情報処理学会論文誌 Vol.54 No.2 710–720 (Feb. 2013)

場所誘因型位置情報付き発言の検出と可視化

蛭田 慎也[1,a)] 米澤 拓郎[1,b)] 徳田 英幸[1,2,c)]

受付日 2012年5月14日, 採録日 2012年11月2日

概要：本研究では，ソーシャルメディア上の位置情報付き発言において，特に現在の場所に誘因された発言を検出する手法について，提案・検証する．GPS による位置情報付き発言を用いてイベント発生場所の検出を行う際，位置情報が付加されていてもその場所とまったく関係のないノイズとなる発言が多いため，イベントの判定結果にもノイズが多くなってしまうことが問題としてあげられる．この問題を解決するため，本研究では位置情報付き発言の中でも，現在の場所で起きた出来事・状況などに誘因されて発言されたものを「場所誘因型位置情報付き発言」と定義し，その検出を行う．その手法として，Twitter の位置情報付き発言を対象とし，取得した発言元情報と発言本文に含まれるキーワードを参照することで場所誘因型位置情報付き発言の検出を実現する．18 名の被験者に実際に Twitter から取得した位置情報付き発言の分類実験を行ってもらい，その結果 82%の精度で場所誘因型位置情報付き発言を正しく判定可能であることを示した．さらに，場所誘因型位置情報付き発言の可視化を行うアプリケーションを作成し，本研究の有効性を示す．

キーワード：アーバンセンシング，データマイニング，位置情報サービス，マイクロブログ，場所誘因型位置情報付き発言

Detection and Visualization of Place-triggered Geotagged Tweets

Shinya Hiruta[1,a)] Takuro Yonezawa[1,b)] Hideyuki Tokuda[1,2,c)]

Received: May 14, 2012, Accepted: November 2, 2012

Abstract: The paper proposes and evaluates a method to detect tweets that are triggered by places where users locate. Recently, many researches address to detect the real-world events from social media such as Twitter. However, geo-tagged tweets often contain noises, which means tweets which are not related to the users' location. These noises are problem for detecting real-world's events. To address and solve the problem, we define the "Place-Triggered Geotagged Tweet", tweets which are related to the users' location, and propose a method to detect it. We design and implement a keyword matching-based detection technique to detect place-triggered geotagged tweet. We evaluate accuracy of the method with 18 participants, and present that our method achieves 82% of accuracy. Additionally, we also present the effectiveness of our method by implementing two applications which visualize place-triggered geotagged tweets.

Keywords: urban sensing, data mining, location-based services, microblogs, place-triggered geotagged tweets

1. はじめに

携帯電話やスマートフォンをはじめとするモバイルデバイスが普及するとともに，Twitter [1] などのソーシャルメディアの利用者が増加している．多くの人々が，日々の生活の中で見たことや感じたことを，モバイルデバイスを用いてソーシャルメディアに投稿し，共有することが可能になった．近年，ユーザが投稿したこのようなデータを分析することで，実世界で発生する様々な出来事の検出を目指

[1] 慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252–0882, Japan
[2] 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252–0882, Japan
a) hiru@ht.sfc.keio.ac.jp
b) takuro@ht.sfc.keio.ac.jp
c) hxt@ht.sfc.keio.ac.jp

す研究が多く行われている [2], [3]．本論文では，実世界で発生する出来事のことを「実世界イベント」と定義する．実世界イベントを検出することによって，混雑箇所を回避するナビゲーションシステムや，人々の移動パターンを都市計画にフィードバックする応用などが可能になると考えられる．

実世界イベントでは，そのイベントがどこで発生したかという空間的な情報が重要となる．インターネット上の仮想空間で発生するイベントと異なり，実世界イベントには必ず地理的，空間的な要素が含まれるからである．ソーシャルメディアの投稿から取得できる空間情報として最も正確なものは，GPS で取得した緯度経度の位置情報である．Twitter などのソーシャルメディアでは，モバイルデバイスに搭載された GPS アンテナにより位置情報を取得し，発言に付加する機能が一般的に利用可能となっている．したがって，緯度経度の情報が付加された発言を収集・解析することで，高精度な空間情報を含む実世界イベントの検出が可能となると考えられる．

一方で，位置情報付き発言には，その場所とまったく関係ない発言が含まれるという問題が存在する．実世界イベントの検出の対象となりうるべき発言では，ユーザが現在地に関連する発言や写真などを共有したいと思った際に，位置情報が付加されていることが期待される．しかし実際には，その場所と関係のない話題で発言する際にも位置情報を付加してしまうケースが多い．よって，実世界イベントの検出を行う際は，位置情報が付加されていても，その場所と関係がないノイズとなる発言が存在することを考慮しなければ，イベントの発生場所を適切に検出することができない．また，解析に関係ないデータを処理に含めてしまうことは，パフォーマンスやスケーラビリティの低下につながる．今後，位置情報付き発言はより増加していくことが予想されるため，効率的なデータ量の削減が求められる．

本研究では，位置情報付きの発言の中でも，現在の場所で起きた出来事・状況などに誘因されて発言されたものを「場所誘因型位置情報付き発言」と定義し，その検出を目的とする．まず，場所誘因型位置情報付き発言の分類を行うため，また位置情報付き発言の中にどの程度場所と関係のない発言がノイズとして含まれているかを把握するために，位置情報付き発言の調査を行った．その結果，発言内容を居場所の報告・食事・天候・帰宅・地震の 5 種類に分類し，場所誘因型位置情報付き発言とした．この結果をもとに，位置情報付き発言をこの 5 種類に分類するため，発言本文に含まれるキーワードなどをもとに判定を行うフィルタを作成した．評価として，第三者の被験者 18 名に正解データの作成を依頼し，分類が妥当であるかと，フィルタによる判定精度の検証を行った．

本研究の貢献は，以下の 3 点である．

- 位置情報付き発言の現状を調査し，整理したこと
- 緯度経度の情報を含み，現在の場所で起きている出来事に誘因された発言を「場所誘因型位置情報付き発言」と定義し，その検出手法を構築したこと
- 評価の結果，18 名の被験者によって位置情報付き発言の分類を行い，提案手法が 82%の精度で場所誘因型位置情報付き発言を検出可能であると示したこと

