情報処理学会論文誌　データベース　Vol.9　No.3　14–24 (Sep. 2016)

Twitterにおける空間差異に基づく群衆の多面的関心分析

若宮 翔子[1,a)]　ヤトフト アダム[2,b)]　河合 由起子[3,c)]　秋山 豊和[3,d)]　荒牧 英治[1,e)]

受付日 2016年3月20日, 採録日 2016年7月5日

概要：Twitter に代表されるソーシャルメディアはイベント検出で頻繁に活用されている．このようなシステムでは，ツイートを特定の時間や位置に割り当てるために，タイムスタンプ（日時）やロケーションスタンプ（緯度経度）などのメタデータが重要な役割を果たしている．一方で，コンテンツに書かれたテキスト文中の時間や位置に関する表現は曖昧な場合があり，メタデータよりも信頼性が落ちることから，十分な活用が困難であった．このため，我々はメタデータと時空間表現の差異分析を可能とする，大規模ソーシャルメディアデータの可視化システムを開発している．本稿では，特に空間に対する群衆の関心を分析するため，ツイートの発信位置（メタデータ）とコンテンツテキスト中の位置表現の差異を可視化するシステムを提案する．実験では，約 3 カ月分の米国で発信されたツイートを用いて 3 種類のデータビューを構築し，空間的な尺度に基づき群衆の空間的関心に関する分析結果の例を示して考察する．提案したフレームワークや考察は，ソーシャルメディアデータの地理的・社会的な側面に関心を持つユーザにとって有用であると考えられ，また，将来的に，テキストの位置情報を用いたメタデータの補完に有用であると期待される．

キーワード：ソーシャルメディア，空間分析，可視化，Twitter

Collective Attention Analysis Based on Spatial Differences in Twitter

Shoko Wakamiya[1,a)]　Adam Jatowt[2,b)]　Yukiko Kawai[3,c)]
Toyokazu Akiyama[3,d)]　Eiji Aramaki[1,e)]

Received: March 20, 2016, Accepted: July 5, 2016

Abstract: Social media data such as tweets in Twitter have been frequently used for detecting real-time events. The spatio-temporal metadata of the social media data such as timestamp and location stamp usually play a key role for assigning tweets to a specific time and space. On the other hand, it is difficult to utilize expressions about location and time in tweet contents since these are sometimes ambiguous and less reliable. In this paper we propose a novel visualization system focused on spatial information for analyzing how users collectively talk about space and for uncovering differences between geographical locations of users and the locations they tweet about. Our exploratory analysis is based on the development of a model of spatial information extraction and representation that allows building effective visual analytics framework for a large scale dataset. We demonstrate examples of analysis results based on a three months-long collection of tweets from USA. The proposed system allows observing many space-related aspects via three types of data views. The system enables to visualize average scope of spatial attention of users. The framework and the findings can be valuable for scientists from diverse research areas and for any users interested in geographical and social aspects of shared online big data. Furthermore, it is expected to be useful to complement metadata using textual location information.

Keywords: microblogs, spatial analysis, visualization, Twitter

1. はじめに

Twitter に代表されるソーシャルメディアは，人々が自

[1] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Ikoma, Nara 630–0192, Japan
[2] 京都大学
Kyoto University, Kyoto 606–8501, Japan
[3] 京都産業大学
Kyoto Sangyo University, Kyoto 603–8555, Japan
[a)] wakamiya@is.naist.jp
[b)] adam@dl.kuis.kyoto-u.ac.jp
[c)] kawai@cc.kyoto-su.ac.jp
[d)] akiyama@cc.kyoto-su.ac.jp
[e)] aramaki@is.naist.jp

ら発信・共有している意見や日常活動の大規模なログデータとして注目を集めており，イベント検出 [1] やユーザ間の情報拡散プロセス分析 [2] などの様々なアプリケーションシステムに活用されている．特に，他のメディアと違って，ソーシャルメディアは，タイムスタンプ（日時）やロケーションスタンプ（緯度経度）といったメタデータが一部に付与されており，これらに基づき，時空間やイベントに基づくトピックの人気度の分析 [3] やローカルイベントの発見 [4], [5] など，時間や位置を用いたマルチモーダルな分析が可能である．ただし，すべてのデータにこのようなメタデータが付与されているわけでない．よって，コンテンツのテキスト文に含まれる時間や位置に関する表現を利用した分析が試みられる場合がある．しかし，ユーザが自由に記述しているため曖昧な場合があり，十分な活用が困難であった．さらに，メタデータはデータが発信された時間や位置であるのに対し，表現はデータが発信された時間や位置だけでなく，イベントの発生時間や位置を指している場合もあれば，ユーザが個人的に関心を持っている時間や位置を指している場合もある．たとえば，先月訪問した場所を思い出したり，次の休暇の旅行計画を立てたり，開催中のイベントや単に興味を持っている場所について意見や感想を述べたりするときに，時間や位置に関する表現を用いるユーザは多い．他にも，突発的なイベントや暦上のイベントが発生した場合には，多くのユーザがいっせいに同じ時間や位置へ関心を向けることがある．

このように，メタデータから得られる時間やユーザの実際の位置は，ユーザの関心と一致することもあれば，異なることもあり，そのパターンについては十分に研究されていない．我々は，人々の興味の対象や焦点となっている時間や位置を，群衆の時空間的関心とし，この俯瞰を試みている．我々はすでに Twitter ユーザの話題や時間についての関心（時間的関心）に着目した分析のための可視化システムを開発している [6]．

本稿では，時間的関心とは直交する群衆の位置についての関心（**空間的関心**と記す）を俯瞰するための可視化システムを構築する [7], [8]．そのために，コンテンツのテキスト中の位置表現に基づく参照位置の推定，ツイートの発信位置と参照位置の差異（空間差異）の判定，空間的関心を表現するためのデータビューの設計・構築を行う[*1]．実際に米国を対象に収集した約 3 カ月分のツイートを分析し，ソーシャルメディアユーザの空間的関心を可視化する．提案システムにより，様々な人々の話題が比較可能になり，また，空間的関心の共通性による場所の特徴付けや分類が可能になる（空間的分析）．さらに，時間経過による空間的関心の変化の分析も可能である（時間を組み合わせた解析）．このように，提案システムの可視化フレームワークは群衆の空間的関心のレイヤを描画し，様々な時空間的な分析や発見を可能にすると期待される．

本研究のポイントは以下の 2 点である．

(1) ソーシャルメディアデータの 3 つの新しい可視化方式を提案した点
(2) 米国のツイートを材料に，人々の空間的関心を可視化して分析例を考察することで，提案する方式の有効性を実証した点

本稿の構成は以下のとおりである．2 章で関連研究をまとめ，3 章でデータモデルと空間差異を分析するための可視化システムの概要について述べる．4 章ではツイートのコンテンツのテキストにおける位置表現の参照位置を推定する手法を述べる．5 章で米国のデータセットを用いて実装した 3 つのデータビューについて説明し，6 章でそれらを用いた分析例を示して考察する．最後に 7 章でまとめと今後の課題を述べる．

