

関連記事の提供によりオンライン議論の活性化を支援する Twitter ボット

草島 将太^{1,a)} 角 康之¹

概要: 近年では、学術的な研究発表会や勉強会の場において、実際の場で行われている発表や議論の他に、その内容について議論したり、メモを取って記録・共有するための手段として、Twitter が広く利用されている。本研究は、Twitter のタイムライン上において、議論を活性化させたり、議論に参加している人々の理解を深めることを目的とする。そこで、タイムライン上の発言や議論について、関連する話題をツイート形式で提供するような Twitter ボットの開発を行った。Twitter ボットは、オンライン議論に参加しているユーザのツイートを分析し、メインピックであると考えられるキーワードを抽出、議論と関連性の深いウェブ上の記事をツイートにより紹介する。この Twitter ボットを用いて、実際の研究発表会や勉強会の場で運用実験を行い、Twitter ボットの有効性を示す。

Twitter Bots that Activate Online Discussion by Providing Related Article

SHOTA KUSAJIMA^{1,a)} YASUYUKI SUMI¹

Abstract: In recent years, we use Twitter to share reactions about live presentations and discuss during recital and study sessions with small messages. For this study, I aimed at observing and creating such discussions and better understand the people using Twitter. I then developed bots that collect data based on remarks, arguments, and analysis and initiate discussions about the events users were attending. They analyze the tweets of users participating in the discussion, extract the keyword that are the main topic, and provide the articles related to the discussion. I operate Twitter bots in the actual discussion, and investigate whether the Twitter bot could create relevant topics of discussion.

1. はじめに

近年、インターネットや SNS の普及に伴い、学術的な会議や研究発表会、勉強会などといった議論の場では、実際に行われる議論の他にオンライン上での議論が行われることも増えてきている。そのような場では、特有のハッシュタグを設け、Twitter 上で議論を行ったり、参加者同士のメモの共有が行われている。この方法は、遠隔地などの実際にその場にはいない人でも議論に参加したり、参加している人のツイートから議論の内容を読み取ることができたり、誰かの発表中でも、その発表を妨げることなくオンライン上で議論を交わすことができる。さらに、ツイート

の形でログデータが残るので、その議論を後から見返すことも容易にできる、といったメリットがある。また、内容についてインターネットで調べながら発表を聞くことができる。しかし、インターネットで検索していたり、ウェブページ探しているうちに研究発表会や勉強会が進行してしまい、理解が曖昧なままとなってしまうかもしれない。

そこで本研究では、オンライン議論の参加者にツイート形式で情報を提供する Twitter ボットの開発を行う。この Twitter ボットは、オンライン議論が行われているタイムライン上のツイートを解析し、議論の中心となっているキーワードを抽出、そのキーワードを用いて、論文掲載サイトやニュースサイトなどで検索した結果を、Twitter ボットのツイートという形でオンライン議論の参加者に提供する。

このような Twitter ボットがオンライン議論に介入し、

¹ 公立はこだて未来大学

^{a)} s-kusajima@sumilab.org

プレゼンテーションの内容との関連が深い過去の論文や、最新のニュースサイトやまとめサイトの記事をリアルタイムに提供することにより、検索する時間や、関連するウェブページを探す時間が省略され、実際の研究発表会や勉強会を聞き逃すことなく関連する情報を得ることができる。また、得られた情報からさらにオンライン議論参加者同士の議論が起こったり、議論参加者の疑問の解消や、理解の促進につながるといった効果が期待される。さらに、議論が活性化することにより、ただ発表を聞くだけでは得られない情報をオンライン議論の中で得たり、Twitter ボットから提供されることが予想されるため、議論そのものをよりよいものにできるといった効果が期待される。

