

SNS・Web 間のシームレスなコミュニケーションシステムの提案

A Simultaneous User Communication System between Microblogs and Web Pages

安井 豪基[♡] 王 元元[◇] 細川 侑士[♡]
河合 由起子[◇] 秋山 豊和[♡] 角谷 和俊[▲]

Goki YASUI Yuanyuan WANG
Yuji HOSOKAWA Yukiko KAWAI
Toyokazu AKIYAMA Kazutoshi SUMIYA

本研究では、ツイートユーザと Web 閲覧ユーザの各々が利用しているメディア (Twitter と Web) を通して、リアルタイムにコミュニケーションが可能なシステムの構築を目指す。具体的には、ツイートの発信場所とページの内容 (場所名) を抽出することで、ツイートと Web ページを対応付ける。対応付けられたツイートはページ上に提示される。これにより、Web 閲覧ユーザは、ページを閲覧しながら内容に関わる場所で発信されたその場所に関するツイートを閲覧できる。また、Web 閲覧ユーザは発信したツイートに対して、メッセージを発信できる。発信されたメッセージは、ツイートユーザに対して、ツイートメッセージとして提示される。本論文では、位置情報に基づいたツイートと Web ページ関連付け手法ならびに Twitter と Web のシームレスなリアルタイム通信システムの構築手法について述べ、実装システムを検証する。

We developed a system to support simultaneous communication between microbloggers (Twitter users) and Web users in real time through both the contents of microblogs (tweets) and Web pages. The system provides a question answering interface attached to Web pages, which allows Web users to chat with each other while browsing the same Web pages, or Twitter users whose tweets are presented with Web pages. To achieve this, the system extracts relationship between heterogeneous media on tweets and Web pages by generating queries based on posted information. Thus, the system can effectively present tweets from the Twitter users to help the Web users easily

- ♡ 非会員 京都産業大学
i1458085@cc.kyoto-su.ac.jp
- ◇ 正会員 京都産業大学
yuanw@cc.kyoto-su.ac.jp
- ♡ 非会員 京都産業大学
i1358103@cc.kyoto-su.ac.jp
- ◇ 正会員 京都産業大学
kawai@cc.kyoto-su.ac.jp
- ♡ 非会員 京都産業大学
akiyama@cse.kyoto-su.ac.jp
- ▲ 正会員 兵庫県立大学
sumiya@shse.u-hyogo.ac.jp

grasp the current situation in real time, and it also can effectively present messages from Web users to help Twitter users immediately obtain useful information or knowledge. In this paper, we discuss our prototype system and evaluate its effectiveness.

1. はじめに

近年、Twitter¹や Facebook²に代表されるソーシャルネットワークサイトを通して、オンライン上でのユーザ間のコミュニケーションを中心としたソーシャルネットワークサービス (以降、SNS と記す) が普及している。また、スマートフォンなどの端末の発達により、場所を問わず任意の時間に情報を発信することができるため、実世界においてスポットに関してリアルタイムに情報の発信を行っている SNS ユーザも多い。それら SNS ユーザのスポット等の場所に関して発信される情報は、リアルタイムでその場所の状況を把握するために役立てることができる。しかし、Twitter にリアルタイムで随時発信されているツイート数は膨大であり、関心のあるトピックのツイートを取得することが難しい。また、ユーザ間のコミュニケーションにおいても、その SNS 内のコミュニティ (フォロワーとフォロー) に限定される場合もあるため、情報の網羅性が低くなってしまふ。例えば、ツイートの内容やハッシュタグでの検索により関連するツイートを取得する手法 [1] では、実空間においてその場所にはない SNS ユーザのツイートも検出する。また、[2] では、位置情報付きツイートをを用いて、緯度と経度を指定し、その場所付近で発信された関連性の高いツイートを取得できる。しかしながら、その場所で発信されたツイートの中で、その場所に関する内容とは無関係のツイートも検出されてしまふ。このように、適切なツイートの取得は手間や時間がかかってしまふ問題点と、網羅性の向上が課題として挙げられる。

そこで、本研究では、実空間で Twitter を用いて情報発信を行っているユーザ (ツイートユーザ) と、実空間のある場所に関連する Web ページを閲覧しているユーザ (Web 閲覧ユーザ) に対して、実空間においてツイートユーザが発信したツイートの位置情報と Web 閲覧ユーザが閲覧している Web ページの内容を対応付けることで、SNS と Web ページといった異なるメディアを利用しているユーザ同士のリアルタイムでコミュニケーション可能なシステムの構築を目標とする。これにより、場所に関して発信されるツイートの効率的な取得と、ツイートユーザ間だけでなく、Web ユーザとも情報提供、取得ができ、網羅性の向上にも繋がる。具体的には、以下の 2 点を実現する。

