

関連情報の提示によるオンライン議論の活性化と理解促進のための Twitter ボット

草島 将太¹ 角 康之¹

概要：近年では、研究発表会や勉強会の場において、実際に行われている発表や議論の他に、その内容についての議論を行ったり、メモを取り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 上で行われているオンライン議論の活性化や、議論参加者の理解の促進を目的とする。そこで、タイムライン上のツイートや議論内容からキーワードを抽出し、関連する話題を検索して、議論に適したウェブページをツイートの形式で提供する Twitter ボットの開発を行う。本稿では、開発した Twitter ボットを実際の勉強会の場で運用し、適切な話題を提供することができているかと、オンライン議論参加者の Twitter ボットへの反応について報告する。

Twitter Bot for Activation of Online Discussion and Promotion of Understanding by Providing Related Articles

SHOTTA KUSAJIMA¹ YASUYUKI SUMI¹

1. はじめに

学会や勉強会などといった議論（ミーティング）の場では、インターネットや SNS の普及に伴い、発表でわからない部分や気になったことをインターネットで調べながら発表を聞いている人や、SNS を用いてオンライン上で意見・感想を述べたり、他の参加者と議論をしている人もいる [1]。しかし、インターネットでの検索や SNS 上での議論を行っているうちにミーティングが進行してしまい、理解が曖昧になってしまうかもしれない。理解が曖昧なままミーティングが進行してしまうと、次の内容も理解できなかったり、議論も起こりにくくなることが心配される。

そこで本研究は、ミーティングの最中にインターネットを用いた調べ物をする手間を省くこと、そして新たな気付きを提供することで、オンライン議論参加者同士の議論を活性化させたり、理解を深めることを目的とする。そのため、オンライン議論の参加者にツイート形式で関連情報をリアルタイムに自動で提供する Twitter ボットの開発を行う。このボットは、オンライン議論中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で検索した結果を、ボットのツイートという形で議論参加者に提供する。

また、開発したボットを実際の学会や勉強会の場で運用し、ボットのツイートにより議論参加者間でのやりとりが新たに発生しているか、ボットのツイートに対する反応（ツイートへのリプライ、リツイート、いいね (Like) 登録など）があるかを観察し、ボットが会議に影響を与えていたかどうかを調査する。

このようなボットがオンライン議論に介入し、発表内容との関連が深い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、実際の発表を聞き逃すことなく関連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、理解の促進に繋がるといった効果が期待される。

2. 関連研究

2.1 グループ議論支援

本研究では、オンライン上でのグループ議論の場で、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

¹ 公立はこだて未来大学

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対話支援環境 AIDE を構築した角らの研究 [2] 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 INGA の提案と評価を行った赤川らの研究 [3] が挙げられる。

AIDE 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る。このシステムは、会話の対話構造を可視化した空間を表示することによって、会議や対話における参加者の共同発想を促進させるという特徴がある。

また INGA 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い、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。

これらは利用者が限定されたシステムであるが、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 エージェントによる情報提供

本研究では、Twitter ボットが会議におけるトピックを推定し、関連する情報を参加者に提供する。このように、ユーザ以外のエージェントがユーザに情報を提供するシステムとして、北村らはマルチキャラクターエージェントによる協調型情報検索システム [4] や、競争型情報推薦システム [5] を提案した。これらは、ユーザの必要とする情報をエージェントとのインタラクションにより学習させる必要があるが、本研究ではトピック推定の対象を個人から議論全体とし、かつこちらが意図的に情報を入力することなく情報を議論に参加しているユーザに提供することを目的としている。

2.3 ミーティングにおける Twitter の活用

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、事前に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 [6] 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、情報の広がりを見視化する「Vital Atlas」を開発した竹内らの研究 [7] 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 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 ボットに関しては、山田が「論文ったー」というボットを開発し、運用している [8]。論文ったーは、Twitter のトレンド情報をもとに、定期的に CiNii で関連のあ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。

