

観光ブログマイニングによる観光行動調査の補完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 ○紀藤 舞華
山梨大学 大学院医学工学総合研究部 正会員 佐々木 邦明

1. はじめに

観光地に訪れた観光客の周遊行動の特性を把握することは、観光地の現状や課題点を知るためには重要である。その観光客の周遊行動を把握する調査方法として、アンケート調査が用いられる。観光行動は希な行動であるため、居住地ベースの調査は困難であり、主に観光地での配布によることが多く、配布場所の選定にサンプル構成が依存する。また、その内容がアンケート設計者の意図に沿って調査が行われる一方、インターネットの普及により、ブログ（blog ; Weblog）サービスが充実してきている。ブログとは、書き手の個人的な体験や日記、または関心をもった製品やニュースについて書き手なりの意見や感想などのコメントを書いた記事を Web 上に掲載するサイトである。ブログは使い方が比較的容易であるため、一般の多くの人々からの情報発信が盛んになっている。その発信された情報の中には、リアルタイム性のある新鮮な情報や「生の声」といったものが含まれている。こういった情報をブログから抽出する研究が行われており、ブログマイニング手法という。現在企業では製品開発や企業活動にブログマイニング手法が活用されている。⁽¹⁾ ブログマイニング手法とはブログの文章データを単語や文節で区切り、それらの出現頻度などを解析することで有用な情報を取り出す分析方法である。

本研究では、ブログマイニングによる観光行動調査を行い、アンケート調査等の補完の可能性について検証することを目的とする。

キーワード：ブログマイニング、観光行動調査

山梨大学土木環境工学科交通工学講座

(〒400-8511 甲府市武田 4-3-11 Tel/Fax 055-220-8671)

E-mail : t07c024@yamanashi.ac.jp

2. アンケートとブログマイニング手法の比較

観光ブログをブログマイニング手法で分析するにあたって、アンケート調査とブログマイニング手法にはどのような違いがあるかについて説明する。

	アンケート	ブログマイニング
サンプリング	観光スポット等でアンケートを配布	ネット上にあるブログより抽出
対象者	配布場所へ訪れた人 アンケートに回答した人	インターネットが使える環境 ブログを書いている
調査項目	アンケートの設計者に委ねられる	ブログに記載されている内容

表1 アンケートとブログマイニングの特性比較

表1に示したように、アンケートは一般に、ある観光スポットでアンケートを配布することになる。つまり、ある観光スポットに訪れていない観光客はアンケートの対象外になるので、配布場所は特に目的の選択を考えた場合には、非常に重要であることが言える。また他の調査項目は、アンケート設計者に委ねられるので、得たい結果は得やすいが、分析を発展しにくいといえる。

ブログマイニングでは、Web 上にあるブログの中から抽出なので、配布箇所には依存しないが、もとも「ブログを書く」という行為自体に、インターネットが使える環境にあり、ブログを開設しているという前提条件があるので、その時点で母集団を代表していない。しかし、テキストマイニングが一般化した現在では、ブログからはアンケートからは得られにくい結果を得ることができると言える。

3. 使用したデータ

検索エンジンから検出されたブログで表示された順番に上から 136 記事のデータを得た。

3.1. 検索ワード

今回対象となる観光地は、山梨県と長野県にまたがる八ヶ岳南麓地域を対象にしているので、まずは「八ヶ岳」で検索を行った。しかし、「八ヶ岳」のみでは八ヶ岳に関する HP が多数出てきて、ブログの検出は難しかったことから、「八ヶ岳」と「行った」で検索を行った。

3.2. 検索エンジン

ブログを検索するにあたって、どの検索エンジンを用いるかによって結果に違いが出てくる可能性がある。図 1 のネットレイティングス⁽²⁾による 2010 年 4 月現在の日本の検索サービス利用動向の調査では、Yahoo! と Google の検索エンジンで 90% を占めていることがわかる。このことから、Yahoo! と Google で検索を行うことにした。

