情報処理学会論文誌 Vol.62 No.1 263–274 (Jan. 2021)

SNSや人口統計に基づく人気度を考慮した施設検索

川崎 仁嗣[1,2,a)] 深澤 佑介[1] 豊田 正史[2]

受付日 2020年4月6日, 採録日 2020年10月6日

概要：Point-of-Interest（POI）の検索は，地図やカーナビアプリでの目的地検索だけでなく，場所に対するチェックインや写真へのジオタグ付けなど様々な用途で利用されている．POI検索では施設名称や施設ジャンルを検索クエリとして利用することも多いが，同一の施設名称であるPOIが存在したり，有名な施設とそうではない施設が混ざって表示されたりすると本来意図するPOIを探し出す際のユーザ体験に悪影響となる．そのため，現在地付近のPOIを優先して提示するなどの方法も考えられるが，ランドマークなどの遠方にある有名なPOIを近隣のPOIよりも上位に表示してほしい場合は多い．我々はソーシャルメディア上でのPOIに対する言及がある投稿の数やPOI周辺エリアの昼間時間帯滞在人口に応じて施設人気度を付与し，施設人気度も加味した表示順序とすることで，多くのユーザにとって想定しているPOIが上位に表示されることを目指した．観光施設検索ログ，およびスマホ向けカーナビアプリの目的地検索ログを用いた検索精度の評価において，いずれのデータセットでも想定するPOIが検索結果のTOP1で検索される割合が13%程度向上することを確認した．

キーワード：施設検索，施設人気度

Point-of-Interest Search Considering Popularity based on Social Network Services and Population Statistics

Satoshi Kawasaki[1,2,a)] Yusuke Fukazawa[1] Masashi Toyoda[2]

Received: April 6, 2020, Accepted: October 6, 2020

Abstract: Point-of-Interest (POI) search is used in not only searching for destinations on maps and car navigation services, it is used for various purposes such as checking-in for places and geotagging of photos. Users often input the facility name as a search query, but some POIs may have the same facility name. Furthermore these search results are oftenly unsorted, so a globaly famous POI is listed with a local POI. This situation affects the user experience when POI searching. To solve this problem, the distance between POIs and the current location is used to rank the POI around the user higher, but it is usual that users want a famous landmark far away is ranked ahead of near POIs that have same name. We assign facility popularity according to the number of posts refering to the POI on social media and the number of people staying around the POI during daylight, and aimed to rank the POI expected for many users higher by using a sort order that also includes facility popularity. In our evaluation, proposed method improved search accuracy about 13% for the search log of tourist facilities and destination facilities in car navigation application for the smartphone.

Keywords: Point-of-Interest search, POI popularity

1. はじめに

Point-of-Interest（以降はPOIと呼ぶ）とは，地図における施設などのユーザが関心を持つ場所のことであり，一般的にイメージされるような商業施設やオフィスビルなど

[1] 株式会社NTTドコモ
NTT DOCOMO, INC., Chiyoda, Tokyo 101–6150, Japan
[2] 東京大学
The University of Tokyo, Meguro, Tokyo 153–8505, Japan
[a)] satoshi.kawasaki.vx@nttdocomo.com

の施設だけでなく，海岸や山などの自然物，観光名所，待ち合わせスポットやランドマークなども含まれる．オンライン地図サービスやカーナビゲーションシステムなどでは，目的地などに設定される施設などがあらかじめ多数のPOIとして登録されており，ルート検索などの出発地や目的地として住所や緯度経度を直接指定するだけでなく，これらのPOIの所在地を目的地などとして指定することができる．POIは地図サービス以外でも，たとえばソーシャルメディアにおいて，その場所にいることをチェックインとして投稿したり写真に場所をタグ付けしたりする際にも用いられることが多い．また，POIにはその場所を示す緯度経度だけではなく，施設名称や住所などの関連する属性情報も含んでいることが多い．これにより，ユーザは施設の名称や電話番号，施設ジャンルなどの覚えやすい情報から，目的とするPOIを検索して指定することができる．

ユーザにとって意図したPOIを検索できることは，これらのサービスにおけるユーザ体験に大きく影響する．所望のPOIが意図通りに検索できなければ，他のサービスへ移行されたり，利用をあきらめられてしまったりすることが想定される．意図したPOIを検索できない要因としては大きく2つあり，1つ目はそもそもユーザが想定する施設などに対応する適切なPOI自体のデータが存在しないこと，2つ目は適切なPOIは存在するものの検索結果としてヒットしない，またはヒットしたものの上位の候補として表示されないことがあげられる．1つ目の要因に対しては，適切なPOIが存在しない場合にユーザ自身に新たにPOIを登録してもらうことで対応が可能であり，Google Maps [1]においては地図上の任意点にピンを立て，施設名称などをユーザが入力してPOIとしての追加が可能である．

一方で2つ目の要因に対しては，主に検索技術の観点で様々な対応策が提供されており，luceneプロジェクトのSolr [2]や蘭Elastic社が開発しているElasticsearch [3]などの実サービスでの利用が多い全文検索エンジンにおいては名称が完全一致ではなくても，名称の一部の形態素や部分文字列が一致していたり，名称以外の別の属性情報での一致で検索させることが可能である．これにより適切なPOIを検索結果候補としてヒットさせることは可能なものの，ユーザに対して上位の検索結果として提示するには不十分な場合がある．具体的には，名称が同一または類似度の高い施設や，名称ではなくジャンルで検索した場合があげられ，既存の検索エンジンの機能だけでは適切なPOIをより上位に表示させることが難しい．名称やジャンルが同一であっても，所在地が異なるのであれば，たとえば現在地から近いPOIを上位に表示する方法も考えられるが，観光地などの情報を検索する場合，必ずしも近いPOIを提示するのが適切とはいえない．

ユーザがPOIの検索を行う場合，より人気があるPOIを上位に出すことにより，一部の例外を除きユーザが意図する検索結果である可能性は高いと考えられる．例外とは，自分の現在地周囲や普段の生活圏内にあるPOIについてより上位に表示されてほしい場合があり，大西らの研究 [4]では生活圏にあるPOIをより上位に提示する手法が提案されている．人気が高いPOIであればソーシャルメディア上で言及されることも多いと考えられることから，我々はソーシャルメディア上でのPOIに対する言及がある投稿の数に応じて施設人気度を付与することで，人気が高いPOIを検索結果の上位に提示させる方法を提案する．

本研究による貢献としては以下があげられる．

- ソーシャルメディア上の投稿データ本文において，POIに関する言及が含まれるかを判別することで施設人気度を付与した．
- SNS上で言及されないPOIに対しても，他の言及されているPOIのデータを学習することで，仮に言及されたとしたときの言及数を推定する手法を提案した．
- 施設人気度を用いることでPOI検索における検索精度の向上があることを確認した．

2. 施設人気度

施設人気度とは施設に対する関心度の高さを表した度数である．多くのユーザが関心を持つ施設には大きい人気度が，ごく一部のユーザのみ関心を持つ施設には小さい人気度が付与されることを期待している．施設に対し人気度が付与されていれば，より多数のユーザにとって関心のある施設情報をユーザに提示することができる．

