

口コミの言語解析による 訪日訪中外国人の訪問パターンと 観光地イメージの関係に関する研究

柴田 立¹・室町 泰徳²

¹非会員 東京工業大学 環境・社会理工学院土木・環境工学系
(〒226-8502 横浜市緑区長津田町4259)
E-mail: shibata.t.ah@m.titech.ac.jp

²正会員 東京工業大学 環境・社会理工学院土木・環境工学系
(〒226-8502 横浜市緑区長津田町4259)
E-mail: muromachi.y.aa@m.titech.ac.jp

本研究では、訪問先の国や訪問回数異なる観光客の訪問先の観光地イメージの差異を把握することを目的とし、TripAdvisorに投稿された口コミに基づき訪問パターンによる観光地イメージを分析した。具体的には、日本の221の観光地と中国の166の観光地に関する口コミを収集し、訪問年や居住地等によるクリーニングを経た、104,982件を用いて分析を行った。日本と中国それぞれへの訪問回数の推定、及び口コミを用いた観光地のクラスタリング、さらに対応分析を行ったところ、中国と日本の観光地イメージの差異や訪問回数による訪問先の観光地イメージの差異を抽出することができた。

Key Words : Text mining, Inbound Tourism, TripAdvisor, Destination Image, Correspondence Analysis

1. はじめに

(1) 背景と目的

我が国では2006年に「観光立国推進基本法」が成立し、2013年からは「観光立国実現に向けたアクション・プログラム」が毎年策定される等、観光立国推進の取り組みが盛んに行われている¹⁾。2016年には「明日の日本を支える観光ビジョン」が策定され、訪日外国人旅行者数を2020年に4000万人、2030年に6000万人とする目標を打ち出した²⁾。2011年以降、訪日外客数は増加の一途を辿り2018年には3119万人となった一方、その前年比増加率は8.7%と2015年の47.1%から減少が続いている³⁾。

この一因としてみずほ総合研究所は旅行先のシフト⁴⁾、JTB総合研究所は渡航先の多様化⁵⁾を挙げた。「明日の日本を支える観光ビジョン」では国際競争力の向上が1つの視点として据えられているが、ここでは我が国の観光産業を革新するにあたり、諸外国に対する我が国の立ち位置を明らかにした上で戦略を検討する必要がある。また、リピーターは1人当たりの旅行支出が高くなる傾向にあり⁶⁾、渡航

先が多様化する中ではこうしたリピーターの確保が重要になると考えられる。

そこで本研究では、中国と日本で個人の行動や観光地イメージの比較を行い、日本へのリピーターが注目する観光要素や他国と競合関係にある分野を抽出することを目的とし、観光客の訪問先と観光地への評価や関心の双方をユーザーの負担なく把握可能である口コミを対象とし分析を行った。

(2) 関連する研究と本研究の位置づけ

訪日観光客を対象とした研究は多く行われており、田中は日本への旅行商品を扱う企業へのアンケート等から外国人観光客の訪日促進策を提案した⁷⁾。野瀬らは日光と箱根にてアンケート調査を行い、観光地に対する評価や観光地の魅力について分析を行うなど⁸⁾、観光地イメージに関する研究を行っている。

近年ではBig Dataを用いた研究が盛んに行われており、Liらは観光研究に用いられるBig Dataをuser-generated content (UGC) data, Device data, Transaction dataに分類した⁹⁾。観光行動を把握する研究については壇辻らによるWi-Fiパケットデータを用いたDevice

dataによる研究¹⁰⁾や、HasnatらによるTwitterジオタグ付きデータを用いたUGC dataによる研究¹¹⁾等がある。一方で、観光地イメージに関する研究についてはChoiらによる旅行会社、旅行雑誌、旅行ガイド、旅行ブログでの観光地イメージの比較¹²⁾や、大久保らによる旅行ガイドブックと口コミで複数観光地について差異を検討した研究¹³⁾等があり、UGC dataが主となっている。観光地イメージや関心を抽出できることがUGC dataの特徴と言える。観光行動と観光地イメージ双方に関する研究としてはSasakiらの研究が挙げられ、国内旅行者を対象にブログの内容から訪問先を、タグから関心を抽出し分析を行った¹⁴⁾。