- 異種コンテンツ間 (ツイート、Web ページ) の関係性抽出
- ツイートユーザと Web 閲覧ユーザのリアルタイムコミュニケーションシステムの構築

ツイートユーザと Web 閲覧ユーザのコミュニケーションが可能になることで、ツイートユーザは他のツイートユーザだけでなく、Web 閲覧ユーザへの問い合わせも可能となり、情報の網羅性の向上が期待できる。また Web 閲覧ユーザは、Web ページに関連する場所についての感想や混雑具合といった情報をツイートユーザから取得でき、リアルタイムに把握することができる。そのため、提案システムでは、位置情報付きツイートを随時取得し、位置情報からその場所に関連する Web ページを閲覧している Web 閲覧ユーザのページ上にリアルタイムで取得した関連ツイートを提示する。Web 閲覧ユーザがメッセージを発信すると、該当 Web ページを閲覧している他の Web 閲覧ユーザに発信され、ツイートユーザにはメッセージがツイートとして提示される (図 1)。なお、ツイートユーザは本研究で提供するサービスのアカウントのフォロワーになる必要がある。

¹<https://twitter.com/>

²<http://www.facebook.com/>

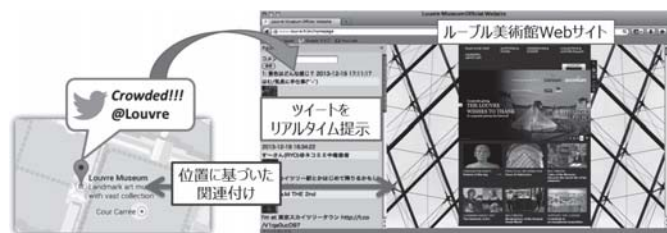


図 1: システム概要図

本論文では、位置情報に基づくツイートおよび Web ページの関連付け手法ならびにリアルタイム通信システムの構築について述べる。また、提案システムを実装し、システム動作について考察する。本論文の構成は以下のとおりである。次章で提案システムの概要を説明し、3章で位置情報付きストリーミングツイートデータの分析手法およびリアルタイム質問応答通信の構築法について述べる。4章で実装したシステムを検証し、5章で関連研究について述べた後、最後に、6章で本研究のまとめと今後の課題と展開について述べる。

2. システム概要

本研究は、場所に関するツイート情報の取得ならびに、ツイートユーザと Web 閲覧ユーザが場所情報に基づきリアルタイムに通信可能なシステムの構築を目指す。

図 1 にシステムの概要を示す。ツイートを発信すると、ツイート閲覧者だけでなく、ツイートの内容と発生場所の位置情報に基づき、関連する Web ページを検出し、それらを閲覧しているユーザのページ上にそのツイートがリアルタイムに提示される。Web 閲覧ユーザにとってはそれら提示されたツイートを閲覧することで、場所に関する現状把握の支援になる。各ツイートに対して返信する場合は、該当するツイートを選択することで、個別に返信できる。ただし、ツイートユーザは本サービスをフォローしている必要がある。

一方、Web 閲覧ユーザが本システムの入力ボックスを用いて情報発信すると、ページを閲覧している他の Web 閲覧ユーザにメッセージが送信される。また、本サービスをフォローしているツイートユーザに対してもツイートとして提示できる。ツイートによる返信は、前者の問合せ同様にツイートだけでなく Web ページ上にも提示され、全てのページ閲覧ユーザはこれらの問合せと返信を閲覧できる。

なお、全てのメッセージは WebSocket サーバを経由するため匿名性が保たれる。図 1 では、ルーブル美術館にいるユーザがツイートを発信した場合に、そのツイートがルーブル美術館のページと関連付けられ、Web ブラウザに提示されている。Web 閲覧ユーザは混雑具合やルーブル美術館の感想など状況に関する問合せができ、一方でツイートユーザはルーブル美術館の開館時間など、その時その場では確認しにくい情報に関してツイートによる問合せが可能になる。

3. 位置情報に基づくツイートと Web ページ間リアルタイムメッセージ通信

図 2 に処理の流れを示す。本研究では、Web 閲覧ユーザとツイートユーザとを Web ページとツイートを通してリアルタイム通信可能にするため、リアルタイムに送信されるツイート（以下ストリーミングツイートとする）ならびに Web ユーザがアクセスしている Web ページの URL を取得する。サーバはツイートユーザが発信したストリーミングツイートを取得し、位置情報に基づいて関連するページを取得し、対応付けの管理を行う。取得した関連ページに Web 閲覧ユーザがアクセスすると、対応するツイートを抽出し、ブラウザへ送信および提示する。なお、ツイートユーザは本サーバからメッセージを受信する際には、本サービスのア

