テキストマイニングを用いた UGC[1] データにみる観光イメージ

――香川県に関する Trip Advisor[2] の和文レビューを事例に――

金　　徳　　謙

（受付　2020年 10 月 26 日）

Ⅰ　は　じ　め　に

近年，観光研究においてインターネット上の Massive Data[3] が注目され，関連研究の増加が著しい。Zeng & Gerritsen（2014）は観光研究における Social Media による貢献の拡大を，また Költringer & Dickinger（2015）や Lu & Stepchenkova（2015）は UGC の重要性を，そして Li et al.（2017）はビッグデータが観光研究に重大な影響を与えたことを，指摘した。また，Li et al.（2017）は，Social Media を含むインターネット上の情報の増加が顧客サービスの改善に役立つと指摘している。他方で，Bucur, C.（2015）は，増え続ける膨大な量のデータが分析をより困難にしていると指摘した上で，膨大な量のデータを集約するため，オンラインプラットホーム構築の必要性を提案している。金（2019）は，日本の観光研究においてもインターネットを介する研究の有効性を論じる初期段階から，インターネットを研究の手段や目的とする発展段階に移っていると指摘している。

このように観光研究においても，インターネットを研究の対象または手段として用いる研究の蓄積が進んでいる。

筆者は，地域における観光情報の発信は観光客のニーズが反映されておらず，地域の要望や希望に寄りついてなされる傾向があることに着目した。インターネット上の観光客に投稿された膨大な量の評価や口コミなどの情報，いわゆる UGC データを収集し分析することで，

1）　User Generated Contents の略で，インターネット上でサイトの利用者により投稿された書き込みや映像，画像の集積を指し，利用者の意識やニーズ，行動の分析に利用されるようになった。

2）　Trip Advisor は2000年に設立され，観光地やレストラン，ホテルなど観光にかかわる各種施設などを対象にユーザー（観光客）によるレビューが大量に投稿されており，観光に必要な情報収集に利用される人気サイトである。現在，世界中で 3 億1,500万人を超える利用者をもつ世界最大の観光情報提供サイトである。

3）　インターネット上で収集する膨大な量のデータを指し，ビッグデータ（Big data）と混用されている。インターネット上から収集しデータを分析する Data Science 分野への注目により Big Data の概念が登場するが，その後大きさを意味する big に替わり，「多い」を意味する massive という表現が登場するようになった。本稿では後者の massive を用いるが，引用においては原文のままの表現を用いた。

観光客の地域に対する観光イメージが解明でき，また，これにより観光客のニーズに合った観光情報の提供が可能になる。

そこで，本研究では2010年から 3 年ごとに芸術祭を開催し国内外からの観光客が増加している香川県を事例に，世界最大の観光情報サイト Trip Advisor に掲載されたレビューデータをもとに地域の観光イメージを明らかにすることを目的に Web Scraping 技法を用いて掲載されているレビューデータを収集し，テキストマイニング[4] 手法を用いて観光イメージの分析を行う。

Ⅱ　研究のフレームワーク

本稿では，図 1 のとおり，データの収集から分析までを 3 段階に分け，進めていく。

第 1 段階は分析に必要なデータを収集する段階であり，分析対象とするサイトからユーザーが投稿した大量のレビューデータを，Web Scraping 技法[5] を用いて収集する段階である。第 2 段階は収集したデータを分析に利用できるよう，形式を整える作業などを行う分析の前処理，いわゆるデータクリーニングを行う段階である[6]。第 3 段階は，前段階で整形したデータを用いてテキストマイニングを行う段階である。

なお，第 3 段階における分析は，次のように 3 段階で構成されている。まず，頻出語の傾向を分析するために形態素解析（図 1 －③），次に，外部変数と形態素との関係性の解明に対応分析（図 1 －④），最後に，頻出語同士の利用傾向の解明に共起ネットワーク分析（図 1 －⑤）を行う。

Ⅲ　デ ー タ 収 集

本章ではインターネット上のユーザーによる書き込みが現在もつづいている UGC データを効率よく収集するため，Web Scraping 技法を用いて収集する範囲や対象，具体的な事項について検討し，収集したデータを概観する。

