情報処理学会論文誌　データベース　Vol.7　No.2　51–60 (June 2014)

時間変化する特徴語によるマイクロブログ地名曖昧性解消

落合 桂一[1,a)]　鳥居 大祐[1]

受付日 2013年12月22日, 採録日 2014年4月3日

概要：本研究では，Twitter などの文章が短いマイクロブログを対象として，位置に関する特徴語を利用した地名の曖昧性解消手法を提案する．従来，同名地名の曖昧性解消には地理的に近い地名との共起が用いられていた．しかし，マイクロブログは文章が短いため，地名以外の単語も曖昧性解消に利用すべきである．そこで，マイクロブログの投稿にはその場所特有のトピックが存在することが多いと考え，地名ごとにその場所特有の単語（特徴語）を利用することで地名の曖昧性解消を行う．特徴語は季節変動などに依存しない定常的なものと，時間の経過によって変化する非定常的なものが存在する．そのため，定常的な特徴語（静的特徴語）を観光案内や Wikipedia の説明文のような静的な文書から抽出し，地名と静的特徴語の共起により曖昧性解消を行う．ここでは季節変動や時期に依存しない特徴語を利用する．一方，非定常的な特徴語（動的特徴語）はマイクロブログの特徴であるリアルタイム性を反映し，場所のトピックが時間とともに変化すると考え，従来手法により曖昧性解消された投稿から地名ごとの特徴語を動的に生成し曖昧性解消に利用する．提案手法の有効性を確認するため，再現率および適合率を評価した．地名に対して人手により正解ラベルを付与し正しく抽出できる数を調査した．その結果，従来手法の地理的に近い地名との共起をベースラインとして，提案手法の有効性を確認した．

キーワード：地名曖昧性解消，マイクロブログ，位置情報，特徴語

Toponym Disambiguation Method for Microblogs Using Time-varying Location-related Words

Keiichi Ochiai[1,a)]　Daisuke Torii[1]

Received: December 22, 2013, Accepted: April 3, 2014

Abstract: In this study, we propose a disambiguation method for toponyms using words related to the location. Conventionally, toponym ambiguition has been resolved by using nearby toponyms based on the hypothesis that geographically-closed toponyms are appeared frequently in the same contenxt. In the case of microblogs, however, words other than toponyms are preferable to be used because short texts of microblogs have less information. To this end, we consider that microblogs have a topic related to the location and propose a method which uses words related to the location ("location-related words") as disambiguators for each toponym. The location-related words are categorized into two groups. One is static words independent of seasonal variations and so on. The other is dynamic one which depends on seasonal variations etc. The dynamic location-related words reflect immediacy of microblog (i.e., the dynamic location-related words vary with time). We evaluated our proposed method by recall and precision using manually labeled data. The result showed that the recall of our proposed method is higher than that of the conventional method.

Keywords: toponym disambiguation, microblog, location, related word

1. はじめに

リアルタイムに情報が共有される Twitter などのマイクロブログや Facebook などの SNS（Social Networking Service）のデータを解析することで，実世界で起こる様々

[1] 株式会社 NTT ドコモ
NTT DOCOMO, INC., Yokosuka, Kanagawa 239–8536, Japan
[a)] ochiaike@nttdocomo.com

なイベントを検出したり，注目されているスポットを抽出したりするなど，今まさに起きている世の中の動向を知ることができる．これらの抽出された情報はユーザにとっても有益なため位置情報サービスでの利用が期待される．このような位置情報サービスを提供するにあたり重要なのは，地名やスポットに関する情報を “リアルタイム” かつ “正確” に収集できることである．

マイクロブログの 1 つである Twitter では，位置に関連するツイートを集める方法は大きく 2 つある．1 つはジオタグ（緯度経度）を付加して投稿されたツイートを利用する方法，もう 1 つはツイート本文をテキスト解析し，地名を抽出して位置と関連付ける方法である．Cheng ら [1] によると全ツイート中 0.42%のツイートのみジオタグが付加されている．日本語の場合，ジオタグ付きツイートの割合はさらに少なく，橋本らの調査 [2] によると，日本語ツイートのうちジオタグが付与されている割合は約 0.18%であった．また，Kitamoto ら [3] によると，ツイートのテキストに地名が含まれる割合は全体の約 12%であり，ジオタグ付きツイートの割合より多い．そのため情報抽出という観点では，ツイートを多く集めるためツイートのテキスト解析を行い，言及されている場所を特定する方法が有効である．テキスト解析で位置と関連付ける場合は地名の曖昧性解消が課題となる．

地名の曖昧性には 2 つの種類がある [4]．1 つは Geo/Non-geo 曖昧性と呼ばれ，地名と同じ表記で地名以外の意味を持つものである．たとえば「松島」という表記は，地名としても人名としても使われる．もう 1 つは Geo/Geo 曖昧性と呼ばれ，表記が同じ地名が複数の地理的な場所に存在するものである．たとえば「日本橋」という表記の地名は東京と大阪に存在する．一般的に文書中の地名の曖昧性を解消するには，(1) 地名の抽出，(2) 場所の特定の 2 段階の処理を行う [5]．第 1 段階処理の地名の抽出では Geo/Non-geo 曖昧性の解消を行うため，CRF（Conditional Random Field）を用いた固有表現抽出が用いられることが多い [4], [6], [7], [8], [9]．第 2 段階処理の場所の特定では Geo/Geo 曖昧性の解消を行う．そのため「1 つのコンテクストで現れる地名は地理的に近い場所を示すことが多い」という仮定のもと，地理的に近い地名（近隣地名）や，地名辞書の階層が隣の地名（たとえば市区町村名であれば 1 つ上の階層の都道府県名など）や 1 つ上の階層が共通の同一階層の地名（たとえば市区町村名であれば同一都道府県内の市区町村名など）との共起により曖昧性解消する方法が多い [3], [4], [6], [7], [9], [10]．ツイートに対して CRF により Geo/Non-geo 曖昧性を解消する手法は，杉谷ら [9] の研究によると Foursquare のスポット名を対象として適合率 0.89，再現率 0.78 という結果であり，また，Liu ら [8] の研究では一般の地名に対して適合率 0.803，再現率 0.775 という結果が報告されている．そこで，Geo/Non-geo 曖昧性を解消した場合でも，その地名が Geo/Geo 曖昧性を持つ場合もあるため，本研究では Geo/Geo 曖昧性の解消を行う手法に着目する．

従来研究 [4], [6], [7] では Web やブログなど，マイクロブログより長い文章を対象としていた．そのため，文書内の単語数がマイクロブログに比べ多く，地名の共起のみでも曖昧性解消が行えていた．しかしながら，マイクロブログは文章が短いため地名以外の単語も曖昧性解消に利用すべきである．

