

# リツイート構造を用いたデマ拡散防止支援手法

中原 英美<sup>†</sup> 富永 一成<sup>‡</sup> 牛尼 剛聡<sup>‡</sup>

<sup>†</sup>九州大学芸術工学部 〒815-8504 福岡県福岡市南区塩原 4-9-1

<sup>‡</sup>九州大学大学院芸術工学研究院 〒815-8504 福岡県福岡市南区塩原 4-9-1

E-mail: <sup>†</sup> ksk.a.232@gmail.com, <sup>‡</sup> ushiama@design.kyushu-u.ac.jp

**あらまし** 近年, Twitter において誤った情報 (デマ) が拡散されるという問題が起こっている. このようなデマは, 主にリツイートと呼ばれる他のユーザの投稿を再投稿することによって拡散される. リツイートされたツイートは大勢のユーザに広がっていくために, 一部のユーザ間で情報の訂正がなされたとしても, 他のユーザはそれを見つげにくく, デマは広がり続ける. これを解決するために, リツイートの回数やコメント付きリツイートに関するコメントを用いることで, 情報がデマであるか否かの判断を支援する手法を提案する.

**キーワード** Twitter, デマ判断, 信憑性

## 1. はじめに

近年, マイクロブログの一種である Twitter[1]の利用者が飛躍的に増加している. Twitter は, 個々のユーザが「ツイート」とよばれる 140 字以内のショートメッセージを投稿し, 閲覧できるコミュニケーション・サービスである. 2012 年 1 月現在, 日本国内の利用者数は約 2100 万人に達した[2]. Twitter は誰でもリアルタイムで情報を発信でき, 「フォロー」という機能を利用することによって自分の興味のある他のユーザを指定し, そのユーザのツイートを閲覧することができる. そのため, 知り合いとの気軽なコミュニケーション手段や, ユーザが興味のある分野の情報収集といった目的で利用されている.

Twitter では, リツイート (RT) と呼ばれる機能によって手軽に情報を広めることができる. リツイートとは他のユーザの投稿を再投稿することである.

リツイート機能はボタン一つで手軽に利用することができるため, 情報を広めるのに有効である. しかし, その一方で誤った情報 (デマ) も急速に広めてしまうという問題がある. 例えば, 東日本大震災の際に, Twitter で, 「放射線対策にイソジンを飲むといい」, 「関西以西でも大規模節電の必要」など, 様々なデマが拡散した. その結果, Twitter 上では正しい情報と誤った情報が混在し, 情報の混乱を招いた.

デマの拡散を防止するための対策として, リツイートされたツイートがデマかどうかを自動的に判断することが考えられる. しかし, 自動的にツイートがデマであるかを高精度に判定することは非常に困難である. そこで, 本研究では, ツイートがデマであるかを自動的に判断するのではなく, ユーザが, リツイートされたツイートに関して, それがデマであるかどうかを判断することを支援する機構によって, リツイートによ

るデマの拡散の被害を軽減させることを目的とする.

Twitter 上でデマが拡散されやすい理由の一つとして, Twitter に投稿できる文章には 140 字以内という制限が設けられていることが挙げられる. 字数の制限がなければ情報ソースを提示することで信頼性を示すことができるが, 字数が制限されているので, 情報のソースまで載せることが難しい. その結果, その情報がデマであっても気付かずにリツイートによって拡散されてしまうことが多い. また, デマのツイートがリツイートされ, 大勢のユーザがデマ情報を含んだツイートを見た場合, 一部のユーザ間でデマ情報への訂正が行われたとしてもその訂正に気付かずにデマのツイートをリツイートするユーザが出現する可能性が高い. このように, デマのツイートをリツイートしたユーザ全員がそのデマの訂正情報を見るときは限らず, Twitter 上で誰かが訂正をしたからといってデマの拡散を防止することは難しい.

本研究では, 上記の問題点を解決するために, リツイートする時点での訂正情報を提示し, そのツイートがデマかどうかをユーザが判断することを支援することによって, デマの拡散を防止する手法を提案する. 提案手法では, リツイートの回数や非公式リツイートについた付加されたコメントを用いる.