施設検索において施設人気度を用いる主な理由としては，同名施設やジャンル名称での検索時の検索結果に順序付けをすることがあげられる．具体例で述べると，検索キーワードとして「清水寺」を指定した場合，多くのユーザは京都にある清水寺が検索結果上位に提示されることを期待する．しかし，「清水寺」は全国に数箇所存在するため，**図1**のように施設人気度が高い「清水寺」を上位に提示することが，多くの場合でユーザが期待する順序となる．単純に距離が近い施設を提示する方法も考えられるが，「清水寺」などのように著名な観光スポットやランドマークの場合は人気度が高い施設を提示することが期待される．また，ジャンル名称での検索の場合は，「居酒屋」などの曖昧な検索キーワードの場合は検索結果となる施設が複数存在する．この場合も同様に，施設の人気度が高い居酒屋を上位に提示することが，多くの場合でユーザが期待する順序であると考えられる．

上記で述べたように，より多数のユーザにとって人気である施設に対して大きな施設人気度を付与するため，我々はSNS上でPOIに関する言及を行っている投稿数と，施設周辺の滞在者数を組み合わせた施設人気度を用いる．施設人気度をSNS上の言及数だけでなく，実際に施設が存在するエリアにおける人口分布のデータを組み合わせるこ

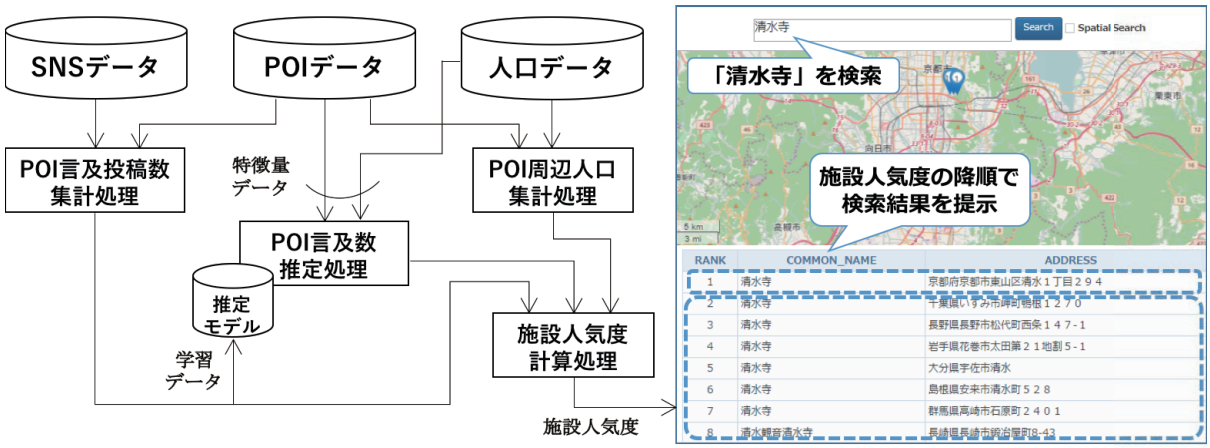

図 1 POI 人気度推定処理の概要
Fig. 1 Overview of POI popularity estimation process.

とで，ネット上での人気度だけでなく，実際に施設周辺の訪問者数を加味した人気度とすることができる．

SNS 上での言及数のようなユーザによる投稿情報を用いる場合，実際にチェックインの投稿が行われる POI は人気のある一部の POI に偏りがちである．このため，大多数の POI はほとんど言及されず，人気度の推定ができないという問題があげられる．

本研究では SNS 上で言及されていない POI に対し，仮に言及されていたときにどれだけの言及数となるかを回帰決定木モデルにより推定する．施設人気度の推定処理を図 1 に示したとおり，「施設人気度生成」と「施設人気度推定」から構成される．施設人気度生成処理では，SNS の投稿データと POI データから施設人気度を生成する．次に，施設人気度推定処理では，生成した施設人気度を学習データ，施設周辺の滞在人口や POI のジャンル，所在地の情報を特徴量として利用する推定モデルを構築することで，人気度が付与できていなかった大多数の POI に対しても言及数を付与する．

3. 関連研究

施設検索における検索結果提示順序の最適化を目的として施設人気度を算出している研究はあまり見られないが，一方で POI をレコメンドするために，ユーザが関心のある POI を算出する研究は数多く見られる．施設人気度の算出方法としては，該当施設への訪問者数を利用する方法が多く，文献 [5] では GPS などの位置測位ログから該当施設への滞留を検知し，滞在者数から算出している．多くの利用者にとって妥当な人気度を算出するためには，特定の集団に偏らずに大量のユーザから位置情報を収集する必要があり，コストなどの点で容易ではない．

3.1 位置情報付き投稿に基づく施設人気度

前述の問題に対し，SNS への投稿情報など比較的に容易に収集可能なデータを利用する方法が考えられる．Yao ら [6] は POI の人気度を POI 周辺 100 m のエリア内における時間帯ごとの滞在人数と，POI のカテゴリにおける時間帯ごとのチェックイン確率とを組み合わせる手法を提案している．エリアの滞在人数についてはタクシーの降車位置と時刻のデータから集計し，カテゴリごとのチェックイン確率については位置情報付き SNS の投稿データから集計しており，後者はカテゴリ単位での集計とすることでデータのスパース性を解決している．この手法では近隣エリアにある同一ジャンルの POI の場合，ほぼ同じ施設人気度が割り当てられるため，周辺にある居酒屋を人気度順でランキングにするような利用方法には適さない．本研究では，SNS 投稿データでの言及が存在しない POI については，POI のジャンルに加え，POI 周辺の施設数や性年代別の滞在人口などの特徴量から回帰決定木モデルによる推定を行う方法とした．

Hsieh ら [7] の研究ではチェックイン回数を人気度として用いているが，新規に開店する施設の人気度を周辺施設との関係から推定している．本研究では，近隣に施設が存在しない観光スポットなどにも人気度を付与するため，周辺施設の情報のみを用いるのではなく，施設カテゴリなど複数異種の情報も利用する．

Ying ら [8] の研究では POI のジャンルごとにチェックインの頻度が異なることから同一ジャンルである POI のチェックイン数の合計に対する該当 POI のチェックイン数を施設人気度として用いることが述べられている．これにより訪問頻度が一般的に低いジャンルの POI（たとえば海水浴場）であっても，他の訪問頻度が高いジャンルの POI と比較し大幅に低い施設人気度となってしまうことがなく，訪問頻度が高いジャンルの POI に偏ってレコメンドがなされてしまうという問題が解決できる．本研究ではチェックインではなく，SNS での POI に関する言及ツイートの数を施設人気度として用いており，チェックイン

のされにくいジャンルであっても言及ツイートが少なくなるわけではないことから，ジャンル間での人気度の正規化は行わない．これにより，同一名称であるがジャンルが異なるPOIであっても，より有名であるPOIに対して大きな人気度を割り当てることができる．

