

ジオタグ付きツイートをを用いた観光スポットの従属関係の分析

佐藤 京也[†] 横山 昌平^{†,‡,‡‡}

[†] 東京都立大学 システムデザイン学部 〒 191-0065 東京都日野市旭ヶ丘 6-6

[‡] 国立情報学研究所 〒 101-8430 東京都千代田区一ツ橋 2-1-2

^{‡‡} 東京大学生産技術研究所 〒 153-8505 東京都目黒区駒場 4-6-1

E-mail: [†]sato-keiya@ed.tmu.ac.jp, ^{‡‡}shohei@tmu.ac.jp

あらまし 観光中に発するツイートには、そのユーザーがその時に訪れている場所に関するツイートと、そこから見える、あるいは連想できる別の場所に関するツイートがある。本研究では後者のツイートを利用して、どこの観光スポットにおいてどこの観光スポットが言及されたのかという関係性を抽出し、スポット間の従属関係を明らかにする。また、この従属関係を分析する事により、エリアにおけるそのスポットの性質を考察することができる。

キーワード ソーシャルメディア, 観光, 地理データ活用

1 はじめに

旅行計画においてメインの観光スポットを決める際、ある観光スポットにおける観光体験だけではなく、そこから見える観光スポットや周囲に存在する観光スポットの情報も観光スポット決定の要因となり得る。また、旅行ではあるエリアにおいて複数の観光スポットに訪れることが多い。

そのためにはメインの観光スポット周囲に存在する観光スポットを再度調べる必要があるが、周囲には有名でないスポットが存在することもあり、全てを把握することは難しい。津屋 [5] によると「旅行に行くために負担しなければならないコストのうちの心理的なコストは、特に旅行プラン作成に密接に関わっている」うえ、旅行中においても、ある観光スポットに訪れた観光客が、そこから見える観光スポットなど、訪れた観光スポットに関係する他の観光スポットを訪れる可能性は高いと考えられ、あるエリアにおける観光スポット間の関係性を抽出することは、観光推薦において重要であると考えられる。

また近年、Twitter¹やInstagram²などのソーシャルメディアの普及によって、ユーザーが旅行中に旅行の様子や観光地に関する情報を投稿することは一般的になった。そのような旅行中の投稿には、位置情報などの地理情報が含まれているものも多く存在する。特に、Twitterにおいて地理情報に関するメタデータが付与されたツイートはジオタグ付きツイートと呼ばれる。

ジオタグ付きツイートを利用した研究にはコロナ禍のソーシャルディスタンスの維持に関する研究 [1] などがある。

またジオタグ付きツイートは旅行や結婚などのポジティブなイベントに関するツイートが多く、ハッシュタグやそれらのイベントに関する言及が多く含まれている。[4] そのため、ツイートが呟かれた地域やスポットに関する情報として利用することができる。実際に、観光推薦に関する研究は多く存在する。例えば Tasse ら [2] は、観光地に関するジオタグ付きツイートから

クリエイティブツーリズムにおいて旅行者がどのようなツアーガイドを求めているかを分析している。また中嶋ら [3] は、ジオタグ付きツイートをを用いてユーザーの観光ルートを抽出し、それらを「食事」、「景観」、「行動」の3つに分類した。そして、それらの中からユーザーの好みに合う観光ルートを推薦する手法を提案している。

他方で、観光に関するツイートは大きく次の二種類に分類することができる。一つ目はユーザーが訪れた観光スポットに関して言及しているツイートで、二つ目は訪れた観光スポットから見える、あるいは連想される別の観光スポットに関するツイートである。

図1に前者のツイートの例を示す。このツイートでは、東京タワーに訪れたユーザーが東京タワーについてのみ言及している。

また図2に後者のツイートの例を示す。このツイートは増上寺で投稿されたものであり、ツイート中では増上寺だけではなく、そこから見える東京タワーについても言及している。またツイート内には増上寺と東京タワーの他に、芝公園を表すハッシュタグが含まれている。これは芝公園が、地理的に増上寺や東京タワーを地理的に包含する観光スポットであるためだと考えられる。