そこで，本研究ではマイクロブログにおいて地名の曖昧性を解消するため，マイクロブログの投稿にはその場所特有のトピックが存在することが多いと考え，地名の共起以外に，地名ごとにその場所特有の特徴語を利用することで地名の曖昧性解消を行う．特徴語は季節変動などに依存しない定常的なものと，時間の経過によって変化する非定常的なものが存在する．そのため，定常的な特徴語（以下，静的特徴語と呼ぶ）を観光案内や Wikipedia の説明文のようなマイクロブログに比べて静的な文書から抽出し，地名と静的特徴語の共起により曖昧性解消を行う．一方，非定常的な特徴語（以下，動的特徴語と呼ぶ）はマイクロブログの特徴であるリアルタイム性を反映し，場所のトピックが時間とともに変化すると考え，従来手法や静的特徴語により曖昧性解消された投稿から地名ごとの特徴語を動的に生成し曖昧性解消に利用する．マイクロブログでは市区町村名より粒度が細かい地名は駅名や観光スポット名以外は使われにくいと考え，本研究では一般的な地名として市区町村名，駅名，観光スポット名を対象とする．また，Geo/Non-geo 曖昧性の解消は，前述のとおり Liu ら [8] および杉谷ら [9] の研究など CRF を用いた固有表現抽出などを行うことで曖昧性を解消できるため対象外とする．

本研究の貢献は以下のとおりである．

- マイクロブログの文章の短さを考慮し，地名以外の単語を利用して曖昧性解消する手法を提案した．曖昧性解消するための単語として静的特徴語と動的特徴語の 2 種類を提案した．
- Twitter を対象に提案手法の定量的な評価を行った．静的特徴語および動的特徴語を使うことで，従来手法と比べ適合率を低下をさせることなく再現率を向上させることができることを定量的に確認した．
- 定性的評価として，静的特徴語および動的特徴語を使って曖昧性解消された事例から，時期によって変化しない特徴語と変化する特徴語のそれぞれが曖昧性解消に有効であることを確認した．

本稿の構成は以下のとおりである．2 章で関連研究について述べる．3 章で提案手法を説明し，4 章では提案手法の有効性を確認するために行った評価実験について述べる．最後に，5 章で本研究のまとめと今後の課題を述べる．

2. 関連研究

2.1 地名曖昧性解消

1 章でも述べたとおり，地名の曖昧性は Geo/Non-geo 曖昧性と Geo/Geo 曖昧性の 2 種類存在する．そして，文書中の地名の曖昧性を解消するには地名の抽出，場所の特定の 2 つの処理を行う [5]．地名の抽出では Geo/Non-geo 曖昧性の解消を行い，場所の特定では Geo/Geo 曖昧性を解消する．Geo/Non-geo 曖昧性は，地名と同じ表記で地名以外の意味を持つ曖昧性で，一般的に CRF を用いて地名とそれ以外の意味の曖昧性を解消する．Geo/Geo 曖昧性は，表記が同じ地名が複数の地理的な場所に存在する曖昧性で，地名の共起を利用して曖昧性を解消する．Geo/Geo 曖昧性の解消方法について，従来研究 [4], [6], [7], [10] では主に以下の 2 つのヒューリスティクスを利用していた．

地名に対する事前知識　人口が多い地名や有名な地名が言及されやすい．

地名の共起　地名が地域名–都道府県名–市区町村名のような階層で保持されている地名辞書において，隣の階層や 1 つ上の階層が共通の同一階層の地名が共起しているかを探索する．あるいは，文書中に複数の地名が出現している場合は，地名どうしの地理的な距離が最小になる場所について言及していると考える．

詳細については文献 [6] を参照されたい．これらの方法では地名の曖昧性解消に同一文書中の他の地名を利用しているが，本研究では地名以外の単語を利用して曖昧性を解消するためこれらの手法とは異なる．

2.2 単語の地域限定性

単語の地理的な局所性を利用した場所の特定に関する従来研究として，手塚ら [11] は郷土料理やお土産品などのオブジェクトや Web ページが持つ地域性を推定する手法を提案している．オブジェクト名で検索して収集した Web ページに対してジオコーディングを行い場所を特定し，オブジェクトと地理的な位置の関連度を混合ガウス分布でモデル化した．奥ら [12] は，グルメ情報サイトなどの位置情報付きの文書から地域限定の語句を抽出する手法を提案している．Web 上での単語の共起頻度に基づいた単語の関連度の指標である WebPMI を用いて地名とその地名に関連する語句を抽出した．たとえば，「松阪」に対する「松阪牛」などである．長岡ら [7] の実世界の位置情報類推に関する研究では，ブログを対象として地名と関連の強い単語を地名との共起度および単語の一般性をもとに生成し，主題となる場所を判断するために関連語を利用している．馬場ら [13] は，地名や施設名などの特定の場所と明確に関連がある単語ではなく，「相撲」や「大仏」といった場所と明確な関連がない単語を検索クエリとして，それに潜在的に関連する場所を抽出する手法を提案している．Flickr のデータを対象に，Flickr の写真に付与された緯度経度，タグの共起を利用して各単語と関連する場所の確率分布を計算した．曖昧性解消は明示的には行わず，写真に付与された緯度経度で場所を特定している．

これらの研究では地名の関連語を利用した曖昧性解消は行っていない．

2.3 マイクロブログを利用した位置に関する研究

Sakaki ら [14] は Twitter への投稿を利用して，地震や台風などのイベントを検出する手法を提案した．地震や台風が発生したことについてのツイートを SVM を用いて判定し，地震や台風についてのツイートに対して時空間の確率モデルを作成し発生場所を特定する．Cheng ら [1] のユーザ位置推定に関する研究では，ツイートに含まれる単語とそのツイートを行ったユーザの位置情報を用いて単語の地理的な分布を作成し，ある地域に特有の単語を抽出する．そして，抽出した地域に特有の単語を利用してユーザの位置を推定する．山口ら [15] は Twitter に投稿されたツイートからのイベント検出と，検出したイベントを使ったユーザ位置推定を行う手法を提案している．地震などの地域的な局所性を持つイベントを利用して Twitter ユーザの位置を推定している．Dalvi ら [16] は，レストランなどのオブジェクトとツイートのマッチングを行う手法を提案している．ユーザとオブジェクトの距離と，オブジェクトに対するツイートをモデル化している．酒巻ら [17] は，ジオタグ付きツイートを対象に，ツイートを位置情報，時刻情報，投稿内容によってクラスタリングすることでユーザの行動を分析する手法を提案した．クラスタリングしたツイートからユーザがよく活動している地点のクラスタを抽出し，投稿内容をナイーブベイズによって「家」「職場」などに分類している．若宮ら [18] はジオタグ付きツイートから人々の移動を抽出し，地域間の近接性を測定する手法を提案している．

