

テキストデータを用いた施策検討の切り札となる 確率モデルの応用

ロコミビッグデータの解析による地域の観光マーケティング検討を事例に

有限責任監査法人 トーマツ
デロイトアナリティクス
野守耕爾

2016年11月18日

デロイトアナリティクスとは、 デロイトトーマツが提供するデータ分析のコンサルティングサービスです

デロイトアナリティクスとは？

デロイトトーマツグループの各事業分野と連携したデータ活用コンサルティングのサービスです

デロイトトーマツグループ

国内40都市に約8,700名の専門家



グループ横断であらゆる業務にアナリティクスを適用

デロイトアナリティクス



- ビジネス領域の専門家と、分析の専門家が協働してサービス提供します
- 既存の手法を適用するだけでなく、新たな分析技術の研究・開発も行っています

提供サービス例

ID-POSデータ	×	キャンペーンの最適化
販売データ	×	マーケティングROI最適化
コールセンターログ	×	顧客接点業務の改善
口コミデータ	×	製品差別化の戦略検討
レセプトデータ	×	保健師の業務効率化
アンケートデータ	×	住みやすい市の総合計画
金融取引データ	×	マネーロンダリング識別
特許データ	×	知財戦略の検討
IoTデータ	×	製品保守業務の効率化
画像データ	×	掲載画像の処理効率化

本日のアジェンダ

1. 確率モデルを応用した新たなテキスト分析

2. 適用事例: 口コミビッグデータの解析による地域の観光マーケティング

事例① 全国地域の観光テーマ分析

事例② 温泉地におけるマーケティング分析

3. まとめ

確率モデルを応用した新たなテキスト分析

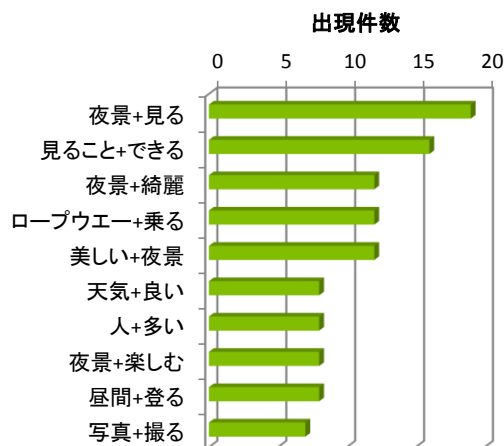
【従来のテキスト分析】

結果の解釈が難しく、状況が変化した場合の結果の再評価はできません

従来の分析例

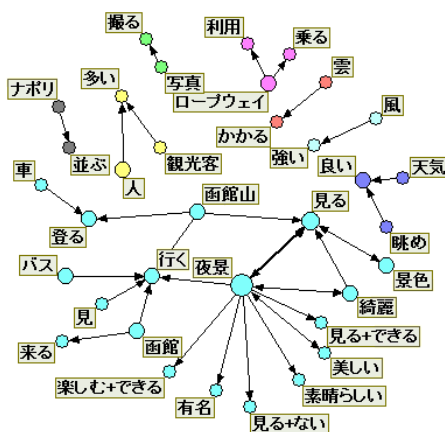
【頻度集計】

単語や係り受け表現の頻度を集計しておおまかな全体像を考察する



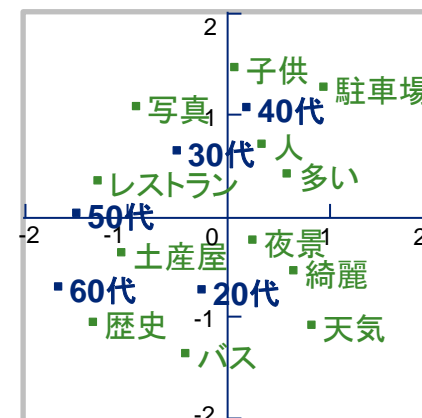
【共起ネットワーク】

同時に出現しやすい単語同士をネットワークでつなぎ、話題を考察する



【コレスポネンス分析】

属性情報と出現単語との対応関係をマッピングし、属性の傾向を考察する



課題

■ 解釈がしづらい

- 単語や表現を単位とした集計ではその周辺状況や背景といった文脈の解釈が難しい
- 同じ意味の表現が複数あり、ある程度グルーピングしないと考察が難しい

■ 現状把握に留まる

- 現状における傾向を考察する分析である
- テキストの内容や属性といった条件が変化すると結果がどの程度変化するか等、状況が変化したときの評価はできない

【課題の解決手段】

他の分析技術を組み合わせることでテキスト分析の課題を解決できます

「解釈がしづらいこと」の解決手段

クラスタリング技術

■ テキストデータの分類

- テキストの内容に基づいてデータを分類する

■ トピックの抽出

- 文脈を考慮した表現の集合体を抽出する
- 「単語」ベースではなく、「話題」ベースで膨大なテキストデータを解釈容易な形に整理する

「現状把握に留まること」の解決手段

モデリング技術

■ テキストデータ内の要因関係の分析

- テキストの話題や、その他属性情報、顧客満足度などの間の定量的な関係を分析する

■ ある条件下での結果のシミュレーション

- 話題が変化したときの効果を推論する
- 効果を最大化させるための要因を把握する

PLSA

確率的潜在意味解析

ベイジアンネットワーク

【PLSA: 確率的潜在意味解析】

データを複数の潜在変数で説明する手法です

PLSAの概要

- 行列データの行の要素xと列の要素yの背後にある共通特徴となる潜在クラスzを抽出する手法である
- 元々は文書分類のための手法として開発されている (Hofman, 1999)
- 各文書の出現単語を記録した文書(行) × 単語(列)という高次元(列数の多い)共起行列データに適用することで複数の潜在トピックを抽出し、文書(行) × トピック(列)という低次元データに変換して文書を分類する

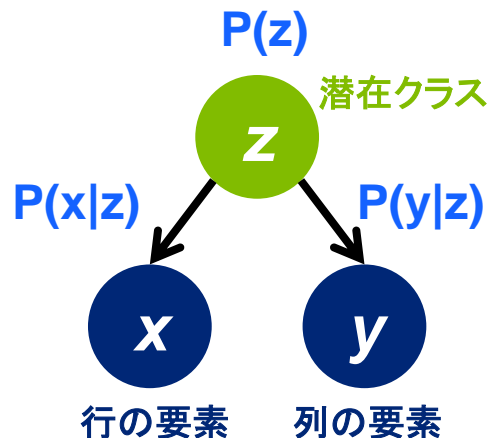
文書ID	単語 1	単語 2	単語 3	...	単語 5,014	単語 5,015
1	0	0	1		1	0
2	1	0	1		0	1
...						



文書ID	トピック 1	トピック 2	...	トピック 11
1	0.09%	0.03%		0.04%
2	0.01%	0.12%		0.06%
...				