以下、2章で関連研究との違いを述べ、3章でシステムの概要を述べる。4章、5章で運用実験での考察を行い、6章で今後の展望について述べ、7章でまとめる。

2. 関連研究

2.1 グループ議論支援についての研究

本研究では、オンライン会議といったグループ議論により、議論を活性化させたり、議論参加者の理解を促進させるといった点から、グループ議論を支援することを目的としている。グループ議論支援に関連する研究としては、共同発想と情報共有を促進する AIDE と呼ばれる対話支援環境を構築した角らの研究 [1] や、会議の場をリフレクションするリアルタイム会議支援システムとして、INGA と呼ばれるシステムの提案と評価を行った赤川らの研究 [2] が挙げられる。AIDE とは、ネットワークを介したチャット機能を有する電子会議システムであり、会話の対話構造を可視化した空間を表示したり、関連するテキストを自動抽出し、仮想的な対話参加者の発言としてディスカッション空間に投げ込むといった働きによって、会議や対話における参加者の協同発想を促進させるという特徴がある。また INGA とは、参加者間の知識継承の円滑化と、研究議論活性化の支援を目的としたシステムである。会議用マイクによって会議中の音声を取得し、発言内容からキーワードの抽出を行う。会議参加者はタブレット端末を通してキーワードを確認することができ、関連性の高い会議資料の電子データを検索することができる。また、該当する資料は各タブレット端末に送信され、そのデータに対し描画することができ、リアルタイムでその内容を共有することができるので、利用者間での電子データへの評価や知識共有が可能になるという特徴がある。この二つのシステムは、利用者が限定されたクローズドな独自システムであるが、本研究は、議論の場としてより一般的に普及している Twitter を用いることにより、議論への参入を容易にし、誰でもシステムを利用できることを目的としている。

2.2 Twitter が用いられた研究

誰もが簡単に利用できる Twitter に関連する研究としては、Microsoft PowerPoint のスライドショーと連動し、前もってノート欄に編集しておいた内容を自動でツイートするためのアドインである「びびつい」を開発した栗原の研究 [3] や、時系列順に表示されるツイートを再帰的にクラスタリングし、情報の広がりや可視化する「Vital Atlas」を開発した竹内らの研究 [4] が挙げられる。本研究では、びびついのように事前に登録しておいた内容をそのまま自動ツイートとしてユーザーについて提供したり、Vital Atlas のようにタイムライン上のツイートを分析するだけでなく、リアルタイムにツイートを分析し、最適な内容を議論参加者に提供することを目指す。

また、議論を活性化させたり、理解を促進させるためには、Twitter のボットがオンライン議論の場にツイートする形で参加者に直接提供する形が望ましい。Twitter ボットに関連する研究としては、「論文ったー」というボットを開発した山田の研究 [5] が挙げられる。論文ったーは、Twitter のユーザ間で盛り上がっている話題であるトレンド情報をもとに、定期的に CiNii で関連の有りそうな内容の論文を検索し、その結果をボットのつぶやきとして一般ユーザに提供するシステムである。しかし情報の提供元が Twitter 全体のトレンド情報なので、議論の対象を絞り込むことができない。そこで本研究では、議論の対象を実際に行われている研究発表会や勉強会に限定し、中心となっている話題を抽出することによって、その議論に適した情報を提供する。

3. Twitter 上での議論を支援するボット

本章では、Twitter ボットのシステムの流れについて説明する。図 1 は、本システムの概念図である。

3.1 ボットの準備

まず、検索結果を自動でツイートするためのボット用アカウントを用意する。このボット用アカウントは、対象とする勉強会や発表会で用いられているハッシュタグが付与されているツイートをリアルタイムで取得し、動作する。

以下、このハッシュタグが付与されており、Twitter 上の検索から誰でも見ることができるツイートによって構成される群をタイムラインと定義する。

3.2 キーワード抽出

ボット用アカウントは、Twitter の Streaming API を用いてハッシュタグを含むツイートをリアルタイムで取得する。Streaming API とは、HTTP 接続を続けることにより、リアルタイムにツイートを取得することができる API である。