これらの研究ではジオタグ付きツイートのみが使われておりツイートに含まれる地名の曖昧性解消は行っていない．

伊川ら [19] はツイートのテキストを解析してツイートを発信した場所を特定する手法を提案している．位置を特定するために，Foursquare などの位置情報サービスを通じて投稿されたツイートを利用し，その前後のツイートと，位置情報サービスからの投稿の類似度を計算し場所を特定している．位置情報サービスの投稿を利用しているため曖昧性解消は行っていない．渡辺ら [20], [21] は，Foursquare から取得したスポット名を利用して，ツイート本文のテキスト解析を行って言及されているスポットを特定する手法を提案している．渡辺らの手法では，「マクドナルド」のように地理的分布が大きいスポット名は除き，「両国国技館」のように場所を一意に特定できる地理的分布が小さいスポットのみを対象としているため，地名の曖昧性は存

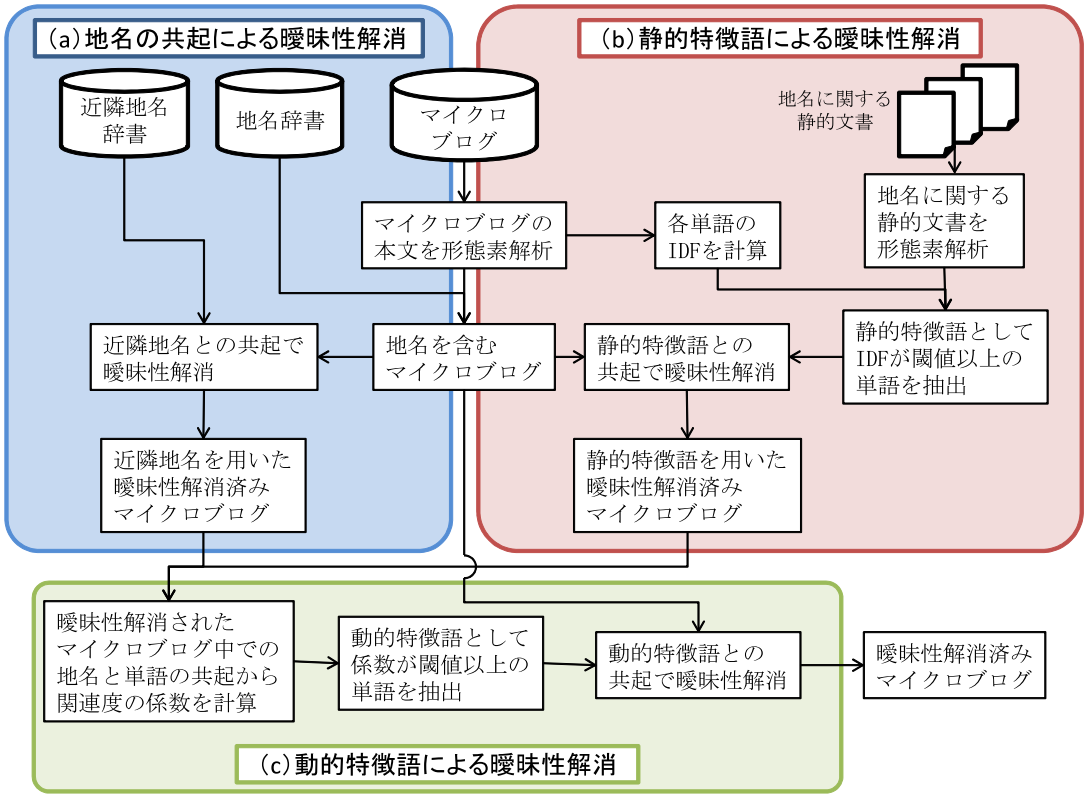

図 1　提案手法の流れ

Fig. 1　Flow of proposed method.

在していない．杉谷ら [9] はスポット名を含むツイートを抽出し，SVM を利用してその場所から投稿したツイートか判定する手法を提案した．Foursquare から取得したスポット名を利用し，CRF を用いてスポット名を抽出した後，近隣地名の共起を利用して曖昧性を解消しており，特徴語は利用していない．Liu ら [8] は K-Nearest Neighbors（KNN）と CRF を組み合わせ Twitter から半教師あり学習で固有表現抽出する手法を提案している．この手法により Geo/Non-geo 曖昧性の解消を適合率 0.803，再現率 0.775 で実現している．Geo/Non-geo 曖昧性の解消を目的としており，Geo/Geo 曖昧性の解消は行われていない．

3.　提案手法

本章では，初めに，地名以外の単語を曖昧性解消に使うことの有効性を確認するために行った予備実験について説明する．次に，提案手法の詳細を説明する．提案手法の流れを**図 1** に示す．図 1 に示すように，提案手法では次の 3 通りの方法で地名の曖昧性解消を行う．

(1) 地名の共起による曖昧性解消（従来手法．図 1 (a)）
(2) 静的特徴語による曖昧性解消（図 1 (b)）
(3) 動的特徴語による曖昧性解消（図 1 (c)）

なお，本研究では地名に対してマルチラベリングを許しており，図 1 (a)，(b)，(c) のそれぞれで異なる位置との関連づけが行われる可能性がある．4 章で述べる精度評価では，このマルチラベリングを含めて評価を行っている．

3.1　予備実験

本節では提案手法である，曖昧性解消に地名以外の単語を使う必要性を確認するための予備実験について説明する．従来研究では Web やブログなど，マイクロブログに比べて長い文章を対象としていた．そこで，文章の長さの違いによって地名の共起割合に違いがあるかを確認するため，Web，ブログ，マイクロブログにおける地名の共起割合を調査した．地名の中でも頻繁に利用されると考えられる日本の駅利用者数の上位 5 件[*1]の駅名（2012 年の上位 5 件は新宿，池袋，渋谷，東京，横浜）を含む Web ページ数，ブログ数，ツイート数を調査した．駅名が 1 つの検索結果件数を 1 として正規化したグラフを**図 2** に示す．Web ページ数は Google[*2]による検索結果件数，ブログ数は Google による検索で URL に “blog” を含む件数，ツイート数は Yahoo!リアルタイム検索[*3]で期間を 2013 年 10 月 12 日から 2013 年 11 月 9 日として検索した件数を利用した．図 2 から，地名の共起（グラフ中の横軸が 2 以上）は Web やブログでは 2 割程度であるが，Twitter では 1 割に満たないことが分かる．そのため，地名以外の手がかりを利用する必要があると考えられる．

*1　http://www.jreast.co.jp/passenger/
*2　https://www.google.co.jp/
*3　http://search.yahoo.co.jp/realtime

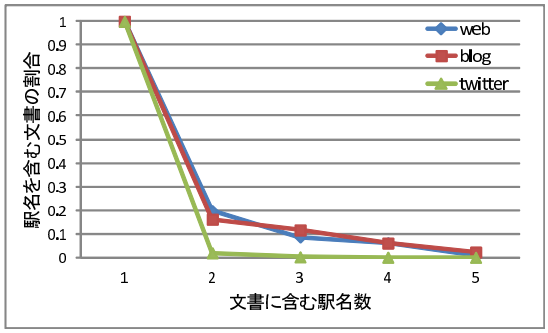

図 2　駅名を含む文書数の割合
Fig. 2　Ratio of document including station names.