本研究では、このようなツイートに着目して観光スポット間の関係性について分析する。ある観光スポット A , B に関する内容を含むツイートにおいて、 A がツイートの主題となっている場合、 A をメインの観光スポットとすると、 B は A における観光体験の一部であると考えられる。本研究ではこのような複数の観光スポットに言及するツイートにおいて、投稿の主題となりやすい、つまりメインの観光スポットとなる観光スポットを主、もう片方の観光スポットを従とする従属関係として定義する。そのようなツイートを抽出することで、観光スポット間の関係性について分析できると考えられる。

本研究では、Twitterのツイートデータに含まれるジオタグ付きツイート中のハッシュタグを利用し、観光スポット間の従属関係を抽出する。また実際にツイートを確認し、抽出した従

1 : <https://twitter.com>

2 : <https://www.instagram.com>



図1 東京タワーに言及するツイートの例



図2 増上寺と東京タワーに言及するツイートの例

属関係が正しいことを確認した。

以下に論文の構成について記す。第3章では観光スポット間の従属関係の抽出方法について述べ、第4章では実際に東京都内の2つのエリアで従属関係を抽出し、その結果を分析する。第5章ではまとめと今後の課題について述べる。

2 関連研究

Twitterに限らず、ソーシャルメディア上の地理情報が付与されたデータはジオソーシャルデータと呼ばれる。ジオソーシャルデータを用いた観光スポットの抽出・推薦や観光スポット間の相性の分析に関する研究は多く存在する。

観光スポットの抽出や推薦の研究として、高田ら[6]はFlickerデータを利用して観光スポット集合の推薦手法を提案している。Flickerでユーザが一日に投稿した写真の位置情報に最も近い観光スポットをユーザが訪れた観光スポットとし、また池田ら[7]の体験表現ルールを用いて各スポットにおける体験表現を抽出する。次に各スポットにおける各体験の重要度を特徴ベクトルとして計算し、これらを利用してメインの観光スポットと相性の良いスポットを推薦する。また、Munenoriら[9]はジオタグ付きツイートを用いた観光スポットと最適な観光時期を特定する手法を提案している。ツイート本文を観光スポットに関するキーワードでフィルタリングし、ツイート中の位置情報からツイート数の多いスポットを観光スポットとする。また、ツイート数推移の移動平均を用いて最適な観光時期を特定している。森田ら[10]は、FlickerとTwitterから取得した位置情報付き画像を観光資源種ごとに分類し、DBSCANによってクラスタリングを行い、各クラスターにおける投稿数といいね数を比較することでカンボジアにおける穴場観光スポットを抽出している。これらの研究は単一の観光スポットや観光スポット集合を対象としており、観光スポット間の関係性については言及していない。本研究では観光スポット間の関係性を抽出している点で異なる。

また、観光スポット間の関係性に関する研究として、伏見ら[8]は、道路ネットワーク上の観光スポットを対象とした回遊性や利便性の向上を目的として、ノード間に実距離が定義されたネットワークにおいて、回遊中心性と利便中心性とと呼ばれる中心性手法の提案を行っている。回遊中心性はあるスポットから他のスポットへ向かう際に立ち寄るのが容易なスポットを表し、利便中心性は多くのスポットから立ち寄りやすいスポットを表す。この研究では道路ネットワークを対象としており、本研究ではツイートを利用している点で異なる。

3 手法

本研究では、観光スポット同士の関係性を抽出するため、ツイートの多くは、ツイート中で言及された観光スポットやそこに関連する観光スポットを表すハッシュタグが含まれている。中でも、複数の観光スポットに関するハッシュタグが含まれているツイートは、ある観光スポットから他の観光スポットに対して言及しているツイートと見なすことができる。

図2に示したツイートのように、多くのツイートにおいてハッシュタグは本文の最後に羅列して投稿されることが多い。また、それらのハッシュタグのうち、本文中で先に出現するハッシュタグは、ツイートの主題と関連度の高いものであると