このようにBig Dataを用いた観光に関する研究は数多く行われているが、訪問観光地や訪問回数等の観光行動と観光地イメージ双方に着目した上で、海外観光地と比較を行った研究はほとんどない。本研究では国による観光地イメージの差異、1回のみでの訪問者とリピーターでの訪問先の差異を検討することで、今後のより効果的な観光施策につながる知見が得られることと考えている。

2. 分析対象データとクリーニング

(1) 分析手順

本研究の分析手順を図-1に示す。本研究で用いる口コミサイト（後述）は観光スポット毎で口コミを投稿するため、まずは世界的な英語版ガイドブックである「Lonely Planet」を用いて対象とする観光スポットの選定を行った。記載のある観光スポットのうちジャンルについてはSights, Activities, Courses, Toursに絞り、その中でも★印（「Must-visit recommendation」と記載）が付された観光スポットのみを対象とした。口コミの収集をした後はクリーニング（詳細は後述）を行い、これを対象として各ユーザーの日本もしくは中国への訪問回数の推定、各観光

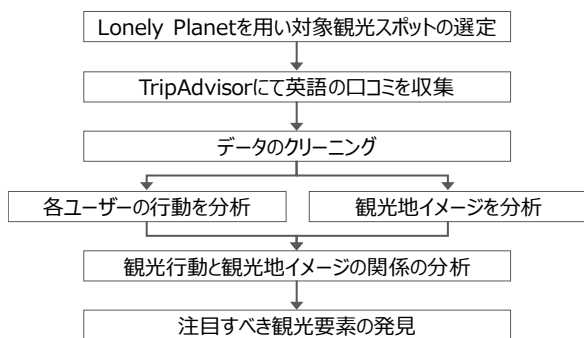


図-1 分析手順

地のイメージの分析を行う。そして観光地への評価を踏まえて観光行動と観光地イメージの関係について分析し、観光施策において重要となる知見を得る。

(2) 分析対象データ

分析には世界最大級の旅行口コミサイト「TripAdvisor」¹⁵⁾に投稿された口コミを用いる。TripAdvisorは旅行前のオンラインの情報収集源としてアメリカとオーストラリアの調査でシェアトップとなっている¹⁶⁾¹⁷⁾。観光スポット毎にページが設けられ、それぞれにユーザーが投稿することが特徴である。内容はユーザー名、居住地、タイトル、コメント、評価（5段階）、訪問時期であるが、居住地、訪問時期の入力は任意であり、記載のない口コミも存在する。TripAdvisorは先述の大久保らの研究の他、Kirilenkoらによる関心と訪問地との関係に関する研究¹⁸⁾、Taccharungrojらによる機械学習を用いた口コミ分析¹⁹⁾等、観光行動及び観光地イメージの研究に多く用いられている。

(3) クリーニング

クリーニングの内容、これによるデータ数の推移を表-1に示す。(1)各観光スポット内で全く同じ口コミを削除した後、(2)ユーザー名が「トリップアドバイザー・メンバー」となっている投稿及び(3)ユーザー名が文字化けしている投稿を除いた。(4)訪問期間は2014～2018年のみとし、それ以外に訪問したとする投稿及び投稿時間未記入の投稿を削除した。次に、いわゆる「サクラ」による投稿を除くため、(5)データ内で投稿が1回のみユーザーによる投稿、(6)訪問回数が多い、または滞在期間が長いと判断されるユーザーによる投稿を削除した。ここで、訪問時期は月ごとであり日付までは記載されないため、連続

表-1 クリーニングによるデータ数の推移

	日本	中国	合計
観光地数			
対象観光地数	227	166	393
データ収集	221	148	369
口コミ数			
(1) 同内容削除	123,667	171,993	295,660
(2) ユーザー不明削除	123,661	171,950	295,611
(3) 名前文字化け削除	123,475	171,752	295,227
(4) 2014～2018 訪問	100,767	126,822	227,589
(5) 1つ投稿者削除	76,587	88,386	164,973
(6) 異常訪問削除	74,180	85,662	159,842
(7) 居住国限定	49,457	55,525	104,982
クリーニング後割合	40.0 %	32.3 %	35.5 %
ユーザー数	14,240	17,915	28,790
1人当たり平均口コミ数	5.21	4.78	5.55