次に、取得したツイートに関して、テキスト解析を行い

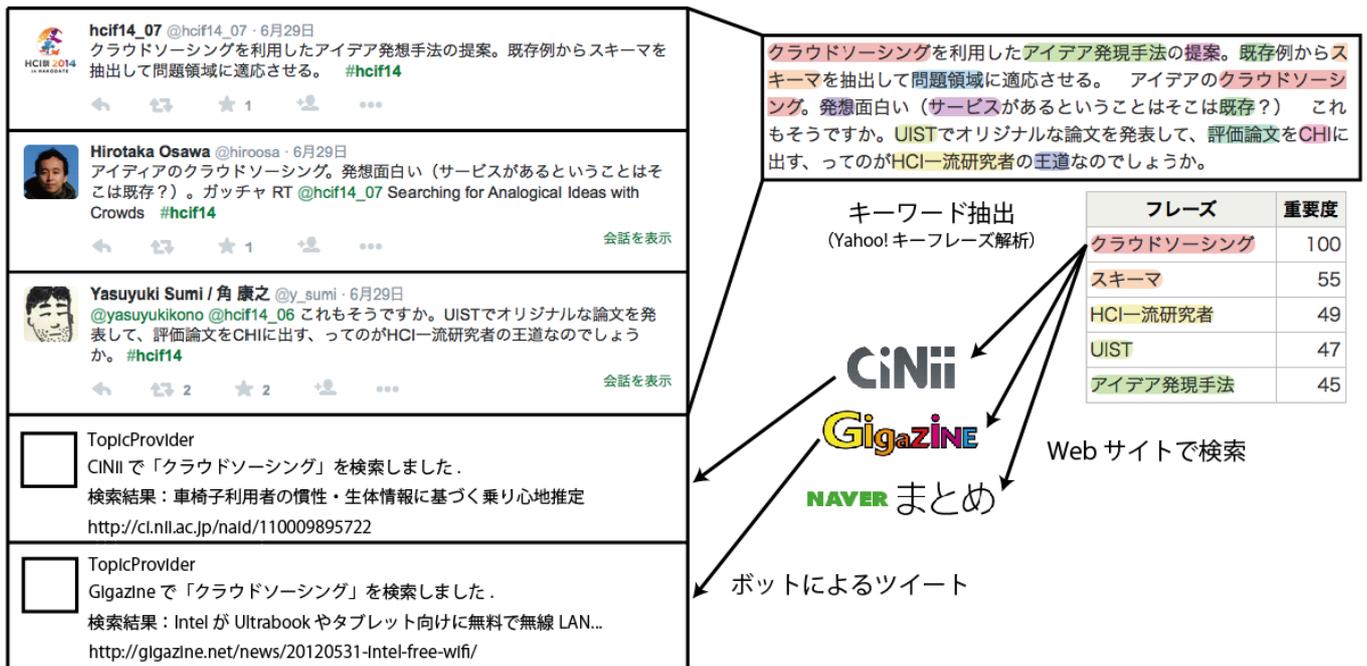


図 1 システム概念図

Fig. 1 System conceptual diagram

キーワードを取得する。まず取得したツイートから、ウェブページへの URL などの余分な情報を削り、文章として保存していく。保存された文章がある程度溜まったところで、テキスト解析を行いキーワードを取得する。今回は、テキスト解析の手法として、Yahoo!JAPAN によって提供されるテキスト解析 API の 1 つである、キーフレーズ解析 API によってツイートの解析を行う。キーフレーズ解析 API とは、指定された日本語文を解析し、その文中から特徴的な表現（キーフレーズ）をいくつか抽出し、100 以下の整数で表される評価値とともに取得することができる API である。しかしこの API を用いた場合、記号などを含む未知語を検出してしまい、有用でないキーワードを抽出してしまう場合も少なくないため、キーワードとして選出する条件として、日本語版 Wikipedia の記事タイトルとして存在するもの、かつ、評価値が最も高かったものとする。