例えば数千列ある高次元のデータでも十数個の潜在トピックで説明することができる

PLSAのグラフィカルモデル



- $P(z)$, $P(x|z)$, $P(y|z)$ の3つの確率が計算される
- 潜在クラスzの数はあらかじめ設定する

※条件付確率 $P(A|B)$
事象Bが起こる条件下で事象Aの起こる確率

$$P(x, y) = \sum_z P(z)P(x|z)P(y|z)$$

PLSAのメリット

行と列を同時にクラスタリングできる

潜在クラスには行の変数と列の変数が同時に所属することができるため、2つの軸の情報から潜在クラスの意味を容易に解釈することができる

ソフトクラスタリングできる

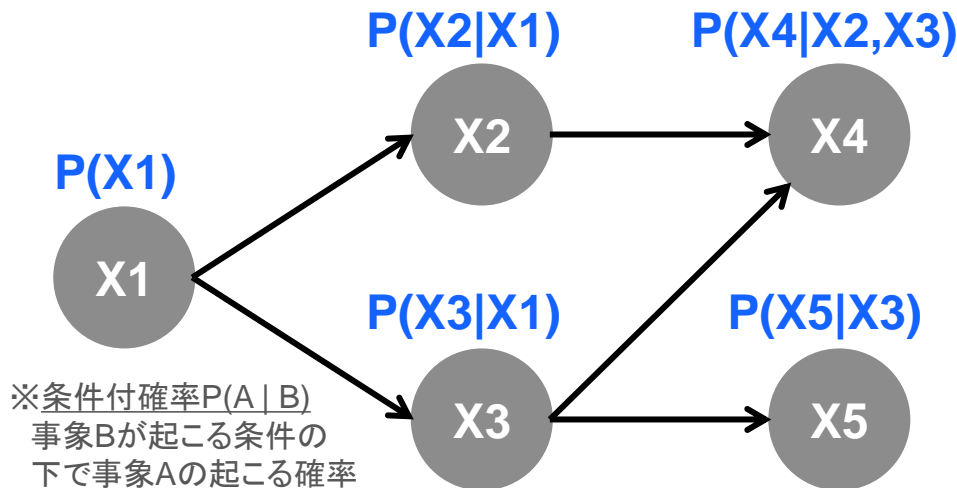
全ての変数が全てのクラスに所属し、その各所属度合いが確率で計算されるため、複数の意味を持つ変数がある場合でも自然と表現できる

【ベイジアンネットワーク】

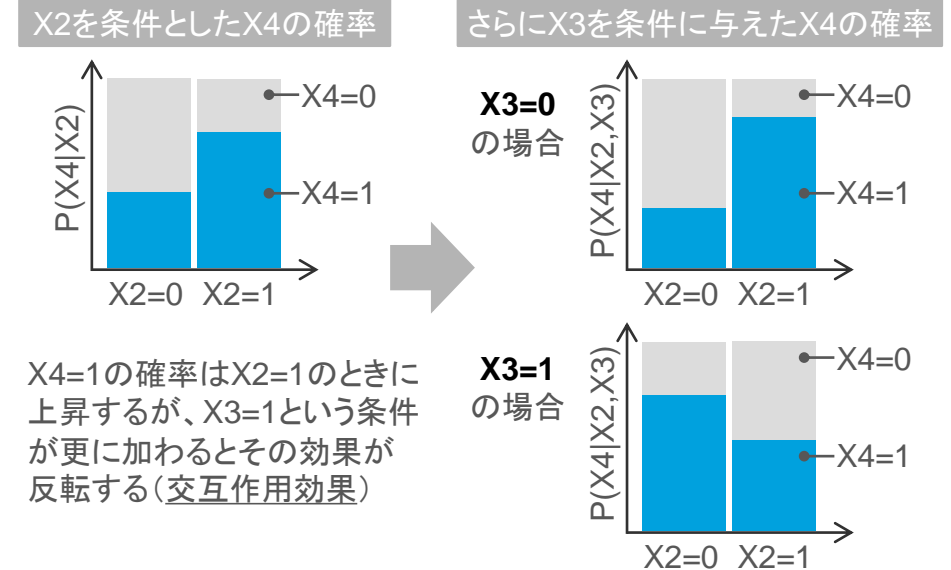
変数の関係をモデル化し、現象の確率シミュレーションをする手法です

ベイジアンネットワークの概要

- 複数の変数の確率的関係をネットワーク構造で表わし、ある変数の状態を条件として与えたときの他の変数の条件付確率を推論することができる
- 目的変数と説明変数の区別はなく、様々な方向から変数の確率推論ができる
- 全ての変数は質的変数(カテゴリカル変数)となるため、量的変数の場合は閾値を設けてカテゴリに分割する
- 確率論の非線形処理によるモデル化のため、非線形の関係や交互作用が生じる現象でも記述できる



