

ネット上の個人投稿の分析による北海道観光の地域特性に関する研究

A Study on the Regional Characteristic of the Hokkaido Sightseeing by Analyzing the Comments in Social Media on the Internet

泉澤 圭亮*

IZUMISAWA, Keisuke

観光は北海道の主要産業の一つであるが、近年は観光入込客数が停滞している。一方で ICT の発達と普及によりさまざまなデータがインターネット上に流通し、蓄積され始めている。その中には個人が直接発信した観光に関するデータが含まれている。本稿は、観光に関するネット上の個人投稿をテキストマイニングの手法で分析することによって、観光客が見いだしている北海道の地域に対する特徴を把握し、類似性や特異性を明らかにするものである。

キーワード：テキストマイニング、個人投稿、観光行動

1. 背景と目的

高齢化社会、人口流出による地方の定住人口減少などによって、地域経済が疲弊している。このような状況において、観光産業は国の成長戦略のひとつに位置づけられている。高い経済波及効果を持ち、同時に雇用創出効果も見込まれる産業だからである。観光は北海道においても主要産業の1つである。しかし、北海道の観光入込客数は停滞している¹⁾。これは、観光客が北海道に求めているものと、北海道の地域が発信しているものとの間に差異が生じているのではないだろうか。この差異を明らかにできれば、今後の北海道の観光振興に役立てることができるものと考えられる。

本研究は、ソーシャルメディア上に蓄積されている観光に関する個人投稿をテキストマイニングの手法で分析することによって、北海道の地域の特徴を把握し、地域ごとの類似性や特異性を明らかにすることを目的とする。その上で、北海道の各地域において、特色を持った観光地形成が可能な施策の提案に寄与できるような手法を提示することを最終的な目的とした。

2. 方法

本研究は主にオープンソースで提供されているツールを分析システムに採用した(表-1)。テキストマイニングにあたっては、分析対象のテキストからノイズになり得るものを一定基準によって除去したり、表記を統一したりするテキストクリーニング処理を実施することが必要である²⁾。本研究では以下の手順によってテキストクリーニング処理を実施した。

はじめに、分析対象テキストに含まれる表記ゆれを減らすため「すべての半角カタカナを全角カタカナへ変換」「すべての全角ローマ字を半角ローマ字へ変換」「すべての大文字および小文字を大文字へ変換」「すべての全角ローマ数字を半角ローマ数字へ変換」処理をした。次に、テキス

*北海商科大学大学院商学研究科博士後期課程

トの形を一定に保つために空白文字と改行を削除した。さらに、URL(Uniform Resource Locator)は分析結果のノイズになる可能性があるため、これを削除した。最後に記号を削除した。また、語句の頻度をもとに分析を行う場合は、テキストの分量にばらつきがあるため語句のテキスト内での重要度を考慮することが望ましいという指摘がある³⁾。本研究はTF-IDF法を使用し、語句の重み付けを行った。本稿では、重み付け処理後の値のことを重要度という。

表-1 本研究の分析システム

分類	ソフトウェア
オペレーティングシステム	Windows 7
形態素解析エンジン	MeCab 0.996
MeCabの追加ユーザー辞書	Wikipedia 辞書 ⁴⁾
統計解析ツール	R 3.0.2
Rの追加パッケージ	RMeCab 0.9994 Igraph 0.7.1
ネットワーク可視化プラットフォーム	Cytoscape 3.1.1
統合開発環境	RStudio 0.98.501

3. テキストマイニングによる観光用語の抽出と対象の分類

観光領域のテキストをテキストマイニングの手法で分析することが妥当であるかについての知見を得るために「観光白書」「観光雑誌」「SNS上の個人投稿」を対象にテキストマイニングの手法を用いて分析を行った。

3.1 観光白書の分析

観光白書は国土交通省および観光庁が発行している公的な資料である。観光白書にはわが国の観光の趨勢が反映されており、同時に観光領域で使用される様々な言葉が使われている。これをテキストマイニングの手法で分析することで、時期の趨勢を明らかにするとともに、結果から観光に関するキーワード群を抽出することを目的とした。分析対象とした観光白書は、国土交通省のウェブページで公開されているうち平成14年版から平成26年版である。

はじめに各観光白書の形態素解析を行った。形態素解析の結果は「名詞」と「形容詞」に分類された語句を残し、それ以外は削除した。品詞細分類から「数・接尾・非自立」に分類された語句も削除した。結果、得られた異なり語句数と延べ語句数は表-2の通りである。この結果を語句頻度により順位付けした。