3.3 ウェブサイトでの検索

得られたキーワードを用いて、論文データベースやニュースサイトなどの指定されたウェブサイトで検索を行う。今回は学術情報データベースである CiNii、ブログ形式のニュースサイトである Gigazine、CGM 型 Web キュレーションサービスである NAVER まとめ の 3 つのウェブサイトを用いた。ウェブサイトでの検索には、そのウェブサイト自身が提供している検索 API を利用するか、XML や JSON を用いてさまざまなウェブサイトの検索結果を取得することができる、Microsoft Azure が提供する Bing Search API を利用した。それぞれのウェブサイトから最

大 10 件の検索結果を取得し、キーワードを取得した際のツイートの内容と、検索結果のウェブページの内容とで類似度を TF-IDF と Cos 類似度推定法を用いて求め、それぞれのウェブサイトでもっとも類似度の高かったものを選出する。

3.4 ボットによるツイート

ウェブサイトによって得られた検索結果を、ボットによるツイートの形でオンライン議論参加者に提供する。利用したウェブサイトが複数であった場合は、1 つのウェブサイトにつき 1 つのツイートで情報を提供する。なお、議論参加者によるツイートが全くなく、キーワードが選ばれなかった場合や、あるウェブサイトの検索結果が 1 つもなかった場合には、そのウェブサイトについては検索結果をツイートしないこととする。

4. ライトニングトーク形式の勉強会での運用実験

2015 年 6 月 27 日に行われた実際の勉強会である CHI 勉強会 2015 で、実際に Twitter ボットを動作させた。CHI 勉強会 2015 とは、CHI2015 で発表された論文について、1 つの論文につき 30 秒で次々と発表していく勉強会である。勉強会は全部で 114 のセッションに分かれており、1 つのセッションにつき 1 人が担当し、そのセッションに含まれる 3~5 つの論文について発表を行った。勉強会は東京にある DMM.make AKIBA と北海道にある北海道大学の 2 会場で実施され、片方の会場で発表が行われているときは、

もう片方の会場ではオンラインで別会場で行われてる発表の音声と使用されているスライドが放送されていた。それぞれの論文の発表のタイミングでは、論文のタイトルと内容紹介についてのツイートがされていた。この勉強会については、東京と北海道という2ヶ所の離れた場所での勉強会であったことや、論文紹介にTwitterが用いられていたことから、勉強会と並行してTwitterを利用する参加者も多いと考えられた。

今回の実験では、1つのセッションの間にされたツイートを1つのまとまりとしてテキスト解析を行い、キーワードを抽出し、CiNii, Gigazine, NAVER まとめの3つのウェブサイトで検索して、それぞれのセッションの終わりにボットのツイートの形でオンライン議論の参加者に提供した。計114のセッションのうち、人間によるツイートが全くないなどといった理由からキーワードを抽出できなかったものを除くと、キーワードが抽出できたセッションは100であった。また、その100のセッションのうち有用なキーワードを得ることができると考えられたのは47と、全体の47%で有益な結果が得られた。なお、3章ではキーワードの抽出について「日本語版 Wikipedia の記事タイトルとして存在するもの」と述べたが、今回の運用実験ではその部分は実装していない。

予想としては、ボットによるツイートに興味を持ったユーザがお気に入り登録やリツイートをしたり、Twitterボットによるツイートが元となる議論が起これると考えられた。しかし、勉強会の期間中にはユーザからの反応はほとんど得られず、ボットがツイートした計233ツイートのうち、議論の参加者からの反応があったのはわずか7ツイートのみであった。

この結果を受けてタイムラインを分析したところ、人間ユーザによる計342ツイートのうち、他のユーザへのリプライや参照ツイートは18しかなく、全体の5%のみであった。これらのことから、Twitterボットがタイムライン上での議論に与えた影響はかなり小さなものであったが、ユーザのTwitterの使用目的は、議論よりも、気になった論文へのメモとしての役割が大きいということがわかった。また、勉強会でTwitterを利用していたユーザはお互いに知り合いというわけではなく、様々な大学や所属のユーザが参加していたため、積極的に他のユーザとのやりとりをしにくい環境であったことが理由として考えられた。

