示方法が非常に重要である。本論文で

は、論述構造グラフの作成アルゴリズムについては深い議論は行わず、ケーススタディ的には成果物を得ることができたが、トピックによる差は大きく汎用的なものとはなっていない。さらに、そのような論述構造グラフを作成することが必須かどうかも検討を必要とするところである。ユーザ支援に最も必要なのは、「反論」関係にある情報の提示である。しかしながら、現状では反論ツイートの識別性能は十分ではなく、また今後性能を向上させていくのも容易なことではない。そのような状況で論述構造グラフを無理に作成するよりは、反論認識で抽出できたツイートを単に並べて提示するだけのほうが役立つ可能性もある。このように様々な手段が考えられる中で有用なアプリケーションを提供するためには、ユーザインタフェースの観点からの考察も避けて通れない問題である。

このように多くの課題が残されてはいるが、元来 Twitter の持つ強力な情報伝達力の活用に向けて、本研究では一定の成果が得られたと考えている。本研究が、先の大震災のような危難の中で、Twitter 上の情報活用に一役買うことになれば幸いである。

発表文献と研究活動

- (1) 大和田裕亮, 岡崎直観, 乾健太郎, 石塚満. 返信・非公式リツイートで表明される論述関係の認識. 情報処理学会 第 75 回全国大会, March 2013 (発表予定)
- (2) 大和田裕亮, 岡崎直観, 水野淳太, 乾健太郎, 石塚満. 返信・非公式リツイートに基づくツイート空間の論述構造解析. 自然言語処理 (投稿中)
- (3) 乾健太郎, 岡崎直観, 水野淳太, 鍋島啓太, 渡邊研人, 大和田裕亮. ツイッターデータの意味的解析による災害情報拡散の分析. 東日本大震災ビッグデータワークショップ最終報告会, October 2012

参考文献

- [1] B. J. Fogg. Prominence-interpretation theory: Explaining how people assess credibility online. In *Proc. 21th International Conf. on Human Factors in Computing Systems(CHI2003)*, 2003.
- [2] C. I. Hovland and W. Weiss. The influence of source credibility on communication effectiveness. *Public Opinion Quartely*, Vol. 15, No. 4, pp. 635–650, 1951.
- [3] B. J. Fogg and H. Tseng. The elements of computer credibility. In *Proc. 17th International Conf. on Human Factors in Computing Systems(CHI1999)*, 1999.
- [4] B. Hilligoss and S. Y. Rieh. Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, Vol. 44, No. 4, pp. 1467–1484, 2007.
- [5] J. Kapoun. Teaching undergrads web evaluation: A guide for library instruction. *College & Research Libraries News*, Vol. 59, No. 7, pp. 522–523, 1998.
- [6] M. Meola. Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *portal: Libraries and the Academy*, Vol. 4, No. 3, pp. 331–344, 2004.
- [7] M. Metzger. College student web use, perceptions of information credibility, and verification behavior. *Computers & Education*, Vol. 41, No. 3, pp. 271–290, 2003.
- [8] E. Hargittai, L. Fullerton, E. Menchen-Trevino, and K. Y. Thomas. Trust online: young adults’ evaluation of web content. *International Journal of Communication*, Vol. 4, No. 1, pp. 468–494, 2010.
- [9] D. H. McKnight and C. J. Kacmar. Factors and effects of information credibility. In *Proc. 9th International Conf. on Electronic Commerce(ICEC2007)*, 2007.
- [10] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proc. 24th International Conf. on Human Factors in Computing Systems(CHI2006)*, 2006.
- [11] トウギャッター株式会社. ツイッターをまとめよう together.
- [12] Y. Duan, Z. Chen, F. Wei, M. Zhou, and HY. Shum. Twitter topic summarization by ranking tweets using social influence and content quality. In *Proc. 24th International Conf. on Computational Linguistics(Coling2012)*, 2012.