## 2. 関連研究

### 2.1 信憑性評価についての研究

山本ら[3]は Google や Yahoo!などのウェブ検索エンジンで検索された結果に関する信憑性の判断支援システムの提案をしている. このシステムでは, 検索結果の信憑性判断のために accuracy (正確さ), objectivity (客観性), authority (権威), currency (鮮度), coverage (詳細性および網羅性) を主要指標として用い, 信憑性を考慮してウェブページをランキングしている. 本

研究の目的は Twitter 上でリアルタイムにつぶやかれるツイートを対象に信憑性を評価することであるため、Twitter の仕組みに適した手法を開発する必要がある。

## 2.2 Twitter についての研究

近年 Twitter を中心に、マイクロブログについての研究が盛んに行われている。

Twitter の特徴的な機能とも言えるリツイート機能を利用した研究の一つとして、吉本ら[4]はツイートの引用回数を考慮して投稿者の重要度を測る手法を提案している。この手法では、リツイートされたツイートは重要なツイートであること、頻繁にリツイート機能を利用しているユーザは引用する情報が豊富であると考えている。それらの情報の重要度と豊富さを要素の一つとして考慮することによって投稿者の重要度を測っている。

また、Twitter 上には膨大な数の情報が存在しているため、情報収集に関する様々な研究が行われている。例えば、巻田ら[5]は有用なアカウントを検索するために、アカウント情報、公式固有のウェブページの有無とその有名度、Twitter アカウントのつぶやきと公式 HP の内容の類似度の3つの要素でアカウントの信頼度を評価している。

ツイートの内容からツイートの種類を機械学習により分類する手法として、Irani ら[6]はリンクされたウェブページの内容を用いてスパムかどうかを分類し、Sankaranarayanan ら[7]は Bag-of-Words を用いてニュースとノイズのフィルタリングを行った。また、Sriram ら[8]は Bag-of-Words だけでなく、著者情報やツイートのテキスト情報も用いて、ツイートを5つのタイプに分類している。

アカウントの信頼度や重要度の計測、機械学習によるツイートの分類についての研究は行われているが、ツイートの内容自体の信頼度についての研究は行われていない。

## 3. アプローチ

### 3.1 リツイートとデマの拡散

Twitter はツイートと呼ばれる最大 140 字のショートメッセージを利用した電子情報メディアである。Twitter では、興味のある情報を提供してくれるユーザを指定（フォロー）することによって、そのユーザが発したツイートを継続的に取得可能である。いまユーザ A がユーザ B をフォローしている場合、ユーザ B はユーザ A の「フォロワー」と呼ばれ、逆に、ユーザ A はユーザ B の「フォロワー」と呼ばれる。個々のユーザは、タイムラインと呼ばれる固有の掲示板を持つ。タイムラインには、自分が発したツイートと、自分のフォロワーが発したツイートが時系列に表示される。ユー

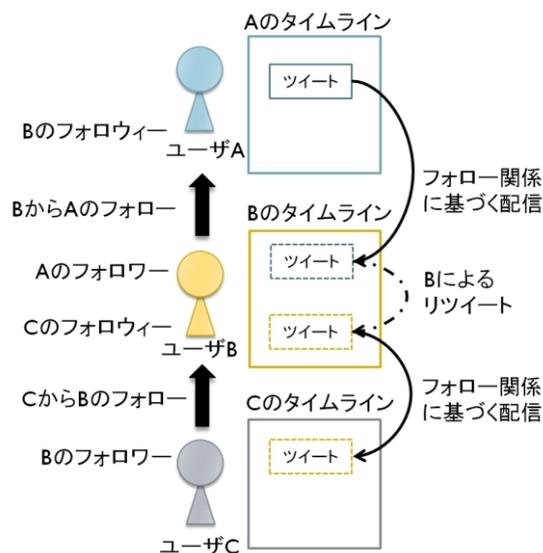


図1 リツイートの概念図

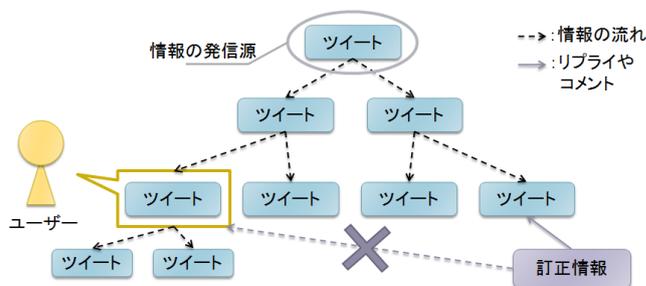


図2 デマのツイートが拡散されていく様子

ザは、自身のタイムラインを見ることによって、フォロワーが発した情報を取得できる。つまり、Twitter とは、基本的に、自分のフォロワーに対してツイートという形で情報を発信し、自分のフォロワーから情報を取得するサービスであると考えられる。