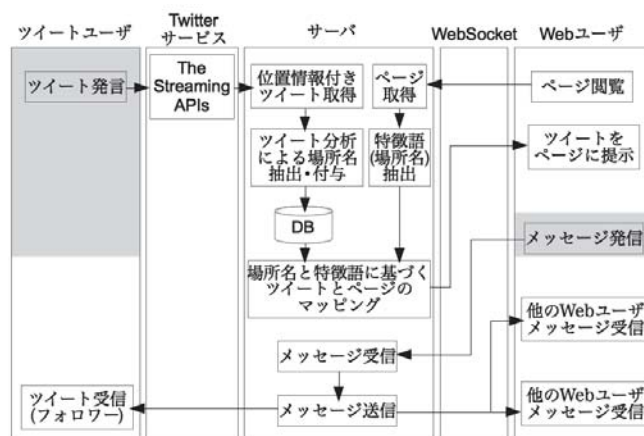


図 2: システム構成図

カウント¹のフォロワーとなっている必要がある、Web 閲覧ユーザは提案システムとなるアドオンを用いる必要がある。

Web 閲覧ユーザがアドオンの入力ボックスにメッセージを入力すると、サーバが受信し、同じページを閲覧しているユーザのブラウザへ送信する。また、Web ページに提示されている各ツイートにのみ直接返信も可能である。直接任意のツイートに送信する場合、サーバはブラウザからメッセージを受信し、ツイートしたツイートユーザ（フォロワー）へ送信する。

全ての送受信はサーバを介するため、匿名性が保たれる。

3.1 ストリーミングツイートデータ取得

本論文では、位置情報に基づく問合せを目的としており、ページとツイートを位置情報に基づき関連付ける。そのためまず、指定地域から重複を除いた緯度経度情報を含むストリーミングツイートを The Streaming APIs を用いて取得する。指定地域は、1度以上異なる南西および北東を指定することで、その 2 点に囲まれた矩形領域のストリーミングツイートを取得できる。次に、取得したストリーミングツイートの緯度経度情報から、Google Place API version 3²を用いて、半径 d m の場所名を取得した。評価実験では、取得した場所名は関連する Web ページ取得の際に検索キーワードとして用いられることと、ツイート発信ユーザの移動も考慮し、 $d=5$ とした。また、ツイート内容を形態素解析し、名詞となる単語を取得する。

以上より、ツイートユーザ id、アイコン画像 URL、緯度、経度、場所名、ツイート内容、単語集合、取得時刻を一定時間管理する。

3.2 ツイートの緯度経度と内容に基づくツイート選別

前節より取得したストリーミングツイートに対して位置情報に基づいた内容判定を行い、ページと関連付ける。ツイートが発信された場所名と関連するかをツイートの内容から判定することで、ツイート発生場所と関係性の低いツイートの除去を行う。

位置情報に基づいたツイート内容判定法は、一定範囲内の一定時間のツイートに多く出現する単語は関連性が高いと考え、場所名に対する特徴語として抽出する。この特徴語を多く含むツイートを場所名に関連するツイートとして選択する。まず、取得したツイート t の位置情報より、半径 d 内に存在する一定時間内のツイート n 個を取得する。次に、下記の式によりツイートの重要度を算出する。まず、ツイート t に出現する各単語 i のツイートに出現する頻度を抽出し、その平均値を算出する。また、特徴的な単語が出現しても単語数が多い場合は、ツイートの重要度が低下するため、シグモイド関数を用いることで、出現頻度の高い単語

¹<https://twitter.com/@RtQAService>

²<https://developers.google.com/place/>

には、さらに重要度の重みを増やす方法を取ることにした。

$$\sum_{i=1}^m \left(\frac{\text{単語 } i \text{ が出現するツイート数}}{\text{ツイート総数 } n} \times \frac{1}{1+e^{-x}} \right) \times \frac{1}{m} \quad (1)$$

m はツイート i に出現する単語総数である。 x は以下の式で求まる単語 i の DF 値である。

$$x = \frac{\text{単語 } i \text{ が出現するツイート数}}{\text{ツイート総数}}$$

最後に、閾値以上のツイート t を位置情報に基づいたツイートとする。

3.3 Web ページの場所名抽出

まず、Web 閲覧ユーザの閲覧している Web ページの URL を取得し、その Web ページのスニペットを取得する。次に、スニペットから出現頻度の高い単語を特徴語として抽出する。また、形態素解析よりその特徴語の中から地名を判別し、該当する単語をそのページの場所名とする。尚、複数地名が抽出された場合は全てを場所名とする。

3.4 場所名に基づく Web ページとツイートの対応付け

3.3 節より Web 閲覧ユーザの閲覧している Web ページの場所名が抽出された。また、3.1 節より、ツイートユーザの位置情報付きツイートを The Streaming APIs を用いて取得し、緯度経度から場所名を取得して、さらに、3.2 節では場所に関連するツイートを選択した。ユーザが Web ページを閲覧すると、場所名から関連するツイートを検索し、Web 閲覧ユーザに提示する。ツイートユーザには、緯度経度情報から場所名を抽出し、その場所名と一致する Web ページを対応づける。なお、DB には取得したツイートおよび抽出した場所名を格納する。これらのツイートと Web ページを場所名に基づき、対応付ける。

3.5 リアルタイム双方向通信

リアルタイム問い合わせシステムを構築する上で、Ajax や Comet, WebSocket といった様々な双方向通信手法が存在する。先行研究として、これまで我々は Web サーバと Web ブラウザ間の通信のための双方向通信として Ajax や Comet を用いてきた [3] が、本研究では、より通信ロスの少ない WebSocket⁵を用いる。