表-2 品詞処理後の形態素解析の語句数

版	異なり語句数	延べ語句数
H14	3,494	17,891
H15	4,040	21,364
H16	4,662	24,507
H17	4,251	22,804
H18	4,278	23,871
H19	4,151	24,903
H20	3,487	18,845
H21	3,030	16,525
H22	3,047	14,979
H23	3,099	16,294
H24	3,052	15,106
H25	3,919	23,015
H26	1,261	5,555

形態素解析の結果から次の知見が得られた。①各観光白書ともに50位以内の語句に「観光」「旅行」「地域」「実施」「推進」「事業」「外国」「日本」「月」「開催」「等」「国際」がある。②特定の版以前に共通して抽出されている語句がある。代表される語句は「ミレニアム宣言」「ハワイ」「低

廉」である。③特定の版以後に共通して抽出されている語句がある。代表される語句は「ビジットジャパン」「日本政府観光局」「東日本大震災」である。また上位 50 位の語句のうち平成 26 年版観光白書に限定的に抽出されている語句に「大会」「オリンピック」「ロンドン」「地方」「英国」「パラリンピック」「プロモーション」がある。いずれの語句も 2020 年東京オリンピックに関係する語句であることが推察され、「地方」「プロモーション」といった語句からは、観光庁が 2020 年東京オリンピックを地方の観光促進の機会として捉えている可能性が指摘できよう。

しかし、「観光」や「旅行」といった語句がそれぞれ単独で用いられているのか、それとも複合語として用いられているのかについては形態素解析の結果からはわからない。これを明らかにするために n-gram による分析を行った。

n-gram は n に指定する値によって単語を連続的に切り出すことができる。バイグラムとトライグラムのどちらが観光白書の文脈を捉えやすいのかを判断するため、形態素解析同様に「名詞」と「形容詞」を対象にバイグラムとトライグラムを行った。バイグラムとトライグラムの結果から、総頻度が 2 以下の n-gram を削除し、数字も削除した。数字を削除したことによって「あ-い」のような空白を持った n-gram が生成されたため、これを削除した。

さらに、「観光」「旅行」「地域」の 3 語句について、形態素・バイグラム・トライグラムの結果を比較した(表-3)。表-3 からバイグラムよりもトライグラムの方が複合語の判断が容易であると考え、トライグラムの結果を使用し分析を進めることにした。

表-3 形態素・バイグラム・トライグラムの比較

形態素	バイグラム	トライグラム
観光	観光-地	観光-地-形成
	観光-立国	観光-立国-懇談
旅行	旅行-者	旅行-者-数
	国内-旅行	国内-旅行-消費
地域	地域-観光	地域-観光-振興
	地域-活性	地域-活性-化

分析にあたりトライグラムを頻度で順位付けした。次に「対-前年-比」「今後-実施-こと」に代表される『統計に恒常的に使用される』『文脈よりも一般的使用の側面が強い』と判断したトライグラムを削除した。この結果から各観光白書の上位 10 位までのトライグラムを抽出した(表-4)。

平成 14 年版の 3 位と平成 15 年版の 5 位に「国際観光-振興-会」がある。平成 16 年版以後は抽出されていないが、平成 16 年版の 3 位と平成 18 年版の 4 位に「国際観光-振興-機構」がある。これは『国際観光振興会』が平成 15 年に『国際観光振興機構』へと改組されたことから、これがテキストマイニングの結果にも反映されたものと考えられる。平成 17 年版は 7 位に「観光-立国-推進」、平成 19 年版から平成 25 年版までは「観光-立国-推進」や「観光-立国-実現」など観光立国に関する結果が抽出されている。さらに、平成 20 年版から平成 24 年版にかけては「国際-会議-開催」や「国際-会議-等」が抽出されている。観光立国推進基本計画において MICE の一環として国際会議についての言及があり、その動きが観光白書にも反映されたものと考えられる。平成 26 年版は 3 位に「訪英-外国-人」、5 位に「英国政府-観光-庁」、10 位に「訪加-外国-人」がある。いずれもオリンピックに関する語句と考えられる。形態素解析結果でも同様の傾向であったが、平成 26 年版観光白書はオリンピック色の濃い編集がなされていると考えられる。

以上の結果から、わが国の観光に関わる政策の変化、社会的背景の変化がテキストマイニングの結果にも反映されていることが確認できた。同時に、テキストマイニングが観光用語の抽出に

表-4 各版の観光白書のn-gram頻度表(10位まで)