図1にリツイートの概念図を示す。図では、ユーザ B がユーザ A をフォローし、ユーザ C がユーザ B をフォローしている状況を表している。ユーザ A が投稿したツイートは、フォロワーであるユーザ B のタイムラインに配信される。これを読んだユーザ B が、そのツイートをリツイートしたとする。このとき、リツイートは、ユーザ B のフォロワーであるユーザ C のタイムラインに配信される。ユーザ C は、さらにこれをリツイートすることにより、ユーザ C のフォロワーにも配信可能である。

Twitter では、ツイートを一般に公開せず、自分が承認したフォロワーのみにツイートのアクセスを可能とするように設定することができる。しかし、現在の Twitter 上のツイートの多くは、一般に公開され、誰でも閲覧できるようになっている。そのため、有名人や企業などは、多くのフォロワーを有している。

Twitter のフォロー関係における情報の流れは単方向である。すなわち、ユーザ A がユーザ B のフォロワーであったとしても、ユーザ B がユーザ A のフォロワーであるとは限らない。したがって、自分のフォロワーが発したツイートから取得した情報は、自分のフォロワーには伝達されていない可能性がある。そこで、自分が Twitter 上で、自分のフォロワーから取得した情報を、自分のフォロワーに伝達したいという場合が多く存在する。自分がフォロワーから取得したツイートを、自分のフォロワーに伝達することを目的に、自分のツイートとして発信することはリツイートと呼ばれる。

リツイートは Twitter の特徴的な機能である。リツイートによって、一人の発した情報がねずみ算的に拡散可能である。リツイート機能によって、個人が発した情報が、瞬時に多くのユーザに伝達されることがある。

ツイートの拡散は図 2 のように、木構造的に行われる。また、先ほど述べたように、ツイートの情報伝播は単方向的に行われる。したがって、基本的には、リツイートされる情報の提供者は、自分が発したツイートが、リツイートによってどれだけ拡散し、どのユーザに配信されたかは解らない。また、リツイートされた情報に対して、それを read ユーザが別のツイートを行うなどで議論や付加情報が提供される場合がある。しかし、上記に述べたような、Twitter の情報伝播の単方向性により、そこで行われた議論は、別の経路でリツイートを受け取ったユーザには伝達されない。

近年、Twitter 上で誤った情報が提供される現象が問題になっている。Twitter では、誤った情報が流れた際に、正確な情報を知っている人がその情報に対する訂正情報をツイートする場合がある。しかし、上記に述べた理由により、与えられた訂正情報は、別の経路で誤った情報を取得した利用者には伝達されない。

本研究では上記の問題点を解決するために、ユーザが自分の read リツイートの信憑性を確認したいと思ったとき、別の経路で配信されているリツイート上で投稿された訂正情報の有無や、リツイート経路上に存在するユーザの信頼度等を確認可能とすることにより、ユーザのリツイートに関する信憑性の判断の支援を行う手法を提案する。

### 3.2 リツイートの種類とデマ判断支援

リツイートによって情報が拡散された場合、もし、そのツイートの真偽性に疑念を感じたユーザが存在する場合は、その疑念をツイッター上で表現することがある。本研究では、リツイートによって拡散された情報に対して、それを read ユーザが疑念を感じた際の反応を抽出し、リツイートの信憑性の判断の支援に利用することを考える。

リツイートは公式リツイートと非公式リツイートに分類される。公式リツイートは発言元がそのまま表示されるのに対し、非公式リツイートは「コメント RT: @username 引用文」という形式で新しいツイートとして投稿される。

公式リツイートでは、Twitter のシステム上でリツイート関係が明示的な関係として管理される。そのことにより、リツイートされた回数の集計が可能であり、元のツイートが削除された場合には、リツイートも削除されるなどの機能が実現されている。しかし、公式リツイートでは、自分が read フォロワーのツイートを、そのままの形で、自分のフォロワーに伝達するだけであり、リツイートを行ったユーザが、そのツイートに関して情報や感想を付加することができない。