ツイートユーザから Web 閲覧ユーザへの配信では、ツイートユーザの位置情報付きツイートを The Streaming APIs を用いて取得し、そのツイートの緯度経度情報から場所名の付与を行い、DB にツイートと場所名を格納する。3.3 節より、Web 閲覧ユーザが閲覧している Web ページの場所名から、場所名とマッチするツイートを DB に問い合わせ、該当するツイート情報を Web 閲覧ユーザに提示する。

Web 閲覧ユーザからツイートユーザへの配信では、3.4 節より、Web 閲覧ユーザは Web ページ上に提示されたツイートに対してサーバを経由し、ツイートとしてメッセージを送信することが可能である。

4. 実装および検証

本研究では、Web 閲覧ユーザとツイートユーザとを Web ページとツイートを通してリアルタイム通信可能なシステムの構築を目的としている。本章では、実装による評価としてプロトタイプを構築し、リアルタイム通信機能の評価ならびにページに提示されるツイートの評価実験を行う。

本プロトタイプでは、サーバは、Apache httpd 2.4, java, php5.5 を用いた。クライアント側は、Javascript を用い、Firefox

⁵<http://gihyo.jp/dev/feature/01/websocket/0001>
<http://www.atmarkit.co.jp/ait/articles/1111/11/news135.html>



図 3: Web 閲覧インタフェース

23.0.1 上で動作確認した。なお、2013 年 9 月 3 日から同年 12 月 14 日までに The Streaming APIs で日本全国のツイートデータを取得した。今回の検証では、上記のツイートデータから平日の 5 日間のツイートデータを使用した。

なお、比較対象として、本手法で用いるデータと同じツイートにおいて、各々の場所名でハッシュタグ検索を行った場合の結果を用いる。

4.1 実装

ストリーミングツイート取得は、指定地域から重複を除いた緯度経度情報を含むストリーミングツイートを The Streaming APIs version 1.1³を用いて取得した。

指定地域は、1 度以上異なる南西および北東を指定することで囲まれた矩形領域のストリーミングツイートを取得できる。評価実験では、関東地方を対象とし、南西 132.2, 29.9, 北東 146.1, 46.20 とした。取得したストリーミングツイートの緯度経度から、Google Place API version 3⁴を用いて、半径 dm 以内の場所名を取得した。評価実験では、取得した場所名は関連する Web ページ取得の際の検索キーワードとして用いられることと、ツイート発信ユーザの移動を考慮して、事前実験より $d=5m$ とした。

サーバとクライアント間のメッセージ通信は、ツイートユーザと Twitter サーバ、Twitter サーバと本システムとし、Twitter サーバと本システム間で通信することで、ツイートユーザに対する受信のインタフェースは既存の Twitter サービスが利用可能となる。Web 閲覧ユーザと本システム間は WebSocket 通信を行い、アドオンを用いた質問応答インタフェースを構築した。

図 3 に、実装したプロトタイプによる Web 閲覧ユーザの表示例を示す。スカイツリーの公式サイトトップページと、ストリーミングツイート（トップページの左横）の提示結果である。ユーザはスカイツリー近辺でリアルタイムに発信されたスカイツリーに関するツイート内容を閲覧することが可能である。なお、これらツイートを各ページごとにログファイルとして保存することで、過去のツイートも閲覧可能である。

Web 閲覧ユーザは、入力ボックスにメッセージを入力し、他の Web 閲覧ユーザに一斉送信可能である。また、任意のツイートを選択することで、特定のツイートユーザへのみメッセージ送信可能である。ただし、メッセージ受信するツイートユーザは本サービスのフォロワーとなっている必要がある。

プロトタイプでのストリーミングツイートのメッセージ提示は 1, 2 秒程度で、リアルタイム通信を確認できた。また、Web 閲覧ユーザのメッセージ提示も同程度で、リアルタイム通信を確認

³<https://dev.twitter.com/docs/streaming-apis>

⁴<https://developers.google.com/places/>

表 1: 実験データ

	場所名	緯度, 経度	ツイート数				
			100m	300m	600m	1km	総数
1	東京スカイツリー (中規模)	35.710023, 139.810702	68	86	148	269	269
2	東京駅 (大規模)	35.681178, 139.766085	247	350	682	3093	3093
3	羽田空港 (大規模)	35.632518, 139.881359	4	115	350	768	768
4	東京ディズニーランド (大規模)	35.359796, 138.727598	20	72	342	679	679

できた。以上より、ツイートユーザと Web 閲覧ユーザがツイートと Web を通してリアルタイム通信できることを確認できた。

4.2 提案システムの運用

Web 閲覧ユーザはツイートユーザのある場所に関するツイートを閲覧できることによってその場所の現状を把握することが可能となる。検証として、東京スカイツリーについて 2013 年 9 月 16 日 17 時 46 分には「オーノー、台風のせいでスカイツリーのぼれない(笑)(笑)」というツイートがあり、9 月 29 日 19 時 09 分には「東京スカイツリーなう。日曜日の今の時間でも 90 分待ちなので、断念 orz」というツイートがあった。Web 閲覧ユーザはそれらのツイートから台風によりスカイツリーが閉鎖しており登ることができないという情報や、夜景を見ようと人がたくさん並んでおり、なかなか登ることができないといった場所に関連する情報を得ることができた。