2. 関連研究

ソーシャルメディアの普及により，人々の振舞い，特徴や社会的なつながりに関する大量のデータを容易に取得可能になっている．Twitter データ分析は，人々の大量の行動ログを利活用する社会科学においてよく行われている [9], [10]．Goonetilleke らはこれまでにツイートを活用して実施された様々な研究について調査している [11]．ツイートにおける話題の多様性は，空間的な観点から人々の関心事を大規模に分析することの価値を高めている．

空間に着目した研究はきわめて重要であり，実際に日常的なコミュニケーションは地理空間を中心話題としたものが多い．たとえば，世界のニュースメディアはどんな場所についても 200 から 300 語で説明する傾向があり，これは他の種類の情報を述べる場合よりも多い [12]．情報アクセスの面でも空間属性との関連性は高い．たとえば，Web 検索において地理的な単語を含んでいる割合は 1/4 以上を占めており，また，全 Web 検索のうち 13%は地理的な特徴を持つことが報告されている [13]．

ソーシャルメディアにおいては，すべてのツイートにロケーションスタンプが付与されているわけではないため，このようなツイートに対する位置推定が 1 つの主要な研究トレンドとなっている．位置推定に関する研究では，ソーシャルグラフをマイニングするアプローチ [14], [15] とテキストコンテンツを解析するアプローチ [16] の 2 つが主流である．後者はローカル語の概念を導入しており，後に拡張されたものが Twitter ユーザの居住地推定に用いられている [17], [18]．このアプローチはトピックモデルに基づく推定手法にも応用されている [19], [20], [21]．統計的なモデルにより，ユーザの履歴，タグや他の属性から自宅の場所を推定したり [22], [23]，ユーザの移動を予測したりすることも可能になっている [15]．さらに，ソーシャルグラフと

[*1] http://delab.kyoto-su.ac.jp/tweet/US7/chartlist.html

コンテンツ解析を組み合わせることで，ユーザが位置情報を非公開にし続けていても，ユーザの位置をより詳細に推定することが可能である [24], [25], [26]．これらの研究はすべて Twitter などで共有されているデータセットを用いて，ユーザの位置を判定可能であることを実証している．

これまでの研究は，イベント検出によりリアルタイム性の高い情報を抽出するために，ソーシャルメディアの有用性を強調してきた [27], [28], [29]．火災や地震検出のような特定のアプリケーションでは，多面的にツイートを探索するための視覚化機能一式も提供されている [1], [30], [31]．さらに，一般的なプラットフォームでは，空間的な観点でデータを分析するための可視化ツールを持ち合わせているものもある [32], [33], [34]．

ロケーションスタンプのようなメタデータとコンテンツ中の位置表現の差異を分析し活用する研究は多数行われている [17], [35], [36], [37]．一方で，我々の知識の及ぶ限りでは，その差異に基づく群衆の空間的関心を抽出し，新たな空間的な知識を探索したり，地域特徴や地域間の関係性を分析可能としたりする可視化システムは存在していない．したがって，提案する群衆の空間的関心分析のための可視化システムは，上記で紹介した関連研究を補完しうるものと位置付けられる．

3. 群衆の空間的関心分析のための可視化システムの概要

群衆の多面的な空間的関心を俯瞰してとらえるために，ツイートにおける空間差異を可視化するシステムを構築する．本章では，ツイートのデータ構造（3.1 節）とデータビューの設計（3.2 節）について述べる．

3.1 データ構造

本研究において，ツイートは 5 つの基本属性の組 $t =$ <*tweet id*, *user id*, *text content*, *time stamp*, *location stamp*> として表現される．*tweet id* と *user id* はそれぞれツイートとユーザを一意に識別するための ID である．*text content* はツイートのコンテンツ内のテキスト文であり，*time stamp* はツイート投稿日時，*location stamp* はツイート投稿位置の緯度経度である．我々が収集し蓄積しているデータセットは，すべて *location stamp* を含むものである．

さらに，これら 5 つの基本属性に基づく空間属性として，<*location mention*, *location diff*> を新たに抽出する（**図 1**）．ここで，*location mention* はテキスト文中で参照されている位置（参照位置）であり，*text content* から抽出される位置表現の曖昧性を除去し，推定される（詳細は 4 章で述べる）．*location diff* はツイートの発信位置と参照位置との差異を表す属性であり，*location stamp* と *location mention* 間のユークリッド距離 d(*location stamp*, *location mention*) として算出される[*2]．また，公式 Twitter アプリからスマートフォンを用いて投稿されたツイートの場合，発信位置の精度は高いと考えられる．しかし，PC を用いて投稿した場合には発信位置の近似値となる．そのため，今回は 1 km 以下のものは誤差と見なし，それ以上の距離のみを差異として考慮する．

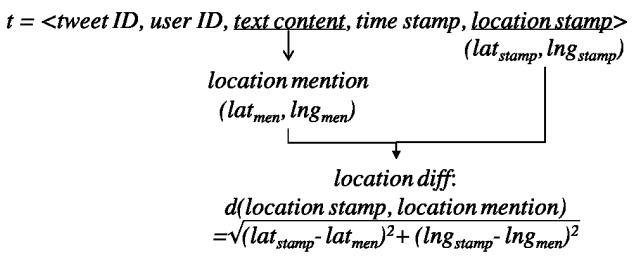


図 1　データモデル
Fig. 1　Data model.

これら 3 つの空間属性（*location stamp*，*location mention*，*location diff*）を用いることにより，以下のような分析が可能である．

- ある場所にいる人々は何を述べているのか．
- ある場所について何を述べているのか．
- 一定の距離離れた場所について何を述べているのか．

3.2 データビューの設計

群衆の多面的な空間的関心を俯瞰してとらえるために，ツイートの 3 つの空間属性と時間属性を組み合わせた 3 つのデータビューを設計・構築する．

データビュー 1（図 3）： 特定の場所（州）にいる人々がある場所（州）にどれくらい関心を向けているのか，何を述べているのかを分析するために設計．2 つの空間属性（*location stamp*，*location mention*）を用いて構築．

データビュー 2（図 4）： ある場所（州）にいる人々が一定距離離れた場所にどれくらい関心を向けているのか，何を述べているのかを分析するために設計．データビュー 1 に比べ，州よりも詳細な場所に対する関心を扱うことができる．2 つの空間属性（*location stamp*，*location diff*）を用いて構築．

データビュー 3（図 5）： 特定の日に一定距離離れた場所に人々はどれくらい関心を向けているのか，何を述べているのかを分析するために，時間属性を組み合わせて設計．空間属性（*location diff*）と時間属性（*time stamp*）を用いて構築．

[*2] 米国全域のような広域を対象とする場合，三平方の定理に基づくユークリッド距離では場所によって誤差が生じる可能性がある．しかし，今回の可視化方式では，州単位での分析・考察が主であり，距離は対数尺度で検討しており，距離の誤差は分析結果に大きな影響を与えないものと考えている．ただし，今後は場所による誤差を考慮した距離算出方式（ヒュベニの公式など）を用いた分析・考察を行う予定である．

