

口コミビッグデータに人工知能を応用した地域観光の次世代マーケティング 観光客の声に基づいた温泉地の特徴と観光客の価値観の確率モデリング

Next Generation Marketing of Regional Tourism Applying Artificial Intelligence to Big Review Data Probabilistic Modeling of Regionality and Tourists' Values Based on Their Voice to Hot Spring

野守 耕爾^{*1}
Koji Nomori

神津 友武^{*1}
Tomotake Kozu

^{*1} 有限責任監査法人トーマツ デロイトアナリティクス
Deloitte Analytics, Deloitte Touche Tohmatsu LLC

This study summarized tourists' voice in some topics by applying text mining and PLSA to tourists review data. The regional characteristics and tourists' values were analyzed based on the topics and a model of satisfaction factors was constructed by applying Bayesian Network. These approaches will be useful for considering tourism marketing strategies.

1. はじめに

地方創生が国の重要政策課題に位置づけられている昨今、観光はその大きな柱の一つである。地域の観光活性化には観光客の誘客促進が不可欠であるが、そのためには地域の観光資源の特徴と観光客のニーズを理解し、効果的なマーケティング戦略を検討することが求められる。その検討には関連するデータが必要であるが、従来は観光入込客数などの観光統計や民間企業の旅行調査結果が利用されていた。また、最近ではスマートフォンのGPSなど位置に関するビッグデータを取得して、観光ルートを抽出する取り組みも盛んである。しかし、これらは観光客の行動の結果を表すものであり、具体的に観光客はその地域で何を何を感じたのかという行動のコンテキストについては把握できない。このように観光客の行動の内容や価値観に関するデータが現状整備されておらず、地域観光において十分なマーケティングが検討されてこなかった。

本研究では、Web上に投稿された観光地の口コミデータの活用を検討する。口コミは観光客の生の声であり、その観光地で実際に行ったことや感じたことなどが詳細に記述されており、観光客の行動のコンテキストを抽出する上で有用な情報源と考えられる。一方口コミはITの普及拡大に伴い膨大な量のデータが蓄積されており、人間が全て目視で確認し、情報を整理するのは困難である。そこで本研究では口コミというビッグデータに人工知能技術を応用することで、観光客の行動のコンテキストを人間が理解しやすい形に整理し、地域観光のマーケティング戦略に資する知識を抽出することを目的とする。

2. 分析データ

本研究では、フォートラベル株式会社が運営する「旅行のクチコミと比較サイト フォートラベル」(<http://4travel.jp/>)における国内旅行の口コミデータを用いて分析する。本データで得られる情報の概要は以下の通りである。

施設情報(施設名称,施設カテゴリ,地域カテゴリ,住所)
投稿者情報(性別,年代)
口コミ情報(タイトル,コメント,評点,同行者,旅行年月)

本研究の内容は有限責任監査法人トーマツの公式見解を示すものではありません。
連絡先:野守耕爾,有限責任監査法人トーマツ デロイトアナリティクス, koji.nomori@tohmatu.co.jp

施設情報の「施設カテゴリ」とは、景勝地、動物園、温泉、ホテル、レストランといった項目情報であり、運営するフォートラベル社が付与している情報である。

本研究では、旅行時期が2011年1月～2015年12月の5年間の口コミデータ(1,072,570件)のうち、「施設カテゴリ」が「温泉」のデータ(12,564件)を分析の対象とする。

3. 分析方法

本研究では、口コミのテキストデータにテキストマイニング及びPLSAを適用することで口コミのトピックを抽出し、そのトピックをベースとして観光地の特徴や観光客層の価値観の特徴を定量的に分析したり、ベイジアンネットワークを適用することで満足度の要因を分析する。各ステップの概要を以下に示す。