2.4 オンラインミーティングからのキーワード抽出

本研究では、ユーザによるツイートから関連の深いウェブページを選出する必要があるため、ユーザのツイートの集合を文書とすると、その中から議論の内容にふさわしいキーワードを抽出する必要がある。

テキスト中からキーワードを抽出する手法としては、これまでに TF-IDF [9] や KeyGraph [10]、SVM による機械学習 [11]、LDA [12]、DTM [13] など多くの方法がある。また、瞬間的なバースト度が高い単語は重要であるという研究もなされている [14]。また、ウェブページのランキングアルゴリズム [15] から、単語の重要度を決定したり [16]、メイントピックを推定する応用例もある [17]。さらに、語彙連鎖構造から文書の要約を行う研究もなされている [18]。それをもとに、羽鳥らは文書からキーワードやトピックの抽出を行っている [19]。

しかしこれらの手法は、キーワードを抽出したい文書の他に別の文書を必要としたり、リアルタイム性に欠けるといった問題点がある。そのため、本研究のようにツイートから逐次キーワードを抽出する必要がある、リアルタイム性が求められる場合には不向きであると考えられる。

3. Twitter 上での議論を支援するボットの提案

本章では、Twitter ボットのシステムの流れと、開発にあたり用いた手法について述べる。図 1 は、本システムの概念図である。

3.1 Twitter ボットの概要

Twitter の用いられ方としては、会議特有のハッシュタグが設けられ、そのハッシュタグを付けてツイートする、というやり方が一般的である。そのため、まず、Twitter

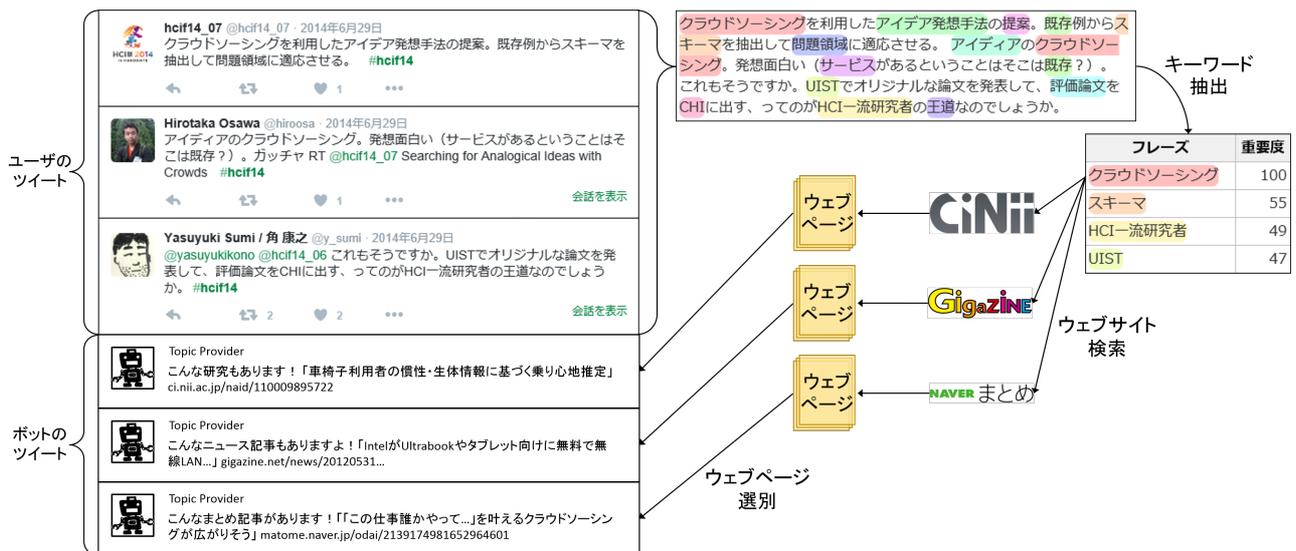


図 1 システム概念図

を用いて、会議に用いられているハッシュタグが付けられたツイートを検索する。検索により得られたツイートを順次保存していき、ある程度の量のツイートを得ることができた、または指定したタイミングになったら、そのツイートの中から会議に関連が深いと考えられるキーワードを1つ抽出する。そのキーワードを用いて論文掲載サイトをはじめとするいくつかのウェブサイトで検索を行い、検索結果として得られたウェブページのタイトルとURLをハッシュタグとともにツイートすることで、オンライン議論参加者に関連情報を提供する。