各データビューは 2 次元平面グラフであり，ツイートを集約する最も小さい単位として設定されたセルからなる．各セルは，属性情報に基づき割り当てられたツイート集合の定量情報（確率質量）と定性情報（特徴語）を有する．これらの抽出方法を 3.2.1 項と 3.2.2 項にそれぞれ示す．さらに，時空間的なまとまりとして，セルを集約したセグメントについて説明する（3.2.3 項）．

3.2.1　確率的質量の算出

セルごとのツイートの量を求めるために，ツイート t がセル $C_{i,j}$ にマッピングされる確率 $P(C_{i,j}|t)$ を算出する．提案システムにおいて，ツイートの発信位置および位置表現より推定される参照位置（詳細は 4 章）は，つねに特定の緯度経度とする．そのため，確率 $P(C_{i,j}|t)$ は $\{0,1\}$ となる．すべての位置を 1 地点として扱うアプローチは最適とはいえないかもしれないが，今回はシステムの効率化を優先し単純化している．

なお，提案モデルでは確率を用いるため，地理的な拡張は容易に可能である．たとえば，参照位置を特定の緯度経度とするだけでなく，より上位階層の地名（州名や “USA” など）の場合には，複数のセルにまたがる可能性があるため，この確率に基づき，ツイートの重みを分散させるなどの処理も可能である．このような拡張については，今後の課題である．

3.2.2　特徴語の抽出

セルごとの群衆の話題を求めるために，セル $C_{i,j}$ のツイート集合において特徴的な語（特徴語）を抽出する．スコアは出現頻度に基づく TF-IDF による重み付けを適用して算出される．すなわち，特定のセルに頻出するが，他のセルには頻出しない単語に対して高いスコアを割り当てる．今回は出現頻度の代わりに，3.2.1 項で算出される確率を用いて修正した TF-IDF を適用する．

具体的には，以下の式により，データビューのすべてのセル $CELLS$ のうち，i 行目 j 列目のセル $C_{i,j}$ における単語 $w \in W$ のスコアを算出する．

$$Score(w, C_{i,j}, T) = \frac{\sum_{t \in T} P(C_{i,j}|t) : w \in W_t}{\sum_{t \in T} P(C_{i,j}|t)} \times \log \frac{|CELLS|}{|C \in CELLS : \exists t \in C : w \in W_t|}$$

なお，w は単語，$C_{i,j}$ は i 行目 j 列目のセル，T はすべてのツイート集合，W_t はツイート t に含まれる単語集合であり，$P(C_{i,j}|t)$ はツイート t がセル $C_{i,j}$ に割り当てられる確率（3.2.1 項において算出）である．

最後に，算出されたスコアに基づき単語を順位付けし，上位語を群衆の話題を表す特徴語として抽出する．現在のシステムでは，データビューのセルをクリックすると，上位 100 件の特徴語とそのスコア，そして割り当てられたツイート ID などのリストを確認できる．

3.2.3　セグメントの設定

あるセルの特徴を分析するためには，そのセルと意味的に関連するセル集合を比較できることが望ましい．そこで，空間的なまとまり（州ごと，距離ごとなど）あるいは時間的なまとまり（1 週間ごとなど）でセルを集約したセグメントを用いる．一定の列に含まれる全セルのまとまりを垂直セグメント，一定の行に含まれる全セルのまとまりを水平セグメントとする．たとえば，データビュー 3（図 5）の垂直セグメント “July 22” は，July 22 から July 28 の列に含まれる全セルの集約情報を有する．

各セグメントにおける特徴語は，3.2.2 項と同様に，確率を用いて修正した TF-IDF を適用して抽出される．このとき，任意のセグメントを 1 つの仮想的な文書と見なし，そのコンテンツを列（または行）に含まれるセルのすべての単語集合 W_{S_i} と見なす．以下の式により，データビューに含まれるすべてのセグメント SEG のうち，k 行目（あるいは列目）の水平（あるいは垂直）セグメント S_k における単語のスコアを算出する．

$$Score(w, S_k, T) = \frac{\sum_{t \in T} P(S_k|t) : \forall C \in S_k : w \in W_C}{\sum_{t \in T} P(S_k|t) : \forall C \in S_k} \times \log \frac{|SEG|}{|S \in SEG : \exists t \in S : w \in W_t|}$$

なお，w は単語，S_k は k 行目（あるいは列目）の水平（あるいは垂直）セグメント，T はすべてのツイート集合，W_C はセグメント S_k 内のセル C のツイートに含まれる単語集合，W_t はツイート t に含まれる単語集合であり，SEG はデータビューを構成するすべてのセグメント，S は SEG における任意のセグメント，$P(S_k|t)$ はツイート t がセグメント S_k に割り当てられる確率である．

最後に，セルの場合と同様に，算出されたスコアに基づき単語を順位付けし，上位語を群衆の話題を表す特徴語として抽出する．現在のシステムでは，データビューのセグメント（データビュー 2（図 4）とデータビュー 3（図 5）の上部と右部）にマウスオーバすると，ポップアップウィンドウに上位 30 件の特徴語が表示される．さらに，セグメントをクリックすると，上位 100 件の特徴語とそのスコア，そして割り当てられたツイート ID などのリストを確認できる．

4.　位置表現に基づく参照位置の推定

本章では，位置に関するキーワード（位置表現）を含むツイートを位置参照ツイートとし，参照位置を推定して，尤もらしい特定の緯度経度にマッピングする手法を示す（**図 2**）．4.1 節ではある位置参照ツイートのコンテンツのテキスト文からの位置表現抽出について，4.2 節では位置表現の参照位置推定による曖昧性除去について述べる．

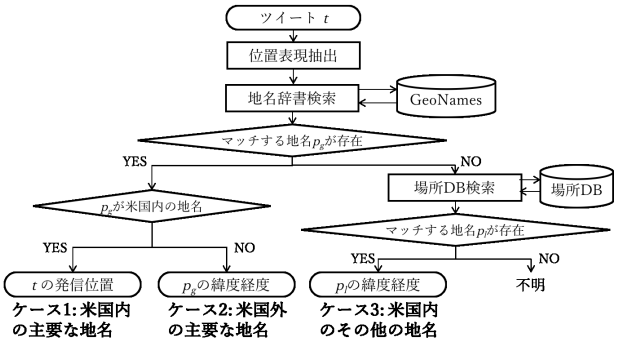


図 2　参照位置の推定による位置表現の曖昧性除去
Fig. 2　Estimation of referred location.