3.2　地名の共起による曖昧性解消

本研究では，まず，従来研究と同様に曖昧性解消の対象となる地名と地理的距離が近い地名や，対象の地名を包含する範囲の地名との共起を利用した曖昧性解消を行う．処理の流れを図 1 (a) に示す．曖昧性解消に利用する地名（図 1 (a) の近隣地名辞書）は，全国の都道府県名，市区町村名，駅名やスポット名称などが利用できる．具体的な近隣地名辞書の構築方法は以下の 2 通りが考えられる．

(1) 地名の階層構造を利用する方法
(2) 地名の緯度経度を利用する方法

方法 (1) については，地名が「地域名」–「都道府県名」–「市区町村名」のような階層構造のデータの場合，市区町村名の曖昧性を解消するための近隣地名として同一都道府県の市区町村名と都道府県名を近隣地名とすることができる．たとえば，北海道と京都府に存在する「円山公園」という地名の場合，北海道内の市区町村名と都道府県名が北海道の円山公園に対する近隣地名となる．方法 (2) では，各地名に対して緯度経度が付与されている場合，曖昧性がある地名の緯度経度から一定の距離内の地名を近隣地名とする．本研究の実験では計算量が少ないという理由で方法 (1) を利用した．

3.3　静的特徴語の生成と曖昧性解消

静的特徴語の生成および静的特徴語を利用した曖昧性解消の流れを図 1 (b) に示す．静的特徴語は，対象の地名についての Wikipedia の紹介文や観光案内の Web ページ，観光情報データベースのような，マイクロブログに比べ静的な文書（以下，静的文書と呼ぶ）から生成する．まず，静的文書を形態素解析し，静的文書中の単語を抽出する．次に，マイクロブログの投稿数をもとに，特徴語抽出によく用いられる IDF [22] を各単語ごとに計算する．単語 w_i の IDF は次式で表される．

$$IDF_i = \log \frac{N}{n_i} \tag{1}$$

ここで，N は文書の総数である．本研究では 1 つのマイクロブログの投稿を 1 つの文書と考える．そのため N は特定の期間のマイクロブログ投稿数となる．n_i はマイクロブログで単語 w_i を含む投稿数である．本研究では IDF の値が閾値以上の単語を，対応する地名に対する静的特徴語として利用する．ここでは，1 つの地名に対して複数の特徴語を抽出する．たとえば，岡山県にある「後楽園」という地名に対して，「日本三名園」「偕楽園」「兼六園」などの単語が静的特徴語として抽出される．後楽園（岡山県）の紹介文と静的特徴語の例をそれぞれ**表 1** および**表 2** に示す．

曖昧性解消は，地名と静的特徴語が共起した場合に，静的文書に対応する地名であると判断する．

表 1　静的文書の例（後楽園（岡山県）の紹介文）
Table 1　Example of static document.

紹介文
後楽園は、江戸時代のおもかげを伝える庭園として歴史が古く昔から多くの方に愛されてきた。金沢の兼六園、水戸の偕楽園とあわせて「日本三公園」と称される日本三名園として有名で岡山のおすすめ観光スポット。

表 2　後楽園（岡山県）の静的特徴語の例
Table 2　Example of static location-related words.

静的特徴語	IDF
日本三名園	6.72904
偕楽園	5.81217
観光スポット	5.27297
兼六園	5.28339

3.4　動的特徴語の生成と曖昧性解消

動的特徴語の生成および動的特徴語を利用した曖昧性解消の流れを図 1 (c) に示す．3.2 節および 3.3 節で説明した曖昧性解消を行って抽出されたマイクロブログから，各地名と関連する特徴語を動的特徴語を抽出する．動的特徴語を生成するため，地名と単語の共起の強さを測る．そのための指標として，共起頻度以外にも様々な指標がある．松尾ら [23] の学会論文における共著関係の研究を参考に，以下の指標を利用した．本研究ではマイクロブログが投稿される期間を考慮している．

共起頻度：$|X_t \cap Y_t|$

ダイス係数：$\dfrac{2|X_t \cap Y_t|}{|X_t| + |Y_t|}$

Jaccard 係数：$\dfrac{|X_t \cap Y_t|}{|X_t \cup Y_t|}$

Simpson 係数：$\dfrac{|X_t \cap Y_t|}{\min(|X_t|, |Y_t|)}$

ここで，$|X_t|$ は地名 X を含む期間 t におけるマイクロブログの投稿数，$|Y_t|$ は地名との関連度合いを計算する対象の単語 Y を含む期間 t におけるマイクロブログの投稿数，$|X_t \cap Y_t|$ は X_t と Y_t の積集合となるマイクロブログの投稿数，$|X_t \cup Y_t|$ は X_t と Y_t の和集合となるマイクロブログの投稿数である．それぞれの指標において閾値をそれぞれ設定し，閾値以上の場合に動的特徴語として利用する．

動的特徴語を利用した曖昧性解消は，曖昧性解消したい地名と動的特徴語が共起した場合に，動的特徴語に対応する地名であると判断する．

4. 評価実験

4.1 実験環境

提案手法による地名曖昧性解消の性能を評価するため，従来手法である地名の共起による曖昧性解消および地名に対する事前知識として地名の人気度による曖昧性解消をベースラインとして，静的特徴語を利用した曖昧性解消，動的特徴語を利用した曖昧性解消に対してそれぞれ実験を行った．マイクロブログデータは 2013 年 5 月 2 日の 1 日分の Twitter データを対象とした．地名辞書は，市区町村名，駅名，観光スポット名を利用した．市区町村名は統計局ホームページ[*4]に市区町村別人口が掲載されている市区町村名を利用した．市区町村名の総件数は 1,918 であり，Geo/Geo 曖昧性がある名称は 38 件である．駅名は「駅データ.jp」[*5]のデータを利用した．駅名の総数は 9,172 件であり，そのうち Geo/Geo 曖昧性が存在する駅名は 908 件である．駅名は駅までを含む表記と，駅を含まない表記（たとえば，「日本橋駅」と「日本橋」）を使用した．駅までを含む表記はピンポイントの地名を表し，駅を含まない表記は駅周辺のエリアを表す地名と考え，この 2 通りを使用した．観光スポット名については NTT ドコモが提供している「ご当地ガイド」[*6]という観光向けアプリで利用されているスポット名称を利用した．スポット数は約 3 万件である．そのうち Geo/Geo 曖昧性が存在するスポット名称は 2,761 件存在する．実験では，提案手法で着目している Geo/Geo 曖昧性が存在する地名で，かつ，1 日のツイート数が 10 以上の地名として，**表 3** に示した 38 地点を選択した．ベースラインとなる地名に対する事前知識を使った方法について，地名の人気度は，市区町村名の場合は人口が多い市区町村名を人気とし，駅名および観光スポット名については Foursquare [*7]のチェックインユーザ数が多いものを人気と判断した．地名に関する静的文書として，市区町村名および駅名については，Wikipedia で該当の地名を説明している文章を利用し，観光スポットについてはご当地ガイドアプリの観光スポットごとの紹介文を利用し静的特徴語の生成を行った．静的特徴語を計算する際の IDF の閾値は経験的に 5 とした．