本論文は，以下のように構成される．まず 2 章で位置情報付き発言による実世界イベントの検出における手法と問題意識について述べ，3 章で位置情報付き発言の実態調査について述べる．4 章で場所誘因型位置情報付き発言の検出を行うアプローチと，本研究の提案するシステムの設計と実装について述べる．また，判定した場所誘因型位置情報付き発言を可視化するアプリケーションについて，5 章で説明する．6 章で評価についての詳細を述べて考察を行い，最後に 7 章で本論文をまとめる．

2. 位置情報付き発言による実世界イベントの検出

本章では，まず本研究の関連研究について述べる．次に，問題意識と本研究の目的について述べる．

2.1 関連研究

ソーシャルメディアを用いて実世界イベントの検出を行う際において，発言が行われた場所をどのように取得するかが問題となる．Sakaki らは，Twitter の発言を取得し，地震の発生をリアルタイムに検出している [2]．この研究では，発言が行われた場所を推定するために，ユーザのプロフィールに登録された静的な位置情報を主に参照している．しかし，ユーザはモバイルデバイスを用いて様々な場所で発言を行うことがあるため，必ずしもプロフィールで公開している場所で発言を行ってはいない．したがって，発言の位置情報が不正確なまま解析を行った結果，イベント発生場所の推定結果も不正確になってしまうという問題がある．実際，地震発生場所の推定結果においては，最も精度の良かった手法でも緯度経度で平均 3.01 度（約 300 km）の誤差が発生している．

一方，緯度経度の位置情報が付加された発言は，実世界イベントをより高精度に検出できる可能性がある．しかし，位置情報付き発言にはノイズとなる発言が非常に多い．つまり，GPS による緯度経度が付加されているからといっても，必ずしもその場所に関連する発言ではないということである．実世界イベントの検出を行う際，このようなノイズを含んだ発言をそのまま用いてしまうと，イベントの判定結果もまたノイズが多くなってしまうことが懸念される．Lee らは，Twitter において緯度経度が付加された発言の規則性を計測し，非日常的なイベントが発生した場合に群衆が発言する傾向を検出することにより，イベント判

定を行っている [3]．また，石川らは，ある地域のある時間において発生するホットトピック検出のため，トピックに対する発言単語のばらつきを吸収する意味的辞書の構築を行っている [4]．藤坂らは，集中型・分散型という 2 つのモデルを定義し，位置情報付き発言の履歴を分析することで，地域空間においてユーザが移動する際の特徴的なパターンを発見することを目的としている [5]．しかし，これらの研究においては，位置情報付き発言が本当にその場所に誘因されたものかどうかは考慮されていない．イベントの検出手法という点で本研究と補完し合うものであるが，目的は異なっているといえる．

一方，Yin らは，Web において複数の競合する情報ソースが存在する場合に真実を推定する手法を提案している [6]．また，Wang らは，ベイズや最尤法を用いて真実の推定を行っている [7], [8]．これらの研究では，発言自体の真偽を推定しているが，発言がその場所に誘因されたものかどうかは考慮していない．

このように，位置情報付き発言は実世界イベントの検出に多く用いられているが，場所とは関係のない発言がノイズとしてそのまま含まれており，検出精度に悪影響を及ぼしている．また，発言の真偽に関する研究は存在するが，我々が注目する，場所に誘因された発言かどうかを検証する試みはこれまで行われていない．

2.2 目的

位置情報付き発言を用いてイベント検出を行うためには，その場所に確かに言及している発言を，ノイズとは区別して扱えるようにすることが重要である．もし場所に言及している発言を検出することができれば，街の中で人々の注目を集めている場所が浮かび上がってきたり，特定の場所に対して短時間に多くの言及があった場合に，何かイレギュラーなイベントが発生したことを検出できたりするなどの応用例が考えられる．さらに，どのような意図で場所に言及しているかまで解析できれば，評価の高い観光スポットを発見したり，ゲリラ豪雨の発生を早期発見するなどへの発展も可能となると考えられる．また，解析において不要なデータを除去することで，パフォーマンスやスケーラビリティの向上にも寄与できる．

本研究では，位置情報付き発言の中でも，現在の場所で起きた出来事・状況などに誘因されて発言されたものを「場所誘因型位置情報付き発言」と定義する．場所誘因型位置情報付き発言の種別を整理し，発言を分類可能にすることを目的とする．

3. 位置情報付き発言の実態調査

本章では，位置情報付き発言の内容を分類するため，まず事前調査の手法について述べ，調査結果を示す．次に，調査結果をもとに，場所誘因型位置情報付き発言の種別を分類する．

3.1 調査手法

事前調査として，位置情報付き発言がどのような内容であるかを調査するため，発言の分類を行った．調査対象は，Twitter において日本周辺の位置情報が付加された発言データを StreamingAPI を利用して取得し，2010 年 12 月に取得した 2,010 件を抽出した．それぞれ発言内容の本文を参照し，発言が行われた状況を最も表していると思われる分類を，著者の中の 1 名の主観的判断により作成した．なお，1 つの発言は複数の項目に分類可能とした．

3.2 調査結果

事前調査によって分類された内容，判定数と割合を**表 1** に示す．1 つの発言につき複数の分類が対応付けられた場合は，それぞれの種別を＋記号で並記した．

調査の結果，現在地に関連する内容としては，Foursquare [9] などのチェックイン系サービスと連携した発言が最も多く，全体の 20.0%を占めることが分かった．チェックイン系のサービスは，ユーザがチェックインを行った際などに連動して Twitter にも発言を行う機能が実装されており，多くのユーザが連動機能を利用している．その際，“I’m at 湘南台駅 (Shonandai Sta.) (藤沢市, 神奈川県) http://example.com/hoge” のような文章が投稿される．この文章は投稿時にユーザが編集することが可能であり，天候について追記していた発言が 1.7%，天候と交通状況について追記していた発言が 0.1%ほど確認された．次

表 1 位置情報付き発言の実態調査の結果
Table 1 Result of survey with geotagged tweets.