した月に訪問したとするロコミがある場合はそれを1回の訪問とみなしてクリーニングを行った。本研究での閾値は日本中国共に訪問回数6回（年1回超）、滞在期間3カ月とした。居住国については、日本への訪問者数が多い20の国・地域のうち中国、香港を除いた18の国・地域とした。これらのクリーニングの末に残った、104,982のロコミを分析に用いる。

3. 分析結果

(1) 訪問地域の分析

ユーザーの訪問先を把握するため、各地域の訪問有無データを作成した。本研究においてはLonely Planetでの中規模の地域分類（日本60地域、中国61地域）を用いる。訪問回数別の平均訪問地域数を表-2に示す。t検定（有意水準1%）を行うと、日本中国ともに訪問回数が多い場合有意に訪問地域数も多いことが分かった。日本と中国で訪問地域数の差についてt検定（有意水準1%）を行うと、訪問回数1回の場合は有意な差が見られなかったのに対し、2回以上の場合は有意に日本の方が多いことが分かった。

次に、各ユーザーの地域別訪問有無（0,1）データを用いてユーザーのクラスタリングを行った。クラスター数を7とした際のクラスタリングの結果を

表-2 平均訪問地域数

	日本	中国
訪問1回のみ	1.76	1.75
訪問2回以上	2.83	2.36

表-3 訪問地域クラスター（日本）

訪問1回のみ		訪問2回以上	
東京	17.1%	東京	10.6%
東京・京都	7.5%	東京・京都	8.6%
京都	22.0%	北海道	15.7%
大阪	11.5%	関西	6.4%
関西	8.8%	東京・大阪	9.6%
広島	9.1%	京都・広島	12.0%
その他	23.9%	その他	37.0%

表-4 訪問地域クラスター（中国）

訪問1回のみ		訪問2回以上	
北京・長城	22.4%	北京・長城	23.6%
香港	25.3%	香港	18.3%
上海	13.3%	上海	14.2%
陝西省	11.4%	陝西省	10.6%
四川省	4.6%	四川省	7.6%
マカオ	5.3%	江蘇省	6.8%
その他	17.5%	その他	18.9%

表-3、表-4に示す。日本については、訪問が2回以上になると「北海道」クラスターが生まれ、「その他」クラスターの割合が上昇していることから、日本に複数回訪れる場合は東京や京都といった大都市圏だけでなく地方にも訪れていることが分かる。一方中国については、「江蘇省」クラスターが生まれるものの、その割合は小さく、「その他」クラスターの割合の上昇も小さいことから、訪問回数による差異が日本に比べて小さいことが分かる。よって、訪問回数が増えた場合、日本は地方も訪れる傾向があるのに対して、中国は大都市を複数訪れる傾向があると考えられる。

(2) 観光地イメージの分析

各地域について観光地イメージを把握するため、ロコミ内容を用いて地域のクラスタリングを行う。分析手順は以下の通りである。

- ① KH Coder²⁰を用いて形態素解析を行い、各ロコミでの各語使用回数（名詞形容詞動詞200語）を出力する。
- ② 名詞動詞同形の単語はまとめる。
- ③ 国ごとで集計し、使用回数が他方の8倍以上の語（“peak”等特定の地域を指す語）を除く。
- ④ 観光スポットごとで集計し、各語についてそれぞれスポットの総語数で除した割合を計算する。
- ⑤ その割合から変動係数を計算し1.5未満の語（“place”のような特徴を表さない語）を除く。
- ⑥ 各ロコミ（最高評価のみ）での各語使用回数（名詞形容詞動詞200語）を出力する。
- ⑦ 地域ごとに集計し、各語についてそれぞれの地域でのロコミの総語数で除した割合を計算する。
- ⑧ ロコミ数が100以上の地域を対象に、Ward法を用いてクラスタリングを行う。