確率推論の例



ベイジアンネットワークのメリット

現象を理解して柔軟にシミュレーションできる

目的変数、説明変数の区別なく変数の関係をモデル化するので、現象の構造を理解でき、推論変数と条件変数を自由に指定して確率推論できる

効果を発揮する有用な条件を発見できる

ある条件のときに効果が現れるような交互作用がある場合でも、確率的に意味のある関係としてモデル化することができる

【新たなテキスト分析】

膨大なテキストの解釈を容易にし、確率シミュレーションが可能です

確率モデルを応用した新たなテキスト分析アプローチ

テキストマイニング

- テキストデータ内の単語とその出現頻度を抽出する
- テキストで記された現状の全体像を把握できる

Text Mining Studio

PLSA 確率的潜在意味解析

- データの背後にある潜在要因を抽出する
- 大量で複雑なデータでも理解可能なレベルに変換できる

産業技術総合研究所
APOSTOOL

ベイジアンネットワーク

- 変数間の関係構造を確率的にモデル化する
- 条件を指定した下での結果の確率を推論できる

BAYONET

単語抽出



類型化
トピック抽出



モデリング

膨大なテキストデータを人間が理解しやすい形に整理できる

テキストの内容における複雑な要因関係を構造化できる

条件を変化させたときの結果の挙動をシミュレーションできる

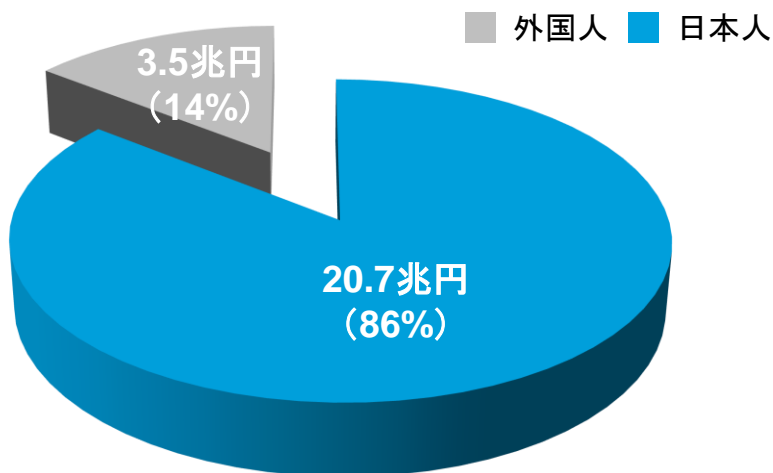
ある事象の発生確率をコントロールする条件を発見できる

【適用事例】
ロコミビッグデータ解析による
地域の観光マーケティング

【地域観光の課題】

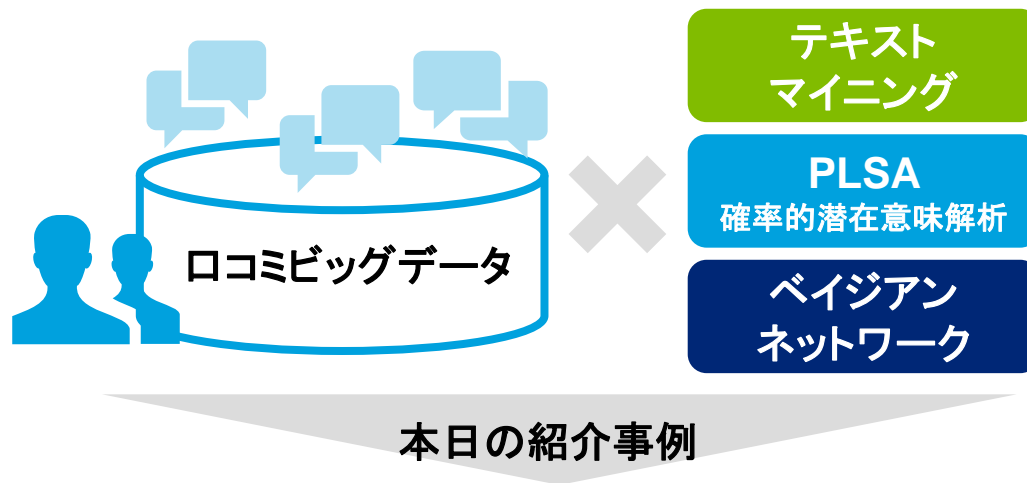
観光客の目線を取り入れた地域の観光マーケティングを検討します

国内における観光市場(2015年)

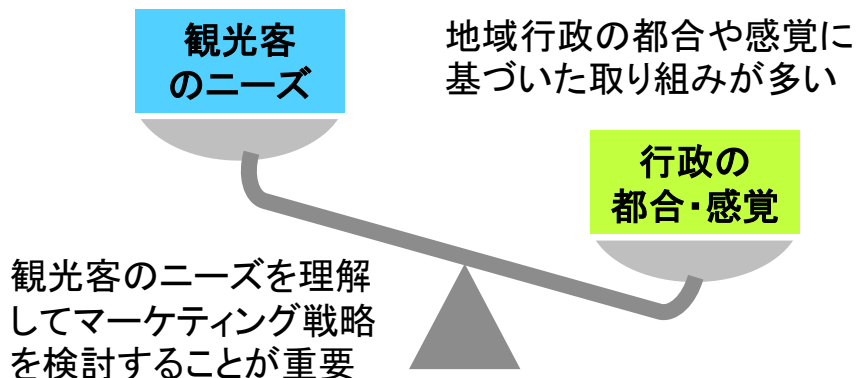


出典: 観光庁「旅行・観光消費動向調査」「訪日外国人消費動向調査」

観光客の生の声“口コミ”の活用



地域観光の課題



1

全国地域の観光テーマ分析

- ①観光テーマの抽出
- ②地域のテーマ性の分析
- ③類似テーマ地域の分析

2

温泉地のマーケティング分析

- ①口コミトピックの抽出
- ②地域の特徴の分析
- ③観光客の価値観の分析
- ④満足度の要因の分析

【使用データ】

旅行口コミサイトの口コミデータを使用します

フォートラベル(<http://4travel.jp/>) の口コミ



函館山のクチコミ 172件

😊 絶景は日没20分後
満足度: ★★★★★ 5.0

👤 わさんぼんさん
男性 / 函館のクチコミ: 1件

旅行時期: 2013/08(約7ヶ月前)



函館に着たらここは外せません。いわゆる100万ドルの夜景です。夜景が有名ですが、昼に来ると函館の景色がはっきり見えるのでこれもまた絶景です。でも夜はやはり感動します。左右に海があるくびれた地形になっているので街の明かりが引き立つんですね。特に日没20分後あたりだと空がいい感じにダークブルーになってとても綺麗です。日没から40分くらいしたら空は真っ暗になってその後あまり景色は変わりません。展望台のいい場所はプロのカメラマンが陣とってしまいますし、休みの日は観光客も多いので、早めの場所取りをおすすめします。

同行者 一人旅

旅行時期が2011年1月～2015年12月の直近5年間の口コミデータ総数
⇒ **1,072,570件**

取得できる主な情報

口コミタイプ	施設所在地	施設所在地域
タイトル	コメント	評点(5点満点)
投稿者の性別	投稿者の年代	同行者
旅行時期(年月)	投稿日時	

北海道 観光 人気ランキング 4,286件



1位 旭川市旭山動物園

カテゴリ: 観光・遊ぶ > 動物園
エリア: 旭川 [地図]
🚗 駐車場に注意
👤 動物の魅せ方が上手です

クチコミ 525件 >

みんなの満足度:
★★★★★ 4.65
アクセス: 3.31
コスト: 4.01
動物・展示物の充実度: 4.43
施設の快適度: 3.88
人混みの少なさ: 2.98

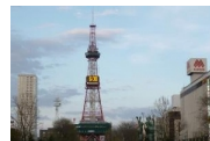


2位 小樽運河

カテゴリ: 観光・遊ぶ > 名所・史跡
エリア: 小樽 [地図]
🌃 夜がオススメ
🚶 混雑

クチコミ 321件 >

みんなの満足度:
★★★★★ 4.54
アクセス: 3.65
人混みの少なさ: 3.10
見ごたえ: 3.67
バリアフリー: 3.14



3位 大通公園

カテゴリ: 観光・遊ぶ > 公園・植物園
エリア: 駅周辺・大通り [地図]
👤 きれいな公園
👤 札幌市内にはあるものの、札幌

みんなの満足度:
★★★★★ 4.54
アクセス: 4.24
人混みの少なさ: 3.40

＜事例①＞ 全国地域の観光テーマ分析

【対象データ】

全国5,553スポットの観光地の口コミデータ169,407件を分析します

分析の対象



2014年8月15日までに投稿された全口コミのうち、10件以上の口コミ投稿のあった観光地を対象とした

【観光テーマの抽出方法】

テキストマイニングとPLSAを応用して日本の観光テーマを抽出します

観光テーマ抽出のアプローチ

テキストマイニングの実行

共起行列の作成

PLSAの実行

テーマの抽出



抽出した係り受け表現が出現するデータ数を観光地毎に集計した、「**観光地×係り受け表現**」の共起行列(クロス集計表)を作成する

係り受け表現

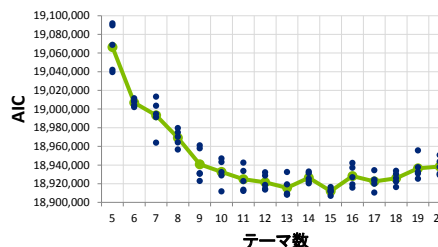
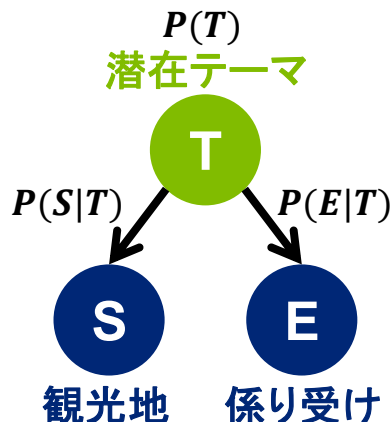
	人↓多い	天気↓良い	写真↓撮る	場所↓ある	⋮
スカイツリー	26	61	38	2	
美ら海水族館	24	10	18	8	
清水寺	41	11	18	3	
首里城公園	11	11	8	3	
...					