静的特徴語および動的特徴語には，形態素解析での品詞が名詞の単語を利用した．形態素解析器には，JTAG [24] を利用した．従来手法である地名の共起による曖昧性解消を行うために利用した近隣地名辞書は，全国の都道府県名と，前述の市区町村名，駅名および観光スポット名を利用した．実験では，Geo/Geo 曖昧性がある地名に対して，近隣地名辞書中で同じ都道府県に存在する地名，および評価対象の地名が存在する都道府県名を共起の対象として利用した．動的特徴語を生成する際のパラメータは経験的に以下のものを利用した．

- 共起頻度：3 回以上で単語の長さが 2 以上
- ダイス係数：共起頻度の条件に加えて係数が 0.005 以上
- Jaccard 係数：共起頻度の条件に加えて係数が 0.001 以上
- Simpson 係数：共起頻度の条件に加えて係数が 0.05 以上

また，動的特徴語の生成対象期間として，2013 年 5 月 2 日の 1 日を対象とした．

表 3　各地名のツイートの正解数およびツイート総数
Table 3　Number of manually labeled tweets and total tweets.

地名	所在 都道府県	緯度	経度	ツイート 総数	正解 ツイート数
万博記念公園	大阪府	34.81051761	135.5300611	171	148
八坂神社	京都府	35.00363757	135.7785049	183	140
円山公園	北海道	43.05275284	141.3085349	214	69
円山公園	京都府	35.00418757	135.7809685	214	108
こどもの国	神奈川県	35.56128151	139.4896776	233	72
大通公園	北海道	43.05979772	141.348053	104	81
後楽園	岡山県	34.66864202	133.9337069	231	32
後楽園	東京都	35.707898	139.751864	231	161
水天宮	東京都	35.68368121	139.7850809	131	72
水天宮	福岡県	33.32036754	130.4960101	131	16
護国寺	東京都	35.719044	139.72754	150	124
鉄道博物館	埼玉県	35.92111664	139.6180577	147	68
駒ケ岳	北海道	42.06563817	140.6775204	98	16
駒ケ岳	秋田県	39.76002469	140.7962197	98	15
駒ケ岳	長野県	35.77212538	137.8248705	98	44
中央公園	高知県	33.56014568	133.5404346	53	10
中央公園	大阪府	34.46987535	135.3939763	53	11
京橋	大阪府	34.696047	135.534253	100	70
京橋	東京都	35.676856	139.770126	100	13
伊達市	北海道	42.4698733	140.8638647	35	20
伊達市	福島県	37.8153492	140.5538056	35	15
元町	兵庫県	34.689602	135.187401	99	42
元町	神奈川県	35.44243	139.650446	99	33
府中	広島県	34.571646	133.236021	50	10
府中	東京都	35.672245	139.4799	50	38
日本橋	東京都	35.682078	139.773516	95	20
日本橋	大阪府	34.667146	135.506635	95	50
横川駅	群馬県	36.336009	138.737926	45	14
横川駅	広島県	34.410173	132.45045	45	18
福島駅	福島県	37.754123	140.45968	36	18
福島駅	大阪府	34.697167	135.486563	36	10
那珂川町	栃木県	36.7627553	140.1445406	44	34
那珂川町	福岡県	33.5013994	130.419305	44	10
郡山	福島県	37.398187	140.389363	56	41
郡山	奈良県	34.648335	135.790441	56	11
金山駅	愛知県	35.142928	136.900517	39	34
青葉区	宮城県	38.2713197	140.8736847	27	12
青葉区	神奈川県	35.5495989	139.5451461	27	14

[*4] http://www.stat.go.jp/
[*5] http://www.ekidata.jp/
[*6] https://www.nttdocomo.co.jp/service/information/map_navi/feature/local_guide/
[*7] https://foursquare.com/

評価の正解データは前述の 38 地点の名称を含むツイートから，地名ごとにツイートをランダムサンプリングしたものを目視確認した．その際にツイートの内容や前後のツイート，ツイート中に含まれる URL の Web ページ，投稿したユーザのプロフィールから同名の地名のうち，どこについて言及しているか判断し，人手により正解ラベルを付与した．人手により正解ラベルを付与したツイート総数は 2,344 ツイートである．各地名のツイート総数，正解ツイート数を表 3 に示す．ツイート総数は評価対象の地名を形態素として含むツイート数であり，そのツイートのうち人手により評価対象の地名に関するツイートを判断したツイート数が正解ツイート数である．なお，評価対象の地名が同名の場合，1 つの評価データに複数の場所のデータが入っている．たとえば，円山公園の評価データには北海道と京都の両方のデータが含まれている．動的特徴語の評価では，動的特徴語の抽出に利用したツイートと正解ラベル付きツイートには重複がないように評価を行った．

4.2　実験結果

本節では実験の定量的および定性的な評価を行う．

4.2.1　定量評価

実験の定量的な評価指標には再現率（recall），適合率（precision）および F 値（F-measure）を利用する．再現率は人手で地名に関連するとラベルを付けたツイートのうち，いくつのツイートを抽出できたかという割合，適合率は抽出したツイートのうち，人手で正解ラベルを付与したツイートの割合である．F 値は再現率と適合率の調和平均によって求められる．計算方法は以下のとおり．

$$\text{F 値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}}$$

各曖昧性解消手法での再現率，適合率および F 値の平均を**表 4** に示す．静的特徴語の評価結果には従来手法で曖昧性解消したデータも含まれている（図 1 (a)，(b) に該当）．動的特徴語については，従来手法および静的特徴語で曖昧性解消を行ったデータも含まれている（図 1 (a)，(b)，(c) すべてに該当）．従来手法は，地名の人気度を利用する方法は，再現率は近隣地名を利用する方法より高く，適合率は近隣地名を利用する方法より低い結果となった．これは，人気のある地名のほうがツイートされることが多く，人気がある地名についてはすべてのツイートを抽出できるため再現率が高くなったと考えられる．しかしながら，適合率については，最も人気がある地名以外はすべて誤りとなってしまうため低くなっている．一方，近隣地名を用いる方法では，地理的に近い地名が共起しやすいヒューリスティクスが有効であり適合率が高くなっている．なお，予備実験では Twitter での地名の共起が 1 割未満となっていたが，これは予備実験の対象地名が 5 つのみであったためであり，本定量評価ではそれよりも多くの地名を使ったため共起の割合が増加していると考えられる．