分析結果を図-2に示す。

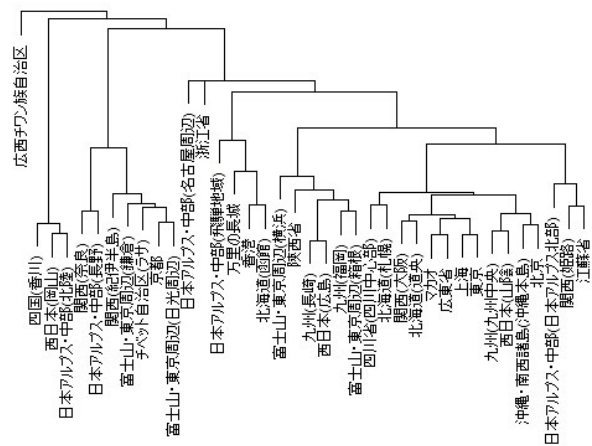


図-2 観光地のクラスタリング

京都・日光・鎌倉、香港・函館、長崎・広島など、要素の近い観光地の組み合わせが抽出できていることが分かる。(1)での訪問地域クラスターと照らし合わせると、京都は左側に、万里の長城や香港が中程に、東京大阪や上海北京が右側にあることから、左側は日本、中程は中国、右側は両国で人気がある地域であることが分かる。クラスター数を11とした際のクラスタリング結果を表-5に示す。

(3) 訪問回数と観光地イメージの関係

訪問回数と訪問先の観光地イメージの関係を見るため、対応分析を行う。ここでは日本1回訪問者(J1)、日本複数回訪問者(J2)、中国1回訪問者(C1)、中国複数回訪問者(C2)、それらを組み合わせた8つの分類(J1C2であれば日本に1回、中国に複数回訪れたユーザー)それぞれについて、11の観光地クラスターに属する地域に最高評価を付けたユーザー数のデータを用いる。結果を図-3に示す。

Dimension1は国軸(正が中国、負が日本)、Dimension2は訪問回数軸と読み取ることができる。ここから、寺院や庭園といった観光要素は日本で人気がある一方、自然や景観の名所といった要素は中

表-5 観光地クラスター

クラスター名	地域数	該当地域
寺院・宮殿	5	京都, 鎌倉, ラサ等
都市	12	東京, 札幌, 四川省等
寺院	2	奈良, 長野
歴史建築	3	松本, 姫路, 江蘇省
博物館	6	広島, 箱根, 陝西省等
自然	1	広西チワン族自治区
人気景観	3	函館, 香港, 長城
庭園	3	岡山, 香川, 北陸
地方ツアー	1	飛騨
鉄道・博物館	1	名古屋
湖	1	浙江省

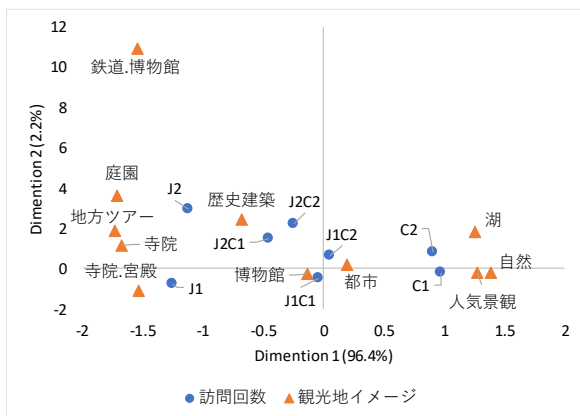


図-3 対応分析結果

国で人気があることが分かる。また、日本は中国に比べ1回訪問者とリピーターとで訪問先に差があり、Dimension2の大きい観光地は「都市」や「博物館」などに対し「歴史建築」、「庭園」など該当地域が少なく、より限定された個人の嗜好に合う観光地であることが分かる。よって、訪日客のリピーターを増やすには、1回目に都市などの大衆向けの観光地を訪れた後に個人の嗜好を考慮した観光地の提案を行っていくことが重要になると考えられる。

(4) 評価と関連のある語の抽出

以上から都市部の観光地は日本中国両国に存在し、かつ人気があることが明らかとなったが、次に、同じイメージの都市部の観光地であっても差別化を図るためにはどのような要素が重要であるのかを、口コミの内容に加えてユーザーによる評価を用いて抽出する。ここでは東京と上海を取り上げ、それぞれの地域に対する口コミを用い、各語の使用有無により評価に有意な差が見られるかt検定(有意水準0.01%)を行う。なお、ここで用いる語は(2)で用いた語のうち、東京と上海の少なくとも一方で全ての口コミの総単語数に占める割合が0.05%以上である語である。この結果より、以下の単語を抽出する。