3.2 投稿内での言及に基づく施設人気度

チェックイン投稿は実施に施設に訪問した際に行われることから，通常の投稿と比較しチェックインに関する投稿は件数が限られている．たとえば，Twitterにおいては，位置情報付きのTweetが全体の0.58% [9] にすぎず，また，Foursquare（現Swarm）でのチェックイン情報をTwitterに連携しているユーザは15.7% [10] と少ない．実際には訪問していないが話題になっている施設の場合は，人気度を正確には推定できていない．そこで，チェックインに関する投稿だけでなく，通常の投稿も用いてPOIに対する言及がある投稿数を集計することで，話題になっている，つまり，ネット上で人気が高い施設に対して大きな人気度が付与されることを期待できる．そのためには，投稿データの本文を解析して，POIの名称が含まれているかを確認することになるが，POIの名称が人名などの地名以外を指す言葉としても用いられていたり，地名だとしても同一の地名が複数箇所に存在したりする場合があり，実際にPOIを指す言葉として用いられているかを判別する必要がある．このような投稿データ本文内の単語が地名を指すものとして扱われているのか否かを判別する手法として，Liuら [11] の研究では条件付き確率場（Conditional Random Field；CRFとも呼ばれる）を用いた形態素の系列ラベリングを行うことが提案されており，落合ら [12] の研究では共起語を用いる手法が提案されている．

人気度を推定するにあたって，チェックイン数ではなく複数の観光情報サイトにおけるコメントや画像，レーティング情報を用いる手法 [13] も提案されている．本研究では人気度の推定対象となるPOIの割合を増加させるために，投稿者数が限られる観光情報サイトではなく，より投稿者数が多いSNSの投稿を用いることとした．

なお，SNSや観光情報サイトなどへの投稿を用いる手法は，そもそもGPSなどの位置情報自体を収集するのが難しいことから提案されてきた手法であり，投稿情報に加えて位置情報を利用する手法は特にみられなかった．

4. SNSと人口統計からの施設人気度生成

施設人気度は，前章までに述べたとおり，ネット上での人気度を表すSNS上での言及数と施設周辺の実際の滞在者数を表す人口分布との組合せで構成される．SNS投稿データでの言及を用いた施設人気度は，チェックインに関する投稿だけでなく，通常の投稿も用いてPOIに対する言及がある投稿数を集計することで算出を行う．実際にユーザがその場所に行ったかどうかは考慮していないが，施設に対する言及を行っているということはSNS上で話題になっている，つまり，人気が高いことを反映していると考えられる．

投稿データの本文を解析してPOIについて言及しているかどうかを判断するうえで，以下の3つの問題が存在する．

(1) 表記ゆれ問題
(2) 人名・同一地名問題
(3) チェーン総称問題

また，施設周辺の実際の滞在者数を表す人口分布の作成方法についても述べる．

4.1 表記ゆれ問題

正式名称の文字数が長い場合や，大学名称やドーム施設のように通称や愛称，略称などが用いられることが多いPOIでは，SNSの投稿データ本文においても正式名称とは異なる表記がされる可能性が高いため，正式名称との一致を判定するだけでは真の言及数よりも少なく算出されてしまう．このような表記ゆれに対応するため，本研究では，それぞれのPOIについてあらかじめ略称などの別称を一覧化したシソーラス辞書を作成しておき，正式名称とは異なる表現であっても言及数として集計対象としている．

4.2 人名・同一地名問題

同じPOIの名称が複数ある場合や，POIの名称やその略称が人名など施設名以外の一般用語としても用いられている場合に，どのPOIに対する言及であるかが曖昧となることがあり，たとえば，「清水寺」は京都府京都市東山区清水にある清水寺を指すこともあれば，千葉県いすみ市岬町鴨根にある天台宗の清水寺を指すことも考えられる．

落合ら [12] の研究では共起語を用いる方法が提案されており，従前の「清水寺」の例であれば，同じ本文内に「京都」，「舞台」などの単語が含まれる場合，京都府にある清水寺を指す可能性が高く，一方で「千葉」，「いすみ市」，「坂東三十三」などの単語が含まれる場合は千葉県の清水寺を指す可能性が高いと判断する．この際に利用する共起語は，POIに関する観光案内説明文から抽出した単語を用いているが，実際に観光案内説明文が存在するPOIは全体のごくわずかであり，大多数のPOIには観光案内説明文は存在しない．各投稿文についてCRFを用いて地名かどうか判定する手法もあるが，言及数を集計するためには大量の投稿データを処理する必要があり現実的ではない．

本研究では，POIに付与されている住所などの属性データや，形態素解析に用いられる形態素辞書を用いて，以下の判断基準に基づいて共起語の選定を行った．なお，形態素辞書で未知語となるなど判定ができなかった場合については，CRFを用いて地名と判定された場合にのみ言及数として集計を行った．

1) POI の名称が他の POI と重複しない，かつ，形態素辞書における品詞が一般名詞ではない（固有名詞，名詞–地名，名詞–組織名など）場合，共起語なし．
（例）スカイツリー，JR 東京駅
2) POI の名称が他の POI と重複する，または，形態素辞書における品詞が一般名詞である場合，共起語として住所の一部（都道府県名，市区名），もしくは，施設ジャンル名を用いる．
（例）宮島（正式名称は「厳島」，
共起語は「広島」，「廿日市」，「景観地」）
2-1) ただし，複数の POI が同一の施設ジャンルの共起語となる場合は，住所や施設ジャンルを用いずに，該当 POI の近隣にある POI の名称を共起語とする．
（例）清水寺（施設ジャンルは「寺」，
共起語は「八坂神社」，「京都駅」），
清水寺（施設ジャンルは「寺」，
共起語は「合羽橋」，「浅草駅」）
2-2) ただし，複数の POI が同一の住所の共起語となる場合は，住所と施設ジャンルの組合せを共起語とする．
（例）吉祥寺（住所の一部は「東京」，
共起語は「駅」かつ「東京」）
吉祥寺（住所の一部は「東京」，
共起語は「寺」かつ「東京」）
3) POI の名称自体に住所の一部（都道府県名，市区名）が含まれる場合，共起語なし．
（例）東京ドーム

4.3 チェーン総称問題

コンビニなど系列店舗が複数存在する POI の場合，特定店舗の POI に関する言及がほとんど行われなず，ユーザはチェーン店であればどの店舗でもよい場合が多いため，具体的な店舗名まで含めた言及をすることは期待できない．本研究では，POI の名称からチェーン名称を抽出し，POI に対する言及ではなくチェーン名称に対する言及数を集計し，これを同一チェーンの各店舗の POI に均等に按分することで 1 店舗あたりの平均言及数を付与した．

4.4 人口分布に基づく人気度

関連研究で示した手法ではタクシーの降車データから POI 周辺エリアを目的地として流入してきた人数を用いて人気度を推計している．本研究では文献 [14] で提案されている携帯電話基地局の運用データを用いたエリアごとの推計人口を用いる．施設周辺の人口が多いほど，施設に訪問する人数は増加することが期待され，また，実際の訪問者数も人口に含まれることから，4.1 から 4.3 節で述べた SNS データに基づくスコアと比較して，実世界における人気度をより反映していると考えられる．

推計人口は集計処理において，性別，年代などの属性別の集計や，集計時間帯を任意に決めることが可能であるが，夜間帯における人口はそのエリアにおける居住者を多く含んでいることから，施設の人気度としては昼間時間帯の人口を用いる．推計人口は総務省統計局が定める 2 分の 1 地域メッシュに準じたエリアごとに算出され，具体的には 1 つのエリアは 500 m × 500 m の矩形領域となる．このため，エリア内には複数の POI が含まれる場合があり，エリア内の POI すべてに同一の値が付与される．最終的には SNS データに基づくスコアと組み合わせて用いるため，同一の値が付与されているものの，同一エリア内の POI でも人気度には差が生じる．