次に，提案手法の再現率，適合率および F 値の平均値について，従来手法との差に関して有意差検定した結果を**表 5** に示す．従来手法は地名の人気度を利用する方法より近隣地名を利用する方法が総合的な性能を表す F 値が高かったため，近隣地名を利用した手法を検定の比較対象とした．表中の数値は両側検定した結果の有意水準のパーセンテージである．静的特徴語について，再現率は有意水準 5%で向上している．一方，適合率は 0.005 低下しているが，検定の結果，有意な差があるとはいえない結果となった．そのため，適合率を維持して再現率を改善したといえる．動的特徴語については，すべての共起の指標で有意水準 0.1%で再現率を向上している．適合率は共起頻度および Simpson 係数を使った場合は低下しているが，ダイス係数および Jaccard 係数を使った場合では有意な差は見られなかった．そのため，ダイス係数および Jaccard 係数を使った場合は静的特徴語と同様に適合率を維持して再現率を改善したといえる．最後に，F 値については，静的特徴語および動的特徴語のすべての指標で数値が向上した．F 値についての有意水準は静的特徴語を使った場合は 1%，動的特徴語を使った場合は 0.1%であった．共起頻度および Simpson 係数は，適合率が低下したが再現率と適合率を総合的に見た場合，従来手法より性能を改善したといえる．

以上の結果から，特徴語を利用することで曖昧性解消の性能を改善できることを定量的に確認できた．また，動的特徴語の共起指標については，どの指標でも F 値が向上していることから，動的に特徴語を生成し曖昧性解消に利用

表 4　各手法の平均再現率，適合率および F 値
Table 4　Mean of recall, precision and F-measure of each method.

手法	再現率	適合率	F 値
従来手法（近隣地名）	0.501	0.935	0.616
従来手法（地名人気度）	0.573	0.575	0.573
静的特徴語	0.529	0.93	0.64
動的特徴語（共起頻度）	0.638	0.905	0.718
動的特徴語（ダイス係数）	0.585	0.924	0.682
動的特徴語（Jaccard 係数）	0.597	0.924	0.693
動的特徴語（Simpson 係数）	0.611	0.906	0.702

表 5　従来手法（近隣地名）との有意差検定の結果．表中の数字（%）は有意水準を表す．
Table 5　Result of statistical significance test between conventional method and proposed method.

手法	再現率	適合率	F 値
静的特徴語	5%	有意差なし	5%
動的特徴語（共起頻度）	0.5%	有意差なし	1%
動的特徴語（ダイス係数）	0.5%	有意差なし	0.5%
動的特徴語（Jaccard 係数）	0.5%	有意差なし	0.5%
動的特徴語（Simpson 係数）	0.5%	5%	1%

表 6　静的特徴語を使って抽出された事例
Table 6　Example of tweets disambiguated by static location-related words.

地名	所在都道府県	投稿日時	本文
後楽園	岡山県	2013/5/2 21:17	ここ一か月ちょいの間に後楽園と 兼六園、二つの大きな庭園を見に行ったんだけども、この二か所、同じ江戸時代の武士が作った庭園なのにこうも趣が違うもんなんだなーとしみじみ
後楽園	岡山県	2013/5/2 16:32	偕楽園、兼六園、後楽園 RT @username [0426]日本三名園 をすべて答えなさい。
万博記念公園	大阪府	2013/5/2 12:42	万博記念公園は 太陽の塔 しか見るもん無いな…空港戻ってビールでも飲むか…
水天宮	福岡県	2013/5/2 15:09	近すぎてかえって行かないオリジナル水天宮で、御朱印頂戴。 (@ 全国 総本宮 水天宮) [pic]: http://t.co/NX6yAaakTk
府中	東京都	2013/5/2 18:51	RT @username: いよいよ明日から、大国魂神社 の例大祭「くらやみ祭」が始まります。５月３日は 武蔵国 府太鼓の演奏が午後１時と３時から、府中囃子の競演が午後６時から、競馬式が午後８時から行われます。ぜひ、お越しください。写真は昨年の様子。＃ fuchu ＃府中 http…
日本橋	大阪府	2013/5/2 1:35	東京チカラめし 宗右衛門町店 (日本橋、近鉄日本橋、長堀橋) ◆４月２７日のランチは東京チカラめしで。大阪は日本橋だけだと思っていたら、どんどん支店も増えているようですね。一度食べたいと思っていました。宗右衛門町は 堺筋 沿いに発見して即・… http://t.co/3JBF08vhuP
横川駅	群馬県	2013/5/2 21:31	関東なので横川駅、と言われると 碓氷峠 のアレを想像するほうの住民です。鉄道文化むら、行ったことありますよ。
金山駅	愛知県	2013/5/2 16:14	金山駅でお父さん待つから アスナル きたら、セカオワが来るみたい！

するという枠組みが有効であるといえる．

4.2.2　定性評価

ここでは定性的な評価を行うため，提案手法により抽出されたツイートの事例を見て評価を行う．まず，静的特徴語を使って抽出された事例を**表 6** に示す．静的特徴語を下線付きの赤字で示している．静的特徴語として，岡山県の後楽園に対して「日本三名園」「偕楽園」「兼六園」が使われていたり，東京都の府中に対して「大国魂神社」が使われていたりするなど，その場所ならではの単語によって曖昧性解消が行われていることが分かる．このような単語は時期によって変わらず，その場所の特徴を表している．

次に，動的特徴語を利用して曖昧性解消された事例を**表 7** に示す．ここでは Jaccard 係数を計算して生成された動的特徴語を使った場合を例示しており，動的特徴語を下線付きの赤字で示している．動的特徴語を使った場合は，大阪府の万博記念公園では「MARKET」，京都府の円山公園では「音楽堂」「しゃぼん玉」，北海道の円山公園では「花見」，東京都の護国寺では「チベット」「フェスティバル」「プロジェクション」「マッピング」「デジタル」「掛け軸」，東京都の府中では「くらやみ祭」などイベントが話題になっており，そのイベントに関連する単語が動的特徴語として抽出されている．また，北海道の大通公園では「報道ステーション」「青山愛」が動的特徴語として使われており，こちらはテレビ放送が話題となっていることが分かる．動的特徴語を利用して曖昧性解消を行う文書と動的特徴語を生成した期間はほぼ重なるものと考えられる．たとえば，前述の大通公園の事例のようにテレビ番組が要因となって話題になり動的特徴語になった場合，テレビが放映されている間投稿が増えるため，その期間の投稿に対して曖昧性解消が有効な特徴語となると考えらる．実際に前述の動的特徴語により曖昧性解消されたツイートの投稿時刻は 2013 年 5 月 2 日 22:03〜22:39 であった．一方，万博記念公園，護国寺，府中の事例のように，お祭りやイベントなど日単位で開催されるものが要因となって動的特徴語が生成された場合は，より長期間のツイートに対して曖昧性解消に有効な特徴語が生成されると考えられる．護国寺の例では動的特徴語により曖昧性解消されたツイートの投稿時刻は 2013 年 5 月 2 日 00:01〜23:52 であった．