また、ツイートユーザが Web 閲覧ユーザに質問を投げかけることもできる。例えば、ツイートユーザが東京スカイツリーにて、「東京スカイツリーは世界で何番目に高いですか」とツイートすると、そのツイートがフォロワーだけでなく、東京スカイツリーに関連する複数の Web ページにおいて、それら多くの Web 閲覧ユーザにも同時に提示されることを確認できた。また、Web 閲覧ユーザの返信はツイートとして、ツイートユーザに提示され、これにより、双方向コミュニケーションが行え、ツイートユーザは効率よく情報取得ができる。

4.3 位置情報に基づくツイート提示

提案システムでは、ツイートとページを位置情報に基づき関連付けることで場所に関する問合せを実現する。提案手法では、ツイートの発生位置の半径 d 内のツイートを対象として、ツイートの内容が位置に基づいているかを判定する。

実験では、半径 d を変化させ、ページに提示されるツイートの適合率、再現率、F 値を検証する。対象としたツイートは、2013 年 9 月 16 日～20 日の月曜日から金曜日の 9 時～14 時、16 時～20 時の期間で、「東京スカイツリー」、「東京駅」、「羽田空港」、「東京ディズニーランド」で発生したツイートを対象とした。表 1 に、実験で使用した各データのツイート総数を示す。

各ツイートに対する正解データの判定は、20 代の大学生 12 人による主観的評価に基づき、ツイート内容が場所に関するツイートとして適しているかどうかで評価を行う。1 つのツイートに対して被験者 5 人が評価 (適当: 1, 不適当: -1, 不明: 0) を行い、平均が 0.6 以上の適しているとし、正解データとした。なお、場所名が含まれているだけでなく、場所に関する感想や場所に関する問合せも正解データとして評価してもらった。この実験における適合率、再現率は以下とする。

提案手法が提示した半径 dm のツイートのうち、

$$\text{適合率} = \frac{\text{被験者が正解と評価したツイート数}}{\text{提案手法が提示した半径 } dm \text{ のツイート総数}}$$

提案手法が提示した半径 dm のツイートのうちの正解データ数

$$\text{再現率} = \frac{\text{提案手法が提示した半径 } dm \text{ のツイートのうちの正解データ数}}{\text{被験者が正解と評価した半径 } 1km \text{ のツイート総数}}$$

また、同じ範囲で場所名におけるハッシュタグ検索を行った場合の適合率、再現率、F 値を求め、比較を行う。

図 4 に取得範囲 d を 100m, 300m, 600m, 1km で変化させた各場所名での適合率、再現率、F 値の結果を示す。実線が本手法における結果で、点線でハッシュタグ検索における結果である。

ハッシュタグ検索の場合、ハッシュタグをつけてツイートを発信しているユーザが少ないため、再現率が低い結果となった。

本手法においては、東京駅の 600m 以内、東京ディズニーランドの 600m から 1km 及び東京スカイツリー、羽田空港において、0.8 以上の再現率を得ることができた。東京スカイツリー、東京駅の 600m 以内においては 0.7、羽田空港の全範囲においては 0.7 以上の適合率を得ることができた。一方で、東京ディズニーランドにおいては、全体的に適合率が低くなっている。これは、内部の建物名やイベント名の単語の種類が多く使われていることにより、特徴語を抽出することができなかったためと考えられる。なお、東京ディズニーランド及び羽田空港の 100m においては、ツイート数が少ないため実験データとしては適さなかった。

4.4 考察

4.3 節の図 4 で示したとおり、平日の 5 日間でツイートの発生位置から半径 d が 100m～1km において、東京駅では 600m、その他の東京スカイツリー、東京ディズニーランド、羽田空港では 1km の各範囲内において、F 値が最も高い結果であった。このことより、平日において、人が密集している公共施設では、本手法によるツイート選別は有効であることが確認できた。

また、F 値の低かった取得範囲に関しては、提案手法でツイートを選別する場合、ツイートにおける特徴語として、名詞となる単語のみを対象としており、形容詞や顔文字など感想のツイートを対象としていないことが影響していたと考えられる。また、対象ページに関しては、今回は場所名を検索キーワードとした検索結果のページを対象としているが、ツイートを店舗ごとの小規模施設に細かく分類するため、店舗などが密集している場所では、周辺の店舗も対象とする必要があると考えられる。周辺店舗の名称の抽出は、実験結果のツイートを検証した結果、ツイートから抽出された出現頻度の高い単語の利用が考えられるが、今後、より多くの施設や駅等を密接度で分類し、検証する必要がある。

