

## WebとSNSのメディア横断型コミュニケーション支援システム

王元元<sup>†</sup> 安井 豪基<sup>†</sup> 細川 侑士<sup>†</sup> 河合 由起子<sup>†</sup> 秋山 豊和<sup>†</sup> 角谷 和俊<sup>‡</sup>

Yuanyuan Wang Gouki Yasui Yuji Hosokawa Yukiko Kawai Toyokazu Akiyama Kazutoshi Sumiya

## 1. まえがき

近年, Twitter<sup>1</sup> や Facebook<sup>2</sup> 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言を行っている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ワー)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いないSNSユーザのツイートも検出できる。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た場所に関する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
- SNSユーザとWebユーザのリアルタイムコミュニケーションシステムの開発

<sup>†</sup>京都産業大学<sup>‡</sup>兵庫県立大学<sup>1</sup><https://twitter.com/><sup>2</sup><http://www.facebook.com/>

ツイートユーザとWeb閲覧ユーザのコミュニケーションを可能とすることで, 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。

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擇することで, 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



図1: Web・SNS間のコミュニケーションシステム概要図

るため匿名性が保たれる。図1では、東京スカイツリーにいるユーザがツイートを発信した場合に、そのツイートが東京スカイツリー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ページ閲覧ユーザは混雑具合やスカイツリー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スカイツリーの風景や展望の場所など、その時その場でしか確認できな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記す)ならびにWeb閲覧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、対応付け管理する。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、本サービスのアカウント<sup>3</sup>のフォロワーとなっている必要があり、Web閲覧ユーザは提案

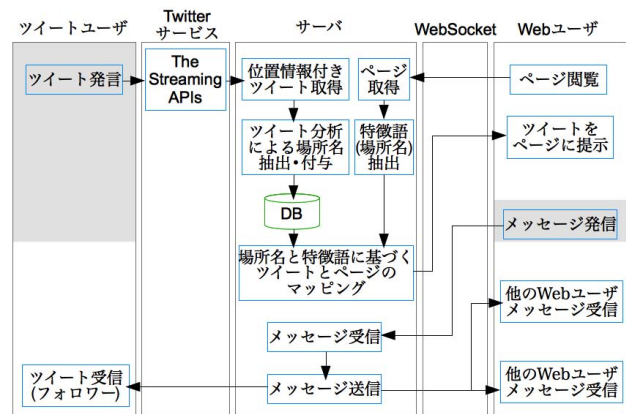


図2: システムの構成図

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

<sup>3</sup><https://Twitter.com/@RtQAService>

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、東西および南北が各々1度以上異なるように、北東および南西の緯度経度を指定することで、その2点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。

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 3<sup>4</sup> 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、下記の式よりツイート  $t$  内における単語  $i$  の出現頻度を抽出し、その平均値を算出する。

$$\sum_{i=1}^q \frac{\text{単語 } i \text{ が出現するツイート数}}{\text{ツイート総数 } n} \times \frac{1}{q} \quad (1)$$

$q$  はツイート  $t$  に出現する単語総数である。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

<sup>4</sup><https://developers.google.com/place/>

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別した。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 WebSocket<sup>5</sup> 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 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、場所に関するツイート情報取得ならびに 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、Firefox23.0.1 上で動作確認した。なお、2013 年 8 月から同年 9 月末までに The Streaming APIs で取得した日本全国のツイートデータは約 20MB であった。

### 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 1.1<sup>6</sup> 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

<sup>5</sup><http://gihyo.jp/dev/feature/01/websocket/0001>

<http://www.atmarkit.co.jp/ait/articles/1111/11/news135.html>

<sup>6</sup><https://dev.twitter.com/docs/streaming-apis>

表 1: 実験データ

| No. | 場所名              | 緯度, 経度                | ツイート数 |      |      |      |      |      |      |
|-----|------------------|-----------------------|-------|------|------|------|------|------|------|
|     |                  |                       | 50m   | 100m | 300m | 600m | 800m | 1km  | 総数   |
| 1   | 東京スカイツリー (中規模)   | 35.710023, 139.810702 | 58    | 68   | 86   | 148  | 216  | 269  | 269  |
| 2   | 東京駅 (大規模)        | 35.681178, 139.766085 | 64    | 247  | 350  | 682  | 1945 | 3093 | 3093 |
| 3   | 羽田空港 (大規模)       | 35.632518, 139.881359 | 1     | 4    | 115  | 350  | 710  | 768  | 768  |
| 4   | 東京ディズニーランド (大規模) | 35.359796, 138.727598 | 10    | 20   | 72   | 342  | 587  | 679  | 679  |