- i. 東京でその語が使用されると有意に評価が高いが、上海ではその逆
 - ii. 上海でその語が使用されると有意に評価が高いが、東京ではその逆
 - iii. 東京でのみその語の使用回数が多く、その語が使用されると有意に評価が高い
 - iv. 東京でのみその語の使用回数が多く、その語が使用されると有意に評価が低い
 - v. 上海でのみその語の使用回数が多く、その語が使用されると有意に評価が高い
 - vi. 上海でのみその語の使用回数が多く、その語が使用されると有意に評価が低い
- iii以降での使用回数の多寡は総単語数に占める割合が0.05%以上であるかどうかを基準としている。分類結果を表-6に示す。ii, ivに該当する単語はなかった。

表-6 評価と関連のある語

i	iii	v	vi
garden	fee	world	queue
food	easy	architecture	price
	peaceful	boat	better
	traditional	river	wait
		display	
		evening	
		spectacular	

i, iiiより東京の評価の要因を見る。共起関係を見ると、これらの単語は「食」、「伝統」、「静かな庭園」、「アクセス」、「安い入場料」に分けられ、食の魅力、神社における日本の伝統的な結婚式、庭園における落ち着いた雰囲気、鉄道駅から観光スポットへのアクセシビリティ、新宿御苑の安い入場料が高評価に結びついていることが分かった。

次にi, v, viより上海の評価の要因を見る。まず上海独自の要素で高評価の要因であるvについて共起関係を見ると、ほとんどが外灘 (The Bund) に関する語であった。東京にはあまりない建築や船等による景観的要素を多く持つ人気の高い観光地となっていることが分かる。"display"は上海博物館、"world"は上海ワールド・フィナンシャル・センターで多く用いられ、こうした観光スポットも高評価の要因となっていると考えられる。一方、viより低評価の要因を見ると、「混雑」が課題となっていることが分かる。観光スポットとの共起関係を見ると豫園、東方明珠電視塔、上海ディズニーランドで混雑に関する低評価のロコミが多く見られた。iに含まれる"garden"は豫園、"food"は上海ディズニーランドと共起関係にあり、混雑による低評価の影響を受けているものと考えられる。"price"については特定の観光スポットとの関係は見られなかったが、価格に不満を持つ観光客が一定数存在することが示唆された。

以上より、東京での伝統的な要素、上海での外灘の景観等、東京上海ともに評価が高い独自の要素を有している一方で、上海では複数の観光スポットにおいて混雑が課題となっていることが分かる。

4. 結論と今後の課題

本研究では、日中間で個人の行動や観光地イメージの比較を行い、日本へのリピーターが注目する観光要素や中国と競合関係にある要素を抽出することを目的として、ロコミデータを用いて観光行動と観光地イメージを分析した。

本研究の主な結論としては、

- (1) 訪日観光客は訪中観光客に対し、複数回訪れた場合の訪問地域数の増加が大きく、地方部への訪問も増加することが分かった。
- (2) 観光地イメージについて分析を行った結果、都市部は日中間でイメージが類似していたのに対し、地方部は独自のクラスターを形成するなど傾向に相違があることが分かった。
- (3) 訪問回数と訪問先の観光地イメージの関係から、

日本に複数回訪れた観光客はより限定された個人の嗜好に合う観光地を訪問する傾向にあることが示唆された。

- (4) 観光地イメージの類似している東京と上海について評価と関連する単語を抽出したところ、東京上海ともに評価が高い独自の要素を有している一方で、上海では観光スポットの混雑が課題となっていることが分かった。

本研究の主な今後の課題として、地域の観光地イメージが分析に用いた観光スポットに大きく左右されてしまっていることが挙げられる。例えば名古屋では「名古屋城」が含まれておらず「リニア・鉄道館」の影響により独自の「鉄道・博物館」クラスターを形成しており、分析に用いる観光スポットの選定に課題が残る。また、評価と関連のある語に関する検討ではロコミでの5段階評価のみを用いてロコミ単位の単語数に依存した分析となっており、分析単位の妥当性に疑問が残る。今後は感情分析を用いて文単位での分析を行う必要がある。最後に、低評価のロコミの数の少なさも課題であり、前述の観光地の問題と同様、データを増やして低評価のロコミに関する分析を行うことが必要である。