発言種別	判定数	割合
位置情報と関係ない発言内容	1,195	59.45%
チェックイン	403	20.0%
天候	203	10.1%
風景・施設の写真	45	2.2%
食事	43	2.1%
天候＋チェックイン	34	1.7%
移動	32	1.6%
帰宅	17	0.85%
事故	9	0.4%
交通	6	0.3%
天候＋写真	6	0.3%
地震	4	0.2%
天候＋交通	3	0.1%
災害	2	0.1%
天候＋移動	2	0.1%
天候＋交通＋チェックイン	2	0.1%
チェックイン＋帰宅	2	0.1%
帰宅＋天候	2	0.1%
合計	2,010 件	

に多かった分類として，天候に関する発言が 10.1%，風景・施設の写真が 2.2%，食事に関する内容が 2.1%，移動中であるという内容が 1.6%確認できた．その他，帰宅したという内容，事故・災害を目撃したという内容，交通状況に関する内容，地震を体感したという内容に加え，天候・帰宅・交通などの組合せがそれぞれ 1%未満の割合で分類された．

一方，全体の 59.45%の発言が，付加された位置情報と関係のない内容であることが確認された．場所と関係ない発言として，“@hiru_pub おはようー 昨日の件はどう？” など，他のユーザに対する返信を行っている例が多かった．これは，つねに位置情報を送信するような設定にしているために，他人へ返信を行った際にたまたま自分がいた場所の位置情報を送信してしまっているものだと思われる．また，“最近、仕事がすごく忙しい。” “腹減って眠い。” など，その場所とはまったく関係ない話題を発言していたりする例もあげられる．このように，位置情報が付加されていてもその場所と関係のない発言が多いという現状では，これらのノイズが位置情報付き発言を用いて実世界イベントの検出を行う際の妨げとなってしまうと考えられる．

3.3 場所誘因型位置情報付き発言の分類

事前調査の結果より，検出の対象とする場所誘因型位置情報付き発言の種別を以下の 5 種類に分類した．

- 居場所の報告：ユーザが，現在地の場所や施設などについて言及している発言．
- 食事：何かを食べた・飲んだなど，食事に関連する発言．
- 天候：現在地の天候に関する発言．
- 帰宅：学校・職場などから帰る，帰宅したという発言．
- 地震：地震を体験した，揺れたなどの発言．

全体の約 20%以上の発言が，Foursquare [9] などのチェックイン系サービスと連携しているものであった．これは，ユーザが現在の場所にいるという事実をソーシャルメディア上でつながりのある人にアピールしているものであるといえ，「居場所の報告」という分類が妥当であると考えた．既存研究において，石川らは，場所に対して意味づけを行うための辞書を作成することが目的とし，Foursquare への投稿は位置情報と URL のみしか含まないため有用なサンプルではないと主張している [4]．しかし，本研究は場所に意味づけを行うだけではなく，その場所で起きた出来事や状況に誘因されて発言したユーザが存在するという事実を取得することも重要だと考えている．たとえば，普段はチェックインが少ない場所に，多くのユーザがチェックインを行っていることが観測された場合，その場所で何かイベントが発生しているのではないかと推測することができると考える．

また，ユーザが現在の場所で行っている行動に注目すると，食事をしているという行動と，帰宅したという行動に分類できる発言が多かったため，「食事」，「帰宅」という分類を行った．さらに，ユーザが現在の場所で見たり感じたりしたことに注目すると，雨に降られたり，良い天気であることを共有する発言が多く見られたため，「天候」という分類を行った．最後に，「地震」という分類を行った．事前調査期間においては大きな地震が発生しなかったため地震を感じたという発言は 0.2%と少数であったが，地震は甚大な被害を及ぼすおそれがあり，特に警戒されるべき対象であると考え，分類に追加した．

一方，風景・施設の写真に関しては，確かにその場所を写してはいるが，必ずしも現在の場所で起きた出来事や状況などに誘因されて投稿されているとはいいきれないと考えたため，場所誘因型の分類からは除外した．また，移動中であるという発言に関しては，これから移動を始める場合，移動している最中の場合，移動し終わった場合など多くの状態が考えられるとともに，現在地の地名を述べているだけの発言と区別が困難であったため，対象外とした．さらに，事故現場・災害現場の目撃情報はそれぞれ 9 件（0.4%），2 件（0.1%）であり，交通機関に関する情報については計 14 件（0.7%）と全体の 1%未満であったため，本研究では分類から除外した．これらの妥当性は，6 章において評価を行う．

4. 場所誘因型位置情報付き発言の検出

本章では，まず場所誘因型位置情報付き発言の検出を行うためのアプローチについて述べる．次に，本研究で提案する場所誘因型位置情報付き発言検出システムの設計と実装について述べる．

4.1 アプローチ

それぞれの場所誘因型位置情報付き発言の種別について，効果的に検出を行うためのフィルタを作成する．主に，キーワードマッチングにより判定を行うアプローチとした．場所誘因型位置情報付き発言であるかを人間の目で判断する場合，第 1 に発言本文に特徴的なキーワードが含まれることを手がかりにしていると考えられるためである．一方，本アプローチでは，あらかじめ用意した 5 種類の分類に適合する発言は検出可能であるが，場所誘因型でありながらこの分類に適合しない発言に関しては検出することができないという制約がある．本研究では，最も単純な手法でどの程度場所誘因型位置情報付き発言を検出可能であるかを実証するために，本アプローチを選択した．

以下にそれぞれのフィルタにおけるアプローチの詳細を述べる．まず，居場所の報告については，チェックイン系サービスの URL が含まれているかという点のみで判定を行った．これは事前調査の結果から，居場所の報告に該当する多くの発言が，チェックイン系サービスを利用したもので占められていたためである．判定手法としては，Twit-

terAPI から取得した発言情報に含まれる，source という項目を参照している．ここにはユーザが発言を行ったクライアント情報が明記されているため，チェックイン系サービスから Twitter に発言が投稿された場合，各サービスの URL が含まれることになる．たとえば，Foursquare から投稿された場合は，＊”\u003Ca href=\”http:\/\/foursquare.com\” rel=\”nofollow\”\u003Efoursquare\u003C\/a\u003E”のような記述となる．本研究では，Foursquare [9]，ロケタッチ [10]，今ココなう！ [11] の URL が含まれる場合，チェックイン系サービスからの投稿であると判定する．