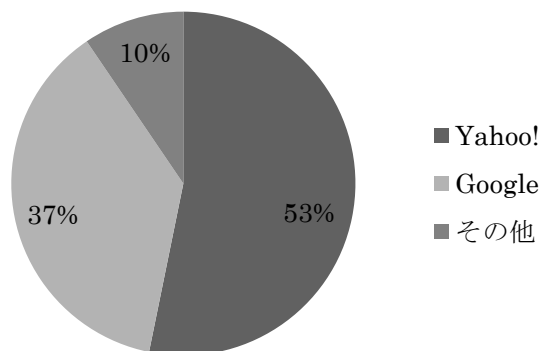


図1 日本の検索サービス利用動向(2010年4月)

また、Yahoo! と Google ではそれぞれ異なった検索結果が出てくる。これは、検索方法が違うからである。検索方法には、一般的にロボット型とディレクト型がある。Yahoo! は、ロボット型とディレクト型を併用していて、Google はロボット型で検索を行っている。今回は、それぞれの検索エンジンで 136 記

事を収集した。そのうち、重複したのは 80 記事 (58.8%) であり、半数以上のブログは、同じキーワードで共通して検索された。今回はこれを統合したものを分析対象とする。

4. 分析方法

ブログの分析は、KHCoder⁽³⁾で行う。KHCoder は、テキストマイニングを行うフリーソフトで、文章データを形態素解析し、その結果を使用して、対応分析・多次元尺度構成法・階層的クラスター分析・共起ネットワークといった分析を行う。

4.1. 分析結果

分析結果では、基礎集計と季節別による分析、アンケートとの比較を行った。ここで言うアンケートとは、平成 21 年秋に八ヶ岳南麓地域で実施されたアンケート⁽⁴⁾のことであり、八ヶ岳南麓地域へ訪れた観光客の周遊行動特性の把握を目的として行った調査である。

図 2 に月別の訪問者数を示した。来訪日時はブログの掲載日時や内容から判断した。八ヶ岳南麓地域は、8 月に来訪者が 1 番多く、冬は少ない傾向にあることが言える。

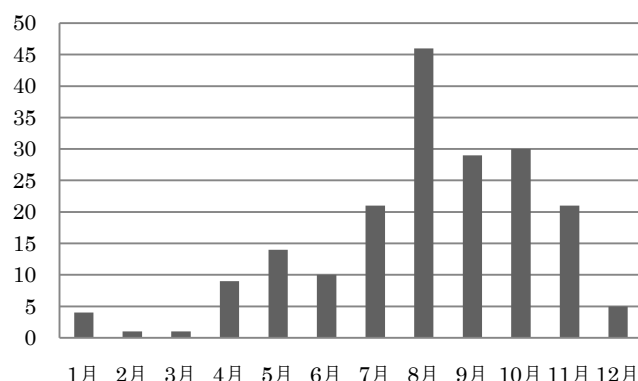


図2 月別訪問者数

図 3 に訪問者の居住地の割合を示した。居住地はブログの内容を用いて判定した。八ヶ岳南麓地域へ訪れるのは、関東が 6 割以上を占め、次いで東海、関西という結果になった。