- [13] L.A.T. Cox Jr and D.A. Popken. Overcoming confirmation bias in causal attribution: A case study of antibiotic resistance risks. *Risk Analysis*, Vol. 28, No. 5, pp. 1155–1172, 2008.
- [14] S. Akamine, D. Kawahara, Y. Kato, T. Nakagawa, K. Inui, S. Kurohashi, and Y. Kidawara. Wisdom: A web information credibility analysis systematic. In *Proc. International Conf. on ACL-IJCNLP 2009 Software Demonstrations*, 2009.
- [15] 水野淳太, 渡邊陽太郎, 村上浩司, 乾健太郎, 松本裕治. 文間関係認識に基づく賛成・反対意見の俯瞰. *情報処理学会論文誌*, Vol. 52, No. 12, pp. 3408–3422, 2011.
- [16] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proc. 20th International Conf. on World Wide Web(WWW2011)*, 2011.
- [17] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proc. 8th Empirical Methods in Natural Language Processing(EMNLP2011)*, 2010.
- [18] M. Gupta, P. Zhao, and J. Han. Evaluating event credibility on twitter. In *Proc. 12th International Conf. on Data Mining(SDM2012)*, 2012.
- [19] Y. Yamamoto and K. Tanaka. Enhancing credibility judgment of web search results. In *Proc. 29th International Conf. on Human Factors in Computing Systems(CHI2011)*, 2011.
- [20] J. Schwarz and M. Morris. Augmenting web pages and search results to support credibility assessment. In *Proc. 29th International Conf. on Human Factors in Computing Systems(CHI2011)*, 2011.
- [21] R. Ennals, B. Trushkowsky, and J. M. Agosta. Highlighting disputed claims on the web. In *Proc. 19th International Conf. on World Wide Web(WWW2010)*, 2010.
- [22] I. Dagan, O. Glickman, and B. Magnini. The pascal recognising textual entailment challenge. In *First Machine Learning Challenges Workshop*, 2005.
- [23] P.D. Turney. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proc. 40th Annual Meeting on Association for Computational Linguistics(ACL2002)*, 2002.
- [24] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proc. of the 1st Workshop on Unsupervised Learning in NLP*, 2011.
- [25] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-dependent twitter sentiment classification. In *Proc. 49th Annual Meeting of the Association for Computational Linguistics(ACL2011)*, 2011.
- [26] A. Hassan, A. Abu-Jbara, and D. Radev. Detecting subgroups in online discussions by modeling positive and negative relations among participants. In *Proc. 9th Empirical Methods in Natural Language Processing and Computational Natural Language Learning(EMNLP-CoNLL2012)*, 2011.

- [27] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *In Proceedings. of EMNLP*, pp. 230–237, 2004.
- [28] Danilo Giampiccolo, Bernardo Magnini, Ido Dagan, and Bill Dolan. The third pascal recognizing textual entailment challenge. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pp. 1–9, 2007.
- [29] Marie-Catherine de Marneffe, Anna R. Rafferty, and Christopher D. Manning. Identifying Conflicting Information in Texts. In *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*. Springer, 2011.
- [30] Hideki Shima, Hiroshi Kanayama, Cheng-Wei Lee, Chuan-Jie Lin, Teruko Mitamura, Yusuke Miyao, Shuming Shi, and Koichi Takeda. Overview of NTCIR-9 RITE: Recognizing Inference in TExt. In *Proceedings of the 9th NTCIR Workshop*, pp. 291–301, 2011.
- [31] Taku Kudo and Yuji Matsumoto. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pp. 63–69, 2002.
- [32] Yotaro Watanabe, Masayuki Asahara, and Yuji Matsumoto. A structured model for joint learning of argument roles and predicate senses. In *Proceedings of the 48th Annual Meeting of the Association of Computational Linguistics*, 2010.
- [33] 江口萌, 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. モダリティ、真偽情報、価値情報を統合した拡張モダリティ解析. 言語処理学会第 16 回年次大会発表論文集 E3-8, 2010.
- [34] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. Enhancing the Japanese WordNet. In *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 1–8, 2009.
- [35] Chikara Hashimoto, Kentaro Torisawa, Kow Kuroda, Masaki Murata, and Jun'ichi Kazama. Large-scale verb entailment acquisition from the web. In *Conference on Empirical Methods in Natural Language Processing (EMNLP2009)*, pp. 1172–1181, 2009.

謝辞

修士研究を進めていく過程で御指導いただき、自由に研究が行える環境と機会を与えていただいた石塚満教授に、大変深く感謝いたします。なかなか自分の主張を曲げずに強硬的な態度を取ることのあった私を温かく見守ってくださったおかげで、納得の行く研究ができました。

研究のテーマという基礎的な部分から、実験や原稿の細部にいたるまで、議論にお付き合いくださり、多大なサポートをしてくださった東北大学乾・岡崎研究室の岡崎直観准教授に深く感謝いたします。本論文は岡崎先生のご指導なしには完成しませんでした。時には教育や人生のことについてなども楽しく話すことができ、研究以外の面でも支えてくださいました。本当にありがとうございました。乾健太郎教授とは直接お会いする機会は少なかったものの、こちらの報告に対し必ず的確なアドバイスを返してくださいました。感謝しております。同研究室研究員の水野淳太博士は、手法比較のための文間関係認識器を提供してくださると共に、研究の相談にも乗ってくださいました。大変感謝いたします。また、岡崎先生に自分のことを紹介してくださった、昨年卒業された田中翔平さんに感謝いたします。

本研究を行うにあたり、貴重なデータを提供してくださった株式会社ホットリンクに感謝いたします。東日本大震災ビッグデータワークショップは、自分の研究成果を活用できる貴重な機会となりました。このような機会を設けてくださったグーグル株式会社の方々に感謝いたします。本研究に多くの有益なコメントをくださった、言語処理学会の編集委員および査読者の方々に感謝いたします。

