

# テキストマイニングによる観光地の定量的評価分析

1190437 越智香那恵

高知工科大学 経済・マネジメント学群

## 1. はじめに

観光庁は「国・地方公共団体・民事事業者等が連携し、訪問外国人旅行者の受入環境の整備・充実を総合的に推進することで、全国的に訪問外国人旅行者の訪問を促進するとともに、満足度を高め、リピーターの増加を図ることを目指して」（同庁HP）、観光地域づくりを推進している。また、観光を通じて地域の活性化を図るために「観光関係者のみならず、地域住民も含めた地域の幅広い関係者が連携し、住んでよし、訪れてよしの魅力ある地域づくり」（同庁HP）を目指している。全国、世界的には、各所に観光地が存在しているが、近年の世界的なIT化の進展と、質量ともに情報データベースが増大しているため、観光地のロコミ情報が入手しやすくなっている。情報の質的なバラツキは否めないが、量的に一定数以上確保できる文字データを解析することで、一定レベル以上の観光地の現状把握と評価分析を行うことができるようになってきている。本稿では、樋口耕一氏の開発した分析ソフトKH-Coderを採用し、先行研究をもとに、テキストマイニングによる観光地の定量的評価分析を試みる。

## 2. 研究の背景

観光客の地域復興や観光地域づくりは、国の情報政策課題とされている。高知県庁の平成29年度県外観光客入込数の推移（図1）を見てみると、一昨年平成29年に7年ぶりの過去最高数440万人を突破している。

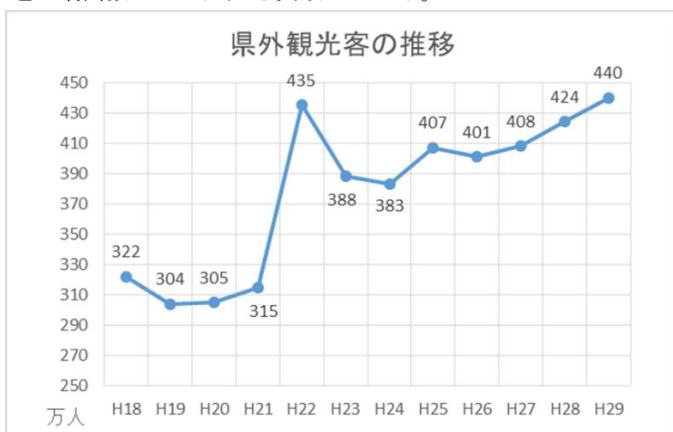


図1 高知県庁県外観光客入込数推移から筆者作成

過去最高数を突破しているものの、継続的な観光客の増加は地域活性化に必要であると考えられる。

文字列を対象としたデータマイニングを「テキストマイニング (text mining)」という。文章データを単語や文節で品詞ごとに抽出し、出現頻度や共制限の相関、出現傾向、時系列などを解析する。2000年ころから数値データ以外の文字列を対象としたデータマイニング手法が発達し、2008年ころからは、経営学/商学や地域活性化の分野でも援用が図られている。国立情報学研究所 (NII) の論文検索サービスCiNii (NII学術情報ナビゲータ) のデータベースを検索すると、テキストマイニング手法を用いたものは2765件の報文が抽出された。1998年ころからの研究発表やジャーナル論文が示されている。

## 3. 目的

本研究の狙いは、テキストマイニング手法を用いて分析をし、観光地の特徴の強化ポイントを明らかにし、今後の観光地の整備やPR活動に反映する等で、地域活性化につなげようとするものである。高知県中部、東部、西部の主要な観光地の印象についてロコミデータをもとにテキストマイニング手法で分析することで観光客のニーズを理解する。分析により得たデータから、高知県の観光客増加、地域活性化に対する効果のある改善策を見つけ出す。