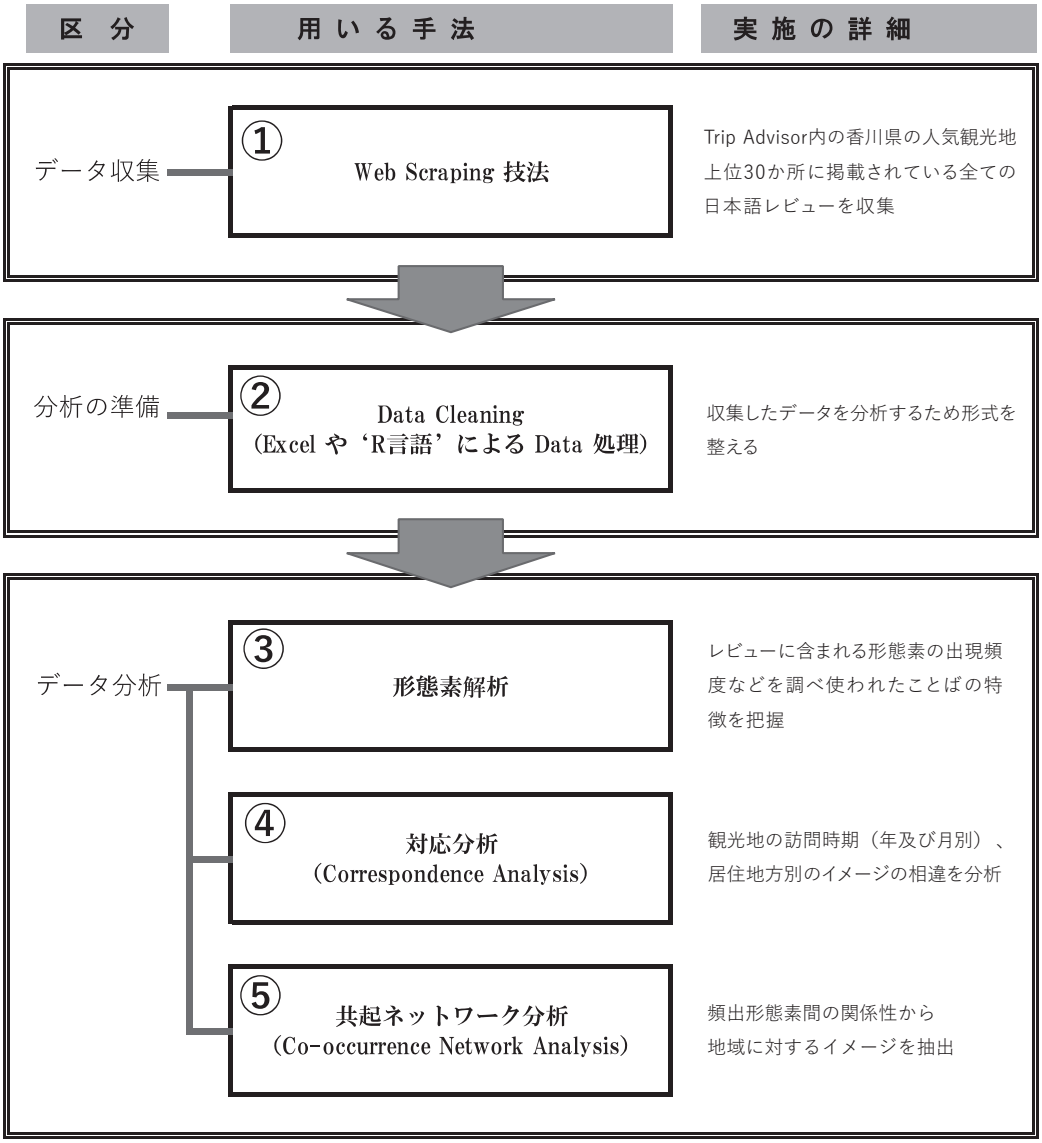
---

4)　テキストマイニングとは，大量のテキストデータから有益な情報を抽出することの総称で，具体的には自然言語処理手法により文書を分割（形態素化）し，形態素の出現頻度や形態素の間の関係性を分析する手法である。本稿ではテキストマイニングにフリーソフトの KH コーダーを用いる。生田（2016）は KH コーダーをツールに学生が作成した文書の分析により学生の感情把握を行った。他にも日本語のテキストマイニングに多用されている。

5)　インターネットの Web Site の制作には HTML 形式に加え，多くの情報を効率よく提供するため，従来の HTML より複雑な構造をもつ XML 形式，HTML5，Java Script などが用いられている。近年は多様化する情報を効率よく提供するため様々な形式が複合的に用いられている。一方で，これらに関する知識があれば，効率よく情報収集することができる。本稿ではこの手法を用いて https://www.tripadvisor.jp 内の大量のレビューデータを収集する。なお，今回の Web Scraping に必要なコードの作成は金（2019）をもとに適宜修正，改良を加えて行った。

6)　データクリーニングには，Web Scraping 技法に用いた「R」言語および Excel を適宜利用した。

図1　研究のフレームワーク



・収集対象

本研究では，世界最大規模の観光情報提供サイトTrip Advisorに掲載されている香川県の観光スポットの上位30件を収集の対象として[7)]，日本語によるレビューをすべて収集する。

7）順位はTrip Advisor社によるものである。レビューの件数や内容などが順位に影響していると考えられるが，一般的な人気順位と大きく異なるものとはいえない。そのため，検索結果を本稿では知名度の高い観光スポットと捉え，サイトを検索すると表示される上位30件をデータ収集の対象と ↗

具体的にはTripAdvisorのトップページから「香川県」をキーワードとして検索し[8)]，データ収集を行った。

・収集方法および内容

香川県を訪れた観光客によるレビューの投稿がつづいているため，UGCの量は現在も増加中である。これらのデータを収集し，分析，応用することで観光全般への好い影響を与えることができるが（Bucur, C.，2015），その一方，手作業で収集することはほぼ不可能であり（金，2019），収集方法に課題が残る。

本稿では，金（2019）が指摘した課題をWeb Scraping技法を用いて解決し，TripAdvisorに掲載されているレビューデータの収集を行った。

また本稿では，主にレビューの内容に焦点を当てているため，収集する項目をレビュー文に加え，投稿者の「居住地」[9)]，「訪問年月」に限定し，対象となる観光スポット30箇所の日本語によるレビューデータをすべて収集した[10)]。データの収集に用いたWeb Scraping技法は，金（2019）のコードを借用し，必要な個所を適宜修正加筆した。収集を行った結果，7,164件のデータとなった。