食事・天候・帰宅の種別については，発言本文に特定のキーワードが含まれるかどうかで判定を行う．キーワードによって判定を行う例として，まず食事の種別について述べる．「食事」というキーワードの類義語辞書を用いて，食事に関連する類義語一覧を作成する．類義語辞書は，weblio 類語辞典 [12] を参照した．作成した類義語一覧を用い，発言本文に「食事」の類義語が含まれる場合に，食事の種別であるという判定を行う．食事フィルタで用いた単語の登録数は 86 個であり，「朝食」「メシ」「食べる」「食う」「ディナー」などの単語が含まれる．次に，「天候」についても同様に類義語辞書を作成し，フィルタを作成した．天候フィルタで用いた単語は 131 個であり，「晴れ」「天気」「曇り」「雨」「寒さ」などが含まれる．さらに，「帰宅」についても同様にフィルタを作成した．帰宅フィルタで用いた単語は 5 個であり，「帰宅」「帰った」「帰着」「帰還」「朝帰り」を登録した．

食事・天候・帰宅については，類義語以外にも関連するキーワードが数多く考えられる．しかし，どのような関連語を選択するかは著者の主観に影響されてしまい，キーワードマッチング手法自体の評価が正しく行えない恐れがあったため，今回は広く公開されている類義語辞書をそのまま利用することにした．

最後に，地震については独自に作成したキーワード一覧が発言本文に含まれるかどうかで判定を行う．発言本文に「揺れた」「ゆれた」「ゆれてる」「地震」「じしん」の 5 個のキーワードが含まれている場合，地震の種別であると判定する．地震が発生した場合，ユーザは漢字変換を行う余裕もなく投稿を行う場合が多く見受けられたため，未変換の単語も登録することで対応した．

取得した発言は，以上のフィルタによってタグ付けを行われた後，判定結果として出力される．各フィルタにおいては，辞書に登録されたすべての単語を含む 1 つの正規表現として保存しており，正規表現が発言本文とマッチした場合，その発言は当該フィルタの種別であると判定される．たとえば，「帰宅した直後に揺れた」という発言を判定する際，帰宅フィルタの正規表現：/帰宅｜帰った｜帰着｜帰還｜朝帰り/ と地震フィルタの正規表現：/揺れた｜ゆれた｜ゆれてる｜地震｜じしん/ にマッチするするため，「帰宅」かつ「地震」であると判定される．このように，互いのフィルタは独立しているため，1 つの発言は複数の種別と判定されうる．

なお，発言本文を形態素解析し，表記の揺れを取り除いたうえで判定を行うアプローチも検討した．形態素解析ツールには mecab を用い，動詞・形容詞を原形に直した後，上記のキーワードフィルタを適用した．約 1 時間分の発言に相当する 3,000 件を解析してみたところ，形態素解析を行わない場合は 40 秒程度で判定が行えたが，形態素解析処理を追加したことで処理時間が約 10 倍増加した．しかし，精度については 6 章で述べる手法と同様に評価を行ったが，形態素解析を行った場合と行わない場合で F 値に顕著な差は見られなかった．そのため，本研究では単純なキーワードマッチングを適用し，その有効性を評価する．

4.2 設計と実装

Twitter などのソーシャルメディアから，位置情報付き発言の情報を取得できることを前提とする．位置情報は，ユーザがモバイルデバイスから投稿を行う際，GPS を用いて緯度経度を取得し，発言に付加していることを想定する．

本研究で提案するシステムは，位置情報付き発言取得モジュール，位置情報付き発言解析モジュール，データベース，可視化アプリケーションで構成される．本システムのモジュール構成を**図 1** に示す．位置情報付き発言取得モジュールでは，Twitter から日本周辺の位置情報が付加された発言の取得を行う．Twitter Streaming API [13] に日本周辺の緯度経度をパラメータとして渡し，リアルタイムに発言を取得している．1 日あたりの取得量は約 5〜7 万件である．位置情報付き発言解析モジュールでは，4 章で述べた手法を用い，場所誘因型位置情報付き発言の検出を行う．取得した発言が場所誘因型である場合は，どの種別に該当するかを判定し，解析結果をデータベースに保存する．可視化アプリケーションは，場所誘因型位置情報付き発言から実世界イベントの検出を支援することを目的とする．実装したアプリケーションについては，5 章で詳細を述べる．

場所誘因型位置情報付き発言の検出および可視化システムの実装環境について述べる．実装および運用に

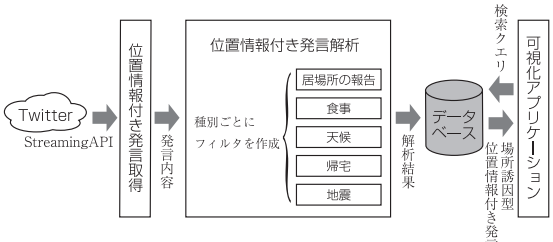

図 1 モジュール構成図
Fig. 1 Module configuration.

は，CPU：Intel Core2 Quad Q9550 @ 2.83 GHz，メモリ：4 GB，HDD：1 TB のデスクトップ PC を用いた．システムのソフトウェア構成は．OS：Debian GNU/Linux，Web サーバ：Apache HTTP Server 2.2.9，解析プログラム：Ruby 1.9.2 p290，可視化プログラム：PHP ver. 5.2.6，GMT 4.5.2，データベース：MySQL ver. 5.0.51a を用いて実装を行った．

5. アプリケーション

本章では，場所誘因型位置情報付き発言の検出を利用したアプリケーションについて述べる．まず，場所誘因型位置情報付き発言の可視化により，実世界イベントの発見支援を可能とするアプリケーションについて述べる．次に，ユーザとの対話性をより重視し，個人の要求に沿った可視化を可能とするアプリケーションについて述べる．

本アプリケーションにおいては，位置情報付き発言解析モジュールのフィルタで検出された発言を場所誘因型，そうでない発言をノイズとして扱っている．しかし，本来は場所誘因型でありながら，フィルタによって取りこぼされてしまった発言もノイズに含まれてしまうという制約がある．

5.1 アニメーション生成による可視化

解析を行った位置情報付き発言を地図上にプロットするアプリケーションを作成した．日本全体，関東，東京など任意の範囲と，任意の時間間隔でプロットすることが可能である．**図 2** では，2012 年 4 月 29 日の位置情報付き発言を 4 時間間隔で抽出した結果をプロットした．A 列はすべての位置情報付き発言をプロットしたもの，B 列は場所誘因型位置情報付き発言（Place-triggered）を青色の○，それ以外のノイズ（Noise）をグレーの×印でプロットしたもの，C 列は場所誘因型位置情報付き発言の種別ごとに印を変えてプロットしたものである．居場所の報告（Report of whereabouts），食事（Food），天候（Weather），帰宅（Return Home），地震（Earthquake）とノイズ（Noise）を示している．プロット結果を連続した画像として書き出し，アニメーションを生成することも可能である．