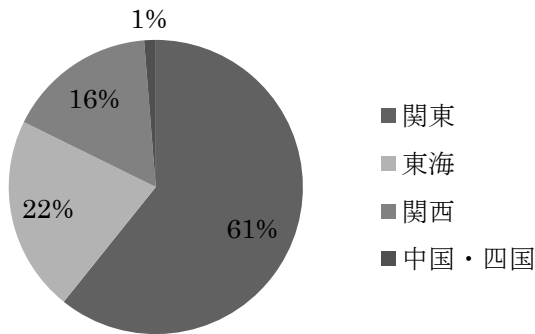


図3 訪問者の居住地

図5には来訪者が日帰りか宿泊かどうかについて示した。日帰り宿泊別もブログの内容を用いて判定した。宿泊が7割と日帰りを大きく上回っていることがわかった。

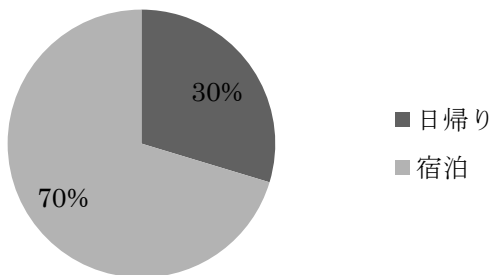


図5 日帰り宿泊別

図6に訪問客がよく訪れている訪問地を20箇所示した。1番訪問者が多かったのは、清泉寮であった。次いで、ハケ岳リゾートアウトレット、萌木の村、ハケ岳倶楽部、まきば公園という順番になった。

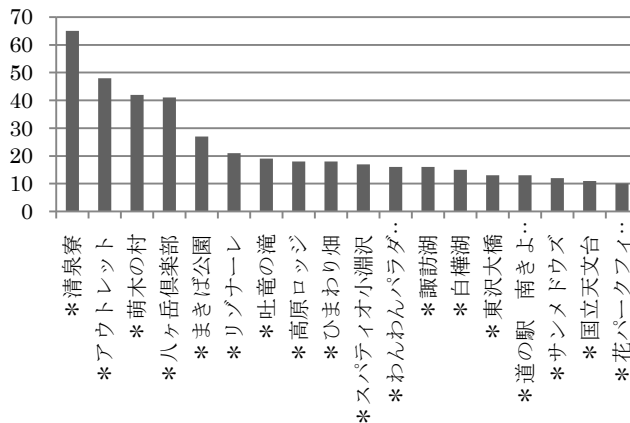


図6 訪問地上位20箇所

アンケート調査では、調査を行った日のデータしか入手することができないが、ブログマイニングでは、いろいろな時期のデータの入手が可能である。この利点を利用して、図2の月別訪問者数から、サンプル数が確保できている7~11月を夏(7月・8月)と秋(9月・10月・11月)と分けて、分析を行った。

まずは、訪問客が多く訪れている場所の割合を示した。夏は、三分の一湧水・シャトレゼ白州工場・花パークフィオーレといった涼しい場所や体験できる場所に来訪客が多かった。秋は、東沢大橋・諏訪湖・白樺湖といった自然が綺麗な場所に多い。

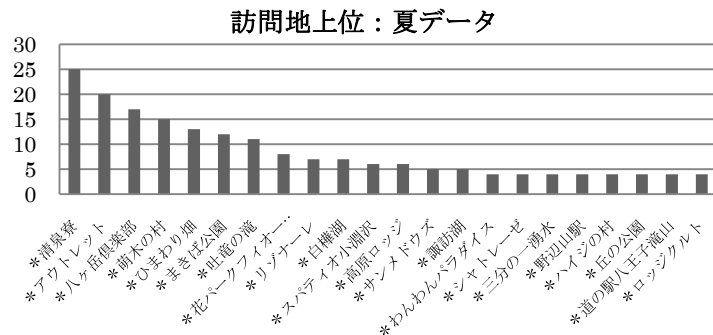


図7 訪問地上位(夏)

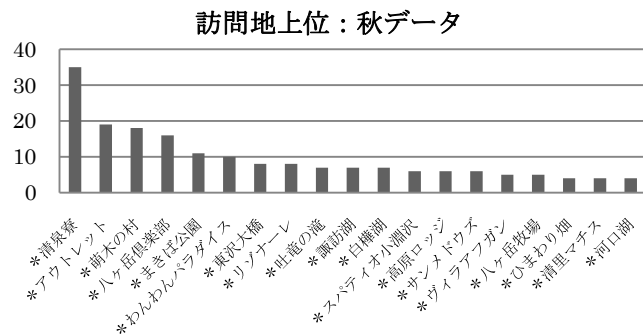


図8 訪問地上位(秋)