Trip Advisorは，いまや旅行先の情報収集に欠かせない人気サイトであるが，2000年に設立されてから日本国内での利用が普及するまでは少々時間がかかり，2000年代の後半に入って利用者が急増している。このことは今回収集したレビューデータからも確認された。訪問年別のレビュー件数は表1で確認できる。香川県に関するレビューデータは2008年から掲載がみられる。また，訪問時期の記載がないデータは338件あった。2010年に第1回瀬戸内国際芸術祭が開催され，その翌年からレビュー件数が急増した。とくに，2016年の第3回瀬戸内国際芸術祭の開催とその前年にはレビュー件数が過去最高を記録し，レビューが急増した。訪問月のレビュー件数は表2のとおりであり，春の花見と秋の紅葉が楽しめる時期やお盆休みを含む夏季休暇期間中にレビュー件数が多いことが分かった。最後にレビュー投稿者の居住地方別レビュー件数は，表3のとおりであり，居住地方は東京都と関東地方が最多で，近畿地方がつづく。

↘ した。他方で，レビューが一部の観光スポットに集中している点は否定できないが，本稿では上位30件を対象に限定することでデータの偏りが分析に大きく影響を及ぼすことはないと判断した。

8） https://www.tripadvisor.jp/Attractions-g298231-Activities-Kagawa_Prefecture_Shikoku.html

9） 居住地とは，実際にはTrip Advisorの登録時に記載した登録地とすることがより正確な表現といえるが，本稿では登録地を居住地とみなして分析する。さらに，全国をいくつかの地方に再分類したうえ，分析を進める。

10） 掲載されているデータはレビューとともに画像データも掲載されていることも多い。そのため，データのTimeラインを分析することで実際訪問された観光コースなどを分析することも可能である一方，個人の特定や行動の解明につながるなど，いわゆる不本意ながら個人情報の収集などが懸念される。本稿では分析に必要な必要最小限の項目に限定して，データ収集を行った。

表1　年[11]別件数

訪問年	レビュー件数
2008	9
2009	10
2010	25
2011	174
2012	287
2013	527
2014	823
2015	1,219
2016	1,178
2017	965
2018	792
2019	737
2020	80
不明	338
合　計	7,164

表2　訪問月別件数

訪問月	レビュー件数
1月	416
2月	396
3月	554
4月	543
5月	780
6月	353
7月	525
8月	757
9月	651
10月	671
11月	676
12月	504
不明	338
合　計	7,164

表3　居住地方[12]別件数

地　方	レビュー件数
北海道	152
東北	74
北陸	64
関東	1,173
東京都	1,604
中部	374
近畿	1,156
中国	312
四国	555
九州・沖縄	221
外国	208
不明	1,271
合　計	7,164

Ⅳ　分　　析

本章では，前章で収集した7,164件のレビューデータをもとに，テキストマイニング手法を用いて分析する。分析には主に，必要なデータの整形に「R」言語，その他形態素解析や対応分析，共起ネットワーク分析に「KHコーダー」を用いて行った[13]。

まず，大量のレビューデータに多用された形態素を抽出する（図1−③の段階）。次に，抽出された形態素と，訪問時期（図1−④の段階）および投稿者の居住地方との関係性を対応分析を用いて把握する。最後に，多用された言葉はどの類の言葉と一緒に使われているのかを把握するため，共起ネットワーク分析を行う（図1−⑤の段階）。

11）　2020年の件数は8月からデータ収集を行ったこととコロナ禍の影響のため，レビュー件数が極端に少ない。

12）　東京都は，関東地方に含まれるが，東京都からの来訪者の特徴を把握するため，本稿では関東地方から東京都を切り離して，分析を行った。

13）「R」言語，「KHコーダー」どちらも無料で自由に利用できるフリーソフトであるが，前者はデータの収集にとどまらず統計分析やデータ整形にも有効で，利用者も多い。多くのテキストマイニング用のソフトが英語を代表する1byte文字に適しており，2byte文字である日本語の分析には工夫が必要で，利用に向けての難易度が高い。そのため今回は，日本語への対応やGUIが充実しており使いやすく，日本語の解析・分析に適している後者の「KHコーダー」を用いた。なお，「KHコーダー」の中での解析には「R」言語との相互利用の利便性を考慮し「Me Cab」を選択して分析を行った。