5. 言及投稿がない POI の施設人気度推定方法

人気度のうち，4 章で述べた SNS データに基づく施設人気度が付与できる POI は，実際に SNS 上で言及されるような有名であったり話題になったりしている施設に限られるため，POI 全体のうち一部に限られる．およそ 3 カ月間の Twitter データに含まれる約 387 万件の Tweet を分析したところ，4 章の手法により POI を含むと判定された Tweet の件数は約 91 万件であり，言及の対象となっている POI の数は約 2 万施設であった．言及数の集計対象としている POI は約 18 万件であることから，Twitter データから言及数を集計できる POI はおよそ 1 割程度であるといえる．

このように SNS データを用いる手法でのデータのスパース性については関連研究の多くで課題となっているが，本研究においては，言及されていない POI について，仮に言及されたときの言及数を回帰決定木のモデルによって推定する手法を提案する．

5.1 提案特徴量

推定 POI 言及数の推定モデルでは，**表 1** に示す 244 項目の特徴量を入力データとして用いる．

1) 施設ジャンル・業種

i 番目の POI P_i に対する施設ジャンル・業種 Cat_{P_i} は訪問者属性ごとの訪問頻度に影響すると考えられ，駅やコンビニのように訪問者属性間での差異が少ない施設ジャンルもあれば，学校や競馬場のように訪問者の年代に偏りが多い施設ジャンルもあるため，これらの違いを説明するための特徴量として用いる．Li らの調査 [15] でも，施設カテゴリがチェックイン数に大きく影響することが示されている．

$$Cat_{P_i} = [G_1, G_2, \ldots, G_{10}, B_1, B_2, \ldots, B_6]$$

ここで，G_j は施設ジャンル，B_k は業種を示すコードであり，1 つの POI に対して複数存在しうる．なお，業種が 1

表 1 推定モデルで利用する特徴量一覧
Table 1 List of features used in the prediction model.

特徴量種別	特徴量数	備考
1) 施設ジャンル・業種	16 項目	1 つの POI に複数ジャンルや業種が付与されうる
2) 施設所在地	4 項目	地域メッシュ番号，最寄り駅距離，市区町村コードなど
3) 周辺 200 m 以内のジャンル別施設数	59 項目	例：小学校，宿泊施設（民宿）
4) 属性ごとの平均滞在人口	117 項目	平休日，時間帯，性年代別
5) 国勢調査における人口	46 項目	性年代属性別
6) 平均滞在人口と国勢調査との人口差分	1 項目	
7) POI データ取得元種別	1 項目	

つであるがジャンルが複数ある POI も存在する一方で，業種とジャンルのいずれも存在しない POI もある．

2) 施設所在地

施設所在地 Loc_{P_i} は地域ごとの訪問傾向の差異を説明するための特徴量として用いており，POI が所在する地域メッシュ番号 $Mesh$，最寄り駅距離 $Dist_{ST}$，都道府県 JIS コード JIS_{PREF}，市区町村コード JIS_{MUNI} を含む．

$$Loc_{P_i} = [Mesh, Dist_{ST}, JIS_{PREF}, JIS_{MUNI}]$$

地域メッシュとは，総務省統計局が定めた一定の緯度経度で区切られる矩形領域のことであり，本研究で利用したデータは 2 分の 1 地域メッシュで区切られている．具体的には 1 つのメッシュは 500 m × 500 m の矩形領域となる．地域ごとの訪問傾向の差異とは，たとえば，都心では駅の訪問頻度は高い一方で，郊外では自動車利用が多いことで駅の訪問頻度が低くなる地域もあると考えられる．最寄り駅距離は，電車を利用して施設に訪問するユーザにとって訪問頻度に影響すると考えられ，同一の地域メッシュ内に同一ジャンルの施設があったとしても，駅からの距離がより近い施設が好まれる傾向が想定される．都道府県や市区町村コードは，同一の地域メッシュであっても行政区境界があることで，メッシュ内にある類似ジャンルの POI であっても訪問傾向が異なることを説明するために用いた．たとえば，区役所の開庁時間が区ごとに異なっており，一方の区役所が閉庁していても別の区の区役所がまだ閉庁しておらず訪問者がいるような場合が想定される．

3) 周辺 200 m 以内のジャンル別施設数

ジャンル別施設数 N_{P_i} は POI の周辺にあるジャンル別に集計した POI 数であり，本研究で利用したデータでは 59 個のジャンルが存在する．

$$N_{P_i} = [Q_{ct1}, Q_{ct2}, \ldots, Q_{ctl}]$$

ここで，Q_{ct_l} は周辺 200 m 以内でジャンル l である POI 数である．ジャンル別施設数は POI が存在する周囲の特性を説明するための特徴量であり，たとえば宿泊施設や観光名所の POI が多いエリアは観光客が多いと考えられ，観光スポットなどの施設への訪問者が多いと考えられる．

4) 属性ごとの平均滞在人口

属性ごとの平均滞在人口 $Stay\_Pop_{P_i}$ は，文献 [6] でのエリア活性度の考え方に類似しており，POI 周辺の平均滞在人口を集計した特徴量であり，平均滞在人口が多いエリアでは，より訪問者数が多くなると考えられる．また，POI のジャンルに応じて性年代ごとに訪問傾向が異なる可能性があるため，属性別に集計した平均滞在人口を用いる．これらの平均滞在人口は，前にも述べたとおり，携帯電話基地局の運用データに基づき，端末在圏数から推計された人口 [14] を用いており，ある時間帯の属性別人口をメッシュ単位で集計することができる．

$$Stay\_Pop_{P_i} = [S_{all}, S_{week}, S_{holiday}, S_{male}, S_{female}, S_{0,8,0,15}, S_{0,8,0,20}, \ldots, S_{d,t,g,a}]$$

ここで，S_{all} は全属性の合計平均滞在人口，S_{week} は平日の合計平均滞在人口，$S_{holiday}$ は休日の合計平均滞在人口，S_{male} は男性のみの合計平均滞在人口，S_{female} は女性のみの合計平均滞在人口である．また，$d \in \{0,1\}$ は平日（0）と休日（1）を，$t \in \{8,11,15,18\}$ は時間帯（8 時～20 時を 3 時間ごとに 4 区分．ただし，11 時～14 時のみ 4 時間）を，$g \in \{0,1\}$ は男性（0）と女性（1）を，$a \in \{15,20,30,40,50,60,70\}$ は年代（20～70 歳台まで 10 歳区切りと 15～19 歳の 7 区分）を示している．

5) 国勢調査における人口

国勢調査における人口 $Census\_Pop_{P_i}$ は，国勢調査によって集計された人口の特徴量である．国勢調査は 5 年に 1 度の頻度で実施されており，2015 年調査（2016 年 12 月公開）の結果が現時点で最新である．属性ごとの平均滞在人口と同様に，人口が多いエリアでは，より訪問者数が多くなると考えられる．国勢調査では居住地に基づき人口を集計しており，エリア内の居住人口を表しているが，属性ごとの平均滞在人口は昼時間帯におけるエリア内の人口であり，エリア外からの訪問者も含めた人口となっている点が差分である．また，国勢調査の場合，携帯電話を所持していない乳幼児や高齢者も集計されていることから，POI のジャンルによっては属性ごとの平均滞在人口よりも訪問傾向への寄与が大きくなることが期待される．

$$Census\_Pop_{P_i} = [C_{all}, C_{male}, C_{female}, C_{0,0}, C_{0,5}, \ldots, C_{g,a}]$$