H14		H15		H16		H17	
トライグラム	順位	トライグラム	順位	トライグラム	順位	トライグラム	順位
外国-人-旅行	1	人-旅行-者	1	外国-人-旅行	1	外国-人-旅行	1
人-旅行-者	1	外国-人-旅行	2	人-旅行-者	2	人-旅行-者	2
国際観光-振興-会	3	日本人-海外旅行-者	3	国際観光-振興-機構	3	外国-人-観光	3
日本人-海外旅行-者	4	旅行-者-誘致	3	交流-空間-づくり	4	人-観光-客	4
安全-確保-対策	5	国際観光-振興-会	5	歴史-的-風土	5	観光-交流-促進	5
観光-安全-確保	6	者-誘致-受入れ	5	観光-交流-空間	5	訪日-外国-人	5
章-観光-安全	6	歴史-的-風土	7	施設-等-整備	7	日本人-海外旅行-者	7
経済-波及-効果	8	旅行-者-訪日	8	国内-宿泊-旅行	8	農林-漁業-体験	7
国際観光-テーマ-地区	8	旅行-者-受入れ	8	都市-農山-漁村	9	観光-立国-推進	7
観光-関係-者	10	者-訪日-促進	10	日本人-海外旅行-者	10	歴史-的-風土	10
施設-等-整備	10			これ-施策-記述	10	観光-案内-所	10
沖縄-観光-振興	10						
H18		H19		H20		H21	
トライグラム	順位	トライグラム	順位	トライグラム	順位	トライグラム	順位
外国-人-旅行	1	人-旅行-者	1	観光-地-住民	1	国内-宿泊-観光旅行	1
人-旅行-者	2	外国-人-旅行	2	地-住民-等	2	外国-人-旅行	2
観光-交流-促進	3	観光-交流-促進	3	外国-人-旅行	3	人-旅行-者	3
国際観光-振興-機構	4	観光-立国-推進	3	人-旅行-者	3	国際-会議-開催	4
日本人-海外旅行-者	5	外国-人-観光	5	国際-会議-開催	5	観光-立国-推進	5
訪日-教育-旅行	6	国内-観光-消費	6	外国-人-観光	6	観光-立国-実現	6
外国-人-観光	7	観光-立国-実現	7	観光-立国-推進	7	立国-推進-基本計画	7
地域-観光-マーケティング	7	外-航-クルーズ	7	市町村-観光-地	7	国際-会議-等	8
歴史-的-風土	9	人-観光-客	9	観光-交流-促進	8	観光-交流-促進	9
施設-等-整備	9	日-中-韓	9	等-情報-提供	9	地域-活性化	9
						層-国内-宿泊	9
H22		H23		H24		H25	
トライグラム	順位	トライグラム	順位	トライグラム	順位	トライグラム	順位
外国-人-旅行	1	訪問-者-意識	1	外国-人-旅行	1	外国-人-旅行	1
人-旅行-者	2	外国-人-旅行	2	人-旅行-者	2	人-旅行-者	1
観光-立国-推進	3	人-旅行-者	3	日本人-国内-宿泊	3	受入-環境-整備	3
国際-会議-開催	4	意識-訪問-者	4	国際-会議-等	4	訪日-旅行-者	4
観光-立国-実現	4	住民-意識-訪問	4	国内-旅行-消費	4	昇-龍-道	5
外国-人-観光	4	観光-立国-推進	6	国内-宿泊-者	6	外国-人-観光	6
観光-地域づくり-人材	4	観光-立国-実現	7	訪日-外国-人	7	日本-海外旅行-者	7
旅行-者-受入れ	8	国際-会議-等	8	客-中心-宿泊施設	7	関係-者-連携	8
国際-会議-等	9	訪日-外国-人	8	観光-立国-推進	9	観光-立国-実現	9
地域づくり-人材-育成	9	魅力-観光-地	8	観光-立国-実現	10	日本人-海外旅行-者	9
地域-観光-振興	9			国際-会議-開催	10	観光-案内-所	9
H26							
トライグラム	順位						
外国-人-旅行	1						
人-旅行-者	1						
訪英-外国-人	3						
訪日-旅行-者	4						
旅行-者-消費	5						
英国政府-観光-庁	5						
VB-英国政府-観光	7						
旅行-者-誘致	8						
旅行-者-受入	9						
国内-宿泊-観光旅行	10						
訪加-外国-人	10						

