

## ブログマイニングによる観光客の意識と行動の導出可能性に関する研究

山梨大学工学部土木環境工学科  
学生会員 ○櫻田 朋也

### 1. はじめに

観光行動において、観光客の個々の行動やそれに付随する行動などを把握することは、観光地の現状把握や課題の早期発見へと繋がるため非常に有効である。また、その場合の調査手法としてはアンケートが一般的に用いられる。しかしながら、観光行動は稀な行動であるのに加えて、その対象が観光客である以上配布場所も観光地であり、居住地ベースでの調査が行えず、かつその内容も配布場所に依存してしまうという問題が生じる。本研究では特定テーマの下で収集した観光ブログを対象にテキストマイニングを行い、観光客の意識と行動の導出可能性について考察する。

### 2. 使用データと分析手法

#### 2.1 使用データ

Google ブログ検索にて検索ワード「八ヶ岳 行った」でヒットした観光ブログ 300 件を収集し、テキストファイルとしてひとつにまとめたものを使用する。

#### 2.2 分析手法

本研究ではフリーソフト Kh-coder と SPSS を用いて分析を行う。Kh-coder にてテキストマイニング、対応分析を行い、階層的クラスタ分析のみ SPSS を利用した。

#### ・テキストマイニング(形態素解析)

文章を意味のある最小の語句(形態素)にそれぞれ分解して集計し、語句の出現頻度や共起性などより情報を抽出する解析方法である。利点としては、自由記述の確認作業の負担大幅減、生の意見の抽出、要点整理などが特徴的である。例文として「お待ちしております。」という一文を形態素解析した結果を表 1 に示す。

表 1. 形態素解析例

文字列	読み	原型	品詞の種類	活用の種類	活用形
お待ち	オマチ	お待ち	名詞-サ変接続		
し	シ	する	動詞-自立	サ変・スル	連用形
て	テ	て	助詞-接続助詞		
おり	オリ	おる	動詞-非自立	五段・ラ行	連用形
ます	マス	ます	助動詞	特殊・マス	基本形
。	。	。	記号-句点		

#### ・対応分析

コレスポンデンス分析とも呼ばれ、クロス集計結果を対象とし、2次元(あるいは3次元空間)上に、似た値を持つ行および系列同士が近くなるように座標をプロットする手法である。プロット点間の相対的な位置関係より、行や列間の関係を探索的に解釈することが出来る。

#### ・階層的クラスタ分析

データの集団を、各々の有する分散や相関、類似度や距離といった情報を用いてグループ分けする手法。その中でも、近似している個体順にグルーピングする手法を階層的クラスタ分析と呼ぶ。本研究では各記事の書き手のホームページへ行き、所有しているカテゴリ毎の記事数を重みとし、それらを趣味選択の傾向として集計した。その後趣味選択が比較的同傾向にあった書き手を最終的に10種類のクラスタに分割した。

### 3. 事例研究

#### 1) 基礎的分析

全体の概要把握として、記事全体における頻出語句の上位 20 語を表 2 にリストアップした。抽出対象は名詞・形容詞・動詞である。図より、名詞として上位に挙がる語句に「八ヶ岳」・「旅行」・「観光」といったようにテーマに沿った内容が顕著となった。また動詞に見る全体の行動には食に関する記述が多かったことが窺える。

形容詞には文章表現上、あるいは好印象であったことを表す「良い」が多く、最上位では食べ物についての

感想の記述が多いという結果になった。

また、全体への対応分析を行うことで採択される行動の概要の把握も行うことが出来た。ここではさらに行動選択の詳細な分布を得るために多次元尺度法にて主要な動詞(本文中にて 120 回以上出現したものをプロットし、図 1 に示した。この結果より、本文中での登山などに

