

ブログデータを用いた八ヶ岳地域の観光行動の 分析と移動中の景観評価における活用可能性

佐々木 邦明¹・桜田 朋也²

¹正会員 山梨大学教授 工学部土木環境工学科 (〒400-8511 山梨県甲府市武田4-3-11)

E-mail:sasaki@yamanashi.ac.jp

²非会員 マサル (〒135-8432 東京都江東区佐賀1丁目9-4)

本研究は八ヶ岳南麓地域の観光行動特性と景観評価について検討を行った。用いたデータは2013年に検索エンジンで集められた八ヶ岳地域を訪れた観光客によるブログのエントリの記述である。その分析はデータの特性からテキストマイニングを用いて行った。

基礎的な分析として2010年に収集したブログデータとの比較を通じて、ブログ記述の安定性を確認し、リピータと初来訪者に着目して観光特性について明らかにした後、八ヶ岳エリアの景観評価がどのように行われているのかについて、テキストマイニング手法による分析可能性を検証した。その結果として、推測された趣味や訪問経験によって観光行動が異なることや、道中や各スポットにおける景観評価が何を対象として、どのような場所で行われているのかについて明らかにできることを示した。

Key Words : *tourist behavior, text mining, landscape evaluation,*

1. はじめに

観光行動は季節変動や個人間の変動が大きいことや、まれな特性によって、その実態を的確に調査することが困難な行動である。日常的・定期的な交通ではホームページ調査によってある程度の量や特性を把握可能であるが、観光行動は稀でかつその特性が非定常であり、その特性をホームページ調査で行うには効率が悪い¹⁾。入り込み型の調査を行う場合には、そのサンプル抽出方法によるバイアス発生が危惧され、入り込み客数調査などを並行して行うことで、補正を実施する例もみられるが²⁾、行動全体の補正が可能なケースは、比較的同質的な観光がなされているという仮定が必要で、非同質な観光行動の場合には、各スポットにおける入り込み数のみの補正となる。しかし、そのようにして様々な補正等が行われた観光行動調査を用いたとしても、観光地の魅力向上にどのように貢献できるのかということを考えると、観光地の様々な課題解決において、交通調査分析を活用方法が提示されているが³⁾、その後の観光地の魅力向上策などにおいて、行動調査を導入することが特に明示的には示されていない⁴⁾。その理由として、観光地の魅力向上を検討する際には、必ずしも集計的・平均的な行動特性が必要とはされず、特定の層の観光客に訴えかける観光地のイメージ形成や、特定の人が反応するストーリー

性を持たせることなど、個別的なマーケティングやそれに対応する詳細なデザイン戦略等が有効であることが指摘されている⁵⁾。とすれば、観光地の魅力向上の戦略・戦術を検討する際に、より個人から多く・深く集められた情報が観光マーケティングには有益なのであろう。しかし現実的に、個人から大量の情報を簡単に得ることは容易ではなく、GPS等を用いた移動軌跡データや、観光情報ウェブのアクセスのログ等から個人のデータを収集することも可能ではあるがそれにも限界はある。佐々木ら⁶⁾はそのような背景のもと、ブログマイニングを用いてその行動特性を分析した。

本研究は佐々木らによる研究の発展として、ブログからのデータ収集可能性を検証する。佐々木らは既存のアンケート調査と比較してブログから得られる行動特性がどのようになっているかを検証した。本研究は、先行研究から3年後に、新たにブログデータを収集し、2時点でのブログデータの特性を比較することで、その安定性を検証する。さらにクラスタ分析を用いてブログの書き手の嗜好の変動を検討したうえで、リピータに着目した行動特性を示す。また、景観評価をするために、「見る」という行動に着目して、その情報がルールベースで抽出可能であるのかについて確認を行い、魅力ある観光地形成のための基礎情報の提供可能性を検討する。

2. データの収集と基礎的な特性

(1) ブログデータの基礎的な特性

本研究で用いるデータは、2013年10月～11月にGoogleブログ検索を用いて「八ヶ岳 行った」でヒットしたブログ上位300にあるエントリを収集した。あわせて各ブログのタグとその該当エントリ数を収集した。収集されたブログのエントリが書かれた年別に集計した結果が図-1である。このように2011年以降に書かれた数が半数を占めており、佐々木らが2010年に実施した際にも収集されたブログは半数以下であった。そのため、多くのデータが変化していることが予想されたため、その安定性について検討を行う。