5. グループ議論での運用実験

2015年12月15日に行われた研究室での勉強会で、実際にTwitterボットを動作させた。この勉強会に参加したのは学部4年生7人と学部3年生8人の計15人であり、勉強会の内容は、学部4年生の4人が、現在自分が行っている研究の概要と現在までの成果、そして今後の展望を他の参加者に発表するものであった。この15人の参加者は

全員Twitterでのオンライン議論も行っており、発表中には他の参加者が話したり、質疑応答の時間を十分に取らない代わりに、Twitter上でメモを取ったり、議論を交わしたりしていた。発表は1人10分程度ずつ行い、勉強会を行った時間はおよそ40分間だった。このうち、本システムは6回のテキスト解析を行い、10のツイートをオンライン議論の参加者に提供した。

以下では、その中でも特徴的な結果が得られたものについて紹介する。

5.1 適切なウェブページを提供することができた例1

表1は、Twitterボットがオンライン議論の場に適切なウェブページを提供することができた例である。なお、上の「User+数字」で表されているものがオンライン議論に参加していたユーザの名前を模式的に表したもので、下の「Bot」が本研究で運用したTwitterボットのツイートを表している。また、全てのツイートにはこの勉強会のハッシュタグが含まれているが、今回はテキスト解析にはその部分を用いていないので省略した。この表のツイートがなされていた間は、実際の勉強会では、音環境を比較することにより会話グループを検出する研究の概要が紹介されていた。このタイムラインをテキスト解析することにより、下の2つのウェブサイトを提供することができた。特にCiNiiでの検索から得られた論文は、発表者が引用として使っていたものである。今回の発表ではこの論文について触れていないにもかかわらず、関連研究として議論の参加者に提供することができていたため、この結果は有用であると考えられた。また、このCiNiiの記事については、3人のユーザからお気に入り登録されていた。

5.2 適切なウェブページを提供することができた例2

表2は、表1と同様に適切なウェブページを提供することができた例である。この表のツイートがなされていた間は、図書館内の本棚を陣地とした陣取りゲームを行うことで、図書館の利用を活性化させようという研究の概要が紹介されていた。Twitterボットが提供したウェブページは、違うゲームではあるが、同じく図書館内でゲーミフィケーションを行う研究が紹介されているものであった。この研究は、発表者自身も知らない内容であったが、関連研究としてとても役に立つものであると勉強会後のアンケートで語っていた。テキスト解析によって得られたキーワードは「ゲーミフィケーション」だが、ユーザのその他の「図書館」や「ゲーム」といったツイートから、ツイートと記事内容とのCos類似度比較によるウェブページ選出は有用な手法であると考えられる。このツイートに興味を持ったユーザも多く、5人のユーザがお気に入り登録していた。さらに、Twitterボットのツイート後、CiNiiの記事についてユーザが言及し、発表者も情報を得ることができていた

表 1 成功例のタイムライン 1

Table 1 Timeline of successful example 1

ユーザ名	ツイート
User12	最初は音環境比較による会話グループの検出と場を読むロボットへの応用って発表ですな
User15	最初はいくんです!!
User04	場を読む
User10	集中線?
User02	場を読むとは
User06	難しそう
User03	場を読むことが苦手な人もいるよね
User12	グループとグループの間を通るのはいいけど、グループの中を通るのはいけない
User01	コミュニケーション空間の認識難しそう
User12	6秒ごとにフーリエ変換する
User10	0.775 っていうのが最適な値なのか
User01	フーリエ変換からのコサイン類似度比較
Bot	これまでには、こんな研究もあります! 「「空気を読むロボット」: コミュニケーション空間を利用した人を動かす HAI デザイン (<特集>人を動かす HAI)」 ci.nii.ac.jp/naid/110009594...
Bot	こんなニュース記事もありますよ! 「純度 100%の真っ暗闇の中で非日常の世界とコミュニケーション ...」 gigazine.net/news/20131222...