当たる「登る」、工房などにおける体験を示す「作る」といった行動は比較的稀な行動であり、「見る」「食べる」

「撮る」といった行動が全体的に採択されやすい傾向にあることが窺える。

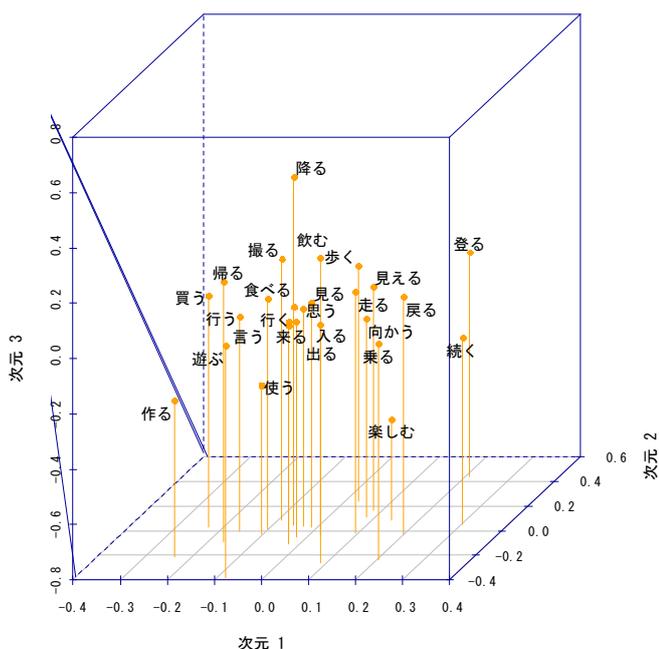


図 1. 多次元尺度法プロット結果：動詞の分布

### 2) 趣味による行動特性の差異の検討

趣味選択の傾向によって差異が無いかを検討するに当たり、まずは階層的クラスター分析によって 300 人を適当数のクラスターに分割した。クラスター分析結果および各クラスターの主要カテゴリを表 2 に示した。また、各クラスター毎の行動の傾向などの概要を、対応分析にて図 2 に示した。各クラスターの分析により、傾向としての特徴が顕著であったのはクラスター 1、3、4、9 であった。共通する傾向として、それぞれの趣味を主要素とした観光地を多く訪れており、それらについても概ね好印象との記述が多かった。来訪月については大半のクラスターが全体の傾向と同様に夏に多く冬に少ないという結果であり、クラスター 4 のガーデニングのみ来訪が温暖な季節に集中しており、

表 2. クラスター分析結果

クラスター 1 (n=18)	スポーツ(76.7%)
クラスター 2 (n=18)	旅行(81.4%)
クラスター 3 (n=20)	ドライブ・ツーリング(54.8%)、日記(8.9%)
クラスター 4 (n=13)	ガーデニング(64.5%)、旅行(8.9%)
クラスター 5 (n=4)	政治・経済(99.2%)
クラスター 6 (n=13)	グルメ(35.6%)、写真(29.8%)
クラスター 7 (n=6)	家族・育児(47.7%)、旅行(21.1%)
クラスター 8 (n=115)	10%前後多数(=雑多なカテゴリー)
クラスター 9 (n=27)	ペット(67.4%)、旅行(9.6%)、日記(9.1%)
クラスター 10 (n=66)	日記(63.3%)、旅行(10.5%)

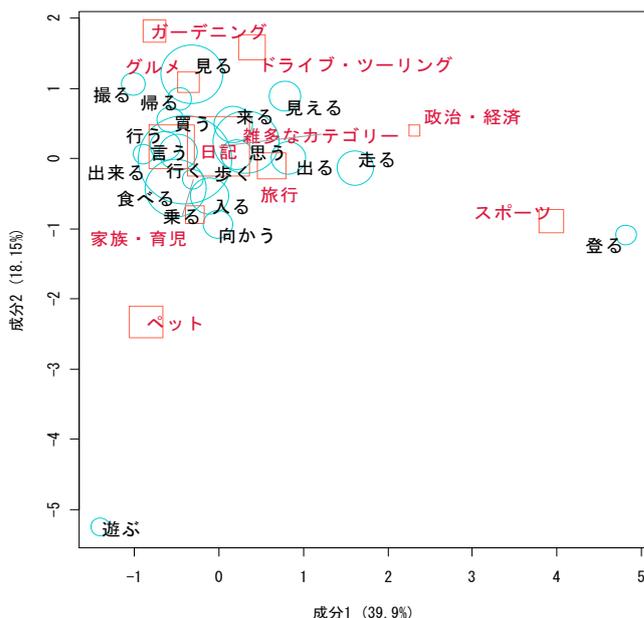


図 2. 各クラスター毎の行動の傾向

シーズンを考慮すれば妥当な結果であることが窺える。

対応分析より、各クラスターにおける行動の傾向が示された。例えば、「登る」という行動は全体からすれば比較的稀な行動であることが分かり、その近辺のクラスターは「スポーツ」である。これは、土地柄からマウンテンスポーツが採択されることが多いための結果であることが窺える。それ以外の多くのクラスターでは比較的行動が共通しており、多くの動詞と近い範囲に出現しているのが分かる。カテゴリ「ペット」については、他のクラスター同様に共通した行動が多い一方で「遊ぶ」という表現を文章中で多く用いているために、若干外れた位置に分布されていた。

