

# トップダウン型共起グラフを用いたブログからの観光地の行動分析

徳久雅人\*1 竹中直人\*2 木村周平\*1 谷本圭志\*3

\*1鳥取大学 大学院工学研究科 情報エレクトロニクス専攻

\*2鳥取大学 工学部 知能情報工学科

\*3鳥取大学 大学院工学研究科 社会基盤工学専攻

tokuhisa@eecs.tottori-u.ac.jp

## 1. はじめに

旅行計画を立てるために、観光地について知る必要がある。その内容として、期待される体験内容や旅行者の種類(親子連れ等)が挙げられる。ここで、旅行中に寄り道をするなど、急な計画作りが必要な場合には、より端的に情報をまとめる必要がある。そこで、本稿では、寄り道をする観光地を指定すると、その特徴を可視化する手法の開発を目的とする。

ブログ記事から観光地の傾向を分析する手法として、キーワードの出現回数を計測する手法は[1],[2]で示された。観光地の素性に対応するキーワードの頻度を Z-得点で正規化して観光地間で比較することが可能である。例えば、飲食に特徴のある観光地であるや、自然散策に特徴のある観光地であるという観光地の内容の方向性が分かる。文書の特徴を可視化するために共起グラフが用いられた。R および Mecab を用いるツールが公開されている[3]。共起グラフを用いると観光地の内容そのものが多数者からの意見として得られる。しかし、共起グラフは注意して読み取る作業を要する。

そこで、本稿では、共起グラフの中から知りたい情報を選出するために、観光地の素性に対応するキーワード(クエリ)を与えて、共起グラフを簡素化して表示する手法を提案する。

## 2. 提案手法

### 2.1 旅行中の状況

旅行中に観光地を調べる状況として以下の3つを想定する。

- (1) ぶらり旅: 見かけた観光地は、何が面白いのか?
- (2) 目的指向: 食事にしたい。何処かあるか?
- (3) 回遊: 次はどこに行こうか?

いずれの状況においても、ある指定する観光地に関して人々に共通する事項を抽出する。例えば、(1) では観光地で見られる人々の行動を、(2) では観光地で見られる人々の飲食に関する行動を、(3) では観光地から観光地への関係を、それぞれ情報源から抽出するものである。

ブログから生成した共起グラフを情報源とし、状況に応じて部分グラフを抽出することで上記の解決になると考える。

### 2.2 トップダウン型共起グラフの提案

本稿では、次の手順で共起グラフを扱うことを提案する。

- (1) 観光地を指定し、ブログ記事集を取得する。
- (2) 単語対の共起スコアを、同時に出現した記事数をベース

として算出する。具体的には、最低出現数を 4 回として単語を限定し、Jaccard 係数で共起スコアを算出する。

- (3) 単語を頂点、共起スコアの得られた単語対を辺として、共起グラフを生成する。
- (4) 分析目的に応じて注目する頂点を指定する。
- (5) 指定した頂点から 1 または 2 の距離で到達可能かつ、共起スコアが一定範囲内という辺および頂点を一定数を限度に残す。

共起スコアの指定範囲、および、辺の数の限度は、閲覧目的に依存する。ここで、(4) で述べた目的に応じた頂点の指定(クエリ)の定義が問題となる。

## 3. 観光地分析用クエリセット

### 3.1 旅行者がふれる情報の体系

旅行者がふれる情報を体系化する。その情報は、旅行者の内面に関する事、旅行の内容に関する事、情報伝播に関する事と分けてとらえる。

旅行者は、動機を持ち、計画をたて、旅行に行き、帰ってくるという過程をたどる。動機は「精神、身体、資産、仲間意識、上下関係、人生観」という観点で解釈する。例えば、安らぎの旅、グルメ旅、巡礼の旅などである。計画は「同伴者、予算、時間、所、品質、気象」に対する条件のもとで立てられ、「調査、予約、携行品準備」を行う。旅行の出発から帰宅までの行動は「移動、携行、見聞、飲食、体験、購入、交流、宿泊」がある。旅行から帰ると動機であった事柄がどれほど対応したかという「効果」を振り返る。また、「感想、記録、お土産配り、レポート」を行う。これらは旅の結果である。