## 4. 分析方法と対象

こうした背景を踏まえ、本稿では、高知県の代表的な観光地である高知城 (県中部)、室戸岬 (県東部)、四万十川 (県西部) を取り上げ、ロコミデータの量的評価分析を試みる。ロコミサイト「じゃらん」からロコミデータを収集し、分析する。解析は、学術研究で援用されているKH-Coderを使用し、軽量テキスト分析を行う。まず、語彙の抽出を行い、抽出語間の特徴について階層的クラスター分析を通じて解析する。次に、抽出語に用いられ方を分析するためにKWICコンコーダンスを実施する。

## 5. ①分析結果「高知城（高知県中部）」

### (1) 頻出語彙抽出

名詞	サ変名詞	形容動詞	固有名詞				
天守閣	126	観光	40	綺麗	54	姫路城	3
階段	79	一望	37	急	25	キツ	2
市内	52	散歩	31	立派	23	ビリ	2
天守	51	現存	29	きれい	16	桂浜	2
歴史	49	展示	16	大変	13	大河ドラマ	2
景色	40	見学	15	素敵	10	明治	2
眺め	29	案内	14	便利	7	伊賀上野	1

#### ①高知城観光客口コミ全体 273 サンプル

図2 高知城全体サンプル（名詞、サ変名詞、形容動詞、固有名詞）：筆者作成

KH-Coderを使用して、頻出語彙の抽出を行った。図2に関して、名詞では、上位に眺めや景観に関連する場所を示す語彙が多い。例えば「天守閣」「景色」「眺め」などである。サ変名詞でも「一望」などのワードが上位にある。また形容動詞も「綺麗」「きれい」が上位に抽出されている。

形容詞	副詞	名詞B	未知語				
良い	50	少し	21	おすすめ	21	♪	10
多い	29	結構	9	まつり	5	ε	3
素晴らしい	19	特に	8	そば	3	ω	3
暑い	16	初めて	6	おかげ	2	▽	2
高い	14	本当に	6	おじさん	2	キツイ	2
小さい	14	意外と	4	さくら	2	ポケモン	2
大きい	13	少々	4	ちゃん	2	((o*))	1
楽しい	11	キンキン	3	つくり	2	(^o)	1
美しい	10	一度	3	いくつか	1	(*)o))	1

図3 高知城全体サンプル（形容詞、副詞、名詞B、未知語）

また図3に関連して、形容詞では、上位に「良い」「多い」「素晴らしい」「美しい」「楽しい」などポジティブな表現の語彙抽出が目立つ。また、高知県が注力しているはずの「おもてなし」に関しては、名詞Bの1つしか抽出できず、県の観光政策との認識のズレがある可能性が指摘できる。また、未知語というカテゴリで♪が10個出てきており、訪問者がポジティブな気持ちであった可能性が高いと思われる。

#### ②高知城観光客男女別（48：52）

名詞	サ変名詞	形容動詞	固有名詞				
天守閣	63	観光	23	綺麗	18	ビリ	2
市内	33	一望	20	立派	8	桂浜	2
天守	32	現存	19	きれい	7	大河ドラマ	2
歴史	30	散歩	17	急	7	明治	2
階段	22	展示	11	大変	5	キツ	1
石垣	16	見学	10	素敵	4	伊賀上野	1
景色	15	案内	8	残念	3	元日	1
眺め	13	説明	7	十分	3	大阪城	1
場所	11	是非	4	結構	2		1

図4 高知城男性（名詞、サ変名詞、形容動詞、固有名詞）

名詞	サ変名詞	形容動詞	固有名詞				
天守閣	63	一望	17	綺麗	35	姫路城	2
階段	57	観光	17	急	18	キツ	1
景色	25	散歩	14	立派	15	鏡川	1
市内	19	現存	10	きれい	10	高知城前	1
天守	19	サービス	8	大変	8	小城	1
歴史	19	アクセス	7	素敵	5	昭和	1
ライトアップ	17	ガイド	7	便利	4	明治維新	1
最高	17	案内	6	好き	3		
眺め	16	説明	6	十分	3		
アイス	13	旅行	6	重要	3		
リン	12	見学	5	必要	3		
市場	10	散策	5	すてき	2		
石垣	10	展示	5	気軽	2		
場所	9	是非	4	結構	2		