3.1 口コミのトピックを抽出する

口コミのテキストデータに対し、テキストマイニングの係り受け分析を実行し、係り受け表現を抽出する。

抽出した係り受け表現群を列に、その係り受けを構成する単語群を行にとり、その単語と係り受けの共起出現頻度を集計した共起行列(クロス集計表)を作成する。これにPLSAを適用することでその表現群を集約し、口コミトピックを抽出する。PLSAの実行により、トピック T を条件とした単語 W の条件付確率 $P(W|T)$ 、トピック T を条件とした係り受け表現 E の条件付確率 $P(E|T)$ 、トピックの出現確率 $P(T)$ が計算される。抽出されたトピックは、 $P(W|T)$ によって重み付けされた単語や $P(E|T)$ によって重み付けされた係り受け表現から意味を解釈する。

3.2 トピックを変数化する

抽出したトピックをベースとして観光地や観光客層別の特徴、満足度との関係性を分析するため、トピックを変数化する。全口コミに対して、出現する係り受け表現とその表現の各トピックにおける重み $P(E|T)$ に基づいて、その口コミが各トピックの話題をどれくらいしているのかというスコアを計算する。

3.3 観光地及び観光客層の特徴を分析する

観光客の声から抽出された各トピックのスコアを温泉地別に集計することで、各温泉地の特徴を観光客目線で考察する。同じく各トピックを投稿者の属性別に集計することで、各観光客層が関心を持っている話題を考察する。観光地の分析から、その温泉地の特徴や周辺温泉地と差別化できる特徴を把握し、また

その特徴に関心を持っている観光客層を観光客の分析から把握することで、各観光地のターゲット層を検討できる。

3.4 満足度に影響を与える要因と効果を分析する

口コミの評点に対して、投稿者の属性やトピック変数との関係をベイジアンネットワークによりモデル化し、満足度に影響を与える要因とその効果を考察する。満足度を高める要因を把握することで効果的なプロモーション施策を検討できる。

4. PLSA の新たな応用

4.1 PLSA の概要と特徴

PLSA(Probabilistic Latent Semantic Analysis)は、文書分類のために開発された次元圧縮手法である[Hofmann 1999]。文書 D とそこに出現する単語 W の間には潜在的な意味クラス C があることを想定し、各文書における単語の出現頻度が記録された「文書」×「単語」の共起行列データを学習し、文書と単語の共通トピックとなるような特徴を見つける手法である。PLSA の実行により 3 種類の確率変数 $P(D|C)$, $P(W|C)$, $P(C)$ が計算され、これにより「文書」×「潜在クラス」という低次元データに変換でき、クラスタリングの手法としても用いられる。

PLSA が他のクラスタリング手法と異なる特徴の一つは、行と列を同時にクラスタリングできることである。一般的なクラスタリング手法は、列をベースに行をクラスタリングする、あるいは行をベースに列をクラスタリングするため、どちらか一方しかクラスタリングできない。一方 PLSA で抽出される潜在クラスには、行の要素と列の要素が同時に所属することができる。これにより、抽出されたクラスは 2 つの軸から構成され、2 つの軸からクラスターの意味解釈をすることができる。

4.2 本研究における PLSA の新たな応用

本来の PLSA の適用では、「文書」×「単語」という共起行列をインプットとするが、行に設定された「文書」はつまりは文書 ID であり、それ自体に意味を持たない情報であるため、抽出された潜在クラスの意味解釈には使用されない情報である。行と列を同時にクラスタリングできる PLSA では、行と列は双方が十分意味を持つ情報で構成すれば、抽出された潜在クラスの意味を 2 つの軸から解釈することができる。

共起行列の構成を工夫して PLSA を適用する例として、「品詞」×「品詞」の共起行列を用いる方法が提案されており、有用な知識が抽出されたことが報告されている [Kameya & Sato 2005][野守 2014]。しかし「品詞」と「品詞」という単語同士の共起だけで潜在クラスを抽出した場合、それぞれの単語は文法的に関連して出現しているとは限らないため、文脈が理解できず解釈が困難なことがある。また筆者らは、全国の観光地の口コミから得られた「観光地」×「係り受け表現」の共起行列に PLSA を適用することで観光地のテーマを抽出しており[野守 2015]、共起行列の軸の一方を「係り受け表現」とすることで、文脈をイメージしやすくクラスの解釈がより容易になった。