旅行の内容として、「催事・イベント」は「学び、文化、スポーツ、総合」という観点で解釈する。例えば、「体験教室、修学旅行、見学」などは学びに関する催事である。「コンサート、パレード、ライトアップ」などは文化である。総合は切り口をかえて説明するための観点で、「前後(前夜祭、祝勝会など)、内容(～ショー系、～会系、～大会系など)、場所(～場系、～館系、～室系、～ドーム系など)」である。

情報伝播に関する事として、「情報源、媒体、内容」という観点で解釈する。情報源は「ゲスト(旅行者のロコミ)、ホスト(事業者の Web サイト)、ブローカー(代理店による広告)」という分け方が可能である。媒体は「会話、パンフレット、新聞、テレビ、作品」などである。内容は「実体験、評判、解説、広告、物語」である。作品・物語とはロケ地巡りや聖地巡礼という

タイプの旅行の背景になる情報である。

以上には下位分類を設けている。つきつめると最下位には固有名詞を割り当てることができる。

本稿は、旅行者の行動を分析する。少し詳しく説明する(図1)。「移動」に関する下位分類は「徒歩, 個人交通, 公共交通」があり「軽車両, 自動二輪, 自動車, バス, 列車, 航空機, 船舶」を経て「ロードバイク, レンタサイクル, ……」と具体的な手段がある。「携行」は旅行鞆などの持参・郵送・預けなどが対応する。「見聞」は「自然, 人工物, 複合体, 歴史, 認定, 非現実, 案内人付き, 催事見物」と分けており, 下位層では「桜, 白鳥, 蛍, オーロラ, 日の出, 高原, 滝, 砂丘, 湾岸線, パークウェイ, スーパー林道, ループ橋, ダム, 棚田, コンビナート, テレビ局, 磨崖仏, 温泉街, アーケード街, 夜景, レトロ, 城, 生家, 廃墟, 聖地, 百選, 文化財, パワースポット, 仮想スポット, 語り部」などに広がる。「飲食」は「料理の素材, ジャンル, 店舗, 座席, 給仕形式」などの中間層を経る。「体験」は「癒やし, 農林水産業, 信仰, 遊び, 催事参加」を経て「温泉, セラピー, 猫カフェ, 梨狩り, 田植え, 釣り, 翡翠探し, 初詣, テーマパーク, キャンプ, 蕎麦打ち, 金箔貼り, 写経, 郷土楽器, トレッキング, ダイビング, シーカヤック, 乗馬, カート, ツーリング, ゴルフ」などに広がる。「購入」は「商品, 場所, おまけ」を経て「お土産, アウトレットパーク, クーポン」などに広がる。「交流」は「相手, 内容, 機会」を経て「同伴者, 地元民, 会話, 協力, 手伝い, ツアー, 会, 教室」などに広がる。「宿泊」は「ホテル, ペンション, 車中泊」などである。以上の行動についてまとめると, 中間層のラベルは 168 種類, この時点での最下位のラベルは 827 種類となった。なお, 最下位ラベルに固有名詞は含まれていない。

体系の作成方法は, 内省により概要を作成し, 観光カテゴリ [1], アンケート項目と結果 [4] を参照しながら不足を補った。

#### (行動)

- (移動 徒歩 個人交通 公共交通)
- (携行 持参 依頼)
- (見聞 自然 人工物 複合 歴史 認定 非現実 案内人付き 催事見物)
- (飲食 料理 形式)
- (体験 癒し 農林水産業 信仰 遊び 催事参加)
- (購入 商品 場所 おまけ)
- (交流 相手 内容 機会)
- (宿泊 施設 特別行為))

図1 行動の中間層

### 3.2 クエリセットの作成

クエリは, ラベルおよび意味コードとする。意味コードは類語大辞典 [5] を用いる。前述したラベルについて意味コードを割り当てる。例えば, 「桜」の意味コードは “0108r咲く” である。多義語は意図するコードのみを残す。例えば「花」は “0108k咲く”, “1806h学ぶ”, “3200j遊ぶ”, “3200m遊ぶ”,

“4217m払う” があり, 「見聞」のためのコードとしては “0108k咲く” を残す。意味コードの付かなかったラベルは, ラベル単位でクエリとして扱う。

前述のラベルに不足することが多義語の解消の作業において見つかる。例えば, “1806h学ぶ” は, 学びの対象としての花(華道)であるが, 「見聞」ではなく「体験」のためのコードとしては妥当である。ゆえに, 体験の下位ラベルの追加となった。さらに, “1806a学ぶ” は, “1806h” の兄弟概念であり, 学ぶ動作である。これも体験の下位に位置することがふさわしい。このコードには「学ぶ, 勉強, 勉学」などの単語があり, それらを最下位ラベルとして追加した。