一方、非公式リツイートでは、Twitter 上でリツイートの関係が明示的に管理されないため、リツイート数の集計や、元のツイートが削除された際に、リツイートも自動的に削除されるような機能は提供されない。しかし、非公式リツイートではユーザはリツイートする際に、ツイートの先頭にコメントとして、付加的な情報を追加したり、感想を付加することが可能である。

本研究では、非公式リツイートに付加されたコメントを利用して、それを read ユーザの、リツイートに対しての疑義や訂正情報を取得し、ユーザのリツイートに関する信憑性の判断の支援に利用する。すなわち、リツイートに対して、それに関する訂正情報が多く付加されている場合、そのリツイートはデマである可能性が高いとして、ユーザに注意を促す。

なお、本研究で目的とするのは、リツイートで配信された情報の信憑性を正確に判断することではない。リツイートで配信された情報がデマであるか、正確な情報であるかを正確に判断するためには、情報のソースの調査や、発信者の信頼性の評価など、複雑で高度な処理が必要であり、それを自動的に行うのは非常に困難である。本研究では、最終的な信憑性の判断はユーザが行うことを前提とする。この場合、ユーザがリツイートされた総ての情報に関して、手動でそれらの処理を行うことは現実的でない。そこで、本研究では、リツイートに与えられたコメントを利用して、リツイートが拡散していく状況で信憑性に疑問が持たれているかどうかを、ユーザが簡単に調査できるようにすることで、信憑性が低い情報に関する注意を促す。

## 4. 提案手法

提案手法では、非公式リツイートに付加されたコメントを、リツイートしたツイートに対する意見として抽出する。抽出したコメントをそのコメントに含まれている単語によって分類し、提示することによってユ

ーザがデマかどうかを判断できるよう支援する。提示する情報は「訂正ツイート」と「危険度」である。これらの詳細を以下の小節に示す。

#### 4.1 訂正ツイート

リツイートに対して、その情報に関して「デマ」など、その信頼性に対して否定的な意見をコメントとして付加している非公式リツイートを「訂正ツイート」と呼ぶ。

訂正ツイートを取得するために、Twitter API[9]の search メソッドを使用し、「username (スクリーンネーム)」+「引用文」をクエリとして検索して得られた非公式リツイートからコメントを抽出する。抽出したコメントが「デマ」「間違い」「ガセ」「釣り」「嘘」という単語を含む場合、そのコメントは引用文の内容に対する訂正情報であると考えられる。そこで、「デマ」などの単語を含むコメントのついた非公式リツイートを訂正ツイートとする。

#### 4.2 危険度

あるツイートがリツイートされたうち、リツイートの中で、訂正ツイートが占める割合のことを危険度と呼ぶ。

訂正ツイート数については上記の訂正ツイートを取得する際にカウントし、公式リツイート数は TOPSY[10]が提供する otter API を使用して取得する。取得した公式リツイート数、訂正ツイート数を用いて危険度を算出する。

具体的には、公式リツイート数を  $O_n$ 、訂正ツイート数を  $C_n$ 、訂正ツイートがリツイートされた回数を  $R_n$  とすると、危険度を以下の式で求める。

$$(\text{危険度}) = \frac{|C_n| + |R_n|}{|O_n|}$$

### 5. 実験

#### 5.1 目的と手順

提案手法の有効性を示すために実験を行った。本実験では、デマであることが分かっているツイートを対象に、そのツイートに関する「訂正ツイート」、「危険度」のデータを取得し、結果を検証した。

#### 5.2 プロトタイプ

実験を行うために、提案手法が動作するプロトタイプシステムを実装した。

本プロトタイプシステムでは、OAuth 認証を使ってユーザのタイムラインを取得し、画面上に表示する。そして、図3のようにツイートごとに確認ボタンが表示される。確認ボタンを押すと、まず、ツイートの username (スクリーンネーム) とステータス ID が引数として渡される。次に引数として渡されたスクリーンネームとステータス ID、そして TOPSY の提供する otter API を利用して確認ボタンが押されたツイートの