本研究では、口コミの中で話されているトピックを顕在化させるため、口コミにテキストマイニングを実行して抽出した係り受け表現群を列に、その係り受けを構成する単語群を行にとり、その単語と係り受けの共起出現頻度を集計した共起行列(クロス集計表)を作成し、これに PLSA を適用する。単語と係り受けを同時にクラスタリングすることで、単語という話題の観点となる軸に基づいて、その観点的具体的な話題となる係り受け表現をグルーピングでき、より文脈上近い係り受け表現でまとめられた口コミのトピックを潜在クラスとして抽出できると考えられる。

5. 口コミのトピック抽出

温泉の口コミにテキストマイニング(係り受け分析)を実行し、「単語」×「係り受け表現」の共起行列を作成して PLSA を適用することで、温泉観光で話題にされているトピックを抽出する。

5.1 係り受け表現の抽出

全 12,564 件の温泉に関する口コミデータのタイトルとコメント本文を結合させたテキストデータに対して、係り受け分析実行し係り受けペアを抽出した。本研究では、観光における観光客の体験や評価に関連する表現を抽出するため、名詞と動詞(サ変接続名詞含む)及び名詞と形容詞(形容動詞含む)という組み合わせの係り受け表現を抽出した。特に出現頻度が 5 件以上の係り受け表現に限定したところ、4,981 件の表現が抽出された。

なお、本研究では、株式会社 NTT データ数理システムの Text Mining Studio 5.0 を使用し係り受け分析を実行した。

5.2 PLSA 適用による口コミトピックの抽出

抽出した係り受け表現を列に、その係り受けを構成する単語を行に取り、一文章当たりの共起出現数をカウントした「単語」×「係り受け」の共起行列を作成し、PLSA の学習データとした。ここで、「ある(動詞)」「や」思う(動詞)など、全体の出現頻度が非常に高く、トピックの解釈において重要な意味を持たないと考えられるものはノイズとして除去した。その結果、「単語: 1,648 語」×「係り受け: 4,804 表現」の共起行列が構成された。

PLSA の実行では、あらかじめクラス数を指定する必要があり、また初期値によって結果が異なる初期値依存性がある。そこで最適なクラス数の解を得るため、クラス数を 15 から 35 まで 1 刻みで変化させ、それぞれ PLSA を 5 回ずつ初期値を変えて実行し、各実行解の AIC を計算した。その結果、AIC がクラス数に対して下に凸のカーブを描き、クラス数 23 の実行解の平均 AIC が最小となったので、クラス数 23 の 5 回の実行解の中で AIC 最小となる解を採用することとした。本研究における PLSA では、単語 W_i と係り受け表現 E_j の背後にある潜在クラス(口コミトピック) T_k を想定しているため、 $P(W_i|T_k)$, $P(E_j|T_k)$, $P(T_k)$ が計算された。

なお、本研究では、国立研究開発法人産業技術総合研究所が開発した APOSTOOL2 の PLSA プログラムを使用した。

抽出された 23 個のクラスについて、所属確率の高い単語と係り受け表現から、そのクラスを意味するトピックを解釈した。単語はそのトピックの観点を示し、係り受け表現はその具体的な内容を示しているといえる。23 個のトピックのうち 2 つを例に、所属確率の高い単語及び係り受け表現についてそれぞれ上位 10 件を列挙したものを表 1 に示す。T04 は古い建物など雰囲気が良いことに関する話題、T 17 は寒い時期などで体を温めることに関する話題と解釈できる。このように 23 個のトピックの意味を解釈した結果を表 2 に示す。観光客は温泉地で様々な話題をしているが、景色や雰囲気、湯の泉質や温度、足湯や砂湯、食事やアメニティなど、23 個の話題に集約されることが分かった。