考えられる。したがって、ツイート中に複数の観光スポットに関するハッシュタグが存在する場合、その中で先に出現するハッシュタグに関する観光スポットはツイート内容に最も関連する、メインの観光スポットであり、後に出現するハッシュタグに関する観光スポットはメインの観光スポットに関係する観光スポットであると考えられる。

本研究ではツイート中の観光スポットを表すハッシュタグの出現順序から、観光スポット間の従属関係を抽出する。

3.1 分析に用いるツイート

従属関係を抽出するために、データセットとしてツイート集合を用いる。まず、従属関係を抽出したいエリアを設定し、そこで呟かれたジオタグ付きツイートを収集する。その中からハッシュタグが2つ以上含まれているもののみ抽出し、これをデータセットとする。

また、次節で説明する従属関係の抽出方法においては、ツイートに含まれる位置情報データは使用していない。しかし旅行に関するジオタグ付きツイートの多くは、投稿された場所の近辺に存在する観光スポットに対して言及していると考えられるのに対し、ジオタグがついていないツイートでは、全く無関係な複数の観光スポットに対して言及している可能性がある。したがって目的の観光スポット周辺で呟かれたツイート以外はノイズとなりうるため、エリアを設定し、その中で呟かれたツイートのみ利用している。

3.2 従属関係の抽出

次に、3.1節で作成したデータセットを利用して、ハッシュタグ間の従属関係を抽出する。

一般に、ツイートに含まれるハッシュタグには、観光スポットに関係のないハッシュタグも多く存在する。そのため、まずデータセットに含まれるすべてのハッシュタグの中から、観光スポットに関するハッシュタグのみを抽出する。

次に、観光スポットに関するハッシュタグを用いて観光スポット間の従属関係の抽出を行う。本研究では、二つの観光スポットに関するハッシュタグペアのうち、どちらの観光スポットがツイート本文中で先に出現しやすいかを計算することで、従属関係の抽出を行う。ハッシュタグ X, Y をどちらも含むツイートのうち、本文中でハッシュタグ X が Y より先に出現するツイートの集合を $tweets_{X>Y}$ とする。まずデータセットに含まれるに対し、 $|tweets_{A>B}|$ と $|tweets_{B>A}|$ を計算する。次に $|tweets_{A>B}| > |tweets_{B>A}|$ の場合、つまり、ハッシュタグ A がハッシュタグ B よりも、より多く先に呟かれている場合、ハッシュタグ A が表す観光スポットを主、ハッシュタグ B が表す観光スポットを従とする従属関係を得る。また、そのようなツイートが存在しない場合は、二つの観光スポットは無関係であるとする。

4 評価

本章では東京都における東京タワー周辺のエリアと、新宿駅周辺のエリアにおいて従属関係の抽出を行う。また、実際にツ

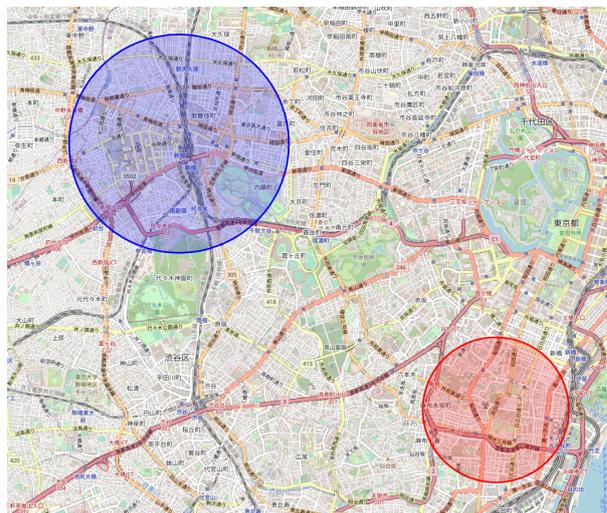


図3 従属関係の抽出を行ったエリア

表1 抽出を行ったハッシュタグのペアとそのエリア

ハッシュタグ 1	ハッシュタグ 2	エリア
東京タワー	芝公園	東京タワー
東京タワー	東京プリンスホテル	東京タワー
増上寺	東京タワー	東京タワー
増上寺	芝公園	東京タワー
歌舞伎町	花園神社	新宿駅
芝生広場	新宿中央公園	新宿駅
東京都庁	新宿中央公園	新宿駅