観光地

観光地: 5,553箇所
係り受け: 6,843種

※係り受けは頻度30件以上を対象

共起行列にPLSAを適用する



情報量基準AICを用いて最適なテーマ数を選択する

各トピックについて以下の3つの確率が計算される

- ① $P(T)$ 潜在テーマの割合
- ② $P(S|T)$ 各テーマにおける観光地の構成割合
- ③ $P(E|T)$ 各テーマにおける係り受けの構成割合

$P(T_2) = 7.4\%$

$P(S T)$	観光地	$P(E T)$	係り受け
3.8%	東京スカイツリー	2.0%	天気⇒良い
1.3%	函館山	1.7%	景色⇒良い
1.3%	東京タワー	1.2%	高台⇒ある
1.2%	五稜郭タワー	1.0%	きれい⇒見る+できる
1.2%	東京都庁舎展望室	1.0%	夜景⇒きれい
1.2%	さっぽろテレビ塔	0.9%	眺める⇒最高
1.1%	函館山展望台	0.9%	階段⇒登る
1.1%	なごみの塔	0.9%	見晴らす⇒良い
1.0%	グラバー園	0.8%	景色⇒見る+できる
0.8%	港の見える丘公園	0.7%	階段⇒上る

$P(S|T)$ と $P(E|T)$ という2つの軸の要素の重みからテーマの意味を解釈する

【観光テーマ】

日本全国17万件の観光客の声は15個のテーマに集約されました

1 **歴史資料
展示
(城など)**

- | | |
|-------|--------|
| 首里城 | 資料→展示 |
| 熊本城 | 場所→ある |
| 名古屋城 | 歴史→展示 |
| 原爆ドーム | 中→入る |
| 松本城 | 歴史→学べる |

2 **展望景色**

- | | |
|----------|---------|
| 東京スカイツリー | 天気→良い |
| 函館山 | 景色→良い |
| 東京タワー | 高台→ある |
| 五稜郭タワー | きれい→見れる |
| 東京都庁展望台 | 夜景→きれい |

3 **動物園・
水族館**

- | | |
|--------------|---------|
| 美ら海水族館 | 大人→楽しめる |
| 旭山動物園 | 動物→いる |
| 上野動物園 | 間近→見れる |
| アドベンチャーワールド* | 子供→楽しめる |
| 海遊館 | 泳ぐ→姿 |

4 **桜・公園**

- | | |
|--------|-------|
| 上野公園 | ある→公園 |
| 大阪城公園 | 広い→公園 |
| 千鳥ヶ淵 | 桜→咲く |
| 日比谷公園* | 桜→きれい |
| 井の頭公園 | 人→多い |

5 **歴史的
建物**

- | | |
|----------|----------|
| 札幌市時計台 | 重要文化財→指定 |
| 北海道庁旧本庁舎 | 歴史→感じる |
| 大浦天主堂 | 場所→ある |
| 道後温泉本館 | 中→入る |
| 旧函館区公会堂 | ある→建物 |

6 **海**

- | | |
|-------|-------|
| 川平湾 | 天気→良い |
| 古宇利大橋 | 海→きれい |
| 古宇利島 | 海→見れる |
| 万座毛 | 青い→海 |
| 厳島神社 | きれい→海 |

7 **買物街**

- | | |
|--------|-------|
| 横浜中華街 | 店→多い |
| 国際通り | 店→ある |
| 浅草寺 | 店→並ぶ |
| 高山の町並み | 人→多い |
| 仲見世通り | 軒→連ねる |

8 **紅葉・寺**

- | | |
|-----|---------|
| 清水寺 | 紅葉→時期 |
| 金閣寺 | 人→多い |
| 東福寺 | ある→お寺 |
| 東大寺 | 世界遺産→登録 |
| 天竜寺 | 紅葉→きれい |

9 **景勝地**

- | | |
|----------|---------|
| 大涌谷 | 天気→良い |
| 高尾山 | 富士山→見れる |
| オシンコシンの滝 | 紅葉→時期 |
| 富士山 | 滝→見れる |
| 華厳滝 | 車→行く |

10 **テーマ
パーク**

- | | |
|------------|---------|
| ディズニーシー | 大人→楽しめる |
| ディズニーランド | 子供→楽しめる |
| ユニバーサルスタジオ | 人→多い |
| 東京スカイツリー | 写真→撮る |
| ハウステンボス | 十分→楽しめる |

11 **花々**

- | | |
|---------|--------|
| ファーム富田 | 花→咲く |
| 兼六園 | 広大→敷地 |
| 四季彩の丘 | 花→楽しめる |
| 足立美術館 | 花→見る |
| ひたち海浜公園 | ある→美術館 |

12 **散策**

- | | |
|-------|--------|
| 青い池 | 遊歩道→整備 |
| 白川郷 | 気持ち→良い |
| 高千穂峡 | 写真→撮る |
| 奥入瀬溪流 | 流れる→川 |
| 忍野八海 | 遊歩道→歩く |

13 **温泉・
アクセス**

- | | |
|--------|--------|
| 道後温泉本館 | 車→行く |
| 羊が丘展望台 | 温泉→入る |
| 血の池地獄 | 場所→ある |
| ファーム富田 | バス→出る |
| 湯畑 | 本数→少ない |

14 **写真
スポット**

- | | |
|-----------|-------|
| 札幌市時計台 | 写真→撮る |
| 東京駅赤レンガ駅舎 | 橋→渡る |
| 東京タワー | 人→多い |
| 渡月橋 | かかる→橋 |
| 東京スカイツリー | 人→いる |

15 **神社**

- | | |
|--------|--------|
| 伊勢神宮内宮 | ある→神社 |
| 出雲大社 | 祀る→神社 |
| 伏見稲荷大社 | 鳥居→くぐる |
| 伊勢神宮外宮 | 人→多い |
| 太宰府天満宮 | 場所→ある |