3.2 ツイートからキーワードの抽出

まず、TwitterのStreaming APIを用いて指定したハッシュタグを含むツイートを取得する。Streaming APIとは、HTTP接続をし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。

取得したツイートに関してテキスト解析を行い、キーワードを取得する。まず取得したツイートの本文から、ハッシュタグやURLなどの余分な情報を削除してテキストデータとして保存していき、保存されたテキストデータ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抽出する。

どの程度の量のツイートを1つの文書としてキーワードを抽出するかは、システムを運用する会議の参加者数や発表などの方法により、適当なものを設定する必要がある。例えばいくつかの小さなセッションに分かれていて、その区切れ目が時間やツイートによりはっきりとわかる形式の勉強会では、1つのセッションごとのツイートを1つの文書としてキーワードを抽出し、セッションの終わりにツイートしたりといったことが考えられる。それ以外の場合では、一定の時間、一定の文字数、一定のツイート数ごと

に1つの文書とするといった方法が考えられる。

キーワードの抽出方法としては、Yahoo!JAPANによって提供されるテキスト解析WebAPIの1つであるキーフレーズ抽出APIを用いた。キーフレーズ抽出APIとは、指定された文章を解析し、その文章中から特徴的な表現（キーフレーズ）を重要度（＝スコア）とともに取得することができるAPIである。

しかし、キーフレーズ抽出APIを用いた場合、記号などを含む未知語などを抽出してしまい、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。そのため、キーワードに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のうち、スコアがもっとも高かったものとする。

日本語版Wikipediaのタイトルの一覧は事前にデータベースにまとめておき、得られたキーフレーズをスコアの高い順に確認し、データベースに存在するものが見つかるまで繰り返す。

3.3 ウェブサイトでの検索とツイート

抽出されたキーワードを用いて、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つのウェブサイトで検索を行う。今回は、議論の中で行われている話題について、関連する過去の研究論文を提供するためにCiNiiを、そして論文にはなっていないまでも、関連する最近の出来事を知ることを促すためにGigazineとNAVERまとめを用いることとした。

ウェブサイトでの検索には、そのウェブサイトが提供している検索APIを利用するか、APIが提供されていないものについては、Microsoft Azureが提供するBing Search APIを利用する。

検索結果として得られるのは関連度順、日付順などが考えられるが、得られた最上位のウェブページが会議の内容に関連したものであるとは限らない。そこで、それぞれのウェブサイトから日付順に最大 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 tf-idf[9] とコサイン類似度推定法 [20] を用いた類似度を求め、それぞれのウェブサイトでもっとも類似度が高かったものを選出する。

この手法を取ることで、スコアがもっとも高かったキーワードのみに依存した検索結果になってしまうことを避け、キーワードとして抽出されなかったが特徴的であり、議論の内容に関連の深い他の語からも、より議論の内容に近いウェブページを選出することができると考えられる。

得られた検索結果について、その記事のタイトルと URL をボットのツイートでオンライン議論に参加しているユーザに提供する。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こととし、ウェブサイトの検索結果が 1 つもなかった場合は、そのウェブサイトについてのツイートはしないこととする。また、キーワード抽出の段階で 1 つもキーワードが得られなかった場合や、得られたキーワードがすべて日本語版 Wikipedia のタイトルとして存在しないものであった場合は、ウェブサイトによる検索も行わず、ツイートもしないこととする。