4.1　位置表現抽出

構文解析の固有表現抽出技術により空間アノテーションが付与された 1 語または連続する 2，3 語を位置表現と見なし，位置表現を含む位置参照ツイートを抽出する．そのために，文献 [38] と同様にノイズを除去する．そして，位置表現の単語のまとまりをより正確に特定するために，位置記述に用いられる標準パターンなどを用いて，経験則に基づく探索を行う．たとえば，米国の場合，2 つの大文字は州の名前を表すというパターン（Texas は TX）や市の名前の後にカンマが続くというパターン（Austin, TX）などを用いる．

4.2　曖昧性除去

次に，抽出した位置表現の曖昧性を除去する．本研究における曖昧性除去は，4.1 節で抽出された位置表現を尤もらしい特定の緯度経度（これを参照位置と呼ぶ）に対応付けることである．もし，あるツイートから複数の位置表現が抽出される場合[*3]，単純化および効率性を重視し，最後の位置表現の曖昧性を除去することとする．これにより，各ツイートにおける位置表現の数にかかわらず，曖昧性除去を 1 度実行するだけでよく，処理コストを大きく削減できる．そうでなければ，ツイートの位置表現のべき集合の要素と同じ数だけ繰り返し曖昧性除去を実行しなければならず，大規模データ分析への適用が難しい．また，効率の問題を無視して含まれるすべての位置表現の曖昧性を除去したとしても，その結果から適切な位置表現を選択することは容易でない．さらに，選択した位置表現以外は無視されるが，そのバイアスは大量のツイートを扱うことによって低減できると考えられる．

具体的な曖昧性除去手法について述べる．まず，位置表現で地名辞書である GeoNames[*4] を検索する．GeoNames には，地名，位置情報（緯度経度）や人口からなる要素 p_g が格納されている．しかしながら，GeoNames のような地名辞書には主要な地名のデータしか格納されておらず，その他の地名（施設など）を示す位置表現の参照位置を推定することが難しい．

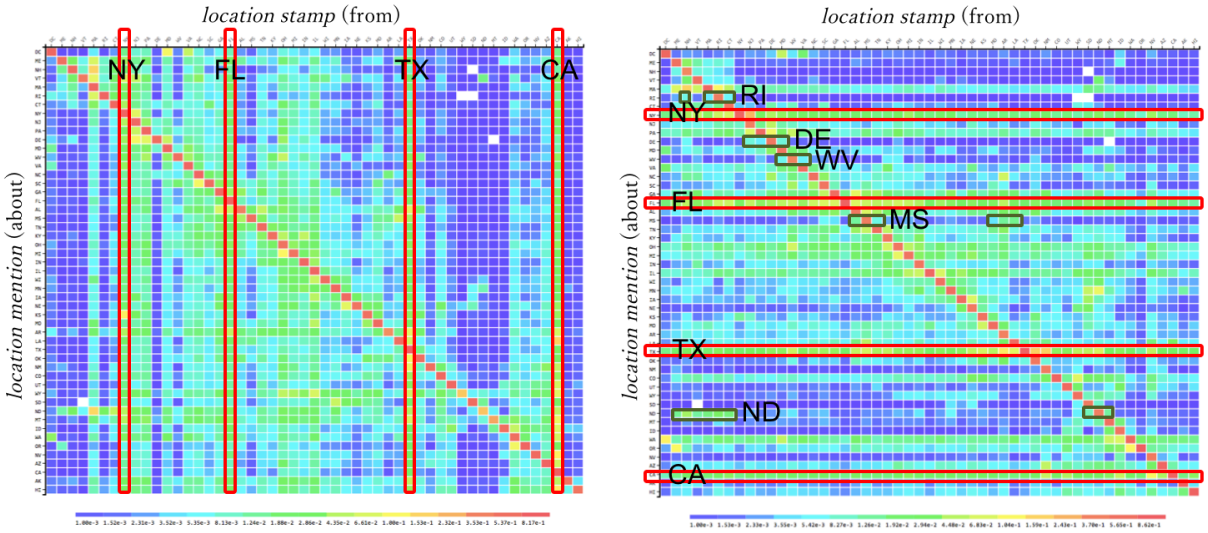
そのため，Foursquare などのチェックイン・アプリケーションを通して投稿されたツイートを用いて場所データベース（場所 DB）を構築し，利用する．一般的に，チェックイン・アプリケーションは，GPS 情報から場所名を推薦し，ユーザが選択した場所名に位置参照パターン（“I’m at” など）を自動的に付与してコンテンツを生成し，Twitter に投稿する．そのため，ツイートのコンテンツのテキスト文に位置参照パターンが含まれる場合，それに続く位置表現は発信位置と同じ場所を指すものと仮定できる．この仮定をもとに，位置表現および発信位置でツイートをまとめ，地名，位置情報（緯度経度）とツイート数からなる要素 p_l を格納し，場所データベースを構築する．

位置表現に基づく参照位置の推定手順（図 2）は以下のとおりである．

(1) 地名辞書検索で GeoNames に位置表現とマッチする地名 p_g があるとき，地名が米国内のものかを判定する．
　ケース 1　米国内の主要な地名：　p_g の緯度経度とツイートの発信位置が同じ地域（州または市）内である場合，位置表現はその地域（州または市）を参照している可能性が高いと見なし，参照位置にツイートの発信位置を割り当てる．そうでない場合は，参照位置に p_g の緯度経度を割り当てる．たとえば，「New York」という位置表現が GeoNames にマッチし，かつ，ツイートの発信位置も New York であれば，発信位置＝参照位置と見なす．一方，同じ条件で，ツイートの発信位置が Boston である場合，参照位置に GeoNames の New York の緯度経度を割り当てる．前者の場合，参照位置としてより詳細な緯度経度を対応付けることができる．なお，地名が複数ある場合には，人口が多く，かつ，発信位置からの距離が最も近い地名を選択する．
　ケース 2　米国外の主要な地名：　参照位置に p_g の地名の緯度経度を割り当てる．
(2) 地名辞書検索でマッチする地名がないとき，場所 DB 検索を行う．
　ケース 3　米国内のその他の地名：　位置表現にマッチする場所 p_l がある場合，参照位置に p_l の緯度経度を割り当てる．候補となる場所が複数ある場合，チェックイン・ユーザ数が多くかつ発信位置からの距離が最も近い場所を選択する．マッチする場所がない場合は，参照位置を不明とする．

なお，地名辞書検索，地理的近接性および場所の重要性レベルに基づく位置表現の曖昧性除去は単純ではあるが，その効果も実証されている．文献 [38] では，できる限り小さい場所や曖昧性のある場所を含む新聞のストーリのコー

[*3] ただし，ツイートのコンテンツのテキスト文は短文であるため，実データを用いて実施した予備調査では，複数の位置表現が含まれるツイートの割合は少ないことが分かっている．
[*4] http://www.geonames.org



(a) X から Y への関心度.　(b) Y に対する X からの人気度.

図 3　データビュー 1：ツイートの起点（州）と終点（州）のヒートマップ

Fig. 3　Data view 1: Heat map of origin-target of tweets from/about prefectures in USA.