イートを確認することで、上記手法で抽出された従属関係が正しいことを確認する。

図3に従属関係の抽出を行ったエリアを示す。図中の2つの円のうち、赤い円が東京タワー周辺、青い円が新宿駅周辺である。東京タワー周辺エリアは緯度 35.658、経度 139.747 を中心とする半径 1km の円であり、新宿駅周辺エリアは緯度 35.691、経度 139.699 を中心とする半径 1.5km の円である。

4.1 観光スポットに関するハッシュタグの設定

一般にツイートには観光スポットに関しないハッシュタグも多く含まれているため、観光スポットに関するハッシュタグを抽出する必要がある。しかし、観光スポットに関するハッシュタグのみを自動で抽出することは困難であるため、本研究では、各エリアにおける主要な観光スポットを表すハッシュタグを手動で設定した。そして Twitter の検索機能を利用して、各ペアのハッシュタグを含む、エリア内で投稿されたツイートのうち、最新の投稿を最大で 50 件取得し、従属関係の抽出を行った。

表1に、抽出に用いたハッシュタグのペアと、そのエリアを示す。また、抽出に用いるハッシュタグペアは、各エリアにおける有名な施設名およびエリア名を表すものを選択した。

4.2 従属関係の抽出

各エリアで投稿されたツイートに対し、上記手法を用いることで観光スポット間の従属関係を抽出した。表2, 3に、それぞれ東京タワー周辺と新宿駅周辺で投稿されたツイートに対し

表 2 東京タワー周辺において抽出した従属関係

観光スポット (従)	観光スポット (主)	全ツイートに対する主の観光スポットが先に出現するツイート数
芝公園	東京タワー	29/50 tweets
東京タワー	増上寺	20/37 tweets
芝公園	増上寺	26/50 tweets
東京タワー	東京プリンスホテル	29/50 tweets

表 3 新宿駅周辺において抽出した従属関係

観光スポット (従)	観光スポット (主)	全ツイートに対する主の観光スポットが先に出現するツイート数
歌舞伎町	花園神社	28/50 tweets
芝生広場	新宿中央公園	16/22 tweets

て提案手法を用いて抽出した従属関係を示す。

ここで抽出した従属関係は、主の観光スポットをメインの観光スポットとするとき、従の観光スポットをそこで得られる観光体験の一部とする関係である。そのため、表 2 より、東京タワーや増上寺と、芝公園という観光スポットのペアでは、芝公園はもう片方の観光スポットを地理的に包含する観光スポットだが、単一ではメインの観光スポットとなりづらく、東京タワーや増上寺を内包するエリアとして抽出されたと考えられる。また増上寺や東京プリンスホテルにおいても、同様に東京タワーがホテルから綺麗に見えるという観光体験として抽出されたと考えられる。

表 3 についても、歌舞伎町は花園神社を包含するエリアであり、単体ではメインの観光スポットとなりづらいため、花園神社が主となっていると考えられる。また芝生広場と新宿中央公園においては、歌舞伎町と花園神社のように新宿中央公園が芝生広場を地理的に包含しているものの、芝生広場はそれ単体でメインの観光スポットにはならず、新宿中央公園における観光体験の一部であるため、新宿中央公園を主とする従属関係が抽出されたと考えられる。

また、「東京都庁」と「新宿中央公園」のペアでは、全 40 ツイート中、「#東京都庁」が先に出現するツイートと「#新宿中央公園」が先に出現するツイートがそれぞれ 20 ツイートずつであった。そのため新宿中央公園は、東京タワーと芝公園のように都庁がよく見える観光スポットではあるものの、新宿中央公園単体としてもメインの観光スポットになることが多いと考えられる。