ここで，C_{all} は全属性の合計人口，C_{male} は男性のみの合計人口，C_{female} は女性のみの合計人口である．また，$g \in \{0,1\}$ は男性（0）と女性（1）を，$a \in \{0,5,10,\ldots,90,95,100\}$ は年代（0〜95 歳台まで 5 歳区切りと 100 歳以上の 21 区分）を示している．

6) 平均滞在人口と国勢調査との人口差分

平均滞在人口と国勢調査での人口との差分 $Diff_{P_i}$ は，属性ごとの平均滞在人口と国勢調査における人口との差分であり，エリア外からの訪問者のみの人口を表していると考えられ，宿泊施設のように居住者は訪問しにくいがエリア外からの訪問者が立ち寄りやすい施設への訪問傾向を説明するための特徴量である．

$$Diff_{P_i} = [S_{all} - C_{all}]$$

7) POI データ取得元種別

POI データ取得元種別 $D\_Src_{P_i}$ は，POI データのデータソース種別であり，データソースごとの特性による人気度の差を説明するための特徴量である．本研究で用いている POI のデータソースは，観光者向け施設の POI データ，居住施設の POI データ，飲食店の POI データ，SNS 投稿データから抽出した POI データなどの種類が存在し，それぞれでジャンルは異なるものの同一名称の POI が含まれることがあり，かつ，ジャンルがどちらも付与されていないことがあった．実際にはジャンルが異なり人気度にも差があることから，人気度推定の際に識別可能とするための特徴量として用いた．

$$D\_Src_{P_i} = [Src]$$

ここで Src はデータソースごとにユニークに割り振られた識別子である．

5.2 推定モデル

推定モデルは，言及の投稿がある POI における言及数を正解値（目的変数）として，POI のジャンルや所在地など 5.1 節で述べた特徴量を説明変数に用いて学習を行う．

特徴量の次元数が高いこと，および，POI によりジャンルの付与数が異なるなど特徴量に欠損値があることから，このようなデータに対しても良好な精度を期待できるランダムフォレスト回帰決定木を用いた．

5.3 推定精度の評価

SNS での言及がなされていない POI に対して仮に言及されたとしたときの推定モデルの精度について評価を行った．評価に用いたデータセットの詳細を**表 2** に示す．

推定モデルの学習および評価では，SNS データにおいて 1 回以上言及されている POI を用いて実施しており，対象となった POI は約 2 万施設である．言及数が既知の POI について，5 分割交差検証で推定精度の評価を行っており，ランダムフォレストの木の数は 20，500，1,000 個で実施して最も精度が高かった 1,000 個の場合で以降の評価を進めた．

表 2 評価で用いたデータセット
Table 2 Datasets used for evaluation.

データ種別	期間	件数
POI データ	2018.06 版	約 18 万施設
SNS（Twitter）データ	2018.6〜8	約 387 万件
国勢調査データ	2015 年調査	約 33,000 件
平均滞在人口データ	2018 年間平均	約 6,100 万件

表 3 5 分割交差検証での推定誤差
Table 3 Estimation error in 5-fold cross-validation.

MAE	RMSE	相関係数	nDCG
7.41	10.52	0.60	0.68

評価指標として，言及数の真値との誤差を平均絶対誤差（MAE），平均二乗誤差（RMSE），相関係数で評価する．また，言及数により POI の提示順を決定するうえで真値の言及数を用いた場合と推定モデルによる言及数を用いた場合とでの順位の誤差を nDCG（normalized Discounted Cumulated Gain）で評価した．nDCG は以下の式で定義される．1 に近づくほど，2 つの順序関係の差異が少ないことを示す．

$$\mathrm{nDCG} = \frac{\sum_{r=1}^{n} \frac{2^{f(r)}-1}{\log_2(r+1)}}{\sum_{r=1}^{n} \frac{2^{y(r)}-1}{\log_2(r+1)}}$$

ここで r は順位，$f(r)$ は r 位の推定モデルによる言及数，$y(r)$ は r 位における真値の言及数である．

精度評価結果は**表 3** に示すとおりであり，MAE が 7 前後，RMSE が 10 前後となっており言及数の推定誤差は無視できないが，相関係数で見ると 0.60 と比較的に相関が見られており，施設人気度での順序付けの差異についても nDCG が 0.68 となっていることから，言及数に基づく施設人気度を用いて提示順の最適化を行う利用方法を想定すれば十分な精度であるといえる．

また，**図 2** に POI 言及数（真値）と推定した POI 言及数との関係を散布図としてプロットした結果を示す．推定した POI 言及数が 40 以下の POI と，40 より大きい POI とで分布の傾向が違うように見えるが，これは POI のジャンルにより分布傾向に差があることが原因と考えられる．駅やテレビ塔，遊園地などのランドマーク系の POI（分布の範囲はおおむね 20〜80）と，飲食店や商店などのランドマーク系以外の POI（分布の範囲はおおむね 7〜40）とで分布の傾向が異なっており，ランドマーク系の POI は真値と推定結果が線形に比例しているように見受けられる一

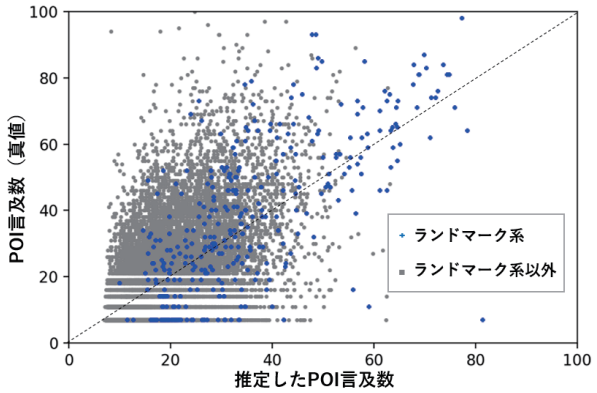

図 2 真値と推定言及ツイート数の散布図
Fig. 2 Scatter plot of ground truth and estimated number of referring tweets.

方で，ランドマーク系以外の POI では推定結果が 40 以下の範囲に集中している．飲食店などの POI はあるエリア内で場所が近接して立地していることが多く，特徴量として利用している平均滞在人口などの値も似たような範囲の数値をとることから，推定結果が一定の範囲に集中してしまったと考えられる．また，全体的に推定した POI 言及数は真値よりも過少になっているように見えるが，これも同様にランドマーク系以外の POI が POI データ全体のうちの大多数を占めていることから，推定結果が 40 以下の範囲に偏って集中することで発生していると考えられる．ランドマーク系の POI に限定すれば推定結果が過少に偏るような傾向は見られなかった．

6. POI 検索精度評価

施設人気度を用いることによる POI 検索精度の向上度合いを確認するため，以下の 2 つのデータセットで検索精度を評価した．全文検索エンジンとしては Apache Solr 6.2.2 を用い，検索結果の上位 1 件までに正解とする POI が含まれる割合，および上位 3 件までに含まれる割合の 2 つで評価を行った．

- 観光施設検索ログテストセット（1,071 件）
- カーナビ目的地検索ログテストセット（3,500 件）

観光施設検索ログテストセットは主に観光施設に関する検索クエリであり，地図データや検索サービスを提供している会社が保有する地図サービスでの検索ログに基づき，検索対象を観光施設に限定したデータセットであり，検索クエリに対応する正解とすべき POI については以下に示す判断基準で目視で付与している．