4. ライトニングトーク形式の勉強会での運用実験と考察

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。以下、実験の概要と結果について述べ、考察を行う。

4.1 CHI 勉強会 2015 について

CHI 勉強会 2015 とは、CHI2015 で発表された論文について、1 本の論文につき 30 秒で次々と発表していく勉強会である。勉強会は 3 本から 6 本の論文から成る 114 のセッションに分かれており、1 つのセッションを 1 人が担当し発表した。この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 会場で開催され、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開始のタイミングでは、論文のタイトルと内容紹介についてのツイートが情報提供用アカウントによりされていた。

この勉強会は、東京と北海道という 2 ヶ所の離れた場所で開催された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。

4.2 実験の概要

この実験では、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 Twitter ボットによるツイートの形でオンライン議論に参加しているユーザに提供した。勉強会に参加しているユーザは共通のハッシュタグを用いて Twitter でメモを取ったり、意見をツイートしていた。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、この運用実験ではその部分は実装しておらず、キープレーズ抽出 API によって得られた結果をそのまま使用し、もっともスコアの高かったものをキーワードとしている。また、ウェブサイトによる検索について、「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 とコサイン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」としたが、この運用実験では実装しておらず、検索結果で得られた最上位のウェブページを選出した。

4.3 実験の結果

勉強会の計 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、キーワードを抽出することが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、有用なキーワードであると考えられるものを得ることができたのは 47 であった。それ以外のセッションでは、「ユーザ」「画面」などのありふれていて、特徴的でないものが含まれてしまっていたり、記号が含まれてしまっているものであった。

4.3.1 適切なウェブページを提供することができた例

表 1 は、この勉強会において Twitter ボットが有用であると考えられるキーワードを抽出し、適切なウェブページをオンライン議論参加者に提供することができた 1 つの例である。ユーザ名のうち、hchokkaido は情報提供用アカウントであり、bot は今回開発した Twitter ボットを示している。なお、すべてのツイートには勉強会のハッシュタグが含まれていたが、キーワード抽出の際には用いないことと、簡略化のため、表への記載を省略した。

このセクションでは、FingerReader と呼ばれる、視覚障がい者が文章を読むのを支援する指装着型デバイスの紹介がなされていた。キーワード抽出では「ActiveReading」を取得することができており、2 つのウェブサイトでの検索結果を提供することができていた。

4.3.2 勉強会を通しての結果

予想としては、ボットによるツイートに興味を持った

表 1 ボットにより有用な情報提供が行われた例 1 (勉強会のタイムラインの一部)

ユーザ名	ツイート
user01	ベルトの UI 化とは、007 っぽいですね
hchokkaido	RegionSpeak: Quick Comprehensive Spatial Descriptions of Complex Images for Blind Users http://t.co/OS6FVB0C2S
user02	体験してみたい https://t.co/rkSmh1G6eq
hchokkaido	FingerReader: A Wearable Device to Explore Printed Text on the Go http://t.co/Cmec3onksU
hchokkaido	Collaborative Accessibility: How Blind and Sighted Companions Co-Create Accessible Home Spaces http://t.co/nQ9sFbakwQ
user03	@hchokkaido active reading を強化するデバイス。良いアイデア。
bot	CiNii で「active reading」を検索しました。検索結果：自然言語量子化を含む文の曖昧性について http://t.co/ywcZX7y2fl
not	Gigazine で「active reading」を検索しました。検索結果：Scientists revealed how brain wor… http://t.co/qhKdBTOBSA

ユーザがお気に入り登録やリツイートをしたり、ボットによるツイートが原因となる議論が起こると考えられた。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、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。図 2 は、ユーザのツイートの傾向を示している。

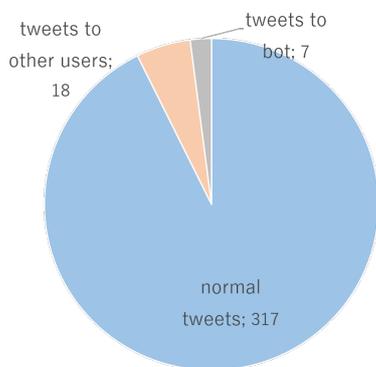


図 2 Tendency of users' tweets

4.4 考察

キーワードの抽出に関しては、全体の 47% でしか有用であると考えられるキーワードを抽出することができなかった。そのため、単純に Yahoo! キーフレーズ解析を用いるだけではなく、なんらかの工夫をすることが必要であると考えられた。