パスを作成し，曖昧性除去手法を適用して適合率，再現率および F 値により評価し，77%の精度を達成したことを報告している．ただし，これはフォーマルな文章である新聞を用いたときの精度であり，データの質が異なるツイートで同程度の精度を達成することは難しいかもしれない．ツイートにおける曖昧性除去の詳細な評価は今後の課題としたい．

5.　可視化システムの実装

本章では，データセット（5.1 節）と実装した 3 種類のデータビューをそれぞれ示す（5.2 節から 5.4 節）．

5.1　データセット

米国を対象に約 3 カ月間（2013 年 9 月 25 日から 2014 年 1 月 17 日）にわたり収集した約 198M（百万）ツイートを用いた．ただし，図 5 や**図 6** の空白（10 月 11 日から 29 日，12 月 13 日から 27 日）は，ネットワークの切断などのため，データクローリングが中断した期間である．

前処理として，単純ベイズ分類器に基づく言語判定手法を適用し，英語以外で書かれたツイートを取り除いた．次に，約 158M の英語ツイートに対して Stanford CoreNLP tagger [39] を用いてテキスト処理および位置表現抽出を行った．具体的には，LOCATION のタグが付与されたものを位置表現として抽出した．その結果，30M ツイートが位置参照ツイートとして抽出された．これらに対して地名辞書の GeoNames と場所 DB を用いた参照位置の推定を行い，曖昧性を除去した．その結果，地名辞書のみの場合よりも約 8%多い，62%のツイートの位置表現の参照位置を推定することができた．このことから，場所 DB を用いることにより，地名辞書のみでは対応できない位置情報を補完可能であることが示された．たとえば，John F. Kennedy International Airport のように地理的には小さいが多くの人が集まる場所や人気のある場所などの情報が補完された．

最終的に，参照位置推定により米国内に割り当てられた約 4.3M の位置参照ツイートをデータセットとして抽出した．なお，本データセットにおいて，位置参照ツイートを発信したユーザの割合は全体の約 28%を占めていた．

これらのデータセットに対し，バックエンドには SCALA [*5] を，フロントエンドには D3.js [*6] を用いて実装した．なお，ツイートのようにノイズが多いテキスト向けに特別に設計された固有表現抽出器 [40] も発見したが，提案システムへの適用は今後の課題とする．

5.2　データビュー 1：発信位置と参照位置に関する関心度

州単位で集約した群衆の空間的関心を俯瞰するために，データビュー 1 の可視化方式を提案する．図 3 はツイートの発信位置（*location stamp*）（横軸）と参照位置（*location mention*）（縦軸）を対比したデータビュー[*7]である．

グラフの横軸は発信位置からのツイートの頻度，縦軸は参照位置に関するツイートの頻度である．なお，地理的な近接効果を示すために，発信位置は左から右に，参照位置は上から下に，同じ並び順で整列されている．

さらに，このデータビューはオプションで行（または列）ごとに正規化する機能を有する．これにより，発信位

[*5] http://www.scala-lang.org
[*6] https://d3js.org
[*7] http://delab.kyoto-su.ac.jp/tweet/US7/TweetHeatMap_cityMatrix.html

置（あるいは参照位置）ごとのツイートの割合を可視化する．図 3 (a) は参照位置（行）ごとに正規化したものである．任意の発信位置の州 c に着目したときに，c から参照位置の各州 l について発言している割合を示している．つまり，特定の州 c にいる群衆が各州 l に向けている相対的な関心度を確認できる．図 3 (b) は発信位置（列）ごとに正規化したものである．任意の参照位置 l に着目したときに，l が発信位置（列）c からそれぞれ参照されている割合を確認できる．すなわち，任意の州 l が各州 c にいる群衆から向けられている相対的な関心度が示されている．

5.3　データビュー 2：位置参照の分布

次に，州よりも細かい空間的な尺度で群衆の空間的関心を俯瞰するために，データビュー 2[*8]（図 4）の可視化方式を提案する．データビュー 1 では，多数あるツイートの発信位置やコンテンツ中での参照位置を州単位で集約したため，州 × 州のマトリクスで描画することができた．一方，ツイートの発信位置やコンテンツ中での参照位置のすべての組合せを 1 つのグラフに描画することは難しい．そこで，データビュー 2 では，発信位置（*location stamp*）に基づく州ごとのデータで距離ベースの差異（*location diff*）を求め，州単位での空間的関心を可視化する．

グラフの横軸は州であり，並び順は地理的な近接性に基づき決定している．縦軸は距離であるが，線形目盛を用いて距離ベースの差異に基づく空間的関心を可視化することは難しい．これは，すべてのとりうる距離を示すためにグラフの縦軸を拡張すると，データの確認や分析がしにくくなるためである．そこで，広域にわたるデータを視覚的に描画するために，グラフの縦軸の *location diff* にはオプションで対数目盛を用いることも可能とした．なお，人は離れた位置を参照するとき，より大きな空間粒度での表現を用いる傾向も報告されており [41]，この点でも対数目盛を用いることは妥当といえる．グラフ下部のヒストグラムは州ごとに発信されたツイート数であり，データの正規化のために用いられ，それぞれ対応する列に合わせて表示される．さらに，各グラフの右側の折れ線グラフ（青色）は，データセット全体における空間差異の集約分布を示す．

5.4　データビュー 3：時間的な位置参照

群衆の空間的関心を時間に着目して俯瞰するために，データビュー 3[*9]の可視化方式を提案する．データビュー 3 では，タイムスタンプに基づく 1 日ごとのデータで距離ベースの空間差異（*location diff*）を求め，1 日単位での空間的関心を可視化する．

[*8] http://delab.kyoto-su.ac.jp/tweet/US7/TweetHeatMap_location_city.html
[*9] http://delab.kyoto-su.ac.jp/tweet/US7/TweetHeatMap_location.html

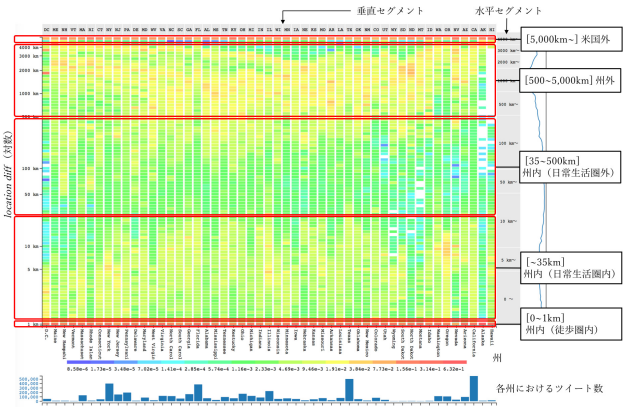


図 4　データビュー 2：州ごとの空間差異のヒートマップ
Fig. 4　Data view 2: Heat map of location differences per states.