1) クエリ内の文字列と POI の名称が完全一致する場合は該当の POI を正解とする．同一名称の POI が複数存在する場合，クエリ文字列に含まれる地名や POI のジャンルなどの属性が一致する POI を正解とする．
2) クエリ内の文字列と POI の名称が完全一致はしないものの類似した名称である場合，ほかに類似した名称の POI が存在しなければ，該当の POI を正解とする．また，クエリ文字列に地名や POI のジャンルなどの属性が含まれる場合は，地名や属性が一致する POI に限定を行い，POI が一意に特定できれば，該当の POI を正解とする．
3) クエリ内の文字列が複合施設を示す場合，複合施設内のテナントを表す POI は不正解とした．この場合，複合施設自体を表す POI のみ正解とする．
4) クエリ内の文字列がチェーン店舗の名称を含む場合，チェーン店のうち具体的な店舗が特定できるクエリであるときのみ，該当の POI を正解とする．
5) クエリ内の文字列において POI の名称が省略されている，または，別称や愛称となっている場合は，本来の正式名称に該当する POI を正解とする．

また，上述の判断基準のいずれにおいても，クエリ文字列内に POI を特定するだけの十分な情報が含まれない場合や，同一の属性を持つ POI が複数存在する場合など，適切な POI を客観的に判断できない場合はテストセットから該当のクエリを除外した．施設人気度はジャンル検索でも精度改善に寄与することは期待されるが，正解とすべき POI を客観的に選定することが難しいことから，本評価においては特定 POI を検索するクエリのみで評価を行った．なお，施設を示す POI とその施設の駐車場を示す POI のように同じ敷地内に存在しているものの，POI データ自体には施設を示す POI が含まれないことがあり，この場合に限っては駐車場などの関連施設を示す POI を正解として扱った．また，一般的に 1 つの施設と見なされる場合であっても，POI データが複数存在する場合（たとえば，「JR 東京駅（京浜東北線）」と，「JR 東京駅（山手線）」は，いずれの POI も正解として扱った．

カーナビ目的地検索ログテストセットは，スマホ向けのカーナビアプリにおける目的地検索での検索ログであり，検索クエリに対応する正解とすべき POI については観光施設検索ログテストセットと同様の判断基準で目視で付与している．テストセットの例を**表 4** に示す．

観光施設検索ログテストセットで正解とすべき POI のうち SNS での言及数が集計できた POI は約 400 個で全体の約 38%，同様にカーナビ目的地検索ログテストセットで

表 4 評価で用いたテストセットの例
Table 4 Sample of test sets for evaluation.

クエリ文字列	正解とすべき POI の名称
東京国際展示場	東京ビッグサイト
免許センター　幕張	千葉運転免許センター
コストコ　幸浦	コストコホールセール 金沢シーサイド店
伊豆市長岡　かつらぎ会館	伊豆の国市長岡総合会館 AXIS かつらぎ

表 5 POI データの例
Table 5 Sample data of POI datasets.

POI-ID	名称	名称（読み）	住所	ジャンル	人気度 ($Score_{tw}$)	人気度 ($Score_{org}$)	人気度 ($Score_{mss}$)
Z0012345	清水寺	キヨミズデラ	京都府 ···	寺，仏閣	100	0	50
Z0054321	清水寺	キヨミズデラ	千葉県 ···	寺	60	0	40

は約 700 個であり全体の約 21%を占めている．残りの 7〜8 割程度の POI については推定モデルを利用することで推定言及数による施設人気度が付与できたことになる．

6.1 施設人気度による精度向上

全文検索エンジンでは，検索対象となる POI データの属性（フィールド），具体的には POI 名称や施設カテゴリ，住所などのテキストを形態素分割し，形態素ごとに，その形態素を含む POI の一覧を辞書（インデックス）化しておき，入力された検索クエリ文字列も同様に形態素に分割して辞書を検索することで，クエリ内の形態素を含む POI を高速に抽出することが可能である．抽出された検索結果の各 POI に対し，各フィールド内の形態素のマッチング度合いに応じて検索スコアが計算され，そのスコアに基づいて検索結果をソートして最終的な検索結果が得られる．

検索結果の表示順を決定するための検索スコアにおいて，検索ロジックにより算出された検索スコアをそのまま用いる既存手法（式 (1)，(2)）と，施設人気度に一定の重みを掛けた値を検索スコアに加算する提案手法（式 (3)）とで比較を行った．

$$k_{i,max} = \arg\max_{k} f(P_i W_k) \tag{1}$$

$$Score_{i,query} = f(P_i W_{k_{i,max}}) + \sum_{k \neq k_{i,max}} 0.01 \cdot f(P_i W_k) \tag{2}$$

$$Score_{i,popularity} = Score_{i,query} + a_1 Score_{i,tw} + a_2 Score_{i,org} + a_3 Score_{i,mss} \tag{3}$$

ここで P_i は POI，w_k は入力クエリ文字列と比較する k 番目のフィールド（属性種別）であり，今回の場合は POI の名称，名称（読み），住所，ジャンルを示し，$f(x, y)$ は POI x におけるフィールド y の類似度を示すスコアを返す関数である．また，$Score_{i,tw}$ は 4.1，4.2 節，および 5 章で述べた手法による言及数に対し自然対数をとった値，$Score_{i,org}$ は 4.2 節で述べた手法による言及数に対し自然対数をとった値，$Score_{i,mss}$ は POI 周辺の昼間時間帯における平均滞在人口の値である．なお，検索スコア $Score_{i,query}$ を算出する際の検索ロジックは提案手法でも同等としており，検索クエリの解釈やマッチング処理の対象となるフィールドは同一としている．

Solr でのスコア計算ロジックはクエリ文字列との類似度を示す Similarity Model として，TF-IDF に基づいた ClassicSimilarity と，TF-IDF を拡張した手法である BM25Similarity が提供されているが，本評価においては，より検索精度が良好であった ClassicSimilarity モデルを用いている．また，検索ロジックによる検索スコア算出には POI データの各属性情報のフィールドを横断検索するため，DisjunctionMaxQuery パーサを用いており，既存手法，提案手法のいずれも重みの値以外は同等である．Solr を含む一般的な全文検索エンジンでは検索速度向上のために，全ドキュメント（POI データ）にスコア付けを行うのではなく，検索結果候補となるドキュメントを抽出した後に，検索結果をソートするためのスコア算出を行うという 2 段階構成となっている．