また、ライトニングトーク形式の勉強会での運用実験では、Twitter ボットのツイートに対するオンライン議論に参加しているユーザの反応はほとんど得られなかったが、タイムラインの分析から、議論に参加しているユーザの 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。理由としては、特有のハッシュタグを用いていたことにより、議論に参加しているユーザがフォロー・フォロワー関係にないより広い範囲のユーザのツイートを見ることができたというメリットがある一方で、ツイートの内容を確認はするものの、知らないユーザとオンライン上で直接議論を交わすことに抵抗があったのではないかとすることも考えられた。

Twitter ボットへの反応が少なかった理由としては、上記のようにユーザの Twitter の利用方法が議論ではなくメモのようなものとなっていることに加えて、この勉強会で Twitter ボットがツイートした内容が、「(ウェブサイト)で(得られたキーワード)を検索しました。」という文章とともにウェブページのタイトルと URL をツイートするだけの無味乾燥なものであったことや、キーフレーズ抽出で得られたキーワードをそのまま使ってしまい、意味不明な単語で検索してしまったことで、全く関係のない内容のツイートも多かったことから、議論参加者の興味を引くことができなかったことが考えられた。

5. グループミーティングでの運用実験

CHI 勉強会 2015 での結果を受けて、キーワードの抽出方法やウェブページの選別方法を工夫し、研究室での勉強会で Twitter ボットの運用実験を行った。以下、この勉強会の運用実験の概要と、特徴的な結果が得られたものについて述べる。

5.1 実験の概要

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。

5.2 実験の結果

運用実験を行った結果、特徴的な結果を得られたものについて、そのときのタイムラインとともにまとめる。

5.2.1 適切なウェブページを提供することができた例

表 2 は、適切なウェブページを提供することができた例である。

表 2 ポットにより有用な情報提供が行われた例 2 (グループミーティングのタイムラインの一部)

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発化ですな
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な？
user10	こんな先行研究あるのね。
bot	これまでは、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としてたね

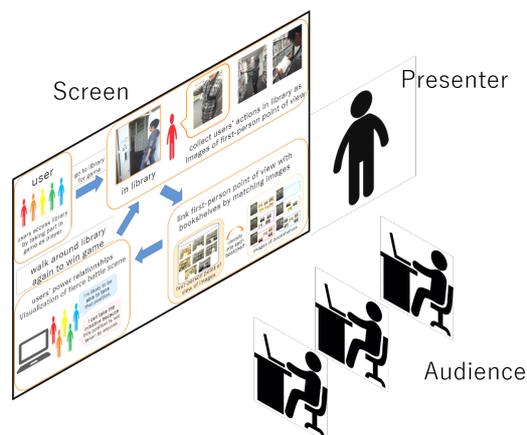


図 3 グループミーティングの様子

表 2 のような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ることを目的とした研究の概要が紹介され

ていた。図 3 は、そのときの勉強会の様子である。ポ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り、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。表 3 は、表 2 の間のツイートを文書とし、キーフレーズ解析を行った際の結果をまとめたものである。スコアがもっとも高かった単語は「ゲーミフィケーション」であり、この単語は日本語版 Wikipedia のタイトルとして存在しているので、この単語がキーワードとして選出された。

表 3 キーフレーズ解析による結果

キーワード	スコア
ゲーミフィケーション	100
図書館利用⇔ゲーミフィケーション	44
実世界	40
陣取り	31
ゲーム	26
目的	24
先行研究	22

また、表 4 は、CiNii で「ゲーミフィケーション」を検索した結果と、得られたウェブページとツイート内容とのコサイン類似度を計算した結果をまとめたものである。得られた 10 のウェブページのうち、図書館でのゲーミフィケーションという勉強会の内容に深く関連したウェブページを選別することができていた。このことから、ツイートと記事内容とのコサイン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た。