表 1 抽出されたトピックとその構成の例

T04		T17					
P(W T)	単語	P(E T)	係り受け	P(W T)	単語	P(E T)	係り受け
12.5%	雰囲気	3.0%	雰囲気->良い	13.7%	体	3.3%	体->温まる
6.8%	建物	1.1%	良い->雰囲気	9.0%	温かる	2.6%	湯->温かる
6.5%	温泉地	1.0%	落ち着く->雰囲気	7.3%	温まる	2.6%	体->洗う
4.1%	きれい	1.0%	レトロ->雰囲気	4.1%	つく	2.2%	芯->温まる
3.0%	古い	0.9%	雰囲気->味わみできる	3.5%	寒い	2.1%	温泉->温かる
2.5%	共同浴場	0.8%	清掃->行き届く	2.9%	日	2.0%	体->温める
1.7%	新しい	0.8%	建物->古い	2.4%	洗う	1.7%	冷える->体
1.7%	風情	0.7%	温泉地->ある	2.0%	冬	1.3%	寒い->日
1.4%	銭湯	0.7%	古い->建物	1.2%	芯	1.2%	寒い->時期
1.2%	旅館	0.7%	野趣->あふれる	1.8%	温める	1.0%	冬->寒い

表 2 PLSA で抽出された 23 個のトピックの解釈

ID	P(T)	トピック名称	ID	P(T)	トピック名称
T01	7.3%	露天風呂からの眺め	T13	4.2%	自己・他社の利用
T02	6.4%	湯船の特徴	T14	4.2%	湯の温度
T03	5.4%	入浴受付の説明	T15	4.1%	足湯
T04	5.1%	雰囲気の良い	T16	3.8%	宿泊でも日帰りでも楽しめる
T05	5.1%	訪問した時間帯	T17	3.4%	体を温めること
T06	5.0%	アクセス	T18	3.4%	砂湯、有名な温泉
T07	5.0%	良い温泉、疲れの癒し	T19	3.4%	充実した食事・休憩所
T08	4.7%	設備の広さ・充実性	T20	3.3%	綺麗な施設、家族で楽しめる
T09	4.5%	安価、無料入浴	T21	3.2%	人の多さ
T10	4.5%	目的地までのルート	T22	2.9%	アメニティの有無
T11	4.4%	泉質・湯量	T23	2.5%	立ち寄り温泉
T12	4.2%	温泉に入ること			

6. トピックの変数化

PLSA で抽出したトピックに対して、各口コミデータがそのトピックをどれくらい話題にしているのかというスコアを付与し、変数として扱える形に処理する。これにより、投稿者の属性別にトピックを集計したり、評点とトピックとの関係をモデル化するなど、トピックに関連する各種統計解析が可能となる。

1 件の口コミレコードは複数の文章から構成されるが、文章によって記述されている話題が異なることがある。そこで本研究では、文章単位にトピックのスコアを計算し、その後レコード単位に各トピックのスコアを集約する。

文章 D_h におけるトピック T_k のスコアとして $P(D_h|T_k)$ を考える。この条件付確率はそのトピックをよく話題にしている文章ほど高くなるが、これをトピックを構成する「係り受け」に基づいて定義すると、 $P(D_h|T_k)$ は式(1)で計算される。係り受け表現 E_j が含まれる文章の数を $n(E_j)$ とすると、 $P(D_h|E_j)$ は $n(E_j)$ の逆数として計算される。 $P(E_j|T_k)$ は PLSA の実行結果によって得られている。また $P(D_h|T_k)$ は多くの係り受け表現が含まれる長い文章では値が高くなる傾向があり、この値だけではトピックと文章の関係の強さが分かりにくい。そこで事後確率 $P(D_h|T_k)$ を事前確率 $P(D_h)$ で除した $P(D_h|T_k)/P(D_h)$ をもって文章 D_h におけるトピック T_k のスコアと定義する。なお $P(D_h)$ は式(2)で算出でき、 $P(T_k)$ は PLSA の実行解によって得られている。

$$P(D_h|T_k) = \sum_j P(D_h|E_j)P(E_j|T_k) \quad (1)$$

$$P(D_h) = \sum_k P(D_h|T_k)P(T_k) \quad (2)$$

7. 観光地及び観光客の特徴分析による誘客ターゲットの設定

各口コミデータに対してそのトピックの話題をしていれば 1、そうでなければ 0 というフラグを付与し、その分布を観光地ごと及び観光客(投稿者)の属性ごとに集計する。これにより観光客の声から生まれたトピックをベースに観光地の特徴を把握し、また各トピックに関心を強く持っている観光客層を把握することで、各観光地のターゲット層を設定することができる。