実装上は, コード単位で管理できるものは最下位ラベルとして展開しない。1 つのコードに含まれる単語の一部のみを「見聞」や「体験」などに対応させる場合に展開した。以上の結果, 表1のとおりとなった。8 つのクエリセットを列挙している。例えば, 「移動」に関する記述を共起グラフから抽出する場合, ノードの単語の意味コードが 63 件のコードのいずれかと一致する, または, 単語の字面が 317 件のラベルのいずれかと一致することが条件となる。

表1 クエリセットの規模

行動分類名	最下位ラベル数	意味コード数
移動	317	63
携行	118	15
見聞	413	403
飲食	553	91
体験	404	368
購入	227	69
交流	151	158
宿泊	55	44

## 4. 実験

### 4.1 実験条件と評価方法

観光地についての情報を抽出する実験を行う。抽出の目的は, 指定する観光地について, (1) 全般的な情報, (2) 移動に関する情報, (3) 携行に関する情報, (4) 見聞に関する情報, (5) 飲食に関する情報, (6) 体験に関する情報, (7) 購入に関する情報, (8) 交流に関する情報, (9) 宿泊に関する情報, である。抽出した情報を共起グラフで表示し, 観光地に行く前に閲覧することを想定し, 表示された情報が参考になり得るか否かを評価する。評価者は 1 名であり主観的な判定となる。

実験対象とする観光地は, 鳥取県内の観光スポットとして「じやらん」に掲載されていること, 同名の観光地が他に存在しないこと, 観光地名を含むブログ記事が 8 件以上存在すること, ブログ記事集にスパム記事や観光地名まとめ記事が含まれていないこと, とする。抽出した情報が妥当であることを判定するため, 評価者が行ったことのある観光地を評価に用いる。ブログ記事は, 2008 年 8 月から 2011 年 4 月までの期間に RSS で発信された記事(本文)である。具体的な観光地名とその記事数を, 表2 に示す。

表2. 実験対象の観光地名リスト

#	観光地名	記事数
1	大山まきばみるくの里	43
2	境港さかなセンター	24
3	江島大橋	42
4	鳥取砂丘オアシス広場	13
5	鳥取砂丘こどもの国	26
6	東郷湖	149
7	夢みなとタワー	61
8	わらべ館	70
9	鳥取城跡	32
10	因幡万葉歴史館	13
11	魚見台	14
12	境水道大橋	50
13	城原海岸	9
14	東郷湖羽合臨海公園	10
15	倉吉パークスクエア	18
16	倉吉博物館	20

## 4.2 結果

### (1) 全般的な情報の抽出結果

出力例として、観光地 #1 の共起グラフを図2 に表示する。中心に「大山まきばみるくの里」が赤いノードで表示されている。その周辺に「ソフトクリーム、食べる、行く、美味しい、牛乳、蒜山」などがあり、ここでの楽しみが正しく表示されている。「渋滞、駐車場、車」から移動手段が分かる。