このツイートに興味を持ったユーザも多く、5 人のユーザがいいね (Like) 登録していた。さらに、ポットのツイート後、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 (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。

5.2.2 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 5 は、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

表 4 CiNii による「ゲーミフィケーション」の検索結果とツイート
内容とのコサイン類似度

記事タイトル	コサイン類似度
過去の写真を利用する日記スタイル単語学習アプリケーション W-DIARY の開発 (教育工学)	0.0847
すれ違い通信を利用したキャラクター育成アプリの実証実験 (ネットワークシステム)	0.057
ゲーム化によるコミュニケーション支援 (マルチメディア・仮想環境基礎)	0.0544
ゲームの手法を用いた情報リテラシー教育の可能性: 海外の事例を中心に	0.0444
ゲーミフィケーションを活用した小1向けアプリ教材が教室に与える影響 (教師教育・教育の情報化/一般)	0.0365
from NTT データ ビジネス現場でのゲーミフィケーション活用技術と検証	0.0134
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	0.2426
話し合いとネゴシエーションを通じたアクティブラーニング-大学教育を題材として-	0.012
ゲーム要素を取り入れた授業デザイン枠組の開発と実践 (教育実践研究論文)	0.0509
部分集合から算出した順位の提示効果 (ライフインテリジェンスとオフィス情報システム)	0.0089

表 5 ボットによる情報提供が失敗した例

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	宇宙といったら☆
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ことができる
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...
user07	ステーキ食べたい

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う結果が得られた。

5.3 考察

この勉強会での運用実験の結果を受けて、参加者が互いに知り合いである研究室内での勉強会を行ったところ、興味深い他のユーザのツイートには気軽にいいね (Like) 登録やリプライを送るといったことが見て取れた。また、表2のタイムラインの例から見て取れるように、Yahoo!JAPANのキーワード抽出 API によるツイート内容からのキーワード抽出と、コサイン類似度比較によるウェブページの選別で、ある程度良い結果は得られていた。

また、表5のタイムラインでボットがツイートした内容の結果から、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、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。

6. Twitter ボットの継続的な利用

第6章での運用実験により、勉強会に関連の深かった記事を Twitter ボットが提供することができ、それをもとに話題が生まれたり、参加しているユーザの気付きに結びつくような結果を得ることができた。しかし、まだ参加したユーザが Twitter ボットの存在に不慣れで、提供された記事をあまり見ず、議論にも強くは結びつかなかったのではないかと考えられた。そこで、開発した Twitter ボットを継続的に運用した場合のオンライン議論に参加しているユーザの反応について調査した。以下、実験の概要と、実験により得られた特徴的な部分についてまとめ、考察する。

6.1 実験の概要

第5章で述べた勉強会に参加したユーザのうち、5人のユーザが Twitter による議論を行った。参加したのは学部4年生5人で、各自の卒業研究についての議論を行い、そのツイートをもとに Twitter ボットが動作するようにした。

今回の運用では、ユーザのツイートがある程度されたらシステムがキーワードの抽出を行い、ウェブサイトを検索するようにした。具体的には、ユーザのツイート本文のうちハッシュタグやリプライ先の ID などの余分な情報を抜いた部分のみをテキストとして保存・追加していき、テキスト全体を UTF-8 に変換した結果が 2000 字を超えた場合にキーワードの抽出を行うようにした。時間などの区切りではなく文字数にした理由としては、今回の議論の形式が、誰かが1人ずつセッション区切りや決められた時間で発表して、その内容について他のユーザがツイートする、といったものではないことが挙げられる。また、文字数を UTF-8 に変換したものを基準とした理由としては、Yahoo!キーワード抽出 API を実行する際に、検索クエリを UTF-8 に変換する必要があったからである。