図3 プロトタイプの画面



図4 デマであるツイートの場合

公式リツイート数を取得する。その後、引用文となる本文を取得し、「スクリーンネーム+引用文」をクエリとして検索し、得られたツイートの内容をデータベースに保管する。そのデータベースの中から「デマ」「間違い」「ガセ」「釣り」「嘘」という単語で検索し、いずれかの単語を含むツイートを取り出し、取り出したツイートの数をカウントし、内容をデータベースに保存する。また、ユーザに視覚的にわかりやすく提示できるよう、100 をかけてパーセント表記になおす。こうして得られた危険度と訂正ツイートを結果として表示する。

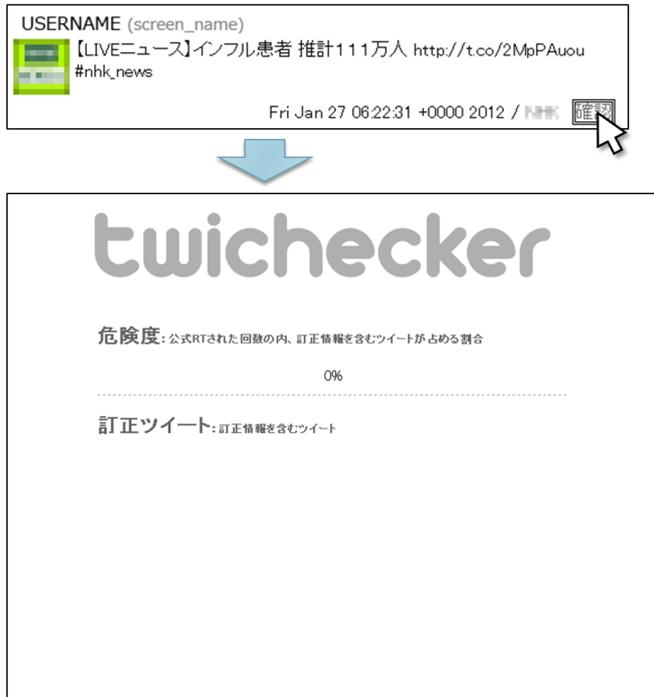


図5 デマでないツイートの場合

今回、本プロトタイプシステムを実装するにあたって利用した otter API を提供している TOPSY とは、リアルタイム検索のサービスである。TOPSY はリアルタイム性の高いメディアに特化して検索を行うことができ、手頃な Twitter 検索のサービスとして利用されている。

実行結果を図4、図5に示す。図4はデマであるツイートの確認ボタンを押した場合、図5はデマでないツイートの確認ボタンを押した場合の実際の結果である。デマであるツイートの場合、危険度が10%となり、訂正ツイートを95件取得し、表示することができた。図5の場合、otter APIによって取得した公式リツイート数は62回であるが、訂正ツイートが0件であったため、危険度は0%となった。また、訂正ツイートは0件であるので、表示されない。

### 5.3 実験方法

本プロトタイプを実際に用いて、取得した訂正ツイートと危険度について、ツイートがデマであるか判断するときに参考になるかについて5段階で評価してもらった。また、その結果を参考に、本プロトタイプによって得られた結果と手動で取得した結果を比較し、精度を確認する。

### 5.4 結果と考察

対象としたデマであることが分かっているツイートの数は2件、それぞれ例1、例2と番号を振り、内容は表1の通りである。

例ごとに「username (スクリーンネーム)」+「引用文」をクエリとして検索して得られた非公式リツイ

	引用文
例1	【拡散希望】マクドナルドで今限定 200円のビッグマックがすごく放射能臭いで気をつけて下さい！特に千切りになっているレタスから、放射能特有のあのジャリジャリ感、喉奥に残る後味、食べたあとの胸焼け感がすごく強いです！安いものには安いなりの理由があります！気をつけて！！
例2	SJの新しいアルバムは"MY SOUL"で11月2日に発表？カンインとキボムはいるけどイトクとイエソンはいない…なんか涙でてきた…嬉しいのに悲しい。やっぱ兄さんも行っちゃうんだ…

表1 対象とした引用元となるツイート

	例1	例2
訂正ツイート数	95	34
公式RT数	906	196
危険度	10%	17%

表2 実行結果

トから「デマ」「間違い」「ガセ」「釣り」「嘘」という単語を含むツイートだけを抽出し、訂正ツイートとみなすことのできる訂正情報を含んだツイートを確認することができた。また、otter APIによって公式リツイート数を取得し、訂正ツイートの数が公式リツイート数に占める割合である危険度を算出した。結果を表2に示す。