### 3) 印象の導出と要因分析

次に観光客の意識、つまり観光行動に対しての印象についての分析を行った。行動の導出の際には動詞に着目したが、印象については形容詞、並びに形容動詞

の出現に焦点をおいて分析を行う。まず対応分析にて、各クラスター毎の印象傾向の有無を検討したが、結果はあまり集約されず、傾向を示す要素は確認できなかった。わずかながら、スポーツのクラスター周辺には「きれい」「素晴らしい」といったような景観への印象の記述が見られた程度である。しかしながら、図の中心付近に「残念」というネガティブな表現が確認された。そこで、次はネガティブな印象とその要因の導出に関してマイニングを行った。日本語の特性上、その表現がポジティブ、あるいはネガティブのどちらに属するかは単語のみでは判断が困難であり、否定助動詞（「～ない」など）との組み合わせも考慮すると、文章の前後も併せて判断しなければならない。そこで、本研究ではネガティブである表現を「残念(形容動詞)」に絞り、これらとの共起関係や語の出現状況より、観光行動におけるネガティブな要素の導出を試みた。抽出語検索より、形容動詞「残念」と共起関係にある語のうち出現の多かったものを表3に示した。

天気	15	富士山	5	本当に	2
見える	15	見る	5	見れる	3
雨	11	掛かる	3	曇る	3
雲	8	台風	3	混む	3

表3. 形容動詞「残念」と共起関係にある語

表3より、ネガティブな要素の多くは天候に関するものであることが窺える。抽出された動詞のうち「見える」「見る」などは否定助動詞との組み合わせであり、その行動が行えず残念であるという表現であった。特定の観光資源についてのネガティブな印象は得られず、ネガティブな要因として抽出されたのは「悪天候」と「想定していた景観・景色が見られなかった」の2点である。

#### 4) 景観分析へのアプローチ

景観分析へのアプローチとして、特定の動詞に着目した分析も行った。本来、景観評価を行う場合にはアンケートによるSD法や主成分分析、写真分析などの手法を用いる。本研究ではテキストからの抽出によって分析を行うため、得られた景観の詳細な評価などは困難であると考えられる。しかしながら、行動の記述がされていればそれらを集計することは十分に可能である。故に本研究では視覚に関する動詞に着目し、景観として得られた対象物、観測場所、対象物への評価の

三点を目視により集計したものをベースとし、他側面からの分析と見解の提示、効率的な集計の提案などにより、マイニング効果による景観分析の可能性について論ずる。

まず行ったのが、コーディングルール利用による景観行動の簡易集計の提案である。先述のように、景観に関する行動はあらかじめ集計してあるため、そのベースに対してコーディングルールがどの程度の補完性を有することが出来るかを検討した。またコーディングにおいて、ベースでの集計は確実にその対象となる景観が得られたケースのみを集計対象としたため、見られなかった分のカウントを防止するために、否定助動詞との組み合わせを除外するようにルールを組んだ。また、視覚に関する動詞として今回集計対象にした語は「見る」「見える」「眺める」「撮る」の4つである。

また先だった問題点として、コーディングルールによる検索は文章単位で行われ、かつKh-coderでは1文章を句読点、あるいは改行(段落)によって判別している。

そのため、「〇〇を見た」といったように、目的語にあたる語が同文中に出現しなければ、その出現をカウントすることが出来ないという問題が存在する。以上の問題を加味しつつ、作成したコーディングルールでの抽出結果とベースとの比較を表4に示した。

景観対象	ベース	コーディングルール利用	補完率
ハヶ岳	67	58	0.86
富士山	50	38	0.7
南アルプス	14	11	0.74
赤岳	5	4	0.8
甲斐駒ヶ岳	5	2	0.4
東沢大橋	5	3	0.6
清里駅	2	1	0.5
高原野菜畑	2	1	0.5
白駒池	2	1	0.5
諏訪湖	2	1	0.5
彫刻の森	2	1	0.5
アウトレット	1	1	1
霧ヶ峰高原	1	0	0
金峰山	1	1	1
最高標高地点	1	0	0
白樺湖	1	1	1
千曲川	1	0	0
蓼科湖	1	0	0
蓼科山	1	1	1
中山峠	1	1	1
瑞穂山	1	0	0
三つ峠	1	0	0
南駒ヶ岳	1	0	0
横岳	1	1	1
合計	169	127	