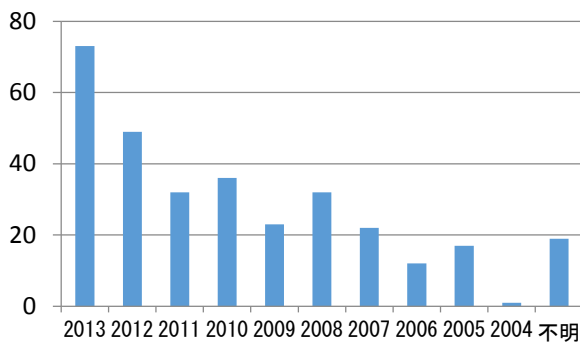


図-1 収集されたエントリの書かれた年の分布

佐々木らが2010年秋に収集したブログデータは、GoogleBlog検索およびYahoo!Japanを用いて同じ「八ヶ岳 行った」をキーワードとして収集されたそれぞれ上位136ブログずつから、重複を除いた192のデータである。これを比較対象として、今回得られているブログデータの特徴を比較で示す。まずエントリの書かれた月別の分布状況を示す。この結果を見ると2010年データと2013年データでは書かれた月の比率に差がないことが読み取れる。わずかに10月と11月の比率に差がみられるが、統計的に有意な差ではない。この変動はブログエントリの収集時期が少しずれているため、それが影響している可能性が指摘できる。

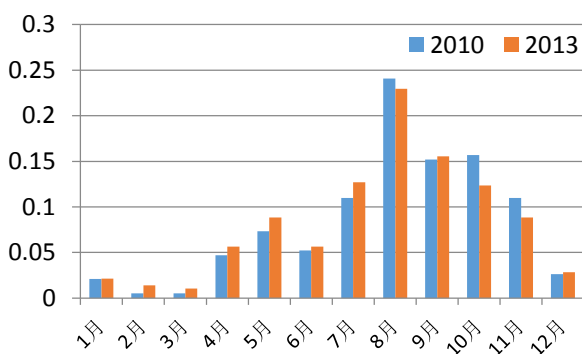


図-2 ブログエントリの月別分布状況

続いて、利用交通手段の判明率について、図-3に示す。2010年は自動車の割合が高いが、2013年は電車の割合が相対的には高くなった。2010年はブログエントリから「車」「駐車場」「高速道路」「電車」「バス」「バイク」「ツーリング」等の等の交通手段と関連するキーワードを抽出し、その前後の文章から移動手段を推測したのに対して、2013年データでは、より簡略化し機械的にルールベースでの手段決定を行った。具体的には「行く」「向かう」等の移動を表す動詞と同じ文に「自動車」「電車」等の移動手段を表す単語が現れた場合に移動手段を推定するというルールで判定を行った。その結果が不明率が高くなった原因である。また、ルールベースの場合に自動車の割合が減少するのは、もともと自動車での観光客の割合が高く、普段から自動車で行くことを前提とした場合には、それをエントリ中で明示することが少なく、手段の単語ルールだけからは、自動車で行くことを当たり前として記述されたような場合には明確には抽出・分類できないと考えられる。

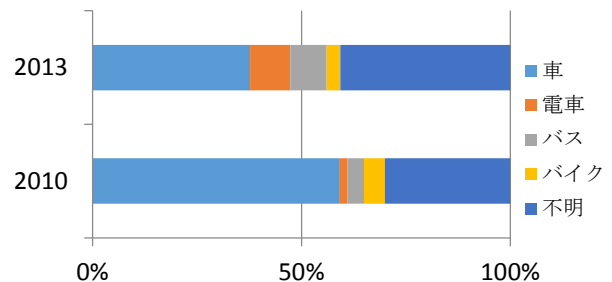


図-3 八ヶ岳までのアクセス交通手段分担率の推計値

続いて目的地の選択率について、固有名詞や施設名がエントリに出てくる回数で検証を行った。2010年のデータで上位20位にあげられた目的地が2013年のデータではどの程度の選択率であったかを図-4に示した。特に2013年データは先に述べたように、コーディングのみでどの程度捕捉できるかを検証することを目的としたため、2010年と違って個別の検証は行っていない。また2010年データは同地域でのアンケート調査との比較のため、コーディングで目的地を抽出したのち、宿泊・日帰りを判定し、宿泊を行っている判断された場合には前後のエントリを確認してそこで得られた他の目的地を追加した。そのような抽出方法の違いによって、2013年は2010年データと比較して狭域の行動になりがちである。そのため諏訪湖・白樺湖といった、一日行動圏としては少し離れた目的地は上位からは消えた。しかしそれを除くと八ヶ岳南麓の観光地では選択率上位の目的地に大きな変動は無く、ワンワンパラダイス、サンメドウズが相対的には順位としては2013年で高くなっているが、それ以外にはほとんど違いが無い。また目的地の選択率が2010年で全般的に高くなった理由は先に述べたように、宿泊と推察