も有効なツールであることも確認できた。

3.2 観光白書と観光雑誌の語句の比較

観光白書はその特質として行政が発行していることが挙げられる。これは一般の観光客が観光行動の際に参考とする観光情報とは異なる。観光客が求める観光情報が具体的にどのような語句で構成されているのであろうか。全国を対象とした観光雑誌の「旅の手帖」と「読売旅行」を分析し、その結果を観光白書と比較することによって媒体間の特徴の違いを明らかにする。

分析対象の観光白書は「3.1 観光白書の分析」で使用した各観光白書を1つのテキストファイ

ルに結合したものを使用した。旅の手帖と読売旅行(以下、観光雑誌という)は、ともに2012年4月号から2012年12月号までを対象とした。両媒体とも紙であることから、分析可能なデジタルデータへ変換した。1ページずつスキャナーで読み込み、コンピューターへ保存した。さらに1ページずつOCR処理し、OCRによるテキストの誤認識があった場合は手作業で修正した。「表紙」「広告ページ」「アンケートページ」などは分析結果のノイズとなる可能性があるため、すべて削除した。以上の処理を行った観光雑誌は、媒体ごとにすべてのページを結合し、保存した。

各媒体とも形態素解析し、語句の重み付けを行った。さらに、品詞の「名詞・形容詞」に分類された語句を残し、すべて削除した。品詞細分類から「数・接尾・非自立・代名詞・副詞可能」に分類された語句を削除した。結果、各媒体の異なり語句数と述べ語句数は表-5の通りとなった。この結果から名詞と形容詞を分け、重要度降順に順位付けし、分析を行った。

表-5 名詞と形容詞の観光白書、旅の手帖、読売旅行の異なり語句数と述べ語句数

媒体	異なり語句数	述べ語句数
観光白書	11,770	237,650
旅の手帖	14,922	63,474
読売旅行	11,795	42,206

各媒体から50位までの名詞を抽出した結果、各媒体の50位以内に共通している語句は「観光」「日本」「月」「開催」「利用」の5語句であった。これらの語句は観光をテーマにした媒体で普遍的に用いられやすい語句と判断し、削除を行い、語句の比較を行った。観光白書は「取組」「整備」「実施」など観光政策に関する語句がある。旅の手帖は「東京」「京都」に代表される特定の場所と結びつきやすい語句がある。「温泉」「海」「秘境」などは資源に関する語句と考えられる。「列車」「青春18きっぷの旅」は移動に関する語句であるとともに、旅の手帖がJRグループの協力を受けている雑誌という特徴があらわれている。読売旅行も特定の場所と結びつきやすい語句がある。「左折」「徒歩」「県道」など移動に関する語句も抽出されているが、列車による移動が語句から類推される旅の手帖と比べ、読売旅行は徒歩や自動車が移動の中心にあると考えられる。

名詞同様に50位までの形容詞を抽出した結果、各媒体の50位以内に共通している語句は「高い」「多い」「大きい」など17語句であった。この17語句を除外し、各媒体を比較した。観光白書は「著しい」「低い」「粘り強い」など様相をあらわす語句がある。さらに「望ましい」「正しい」など期待を表す語句がある。旅の手帖は「おいしい」「すごい」「旨い」などの評価語句がある。特に「おいしい」「旨い」などは食に関する評価語句である。読売旅行も同様の傾向を有し、かつ「むさい」「やばい」といった口語も用いられている。口語は日常生活で一般的に用いられる言葉であることから、情報を伝わり易くするための特異的な編集がなされていると考えられる。

以上の結果から、観光白書は観光政策に関する語句が上位にあり、様相や期待をあらわす語句があった。観光雑誌は場所と結びつきやすい語句、移動に関する語句など多様な表現があった。さらに、口語表現を用いることによって、読者へ情報が伝達しやすくなるような特異的な編集がなされていることが分かった。

3.3 ソーシャルメディア上に投稿された公園のクチコミ

観光白書、観光雑誌はともに編集され、事業者側からの視点を中心にある。一方、ソーシャルメディアは個人投稿が可能で、不定型な情報で満たされている媒体である。北海道庁の調査から、北海道を訪れる観光客は北海道の自然環境に高い関心を寄せていることが明らかとなっている⁵⁾。

観光客のみならず地域住民にとっても身近な自然環境として都市公園がある。旅行系 SNS(Social Networking Service)のフォートラベルとトリップアドバイザーに投稿された札幌市と東京都にある各5つの公園のクチコミの分析と比較を通し、都市公園の景観とその印象が地域の観光イメージにどのように反映されているのかについて検討を行った。