表 4　頻出形態素

順位	抽出語	出現回数	順位	抽出語	出現回数	順位	抽出語	出現回数
1	思う	2,298	51	建物	441	101	出る	270
2	行く	2,249	52	展望	440	102	橋	269
3	見る	1,870	53	たくさん	428	103	上る	267
4	高松	1,431	54	四国	416	104	朝	266
5	時間	1,362	55	本宮	410	105	勧める	262
6	良い	1,125	56	展示	409	106	ガイド	261
7	登る	1,039	57	楽しい	406	107	大きい	261
8	公園	924	58	フェリー	403	108	堀	259
9	人	916	59	綺麗	401	109	施設	258
10	観光	893	60	言う	398	110	紅葉	253
11	島	875	61	無料	390	111	徒歩	251
12	景色	868	62	高い	383	112	小豆島	250
13	場所	829	63	天守	366	113	瀬戸内	248
14	美術館	821	64	入場	363	114	園内	246
15	歩く	803	65	楽しむ	360	115	残念	246
16	階段	798	66	眺め	360	116	歴史	245
17	庭園	718	67	駅	346	117	琴	244
18	作品	687	68	眺める	345	118	おすすめ	243
19	多い	682	69	奥	344	119	時期	243
20	オリーブ	682	70	雰囲気	339	120	説明	243
21	駐車	670	71	直島	336	121	子供	242
22	訪れる	663	72	散策	329	122	特に	241
23	海	651	73	有名	327	123	石段	238
24	入る	643	74	利用	327	124	結構	231
25	楽しめる	639	75	客	326	125	素敵	231
26	バス	629	76	最高	326	126	今回	230
27	感じる	615	77	大変	326	127	季節	229
28	アート	611	78	訪問	326	128	聞く	229
29	城	580	79	乗る	322	129	今	227
30	見える	564	80	電車	319	130	昔	227
31	途中	559	81	きれい	318	131	価値	225
32	素晴らしい	559	82	屋島	318	132	桜	225
33	出来る	536	83	映画	316	133	船	225
34	広い	514	84	香川	314	134	知る	225
35	行う	507	85	見学	313	135	池	225
36	道	505	86	店	309	136	長い	225
37	感じ	499	87	渡る	303	137	初めて	224
38	前	498	88	空間	301	138	立派	224
39	日本	485	89	参拝	300	139	松	223
40	石垣	479	90	少ない	297	140	一番	222
41	瀬戸内海	467	91	案内	295	141	行ける	222
42	近く	467	92	自転車	291	142	散歩	220
43	土産	466	93	丸亀	289	143	一つ	219
44	天守閣	466	94	栗林公園	288	144	回る	219
45	車	462	95	中	286	145	体験	215
46	瀬戸大橋	460	96	買う	286	146	必要	215
47	来る	450	97	スポット	284	147	感動	212
48	美しい	446	98	芸術	279	148	小さい	211
49	少し	446	99	食べる	278	149	違う	210
50	写真	443	100	参道	270	150	自然	210

本章では，上述した手順により，香川県を訪れた観光客が抱く香川県の観光イメージを明らかにする。

1　形態素解析

本節では収集した7,164件のレビューデータをもとに形態素解析を行った。表４はレビュー文中で確認された上位150位までの形態素を表したものである。

今回の解析では「思う」が出現回数2,298回で最上位となった[14)]。その他，地名や，行動が読み取れる表現，イメージが読み取れる表現などが頻出形態素として抽出された。

2　対 応 分 析

外部変数と形態素との関係性を明らかにするため，差が顕著な60語をもとに，また，外部変数として訪問年・訪問月・居住地方を加え，分析を行った。

なお，対応分析の全ての結果図は外部変数間の差の視認性を高めるため，中心から外側に30％傾斜配置（拡大）したプロット図にして表した。

(1)　訪問年との関係性

本節では，投稿者が訪問した年に注目し，レビュー文中の形態素間の差が顕著な60語をもとに訪問年との関係性を分析した。その結果は，「時期」（成分１），「瀬戸内国際芸術祭」（成分２）の２要素で説明でき，両者の出現の有無で次のように説明できる（図２）。

成分１は，2016年を起点に前後で両分し説明できる。また，2016年放映のNHK大河ドラマ真田丸によるお城ブーム，および瀬戸内国際芸術祭による現代アートや島のブームの起点であると説明でき，2016年以降は島しょ部やお城への関心が高まったことが分かる。他方で成分２は，2010年から始まった瀬戸内国際芸術祭の開催を起点に前後で両分して説明できる。成分２の上部は，レビューの中心が現代アートと島しょ部に置かれていることが分かる。