なお、「そのトピックの話題をしている」という判断は、算出したトピックスコアが 3 以上と設定した。スコアは事前確率に対する事後確率の比率を表しているため、1 以上であればそのトピックとの関係があるとすることはできるが、実際にコメントとスコアの対応を確認したところ、スコア 1 付近の口コミにはそのトピックの話題がされていないと思われるノイズコメントも存在しており、スコア 3 以上となればほとんどがそのトピックを象徴するコメントであった。本研究では抽出したトピックの各変数に対する影響を考察するため、なるべくノイズは除外し、そのトピックを象徴する口コミのみ「そのトピックを話題にしている」と判断することとし、その閾値を厳しく 3 に設定した。

7.1 観光地の特徴分析

口コミの対象となった施設が所在する地域カテゴリについて、トピックスコアのフラグ=1 の割合を集計し、全体の割合と比較した。結果を表 3 に示す。表 3 は口コミ件数が上位の地域カテゴリに限定したものであるが、例えば九州で有名な別府温泉、黒川温泉、由布院温泉を比較すると、全体と比較して別府温泉は「T18:砂湯、有名な温泉」の話題が多く、黒川温泉は「T04:雰囲気の良い」や「T16:宿泊でも日帰りでも楽しめる」の話題が多く、由布院も「T04:雰囲気の良い」の話題は多いが、「T21:人の多さ」の話題も多いことが分かる。これらの温泉地は互いに近い場所にあるが、このように比較することで観光客の目線ではそれぞれ何が差別化要素となっているのか把握することができる。由布院は美しい自然や雰囲気の良さで人気の温泉地だが、観光地化が進み、観光客の多さが目立っていることが考えられる。例えば平日への観光客の分散や、駐車場整備や交通対策といった施策の検討が考えられる。

表 3 温泉地域別でみる各トピックの該当割合

全体	地域カテゴリ	別府温泉	道後温泉	城崎温泉	指宿	草津温泉	熱海	有馬温泉	山形市	横濱温泉	諏訪	黒川温泉	由布院温泉
12,564	口コミ件数	599	449	233	198	158	154	124	119	111	110	109	104
23%	T01	14%	8%	16%	30%	17%	14%	26%	7%	37%	13%	19%	19%
16%	T02	14%	12%	17%	9%	9%	5%	13%	16%	19%	26%	8%	12%
10%	T03	12%	12%	11%	11%	8%	5%	8%	16%	11%	11%	8%	12%
26%	T04	31%	42%	46%	14%	32%	26%	23%	17%	39%	31%	52%	37%
26%	T05	21%	42%	23%	30%	25%	16%	29%	50%	38%	15%	33%	29%
25%	T06	21%	12%	24%	23%	21%	30%	16%	20%	38%	25%	24%	19%
28%	T07	19%	20%	20%	24%	25%	19%	20%	38%	43%	31%	30%	26%
23%	T08	14%	18%	26%	16%	19%	6%	22%	20%	44%	20%	14%	11%
21%	T09	20%	22%	20%	9%	27%	13%	23%	18%	15%	31%	15%	21%
18%	T10	18%	12%	9%	15%	16%	28%	10%	20%	25%	12%	20%	23%
23%	T11	24%	10%	19%	10%	34%	22%	39%	14%	19%	36%	21%	19%
19%	T12	22%	32%	21%	22%	26%	5%	25%	29%	26%	18%	18%	21%
23%	T13	23%	31%	14%	18%	23%	17%	35%	21%	22%	27%	24%	18%
21%	T14	26%	22%	27%	17%	37%	19%	23%	30%	17%	22%	16%	23%
23%	T15	26%	21%	32%	16%	30%	34%	27%	13%	20%	40%	31%	27%
22%	T16	22%	14%	24%	14%	20%	12%	39%	13%	31%	15%	40%	25%
19%	T17	16%	14%	13%	13%	18%	8%	19%	33%	18%	19%	13%	12%
22%	T18	41%	41%	22%	73%	22%	22%	27%	26%	29%	26%	19%	19%
19%	T19	18%	15%	6%	13%	8%	7%	21%	19%	40%	12%	20%	13%
22%	T20	15%	11%	20%	9%	29%	7%	24%	18%	36%	25%	22%	13%
18%	T21	20%	26%	20%	16%	20%	16%	29%	22%	14%	18%	24%	26%
17%	T22	19%	16%	12%	14%	17%	7%	11%	14%	19%	16%	10%	20%
19%	T23	16%	11%	15%	2%	15%	12%	15%	16%	30%	18%	14%	14%