図5 高知城女性（名詞、サ変名詞、形容動詞、固有名詞）

次に、高知城全体のサンプルを男女別に分析する。図4と図5に示す抽出語彙では、男性と女性の名詞、サ変名詞、形容動詞、固有名詞の頻出語彙について示している。男女とも名詞の最上位には「天守閣」がある。また、男性の場合「階段」と「石垣」というワードは連続した順位になっている。しかし、女性は「階段」が圧倒的に上位に来ている。これは、男性に関しては、「石垣」そのものや眺望を理由に高知城に訪れているのに対して、女性は階段を登ることそのものへのネガティブな感情が先行している可能性があるのではないだろうか。また、女性には「アイス」の単語が上位に出ており、これも女性の来訪目的や印象を示す語彙として特徴的である。

動詞	形容詞	副詞	名詞B				
思う	36	良い	28	少し	11	おすすめ	7
行く	32	多い	15	少々	4	おじさん	2
見る	29	素晴らしい	14	特に	4	さくら	2
登る	19	高い	8	結構	3	いくつか	1
楽しめる	14	広い	6	意外と	2	かん	1
感じる	14	暑い	5	実際	2	すし	1
見える	10	小さい	5	当然	2	そば	1

図6 高知城男性（動詞、形容詞、副詞、名詞B）

動詞	形容詞	副詞	名詞B				
行く	66	良い	22	少し	10	おすすめ	14
登る	37	多い	14	結構	6	まつり	5
見る	31	暑い	11	本当に	6	おかげ	2
思う	28	小さい	9	初めて	5	そば	2
歩く	14	楽しい	8	特に	4	ちゃん	2
楽しめる	12	大きい	8	キンキン	2	おじいちゃん	1
訪れる	12	美しい	8	意外と	2	おなじみ	1
行う	11	高い	6	一度	2	おもてなし	1
感じる	8	少ない	5	昔ながら	2	かたち	1

図7 高知城女性（動詞、形容詞、副詞、名詞B）

次に、図6および図7では、男女別の動詞、形容詞、副詞、名詞Bを示す。ここで特徴的なのは、動詞の頻出語彙である。男性の場合「思う」「見る」「感じる」「楽しめる」などの感覚的・主観的な動詞が上位に抽出されている。また、女性



れた。しかし、「急」や「きつい」の前後にある言葉を見るとほとんどがネガティブなロコミというよりは、「階段はきついですが、」などといった逆接を使いその後続く言葉はポジティブなロコミや、「急なので注意が必要」などの言葉で次に高知城を訪れる人への情報提供という形のものが多いことがわかった。

## 5. ②分析結果「室戸岬（高知県東部）」

### (1) 頻出語彙抽出

#### ①室戸岬観光客ロコミ全体 88 サンプル

名詞	サ変名詞	形容動詞	固有名詞	
灯台	28 観光	10 雄大	13 室戸岬	18
景色	21 ドライブ	9 自然	8 最御崎寺	7
海岸	20 整備	9 きれい	6 足摺岬	7
遊歩道	20 展望	9 キレイ	6 海岸線	6
夕日	14 感動	7 残念	6 エボシ	1
場所	13 散策	7 綺麗	6 桂浜	1
カップル	10 駐車	7 有名	4 高知空港	1
眺め	10 施設	4 格別	3 佐田岬	1
天気	10 満足	4 好き	3 日和佐	1

図 1 4 室戸岬全体サンプル (名詞、サ変名詞、形容動詞、固有名詞)