された行動は滞在期間のデータを用いたために目的地数が増加したためと考えられる。

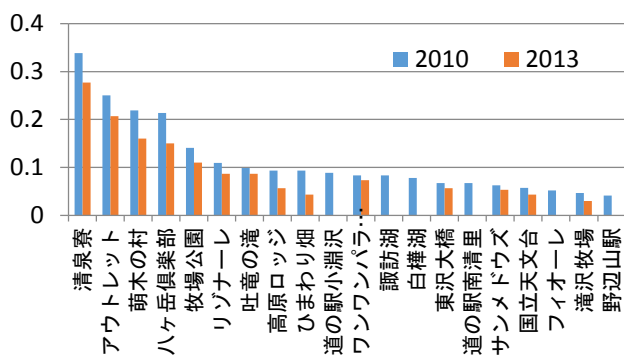


図4 地名・施設名のブログエントリーへの登場率

続いてリピーター、初来訪者の推計を行った。推計方法は「初」「初めて」等の初来訪を類推する語や「〇年ぶり」「また」等のリピーターを類推させる単語検索から大まかな分類を行い、最終的にはエントリーの文章内容を検討して判断した。いずれの記述もないものは不明としてある。この結果を図-5に示す。ここでは2時点での比率に統計的な差は無く、エントリーの内容を確認したために、ほぼ同様の結果となった。

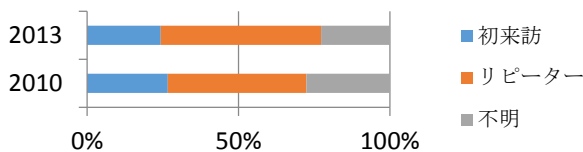


図5 ブログの書き手のリピーター/初来訪者率の推計

ここまでの結果からは、類似の手法を用いたときにはおおよそ同じ結果が得られること、細かな内容を検討すると不明率が低下し、その不明率は対象とする属性に影響されることが示唆される。ただし、目的地に関してはこの2基の間ではほとんど変わりが無く、コーディングルールのみでおおよその目的地が判明可能である。

(2) ブログの書き手の嗜好の変動

先行研究ではブロガーの嗜好をブログエントリーの分類タグをベースに分類し、目的地の選択にその有効性を示している。そこで、本研究では、2013年に収集したデータに同じ手法を用いて、書き手の嗜好に変化が生じているのかを検証した。この時に用いたタグを集約するカテゴリは先行研究と同様として、表-2に示した。これらのカテゴリに属するエントリーの比率を、そのブログの書き手の嗜好を示す特性値として、ユークリッド距離によるクラスタ分析を行い、結合距離の変化を判断してにクラスタ数を決定し、属するブログ数を求めた。

表-1 ブログにあるタグの集約カテゴリ

ガーデニング	イベント	オフ会
インテリア・雑貨	旅行	グルメ
日記	ドライブ・ツーリング	環境
パソコン	買い物	写真
スポーツ	政治・経済	ペット
製品	家族	散歩
映画・TV・本	健康	料理

分類されたクラスタ数は2010年では雑多なカテゴリを含めて7つであるが、2013年では2010年では見られなかったグルメ、家族、政治・経済のことがほとんどのブログが登場し、それらが別のクラスタとして形成された。その結果を図-6に示す。スポーツとドライブ・ツーリングの比率が2010年よりも2013年では上がったが、雑多なカテゴリを含めて大きな違いは無く、より広く分布するようになってきた。特に食べ物に関して特化したブログなどはこの数年スマートフォンの普及に伴って食べ物の写真を中心にブログ・SNSを構成する人が増えていることが指摘され、それによってこのカテゴリが新たに形成されたと考えられる。家族についても同様の傾向があると考えられる。