5. 関連研究

近年、Twitter をテキストマイニングの対象とした研究は活発に行われており、Twitter に投稿されたツイートを分析することでイベントの検出や位置情報の取得を試みた研究も数多くある。

Arakawa ら [4] は位置情報ツイートから位置依存性の高い文字列を抽出する手法を述べている。位置情報ツイートから得たエリアを 100 キロ四方のグリッドに分割し、それぞれのグリッド内のツイート含有率を計算し、ツイート含有率がある閾値を超えたエリアを最終的に 1 キロ四方のグリッドまで走査することにより、1 つのキーワードに対して複数の位置依存性を抽出することができる。この研究では、位置情報とツイートのコンテンツを対応付けている。本研究でも、位置情報とツイートのコンテンツ内容に関連付けているが、こちらは、特定の場所や建造物を中心とした位置の重要性の高い文字列の抽出を行っている。また、Yamamoto ら

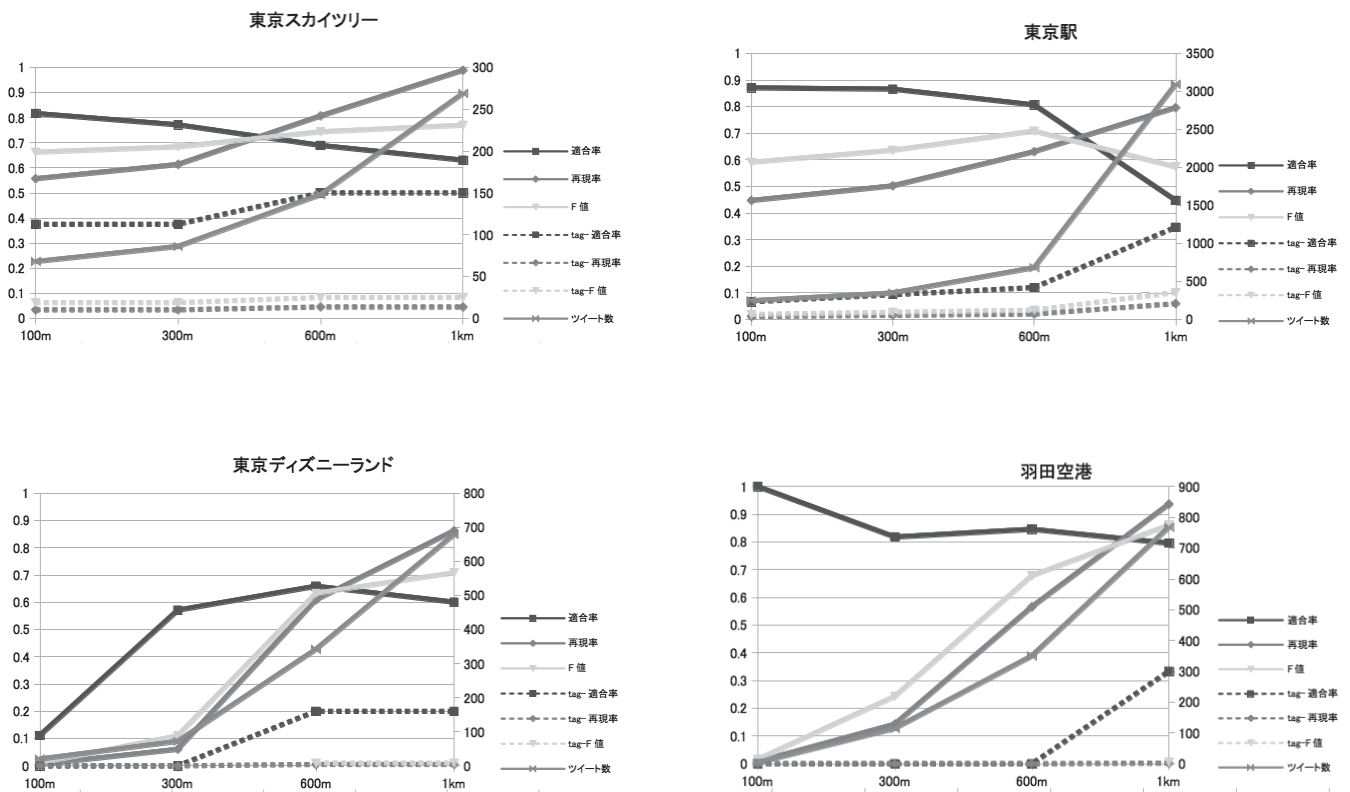


図 4: 位置情報に基づき提示されたツイートに対する適合率, 再現率, F 値の結果

の Twitter に投稿された実生活情報から有用性の高いものを抽出し局面に応じた記事をユーザに提示するシステム [5, 6] やツイートから地震や台風などのイベントの検出を試みた研究として榊らの研究 [7] がある. Twitter のタイムラインを監視しておくことでリアルタイムでイベントの検出を行い, 高い精度を得られた. これらの研究では, コンテンツベースでの抽出を行っているが, 本研究では, コンテンツ内容と緯度経度情報の関連付けを行っている. また, Lee らの研究 [8] では, イベント検出対象となる地域をいくつかの小さな地域に分割し, ツイート数, ユーザ数, ユーザの移動状況の 3 点を分析し, その地域の通常時の状態を推測する. そして, 通常の状態とは異なった多くのツイートが投稿された場合, イベントが発生したとみなしている. Nichols らの研究 [9] は, ツイートのコンテンツ内容の変化に注目しており, 更新の量の急増などでイベント内の重要な瞬間の識別を行う. Ribeiro らの研究 [10] では, ツイート内容を識別し緯度経度から区域でのイベント発生を検出する. 本研究では, ツイートの内容に着目し, 特定の単語の出現頻度が高くなれば, イベントが発生したと見なししている.

位置情報付き画像ツイートをを用いてイベント検出を試みた研究として, Nakaji ら [11] はあらかじめ特定のキーワードや期間を設け, 位置情報付きツイートを収集し, 解析することで画像付きのイベント検出を試みた. Kaneko ら [12] はイベントのキーワードをシステムにより自動的に抽出することで多くのイベントを抽出することで未知のイベントのキーワードを得られるようにし, キーワードを用いて収集した画像を解析することで, ユーザが知らないイベントでも画像により視覚的にとらえることができるようにした. これらの研究は, それぞれ位置情報ベースとコンテン