※各テーマに関係の強い観光地と表現の上位5位をそれぞれ列挙しており、同じ行にある観光地と表現は互いに対応しているのではない。

【地域のテーマ性の計算】

ベイズの定理を適用して地域のテーマ性を定量的に把握できます

観光地の確率

PLSAで計算された確率

$$P(S|T)$$

各テーマ T における
観光地 S の構成割合

※口コミ件数が多いほど
確率値も高い傾向がある

条件の
反転

ベイズの定理

$$P(T_k|S_i) = \frac{P(S_i|T_k)P(T_k)}{\sum_k P(S_i|T_k)P(T_k)}$$

$$P(T|S)$$

各観光地 S における
テーマ T の構成割合

地域の確率

地域単位に集約された確率

$$P(R|T)$$

各テーマ T における
観光地域 R の構成割合

条件の
反転

ベイズの定理

$$P(T_k|R_h) = \frac{P(R_h|T_k)P(T_k)}{\sum_k P(R_h|T_k)P(T_k)}$$

$$P(T|R)$$

各地域 R における
テーマ T の構成割合

確率の集約

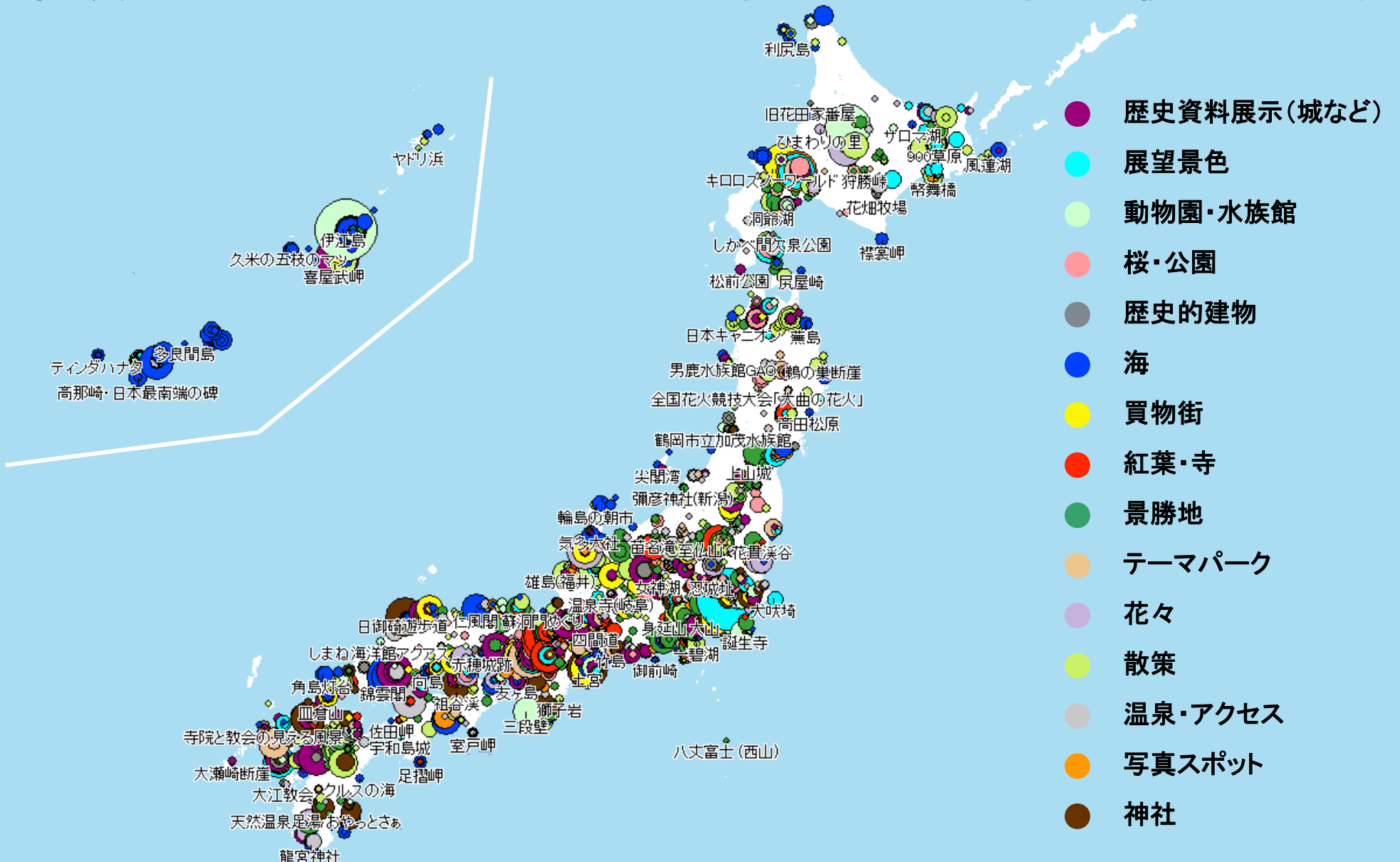
各地域 R に対して、そこに含まれる
観光地 S の確率 $P(S|T)$ を合計する