■全体の1.1倍以上 ■全体の1.4倍以上 ■全体の1.7倍以上 ■全体の2.0倍以上

7.2 観光客層の特徴分析

口コミ投稿者の属性情報である性別と年代、及び同行者について、トピックスコアのフラグ=1 の割合を集計し、全体の割合と比較した。結果を表 4 に示す。例えば 7.1 で取り上げた九州の温泉地で特徴的だった T04 と T18 について確認すると、「T04:雰囲気の良い」の話題については、女性 60 代やシニアの夫婦旅行者、乳幼児連れ家族旅行者の関心が強く、「T18:砂湯、有名な温泉」の話題については、女性全般やシニアの夫婦旅行者、家族旅行者の関心が強いことが分かる。

表 4 観光客の属性別でみる各トピックの該当割合

全体	観光客の属性	女性										男性				同行者			
		20代	30代	40代	50代	60代	20代	30代	40代	50代	60代	乳幼児連れ家族旅行者	家族旅行者	カップル・夫婦(2人)	友人	一人旅			
12,564	口コミ件数	577	1,733	1,554	762	192	337	1,623	2,509	2,172	903	349	1,817	2,436	543	1,408	3,548		
23%	T01	24%	25%	26%	25%	26%	20%	23%	20%	24%	26%	25%	24%	24%	24%	27%	20%		
16%	T02	12%	14%	17%	13%	16%	9%	16%	20%	15%	17%	14%	13%	17%	16%	14%	16%		
10%	T03	9%	10%	10%	11%	6%	7%	11%	13%	8%	10%	9%	10%	11%	11%	11%	11%		
26%	T04	27%	27%	26%	26%	31%	24%	27%	24%	24%	25%	30%	27%	28%	33%	25%	25%		
26%	T05	32%	30%	28%	29%	21%	31%	24%	24%	24%	18%	31%	27%	27%	23%	29%	23%		
25%	T06	23%	25%	21%	27%	23%	28%	30%	26%	24%	23%	21%	24%	24%	26%	26%	26%		
28%	T07	26%	32%	31%	31%	36%	28%	28%	26%	25%	23%	33%	31%	31%	39%	28%	30%		
23%	T08	27%	26%	23%	27%	15%	22%	23%	28%	22%	19%	32%	25%	23%	22%	26%	21%		
21%	T09	20%	22%	22%	20%	19%	18%	21%	23%	17%	23%	23%	19%	20%	27%	21%	22%		
18%	T10	14%	14%	17%	18%	16%	16%	20%	20%	19%	21%	13%	17%	17%	18%	16%	21%		
23%	T11	14%	21%	23%	21%	31%	14%	24%	25%	25%	30%	20%	22%	26%	28%	21%	25%		
19%	T12	18%	20%	21%	19%	19%	14%	19%	20%	17%	18%	20%	18%	21%	19%	20%	18%		
23%	T13	26%	27%	24%	28%	24%	21%	23%	21%	21%	23%	25%	23%	23%	23%	23%	23%		
21%	T14	24%	25%	23%	21%	20%	17%	22%	21%	20%	16%	26%	21%	23%	22%	22%	21%		
23%	T15	21%	24%	24%	25%	26%	19%	27%	23%	23%	19%	27%	26%	24%	23%	23%	24%		
22%	T16	24%	24%	22%	21%	20%	17%	23%	20%	20%	25%	26%	25%	24%	32%	22%	19%		
19%	T17	22%	22%	21%	23%	18%	15%	18%	20%	17%	16%	20%	20%	21%	21%	21%	18%		
22%	T18	27%	25%	25%	24%	24%	23%	21%	21%	19%	18%	23%	24%	23%	27%	23%	20%		
19%	T19	23%	22%	22%	21%	13%	15%	18%	19%	16%	18%	28%	19%	21%	20%	21%	18%		
22%	T20	24%	25%	23%	23%	13%	18%	21%	23%	19%	21%	37%	25%	22%	24%	22%	21%		
18%	T21	20%	21%	19%	19%	20%	17%	19%	16%	20%	17%	18%	19%	20%	15%	19%	18%		
17%	T22	21%	21%	20%	20%	13%	14%	13%	16%	13%	16%	19%	16%	19%	18%	17%	15%		
19%	T23	22%	21%	18%	20%	17%	15%	19%	18%	18%	18%	20%	19%	19%	22%	18%			