5 おわりに

本論文では、訪れた観光スポットから見える、あるいは連想される別の観光スポットに関するツイートにおいて、主題となる観光スポットを主、そうでない観光スポットを従とする従属関係を定義した。ツイート中のハッシュタグは、ツイートの主題との関連度が高いものが先に現れるという仮説をもとに、観光スポットを表すハッシュタグを利用して、観光スポット間の従属関係の抽出を行った。また実際に東京都内の 2 つのエリアにおいてジオタグ付きツイートを利用し、ハッシュタグの出現順序に基づいて観光スポット間の従属関係を抽出し、それらの関係性について考察した。

今後の課題として、従属関係の詳細な分析がある。本研究では従属関係をメインの観光スポットと、観光体験の一部として存在する観光スポットとして定義したが、観光体験の一部とは、メインの観光スポットから景色として見える場合や、メインの観光スポットが存在するエリアである場合など複数の場合が存在する。これはツイート中の位置情報から、観光スポット間の距離や観光スポットの大きさなどを特定することで分離できる可能性がある。また、本研究では 2 つの観光スポット間の関係性のみ着目しているが、あるエリアにおける複数の従属関係から有向グラフを構築することで、エリア全体の性質を分析できる可能性がある。今後はこれらの課題に考慮して観光スポット間の関係性の分析を行っていく。

文 献

- [1] Xu P, Dredze M, and Broniatowski DA, "The Twitter Social Mobility Index: Measuring Social Distancing Practices With Geolocated Tweets", *J Med Internet Res* 2020;22(12):e21499, 2020.
- [2] TASSE, Dan and HONG, Jason I, "Travel Guides for Creative Tourists, Powered by Geotagged Social Media" *arXiv preprint arXiv:2112.12009*, 2021.
- [3] 中嶋勇人, 新妻弘崇, and 太田学, "位置情報付きツイートを利用した観光ルート推薦", *研究報告データベースシステム (DBS)* 2013.28 (2013): 1-6., 2013.
- [4] Julie Jiang, Jesse Thomason, Francesco Barbieri, and Emilio Ferrara, "Geolocated Social Media Posts are Happier: Understanding the Characteristics of Check-in Posts on Twitter", *arXiv preprint arXiv:2207.10887*, 2022.
- [5] 津谷篤, "人が旅をする動機の感性評定結果を用いた多様性のあるドライブ旅行プラン作成", *日本感性工学会論文誌* Vol. 10, No. 3, pp. 433-443, 2011.
- [6] 高田 盾作 and 北山 大輔, "観光スポット間の相性を用いたスポット集合推薦", 第 14 回データ工学と情報マネジメントに関するフォーラム, D41-2, 2022.
- [7] 池田佳代, 田邊勝義, 奥田英範, and 奥雅博, "Blog からの体験情報抽出", *情報処理学会論文誌*, 49.2: 838-847., 2008.
- [8] 伏見卓恭, 齊藤和巳, 武藤伸明, 池田哲夫, and 風間一洋, "実距離を考慮した中心性指標の提案と重要観光スポット抽出への応用", *人工知能学会論文誌*, 30-6..JWEIN-C., 2015.
- [9] Takahashi, Munenori and Endo, Masaki and Ohno, Shigeyoshi and Hirota, Masaharu and Ishikawa, Hiroshi, "Automatic detection method of tourist spots using SNS", *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*. p. 91-96. 2020.
- [10] 森田 格 and 荒木 光一, "シームリアップにおける Twitter・Flickr の画像・位置情報等を活用した観光資源種毎の穴場スポット抽出", *人工知能学会全国大会論文集* 第 34 回 (2020). Vol. 一般社

団法人 人工知能学会. p. 4Rin161-4Rin161., 2020.

- [11] S. Brin and L. Page, “The Anatomy of a Large-Scale Hypertextual Web Search Engine”, Seventh International World-Wide Web Conference (WWW 1998), 1998.