$$P(R_h|T_k) = \sum_i P(S_{hi}|T_k) \\ (S_{hi} \in R_h)$$

観光地毎、地域毎に15個の観光テーマの構成割合(全体で100%)
を計算することによって、それぞれのテーマ性を定量的に把握できる

【観光テーママップ】

各観光地のテーマをマッピングして全国のテーマ分布を可視化できます

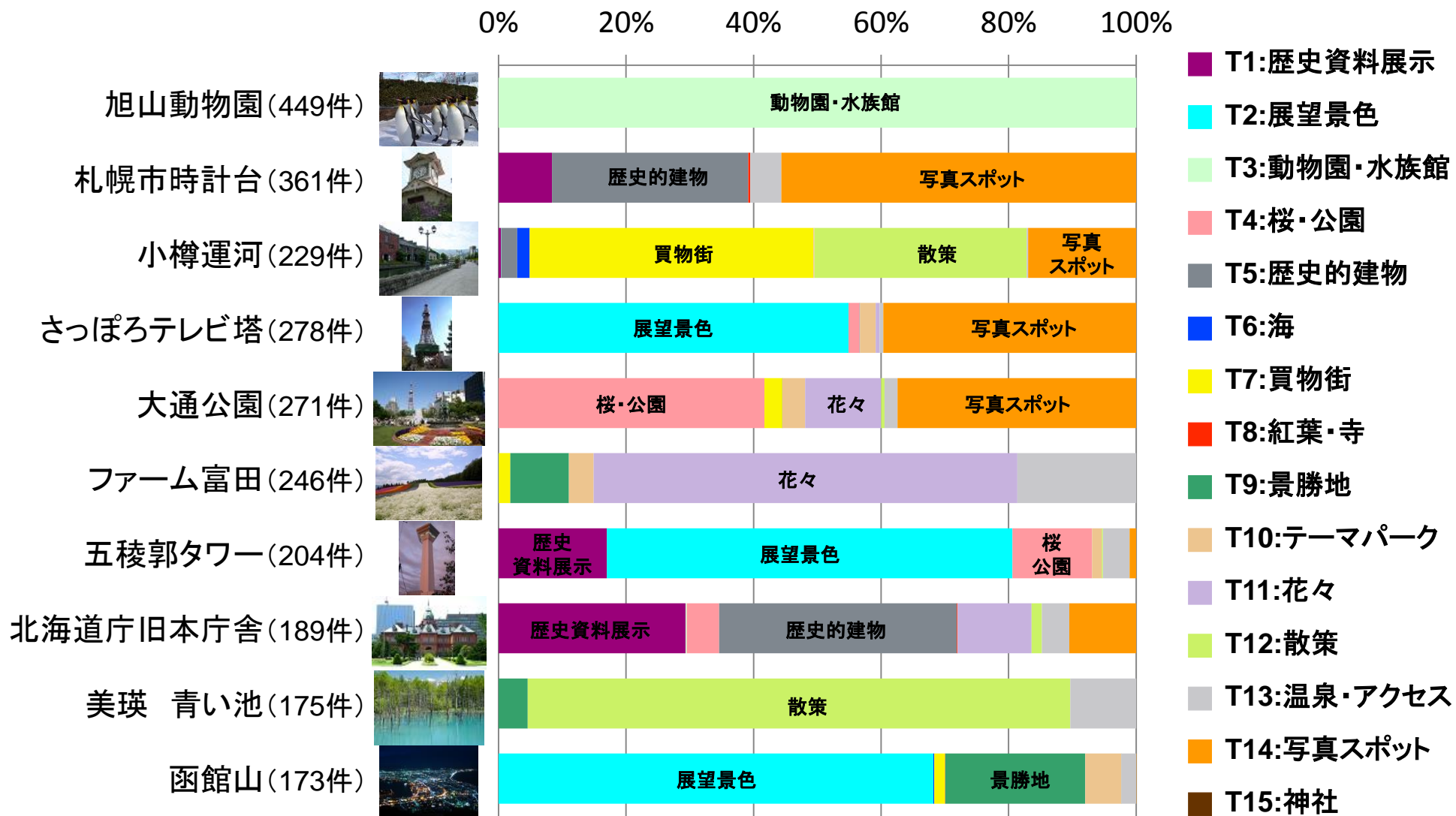


※円の大きさは口コミの多さを表す

【観光地のテーマ性】

各観光地が持つ観光のテーマ性を定量的に把握できます

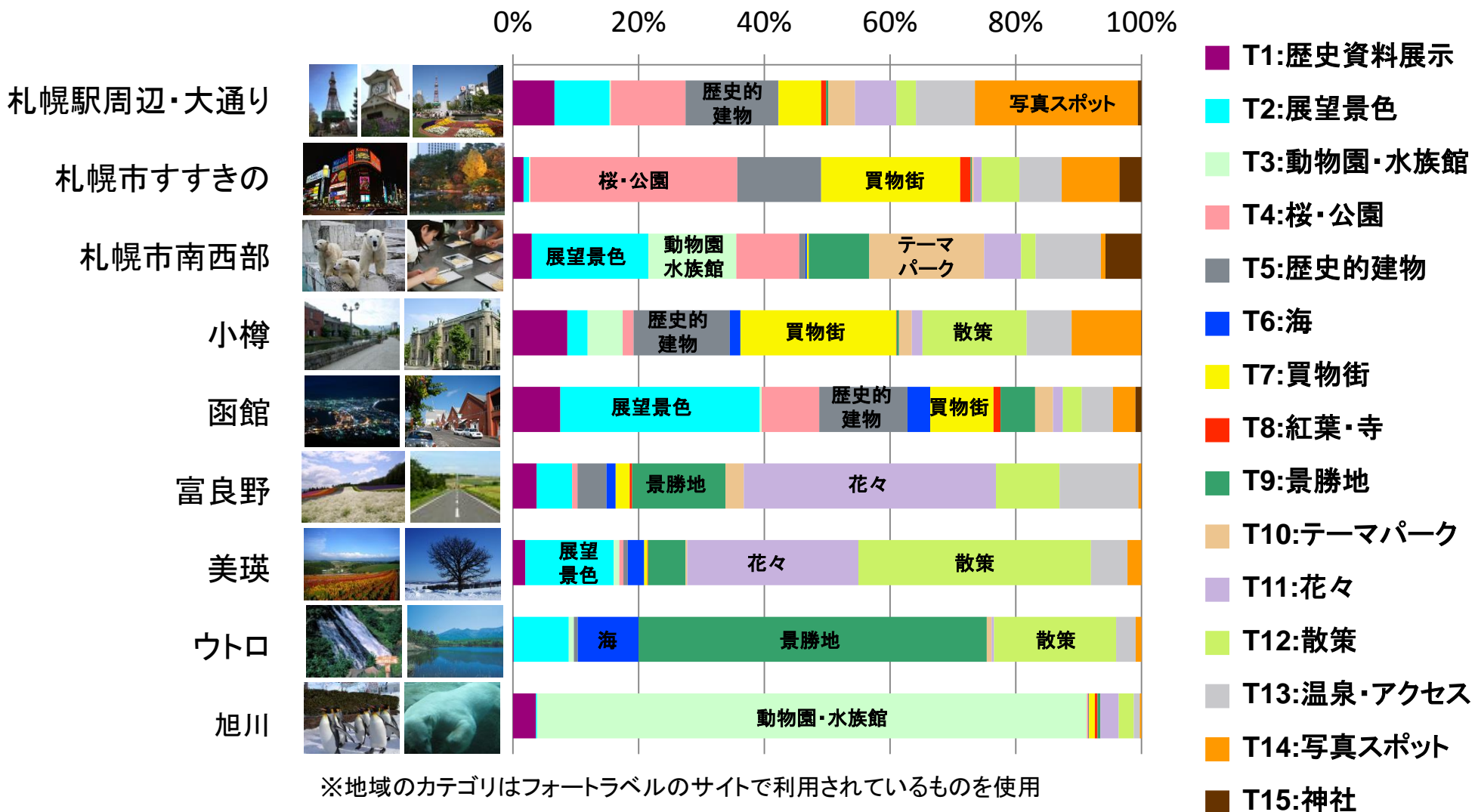
北海道の主要な観光地における15個のテーマの構成割合 P(T|S)



【地域のテーマ性】

各地域が持つ観光のテーマ性を定量的に把握できます

北海道の主要な観光地域における15個のテーマの構成割合 P(T|R)

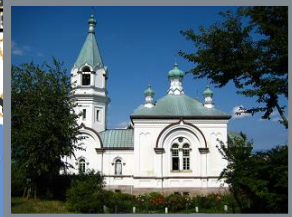


函館エリアの観光マップ

- 円の大きさは口コミの多さを表す
- 各観光地のテーマは、P(T|S)の最も大きいテーマを割り当てた



- T1: 歴史資料展示
- T2: 展望景色
- T3: 動物園・水族館
- T4: 桜・公園
- T5: 歴史的建物
- T6: 海
- T7: 買物街
- T8: 紅葉・寺
- T9: 景勝地
- T10: テーマパーク
- T11: 花々
- T12: 散策
- T13: 温泉・アクセス
- T14: 写真スポット
- T15: 神社