まず、室戸岬のロコミ全体の頻出語彙抽出を行った。ここでは、図 1 4 で室戸岬ロコミ全体の名詞、サ変名詞、形容動詞、固有名詞の語彙抽出を示している。名詞に「灯台」「景色」「夕日」が上位にあることから、灯台を見に来ている人、景色を見に来ている人が多いことがわかる。また景観などに関すること以外にも「カップル」というワードも目立っている。これは室戸岬が「恋人の聖地」とも呼ばれていることから、「カップル」というワードが多いのではないかと考えられる。サ変名詞では、「観光」に続き、「ドライブ」が上位にある。観光目的の来訪者とドライブのための来訪者があるように考えられた。形容動詞を見ると、「雄大」「きれい」「綺麗」というポジティブなワードの中に、「残念」というネガティブなワードも上位に入っている。室戸岬が残念に感じたためにこの語彙が出ている可能性も考えられる。

名詞B	
おすすめ	11
しぶき	2
そば	2
うってつけ	1

図 1 5 室戸岬全体サンプル (名詞B)

また、図 1 5 より名詞Bには、「おすすめ」の語彙が圧倒的に多いという特徴がある。来訪した人が次に来る人たちのためにロコミという形で情報を発信しているのではないかと

考えられる。

#### ②室戸岬男女別 (52 : 48)

名詞	サ変名詞	形容動詞	
灯台	20 展望	9 雄大	8
遊歩道	16 観光	7 綺麗	4
海岸	14 整備	7 キレイ	2
景色	11 散策	5 快適	2
夕日	10 ドライブ	4 格別	2
場所	9 感動	4 真っ青	2

図 1 6 室戸岬男性 (名詞、サ変名詞、形容動詞)

名詞	サ変名詞	形容動詞	
景色	10 ドライブ	5 自然	7
カップル	9 感動	3 きれい	5
天気	9 観光	3 残念	5
灯台	8 駐車	3 雄大	5
地球	7 一望	2 キレイ	4
海岸	6 記念	2 穏やか	2

図 1 7 室戸岬女性 (名詞、サ変名詞、形容動詞)

次に、室戸岬全体サンプルを男女別で分析する。図 1 6、図 1 7 には、名詞、サ変名詞、形容動詞を示している。名詞に注目すると、女性には「カップル」というワードが 2 番目に多いワードとして出ている。これは男性にはないワードだった。また、男性と女性で「景色」「灯台」の 2 つが逆に位置していることが図からわかる。形容動詞を見ると、基本的にはポジティブなワードだが、女性に「残念」というネガティブな言葉が比較的多く出ている特徴があった。

#### ③室戸岬年代別 (29 歳以下 : 60 歳以上 = 53 : 47)

名詞	サ変名詞	形容動詞	
景色	12 展望	6 自然	7
天気	9 ドライブ	5 きれい	6
灯台	9 観光	5 雄大	6
カップル	8 散策	4 キレイ	4
写真	8 駐車	4 残念	4
場所	8 一望	3 好き	3
スポット	6 整備	3 素敵	3
パワー	5 満足	3 壮大	3
海岸	5 ゴツゴツ	2 穏やか	2

図 1 8 室戸岬 29 歳以下 (名詞、サ変名詞、形容動詞)

名詞	サ変名詞	形容動詞	
灯台	19 感動	7 雄大	7
遊歩道	16 整備	6 綺麗	4
海岸	15 観光	5 真っ青	3
夕日	12 ドライブ	4 真っ白	3
景色	9 散策	3 キレイ	2
景観	7 施設	3 快適	2
ホテル	6 駐車	3 残念	2

図 1 9 室戸岬 60 歳以上 (名詞、サ変名詞、形容動詞)

次に、室戸岬全体サンプルを年代別で分析する。図 1 8、図 1 9 に名詞、サ変名詞、形容動詞を示している。名詞を見てみると、60 歳以上、29 歳以上で上位に来ている語彙が全体的に異なっている。例えば、60 歳以上は「灯台」「遊



歩道」が上位だが、29歳以下は「景色」「天気」が上位にある。また、29歳以下には「カップル」「スポット」「パワー」というワードが出ている。これは60歳以上には2つしか出ていないという結果だった。