表-6は分析対象の公園とクチコミ数である。クチコミ数はフォートラベルとトリップアドバイザーのクチコミ数を合算したものである。

札幌市の公園		東京都の公園	
公園名	クチコミ数	公園名	クチコミ数
大通公園	593	日比谷公園	257
中島公園	229	代々木公園	185
円山公園	166	水元公園	73
モエレ沼公園	62	上野恩賜公園	425
滝野すずらん丘陵公園	55	昭和記念公園	210

取得した各公園のクチコミは公園ごとに1つのテキストファイルに統合し、これを保存した。さらに「公園」「北海道」「札幌市」「札幌」「東京都」「東京」に一致する文字列を削除した。この処理はテキスト解析の冗長性を減ずるためである。以上の処理を行ったテキストファイルを「名詞・動詞・形容詞」の3品詞を対象に、語句の総頻度3以上に限定し、形態素解析した。各公園のクチコミ数にはばらつきがあるため、重み付けを行った。この結果から『数字のみの語句』『1文字のローマ字の語句』を削除し、重要度降順で順位付けし、各品詞の30位以内の語句を対象に比較した。表-7は形態素解析の結果から得られた各公園の主な語句である。

公園	主な語句
大通公園	雪まつり、イルミネーション、ビアガーデン、とうもろこし、噴水、食べる、飲む、賑わう、細長い、美味しい
中島公園	豊平、館、文学、天文台、コンサートホール、KITARA、ポート、散歩、ホテル、親しむ、癒やす、心地よい、古い
円山公園	神宮、円山動物園、円山原始林、桜、花見、ジンギスカン、梅、登山、エゾリス、咲く、茂る、深い
モエレ沼公園	ピラミッド、ガラス、海、プレイマウンテン、イサムノグチ、レンタサイクル、子供、遊具、登る、登れる、怖い
滝野すずらん丘陵公園	子供、遊具、水遊び、キャンプ場、チューリップ、アシリベツの滝、コスモス、転がす、遊べる、咲く、安い
日比谷公園	銀座、千代田区、有楽町、噴水、ベンチ、オアシス、鉄道フェスティバル、日比谷公会堂、食べる、勤める、古い
代々木公園	原中、渋谷、小田急線、明治神宮、オリンピック、練習、ダンス、サイクリング、フリーマーケット、花見
水元公園	菖蒲、花、釣り、金町、江戸川、バーベキュー、咲き乱れる、遊ぶ、だだっ広い
上野恩賜公園	美術館、博物館、桜、満開、パンダ、上野動物園、咲く、飲む、臭い、懐かしい、古い
昭和記念公園	コスモス、チューリップ、ポピー、日本庭園、プール、レンタサイクル、イルミネーション、咲く、遊ぶ、安い

形態素解析で得られた特徴から公園を分類するためにクラスター化を行った。結果は、形態素解析から得られた特徴により以下の3つのクラスターに分けた。

- ①水元公園、モエレ沼公園、昭和記念公園、滝野すずらん丘陵公園
- ②円山公園、上野恩賜公園
- ③日比谷公園、中島公園、代々木公園、大通公園

①はアクティビティ性の高さや景観性の高さが特徴の公園であると考えられ、野外活動型の公園として活用されている。②は動物園や文化施設など複数の特徴を持つ公園であると考えられ、総合型の公園として機能している。③はイベント性の高さや地域住民の憩いの場が特徴の公園であると考えられ、地域密着型の公園である。

このように、クラスター化することにより対象をいくつかに分類できることが分かり、そこに共通するものを見いだせることも分かった。以上の結果から、観光領域における語句抽出と対象分類について、テキストマイニングの手法が極めて有効な方法であることが分かった。そこで、この手法を北海道観光に適用し、来道観光客が北海道へ向けている視点を明らかにする。

4. 個人旅行記に見る北海道の地域イメージと分類

地域そのものに対する視点という点においては、クチコミの分析にすべてを委ねることはできない。クチコミは特定の観光対象に向けられた視点であって、地域としての視点ではないからである。地域としての視点が見いだせる媒体に個人旅行記がある。公園のクチコミ分析にも利用したフォートラベルに旅行記というコンテンツがある。これはブログであり、旅行者の幅広い旅行あるいは観光体験について語られている。そこで、北海道旅行を対象とした個人旅行記において、観光客が北海道という地域に具体的にどのような視点を向けているのかについて明らかにする。