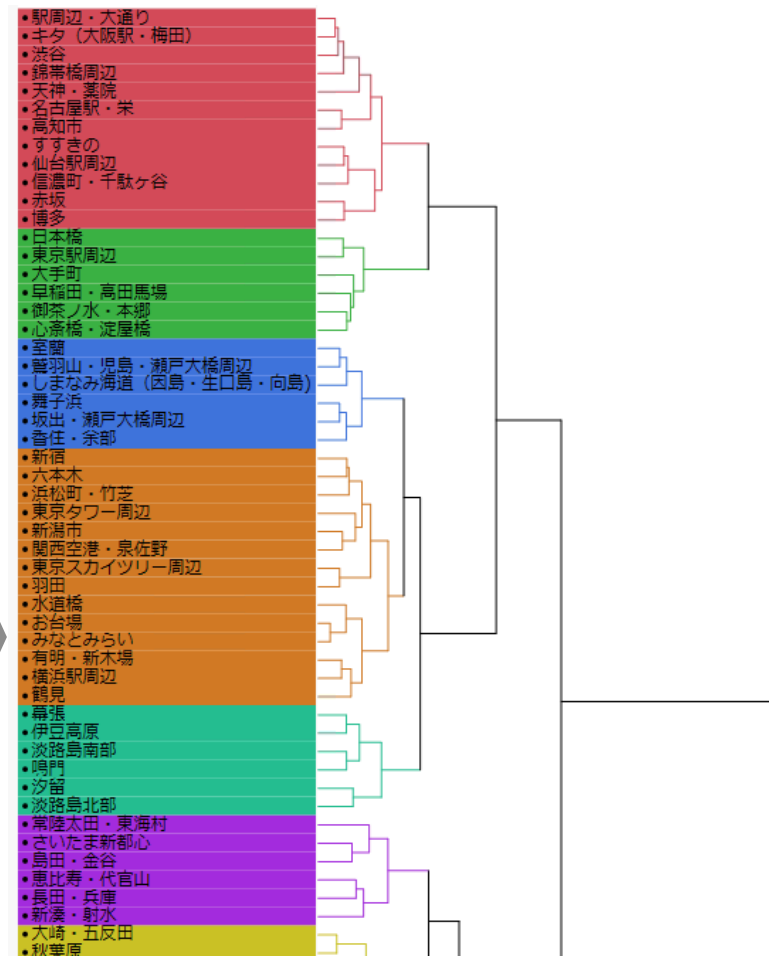
【地域クラスタリング】

地域のテーマ性のデータをクラスタ分析して類似地域を発見します

テーマの類似する地域のクラスタリング

- 地域別に計算した15個のテーマの構成割合のデータに基づいて、階層クラスタ分析を実行する
- 地域間の距離を計算することで、テーマ性が類似する地域を定量的に把握する
- 地域のカテゴリはフォートラベルのサイトで利用されている区町村カテゴリを使用している(927箇所)

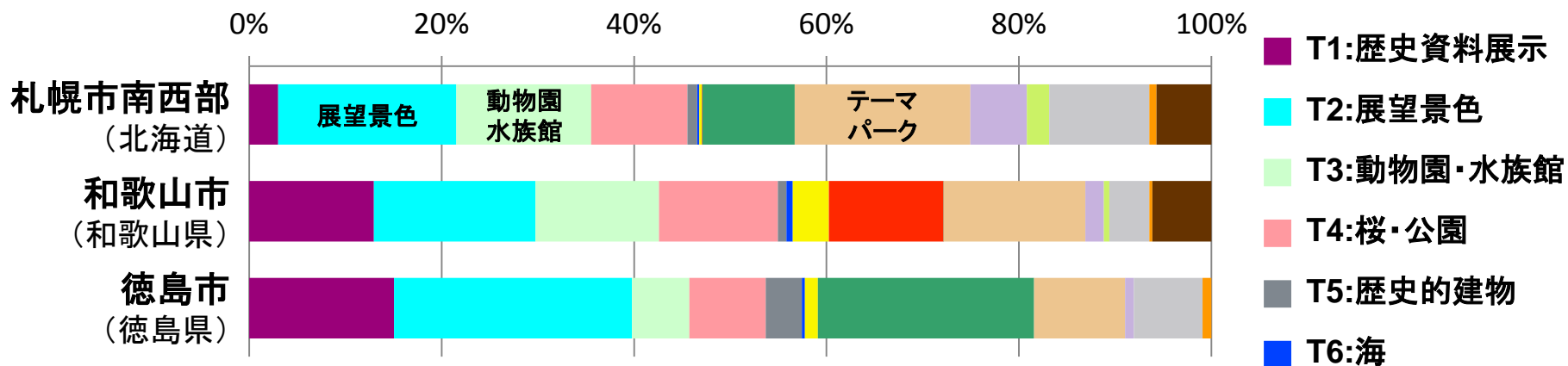
都道府県	地域	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12	T13	T14	T15
北海道	函館	7.5	31.8	0.3	9.1	14.0	3.6	10.1	1.1	5.5	2.8	1.6	3.0	4.9	3.6	1.0
岩手	盛岡	11.2	4.8	0.7	25.6	13.1	0.0	3.9	0.1	1.1	0.0	7.5	7.4	1.8	8.4	14.4
宮城	石巻	10.7	25.7	0.0	18.5	13.1	0.0	0.0	0.0	2.5	10.5	1.7	1.9	6.1	0.0	9.3
福島	会津若松	38.6	13.4	0.4	11.9	11.7	0.0	1.9	3.1	5.1	0.1	3.5	3.8	3.4	0.1	3.0
栃木	那須	0.7	2.0	33.3	2.1	0.5	0.0	0.9	0.0	23.3	10.6	1.9	13.5	7.1	1.1	3.2
東京	浅草	1.0	1.9	0.2	3.7	9.2	0.1	47.4	10.0	0.0	3.5	0.3	1.3	0.4	13.1	7.7
神奈川	横須賀	12.4	5.3	5.4	20.5	3.4	19.2	6.9	0.3	2.0	7.9	9.1	2.7	3.5	0.0	1.5
長野	松本	33.9	7.0	0.0	10.0	9.0	0.0	11.7	5.3	6.0	2.0	2.2	2.6	0.4	2.6	7.3
石川	金沢	8.2	1.4	0.1	4.9	13.2	0.1	18.3	8.0	1.4	3.8	17.1	12.1	2.7	5.7	3.0
岐阜	飛騨	0.0	0.0	3.4	4.6	0.0	0.0	31.3	0.0	0.0	0.0	0.0	57.2	3.5	0.0	0.0
静岡	伊東温泉	2.4	10.9	1.5	2.9	7.5	24.5	0.1	0.0	14.3	0.0	1.8	28.4	3.4	2.4	0.0
滋賀	彦根	38.9	10.1	0.8	7.7	1.5	0.0	8.5	17.4	2.2	4.0	3.4	3.4	0.0	2.1	0.0
京都	宇治	7.6	0.1	0.4	4.8	8.6	0.7	4.7	37.2	0.6	2.1	6.9	7.8	0.1	10.0	8.5
大阪	難波・天王寺	1.3	9.9	6.3	3.0	4.3	0.1	37.0	5.4	0.3	6.7	3.3	1.8	0.5	12.4	7.7
兵庫	明石	17.4	17.2	0.0	41.5	3.8	6.6	1.7	0.9	0.0	2.6	0.7	1.0	0.0	6.6	0.0
奈良	飛鳥	8.7	11.9	0.3	11.3	13.5	1.0	1.0	22.9	0.5	0.5	1.5	18.9	7.5	0.0	0.5
和歌山	白浜	0.1	5.1	35.5	0.6	0.5	36.5	0.9	0.0	0.0	7.3	0.2	7.8	4.0	0.7	0.8
鳥取	境港	7.6	2.9	0.0	0.0	1.6	0.3	32.8	0.0	2.8	30.2	0.4	0.0	11.8	3.5	6.2
岡山	倉敷	7.9	2.4	0.5	0.3	17.4	0.0	41.3	0.2	0.0	0.5	8.3	16.6	1.0	2.1	1.6
広島	広島駅周辺	0.0	0.0	0.7	8.1	3.7	0.0	0.0	0.0	0.0	0.0	76.6	6.0	0.0	4.8	0.0
愛媛	松山	33.9	22.4	0.3	4.0	9.4	0.2	1.5	3.0	17.7	0.0	3.1	0.4	2.9	0.0	1.2
福岡	中洲	8.7	10.5	1.4	10.7	9.2	0.8	4.8	1.3	0.7	0.4	2.6	0.0	2.2	6.5	40.2
長崎	佐世保	22.4	29.0	0.0	2.7	0.0	20.2	0.0	0.0	15.1	0.8	1.4	1.1	3.7	3.7	0.0
熊本	阿蘇	0.2	8.7	4.5	1.9	1.1	0.5	0.9	0.0	37.7	4.5	0.2	26.8	7.4	0.4	5.2
大分	別府温泉	0.0	6.9	9.4	0.4	5.0	2.1	2.7	0.0	6.5	9.4	3.3	6.0	46.9	1.2	0.2
沖縄	宮古島	1.3	7.3	1.1	0.8	1.5	73.1	1.2	0.0	0.9	0.8	1.5	3.1	2.6	4.0	0.8
....															