このように，場所のトピックが動的特徴語として抽出されており，それが曖昧性解消に効果があることが分かる．

5.　おわりに

本稿では地名ごとにその場所特有の特徴語を利用することで地名の曖昧性解消を行う手法を提案した．マイクロブログでは文章が短く地名の共起が少ないため，従来手法の地名の共起以外に場所ごとの特徴語を利用した．場所特有の特徴語にはマイクロブログの投稿にはその場所特有のトピックが反映されていると考え，時間経過にかかわらずその場所の特徴を表す静的特徴語と，イベントなどの時間経過によって変化する動的特徴語の 2 種類を用いた．動的特徴語ではマイクロブログの特徴であるリアルタイム性を反映した単語を利用している．提案手法に対して定量的，定性的な評価実験を行い有効性を確認した．動的特徴語については，地名と単語の関連度の指標として，共起頻度，ダイス係数，Jaccard 係数，Simpson 係数により地名との関連度を計算し，指標ごとの再現率，適合率の比較を行った．

表 7　動的特徴語を使って抽出された事例
Table 7　Example of tweets disambiguated by dynamic location-related words.

地名	所在都道府県	投稿日時	本文
万博記念公園	大阪府	2013/5/2 20:20	明日は、FM802 and FM COCOLO 765 主催のフリーマーケット「FUNKY MARKET」に参加出店します。万博記念公園のお祭り広場でします。お立ち寄りくださいませ。
万博記念公園	大阪府	2013/5/2 20:00	明日万博記念公園で FM802 × FM COCOLO 主催のフリーマーケット「FANKY MARKET 」開催します！ 時間 9:30 17:00、入園料 20 円のみ！ 私も mamaia。で参加します□お暇な方はぜひ http://t.co/45ksiky2p9...
円山公園	京都府	2013/5/2 5:42	RT @username: SOLE CAFE は、5/26(日)@円山公園 音楽堂 にて開催される「Rainbow's End 2013」にてフード店として参加させて頂きます。SOLE CAFE でライブ御出演して頂いている方々もたくさん御出演の素敵フェス！心よりお待ちしております。…
円山公園	京都府	2013/5/2 9:15	RT @username: ☆ しゃぼん玉 パレード第 4 弾☆では、ステージ企画のショートスピーチ希望者、大募集中です！！！５月１２日（日）１３：００円山公園ラジオ塔前広場 http://t.co/V5E5gz1zwL
円山公園	京都府	2013/5/2 12:59	FLAKE でのチケット販売に 5/26 に円山公園 音楽堂 で行われる五味やタンテ、YeYe、UNCHAIN 等が出演する RAINBOW'S END 2013 の前売り取り扱いを追加！その他販売中のチケットはコチラ http://t.co/YdSFurXoCQ
円山公園	北海道	2013/5/2 11:46	さっき円山公園を通ったが、桜のさの字も咲いてませんでした。でもジンギスカンで 花見 してる団体がいました。我慢大会か？
円山公園	北海道	2013/5/2 15:47	地下鉄東西線 円山公園まであと少し。寒い
円山公園	北海道	2013/5/2 19:22	花見 の名所、すぐそこの円山公園は今の期間だけ火気許可。期限ギリギリまで咲かなかったりしないかな…
護国寺	東京都	2013/5/2 4:19	どうやら風邪こじらせて頭痛いですけど今日はインド大使館行くがてら護国寺の チベット・フェスを覗いてみたいと思います☆
護国寺	東京都	2013/5/2 8:49	RT @username: チベットフェスティバル2013 in Tokyo に福島で知り合った素晴らしいアーティスト山作戦さんのステージで 5/5 参加させて頂くことになりました。幻想的な光に包まれる夜の護国寺の前で繰り広げられる演奏、ぜひお越し下さい。https:/…
護国寺	東京都	2013/5/2 9:23	Ｆヨコがちゃんと入らず聞きにくいので、聞きやすい局はどこ？と回していたら、チベットの話をしていたこの局はどこかしら。護国寺の チベットフェスティバル 行きたいな。渡辺一枝さんの『消されゆく チベット』という本を買ったばかり。
護国寺	東京都	2013/5/2 21:17	護国寺の プロジェクションマッピング、空いてる！ @ 護国寺 (Gokoku-ji Temple) http://t.co/IdXjM8bvIS
護国寺	東京都	2013/5/2 21:26	RT @username: 護国寺の チベット フェス、19：00 から 22：00 の デジタル掛け軸 が面白すぎる。護国寺の本堂に投射されたサイケデリックアートとフリージャズ！！　これは見物ですぞ！！　（観覧無料）
大通公園	北海道	2013/5/2 10:33	5/12(日)【ぼくとわたしの未来行進】大通公園西 6 丁目 に集合！ 12:00 オープニング → 13：00 開会 → 14：00　パレード開始！詳細は HP をご覧ください。 http://t.co/O94FikgdCa HP 下に拡散ツイートボタンもあります。 #未来行進
大通公園	北海道	2013/5/2 22:03	あれ、報ステ。青山愛 アナが大通公園で中継中！
大通公園	北海道	2013/5/2 22:04	報道ステーション、大通公園から生中継！
大通公園	北海道	2013/5/2 22:06	報道ステーション お天気今日は大通公園からなのか！
大通公園	北海道	2013/5/2 22:06	青山愛 さんが大通公園に！
京橋	大阪府	2013/5/2 0:31	RT @username: ブログが更新されました。ブログタイトル： キャラメルパッキング”KADOYAN”のブログ記事タイトル： 京橋 ベロニカ ♪♪▼ブログを見る http://t.co/WWGBli68ec
京橋	大阪府	2013/5/2 0:31	今朝の 1 曲。堺出身在住のピアノとサックスの女性 2 人組 tricolore さんで「サヨウナラ」http://t.co/VVAi8jmB9M つい最近完成した PV。昨年の５月２０日京橋 ベロニカ での１ st ワンマンライブでこの曲の MC と歌を聴いて自然と涙したのが思い出されます。
郡山	福島県	2013/5/2 9:56	RT @username: NEW ステッカー間に合った！！明日 5/3 は郡山 PEAK ACTION でライブです！お時間ある方は是非☆ http://t.co/1l178y8t0B
日本橋	大阪府	2013/5/2 0:02	実家帰る予定がだいぶ先になったから余裕あったらいおりん誕生日に日本橋 ナムコ 行くのもあり
府中	東京都	2013/5/2 17:56	くらやみ祭り の準備が着々と進んでる＼(ˆ o ˆ)／今年もこの時期がきたあぁああ！！府中のビッグイベント！！
福島駅	福島県	2013/5/2 8:12	RT @username: 日付が変わって明日です。【金曜行動】『NoNukes!よりみち 音楽会』～あなたの音が声になる～[とき] ５月３日（金） １８時～１９時 [ところ] 福島駅東口 AXC 向かい 街なか広場西側歩道　途中参加、途中離脱ＯＫ!楽器や 鳴り物、プラカード を持ち寄り…