次に、ブログカテゴリーを用いて分析を行った。ブログカテゴリーとは、共通の話題を持った記事の分類で、記事を整理するために用いられる。これを見ることにより、どういう趣味の人が来訪しているのかがわかる。表1は、季節別にカテゴリーの分析を行い、抽出数の多い10カテゴリーを上から順に並べた。夏データのカテゴリーを見ると、ハンドメイド・料理・手作りといった家庭的なカテ

ゴリーが多かった。秋データは、旅行・お出かけ・旅といったカテゴリーが多かった。

カテゴリ：夏	語数	カテゴリ：秋	語数
日記	39	旅行	32
映画	22	お出かけ	28
お出かけ	18	日記	26
旅行	17	旅	24
風景	14	車中泊	12
ハンドメイド	12	料理	10
星空	12	散歩	9
料理	11	映画	8
カフェ	10	日本	8
手作り	10	イベント	7
：		：	

表 2 季節別カテゴリー表（左：夏，右：秋）

図 9 に平成 21 年秋に八ヶ岳南麓地域で実施されたアンケートとブログマイニングによる訪問地の割合を示した。アンケートとブログマイニングで大きく値に差が出た訪問地は、清泉寮、スパティオ小淵沢、東沢大橋であり、いずれもアンケートの配布場所であった。

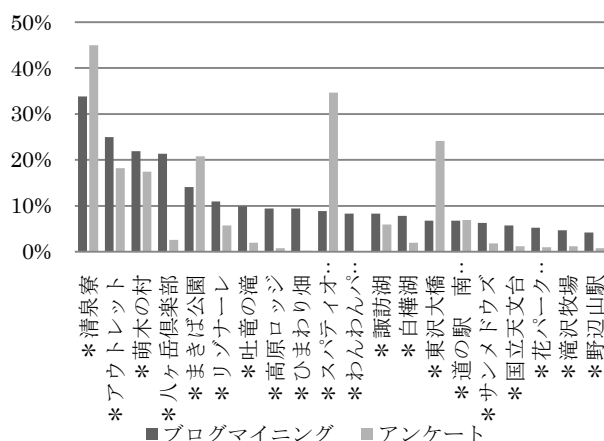


図 9 訪問地上位比較

4.2. 考察

季節別の比較においての、カテゴリーを用いての分析では、夏は家庭的なカテゴリーになったが、これは主婦層にあたる人がブログの書き手と考えられ、

先程の訪問地上位の結果と組み合わせてみると、夏休みに家族と涼しい場所や体験できる場所へ行っていると言える。同様に、秋には旅行好きの書き手が自然の綺麗な場所へ訪れるために八ヶ岳南麓地域へ来ていると考えられる。

アンケートとの比較では、ブログマイニングの結果と大きく値に差が出た、清泉寮、スパティオ小淵沢、東沢大橋はいずれもアンケートの配布場所である。このことから、アンケートの結果は配布場所に依存すると言える。

5. まとめと今後の課題

今回はブログマイニングによる観光行動調査を行った。単純集計では、来訪者の傾向を抽出することができ、季節別での比較では、季節により訪問地や来訪者の趣味に違いが出ていることが明らかになった。また、アンケートと比較を行った結果、配布場所への来訪において大きな差が見られ、アンケート結果が配布場所に依存していることが明確になった。

今後の課題は、アンケートによる観光行動調査の結果を踏まえて、今回行ったブログマイニングによる観光行動調査の結果と照らし合わせて、補完の可能性を示していきたい。

6. 参考文献

- 1) 上田隆穂，黒岩祥太，戸谷圭子，豊田裕貴：テキストマイニングによるマーケティング調査，講談社，2005
- 2) ネットレイティングス株式会社：
<http://www.netratings.co.jp/>
- 3) KHCoder：<http://khc.sourceforge.net/>
- 4) 鈴木大輝：“八ヶ岳南麓地域における観光周遊行動に関する調査分析”，山梨大学 工学部 卒業論文，(2010)