16 件の観光地のうち 12 件は、行ったことのある者からみて妥当な情報が含まれていると評価されたので、抽出の精度は 75 % であるといえる。

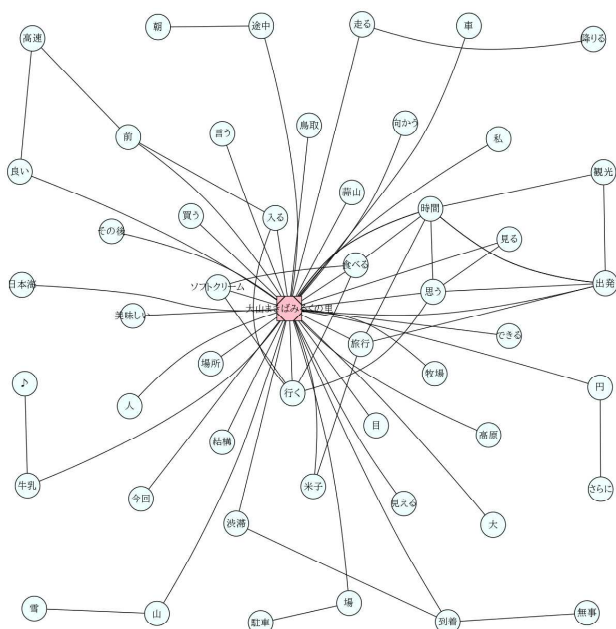


図2. 「大山まきばみるくの里」の共起グラフ

良かった点は、観光地に行くで見られる物事が想像できること、周辺の観光地が分かることが挙げられた。悪かった点は、子供の遊び場所ということ以上に詳しい情報がない(#5)、特別な何かのイベントの記述に影響を受けており本来の展示物の情報が引き出されていない(#10)、そもそも情報が乏しい(#13, #14)であった。

### (2) 移動に関する情報の抽出結果

車の利用が各観光地から読み取ることができた。例えば、「車、自動車、駐車、道路」等のクエリのヒットとその周辺単語が参考になった。但し、ほとんどの人が車で行くところ(#4, #10, #11)ではブログにあえて交通手段を記載しないと思われる。一方、東郷湖(#6)では、「駅、下車、タクシー、船」等のクエリがヒットしており、公共交通機関の利用が伺える。また、「高速」・「湯原」の共起や、「自動車」・「院庄」の共起を根拠として、利用したインターチェンジが伺える。

「JR 駅」の利用が不便なところ(#15)にて「駅」が抽出された。これは一つの記事で1日の旅が記載されていたことが原因である。幾つかの観光地を周遊したことを記載する記事を適切に分割する技術が今後必要になる。

### (3) 携行に関する情報の抽出結果

今回の 16 件の観光地においては携行品についての情報は得られなかった。別の観光地についての実験においては「サイクリング」・「積む」の共起から、「自転車」を携行品としている状況が抽出できた。

### (4) 見聞に関する情報の抽出結果

詳細な情報が得られることがあった。江島大橋(#3)は「べた踏み坂」で有名になったところである。「夕方」・「きれい」という見所が得られた。それ以上に注目する点は、「牡丹、高麗人参」である。これは特産であり、それを鑑賞に来る観光客がいる。知る人ぞ知るタイプの抽出ができた。

周遊に有益な情報が得られた。鳥取城跡(#9)に関して「仁風閣、美術館」、境水道大橋(#12)に関して「美保神社、妖怪、商店街」の各周遊に誘導可能である。

### (5) 飲食に関する情報の抽出結果

飲食が可能な各地で妥当な情報が得られた。「牛乳、ソフトクリーム」(#1)、「鯉、卵、魚、特産品」(#6)、「酒、ヨーグルト、刺身、バイキング」(#7)、「ママランチ、弁当」(#8)である。#8は幼児・児童向けの屋内施設であり、雰囲気を読み取れる。

### (6) 体験に関する情報の抽出結果

#5について詳しい内容が上記(1)で抽出できなかったが、ここでは少し詳しくなった。例えば「子供、遊び、負ける、遊具」等から、子供向け遊具を備えた施設であることが読み取れた。その他の観光地ではイベントの開催が盛んなところを読み取れた。

### (7) 購入に関する情報の抽出結果

購入の行動が観察されなかった。例えば、生鮮食品の購入

が期待されるが(#2),「結局,スーパー,買う」からウインドウショッピングに留まっている様子が想像された。

## (8) 交流に関する情報の抽出結果

「家族連れ」や「友達」のように同伴者や交流相手を読み取れた(#1, #8)。「講演」や「絵画教室」のように機会を読み取れた(#7, #16)。「宴会, 集合, 友好」など多種のクエリがヒットしたことから, 交流の盛んな所であることが伺える(#6)。

## (9) 宿泊に関する情報の抽出結果

「宿泊, ホテル」に関するクエリがヒットしたところは, #1, #3, #6, #7, #12であった。泊付き旅行として組み込まれやすい観光地であることが伺える。

以上についてまとめる。共起グラフを通じて, 全般的情報, および, 各目的別情報を抽出することの可能性が確認できた。共起スコアを用いるため, 多数意見が抽出されたと解釈できる。各観光地ごとに目的別のクエリが特徴的なヒットの仕方となり, その目的が関係しそうな観光地であるか否かが想像できる。旅先の選択の支援になると期待できる。