図 3: Web 閲覧インターフェース

version 3<sup>7</sup> を用いて、半径  $dm$  以内のツイートを取得した。

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送受信のインターフェースは既存の 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ー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メッセージ送信が可能である。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 を通じてリアルタイム通信が可能であ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館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、そのツイートが東京スカイツリーに関連する Web ページにアクセスしている Web 閲覧ユーザに提示される。Web 閲覧ユーザは、すぐさま解答を調べることで、その質問に対する返信をする。返信はツイートとして、ツイートユーザに提示される。これにより、双方向にコミュニケーションを行うことが可能となる。

#### 4.3 場所に関するツイート選別手法の検証

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。本手法は、ツイートの発生位置が特定の位置の中心から半径  $dm$  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  $dm$  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。ただし、東京ディズニーランド及び羽田空港の 100m 以内においては、ツイート数が少ないため実験データとしては適さな

<sup>7</sup><https://developers.google.com/places/>

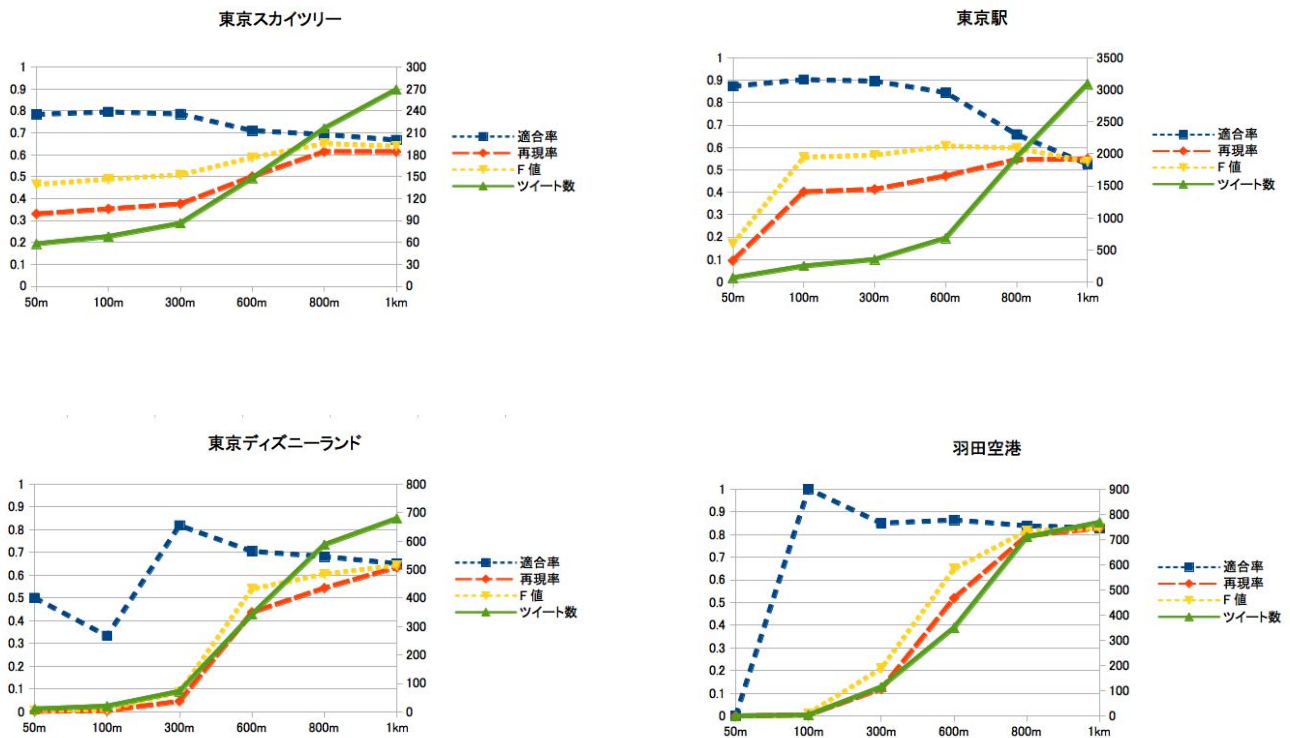


図 4: 位置情報に基づき取得範囲  $d_m$  において提示されたツイートに対する適合率, 再現率, F 値の結果

いため対象から除外した。

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 人が評価を行い、5 人中 3 人以上が適しているとしたツイートを場所名と関連する適した正解データとした。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。