具体的なスコア算出の流れを，検索クエリとして「京都市清水寺」の文字列を与えた場合を例にして説明する．検索対象の POI データは**表 5** であるとすると，検索クエリに対する各フィールド（名称，名称（読み），住所，ジャンル）との類似度 $f(P_i W_k)$ が計算される．POI-ID が Z0012345 の POI については，「名称」や「名称（読み）」，「住所」フィールドにおいて検索クエリとマッチする形態素を含むため，マッチする形態素の TF-IDF 値が類似度として算出される．今回のデータセットにおいては，「名称」フィールドにおける「清水寺」の形態素は 10.73 の類似度となった．同様に，POI-ID が Z0054321 の POI にも，「名称」や「名称（読み）」フィールドにおいて類似度が算出される．なお，この 2 つの POI は「名称」や「名称（読み）」が完全に同一のため，それぞれ算出される TF-IDF 値は 2 つの POI の同一フィールドどうしでは同じ値となる．また，フィールドごとに重みパラメータがあり，TF-IDF 値に対し重みが掛け合わされた結果が該当フィールドの類似度となる．各フィールドに対する類似度が算出できたら，類似度が最も大きいフィールド $k_{i,max}$ の類似度 $f(P_i W_{k_{i,max}})$ に対し，その他のフィールドの類似度に一定の重み（0.01）を掛けて足しこんだ値（$\sum_{k \neq k_{i,max}} 0.01 \cdot f(P_i W_k)$）を加算して最終的な検索スコア $Score_{i,query}$ として出力する．「名称」，「名称（読み）」，「住所」のいずれも検索クエリに含まれる形態素とマッチする形態素は 1 つであるが，フィールドごとに重みのパラメータが掛け合わされるため，このパラメータの大小関係により，どのフィールドが最大値となるかが影響を受ける．

また，提案手法における施設人気度を考慮した検

表 6 評価で用いた重み値
Table 6 Weight values uesd for evaluation.

a_1	a_2	a_3
1,325	50.5	1

表 7 観光施設検索ログテストセットにおける検索ヒット率
Table 7 Results of search hit rate for sightseeing test sets.

	P@1	P@3
施設人気度なし	60.0% （SNS 言及数 POI）28.8% （推定言及数 POI）31.2%	74.2% （SNS 言及数 POI）32.0% （推定言及数 POI）42.2%
施設人気度あり	73.3% （SNS 言及数 POI）34.0% （推定言及数 POI）39.3%	82.4% （SNS 言及数 POI）35.6% （推定言及数 POI）46.8%

索スコアは，検索ロジックにおいて算出された検索スコア（$Score_{i,query}$）に対して，式 (3) で示すとおり施設人気度に重みのパラメータを掛けた値を加算して得られる．POI-ID が Z0012345 の POI については，$Score_{i=Z0012345,query} + a_1 100 + a_2 0 + a_3 50$ のように算出される．施設人気度は生成手法ごとに個別に保持しており，検索スコアで加算される際には，種別ごとに異なる重みのパラメータと掛け合わされる．このパラメータを調整することで，POI 周辺の滞在人数を重視した人気度とするか，SNS 上での話題性を重視した人気度とするかを変化させることができる．

式 (3) における重み（a_1, a_2, a_3）は，テストセットとは別に，POI データのうち有名な POI に対して目視で作成した学習用テストセット（2,244 件）を用いて，正解の POI が検索結果上位となるスコア（$Score_{i,popularity}$）になるようパラメータ探索した結果を用いており，今回の評価においては a_3 を 1 としたときの相対値で示すと**表 6** であった．さらに，検索スコア（$Score_{i,query}$）の算出においてフィールドごとに形態素が一致した際に加算されるスコアにも重みが掛けられており，パラメータ探索の際にあわせてチューニングされている．

施設人気度を用いずに検索スコアのみを利用する場合においても，フィールドごとに形態素が一致した際に加算されるスコアの重みは，同一の学習用テストセットを用いてパラメータ探索した結果を用いており，施設人気度なしの状態で最も検索精度が高い状態としている．

観光施設検索ログテストセットでの結果を**表 7** に，カーナビ目的地検索ログテストセットでの結果を**表 8** に示す．P@1 は検索結果上位 1 位に正解とすべき POI が提示される割合，P@3 は検索結果上位 3 位までに正解とすべき POI が最低でも 1 件は提示される割合を意味する．なお，SNS での言及数がない場合に推定した言及数を用いたことによる効果を確認するため，正解とすべき POI が，SNS 言及数がある POI（SNS 言及数 POI）と推定言及数を付与した POI（推定言及数 POI）の場合とで検索ヒット率の内訳

表 8 カーナビ目的地検索ログテストセットにおける検索ヒット率
Table 8 Results of search hit rate for car navigation test sets.

	P@1	P@3
施設人気度なし	48.0% （SNS 言及数 POI）13.0% （推定言及数 POI）35.0%	63.1% （SNS 言及数 POI）17.0% （推定言及数 POI）46.1%
施設人気度あり	61.3% （SNS 言及数 POI）18.2% （推定言及数 POI）43.7%	70.6% （SNS 言及数 POI）19.6% （推定言及数 POI）50.7%

もあわせて示す．既存研究では SNS での言及数を利用した施設人気度が提案されているが，SNS 言及数がある POI のみでの検索ヒット率はあまり高くないことが分かる．いずれのテストセットにおいても，検索スコアに対し施設人気度を加算した場合に検索精度の向上がみられており，検索結果の表示順序を決める際に施設人気度を用いることでより適切な POI を上位に表示できていることが分かる．

2 つのテストセットでの結果を比較すると，P@1 や P@3 での精度向上は同程度に見えるが，その内訳をみると，カーナビ目的地検索ログテストセットの方が正解とすべき POI における SNS 言及数 POI の占める割合が低くなっている．観光施設検索ログテストセットでは，観光施設に関するクエリに限定していることから，正解とすべき POI が SNS で言及されやすい一方で，カーナビ目的地検索ログテストセットでは，観光施設以外に商業施設や駐車場などの検索クエリが含まれており，正解とすべき POI が観光施設の場合と比較して SNS で言及されにくいことに起因している．したがって，SNS での言及数に加えて推定言及数を付与するという提案手法であれば，観光施設以外の言及されにくい施設であっても，より検索されやすい人気の施設を上位に表示させることができていると考えられる．

6.2 SNS と人口統計を用いた施設人気度の妥当性

施設人気度として，提案手法では Twitter の本文における POI 名称などの言及数，および，周辺の人口分布を用いている．一方で，カーナビにおける検索ログでの対象 POI の検索回数についても人気が高い POI であれば検索回数が増えることが想定されるため，施設人気度と相関があると考えられる．

ここでは，施設人気度として，Twitter での言及数を用いる場合，平均滞在人口を用いる場合，検索ログの検索回数を用いる場合，検索回数と平均滞在人口を組み合わせた場合，Twitter での言及数と平均滞在人口を組み合わせた場合とで比較を行った．既存研究では POI レコメンドのためにチェックイン投稿からチェックイン確率を算出し利用しているが，本研究においては POI 検索結果を最適化するため，検索回数と平均滞在人口を組み合わせた場合での評価を実施した．検索クエリとして，カーナビでの目的地検索ログログテストセットを用いた場合の評価結果を**表 9**

表 9 人口分布による人気度と検索回数による人気度での検索ヒット率比較

Table 9 Results of search hit rate compared with population based popularity and search query count based popularity.