参考文献

- 1) 観光庁：観光立国推進基本法，
<http://www.mlit.go.jp/kankocho/kankorikkoku/index.html>，2019年9月25日アクセス
- 2) 明日の日本を支える観光ビジョン構想会議 (2016)：明日の日本を支える観光ビジョン，
<https://www.mlit.go.jp/common/001126598.pdf>，2019年9月25日アクセス
- 3) 日本政府観光局：日本の観光統計データ，
<https://statistics.jnto.go.jp/graph/#graph--inbound--travelers--transition>，2019年9月25日アクセス
- 4) みずほ総合研究所 (2019)：2020年のインバウンド客数目標達成に黄信号，
<https://www.mizuho-ri.co.jp/publication/research/pdf/insight/jp190708.pdf>，2019年9月25日アクセス
- 5) JTB総合研究所 (2019)：訪日外国人動向2019 - 観光統計 - JTB総合研究所，
<https://www.tourism.jp/tourism-database/stats/inbound/>，2019年9月25日アクセス
- 6) 観光庁 (2018)：訪日外国人旅行者の訪日回数と消費動向の関係について，
https://www.mlit.go.jp/kankocho/news02_000346.html，2019年9月25日アクセス
- 7) 田中賢二(2007)，外国人観光客の訪日促進策に関する研究，運輸政策研究，Vol.10, No.1, pp.11-21
- 8) 野瀬元子・古屋秀樹(2008)：日光，箱根における外国人観光者と日本人観光者の評価特性分析，都市計画論文集，No.43 (3)，pp595-600

- 9) Jingjing Li, Lizhi Xu, Ling Tang, Shouyang Wang, Ling Li (2018) : Big data in tourism research: A literature review, *Tourism Management*, Volume 68, pp.301-323
- 10) 壇辻貴生・杉下佳辰・福田大輔・浅野光行 (2017) : Wi-Fiパケットデータを用いた観光客の滞在時間特性把握の可能性に関する研究, *都市計画論文集*, Vol.52, No.3, pp.247-254
- 11) Md Mehedi Hasnat, Samiul Hasan (2018) : Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data, *Transportation, Research Part C: Emerging Technologies*, Volume 96, pp.38-54
- 12) Soojin Choi, Xinran Y.Lehto, Alastair M.Morrison (2007) : Destination image representation on the web: Content analysis of Macau travel related websites, *Tourism Management*, Volume 28, Issue 1, Pages 118-129
- 13) 大久保立樹, 室町泰徳 (2014) : 旅行ガイドブックと口コミの言語解析による訪日外国人の観光地イメージに関する研究, *都市計画論文集*, Vol.49, No.3, pp.573-578
- 14) Kuniaki Sasaki, Kazuo Nishii (2012) : Study of Blog Mining for Examination of Tourist Travel Behavior in Japan: *Transportation Research Record Journal of the Transportation Research Board* Volume 2285 Issue 1, pp.119-125
- 15) Trip Advisor, <https://www.tripadvisor.jp/>, 2019年9月25日アクセス
- 16) ブルームーン・マーケティング株式会社 (2017) : アメリカ人1,000人に聞いた『海外旅行をする時に情報収集するメディア』ランキング, <http://bluemoonmarketing.jp/2017/09/newsrelease16/>, 2019年9月25日アクセス
- 17) JTB , オーストラリアからの訪日旅行に関する調査研究2017 (2019) : <https://www.jtb.co.jp/inbound/market/2019/australia-report-2017-2/>, 2019年9月25日アクセス
- 18) Andrei P.Kirilenko, Svetlana O. Stepchenkova, Juan M.Hernandez (2019) : Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews, *Tourism Management*, Volume 72, pp.400-410
- 19) Viriya Taecharungroj, Boonyanit Mathayomchan (2019) : Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand, *Tourism Management*, Volume 75, pp.550-568
- 20) KH Coder: 計量テキスト分析・テキストマイニングのためのフリーソフトウェア, <http://kncoder.net/>, 2019年9月30日アクセス

(2019.10.04受付)

THE RELATIONSHIP BETWEEN TOURISTS' VISITING PLACES AND DESTINATION IMAGES : REVIEW SITE ANALYSIS

Tatsuru SHIBATA and Yasunori MUROMACHI