今後の課題は，従来手法の地名曖昧性解消が誤ると動的特徴語も誤るため，動的特徴語の信頼性を判定する仕組みを検討したり，Rauch ら [25] のように近隣地名と地名人気度の両方を利用することで動的特徴語のもとになるツイートの曖昧性解消の性能を向上させたりすること，動的特徴語の生成対象とするマイクロブログの期間をアプリケーションに応じて適切な期間を検討することなどである．

謝辞　本稿を作成するにあたりご助言いただいた筑波大学准教授手塚太郎先生に記して謝意を表する．

参考文献

[1] Cheng, Z., Caverlee, J. and Lee, K.: You Are Where You Tweet: A Content-based Approach to Geo-locating Twitter Users, *ACM CIKM '10*, pp.759–768 (2010).

[2] 橋本康弘，岡　瑞起：都市におけるジオタグ付きツイートの統計，人工知能学会誌，Vol.27, No.4, pp.424–431 (2012).

[3] Kitamoto, A.: Toponym-based Geotagging and Disambiguation for Social Media on Earthquake and Weather Events, *10th International Conference on ISCRAM 2013* (2013).

[4] Amitay, E., Har'El, N., Sivan, R. and Soffer, A.: Web-a-where: Geotagging Web Content, *Proc. ACM SIGIR '04*, pp.273–280 (2004).

[5] Qin, T., Xiao, R., Fang, L., Xie, X. and Zhang, L.: An Efficient Location Extraction Algorithm by Leveraging Web Contextual Information, *Proc. 18th ACM SIGSPATIAL*, pp.53–60 (2010).

[6] Leidner, J.L.: Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names, Ph.D. Thesis, University of Edinburgh (2007).

[7] 長岡　諒，松本光弘，沼尾正行，栗原　聡：Web における実世界の位置情報類推に関する研究，第 23 回人工知能学会全国大会論文集 (2009).

[8] Liu, X., Zhang, S., Wei, F. and Zhou, M.: Recognizing Named Entities in Tweets, *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, HLT '11*, pp.359–367 (2011).

[9] 杉谷卓哉，白川真澄，原　隆浩，西尾章治郎：教師あり機械学習を用いたツイート投稿時のユーザ位置推定手法，情報処理学会研究報告　DBS, Vol.158 (2013).

[10] Serdyukov, P., Murdock, V. and van Zwol, R.: Placing Flickr Photos on a Map, *Proc. 32nd International ACM SIGIR*, pp.484–491 (2009).

[11] 手塚太郎，近藤浩之，田中克己：混合ガウス分布を用いたウェブコンテンツの地域性推定とオブジェクトレベルローカルサーチ，情報処理学会論文誌　データベース，Vol.1, No.1, pp.13–25 (2008).

[12] 奥　健太，西崎剛司，服部文夫：地域限定性スコアに基づく位置情報付きコンテンツからの地域限定語句の抽出，情報処理学会論文誌　データベース，Vol.5, No.3, pp.97–116 (2012).

[13] 馬場雪乃，石川冬樹，本位田真一：Folksonomy 上のタグと関連する場所の抽出，人工知能学会論文誌，Vol.27, No.1, pp.1–9 (2012).

[14] 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).

[15] 山口祐人，伊川洋平，天笠俊之，北川博之：ソーシャルストリームからのイベント検出とユーザ位置推定の統合，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 2013），pp.A5–2 (2013).

[16] Dalvi, N., Kumar, R. and Pang, B.: Object Matching in Tweets with Spatial Models, *Proc. 5th ACM WSDM*, pp.43–52 (2012).

[17] 酒巻智宏，岩井将行，瀬崎　薫：マイクロブログのジオタグを用いたユーザの行動パターンの推定に関する研究（行動解析，第 2 回集合知シンポジウム），電子情報通信学会技術研究報告，NLC，言語理解とコミュニケーション，Vol.110, No.400, pp.37–42 (2011).

[18] 若宮翔子，李　龍，角谷和俊：Twitter における群衆の経験に基づく近接地域検索システム，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 2013），pp.A3–3 (2013).

[19] 伊川洋平，榎　美紀，立堀道昭：マイクロブログのメッセージを用いた発信場所推定，第 4 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 2012），pp.F7–2 (2012).

[20] 渡辺一史，大知正直，岡部　誠，尾内理紀夫：Twitter を用いた実世界ローカルイベント検出，楽天研究開発シンポジウム (2011).

[21] Watanabe, K., Ochi, M., Okabe, M. and Onai, R.: Jasmine: A Real-time Local-event Detection System Based on Geolocation Information Propagated to Microblogs, *Proc. 20th ACM CIKM*, pp.2541–2544 (2011).

[22] 北　研二，津田和彦，獅々堀正幹：情報検索アルゴリズム，共立出版 (2002).

[23] 松尾　豊，友部博教，橋田浩一，中島秀之，石塚　満：Web 上の情報からの人間関係ネットワークの抽出，人工知能学会論文誌，Vol.20, No.1, pp.46–56 (2005).

[24] 今村賢治，齋藤邦子，浅野久子：テキストからの知識抽出の基盤となる日本語基本解析技術，NTT 技術ジャーナル，Vol.20, No.6, pp.20–23 (2008).

[25] Rauch, E., Bukatin, M. and Baker, K.: A Confidence-Based Framework for Disambiguating Geographic Terms, *Proc. HLT-NAACL 2003 Workshop on Analysis of Geographic References*, pp.50–54 (2003).

落合 桂一　（正会員）

2006 年千葉大学工学部情報画像工学科卒業．2008 年同大学大学院博士前期課程修了．同年株式会社 NTT ドコモ入社．SNS および位置情報データ解析の研究開発に従事．日本データベース学会会員．

鳥居 大祐　（正会員）

2001 年京都大学工学部情報学科卒業．2006 年同大学大学院社会情報学専攻にて博士（情報学）を取得．現在，株式会社 NTT ドコモにて，データマイニング，検索，リアルタイム処理，位置情報解析に取り組む．

（担当編集委員　北山 大輔）