(User15 が発表者であった)。このように、新たな気付きを提供してくれている点で、この結果は優れたものであるといえる。

5.3 適切なウェブページを提供することができなかった例

表 4 は、これまでの 2 つの例とは違い、適切なウェブページを提供することができなかった例である。この表のツイートがなされていた間は、個人の連想から連想辞書を作り、書籍の検索に応用させる研究の概要が紹介されているものであった。テキスト解析では「連想」というキーワードを抽出することができていた。CiNii では貪欲探索法に基づいたヒステリシス連想メモリにおける進化的学習アルゴリズムを考察した論文を提供していたが、Gigazine では、脂質を多く含む食事はそうでない食事よりも長期記憶に残りやすいといったほとんど勉強会の趣旨とは関係のないような記事を提供してしまっていた。このように、勉強会の内容に関連深いキーワードを取得することはできていても、提供するウェブページが適切なものであるとは限らないといった結果が得られた。内容と関連性が薄いウェブページに関しては、たとえ取得することができたとしても、ツイートとの類似度で閾値を設け、Twitter ボットはツイートしないなどの処理が必要かと考えられた。

また、このときのタイムラインにおいて、Yahoo! キーフレーズ抽出 API によって得られたキーワードについてまとめたものが表である。この結果によると、実際に検索を

表 2 成功例のタイムライン 2

Table 2 Timeline of successful example 2

ユーザ名	ツイート
User12	ゲーミフィケーションを用いた図書館利用の活性化ですな
User12	実世界の図書館をもっと利用させたい
User12	陣取りゲームを用いる
User14	ゲーミフィケーションで釣る?
User03	図書館利用⇔ゲーミフィケーション
User04	本との偶然の出会いって何かいいね
User08	写真投稿のはもうあるんだね
User12	最初はゲームの陣取りが目的だが、次第に図書館に足を運ぶこと自体が目的になってくれるのではないか?
User10	こんな先行研究あるのね。
Bot	これまでには、こんな研究もあります! 「大学図書館における「脱出ゲーム」とゲーミフィケーションの可能性」 ci.nii.ac.jp/naid/120005588...
Bot	こんなまとめ記事があります! 「#Gamification Geeks 2015.12.08 :: iglobe Inc.」 matome.naver.jp/odai/214494283...
User01	脱出ゲームと図書館利用の文献面白そう!
User15	@User01 そんなのあるの?
User01	@User15 さっきあったよ
User14	@User15 @User01 プロバイダーがついてたね

表 3 失敗例のタイムライン

Table 3 Timeline of failure example

ユーザ名	ツイート
User09	面白そう
User12	個人の読書体験に基づいた書籍探索ですな
User03	宇宙といたら☆
User12	連想は人それぞれなので、それを参照する
User06	絵がかわいい
User14	特定の分野に造詣の深い人の辞書なら専門的な検索語句が得られると・・・
User08	これいいな 調べようにもキーワードわからないときとか便利そう
User12	抽象的なものだけではなく、専門的なものも見るができる
User05	自分では思いつかない単語・・・!
User15	ヒモ理論って聞いたことはあるけど実際どんなものなんだ?
Bot	これまでには、こんな研究もあります! 「ヒステリシスニューラルネットの進化的学習 (ニューロコンピューティング)」 ci.nii.ac.jp/naid/400205244 ...
Bot	こんなニュース記事もありますよ! 「あの時のステーキはうまかった…」脂っこい食品 ...- GIGAZINE」 gigazine.net/news/20090502...