	P@1	P@3
平均滞在人口	53.1%	66.2%
検索回数	56.4%	68.1%
検索回数（ジャンル別）＋平均滞在人口	56.8%	68.0%
Twitter 言及数	60.0%	69.5%
Twitter 言及数＋平均滞在人口	61.3%	70.6%

に示す．この結果より，平均滞在人口や過去の検索回数を用いる手法と比較し，Twitter の言及数を用いた場合の方がより適切な POI を上位に表示できている．また，提案手法である，Twitter での言及数と平均滞在人口を組み合わせた場合が最も精度が高くなっており，施設人気度として妥当であると考えられる．

一方で，P@3 の精度からは，観光施設検索ログテストセットの場合で 2 割程度，カーナビ目的地検索ログテストセットでは 3 割程度の検索クエリにおいては適切な POI を上位 3 件以内に提示できていないことになる．一般的な全文検索エンジンでは，クエリ文字列に対する検索結果候補となる POI を抽出し，その後に検索結果候補の POI に対して検索スコアを算出して上位の結果から提示を行っている．この点を考慮すると，適切な検索ができなかった原因としては，おおよそ以下に大別できる．

1) 正解とすべき POI を検索結果候補として抽出できていない場合．このような事象は，正解とすべき POI の名称と検索クエリ文字列とで形態素の一致がない，または，一致数が閾値以下になると発生する．これを改善するためには，形態素が適切に分割されるよう形態素解析辞書の見直しや，略称などの類義語辞書の拡充が必要となる．
2) 正解とすべき POI を検索結果候補として抽出できているものの，検索スコアが低く，施設人気度を加味しても期待するほど順位が向上していない場合．これを改善するためには，施設人気度に対する重みのパラメータを大きくすることが考えられるが，パラメータを変更することで他の検索結果に悪影響を及ぼす可能性が高い．施設人気度の重みを大きくしすぎると，検索クエリとのマッチング度合いよりも人気度のみを考慮した順序となってしまう．本評価においては重みのパラメータは，学習用テストセットでの検索精度が最も高くなる値を用いている．

適切な検索ができなかったクエリにおける検索結果を確認したところ，およそ 8 割のクエリが 1) のパターンで検索ができていないことが分かった．残りは 2) のパターンであるが，施設人気度を加算する前の検索スコア自体が低いことが根本的な原因であり，こちらも解決策としては形態素のマッチング度合いを上げるために類義語辞書などの拡充が必要となる．

7. おわりに

本研究では，POI 検索における精度改善をめざし，ソーシャルメディア上での POI に対する言及数，および，POI 周辺の滞在人口分布に基づき施設人気度を付与することで，人気の高いと考えられる施設を優先した検索を実現した．これにより，特に同名の施設において，既存手法のような距離が近い POI を優先する方法に比べ，ユーザが想定する検索結果を返却できる割合が増加したことで精度改善が図られた．また，ソーシャルメディア上での POI に対する言及が見られない場合においても，仮に言及されたとしたときの言及数を推定することで，大多数の言及数が付与できない POI に対しても施設人気度を付与することができた．

今後の課題として，現在は 1 つの POI に対して付与した施設人気度は昼夜を問わず変化しないが，実際には時間帯に応じて POI の施設人気度は変化すると考えられるため，動的に施設人気度を生成できるよう言及数の推定モデルを拡張することを検討している．

参考文献

[1] Google, Google maps, available from ⟨https://www.google.com/maps/⟩.
[2] Apache ソフトウェア財団，Apache solr，入手先 ⟨https://lucene.apache.org/solr/⟩.
[3] Elasticsearch B.V., Elasticsearch, available from ⟨https://www.elastic.co/products/elasticsearch⟩.
[4] 大西杏菜，川崎仁嗣，神山 剛，伊藤 駿，深澤佑介：ユーザの生活圏を考慮した施設情報検索の検討，Technical Report 13 (Aug. 2019).
[5] Shivendra, T. and Saroj, K.: User category based estimation of location popularity using the road gps trajectory databases, *Geoinformatica: Intl. Journal*, Vol.4, No.2 (2014).
[6] Yao, Z., Fu, Y., Liu, B., Liu, Y. and Xiong, H.: Poi recommendation: A temporal matching between poi popularity and user regularity, *2016 IEEE 16th International Conference on Data Mining* (*ICDM*), pp.549–558 (Dec. 2016).
[7] Hsieh, H.-P., Lin, F., Li, C.-T., Yen, I.E.-H. and Chen, H.-Y.: Temporal popularity prediction of locations for geographical placement of retail stores, *Knowl. Inf. Syst.*, Vol.60, No.1, pp.247–273 (2019).
[8] Ying, J.J.-C., Lu, E.H.-C., Kuo, W.-N. and Tseng, V.S.: Urban point-of-interest recommendation by mining user check-in behaviors, *Proc. ACM SIGKDD International Workshop on Urban Computing, UrbComp '12*, pp.63–70, ACM (2012)
[9] Lee, K., Ganti, R.K., Srivatsa, M. and Liu, L.: When twitter meets foursquare: Tweet location prediction using foursquare, *Proc. 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MOBIQUITOUS '14*,

pp.198–207, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2014).
[10] Chen, Y., Zhuang, C., Cao, Q. and Hui, P.: Understanding cross-site linking in online social networks, *Proc. 8th Workshop on Social Network Mining and Analysis, SNAKDD'14*, pp.6:1–6:9, ACM (2014).
[11] 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 - Volume 1, HLT '11*, pp.359–367, Association for Computational Linguistics (2011).
[12] 落合桂一，鳥居大祐：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，情報処理学会論文誌データベース（TOD），Vol.7, No.2, pp.51–60 (June 2014).
[13] Yang, Y., Duan, Y., Wang, X., Huang, Z., Xie, N. and Shen, H.T.: Hierarchical multi-clue modelling for poi popularity prediction with heterogeneous tourist information, *IEEE Trans. Knowledge and Data Engineering*, Vol.31, No.4, pp.757–768 (2019).
[14] 寺田雅之，永田智大，小林基成：モバイル空間統計における人口推計技術，NTT DOCOMO テクニカル・ジャーナル，Vol.20, No.3, 一般社団法人電気通信協会 (Oct. 2012).
[15] Li, Y., Steiner, M., Wang, L., Zhang, Z. and Bao, J.: Exploring venue popularity in foursquare, *2013 IEEE Conference on Computer Communications Workshops* (*INFOCOM WKSHPS*), pp.205–210 (2013).

川崎 仁嗣

株式会社 NTT ドコモクロステック開発部勤務．2008 年筑波大学システム情報工学研究科博士前記課程修了．同年株式会社 NTT ドコモ入社．モバイルコンピューティング，端末セキュリティ，分散システムに関する研究に従事．

深澤 佑介 （正会員）

2004 年東京大学大学院工学系研究科修士課程修了．同年株式会社 NTT ドコモ入社．2011 年東京大学大学院工学系研究科博士後期課程修了．東京大学人工物工学研究センターにて協力研究員（2011～2016 年）および客員研究員（2016～2019 年）を兼任．2019 年より早稲田大学イノベーション研究所招聘研究員を兼任，現在に至る．Web マイニング，パーソナライゼーション，確率モデルに関する研究開発を行っている．IEEE，人工知能学会各会員．博士（工学）．

豊田 正史 （正会員）

東京大学生産技術研究所教授．1994 年東京工業大学理工学部情報科学科卒業．1996 年同大学大学院情報理工学研究科修士課程修了．1999 年同大学院情報理工学研究科博士課程修了．博士（理学）．同年科学技術振興事業団計算科学技術研究員．2001 年東京大学生産技術研究所学術研究支援員，2006 年同特任助教授，2008 年同助教授，2009 年同准教授を経て，現在に至る．ウェブ，ソーシャルメディア，IoT データ等のインタラクティブな可視化・解析の研究に従事．ACM，IEEE CS，電子情報通信学会，日本ソフトウェア科学会各会員．