グラフの横軸はツイートのタイムスタンプに基づく日付，縦軸はデータビュー 2 と同様に対数目盛による距離（*location diff*）である．グラフ内の縦の破線は週ごとの区切り，実線は月ごとの区切りである．グラフ下部の折れ線グラフ（青色）は各日のツイート総数であり，折れ線グラフ（オレンジ色）は参照位置（*location mention*）を含むツイート割合である．なお，データビュー 2 のグラフは州ごとにデータを集約した結果であるのに対し，データビュー 3 のは米国内のすべてのデータを集約した結果である．

さらに，データビュー 3 は，特定の州に焦点をあてた可視化結果を表示可能なフィルタリング機能を有する．これにより，特定の州に関わるイベントの発見や分析が可能である．図 6 に NY に焦点をあててフィルタリングした結果を示す．

6.　データビューを用いた分析例と考察

本章では，空間的な尺度を設定し（6.1 節），提案した 3 つのデータビューを用いた分析例を示す（6.2 節から 6.6 節）．

6.1　考察に用いた尺度

空間的な尺度の粗いレンジとして，0 km から 500 km（州内），500 km から 5,000 km（米国内），5,000 km 以上（米国外）を設定した．さらに，州内における細かいレンジを，0 km から 1 km（徒歩圏内），35 km まで（日常生活圏内），500 km まで（日常生活圏外）とした．データビュー 1（図 3）では，米国内の州単位での分析例を示す．データビュー 2（図 4）とデータビュー 3（図 5）では，州よりも細かいレンジでの米国内および米国外も含めた分析例を示す．また，データビュー 3 で組み込んだ時間属性に関しては，主に平日（月曜日から金曜日），休日（土曜日と日曜日），イベントや休暇という時間のレンジで結果を検討する．

6.2　徒歩圏内（0 km から 1 km）

データビュー 2 とデータビュー 3 では，このレンジにあたる最下部全体に濃い赤色のセルが広がっている．場所や

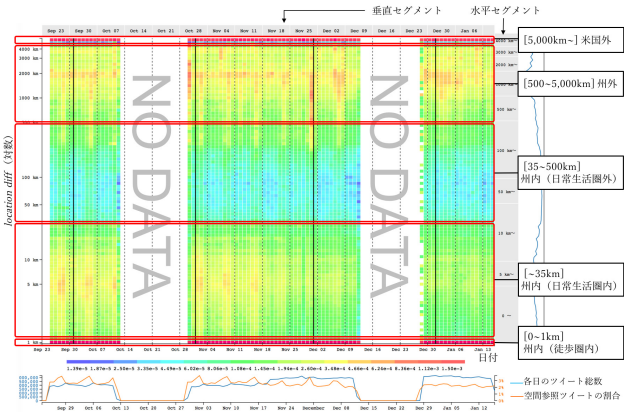


図 5　データビュー 3：日ごとの空間差異のヒートマップ

Fig. 5　Data view 3: Heat map of location differences over time with $1.5e^{-3}$ limit z-value.

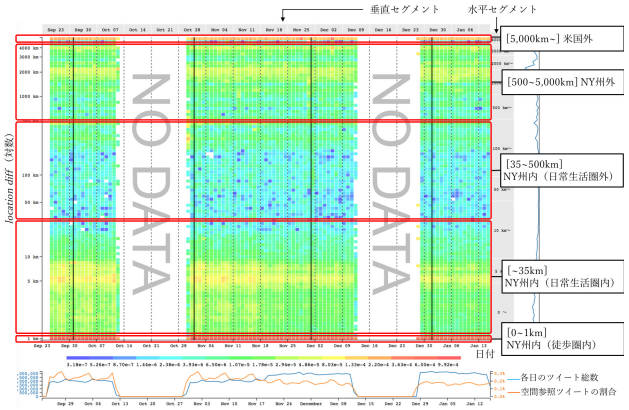


図 6　データビュー 3：NY に焦点をあてた日ごとの空間差異のヒートマップ

Fig. 6　Data view 3 (focused on NY): Heat map of location differences over time.

時間によらず，群衆の空間的関心が高いことが分かる．ツイートの発信位置に固有の不正確さをふまえ，差異が 1 km 以下のツイートをまとめたこと（3.1 節）や，チェックイン・アプリケーションを介して発信されたツイートが影響した結果とも考えられる．一方で，時間ベースの分析における「今」の概念 [6] と同様に，空間的関心を占有している「ここ」の概念を支持するものである．

6.3　日常生活圏内（35 km まで）

データビュー 2 のこのレンジでは，州により多少の差は存在するものの，全体的に黄緑色から橙色のセルが広がっている．このことから，群衆の空間的関心は中程度からやや高めであるといえる．

データビュー 3 においても，全体的に黄緑色から黄色のセルが広がっており，このレンジにおける群衆の空間的関心は中程度からやや高めであるといえる．さらに，11 月半ばまでは，平日から金曜日・土曜日にかけて，1 km から 10 km 前後にわたり黄色のセルが広がり，5km あたりには橙色や赤色のセルも確認される．しかしながら，11 月半ば以降はその特徴が小さくなり，このレンジにおける群衆の空間的関心が相対的に小さくなっている．

米国における通勤時間の統計[*10]や日常生活における行動範囲に関する報告[*11]によると，自宅から最寄りのショッピングセンタや学校までの距離は 5 km 以内，勤め先までの距離は 35 km 以内である人が多いという結果が示されている．これらのデータビューを分析することにより，この調査報告で示されているよりも多様な，米国人の日常生活圏に基づく関心の幅やその変化をとらえることができる．実際に，このレンジにおける水平セグメントの特徴語には，“restaurant,” “grill,” “home,” “hospital” など，日常生活に関するものも見られた．

6.4　州内かつ日常生活圏外（35 km から 500 km）

データビュー 2 とデータビュー 3 ともに，水色や青色のセルや緑色のセルが目立ち，このレンジにおける群衆の空間的関心は中程度からやや低めであるといえる．また，データビュー 3 で平日と休日に分けて確認したところ，休日のこのレンジには緑色のセルや黄色のセルが多く見られた．水平セグメントの特徴語を確認したところ，“road,” “tomorrow,” “weekend,” “drive” などの特徴語のスコアが高くなり，他にも “friend” や “trip” といった日常生活圏内のレンジでは抽出されなかった特徴語が見られた．

6.5　米国内の他の州（500 km から 5,000 km）

まず，データビュー 1 を用いて米国内の州を単位とした群衆の空間的関心を俯瞰する．他の州の人々から関心を向けられており（図 3 (b) で緑色や黄色のセルが占めている），かつ，その州にいる人々も他の州に関心を向けている州（図 3 (a) でも緑色や黄色のセルが占めている）として，New York（NY），Florida（FL），Texas（TX）や California（CA）などがあげられる．また，地理的に近接する州の人々がお互いに関心を向けている州（図 3 (a) と図 3 (b) で部分的に緑色や黄色のセルが占めている）として，Mississippi（MS），West Virginia（WV），North Dakota（ND），Rhode Island（RI）や Delaware（DE）などが確認された．さらに，他の州にはほとんど関心を向けていない州（図 3 (a) で青色のセルが占めている）であるが，他の州の人々から関心を向けられている州（図 3 (b) で図 3 (a) よりも相対的に高いスケールの色のセルが占めている）として，Hawaii（HI）や Alaska（AK）などが見られた．

次に，データビュー 2 とデータビュー 3 のこのレンジには，州内かつ日常生活圏外に比べて黄色のセルが多く見られ，このレンジに対する関心が相対的に高いという傾向が確認された．

さらに，データビュー 3 では，米国内で開催されるス

*10 http://www.statisticbrain.com/commute-statistics/
*11 http://www.ers.usda.gov/media/1807325/eib138.pdf

ポーツなどのイベント（アメリカンフットボールの試合など）に連動した群衆の空間的関心が観察された．このレンジにおいて，相対的に大きい空間的関心が見られた 10 月 30 日と 11 月 30 日について詳しく調査した．