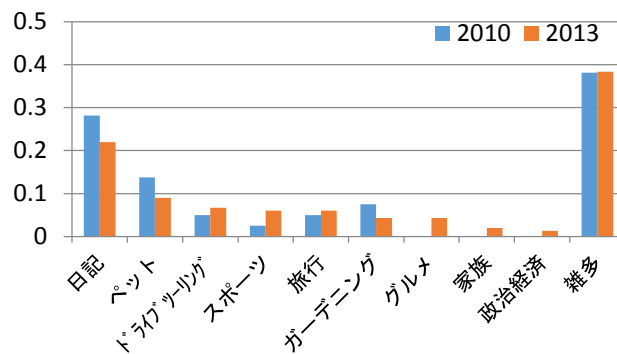


図6 エントリー分類タグに基づくクラスタ

3. 観光行動の抽出と分析

2章ではクロールされたデータが、3年経過時点で半数以上が入れ替わったにもかかわらず、目的地の選択やリピーター率が大きく変わることが無いことを示した。そこで本章では、このデータを用いて観光行動においてどのような活動が行われていたかを検証する。テキストマイニングは基本的には単語に着目して分析を行うことになるため、今回は観光における活動を示す単語に着目した。表-2は対象とするブログエントリーデータにおける名詞、動詞、形容詞の頻出単語上位20位を示す。太字は動詞を表し、斜体は場所に関わる単語である。ハケ岳が一番多

いのは当然として、上位には一般的な動詞や名詞が続いている。

表-2 対象データでの頻出単語上位20

抽出語	出現数	抽出語	出現数	抽出語	出現数
ハケ岳	1568	道	398	高原	354
行く	1071	時間	393	来る	331
思う	642	前	383	観光	330
見る	591	写真	381	入る	330
食べる	585	車	372	言う	326
旅行	425	今回	371	場所	322
店	415	良い	357		

この単語の使われ方を分析するのであるが、2章ではブログの書き手をリピーター、初来訪者に分類した。これらの特性で目的選択等に変化があるかを確認するため、各エントリーにそのフラグを付けた上で、登場する地名に着目して対応分析によって2次元化したものが図-7である。図中のリピーター、初来訪および不明がプロットされ、それらと関連の深い地名が近くに位置されている。ここから、リピーターが多く訪ねる場所や初来訪者が多く訪ねる場所が示される。例えば、ハケ岳倶楽部、高原ロッジなどはリピーターからの距離が近い一方、シャトレレーゼ、白駒池、東沢大橋、ひまわり畑などは初来訪と距離が近くリピーターとは距離が遠くなっている。これは各施設の特徴にも依存するが、リピーターは自然景観・環境を楽しむスポットよりも、ロッジやハケ岳倶楽部等の飲食を伴ったバードウォッチングやガーデニングが行えて、滞在時間の長い施設が目立つ。その一方、例えば東沢大橋はハケ岳の眺めが非常に良い景観スポットであるが、リピーターからは遠くなっている。現在多くの観光施設がリピーターの獲得を目指しているが、複数回の来訪でも飽きさせない工夫が行われる施設にリピーターが多いと推察できる。続いて同じ分析を動詞に対して行ったものを示したのが、図-8である。ここではリピーターと初来訪者ではどのような活動の違いがあるかを見た。例えば「食べる」という動詞は出現回数上位であるが、初来訪・リピーターと等距離で同じように使われることが推察される。一方、初来訪者との距離が近い行動として「撮る」があげられる。初来訪者は景色にひかれるなどで写真を撮影する機会が多いと思われる。その一方、リピーターに近い特色のある動詞は楽しむ、違う、走る、登る等である。山岳地域でもあるため、登るという単語や走る、というスポーツを連想させる動詞や、「違う」という何かとは異なることを体験したことが想像される。このような活動を用意することが、リピーター