場所誘因型位置情報付き発言の可視化を行ったことで，興味深い事例を発見することが可能となった．時系列でプロットを行ったことにより，左下のプロットにおける A-1–A-5 で囲った発言群が，一定速度で北東に移動していることが明らかになった．このままでは意味のある発言であるか否かは推定できないが，中央の列でプロットした結果が示すように，場所誘因型ではないノイズであると判定された．実際に調べてみると，位置情報付き発言を行うフォロワーを追いかける Bot であることが判明した．この Bot の発言は，特にその場所の位置情報と関係するものではなく，機械的に生成されたものであった．また，場所誘因型位置情報付き発言の種別を区別してプロットを行った結果，C-1 で示すプロットの時間帯で，地震に関する発言が急激に出現していることが分かった．2012 年 4 月 29 日 19 時 28 分ごろ，千葉県北東部を震源とするマグニチュード 5.8，最大震度 5 弱の地震が発生しており，実際の地震を明らかに反映しているということがいえる．このように，場所誘因型位置情報付き発言をプロットし可視化を行うことで，実世界イベントの検出が容易に実現可能になったと考えられる．

5.2 Web ベースのインタラクティブな可視化インタフェース

Google Maps 上に任意の時間・範囲の発言をプロットするアプリケーションを作成した．**図 3** に，可視化インタフェースのスクリーンショットを示す．左の地図上には，場所誘因型位置情報付き発言の種別ごとに色分けされたピンがプロットされている．場所誘因型でないと判定された発言は，グレーのピンでプロットされる．ピンをクリックすることで，発言の本文や付加されている写真などの詳細な情報も表示可能である．右上のパネルでは，発言を表示する日時や時間帯を簡単に調節できる．右下のチェックボックスでは，5 種類の場所誘因型位置情報付き発言の種別に加え，発言に写真が付加されているかの計 6 項目でフィルタリングが可能である．地図をスクロールするごとに，表示されている範囲の緯度経度における発言のみをデータベースから取得するため，インタラクティブな操作感覚を実現した．

前述のプロットツールとは異なり，発言の本文や写真まで表示することで，より詳細なイベントの判定を行うことが可能である．たとえば，食事かつ写真で検索することで，よく食事の写真が共有されているスポットを発見したり，天候かつ写真で検索し，ゲリラ豪雨の様子を発見することなどが実現した．**図 4** に，2012 年 8 月 6 日の夕刻に関西地域で発生したゲリラ豪雨の様子を発見した例を示す．日時と「天候」の種別で絞り込むことで，特定の地域・時間において顕著に発言が増加したことから，ゲリラ豪雨が発生していたことが確認された．

Google Maps 上にプロットするアプリケーションでは，API の制限やパフォーマンス上の問題から，同時に 200 件までしかプロットできないという制約がある．本研究の提案する手法によってノイズを除去することで，より重要な発言のみプロットを行うことが可能となる．今後は Web アプリケーションとして公開することで，より多くのユーザに使ってもらうことを目指している．