4.1 個人旅行記による北海道 30 地域の分類

フォートラベルの北海道旅行記の中から、2007年から2013年に分類されている個人旅行記をいいね順に並べ替え、各年最大100件ずつ取得し、年別にテキストファイルとして保存した。テキストファイルはテキストクリーニング処理し、1つのテキストファイルに結合し、保存した。保存したテキストファイルから北海道内の地域名を抽出するため、品詞を「名詞」、品詞細分類を「固有名詞」に限定し、形態素解析した。結果は北海道内の市町村名およびフォートラベルのエリア名と照合し、頻度降順に30地域抽出した(表-8)。このとき「札幌」と「定山溪」のように市町村名とそこに包含されている地域名や観光対象が両方抽出されている場合は、頻度の高い方を対象とした。また頻度は高くともフォートラベルの旅行記の投稿数がゼロの年がある地域は除外した。表-8の個人旅行記を取得条件は北海道旅行記と同様にし、フォートラベルから取得した。

表-8 分析対象の道内30地域名(地域名はフォートラベルのエリア名に準ずる)

ニセコ	旭岳・天人峡・白金	旭川	釧路	厚岸・霧多布	根室	札幌	支笏湖
室蘭	斜里	小樽	石狩・当別・厚田	積丹半島	川湯温泉	増毛・雨竜	帯広
大沼・駒ヶ岳	稚内	鶴居・標茶	洞爺湖	苫小牧	函館	美瑛	富良野
網走	夕張	余市	羅臼	利尻島	礼文島		

各地域の個人旅行記は文章整理後、テキストクリーニングし、分析用テキストとして保存した。分析用テキストは品詞に「名詞・動詞・形容詞」を指定し、形態素解析した。結果は語句の重み付けし、品詞細分類から「一般・サ変接続・固有名詞・ナイ形容詞語幹・形容動詞語幹・自立」のいずれかに分類された語句以外は削除した。さらに地域ごとに名詞・動詞・形容詞に分割し、重要度降順に順位づけし、品詞ごとに順位昇順に100位までを分析対象とした。形態素解析の結果から、主な『その地域でのみ100位以内にある語句』と『複数の地域で100位以内にあるが、その地域では10位以内にある語句』を抽出し、地域ごとの特徴をまとめたものが表-9である。

形態素解析の結果得られた特徴から30地域を分類するためにクラスター化を行った。結果は、5つのクラスターに分けた。①クラスターIは『利尻島、礼文島』で構成され「自然景観(山、植物)」への視点がある。②クラスターIIは『余市』単独で構成され「酒」と「人物」の視点が明確である。③クラスターIIIは『石狩・当別・厚田、増毛・雨竜、旭岳・天人峡・白金、美瑛、富良野、小樽、積丹半島、夕張、旭川、札幌、帯広、苫小牧、支笏湖、ニセコ、洞爺湖、大沼・駒ヶ岳、函館、室蘭、根室、稚内』で構成され、もっとも大きなクラスターである。札幌や小樽などの都市が含まれていることから「北海道観光の象徴」と見なされている可能性がある。④クラスターIVは『羅臼、斜里、網走』で構成され、いずれも道東に位置し「自然景観(流氷、知床)」に視点がある。⑤クラスターVは『厚岸・霧多布、川湯温泉、釧路、鶴居・標茶』で構成され、クラスターIV同様に道東地域の集団である。クラスターIVが道東の北部に位置する集団であるの