の獲得に影響していると考えられる。

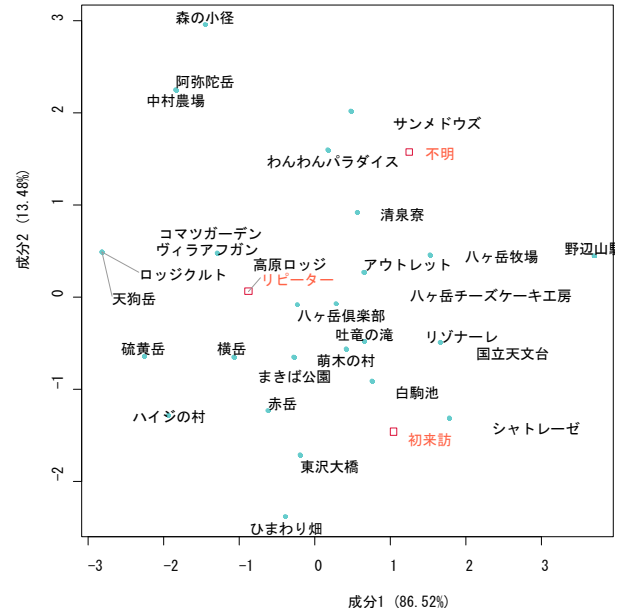


図-7 リピーター、初来訪者と訪問先の関連性

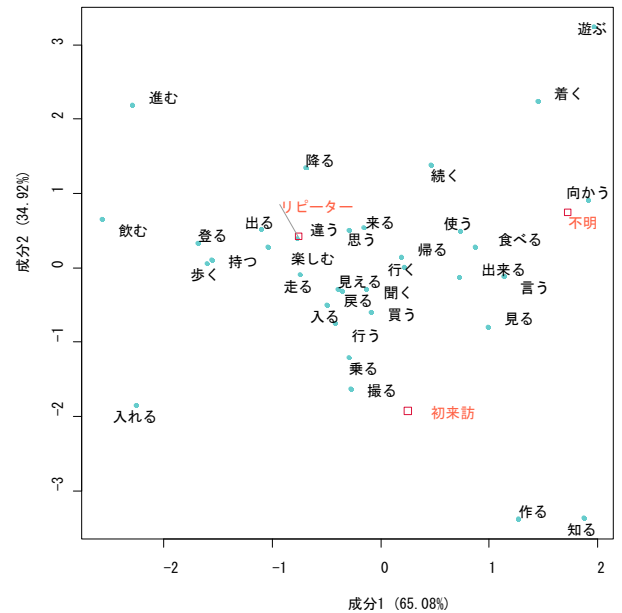


図-8 リピーター、初来訪者と活動の関連性

より具体的に特徴のある行動として「撮る」「食べる」「買う」をあげ、先に行ったブログカテゴリによるオーナーの嗜好との関連性を見たのが図-9である。これらの図の丸と四角の大きさはそれぞれの単語の出現数とカテゴリ分類に属するオーナーの数を示している。これから「撮る」はドライブ・ツーリングおよびガーデニングのカテゴリと近く、「食べる」はグルメはもちろんのこと、雑多なカテゴリとも近くよくつかわれる。一方これらの単語とほとんど無縁なカテゴリとして政治・経済がある。またスポーツカテゴリもこれらの単語はあまり使われず、より特化した行動(動詞)が使われている。

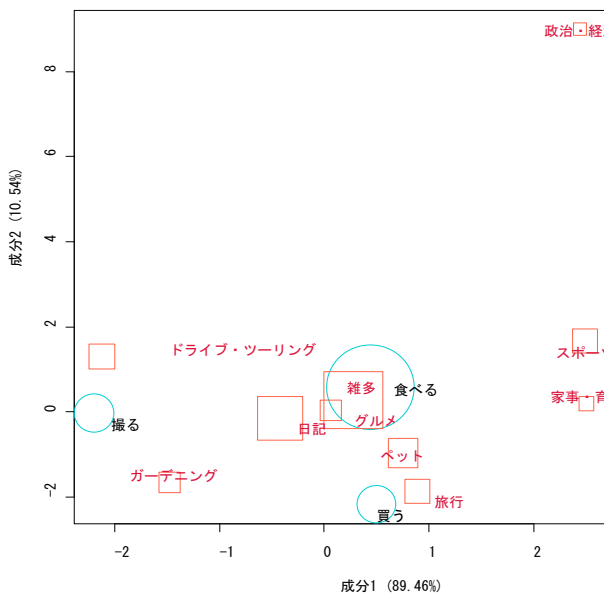


図-9 ブログの書き手の嗜好と行われた活動

4. 観光地における景観評価への適用可能性

(1) テキストマイニングと景観分析

観光地の地域全体の魅力度を高めるためには、定点からの景観のみでなく移動時の風景・移動時の景観の魅力向上も欠かせない。「移動することで変化する景色」「徐々に変わっていくデザイン」はシーケンス景観として建築分野などで多くの研究成果がある。多くの場合、建築構造物の線や面、光などの空間構成要素などのコントロール可能な様々な要素の関係性によって表されている。ブログにおいては頻出単語に「道」が出るように、観光施設間の移動についても多くのコメントがされている。そこで、本章では観光ブログのテキストマイニングによる移動景観についての分析可能性を検証する。特に景観を言語で記述する場合には、その行動が動詞、印象などは形容詞・形容動詞、そして対象物は名詞や固有名詞で表されることになる。本章ではこの点に着目して、単語をベースとした観光行動における意識・行動の導出として、移動景観分析可能性を検討する。