ツベースで別々に取り扱っているが, 本研究では, この 2 つを同時に取り扱う. Takemura ら [13] は, Twitter ユーザを, 広く一般のユーザが興味を示す情報を発信するのか, 一部のユーザのみが興味を示す情報を発信するのかの範囲を示すため, 対象局所性と定義される指標を用いた分類を行う手法を提案している. 本研究では, Twitter の位置情報と内容に基づいて発信されたツイートが発信された場所に関連しているかを判別する.

オンライン上でのユーザ間のコミュニケーションを行う研究として, 質問回答サイトの回答を対象にした研究として, Yahoo!知恵袋を対象にして知恵袋の質問回答情報をクラスタリングし, クラスタごとに機械学習を行って最も質問に適した回答となりうる可能性が高い回答を判定する手法を述べた [14] や, 教師つき負例と教師なし正例からなる学習コーパスからの SVM 学習器を作成し, 不適切な回答の発見を半自動化するシステムの作成を行った [15] がある. また, ある質問に対して一つ以上の回答の組 (以下, QA コンテンツと記す) は急激に増えている. QA コンテンツは質問に詳しい専門家がベストアンサーを決めているわけではなく, 閲覧ユーザの投票で決定したり, 質問者自らが決定するため, 質問に対する回答が不十分な場合がある. そこで, 高田ら [16] は Web 情報を用いてコンテンツを補完することで, QA コンテンツの利用者が回答の信憑性を確認したり, 補足的な情報を得ることができると提案している. 本研究では, ツイートユーザと Web 閲覧ユーザ間でリアルタイムコミュニケーションが可能なシステムの構築を目標とする.

6. まとめ

本研究では、実空間で SNS 等を用いて情報発信しているツイートユーザと、異なる場所で関連する Web ページを閲覧しているユーザに対して、実空間の位置と Web コンテンツの内容とを対応付けることで、異なる場所で異なるメディアを利用しているユーザ間のリアルタイム双方向通信の実現を目指し、位置に基づくリアルタイム問合せシステムを構築した。評価実験ではプロトタイプを構築し、ツイートユーザ密度の異なる 4 地点において、各地点に対するツイートとページの関連性を検証した。実験結果より、東京スカイツリーや空港、駅等の密集地において、本手法が有効であることを確認できた。今回はシステムにおいての前実験として、Web 閲覧ユーザに適したツイートが提示されるかの検証を行った。

今後は、実際にシステムを被験者に操作してもらい、情報取得に有用であったかといった、システムが Web 閲覧ユーザ及びツイートユーザに対して有効であるかの検証を行う予定である。また、Web 閲覧ユーザの Web 上に必要な情報についてのアンケートを行った。アンケートの結果より、その場所で現在行われているイベント情報、その場所の今後の天気の情報やその場所へのアクセス情報といったものが、Web 閲覧ユーザにとって有益であることが確認できた。今後このような集約情報の提示の実装も行う予定である。

[謝辞]

本研究の一部は、総務省戦略的情報通信研究開発推進事業 (SCOPE) および JSPS 研究費基盤研究 B (26280042) の助成を受けたものである。ここに記して謝意を表す。

[文献]