(2) 階層的クラスター分析

室戸岬のロコミを全体的に見ると、高知城同様、室戸岬の灯台、景色に関するポジティブなワードがほとんどだったが、形容動詞には、「残念」というワードが上位に出ている。ここでは、「残念」というワードに注目する。



図20 室戸岬全体の階層的クラスター分析

図20では、室戸岬ロコミ全体の階層的クラスター分析の結果を示している。「残念」に繋がっているワードは、「晴れる」「思う」「天気」「雨」となっている。この結果から考えられるのは、「残念」というワードは「天気」に関するワードだということだ。頻出語彙抽出では、室戸岬に対するネガティブワードかと思われたが、階層的クラスター分析で見ると「天気が残念～」「雨が降っていて残念～」などのロコミになっていたように考えられる。

(3) KWICコンコーダンス

階層的クラスター分析で出た結果を確かめるためにKWICコンコーダンスで確かめる。

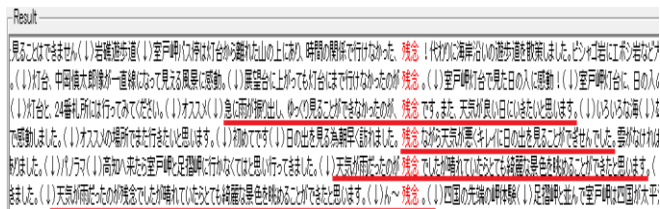


図21 室戸岬KWICコンコーダンス「残念」検索全結果

図21に「残念」の検索結果を示している。やはり残念に続いているのは「天気」に関する言葉であることが確認できた。

5. ③分析結果「四万十川（高知県西部）」

(1) 頻出語彙抽出

①四万十川観光客ロコミ全体147サンプル

名詞	サ変名詞	形容動詞	固有名詞
清流	50 観光	16 綺麗	64 沈下橋
流れ	30 遊覧	14 自然	27 四万十川
景色	27 ドライブ	11 雄大	15 勝間
場所	24 感動	9 有名	11 岩間
最後	21 体験	9 穏やか	10 予土線
カヌー	18 満喫	9 十分	8 プリプリ
上流	13 運転	6 素敵	8 マーチ
風景	12 期待	6 大変	8 酒匂川
季節	10 想像	6 広大	7 仁淀川

図22 四万十川全体（名詞、サ変名詞、形容動詞、固有名詞）

まず、四万十川のロコミ全体の頻出語彙抽出を実施した。図22では、名詞、サ変名詞、形容動詞、固有名詞の頻出語彙について示している。形容動詞を見ると「綺麗」「雄大」「有名」などといったワードが上位に来ている。全体的にはポジティブなイメージがあることがわかる。名詞に注目すると、「清流」「流れ」「景色」という言葉が上位に来ている。固有名詞に「四万十川」よりも「沈下橋」というワードが上位に来ているため、沈下橋を目的に四万十川に来ている観光客が比較的多いのではないかと考えられる。

地名	
日本	22
四万十	21
高知	13
四国	9

図23 四万十川全体サンプル（地名）

次に図23より、地名を見てみると、「日本」というワードが「四万十」よりも上位に来ている。四万十川は、「日本三大清流」の一つとされており、「日本最後の清流」とも呼ばれているため、「日本」というワードが目立っているのだと考えられる。

②四万十川男女別（47：53）

名詞	サ変名詞	形容動詞
清流	32 観光	9 綺麗
流れ	16 ドライブ	7 自然
場所	14 遊覧	7 雄大
最後	13 運転	6 十分
景色	12 感動	6 有名
カヌー	8 期待	5 大変
上流	8 生活	5 透明
季節	7 満喫	5 穏やか
道路	7 想像	4 広大
国道	6 増水	3 残念
ホテル	5 旅	3 ダイナミック