6. 評価

本研究における評価は大きく 2 つの観点から行った．まず，場所誘因型位置情報付き発言の分類が妥当であるかど

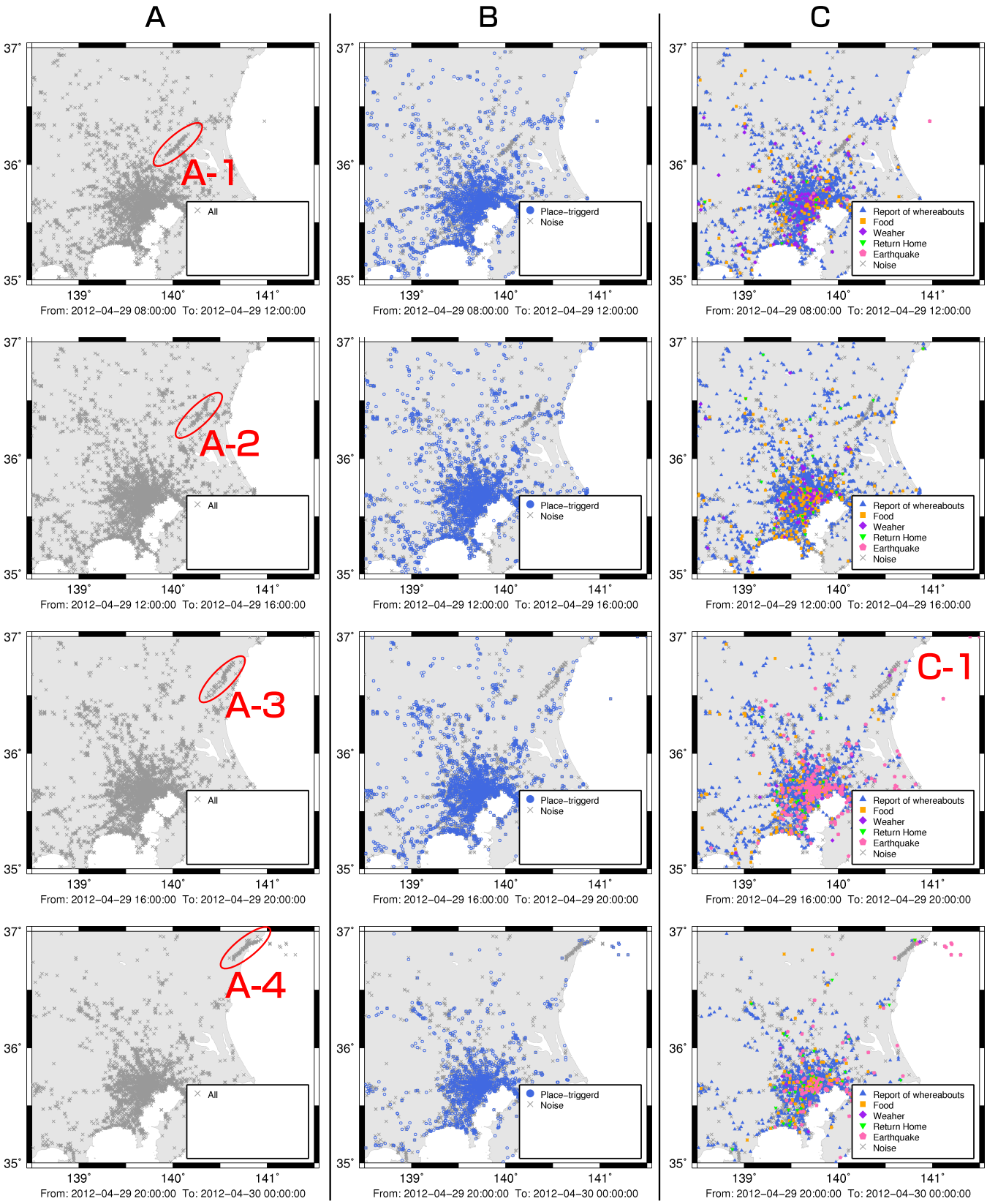

図 2 場所誘因型位置情報付き発言のプロット結果

Fig. 2 Plotted result of place-triggered geotagged tweets.

うかを検証するため，18 名の被験者に対し，実際に Twitter から取得した位置情報付き発言を提示し，それが場所誘因型であるかどうかの分類を行ってもらった．次に，本研究で提案する位置情報付き発言解析モジュールの性能を評価するため，前述した 18 名による判定結果を正解データとし，システムの判定した結果との比較を行った．以下に，

図 3 インタラクティブな場所誘因型位置情報付き発言可視化インタフェース

Fig. 3 Interactive visualization interface for place-triggered geotagged tweets.

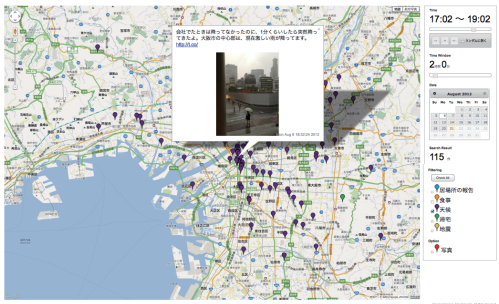

図 4 ゲリラ豪雨の様子を検出した例

Fig. 4 Example of heavy rain event detection.

それぞれの詳細を述べる．

6.1 第三者による位置情報付き発言の分類

6.1.1 評価方法

事前調査では，著者本人の主観のみによって場所誘因型位置情報付き発言の分類を行ったが，その分類が妥当であるかどうかを検証するため，より多くの人によって発言の分類を行ってもらった．被験者は，20 代の男性 12 名，20 代の女性 4 名，30 代の男性 2 名の計 18 名であった．分類を行ってもらう対象は，2012/01/01 00:00:00〜2012/03/31 23:59:59 の期間に取得した計 4,254,257 件の位置情報付き発言の中から，被験者 1 人あたり 500 件をランダムに抽出した．

評価用のツールとして，**図 5** に示すような，Web ベースのアプリケーションを作成した．評価実験を行う際，被験者に対しあらかじめ場所誘因型位置情報付き発言のコンセプトを説明した．被験者には，発言本文，時間，発言に設定された位置情報をマップにプロットしたものを提示する．これらの情報から，まずは発言が場所に誘因されているか，されていないかを判断してもらうように伝えた．発言が場所誘因型であると思った場合には，居場所の報告，食事，天候，帰宅，地震に分類できるかどうかを尋ねる．どの種別にも分類できないが，明らかに場所誘因型であると思われる場合にはその他を選択してもらう．一方，場所誘因型ではないと判断した場合は，何も選択せずに次に進んでもらうという流れで評価を行った．複数の種別に該当すると思われる場合は，重複して選択することを可能にしている．

図 5 発言分類評価用アプリケーション

Fig. 5 Application for evaluation of tweet classification.

表 2 第三者による場所誘因型位置情報付き発言の分類結果

Table 2 Result of place-triggered geotagged tweets classification by third person.

発言種別	判定数	割合
居場所の報告	4,300 件	86%
食事	664 件	13%
天候	255 件	5.1%
帰宅	61 件	1.2%
地震	44 件	0.88%
その他	354 件	7.0%
合計	5,678 件	

なお，「天気」に関しては今後のより詳細な判別手法の構築を試みるため，「良い天気」「悪い天気」の 2 つに分割して選択させた．本論文においては，場所誘因型位置情報付き発言の種別としては「天気」として扱うため，両者の分類は区別せず，1 つの「天気」という分類にまとめた分析を行った．

6.1.2 評価結果

計 8,988 件の発言に対して判定結果が得られた．本来，1 人あたり 500 件依頼しているため 9,000 件の評価結果が得られるはずだが，ユーザが評価用アプリケーションを操作する際，ボタンを連打してしまったり，正しく URL が指定されないことがあり，1 人あたり約 1 件ほどの取りこぼしが発生してしまった．

場所誘因型であると判定された発言は 5,027 件 (55.93%)，場所誘因型ではない（ノイズ）と判定された発言は 3,961 件（44.07%）という結果になった．場所誘因型と判定された発言の内訳を**表 2** に示す．1 つの発言につき複数の種別を回答することが可能なため，判定された種別を合計した結果が 5,678 件となった．

6.1.3 考察

実験の結果，場所誘因型位置情報付き発言として判定さ

れた発言のうち，延べ 5,324 件が本研究が分類したカテゴリ内に属することが分かった．一方，延べ 354 件の発言が，あらかじめ用意した 5 種類の種別以外であるが，被験者によって場所誘因型だと判定された．たとえば，“○○を買いに行く” や “○○のイベントに参加” などの行動を表す発言や，その場所の天気を自動で発言する Bot などが，その他と判定されている場合が多かった．しかし，判定を行う人によって基準が大きく異なっており，特定のカテゴリのキーワードとして傾向を見いだすことは困難であった．これらの中から，新たな種別として分類することが可能かどうかは，今後の検討課題であると考えられる．