本プロトタイプを実際に利用してもらい、被験者5名に訂正ツイートの内容と危険度の有効性の評価を行ってもらった。本実験では、本プロトタイプによって抽出された訂正ツイートと危険度が、引用元となるツイートがデマであるかをユーザが判断する際に参考になるかという点について5段階で評価してもらった。今回の評価実験に使用したツイートの内容は「iPhone持ってる人！ダイヤルで\*#06#を押して、七桁目と八桁目の数字を確認してください。02,20は中国産。質が劣る。08,80はドイツ。質はまあまあ。01,10はフィンランド。質良し。00,03,04は元々の工場。質かなり良し。13はアゼルバイジャン。質最高に悪い。」である。

結果を表3に示す。数値が低いほど有効性が低く、数値が高いほど有効性が高いことを表している。

アンケートの結果によると、危険度の評価は2.4という、全体の評価の平均値よりも数値が低く、低い評価となった。これは、本プロトタイプによって算出された数値だけを見せられてもデマであると判断していか不明瞭であることが原因だと考える。引用元のツイ

	評価の対象	評価の平均
	危険度：8%	3.6
1	こ、これはデマっぽいので、デマッター RT @minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」	4.2
2	デマ乙 RT @minari0802 ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」	3.0
3	RT @minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」 ↑のツイートはデマで「おにぎりおにぎりちよいと詰めて♪」で合ってるんだとさ	2.6
4	デマらしいですよ RT @Kirrmmi: うそ?! (((; ° D° )))) “@minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」”	3.2
5	RT @sasakimitsuhiro: ええっ!? デマ!? “@n0zak1: デマらしいですよ? RT @sasakimitsuhiro: ええっ!? “@minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」”	3.2
6	これも間違いらしい。正解は、上です! RT @minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」	3.8
7	RT @ohnotoyo: このネタ、嘘らしい RT @hairmakelina: 目からうることはこのこと。です! RT @ohnotoyo: ビックリした(^_^;) RT @minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」	3.6
8	嘘だ!! RT @Yukari_0911: なんですって… RT @jamsnisin: まじか、、 RT @owaru_kirastar: 衝撃だっ QT @minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」	4.2
9	握りは間違い http://t.co/lEidN7Rh RT @macchaka: (; ° 0°) RT @celery_a: えっ(° D°;) RT @minari0802: ここ数年で一番の衝撃 ×「おにぎりおにぎりちよいと詰めて♪」 ○「おにぎりを握りちよいと詰めて♪」	2.0
	評価の平均値	3.4

表3 アンケート結果

イトがデマであった場合の危険度の平均値をとり、その数値よりも危険度が高い場合ならデマである可能性が極めて高いなど、基準を設ける必要があることが

わかった。

また、訂正ツイートについては、9番目のツイートの評価が一番高かった。これは参考になり得る URL が一緒に呟かれているからであると考えられる。評価が低いツイートを見てみると、「～らしい」や「～っぽい」などといった曖昧な表現を使っているからであると考えられる。アンケートの結果を見ると、2番目のツイートの評価に関しては、被験者ごとに大きな差が見られた。これは2番目のツイートの内容に「デマ乙」というネットスラングが含まれているため、ユーザがこのツイートに対して受けた印象に差が出たことでそのような結果になったと考えられる。つまり、口語的表現やネットスラングを考慮し、曖昧な表現を含む訂正ツイートなどをノイズとみなし、取り除くと、より有効性の高い訂正ツイートが抽出できると言える。

アンケートの結果によると、上記で述べたように、今回実装したプロトタイプシステムには、危険ツイートについてはツイートがデマであった場合に算出される危険度の平均値をとるなど、基準を設け、訂正ツイートについてはノイズを除くように改良する必要があることが分かった。

アンケートとは別に、手動で抽出したツイートを確認したところ、不適切だったものがそれぞれ1～2件発見された。例えば表3に示されている5番目と8番目のツイートである。これは引用元の内容を信じたうえで、なかなか信じることのできないような内容に驚いているユーザがつけたコメントであった。