## 4.3 考察: 未知の観光地の評価

上述の評価実験では, 評価者が訪問経験のない観光地を除外した。本節では, 未知の観光地について全般的な情報を抽出し, じゃらんネットのロコミと照合し, 抽出結果の妥当性を考察する。

10 箇所の観光地「OU ランド(15), 金持神社(163), 妖怪神社(188), 植田正治写真美術館(154), 那岐山(230), 米子城跡(28), 上蒜山(123), 石谷家住宅(34), 大神山神社奥宮(38), 海とくらしの史料館(8)」について調べた(括弧内は記事数)。ロコミと同じキーワードが得られた共起グラフが 4 件, 共起グラフでは異なるキーワードだが想像される雰囲気がロコミと一致したところが 5 件, 抽出不足が 1 件であった。「海とくらしの史料館」はその抽出不足の観光地である。適切に抽出ができなかったので寄り道先として選択されないだろう。他 9 件は嗜好が合えば訪れるという判断をするかもしれない。

## 5. 関連研究

共起グラフの中から知りたい情報を選出するために, 本稿では, 観光地の素性に対応するキーワード(クエリ)を用いて, 共起グラフのノードを強調するようにした。

共起グラフを用いると観光地の内容そのものが多数意見として分かる。ここで, 共起性に基づく分析は, 事実(現地アンケートによる分析)と必ずしも一致するわけではないという[6]。文献[6]では, ブログから抽出したデータとアンケート調査結果を比較した。交通手段, レポート, 宿泊, 性別, 居住地, 嗜好性, 訪問先, 周遊(目的地対)の項目に着目した。多く項目で一致が見られたが共起性により判定した周遊については一致するとは限らないという。ブログには全ての観光地を記述しないためである。

ただし, 周遊分析のための正確さと観光地候補選択のため

の正確さには要求精度が異なるだろう。観光の前にネタバレをしたくないがハズレを避けたいという心境において, 多数者からの意見を概観し, 観光地の選択を行うことは妥当と思われる。ゆえに, 共起グラフは一つの方法として妥当である。

旅行者に伝える情報の詳しさを考える。本稿では共起数を記事数で数えた。文献[3]では文単位であり, 文献[7]では格関係が考慮されている。記事単位のほうが共起とみなすウインドウ幅が広いため得られる情報が粗い(個別的信息が得られにくい)。さらに, 旅行者に他の旅行者の賑わいを伝え多すぎる情報を伝えないという方法は文献[8]がある。情報閲覧者の目的に依存し, いずれも有益な手法である。

## 6. おわりに

本稿では, 共起グラフを用いてブログから得られる観光情報を簡素に可視化する手法を示した。観光地に関わる全般的情報から, 移動・見聞・飲食など 8 つの観点に関わる情報までをそれぞれで表示した。評価実験によると, 実際の観光地の印象に近い場合が 75% であり, 各観点での抽出も印象通りであった。以上により, クエリにより部分的に抽出された共起グラフによる可視化を通じて, 観光地についての情報を目的に応じて旅行者に伝えることの可能性が確認できた。

## 参考文献

- [1] 三田村保, 岩佐渉, 湯川恵子, 大堀隆文: ブログを利用した観光情報の調査分析, 観光と情報, Vol.4, No.1, pp.57-65, 2008.
- [2] 三田村保, 大町清隆, 大堀隆文: Web検索による北海道観光情報の分析, 観光と情報, Vol.1, No.1, pp.47-53, 2005.
- [3] 石田基広: R によるテキストマイニング入門, 森北出版, 2008.
- [4] 安田亘宏, 川口賢次, 相田正己: 旅人の本音, 「日本」の旅に関するアンケート集, 彩流社, 2009.
- [5] 柴田武, 山田進(編): 類語大辞典, 講談社, 2002.
- [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, 2011.
- [7] 中嶋勇人, 太田学: 旅行ブログ記事からの名所とその付随情報の抽出, 第5回データ工学と情報マネジメントに関するフォーラム, DEIM Forum 2013, B8-4, 2013.
- [8] 倉田陽平, 杉本興運, 矢部直人: あえて案内しない着地型観光案内 - 観光関心点群の抽出と活用, 第19回地理情報システム学会学術大会, 地理情報システム学会講演論文集 19, 2010.