また，事前調査では過半数の発言をノイズとして判定していたが，評価の結果，ノイズは 44.07%と差異が生じた．この大きな原因は，事前調査の時点ではチェックイン系サービスと連携している発言のみを「居場所の報告」として扱っていたが，実際にはチェックイン系サービスを利用していなくても，居場所について言及している発言が多く見受けられ，それを被験者が居場所の報告に分類したためである．たとえば，“病院なう．41 度熱がある.” といった発言が居場所を報告している例としてあげられる．評価実験では，このような非チェックイン系サービスによる居場所の報告が多く判定された結果，ノイズとなる発言が減少したと考えられる．なお，表 2 の分類結果において，場所誘因型位置情報付き発言に占める居場所の報告の割合は 86%という結果になり，多くのユーザが位置情報を居場所の報告に利用しているということが分かる．一方，今回の実験では，地震のカテゴリに分類された場所誘因型位置情報付き発言は，0.88%と，非常に少なかった．これは，評価対象の発言の中に含まれる地震に関する発言の割合が，事前実験の際の割合よりも少なかったことが理由としてあげられる．これらのことから，本研究で分類した場所誘因型位置情報付き発言の分類はおおむね妥当であったと考えられるが，一方でその他の合計 7.0%は地震や帰宅に関する発言の割合よりも多かったため，前述したようにより適切な分類方法が求められることが分かった．

6.2 位置情報付き発言解析モジュールの評価

6.2.1 評価方法

本研究で提案する位置情報付き発言解析モジュールの性能を評価するため，前節で取得した判定結果を正解データとし，システムの判定した結果との比較を行った．評価対象として 18 名の被験者に判定を行ってもらった計 8,988 件の発言を用い，位置情報付き発言解析モジュールにより判定を行った．

評価項目として，位置情報付き発言が場所誘因型であるかどうかの正解率を求める．その際，False positive 率と False negative 率を示す．また，場所誘因型位置情報付き発言の各種別に対して，以下に定義する適合率，再現率，F 値を求める．

- 適合率（precision）
  解析対象のうち正解データと適合していた発言数を R，解析対象の総発言数を N とし，**適合率** $precision = R/N$ を求める． 適合率が高いほど，システムの判定した結果にノイズが少ない．
- 再現率（recall）
  解析対象のうち正解データと適合していた発言数を R，正解データの総数を C とし，**再現率** $recall = R/C$ を求める． 再現率が高いほど，正解データからの取りこぼしが少ない．
- F 値（F-measure）
  precision と recall の調和平均である，**F 値** $F\text{-}measure = \frac{2 \cdot precision \cdot recall}{precision + recall}$ を求める． F 値が高いほど，全体としての性能が良いといえる．

6.3 評価結果と考察

位置情報付き発言解析モジュールによって場所誘因型であると判定された発言は 3,799 件（42.27%），場所誘因型ではないと判定された発言は 5,189 件（57.73%）であった．場所誘因型と判定された発言の内訳と，各種別における判定結果を**表 3** に示す．1 つの発言につき複数の種別が判定されうるため，判定された種別を合計した結果が 3,938 件となった．

まずは本研究の提案するシステム全体としての評価を考察する．**表 4** に示すように，82%の精度で場所誘因型位置情報付き発言を正しく判定可能であった．False positive 率は 2.18%であり，ノイズをほぼ除去することが可能になったといえる．一方，False negative 率は 15.84%を示しており，場所誘因型と見なされる発言を取りこぼしていることが分かる．

次に，それぞれの場所誘因型位置情報付き発言の種別に

表 3 システムによる場所誘因型位置情報付き発言の分類結果
Table 3 Result of place-triggered geotagged tweets classification by system.

発言種別	判定数	割合	適合率	再現率	F 値
居場所の報告	3,561 件	93%	93.18%	77.16%	84.42%
食事	220 件	5.8%	53.6%	17.8%	26.7%
天候	93 件	2.4%	57%	21%	30%
帰宅	26 件	0.68%	54%	23%	32%
地震	38 件	1.0%	76%	66%	71%
合計	3,938 件				

表 4 場所誘因型位置情報付き発言の正解率
Table 4 Answer rate of place-triggered geotagged tweets.

	Positive	Negative
True	40.09%	15.84%
False	2.18%	41.89%

ついて考察する．位置情報付き発言解析モジュールの判定結果を表 3 に示す．

居場所の報告は，検出した場所誘因型位置情報付き発言に対する割合の 93%を占めた．これは表 2 に示した被験者による分類と比較すると，7 ポイントほど高い結果となった．検出精度については F 値が 84.42%を示し，高い精度で検出可能であったといえる．93.18%の適合率に対し，再現率が 77.16%と取りこぼしが発生した原因としては，システムはチェックイン系サービスと連動した発言のみを居場所の報告として扱っていたが，実際にはチェックイン系サービスを用いないで居場所を報告している発言が存在していたためである．取りこぼした発言の例として，「富士 SA なう。夕食にしよう。」という発言があげられる．この発言に対し被験者は「居場所の報告」のみと判定したが，システムはチェックインサービスの URL が含まれていないため居場所の報告とは判定せず，本文中の夕食というキーワードから「食事」と判定した．このように，被験者は発言本文を見て主観的に判断しているため，キーワードや URL が含まれていても文脈や個人の感覚の差異によって，判定結果に大きく影響することが分かった．さらに，「○○なう」など Twitter 特有の発言形式において，○○には場所の名詞だけでなく様々な動詞も含まれることが確認された．したがって，非チェックインの発言から居場所の報告を検出するためには，単純なキーワードマッチングのみでは実現が難しく，場所を表す表現を抽出したり，文脈を考慮したりすることが必要だと考える．また，荒川らは，Twitter の位置情報付き発言における本文と位置情報の相関関係を分析している [14], [15]．これらの提案手法が場所誘因型位置情報付き発言の種別を特定する際にも応用可能かどうか，今後の課題として検討していく．

食事・天候・帰宅については，それぞれ 5.8%，2.4%，0.68%と，被験者による分類結果の割合と比べて約半分程度にとどまった．検出精度としては適合率が 54%〜57%と，約 4 割のノイズを含む結果となった．特に再現率が低下した原因として，種別を特定するために類義語から生成した辞書を用いていたが，実際はそれ以外の単語でも食事や天候などを表現している例が多く見られたためだと考えられる．たとえば，「お好みソースでオムライス作った♪(´▽`) http://[写真 URL]」という発言は，被験者によって食事と判定されたがシステムによって検出できなかった．この発言を正しく食事という種別に分類するためには，「ソース」や「オムライス」などが食材や料理名であると判定するか，写真に料理が写っていることを認識することが求められる．そのため，今後は単純なキーワードマッチングだけではなく，より詳細な本文解析を行っていく必要がある．