- [1] Shingo Tajima, Taketoshi Ushiyama: A Method for Composing Ad-hoc Following Networks on Twitter for Sharing Information among Event Participants, *International Journal of ADADA*, Vol. 17, No. 4, pp. 199-124, 2014.
- [2] 奥 健太, 橋本 拓也, 上野 弘毅, 服部 文夫: 地理情報推薦のための観光スポットと位置情報付きユーザ生成コンテンツの対応付け手法の提案, 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2013), A3-2, 2013.
- [3] 松井 優也, 河合 由起子: 人と情報の検索および相互作用を目指したソーシャルサーチシステムの研究開発, *日本ソフトウェア学会コンピュータソフトウェア (ソフトウェア論文)*, Vol. 28, No. 4, pp. 196-205, 2011.
- [4] Yutaka Arakawa, Shigeaki Tagashira and Akira Fukuda: Relationship Analysis between User's Contexts and Real Input Words through Twitter, *IEEE Globecom 2010 Workshop on Ubiquitous Computing and Networks (UbiCoNet 2010)*, pp.1813-1817, 2010.
- [5] Shuhei Yamamoto and Tetsuji Satoh: Two Phase Extraction Method for Multi-label Classification of Real Life Tweets, *In Proc. of the 15th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2013)*, pp. 16-25, 2013.
- [6] Shuhei Yamamoto and Tetsuji Satoh: Two Phase Extraction Method for Extracting Real Life Tweets using LDA, *In Proc. of the 15th Asia-Pacific Web Conference (APWeb 2013), Lecture Notes in Computer Science 7808*, pp. 340-347, 2013.
- [7] Takeshi Sakaki, Makoto Okazaki and Yutaka Matsuo: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *In Proc. of the International World Wide Web Conference (WWW 2010)*, pp. 851-860, 2010.
- [8] Ryong Lee and Kazutoshi Sumiya: Measuring Geographical Regularities of Crowd Behaviors for Twitter-based Geo-social Event Detection, *In Proc. of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN 2010)*, pp. 1-10, 2010.
- [9] Jeffrey Nichols, Jalal Mahmud and Clemens Drews: Summarizing Sporting Events Using Twitter. *In Proc. of the 2012 ACM International Conference on Intelligent User Interfaces (IUI 2012)*, pp. 189-198, 2012.
- [10] S. S. Ribeiro, C. A. Davis, D. R. R. Oliveira, W. Meira, T. S. Goncalves and G. L. Pappa: Traffic Observatory: A System to Detect and Locate Traffic Events and Conditions Using Twitter. *In Proc. of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN 2012)*, pp. 5-11, 2012.
- [11] Yusuke Nakaji and Keiji Yanai: Visualization of Real World Events with Geotagged Tweet Photos. *In Proc. of IEEE ICME Workshop on Social Media Computing (SMC 2012)*, pp. 272-277, 2012.
- [12] Takamu Kaneko and Keiji Yanai: Visual Event Mining from Geo-tweet Photos, *IEEE ICME Workshop on Social Multimedia Research (SMMR 2013)*, pp. 1-6, 2013.
- [13] Hikaru Takemura and Keishi Tajima: Tweet Classification Based on Their Lifetime Duration, *In Proc. of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)*, pp. 2367-2370, 2012.
- [14] 西原 陽子, 松村 真宏, 谷内田 正彦: QA サイトにおける質問に適した回答の判定, *言語処理学会 NLP 若手の会第 2 回シンポジウム*, 2007.
- [15] Daisuke Kobayashi and Naohiro Matsumura: Automatic Gender Estimation of Bloggers' Gender, *In Proc. of International Conference on Weblogs and Social Media (ICWSM 2007)*, pp. 279-280, 2007.
- [16] 高田 夏希, 山本 裕輔, 小山 聡, 田中 克己: 質問応答コンテンツに対する Web による回答補完, 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2009), C4-6, 2009.

安井 豪基 Goki YASUI

京都産業大学大学院先端情報学研究所博士前期課程在学中。データベースシステム, Web マイニングの研究・開発に従事。

王 元元 Yuanyuan WANG

京都産業大学コンピュータ理工学部研究員。博士 (環境人間学)。マルチメディア・データベース, e ラーニングシステム, Web 情報検索の研究・開発に従事。情報処理学会, 日本データベース学会, 人工知能学会等各会員。

細川 侑士 Yuji HOSOKAWA

京都産業大学大学院先端情報学研究所博士前期課程在学中。データベースシステム, Web マイニングの研究・開発に従事。情報処理学会学生会員。

河合 由起子 Yukiko KAWAI

京都産業大学コンピュータ理工学部准教授。博士 (工学)。情報推薦, 検索技術, 信頼性分析および 3 次元コンテンツ・ユビキタスコンテンツへの応用技術に関する研究に従事。情報処理学会, 電子情報通信学会, 日本データベース学会等各会員。

秋山 豊和 Toyokazu AKIYAMA

京都産業大学コンピュータ理工学部准教授。博士 (工学)。分散システムの研究に従事。電子情報通信学会, IEEE CS 等各会員。

角谷 和俊 Kazutoshi SUMIYA

兵庫県立大学環境人間学部環境人間学科教授。博士 (工学)。Web 情報システム, 情報配信システム, 放送コンピューティング, モバイル・マルチメディア, マルチメディア・データベースに関する研究に従事。IEEE Computer Society, ACM, 情報処理学会, 電子情報通信学会, 日本データベース学会等各会員。