まず，10 月 30 日は野球のメジャーリーグベースボール（MLB）における優勝決定戦（ワールドシリーズ）で，ボストン・レッドソックスが地元ボストンで 95 年ぶりに優勝を決めた日であった[*12]．米国中の野球ファンの注目を集める一大イベントであり，テレビ視聴者も多かったと思われる．そのため，空間的関心が高くなり，その幅も広がったものと推測される．実際に，10 月 30 日のこのレンジのセルには，“win,” “#worldseries,” “#bostonstrong,” “congrat” などの特徴語が多く見られた．

次に，11 月 30 日は米国の祝祭日の 1 つである感謝祭の日であった．この日は，家族や親戚が集うための行事と位置付けられており，実家に帰る人々の増加などにより，空間的関心が高くなり，その幅も広がったと予想される．定性的な観点からも，その週を通して感謝祭の話題が増え，垂直セグメントには “Thanksgiving,” “turkey,” “family” などの特徴語が含まれ，それらのスコアも高かった．

なお，このレンジにおいて，データビュー 2 の一番右列のハワイ州（HI）のほとんどが空白となっているが，これはハワイが他の州から離れた位置に存在するという米国の地理的なトポロジによる結果である．さらに，データビュー 3 における 2,000 km あたりの赤色のセルによる水平線は，“USA” という位置表現の参照位置を米国の中心地点に割り当てたことによる影響が大きい．このように，“USA” や州名など上位階層の地名を参照位置として割り当てて算出した距離の扱いや可視化方式に関しては，今後さらに検討する余地がある．

6.6　米国外（5,000 km から）

データビュー 2 とデータビュー 3 では，このレンジにあたる最上部全体に濃い赤色のセルが広がっており，場所や時間によらず，このレンジにおける群衆の空間的関心が高いことをうかがうことができる．ただし，今回は単純化のために，米国以外に向けられた群衆の関心を最上部のセルに集約しており，データビューの定量的な側面だけで結果を分析することは難しい．しかしながら，定性的な結果である水平セグメントの特徴語と合わせた分析により，主にどの国に対して関心が寄せられていたかを俯瞰できる．今回の例では，“Sydney,” “UK,” “China,” “Tokyo” などの国名が特徴語として抽出された．

7.　おわりに

本稿では，空間的な観点に着目したビッグデータ分析のための新たな可視化システムを構築した．これにより，ツイートの発信位置とコンテンツのテキスト文中の参照位置に基づく空間差異を多面的に分析するために，群衆の空間的関心を可視化した．約 3 カ月にわたり米国で発信された 2 億弱のツイートを用いて 3 つのデータビューを設計・構築し，分析結果の例を空間的な尺度ごとに考察した．これにより，空間差異に基づく群衆の空間的関心を明らかにすることで，コンテンツのテキスト文の位置に関する表現を用いたメタデータの補完も可能となり，また，イベント検出などに活用可能であることが示唆された．さらに，これまでの地理と人々の行動に関する仮説を調査，強化および確証することが可能であると期待される．

ただし，提案システムを用いた分析は探索的であるため，いくつかの制限があることが分かっている．今後の課題として，特定の位置間の差異に関連する語彙の範囲を分析する予定である．たとえば，旅行などの活動は長距離との関連が強く，買い物などの活動はより身近な距離圏に関連しているといった距離と行動に関する語彙との相関を明らかにする．すでに特定の距離と特定の位置に関連する特徴語抽出システムは構築しているため，明白な位置表現が不足しているテキスト文を対象に，自動的に位置に関する情報を推測する予定である．

謝辞　本研究の一部は，JSPS 科研費 16K16057，16H01722，15K00162 により実施した．ここに記して謝意を表す．

[*12] http://www.sponichi.co.jp/baseball/news/2013/10/31/kiji/K20131031006917910.html（2016 年 6 月 9 日閲覧）

参考文献

[1] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *Proc. International Conference on World Wide Web* (*WWW*), pp.851–860 (2010).

[2] Chen, Y., Amiri, H., Li, Z. and Chua, T.-S.: Emerging Topic Detection for Organizations from Microblogs, *Proc. International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR*), pp.43–52 (2013).

[3] Ardon, S., Bagchi, A., Mahanti, A., Ruhela, A., Seth, A., Tripathy, R.M. and Triukose, S.: Spatio-temporal and Events Based Analysis of Topic Popularity in Twitter, *Proc. ACM International Conference on Information and Knowledge Management* (*CIKM*), pp.219–228 (2013).

[4] Lee, R., Wakamiya, S. and Sumiya, K.: Discovery of unusual regional social activities using geo-tagged microblogs, *World Wide Web*, Vol.14, No.4, pp.321–349 (2011).

[5] Valkanas, G. and Gunopulos, D.: How the Live Web Feels About Events, *Proc. ACM International Conference on Information and Knowledge Management* (*CIKM*), pp.639–648 (2013).

[6] Jatowt, A., Antoine, E., Kawai, Y. and Akiyama, T.: Mapping Temporal Horizons: Analysis of Collective Future and Past Related Attention in Twitter, *Proc. International Conference on World Wide Web* (*WWW*), pp.484–494 (2015).

[7] Antoine, E., Jatowt, A., Wakamiya, S., Kawai, Y. and Akiyama, T.: Portraying Collective Spatial Attention in Twitter, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.39–48 (2015).

[8] Wakamiya, S., Jatowt, A., Kawai, Y. and Akiyama, T.: Analyzing Global and Pairwise Collective Spatial Attention for Geo-social Event Detection in Microblogs, *Proc. International Conference Companion on World Wide Web* (WWW Companion), pp.263–266 (2016).

[9] Vieweg, S., Hughes, A.L., Starbird, K. and Palen, L.: Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness, *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI*), pp.1079–1088 (2010).

[10] Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Alstyne, M.V.: Life in the network: The coming age of computational social science, *Science*, Vol.323, No.5915, pp.721–723 (2009).

[11] Goonetilleke, O., Sellis, T., Zhang, X. and Sathe, S.: Twitter Analytics: A Big Data Management Perspective, *SIGKDD Explor. Newsl.*, Vol.16, No.1, pp.11–20 (2014).

[12] Leetaru, K.H.: Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space, *First Monday*, Vol.16, No.9 (2011).

[13] Gan, Q., Attenberg, J., Markowetz, A. and Suel, T.: Analysis of Geographic Queries in a Search Engine Log, *Proc. International Workshop on Location and the Web* (*LOCWEB*), pp.49–56 (2008).

[14] Backstrom, L., Sun, E. and Marlow, C.: Find Me if You Can: Improving Geographical Prediction with Social and Spatial Proximity, *Proc. International Conference on World Wide Web* (*WWW*), pp.61–70 (2010).

[15] Cho, E., Myers, S.A. and Leskovec, J.: Friendship and Mobility: User Movement in Location-based Social Networks, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.1082–1090 (2011).

[16] Cheng, Z., Caverlee, J. and Lee, K.: You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users, *Proc. ACM International Conference on Information and Knowledge Management* (*CIKM*), pp.759–768 (2010).

[17] Kinsella, S., Murdock, V. and O'Hare, N.: "I'M Eating a Sandwich in Glasgow": Modeling Locations with Tweets, *Proc. International Workshop on Search and Mining User-generated Contents* (*SMUC*), pp.61–68 (2011).