■全体の1.1倍以上 ■全体の1.2倍以上

7.3 ターゲットの設定

7.1 と 7.2 の結果を総合すると、砂湯が特徴である別府温泉は女性客やシニア夫婦旅行者や家族旅行者がターゲット層に

なりえ、雰囲気の良い特徴である黒川温泉と由布院温泉は60代やシニア夫婦旅行者、乳幼児連れ家族旅行者がターゲット層になりえと考えられる。このように温泉地の特徴を把握し、またその特徴に関心を持っている観光客層を把握することで、各観光地のターゲット層を設定することができる。

8. 観光客満足度の要因分析による効果的なプロモーション設計

口コミで投稿者が付与した評点に対して、投稿者の属性や抽出したトピックとの関係をベイジアンネットワークによりモデル化し、満足度に影響を与える要因とその効果を考察する。

8.1 満足度モデルの構築

フラグを付与した各トピック変数と観光客の各属性変数(性別、年代、同行者の各カテゴリをフラグ化した変数)を親ノードの候補に設定し、口コミの評点を子ノードに設定して両者の確率的関係をベイジアンネットワークでモデル化した。なお口コミの評点は1点から5点まで0.5刻みで投稿者により付与されているが、本研究では4.5点以上の高満足か否かというフラグ変数を子ノードの対象とし、特に高満足を得られる要因についてモデル化した。構築したモデルを図1に示す。図1より、高満度に影響を与えるのは、「性別」「カップル・夫婦旅行」「T11:泉質・湯量」「T17:体を温めること」「T18:砂湯、有名な温泉」であった。

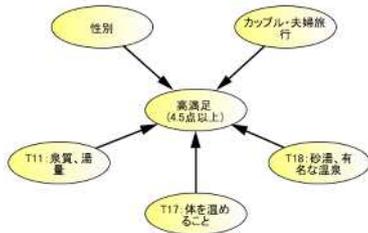


図1 満足度モデル

8.2 モデルの確率推論による高満足の効果の考察

構築したモデルを用いて、各変数の状態を指定した条件下での高満足の確率を推論した。結果を図2に示す。図2より関係のあった3つのトピックは話題にされることで高満足の確率が上昇しており、性別では女性の方が、同行者はカップル・夫婦旅行の方が高満足の確率が上昇する結果となった。また、性別と同行者を組み合わせた条件で推論をした結果を図3に示す。図3より男性はカップル・夫婦旅行の方が、女性はカップル・夫婦旅行でない方が高満足の確率が高く、誰と温泉旅行に行くかという要因が高満度に与える影響は男女で異なることが分かる。続いて、その満足度の高い男性のカップル・夫婦旅行と女性のカップル・夫婦でない旅行をそれぞれ条件としたときの、3つのトピックの高満足への影響を推論した結果を図4に示す。図4より、男性のカップル・夫婦旅行で最も高満足の確率を押し上げたのは「T17:体を温めること」であり、女性のカップル・夫婦でない旅行で最も高満足の確率を押し上げたのは「T18:砂湯、有名な温泉」であった。このように男女によって満足度への価値観が異なることが定量的に把握できる。