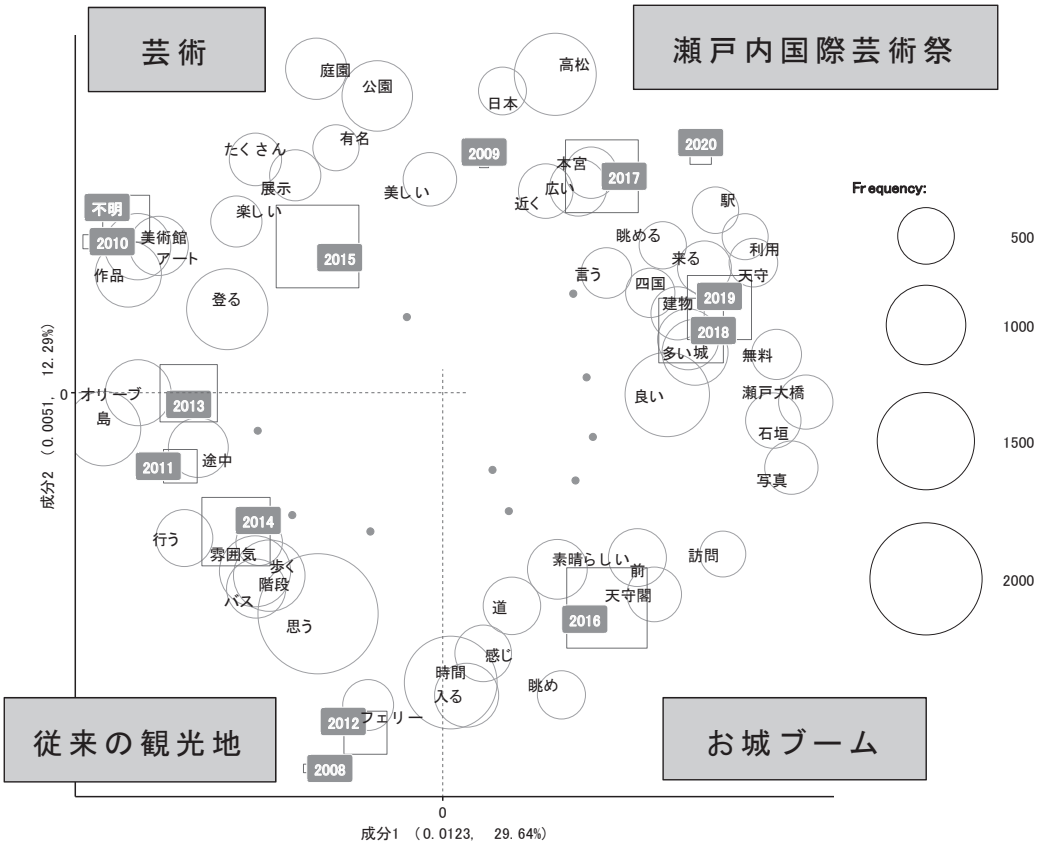
(2)　訪問月との関係性

本節では，投稿者が訪問した月に注目し，レビュー文中の形態素間の差が顕著な60語を用いて，訪問月との関係性を分析した。その結果は，「季節」（成分１），「休暇」（成分２）の２要素で説明でき，両者の出現の有無で次のように説明できる（図３）。

訪問月別の分析から，春と秋を中心とする花見や紅葉を楽しむ行楽や，お盆や年末年始，学校の休業期間を利用した家族旅行などが中心となっている観光行動が再確認された。成分

14）「思う」は，「言い切らない」日本語特有の表現といえ，文中に多用されているが，対象への特定のイメージを抱いていることを読み取ることができると判断し分析に用いた。

図 **2**　訪問年への対応関係



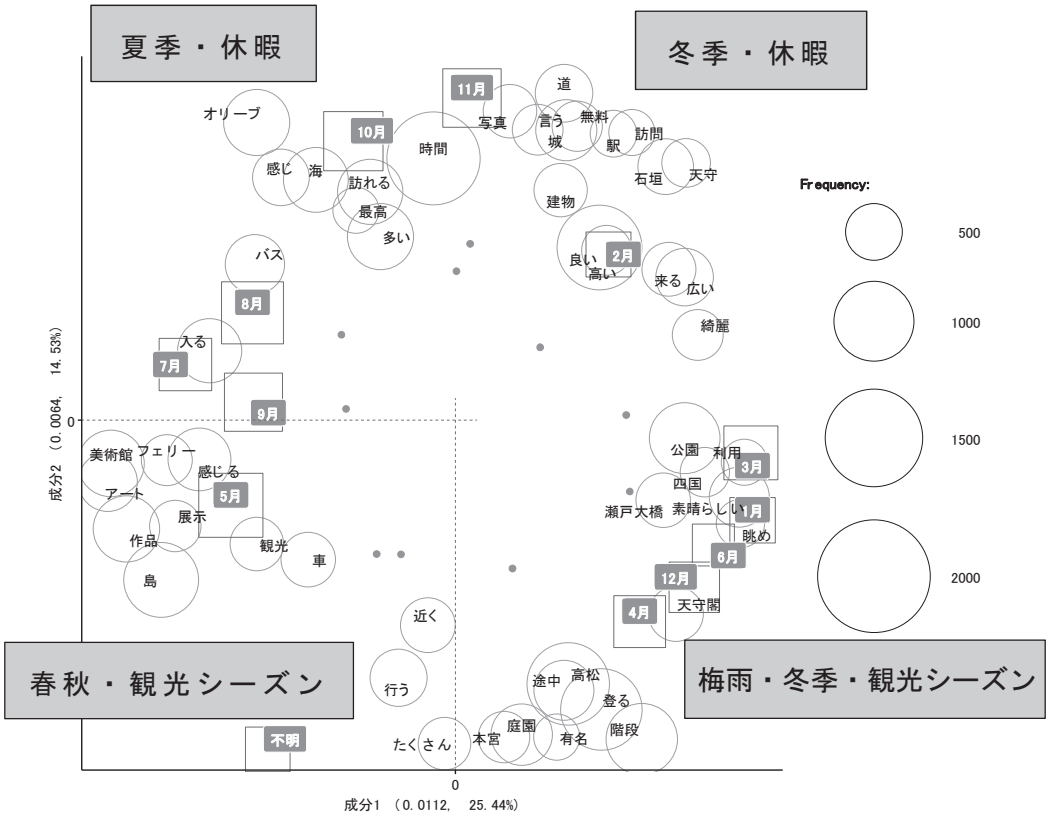
1は，花見や紅葉などの観光が盛んな春，秋の行楽シーズンと，梅雨や冬季の行楽期に両分できること，成分2は，家族旅行が可能となる休暇期間に両分できること，が分析結果図から確認できる。

(3)　居住地方との関係性

本節では，投稿者の居住地域に注目し，レビュー文中の形態素間の差が顕著な60語を用いて，居住地域を9地方[15)]と外国，不明の11カテゴリーに再分類し，居住地方との関係性を分析した。その結果は，「ブーム」（成分1）と「芸術」（成分2）の2要素で説明でき，両者の出現の有無で次のように説明できる（図4）。