(2) 景観要素の抽出

景観評価を抽出する際には、対象物が何か、また観測場所はどこか、対象物への評価がどうであったかを検証することになる。しかしながらテキストマイニングは一般にルールベースで発見を行うため、景観評価が複雑な表現を伴う場合には、的確に抽出することが困難になる。仮に地名や対象物が出現していても、それらが景観として得られたものかどうかを判断するためには前後の文章との内容の比較が必要である。そのため、まずは視覚に関する動詞として「見る」「見える」「眺める」「撮る」

の4つの動詞に着目した。これらの単語がブログテキストに出てくる場合の多くは、これらは書き手が何か記録に残したい対象物があると考えられる。よってこれらの単語で文書検索を行い、景観として評価されているかを確認し、その場合には対象物、観測場所、対象物への評価の3点を集計した。集計に際しては実際に視覚に関する動詞が使用されている文章とその前後に目を通して判定して集計を行い、結果を表-3、表-4、表-5にそれぞれ上位20位までをまとめた。

表-3 景観として得られた対象物

場所	件数	場所	件数
八ヶ岳	67	自然	3
富士山	50	滝	3
アルプス	14	虹	3
星	9	朝焼け	2
景色	8	清里駅	2
赤岳	5	下界	2
甲斐駒ヶ岳	5	高原野菜畑	2
紅葉	5	白駒池	2
東沢大橋	5	諏訪湖	2
ひまわり	5	彫刻の森	2

表-4 景観の観測場所に関する単語

場所	件数	場所	件数
東沢大橋	32	ビーナスライン	5
道中	26	アウトレット	4
清泉寮	12	観音平	4
野辺山	9	木賊峠	4
明野ひまわり畑	7	国立天文台	4
飯盛山	6	まきば公園	4
美しい森	6	赤岳	3
霧ヶ峰高原	6	阿弥陀岳	3
平沢峠	6	サントリー	3
清里	5	サンメドウズ	3

表-5 対象物への評価・印象の単語

印象	件数	印象	件数
無し	73	楽しい	5
きれい	27	くつきり	3
良い	21	興奮する	3
残念	13	すごい	3
美しい	11	素晴らしい	3
最高	11	美味しい	2
絶景	11	かわいい	2
雄大	7	シルエット美人	2
大きい	6	素敵	2
感動	6	バッチリ	2

対象物では八ヶ岳・富士山・アルプスをはじめとして山岳地であるために山が対象として多く評価されている。それを行った場所は目的地としても多く上がった東沢大橋、清泉寮に加えて、「道中」があげられた。これは移動中に見たものについて評価がなされたケースを道中と表現したが、移動中に見た景観についてのコメントは2番目に多くなることが示された。この他にもビーナスライン、平沢峠、木賊峠など道路が地点の単語はほかにも見受けられる。またその評価は「無し」が最も多く、他にも「きれい」「良い」等の単語が得られた。このように実際のテキストに個別にチェックを行い、対象物や評価を抽出することは可能であるが、実際のテキストマイニングでは、個別のチェックを常に行うことは難しい。そこで、様々なルールを導入してその適切な導出可能性を検討した。具体的には対象を適切にルールで発見できるのかを取り上げてその結果を示す。

コーディングルールとしては「見る・見える」「眺める」「撮る」の動詞に着目し、いくつかのルールを用いて検討してもっとも適切に対象物を取り出せたルールを採用した。それは、例えば「見る」の場合には、連用形であるもしくは「ない」という否定を伴った場合には取り除くというルールであった。そこから連用系の対象になる（固有）名詞を取り出した。その上位10位までの捕捉率を示したのが表-6である。表中のコーディングルールとは先に示した対象物の単語を特定するルールで抽出された回数であり、ベースは表-3に示した目視で確認した対象物の出現数である。上位10位までのルールによる抽出率はベース比で60%以上である。ただし複数回記述された場合それをそれぞれカウントするため、100%を超えたり、また実際のブログエントリ数の割合ではないことに注意が必要である。