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

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

図 4 に取得範囲  $d$  を 50m, 100m, 300m, 600m, 800m, 1km で変化させた各場所での適合率, 再現率, F 値の結果を示す。

東京スカイツリー, 東京駅の半径 50m~600m 以内, 東京ディズニーランドの半径 300m~600m 以内において 0.70 以上, 羽田空港の 50m を除く全範囲において, 0.80 以上の適合率を得ることができ, 全体の平均で 0.79 であった。再現率は, 東京駅的全範囲, 他の場所での 600m 以内において再現率が 0.60 以下となり, 1km においての平均は 0.65 となった。また, F 値の平均値は 0.66 であった。以上より, ツイート数が十分に取得できない場所が存在することを確認した。本システムでは, リアルタイムな SNS と Web 間のコミュニケーションが重要なため, Web 閲覧ユーザに常に新しいツイートを提示することが望ましいため, ツイートを十分な数取得する必要がある。具体的には, ツイートを 1 時間に 2, 3 件以上, 1 週間で 300 件程度は取得する必要があると考える。よって, 東京スカイツリーといった中規模施設である場所において範囲半径 600m ではこの数を満たすことができない。そこで, ツイートの取りこぼしを少なくするために, 再現率を上げる必要があり, 4.4 節で改善案を提示する。

#### 4.4 中規模施設における再現率の改善

3.2 節で扱った式 (1) では, ツイート内に特徴的な単語が出現しても, 同ツイート内に他に多くの単語が出現すると, 平均値が下がり, 閾値以下になってしまう。そこで, 単語の重み付けにシグモイド関数を加えて, 特徴的