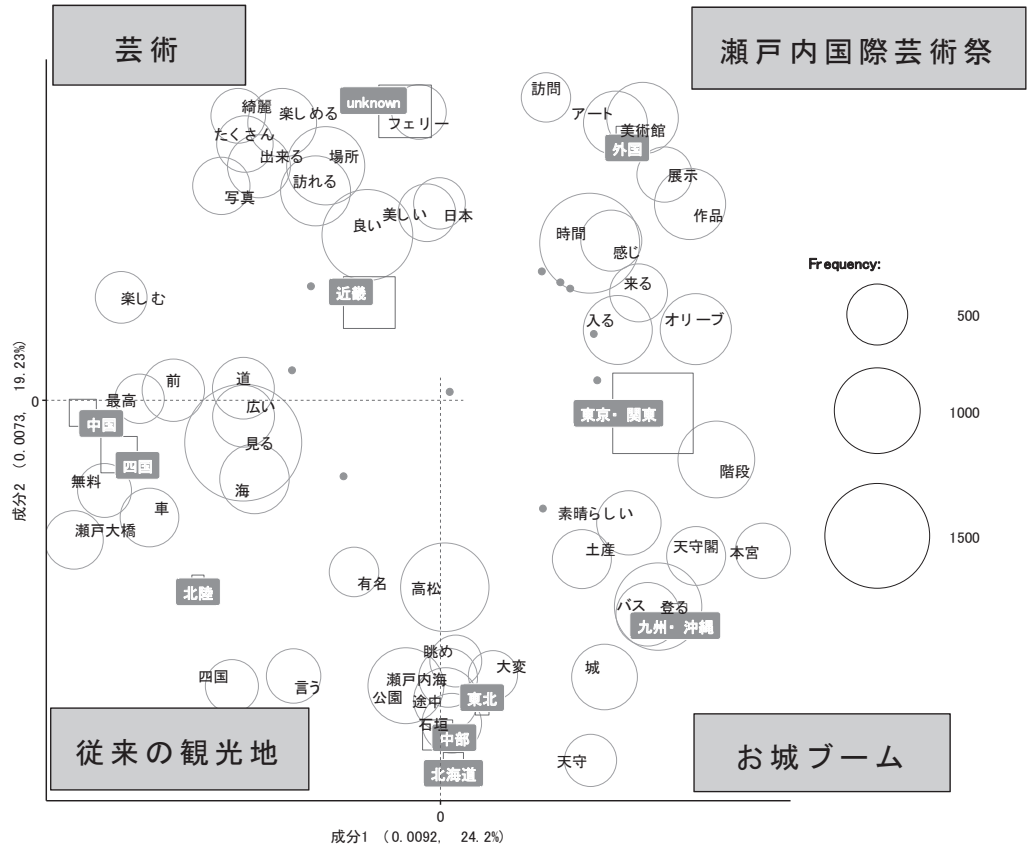
15)　47都道府県を，北海道，東北，北陸，東京・関東，中部，近畿，中国，四国，九州・沖縄の9地方に再分類した。

図3　訪問月への対応関係



成分1は，2010年から3年毎に開催された瀬戸内国際芸術祭による現代アートと島のブームに加え，2016年放映されたNHK大河ドラマ真田丸によるお城ブームが影響し，レビュー文中にアートや美術館，オリーブ，天守閣，城などの表現が頻出していることが分かる。また外国に居住している投稿者はアートや島，九州・沖縄に居住する投稿者はお城，東京・関東地方の投稿者はアートや島とお城の両者に関心をみせていることが分析から明らかになった。他方で，中四国地方に居住する投稿者が従来の観光スポット，北海道や東北，中部地方などの国内の遠隔地からの投稿者が従来の観光スポットや各種ブームで知名度が高くなった人気観光スポットに関心を寄せていることが分かった。このように，訪問機会が少ない遠方からの観光客の場合，できるだけ多くの観光スポットを回遊する一般的な観光者の行動特徴が再確認される結果となった。