地震に関しては，F 値が 71%を示し，比較的精度良く検出できたといえる．これは，地震を報告する発言がある程度限定された単語であることが多いため，キーワードマッチングのアプローチが適していたためと考えられる．

最後に，場所誘因型位置情報付き発言の検出精度について議論を行う．前述のとおり，82%の精度で場所誘因型位置情報付き発言を判定可能であったが，検出した場所誘因型位置情報付き発言の 93%は居場所の報告で占められており，チェックイン系サービスの判定精度に依存していると考えられる．今後は，本来は場所誘因型でありながら取りこぼされてしまった発言や，本システムが誤検出してしまったノイズを減少させるべく，場所誘因型位置情報付き発言の判定手法を改良していくことを課題とする．また，検出した場所誘因型位置情報付き発言から，実世界イベントを検出する手法についても検討を重ねていく．

7. まとめ

ソーシャルメディアの位置情報付き発言から実世界イベントの検出を行う際，位置情報が付加されていてもその場所と関係のない発言がノイズとなる点が問題となる．本研究において，位置情報付き発言の中でも，現在の場所で起きた出来事・状況などに誘因されて発言されたものを「場所誘因型位置情報付き発言」と定義し，その検出を行った．場所誘因型位置情報付き発言を「居場所の報告」,「食事」,「天候」,「帰宅」,「地震」の 5 種類に分類し，発言本文に特徴的なキーワードが含まれることを手がかりに検出手法を構築した．評価として 18 名の被験者によって位置情報付き発言の分類を行い，提案手法が 82%の精度で場所誘因型位置情報付き発言を検出可能であると示した．また，場所誘因型位置情報付き発言の可視化を行うアプリケーションを作成し，実世界イベントが検出可能である例を示した．以上のことから，位置情報付き発言のノイズを除去することにより，実世界イベントの検出に貢献したと結論づける．今後の課題として，単純なキーワードマッチングだけではなく高速かつ詳細な本文解析を行っていくことで，より正確なノイズ除去を行うことがあげられる．さらに，検出した場所誘因型位置情報付き発言を用いて，より効果的に実世界イベントの検出を行う手法を構築してゆくことを目指す．

謝辞 本研究の一部は株式会社 NTT ドコモとの共同研究として，また独立行政法人情報通信研究機構にご支援いただいた．

参考文献

[1] Twitter, Inc.: Twitter (online), available from ⟨http://www.twitter.com/⟩ (accessed 2012-05-14).
[2] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *Proc. 19th International Conference on World Wide Web*, pp.851–860 (2010).
[3] Lee, R. and Sumiya, K.: Measuring Geographical Regu-

larities of Crowd Behaviors for Twitter-based Geo-social Event Detection, *Proc. 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pp.1–10 (2010).
[4] 石川翔太，荒川　豊，田頭茂明，福田　晃：マイクロブログを用いた地域におけるホットトピック検出手法の検討，マルチメディア通信と分散処理ワークショップ，DPSWS2011, pp.77–82 (2011).
[5] 藤坂達也, 李　龍, 角谷和俊：マイクロブロガーの移動履歴を用いた地域特性分析，情報処理学会研究報告，DBS-149, Vol.2009, No.17, pp.1–8 (2009).
[6] Yin, X., Han, J. and Yu, P.: Truth Discovery with Multiple Conflicting Information Providers on the Web, *IEEE Trans. Knowledge and Data Engineering*, Vol.20, No.6, pp.796–808 (2008).
[7] Wang, D., Abdelzaher, T., Ahmadi, H., Pasternack, J., Roth, D., Gupta, M., Han, J., Fatemieh, O., Le, H. and Aggarwal, C.: On Bayesian interpretation of fact-finding in information networks, *Proc. 14th International Conference on Information Fusion*, pp.1–8 (2011).
[8] Wang, D., Kaplan, L., Le, H. and Abdelzaher, T.: On truth discovery in social sensing: A maximum likelihood estimation approach, *Proc. 11th International Conference on Information Processing in Sensor Networks, IPSN '12*, pp.233–244 (2012).
[9] FOURSQUARE LABS, INC.: Foursquare (online), available from ⟨http://foursquare.com/⟩ (accessed 2012-05-14).
[10] NHN Japan Corp.：ロケタッチ（オンライン），入手先 ⟨http://tou.ch/⟩（参照 2012-05-14）.
[11] fujita-lab.com：今ココなう！(β)（オンライン），入手先 ⟨http://imakoko-gps.appspot.com/⟩（参照 2012-05-14）.
[12] Weblio, Inc.：weblio 類語辞典（オンライン），入手先 ⟨http://thesaurus.weblio.jp/⟩（参照 2012-05-14）.
[13] Twitter, Inc.: Streaming API (online), available from ⟨https://dev.twitter.com/docs/streaming-api⟩ (accessed 2012-05-14).
[14] 荒川　豊，田頭茂明，福田　晃：Twitter におけるコンテキストと単語の相関関係分析，情報処理学会研究報告，MBL-53, Vol.2010, No.50, pp.1–7 (2010).
[15] 荒川　豊，田頭茂明，福田　晃：Twitter を用いたコンテキストと入力文字列の相関関係分析，情報処理学会論文誌，Vol.52, No.7, pp.2268–2276 (2011).

蛭田 慎也

2011 年慶應義塾大学総合政策学部卒業．現在，慶應義塾大学大学院政策・メディア研究科修士課程．主に，ユビキタスコンピューティングシステム，サイバーフィジカルシステムの研究に従事．ACM 学生会員．

米澤 拓郎　（正会員）

2007 年慶應義塾大学大学院政策・メディア研究科修士．2010 年慶應義塾大学 Ph.D.（政策・メディア）．現在，慶應義塾大学大学院政策・メディア研究科特任助教．主に，ユビキタスコンピューティングシステム，インタラクティブシステム，センサネットワークの研究に従事．ACM，日本ソフトウェア科学会各会員．

徳田 英幸　（正会員）

1977 年慶應義塾大学大学院工学研究科修士．1983 年ウォータールー大学 Ph.D.（Computer Science）．同年カーネギーメロン大学計算機科学科勤務．1990 年同学科研究准教授．現在，慶應義塾大学大学院政策・メディア研究科委員長．主に，ユビキタスコンピューティングシステム，オペレーティングシステム，分散システムに関する研究に従事．IEEE，ACM，日本ソフトウェア科学会各会員．