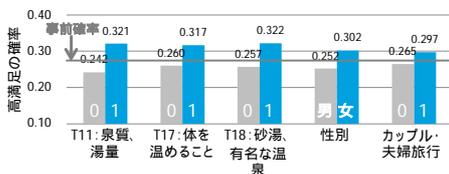


図2 各変数を条件としたときの高満足確率の推論

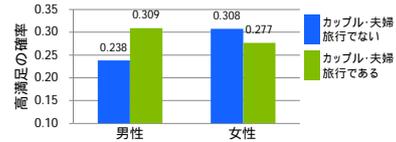


図3 性別と同行者を条件としたときの高満足確率の推論

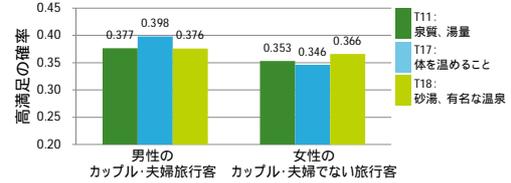


図4 男性のカップル・夫婦旅行者と女性のカップル・夫婦でない旅行者で各トピックを条件としたときの高満足確率の推論

8.3 モデルの結果を用いた効果的なプロモーション設計

観光客の属性によって高満足を向上させるトピックが異なることが分かったが、この結果により、ターゲットに応じた効果的なプロモーションを検討できる。例えば、男性向けの雑誌では、恋人との温泉旅行を提案し、特に寒い中で体を温めるということや雪の中で湯気が立ちのぼる露天風呂の写真を通じてPRしたり、女性向けの雑誌では、友人や家族との温泉旅行を提案し、特に砂をかけて体を蒸す砂湯を楽しむ姿のPRなどが考えられる。

9. まとめ

本研究では観光客の行動のコンテキストを理解し、地域の観光マーケティング戦略に資する知識を抽出するため、観光客の声である口コミにPLSAを適用してトピックを抽出し、そのトピックをベースとして観光地や観光客層の特徴を分析したり、口コミの評点との関係をベイジアンネットワークによりモデル化して満足度の要因を明らかとし、その効果を考察した。

本アプローチにより、地域は自身の特徴と周辺地域の特徴を観光客の目線で理解することができ、何がその地域で特徴的な観光価値となっており、それに関心を持つのはどのような観光客層なのかということを定量的に把握し、その地域の誘客ターゲット層を検討することができる。また、満足度向上に影響を与える要因と効果を定量的に把握することができるので、ターゲット層に応じて、効果的なプロモーションを検討することができる。このように観光客の行動の内容や価値観に関するデータに人工知能技術を応用して知識化することで、地域観光のマーケティング戦略をエビデンスベースに実施することが期待できる。

参考文献

- [Hofman 1999] Hofmann, T.: Probabilistic latent semantic analysis, Proc. of Uncertainty in Artificial Intelligence, pp. 289-296, 1999.
- [Kameya & Sato 2005] Kameya, Y., & Sato, T.: Computation of probabilistic relationship between concepts and their attributes using a statistical analysis of Japanese corpora, Proceedings of Symposium on Large-scale Knowledge Resources, 65-68, 2005.
- [野守 2014] 野守耕爾, 神津友武: 三位一体アプローチによるテキストデータモデリング法の開発 宿泊施設の口コミデータを用いた評価推論モデルの構築, 2014年度人工知能学会全国大会論文集, 2014.
- [野守 2015] 野守耕爾, 神津友武: 口コミデータにPLSAを適用した観光客目線による観光地分析, 2015年度人工知能学会全国大会論文集, 2015.