図 4　居住地域への対応関係



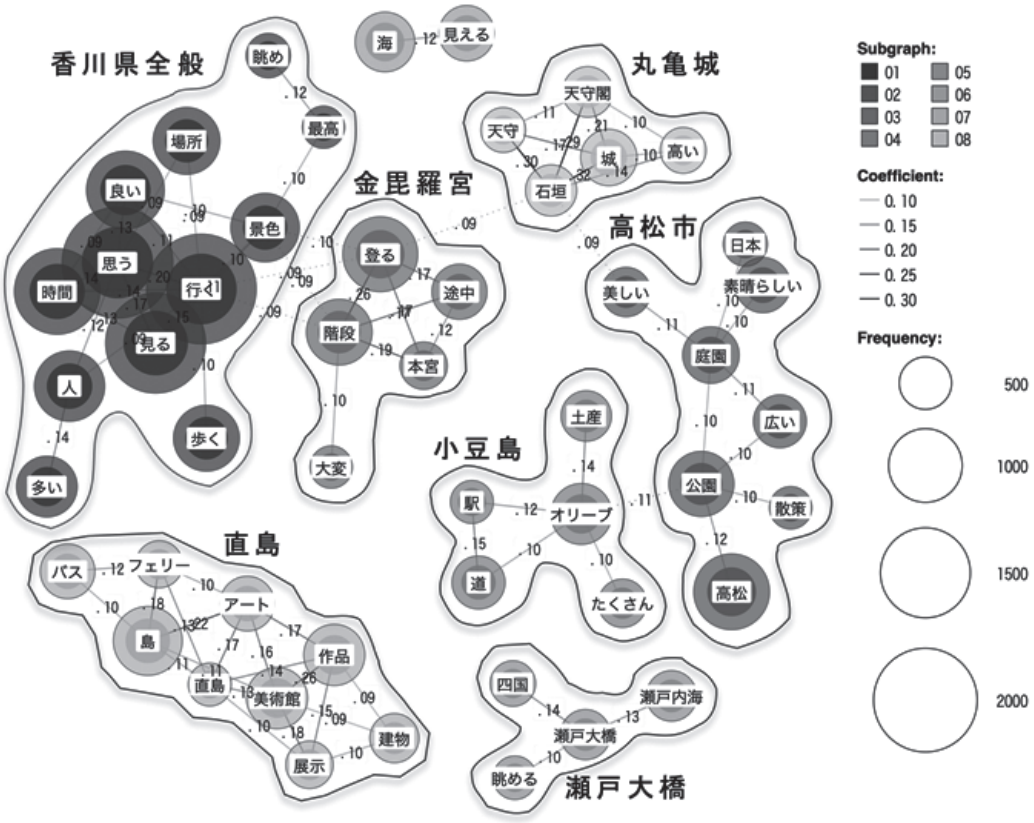
3　共起ネットワーク分析

本節では，レビュー文中で同時に利用される「語」[16] を分析することで投稿者がもつ香川県に対するイメージを解明するために共起関係と語（node）間の関係（edge）を解明するために中心性を分析した。

まず，共起ネットワーク分析を行った結果，投稿者は香川県を 8 つのカテゴリーで捉えていることが分かった（図 5）。カテゴリーは，香川県全般，高松市（栗林公園など），丸亀城，金毘羅宮，小豆島，直島，瀬戸大橋，海の風景の 8 区分である。海岸遠景全般をさす海の風景を除いた 7 区分は場所の特定ができ，投稿者は香川県を具体的に 7 つのカテゴリーに区分して認識していることが明らかになった。また，香川県が海の見える海岸遠景のある場所（観光地）として認識されていることも分かった。

16）「語」は 1 つ以上の形態素で構成される。形態素がそれ以上分解できないのに対し，語は分解できる場合がある。また，前者は文法的，後者は意味的，分析に用いられる。

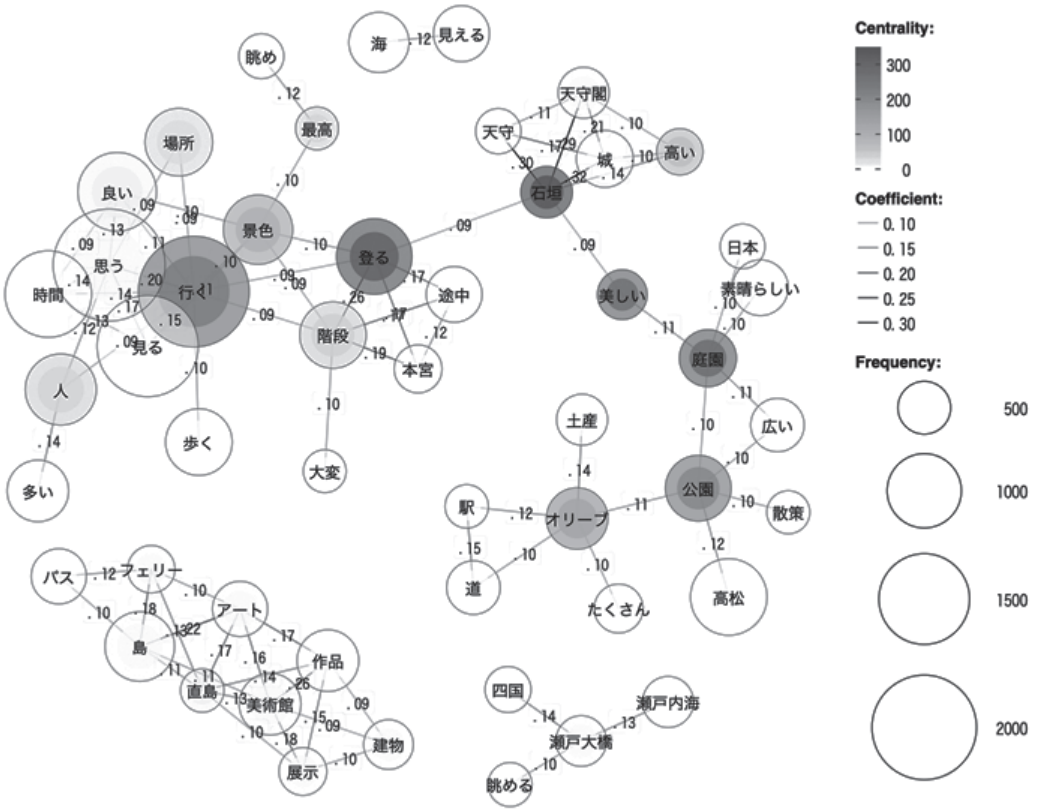
図 5　共起ネットワーク[17)]