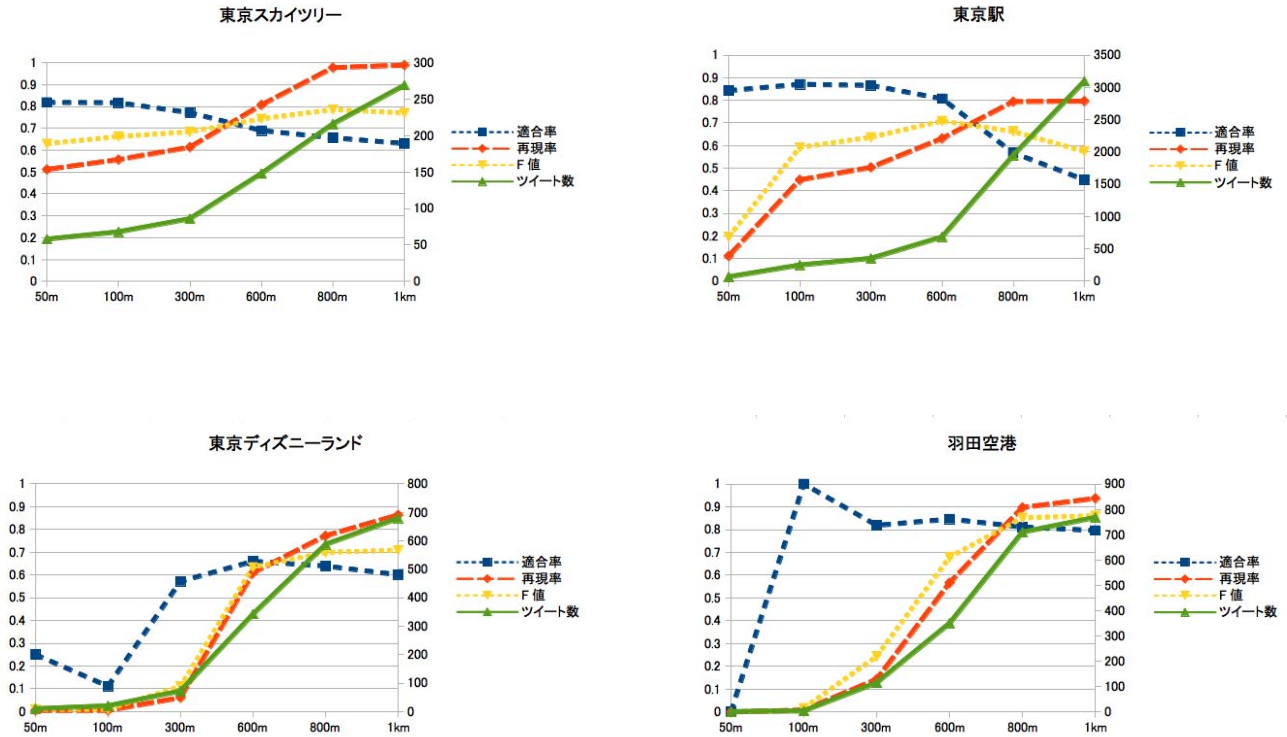


図 5: 式 (2) を適用した場合の取得範囲 \$d\_m\$ においてツイートに対する適合率, 再現率, F 値の結果

な単語に関してはさらに重みを増す式 (2) を提案する.

$$\sum_{i=1}^q \left( \frac{\text{単語 } i \text{ が出現する ツイート数}}{\text{ツイート総数 } n} \times \frac{1}{1 + e^{-x}} \right) \times \frac{1}{q} \quad (2)$$

なお,

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

で求まる単語 \$i\$ の DF 値である. 図 5 にこの式を適用した場合についての適合率, 再現率, F 値の結果を示す.

式 (1) で行った結果の図 4 と比較すると, 適合率の平均が 0.79 から 0.67 と低下したが, 再現率の平均が 0.65 から 0.89 と大きく改善され, F 値の平均も 0.66 から 0.76 と上昇が確認できた.

#### 4.5 考察

4.3 節, 4.4 節の実験より, 東京駅, 羽田空港, 東京スカイツリーにおいては, 良好な結果であることが確認できる. しかしながら, 東京ディズニーランドでは, 有効性は確認できなかった. この結果より, 場所によって建物の規模や人口密度に対して, 適切な取得範囲半径 \$d\_m\$ の設定, 及び式 (1) (2) の適用が必要と考えられる. 中規

模施設の東京スカイツリーでは \$d=1\text{km}\$, その他大規模施設の東京駅及び東京ディズニーランドでは \$d=600\text{m}\$, 大規模施設であるが人口密度が低い羽田空港では \$d=1\text{km}\$ とすることで, 高い適合率, 再現率を得ることが確認できた. 以上より, 人が密集している場所においては, シグモイド関数を用いた式 (2) を適用し, 大規模施設で密集度の低い場所では式 (1) を用いることが有効であることが明らかとなった.

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, 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. 実験結果のツイートを検証した結果, 周辺店舗の名称の抽出は, 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を関連付けているが、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、中地ら[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らの[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。マスメディアに対する意見を対象とした研究として、テレビ番組に対する意見を持つSNSユーザを、電子番組表及びテレビ番組の字幕テキスト番組公式の特徴語群及びSNSユーザが生成する番組の特徴語群が、番組の放送時間帯に投稿されるメッセージに含まれるかによって、リアルタイムに検出する手法を提案している[13]がある。本研究では、予めツイートを取得しておき、位置情報により、場所ごとの特徴語を決定しておく、それを基に新たなツイートが発信されれば、緯度経度情報から場所を求め、その場所の特徴語を含むかで、リアルタイムに検出する方法を提案している。Takemuraら[14]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、質問応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[15]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[16]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[17]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手法を提案している。本研究では、SNSユーザとWeb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、駅等の中大規模施設において、本手法の有効性を確認できた。

今後、さらに多地点での検証ならびに周辺名称を用いた関連Webページ検索による拡張を行う予定である。また、本システムが収集したツイートと、ハッシュタグによる全関連ツイートとを比較し、本システムによるツイート収集結果の妥当性の検証を行う予定である。

## 謝辞

本研究の一部は、総務省戦力的情報通信研究開発制度(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。

## 参考文献

- [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, 日本ソフトウェア科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] 山本 祐輔, 浅井 洋樹, 上田 高德, 秋岡 明香, 山名 早人: テレビ番組に対する意見をもつ Twitter ユーザのリアルタイム検出, 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), C1-4, 2013.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 細川 侑士, 西村 直也, 白石 優旗, 河合 由起子, 秋山 豊和: ユーザ行動抽出に基づく訪問者に対する災害等イベント情報発信システムの提案, 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), B5-1, 2013.