平均62%程度  
(上位10まで)

全体の補完割合  
・・・52.9%

表4. ベース結果とコーディングの比較

先述の問題点より、目的語に当たる景観対象は多くを抽出できた一方で、その観測場所や印象については大半が同文中に存在せず、コーディングルールでの抽出は極端に数が少なく不可能であると判断出来たため、

表4には景観対象のみの抽出結果を示した。また、抽出対象は固有名詞のみとした。

また、分類されたクラスターによって景観に関する行動に差異が生じるかについて検討したものを、対応分析にて図3に示した。

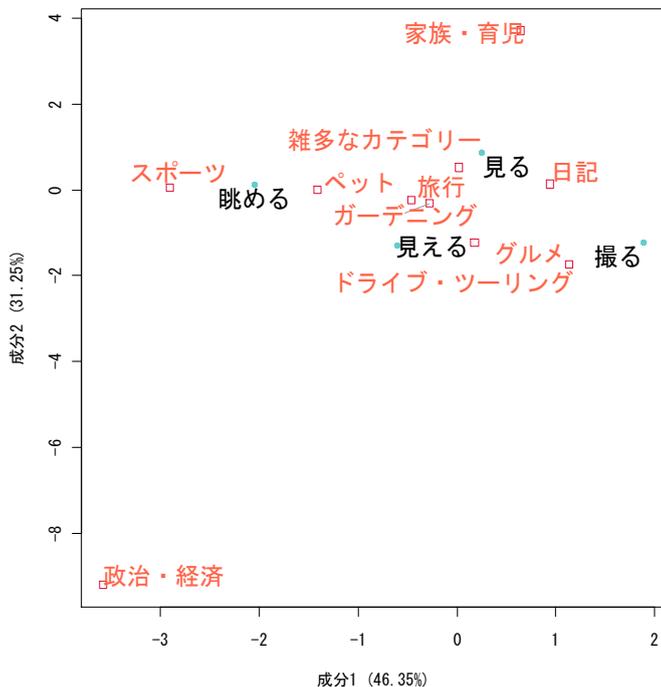


図3. 各クラスターと景観に関する行動比較

表4より、コーディングルールの補完率はおよそ53%という結果が得られた。視覚に関する動詞と固有名詞との組み合わせを本文中より抽出した場合、つまり景観対象とされるものが何なのかを得たいとき、およそ半分程の精度であればそれらを特定出来ていることが窺える。また、景観対象とされるものの多くは遠方の山々や観光資源、あるいは観光地そのものであり、加えてブログという特性上か、それらに対する印象に関しても比較的単純な表現が多く、複雑な表現が避けられている印象が得られた。また図7より、それぞれの動詞とクラスターの共起性から何点か特徴が窺える。例えば、比較的中心より外れている「スポーツ」のクラスターは、山岳からの「眺め」を多く記述するために動詞「眺める」に近い位置にプロットされる。同様に、行動として「写真撮影」は「ドライブ・ツーリング」のクラスターに属する書き手に多いという傾向が窺える。

## 5. おわりに

ブログマイニングによって観光客の行動や意識の導出に関する分析をいくつか行った。アンケートとの大

きな違いとして、回答者が実際に行った行動をブログマイニングでは動詞という形で得ることが出来るので、行動分析の分野において今回用いた手法は非常に効果的に感じられた。一方で意識、つまり観光行動に関する印象については頻出の形容詞および形容動詞から分かるように、記述の形式が自由である分単調なものが多い。また、今回は集計対象であるテーマが「特定地域の観光」と非常に範囲が広く、特定の観光要素に対する評価などが十分に行えなかった。

しかしながら、最も来訪が多かった観光地のみを対象に分析を行ったところ、名物の特定や評価・印象、行動の傾向などが得られたので、特定観光地のみを対象とした場合にはより詳細な分析が可能であると考えられる。

## 参考文献

1. 紀藤舞華：観光行動調査とテキストマイニングの相互補完性に関する研究 山梨大学卒業論文 2011
2. (株)NTT データ数理システム:テキストマイニングスタジオ

<http://www.msi.co.jp/tmstudio/whyTextMining.html>

3. 東北農業研究センター：テキストマイニングを用いた住民による都市イメージの分析 2010
4. 小路剛志ら：景観評価指標を用いた都市河川の景観分析 国土技術政策総合研究所 2005
5. 上條友也：テキストマイニングを用いた観光対象としての工場景観の特徴に関する研究 千葉大学卒業論文、2009