表-9 形態素解析の結果から得られた各地域の特徴

地域	主な特徴(カッコ内は代表的な語句)
ニセコ	自然景観への視点(羊蹄山、神仙)、レジャーへの視点(ニセコビレッジ、ニセコアンヌプリ国際スキー場)
旭岳・天人峡・白金	自然景観への視点(姿見の池、鏡池、チングルマ、黒岳、大雪山)
旭川	都市への視点(旭山動物園、ペンギン、アザラシ、オオサミ、レッサーパンダ)
釧路	自然景観への視点(釧路湿原、細岡展望台、カヌー)、都市への視点(和商市場、釧路市動物園)
厚岸・霧多布	自然景観への視点(霧多布湿原、アゼチ、琵琶瀬)、食への視点(牡蠣)
根室	自然景観への視点(納沙布岬、温根沼)、場所への視点(北方領土、歯舞諸島)、食への視点(エスカロップ、花咲ガニ)
札幌	都市への視点(テレビ塔、大通公園、円山動物園、さっぽろ雪まつり、イルミネーション)
室蘭	自然景観への視点(室蘭八景、大黒島、トッカリシヨ、イタンキ浜)、食への視点(室蘭焼き鳥、カレーラーメン)
支笏湖	自然景観への視点(洞門、恵庭岳、オタコネ湖、支笏湖温泉、温泉旅館)
斜里	自然景観への視点(知床、オシンコシンの滝、知床五湖、斜里岳、流水)
小樽	歴史的背景への視点(運河、小樽運河、手宮線、オルゴール堂、日本銀行、ウォール街、雪あかりの路、日本郵船)
石狩・当別・厚田	レジャーへの視点(釣り、ワカサギ、浜益川、釣果、河口、竿、釣り場、石狩川)
積丹半島	自然景観への視点(神威岬、島武意海岸、積丹岬)、食に関する語句(ウニ)
川湯温泉	自然景観への視点(硫黄山、イツツツジ、ハイマツ、美幌峠)、温泉に関する語句(温泉、砂湯、足湯)
増毛・雨竜	自然景観への視点(ひまわり、雨竜沼湿原)、酒への視点(國稀酒造、酒蔵)、レジャーへの視点(キャンプ場)
帯広	文化的景観への視点(ガーデン、紫竹、六花の森、北海道ガーデン街道)、レジャーへの視点(競馬、ばんえい)
大沼・駒ヶ岳	自然景観への視点(大沼公園、駒ヶ岳、大沼国定公園)、レジャーへの視点(SL 函館大沼号、蒸気機関車)
稚内	場所への視点(宗谷岬、最北端)、自然景観への視点(ノシャップ岬、利尻富士)、歴史的背景への視点(サハリン、樺太)
鶴居・標茶	自然景観への視点(タンチョウ、釧路湿原、サルボ)、レジャーへの視点(SL 冬の湿原号)
洞爺湖	場所への視点(昭和南山、ウィンザーホテル、スイートルーム、洞爺湖温泉、サンパレス)
苫小牧	自然景観への視点(樽前山、ウトナイ湖、外輪山)、移動への視点(乗船、太平洋フェリー、船内)
函館	歴史的背景への視点(五稜郭、ベイエリア、旧函館区公会堂)、食への視点(ラッキーピエロ、ハセガワストア)
美瑛	場所への視点(マイルドセブン、四季彩の丘、ケンとメリーの木、セブンスター、ラベンダー)
富良野	文化的景観への視点(ラベンダー)、場所への視点(風のガーデン、北の国から)、食への視点(オムカレー)
網走	自然景観への視点(流水、網走湖、オホーツク流水館)、歴史的背景への視点(網走監獄、網走刑務所、監獄)
夕張	歴史的背景への視点(石炭、炭鉱、石炭の歴史村、森林鉄道)、場所への視点(幸福の黄色いハンカチ)
余市	酒への視点(蒸溜、ニッカウヰスキー、試飲、原酒、造る、芳ばしい)、人物への視点(リタ、竹鶴政孝、リタロード)
羅臼	自然景観への視点(羅臼湖、知床、流水、羅臼岳)、レジャーへの視点(スノーシュー)、場所への視点(国後島)
利尻島	自然景観への視点(利尻山、姫沼、オタトマリ、レラシモリ)、食への視点(ミルビス)
礼文島	自然景観への視点(桃岩、スコトン岬、ゴロタ、レブンアツモリソウ)、場所への視点(桃岩荘ユースホステル)

に対し、クラスターVは南部に位置する集団である。「自然景観(湿原)」と「温泉」に視点がある。

以上の結果から、北海道観光は総合的な特徴に「自然景観への視点」がある。しかし、その対象は同一ではなく、地域によって異なっている。これらの視点を作り出している地域資源が具体的にどのようなもので、それが他の地域資源とどのような関係性を構築しているのであろうか。

4.2 北海道の地域構造の分類

観光客にとっての北海道観光の中心点がどこであるのかを把握するために、はじめに道内の地域構造を明らかにする。フォートラベルの北海道旅行記から2007年から2013年に分類されている北海道旅行記を全件取得した。これをテキストクリーニングし、ネットワーク図を作成するために品詞を「名詞」に限定し、頻度2以上を対象にバイグラムを取得し、n-gram頻度表として保存した。n-gram頻度表を道内179市町村名と照合し、道内の市町村名同士のバイグラムを抽出し、これをもとにネットワーク図を作成した。さらに、媒介中心性を求め、媒介中心性値が500以上の地域を抽出した。結果「札幌、旭川、富良野、函館、釧路、帯広、苫小牧、小樽、網走、稚内、北見、ニセコ、美瑛」の13地域であった。これら13地域は観光客から北海道観光の中心的役割を見いだされていると考えられる。そこで、フォートラベルにある13地域の個人旅行記のうち、2007年から2013年に分類されているものを全件取得し、ネットワーク分析を行った。