6.2 運用実験の結果

主にそれぞれのユーザの卒業研究について、特に卒業論文とそれに伴う実験や結果についての議論がなされていた。表 6 は中でも特徴的だった部分のタイムラインである。

表 6 Twitter ボットを継続利用した時のタイムライン

ユーザ名	ツイート
user04	ぼくを苦しめるもの。それが onomatopoeia
user01	onomatopoeia (擬声語)
user02	結局オノマトペの実験の結果はどうなったんや？
user05	onomatopoeia ってつい発音したくなるよね
user04	onomatopoeia は僕を苦しめます
user05	おのまとぺいあー
user01	オノマトペは Wikipedia にないけど、擬声語は Wikipedia にある
user04	onomatopoeia はね 一応環境によって同じ動きでも変化が見られたよ
user04	あとは、キャラの属性とどう結びつけようか
user05	視線計測は結局悩んだ時とそうでない時の時間の長さくらいしか解るところなかったな〜
user02	@user04 結局そこで落ち着いたの？
bot	こんな研究もあります！「保育内容「表現」(身体表現)の指導と内容_動きとイメージをつなげる言葉かけについて」 ci.nii.ac.jp/naid/400188785 ...
bot	こんなニュース記事もありますよ！「オノマトペの用例を自動収集したオンライン辞典 ...- GIGAZINE」 gigazine.net/news/20080229 ...
bot	こんなまとめ記事があります！「【料理・食材編】なるほど、意味がわからん...となってしまうツイッターの bot たち【料理・食材編】」 matome.naver.jp/odai/214089614 ...
user05	あれ、既に研究されてた...？
user01	オノマトペの辞典あるやんけww
user04	これは、用例を載せてあるだけで、本研究は実際の動きのデータからそれがちゃんといえるかを分析しています。たぶん

user04 によるオノマトペ (擬声語) についての研究と、その実験についての議論がなされている。このタイムラインでは触れられていないが、研究とは、オノマトペと動作を結びつけ、どのような動作をしているときにはどのようなオノマトペが用いられるのかを分析するというものである。キーワードとして抽出されたものは「擬声語」で、3つのウェブサイト全てで検索結果を得ることができていた。その中でも Gigazine による検索で得られた記事について user01 と user05 が反応し、そのツイートについて user04 が記事の内容と自身の研究についての違いについて説明していた。

6.3 考察

今回の運用実験では、第 5 章で述べたような勉強会のと

きよりも、ユーザ間のやりとりが多く、Twitter ボットによる関連記事のツイートについての発言も多かった。理由としては、以前と比べてオンライン上でグループ議論を行うことにユーザが慣れたという点や、Twitter ボットがどのようなツイートをしてくれるのかがわかってきたという点、前回のように時間区切りの発表ではなく、時間に自由があって Twitter ボットが提供した記事を読覧することができた点が考えられた。オンライン議論の後、参加していたユーザに聞いてみると、「以前よりも (Twitter ボットがツイートした) 記事をよく見ることができた」との意見を得ることもできた。今回の実験からは、回数を重ねることでこの形式での勉強会や議論に慣れ、Twitter ボットが提供したウェブページを読む余裕が生まれれば、さらなる議論の発展や理解の促進に繋がるのではないかということが予想された。

また、1つのキーワードにつき3つすべてのウェブサイトによる検索結果をツイートされても、時間的にすべてを見る余裕がないとの意見を得ることもできた。このことから、得られたものの中からさらに1つか2つ程度に選別することや、より短時間でユーザに内容を把握してもらうことができるような情報提供の方法を考える必要があるということが、今後の課題として判明した。

7. おわりに

7.1 まとめ

本研究では、学術的な会議や研究発表会、勉強会などといった場において、実際の議論の場と並行して行われるオンライン上での議論について、オンライン議論に参加しているユーザ同士の議論の誘導、活性化や、理解の促進を目的として、関連情報をツイートする Twitter ボットの開発を行った。