さらに，カテゴリーの形成にどの語が中心的な役割を果たしているのかを解明するため，いわゆる共起中心性を分析した。結果は図 6 に示した。古くから観光スポットとして認知されている階段や本宮で表現された「金毘羅宮」，天守閣や石垣で表現された「丸亀城」，日本庭園や公園で表現された「栗林公園」，およびオリーブで表現された「小豆島」がもっとも香川県を代表する中心的な語（node）であることが確認できる。他方で，現代アートで国内外に知られる「直島」や本州と四国をつなぐ「瀬戸大橋」は香川県を代表する観光スポットであることに変わりはないが，直島と瀬戸大橋が香川県の観光イメージ形成における中心的な存在にはなってないことも今回の分析から明らかになった。

17)　共起ネットワーク図は，Subgraph の色でカテゴリーを，Coefficient の数字と線の太さで語（node）間の関係（edge）を，また Frequency の大きさで出現頻度が確認できる。

図 6　共起中心性



V　ま　と　め

本稿では，地域における観光情報の発信に観光客のニーズの反映に向け観光客がもつ地域への観光イメージの解明の必要性に着目し，Trip Advisor に掲載されたレビューデータ（UGC データ）を収集し，形態素解析，対応分析と共起ネットワーク分析を行った。

データ収集は，Web Scraping 技法により行った。Web Scraping に必要なコードは金（2009）をもとに適宜修正し作成した。収集データは7,164件となった。

分析には，大量のテキストデータを自然言語処理し文章を形態素に分割し出現頻度や関係性を分析するテキストマイニング手法を用いた。

はじめに形態素解析を行い，レビュー文中で頻出する形態素を表 4 に示した。これをもとに，投稿者の訪問した年や月，居住地方を外部変数とし，変数間の関係性を明らかにするため対応分析を行った。形態素と訪問年との分析からは，「時期」と「瀬戸内国際芸術祭」の 2 成分で説明でき，2 次元プロットにて対応関係の説明が可能となり，2010年の現代アートブー

ムや2016年のNHK大河ドラマ真田丸によるお城ブームの影響が確認された。訪問月との分析からは，「季節」と「休暇」の２成分で説明でき，２次元プロットにて対応関係の説明が可能となり，春の花見や秋の紅葉を楽しむ行楽，お盆や年末年始，学校の休業期間を利用する旅行による影響が確認された。形態素と居住地方との分析からは，「ブーム」と「芸術」の２成分で説明でき，２次元プロットにて対応関係の説明ができ，2010年の現代アートブームや2016年のお城ブームが定着し香川県に対するイメージ形成に影響したことが確認された。

次に，レビュー文中での形態素の利用傾向と中心性を把握するため，共起ネットワーク分析を行った。形態素間の関係性から，観光客は香川県観光を，８区分されたイメージとして認識していることが明らかになった。また，文中の形態素の中心性を分析した結果は，従来からの観光スポットである金毘羅宮や丸亀城，栗林公園，小豆島がもっともレビュー文中で香川県を代表する中心的な位置づけであったことが確認でき，香川県の観光イメージは，伝統的な観光スポットを基本としながら，新しい観光としての現代アートの直島，海の風景と瀬戸大橋になっていることが分かった。

本稿では，観光客が観光地をどのように区分し，イメージしているのかを，観光地に対するレビューデータを収集，分析することで，従来の分析手法とは異なる観光者を視座に分析を行った。また，データの収集手法をはじめ，分析の手法においても膨大なテキストデータの分析に適したテキストマイニング手法を用いることで，観光研究における新たな分析手法を提示した。

付記

本研究は科学研究費（課題番号20K12445）の助成による研究成果の一部です。

参　考　文　献

生田和重（2016）：学生が作成したキャリアプランに込められた感情の把握，大学教育ジャーナル，第13号，pp. 48−55.

金　徳謙（2015）：観光資源の利用実態の解明に向けた画像ビッグデータの空間分析，中四国承継学会第56回研究発表大会，2015年12月

金　徳謙（2016）：画像ビックデータ分析に基づく香川県の観光潜在力の分析，香川大学経済論叢，第88巻第４号，pp. 463−484.

金　徳謙（2019）：Massive Dataの収集・分析手法を用いた観光イメージ分析――宮島に関するTrip Advisorの英文Reviewsを事例に――，修道商学，第59巻第２号，pp. 115−132.

若山公威（2016）：ツイートからの観光ルート抽出，名古屋外国語大学外国語学部紀要，第50号，pp. 167−177.

Amaro, S., Duarte, P., & Henriques, C.（2016）. Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, *59*, 1−15.

Bucur, C.（2015）. Using Opinion Mining Techniques in Tourism. *Procedia Economics and Finance*, *23*, 1666−1673.

Fang, B., Ye, Q., Kucukusta, D., & Law, R.（2016）. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, *52*, 498–506.

Guo, Y., Barnes, S. J., & Jia, Q.（2017）. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483.

Költringer, C., & Dickinger, A.（2015）. Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research*, *68*, 1836–1843.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L.（2018）. Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323.

Lu, W. L., & Stepchenkova, S.（2015）. User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. *Journal of Hospitality Marketing & Management*, *24*, 119–154.

Mak, A. H. N.（2017）. Online destination image: Comparing national tourism organisation's and tourists' perspectives. *Tourism Management*, *60*, 280–297.

Marine-Roig, E., & Anton Clavé, S.（2015）. Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, *4*, 162–172.

Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M.（2017）. A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, *54*, 771–785.

Zeng, B., & Gerritsen, R.（2014）. What do we know about social media in tourism? A review. *Tourism Management Perspectives*, *10*, 27–36.