[18] Chang, H.-W., Lee, D., Eltaher, M. and Lee, J.: @Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage, *Proc. International Conference on Advances in Social Networks Analysis and Mining* (*ASONAM*), pp.111–118 (2012).

[19] Eisenstein, J., O'Connor, B., Smith, N.A. and Xing, E.P.: A Latent Variable Model for Geographic Lexical Variation, *Proc. Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp.1277–1287 (2010).

[20] Hong, L., Ahmed, A., Gurumurthy, S., Smola, A.J. and Tsioutsiouliklis, K.: Discovering Geographical Topics in the Twitter Stream, *Proc. International Conference on World Wide Web* (*WWW*), pp.769–778 (2012).

[21] Yuan, Q., Cong, G., Ma, Z., Sun, A. and Thalmann, N.M.: Who, Where, When and What: Discover Spatio-temporal Topics for Twitter Users, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.605–613 (2013).

[22] Hecht, B., Hong, L., Suh, B. and Chi, E.H.: Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles, *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI*), pp.237–246 (2011).

[23] Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P. and Almeida, V.: Beware of What You Share: Inferring Home Location in Social Networks, *Proc. IEEE International Conference on Data Mining Workshops* (*ICDMW*), pp.571–578 (2012).

[24] Li, R., Wang, S. and Chang, K.C.-C.: Multiple Location Profiling for Users and Relationships from Social Network and Content, *Proc. VLDB Endow.*, Vol.5, No.11, pp.1603–1614 (2012).

[25] Li, R., Wang, S., Deng, H., Wang, R. and Chang, K.C.-C.: Towards Social User Profiling: Unified and Discriminative Influence Model for Inferring Home Locations, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.1023–1031 (2012).

[26] Sadilek, A., Kautz, H. and Bigham, J.P.: Finding Your Friends and Following Them to Where You Are, *Proc. ACM International Conference on Web Search and Data Mining* (*WSDM*), pp.723–732 (2012).

[27] Weng, J. and Lee, F.: Event Detection in Twitter, *Proc. International AAAI Conference on Weblogs and Social Media* (*ICWSM*), pp.401–408 (2011).

[28] Ritter, A., Mausam, Etzioni, O. and Clark, S.: Open Domain Event Extraction from Twitter, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.1104–1112 (2012).

[29] Li, C., Sun, A. and Datta, A.: Twevent: Segment-based Event Detection from Tweets, *Proc. ACM International Conference on Information and Knowledge Management* (*CIKM*), pp.155–164 (2012).

[30] Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S. and Miller, R.C.: Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration, *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI*), pp.227–236 (2011).

[31] Abel, F., Hauff, C., Houben, G.-J., Stronkman, R. and Tao, K.: Twitcident: Fighting Fire with Information from Social Web Streams, *Proc. International Conference on World Wide Web* (*WWW Companion*), pp.305–308 (2012).

[32] McMinn, A.J., Tsvetkov, D., Yordanov, T., Patterson, A., Szk, R., Rodriguez Perez, J.A. and Jose, J.M.: An Interactive Interface for Visualizing Events on Twitter, *Proc. International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR*), pp.1271–1272 (2014).

[33] Morstatter, F., Kumar, S., Liu, H. and Maciejewski, R.: Understanding Twitter Data with TweetXplorer, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.1482–1485 (2013).

[34] Musleh, M.: Spatio-temporal Visual Analysis for Event-specific Tweets, *Proc. ACM SIGMOD International Conference on Management of Data* (*SIGMOD*),

pp.1611–1612 (2014).
[35] Zhang, W. and Gelernter, J.: Geocoding location expressions in Twitter messages: A preference learning method, *Journal of Spatial Information Science*, No.9, pp.37–70 (2014).
[36] Schulz, A., Hadjakos, A., Paulheim, H., Nachtwey, J. and Mühlhäuser, M.: A Multi-Indicator Approach for Geolocalization of Tweets, *Proc. International AAAI Conference on Weblogs and Social Media* (*ICWSM*), pp.573–582 (2013).
[37] Leetaru, K.H.: Fulltext Geocoding Versus Spatial Metadata for Large Text Archives: Towards a Geographically Enriched Wikipedia, *D-Lib Magazine*, Vol.18, No.9/10 (2012).
[38] Pouliquen, B., Kimler, M., Steinberger, R., Ignat, C., Oellinger, T., Fluart, F., Zaghouani, W., Widiger, A., Forslund, A.-C. and Best, C.: Geocoding multilingual texts: Recognition, disambiguation and visualisation, *Proc. International Conference on Language Resources and Evaluation* (*LREC*), pp.53–58 (2006).
[39] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. and McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit, *Proc. Annual Meeting of the Association for Computational Linguistics* (*ACL*), pp.55–60 (2014).
[40] Ritter, A., Clark, S., Mausam and Etzioni, O.: Named Entity Recognition in Tweets: An Experimental Study, *Proc. Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp.1524–1534 (2011).
[41] Camossi, E., Bertolotto, M., Bertino, E. and Guerrini, G.: A Multigranular Spatiotemporal Data Model, *Proc. ACM International Symposium on Advances in Geographic Information Systems* (*GIS*), pp.94–101 (2003).

若宮 翔子　（正会員）

奈良先端科学技術大学院大学博士研究員．2013 年兵庫県立大学大学院環境人間学研究科博士後期課程修了．2014 年京都産業大学コンピュータ理工学部研究員．2015 年より現職．博士（環境人間学）．主にソーシャルメディアデータ分析の研究に従事．日本データベース学会会員．

ヤトフト アダム　（正会員）

京都大学大学院情報学研究科社会情報学専攻特定准教授．2005 年東京大学大学院情報理工学系研究科電子情報学博士後期課程修了．博士（情報学）．主にウェブ検索，ウェブアーカイブマイニングの研究に従事．ACM 会員．

河合 由起子　（正会員）

京都産業大学コンピュータ理工学部准教授．2001 年奈良先端科学技術大学院大学情報科学研究科情報システム学博士後期課程修了．同年独立行政法人情報通信研究機構，2006 年京都産業大学理学部コンピュータ科学科講師を経て 2008 年より現職．博士（工学）．情報推薦，Web マイニング，信憑性分析の研究に従事．電子情報通信学会，日本データベース学会各会員．

秋山 豊和　（正会員）

京都産業大学コンピュータ理工学部准教授．博士（工学）．主に分散システム・インターネットミドルウェア（セキュリティ，大規模分散処理），キャンパス情報システム（認証・認可基盤）の研究に従事．電子情報通信学会，IEEE CS 等各会員．

荒牧 英治　（正会員）

奈良先端科学技術大学院大学特任准教授．博士（情報理工学）．2005 年東京大学大学院情報理工学系研究科博士課程修了．2005 年東京大学医学部附属病院特任助教，2008 年東京大学知の構造化センター特任講師，2011 年京都大学デザイン学ユニット特定准教授を経て 2015 年より現職．医療情報学，自然言語処理の研究に従事．言語処理学会，日本認知科学会，医療情報学会等各会員．

（担当編集委員　北本 朝展）