次に、非公式リツイートに付加されたコメント全てを手動で確認してみたところ、抽出したコメント以外に訂正情報とみなすことのできるコメントがあることが分かった。評価実験で取り上げたデマのツイートを例に挙げて、以下に具体的に述べる。このツイートを非公式リツイートする際に付加されたコメントの一部を図6で示す。この5件のコメントのうち、「デマ」などの単語を含むコメントだけ抽出した場合、抽出されるのは「デマです。iPhoneのリアルパネルをよく見ましょう」というコメント1件のみである。しかし、抽出されたコメント以外にも、「これ結局意味ないらしいよ」というコメントも引用元の訂正情報とみなすことができる。

このように「デマ」などの単語で抽出できなかった訂正情報は、「△△は○○である」という引用元の内容に対し、非公式リツイートによって「これはデマです RT: @username △△は○○である」のような、「デマ」などの特徴的な単語を含んだ訂正情報ではなく、「△△は□□である RT: @username △△は○○である」のような、引用元の内容にのみ適用し、引用元の内容以外に対して汎用性の無い文章で訂正しているために抽出

できなかったと考えられる。

## 6. おわりに

本稿では、Twitterにおいてデマかどうかをユーザが判断することを支援するための基準として危険度・訂正ツイートを提案した。提案手法では、非公式リツイートに付加されたコメントを用いて、ユーザに提示する二つの情報、「訂正ツイート」「危険度」を取得する。

提案手法を実装し、デマであるツイートを取り上げて実際に検証したところ、訂正情報が非公式リツイートに付加されたコメントに含まれていることが確認できた。しかし、今回実装したプロトタイプで取得できた訂正ツイートには引用元のツイートがデマであるかを判断する際に有効なものあまり有効でないものの両方が含まれていた。今後、有効である訂正ツイートの特徴の一つである、「信頼できるサイトのURLを含む訂正ツイート」を上位に表示し、判断材料とするにはあまり適さない「曖昧な表現を含むツイート」をノイズと見なし、除去することにより、有用性が向上すると考えられる。また、訂正ツイートの中には「デマ」などの単語を含まずに、引用元の文章の内容を否定しているツイートや、非公式リツイートを使わずに、「○○というツイートはデマなので気を付けてください」などのツイートによってデマが広がることを防止するケースもある。このような訂正ツイートも取得することができれば、ユーザがデマかどうかを判断する際に、より良い判断材料となる情報を提示することができると考えられる。

危険度については、本プロトタイプでは訂正ツイート数が公式リツイート数に占める割合としたが、何%以上だとデマであると判断するべきかという基準を設けていなかったために、被験者実験による有用性の評価が低下してしまったと考えられる。デマであると判断するべき危険度の基準を設定するために、ユーザがデマであるかどうかを判断した結果をフィードバックし、デマであると判断されたツイートの危険度の平均値を算出し表示することが考えられる。

## 参 考 文 献

- [1] Twitter, <http://twitter.com>.
- [2] Google 社提供の DoubleClick Ad Planner を用いた調査より, <http://www.google.com/adplanner/>
- [3] 山本祐輔, 田中克己, “ウェブ検索結果の信憑性判断支援”, WebDB Forum 2010.
- [4] 吉本和紀, 鈴木優, 吉川正俊, “マイクロブログにおける他者への影響を考慮した投稿者の重要度推定手法”, DEIM Forum 2010.
- [5] 券田孝晴, 西山裕之, “アカウント情報の信頼度を考慮した Twitter アカウント検索システムの設計と実装”, 情報処理学会第 73 回全国大会.
- [6] D. Irani, S. Webb, C. Pu, and K. Li, “Study of trend-stuffing on twitter through text classification,” Proceedings of 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference, 2010.
- [7] J. Sankaranarayanan, H. Samet, B.E. Teitler, M.D. Lieberman, and J. Sperling, “Twitter stand: News in tweets,” Proceedings of 17<sup>th</sup> ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp.42–51, 2010.
- [8] B. Sriram, D. Fuhry, and M. Demirbas, “Short text classification in twitter to improve information filtering,” Proceedings of 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.841–842, 2010.
- [9] Twitter API, <http://apiwiki.twitter.com/Twitter-API-Documentation>.
- [10] TOPSY, <http://topsy.com/>