各地域の個人旅行記は文章整理し、品詞を「名詞」に限定し、バイグラムを取得した。バイグラムから品詞細分類が「一般・サ変接続・自立・固有名詞・形容動詞語幹・ナイ形容詞語幹」のいずれかの組み合わせであるもの以外は削除し、組合せ頻出の上位100組を対象にネットワーク図を作成した。各ネットワーク図の最大連結成分に着目し、語句の連結パターンによりパターン

A からパターン E に分類した(表-10)。

パターン	該当する地域			
A	札幌	函館	小樽	ニセコ
B	旭川	富良野	釧路	帯広
C	網走	苫小牧		
D	北見			
E	美瑛	稚内		

パターン A は地域への明らかな一極集中であり、地域全体が観光の目的を構成している(図-1)。パターン B は広域的な観光行動が特徴であり、観光対象にテーマ性が見いだされていると考えられる(図-2)。テーマ性を高めていくことが必要である。パターン C は地域と密接な関係にある観光対象がある一方で、注目度の高い観光対象が孤立して存在している(図-3)。このような観光対象同士を繋げることが必要である。パターン D は観光対象の独立性が高く、資源同士の連携が図られていないと考えられる(図-4)。地域資源を整理し、再確認することが必要である。パターン E は観光対象が地域を代表するシンボリックな役割を見いだされている(図-5)。シンボルと地域の関係性を高めていくことが必要である。以上の結果から、北海道の地域の類似性と特異性の構造を明らかにすることができた。しかし、観光客が観光行動の際に参考とする媒体はネット上の個人旅行記に限らない。観光客向けの媒体に旅行雑誌がある。旅行雑誌は編集された媒体であるため、ネット上の個人旅行記とは発信されている観光情報や観光行動が異なると考えられる。そこで、ネット上の個人旅行記と旅行雑誌の構造の違いを明らかにする。

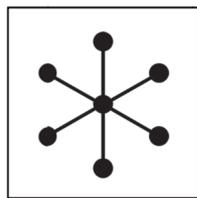


図-1 パターンA

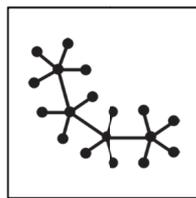


図-2 パターンB

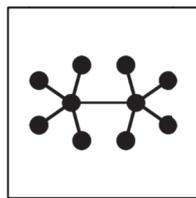


図-3 パターンC

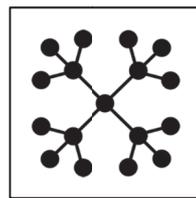


図-4 パターンD

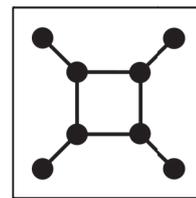


図-5 パターンE

4.3 旅行雑誌と個人旅行記の構造の差異

ネット上の個人旅行記は、フォートラベルで2013年に分類されている北海道旅行記を全件取得し、分析対象とした。旅行雑誌は「まっふる北海道'13」を分析対象とした。まっふる北海道'13は紙媒体であるため、「旅の手帖」と同様の手順でデジタルテキストデータへ変換し、不要なページは削除した。両媒体を対象に品詞を「名詞」に限定し形態素解析を行い、品詞細分類から「一般・サ変接続・自立・固有名詞・形容動詞語幹・ナイ形容詞語幹」に分類された語句を抽出した。結果は頻度で順位付けし、昇順に100位までを分析対象とした。形態素解析の結果から、ネット上の個人旅行記は「移動に関する語句」と「宿泊に関する語句」が特徴的に見いだせる。「移動に関する語句」は「到着」「バス」「駅」「出発」「車」などがある。宿泊に関する語句は「ホテル」「部屋」「宿泊」がある。まっふる北海道'13は「食に関する語句」が特徴的に見いだせる。「料理」「野菜」「丼」「魚介」「食材」「カフェ」などがある。

形態素解析で得られた特徴が文中においてどのような語句と関係性を構築し、イメージ形成しているのかを明らかにするためネットワーク分析を行った。両媒体とも名詞のバイグラムを取得し、品詞細分類が「サ変接続・ナイ形容詞語幹・一般・形容動詞語幹・固有名詞」のいずれかの組合せであるものを抽出し、頻度降順に100組を抽出し、ネットワーク図を作成した。両ネットワーク図の最大連結成分に着目したところ、ネット上の個人旅行記は「北海道」「札幌」「ホテル」