行った「連想」というキーワードは、得られたスコアの順では5番目である。それより上位の部分でも、「検索語句」といった有用と思われるキーワードは存在している。しかし、今回の実装では、日本語版 Wikipedia の記事タイトルとして存在するもの以外は除外してしまっていたので、このような結果が得られていた。「ヒモ理論」や「辞書」などは一見記事タイトルとして存在していそうだが、ヒモ理論と同じ意味の「弦理論」は記事タイトルとして存在していたり、辞書は存在していないが、「辞典」や「国語辞典」なら存在していたりと、日本語版 Wikipedia の記事タイトルだけではカバーしきれていない部分もあることがわかった。

表 4 得られたキーワードとスコア

Table 4 Keyword and score

キーワード	スコア
ヒモ理論	100
造詣	98
検索語句	92
辞書	77
連想	75
読書体験	74
単語	70
書籍	66
分野	65
特定	64

6. 今後の展望

CHI 勉強会 2015 での運用実験では、キーワード抽出の際に日本語 Wikipedia の記事タイトルとして存在するもののフィルタリングをかけていなかったため、記号などを含む意味不明な節や、未知語を含むものを抽出してしまうことが多かった。一方、研究室での勉強会での運用実験ではそのフィルタリングを用いていたので、すべて意味のある、勉強会の内容に関連するキーワードを取得することができていた。このようなフィルタリングは有用であると考えられた。しかし、5章の適切なウェブページを提供することができなかった例について、「ヒモ理論」や「検索語句」というキーワードも得ることができていたが、日本語 Wikipedia の記事タイトルによるフィルタリングに弾かれてしまい、「連想」というキーワードを取得していた。現行のフィルタリングにより有用なキーワードが抜け落ちてしまうのは避けたいので、フィルタリングに別の名詞辞書を追加するなどといった手法で改善していきたい。

また、勉強会の内容とまったく関係のないウェブページを提供してしまうことも少なくなかった。なので、取得したウェブページをツイートする前に、ウェブページの内容を読み込み、本当に現在行われているオンライン議論の流れに沿っているかを判断するなどといった手法で、ツイ

トするウェブページをフィルタリングできるような工夫をしていきたい。

Twitter ボットがツイートするタイミングについても、時間間隔やユーザによるツイートの量で判断したり、ユーザによるツイートから議論の切れ目を推定し、そのタイミングでツイートを行うなど、適切なタイミングについて調査を行っていきたい。

7. おわりに

本研究では、Twitter でのオンライン議論において、議論に参加しているユーザのツイートを解析し、関連情報を提供する Twitter ボットの開発と、その運用実験を行った。その結果、グループ議論においておよそ半数の Twitter ボットによるツイートが議論に適したものであり、オンライン議論参加ユーザからの良い反応を得ることができた。しかし、ツイート解析によりあまり適切でないツイートを取得してしまったり、議論内容に沿っていないウェブページを提供してしまうことが少なくなかった。今後は、キーワードの抽出方法、適切なウェブページの選別方法を改善し、Twitter ボットがツイートする適切なタイミングについて調査を行っていきたい。

参考文献

- [1] 角康之, 西本一志, 間瀬健二. “協同発想と情報共有を促進する対話支援環境における情報の個人化.” 電子情報通信学会論文誌, no. 7 (1997): 542-550.
- [2] 赤川龍之介, 由井蘭隆也. “会議の場をリフレクションするリアルタイム会議支援システム「INGA」の提案と評価.” 研究報告グループウェアとネットワークサービス (GN), 2013-GN-86, no. 18, 1-8.
- [3] Kazutaka Kurihara. “ぴぴつい-PPTwi-” 入手先 (<https://sites.google.com/site/pptwiofficial/>) (参照 2015-11-10).
- [4] 竹内俊貴, 谷川智洋, 西村邦裕, 廣瀬通孝. “Twitter における情報の広がり可視化と分類.” 第 17 回サイバーワールド (CW) 研究会, (2010): 1-4.
- [5] 山田俊幸. “論文ったー” 入手先 (<https://twitter.com/ronbuntter>) (参照 2015-11-10).