手法としては、会議特有に用いられているハッシュタグが付けられたツイートをリアルタイムで取得し、その中から特徴的であると考えられる単語をキーワードとして抽出、そのキーワードを用いていくつかのウェブサイトで行い、検索を行い、議論の内容にもっとも関連の深い記事やウェブページを tf-idf による重み付けとコサイン類似度推定法を用いて選別し、得られたウェブページのタイトルと URL を Twitter ボットがツイートする、という流れである。この Twitter ボットを用いていくつかの勉強会の場で運用実験を行い、適切なウェブページを提供することができているかと、オンライン議論に与える影響について考察を行った。

分析の結果としては、知らない人も参加しているようなオンライン議論の場では参加者同士の議論はあまり見られず、自身のメモ代わりとして Twitter が用いられていたことがわかり、議論を発生、活性化させるには至らなかったが、参加者がお互いに知っている人同士である場では、ある程度の議論が見られ、Twitter ボットによるツイートが

新たな気付きを生み出したり、議論を起こしたりといった結果が見て取れた。

7.2 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。このことから、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うような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うような手法で、ツイー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。

参考文献

- [1] Parra Denis, et al. "Twitter in academic events: a study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences." *Computer Communications* 73 (2016): 301-314.
- [2] Yasuyuki Sumi, Kazushi Nishimoto, and Kenji Mase. "Facilitating human communications in personalized information spaces." *AAAI-96 Workshop on Internet-Based Information Systems*. 1996.
- [3] 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.
- [4] Yasuhiko Kitamura, et al. "Multiple character-agents interface: An information integration platform where multiple agents and human user collaborate." *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2*. ACM, 2002.

- [5] Yasuhiko Kitamura, Toshiki Sakamoto, and Shoji Tatsumi. "A competitive information recommendation system and its behavior." *Cooperative Information Agents VI*. Springer Berlin Heidelberg, 2002. 138-151.
- [6] Kazutaka Kurihara. "びびつい -PPTwi-." <https://sites.google.com/site/pptwiofficial/en>. 2016-04-18.
- [7] 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.
- [8] Toshiyuki Yamada. "論文ったー." <https://twitter.com/ronbunter>. 2016-04-18.
- [9] Gerard Salton, and Michael J. McGill. "Introduction to modern information retrieval." McGraw-Hill, Inc. (1986).
- [10] Yukio Ohsawa, Nels E. Benson, and Masahiko Yachida. "KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor." *Research and Technology Advances in Digital Libraries*, 1998. ADL 98. *Proceedings. IEEE International Forum on IEEE*, 1998.
- [11] Tsutomu Hirao, et al. "Extracting important sentences with support vector machines." *Proceedings of the 19th international conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 2002.
- [12] David M Blei., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.
- [13] David M Blei., and John D. Lafferty. "Dynamic topic models." *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006.
- [14] Jon Kleinberg. "Bursty and hierarchical structure in streams." *Data Mining and Knowledge Discovery* 7.4 (2003): 373-397.
- [15] Sergey Brin, and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine." *Seventh International World-Wide Web Conference*. 1998.
- [16] Wei Wang, Diep Bich Do, and Xuemin Lin. "Term graph model for text classification." *Advanced Data Mining and Applications*. Springer Berlin Heidelberg, 2005. 19-30.
- [17] Mario Kubek, and Herwig Unger. "Topic detection based on the PageRank's clustering property." *IICS 11* (2011): 139-148.
- [18] Gonenc Ercan, and Ilyas Cicekli. "Using lexical chains for keyword extraction." *Information Processing & Management* 43.6 (2007): 1705-1714.
- [19] Jun Hatori, Akiko Murakami, and Jun'ichi Tsujii. "Multi-topical discussion summarization using structured lexical chains and cue words." *Computational Linguistics and Intelligent Text Processing*. Springer Berlin Heidelberg, 2011. 313-327.
- [20] 北研二, 津田和彦, 獅々堀正幹. *情報探索アルゴリズム*, 共立出版, 2002 年.