表-6 ルールによる対象物の導出可能性

	コーディングルール	ベース	補完率(%)
八ヶ岳	64	67	96%
富士山	41	50	82%
アルプス	12	14	86%
星	20	9	222%
景色	27	8	338%
赤岳	4	5	80%
甲斐駒ヶ岳	3	5	60%
紅葉	10	5	200%
東沢大橋	3	5	60%
ひまわり	8	5	160%

以上のように、対象物については、動詞を特定した時点でおおむね補足可能であり、何が「見る」「撮る」の対象となっているのかはルールベースのコーディングでおおよそ把握可能である。同時に、

印象表現は、繰り返しがほとんどないため、形容詞を指定することで、おおむね抽出可能であると考えられる。ただし、先に示したように「評価無し」が非常に多く、あったとしても複雑な表現が単純な「きれい」や「良い」などに集約されてしまい、景観評価としてどこまで有効な分析になりうるかは今後の課題となろう。

5. おわりに

本研究では、2010年に行ったブログマイニングによる観光行動分析を発展させ、3年後に同様のブログクローリングを行って、得られたデータの特性をもとにその安定性を確認した。その結果、半数以上が新たなブログから得られたにもかかわらず、おおむね同様の属性や目的地選択率が得られることが明らかになった。同様にタグをベースとした書き手の嗜好をクラスターに分類したが、そのクラスターは前回よりもより広く分布するようになっていた。統計的に有意なほどの差ではないが、以前よりもブログの書き手が広がるにつれて、嗜好が拡散して来ていると考えられる。

続いて、2013年に得た新しいデータで、リピーターと初来訪者に注目して、固有名詞および動詞で両社との類似性を見たところ、リピーターは自然景観よりも特有の活動が可能な施設との関連が近く、また活動では、初来訪者が多く「撮る」という単語を用いているが、リピーターではほとんど関連性が無いなど、動作を表す動詞においても差が見られた。書き手の嗜好と動詞の関連性を見たところ、ドライブ・ツーリングクラスターやガーデニングクラスターでは「撮る」という動作が多くつかわれているが、「食べる」と「買う」はお互いに近くに位置し、かつ「雑多」「日記」や「グルメ」、「旅行」といった一般的なカテゴリのエントリにおいて多く用いられることが示された。

最後に観光魅力度向上に欠かせない、景観分析への適用可能性を見るため、「見る」という動作を表す4種類の動詞がどのような単語を伴っているかを示した。またそれが純粹にルールである程度正確に抽出可能であることを示し、景観評価の大まかな傾向を機械的に抽出することの可能性を示した。一般性を持たないブログマイニングではあるが、個人の嗜好や特性を把握することができ、そこからリピーターを増加させるためのヒントや、景観評価がどこでどのように行われているのかを把握することができ、魅力的な観光地形成には有益であると考えられる。

ただし、観光行動がブログに書かれる場合には、一般にテキストだけでなく、写真とセットで書かれることが

多い。そのためにテキストマイニングだけでは十分でなく、画像を含めたブログ全体の検証を行っていくことが望ましいと考えられる。

参考文献

- 1) 北村隆一・森川高行編著，交通行動の分析とモデリング，技報堂出版，2002.
- 2) 高尾秀樹，末祐介，山本清二，京阪神都市圏における休日の観光地間回遊行動パターンに関する基礎的分析，土木計画学研究・講演集，36，2007.
- 3) 国際交通安全学会編，魅力ある観光地と交通，技報堂出版，1998.
- 4) 国土総合研究機構観光まちづくり研究会著，観光まちづくりのエンジニアリング，学芸出版社，2009.
- 5) 西村幸夫編著，観光まちづくり，学芸出版社，2009.
- 6) 佐々木邦明，紀藤舞華，山崎慧太，ブログマイニングからの行動データの抽出・分析可能性とアンケート調査との比較，土木計画学研究，講演集，No43，CD-ROM，2011.

A STUDY ON EXCURSION AND LANDSCAPE EVALUATION USING BLOG MINING OF THE YATSUGATAKE AREA

Kuniaki SASAKI and Tomoya SAKURADA