修士論文提出までのおよそ3か月間、食生活を支えてくれた中央食堂のF定食に深く感謝します。ライス・豚汁・野菜炒め400g・冷奴・ほうれん草のおひたしで480円というコストパフォーマンスは最高でした。この2年間の運動生活を支えてくれた、御殿下記念館の設立・運営に関わる全ての方々に感謝します。フリーウェイトも含め、こんな施設が年間9000円ほどで利用できるとは信じられません。退出する時に名前を言わずとも笑顔で利用証を返してくれた窓口のお兄さん、ちょっと恥ずかしかったですがありがとうございます。

研究室で1人寂しくしている私の孤独を和らげ、研究に関する議論にも積極的に付き合ってくれた同期の稲垣隆史君に感謝します。幸か不幸か同じ就職先になってしまいましたが、今後もよろしくお願ひ。研究室で1人寂しくしている私がSkypeで絡んだ時に、忙しいにも関わらずよく付き合ってくれた浅見・川原研究室の加藤拓也君と、江崎研究室の石橋尚武君に感謝します。ありがとう。同じくTwitterで一番よく絡んでくれた江崎研究室のエースこと小坂良太君に感謝します。ありがとう。研究室に配属された卒論生の新谷君・砂川君・田中君・西野君。あまり研究室で見かけた記憶はないけど、麻雀は楽しかったですね。ありがとう。研

研究室にいるのが寂しくてフラフラと顔を出した時に温かく迎えてくれた護身武道空手部のみんな，ありがとう．もう1度強い東大を復活させることを期待しています．1人寂しく研究をしなければならぬ時に支える力になってくれたのは，卒論の時に配属された森川研究室での厳しい教えでした．同研究室の皆様へ感謝いたします．いいトシして研究室で1人寂しくしている私を大学院まで出してくれた両親に深く感謝します．心穏やかに余生を過ごされることを祈っています．

最後に，研究を進めるにあたってこの2年間お世話になった全ての方々に対して，改めて心から感謝の意を表します．

付録 A

反論表現の一覧

4.3.4 節で示した反論表現辞書について、収録した反論表現の一覧を以下に載せる。

表 A.1. 反論表現辞書

RT しないで下さい	やめろ
RT は慎重に	よく考えてツイートすべき
RT やめてもらえませんか	よけいなこと書くな
RT を取り消して下さい	リツイートするのは慎重にしたほうがいい
お控えになられたほうが良い	安易なリツイートは控え
お前にそれがわかってんのか	意味がわからん
こういう話は広めない方が良い	意味わからん
このような事実はありません	違います
この程度でバタバタするなよ	何の情報源もない噂
これ、デマです	気をつけられた方が良い
これデマね	元ツイートを削除していただけませんか
これは違います	元のツイートの削除もして頂けないでしょうか
これ以上首を突っこむべきではない	言い過ぎだと思います
せめて公式 RT してください	言うな
そういうレベルの話じゃない	誤解を招く表現は止めてください
ソースの確認お願いします	誤解を生む RT は控えて、RT 元は吟味していただきたい
そのようなことはありません	混乱を招くような post は控えねばなりません
そのようなことは無い	混乱を避けるために削除して下さい
そのような事実はない	最低です
そのような心配はありません	削除したほうが良いと思われま
それはデマです	削除訂正をおねがいします
そんなことないでしょう	削除訂正をお願いします
そんなことはありません	使用しないで下さい
そんなことはない	事実ではありません

そんなこと言っている場合ではない	自分の一言の影響考えろ
そんな情報はまだない	信じるな
そんな法律はない	信用しない
だから何なの？	真偽の分からない情報を拡散するのはやめた方がいい
ツイートされるのを控えた方が	煽るな
ツイートの削除をお願いします	煽ればいいってもんじゃねーぞ
ツイートを削除していただきたい	全くそういう事実はない
ツイートを削除していただけますか	全くそういう情報はない
ツイートを削除してください	速やかに post の削除を
ツイートを削除願います	訂正していただいてもよろしいでしょうか
ツイート削除をお願いします	納得するわけない
テキトーな情報流すな	馬鹿はおめーだ
デマではないと思います	発言をしないで下さい
デマというのはウソだ	批判もいい加減にしてほしい
デマという報告多数です	必ず根拠を確かめて
デマなのでもう広めないでください	不安を煽る表現はなるべく避けよう
デマをばら撒いて不安を煽るのだけは辞めて頂きたい	不安を煽らないでほしい
デマをリツイートするなんて最低	不安を煽るな
どうかと思います	不安煽りすぎじゃないか
どうかと思う	不安煽る様な争うな
どーでもいい	風評被害はやめて
なにいったんだ	未確認情報は流すな
はあ？	問題のツイート削除してください
ハア？	様々な情報に惑わされぬよう
もっと有益となる情報を流布しろ	流れた情報は鵜呑みにせず一呼吸しまずは情報元を確認しましょう
やめましょうよ	流言飛語垂れ流すのも大概にして下さい