ていなかったため、KWICコンコーダンスで検索してみたところ、結果はほとんど「遊覧船」に関連していることがわかった。

## 6. 結論

以上から、観光地の口コミを全体的に見ると、ポジティブなものが多いことがわかった。

ポジティブなものの中に、少ないがネガティブワードも見られた。しかし、ネガティブなことだけを口コミで書いている人は少なく、ネガティブなワードを使って、次回訪れる人に向けて、情報提供や注意喚起をしたり、「階段はきつかったが、達成感があった」などのネガティブ以上の収穫があったりと誰かに伝える口コミをしている人が多いことが結果として出ている。ネガティブワードではあるが、観光地そのものに対するものではないとわかる。

ところが、情報提供などのためとは言え、ネガティブなワードが出ていることに変わりはない。ネガティブワードがわかりやすく出ていた高知城、室戸岬で必要なのは、「整備」ではないだろうか。今までも観光地の整備はされているが、今回ネガティブが抽出されたことで、より良くなる整備方法を見つけることができると考える。例えば、高知城は階段が急であるならば、登りやすくなるようにエレベーター、エスカレーターのようなものを取り付けてはどうだろうか。室戸岬の場合、残念という口コミに関連づいていた「雨」が降っているときの「足場の水はけがよくない」といった口コミがあった。そのため、水はけがよくなるよう整備をすることが必要ではないかと考える。確かに、整備を行えば、その分高知城の「歴史」の部分の少し欠いてしまうかもしれない、室戸岬も「自然」という部分を少し減らしてしまうかもしれない。その部分の兼ね合いが難しいため、考えていく必要がある。

## 7. 今後（ヒアリングも含めて）

今後の目標として挙げられるのは次の4つだ。

### ① 口コミ評価の向上

口コミ評価は、高評価であればあるほど観光客の増加が見込まれると考えられる。口コミの評価が高ければ、口コミサ

イトの上位にも表示されるため、多くの人が見られるようになるだろう。また、口コミ評価の向上はリピーターの増加も期待できるのではないだろうか。

### ② 口コミ数の増加

口コミサイトは今ではインターネットが普及しているため、多くの人が見ることができる。口コミ評価の向上に合わせて、「その観光地に実際にいった人の声」も増加すれば、行きたいと思う人が増えて観光客増加が見込めるだろう。

### ③ アンケート調査のテキストマイニング分析の実施

高知県では、各観光地にてアンケート調査を行っているが、5段階評価の集計を行っているだけのことだ。自由記述に書かれた実際の声をテキストマイニング手法で分析することで口コミサイトだけでは得られない情報を得ることができるのではないだろうか。

### ④ 観光地の整備について

「歴史や自然を残す」という面で、整備をどこまですべきなのかは今までも考えられてきている。特に高知城について考えたい。本来お城というのは敵が攻めにくいように急な階段などを用いて、造られている。しかし、観光客の立場からすれば、急な階段は、登りにくくきついものがある。高知城も当時の造りのまま残っているお城の1つだ。観光客の中には歴史を見に来ている人もいるが、歴史を大切にするか、観光地として観光しやすい形を取るか、考える必要がある。

今回、高知県の観光地3つのデータを集め分析した。頻出語彙抽出の結果で出たものの一部に注目して、階層的クラスター分析、KWICコンコーダンスを用いて分析を行ったが、より細かく調べることが必要だと感じた。また、今回3か所だけでなく、他の高知県の観光地、全国各地の観光地を調べれば、比較対象が多くなるため、よりもっと高知県の観光客増加、地域活性化への貢献ができるだろう。

[謝辞]

本研究を進めるにあたり、ご指導をいただいた高知工科大学経済・マネジメント学群桂信太郎教授に感謝いたします。

また、ヒアリング調査にご協力いただいた高知県観光振興部  
地域観光課、高知市商工観光部観光振興課の皆様には感謝申し  
上げます。

**【参考文献】**

[1] 高知県庁観光振興部観光政策課「平成29年度県外観  
光客入込動態調査報告書」

[2] 樋口耕一著『社会調査のための計量テキスト分析』ナ  
カニシヤ出版（2014）

[3] ロコミサイト「じゃらん」

<https://www.jalan.net/kankou/>