【類似地域の発見】

札幌市南西部のように展望景色や家族で楽しめる施設がある地域は？

札幌市(南西部)とテーマ性が類似する地域



札幌市南西部

もいわ山展望台



円山動物園



白い恋人パーク



和歌山市

和歌山城からの景色



お城の動物園



和歌山マリーナシティ



徳島市

眉山(びざん)



とくしま動物園



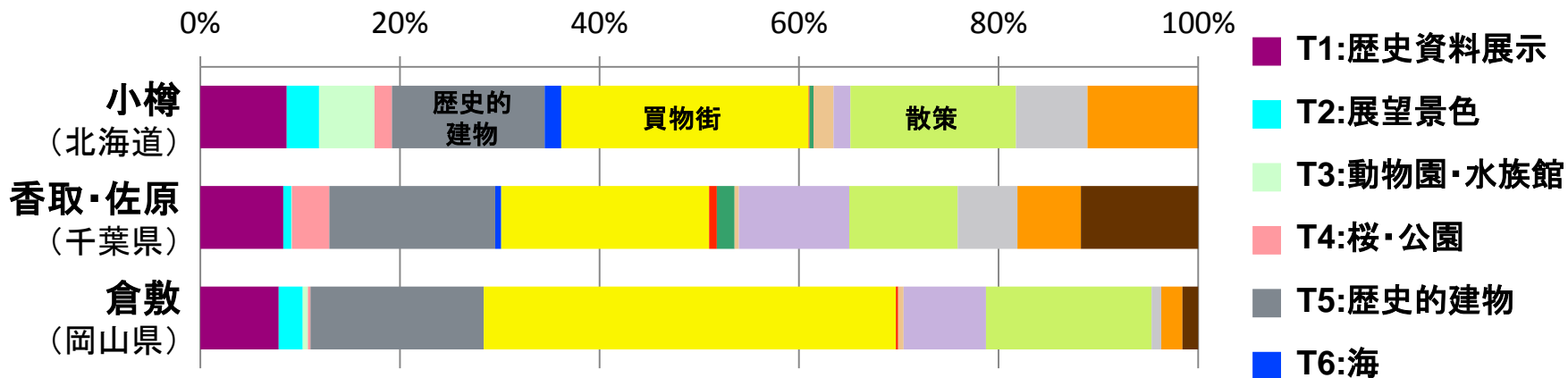
阿波おどり会館



【類似地域の発見】

小樽のように歴史的な建物があり、散策して買物が楽しめる地域は？

小樽とテーマ性が類似する地域



小樽

日本銀行
旧小樽支店



小樽運河



旧手宮線
跡



香取・佐原

佐原三菱
館



小野川沿
いの歴史的
的町並み



樋橋
(とよはし)



倉敷

中国銀行
倉敷本町
出張所



倉敷川



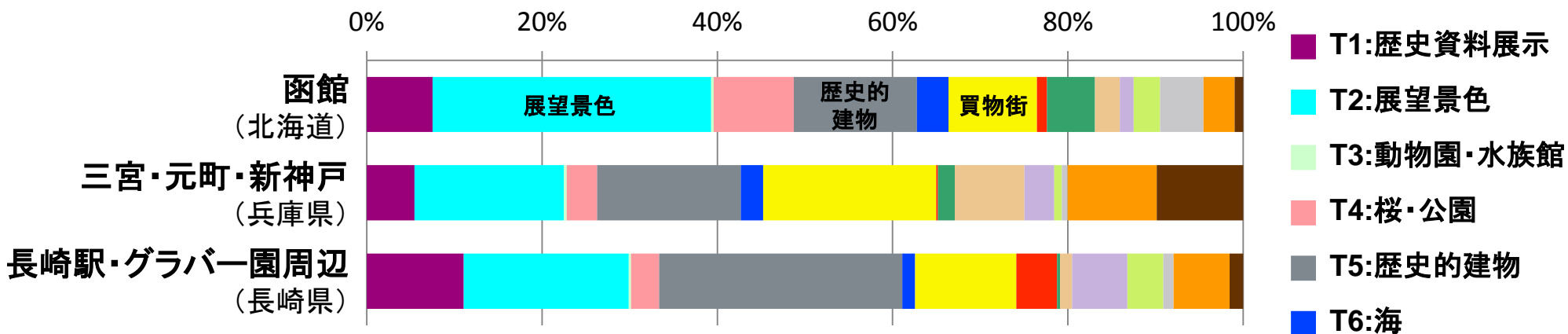
倉敷美観
地区



【類似地域の発見】

函館のように展望景色や歴史的な建物、買物が楽しめる地域は？

函館とテーマ性が類似する地域



函館

函館山



聖ヨハネ教会



函館朝市



三宮・本町・新神戸

神戸ポートタワー



風見鶏の館



南京町



長崎駅・グラバー園周辺

竜馬のぶつ像



大浦天主堂



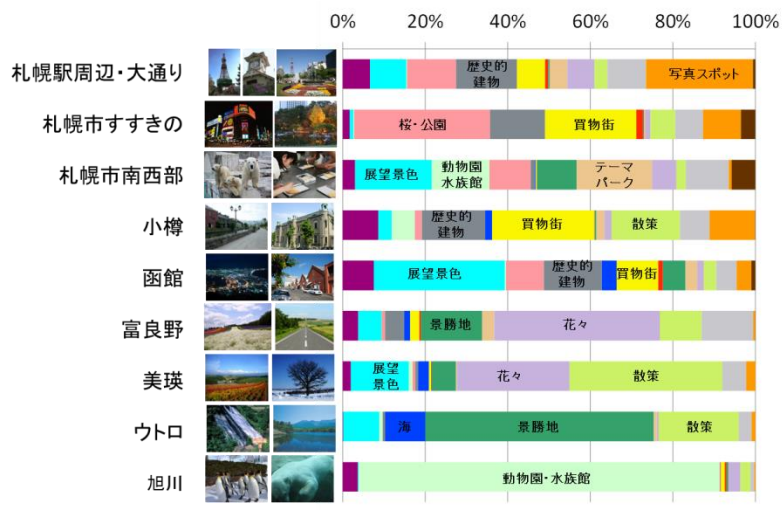
長崎新地中華街



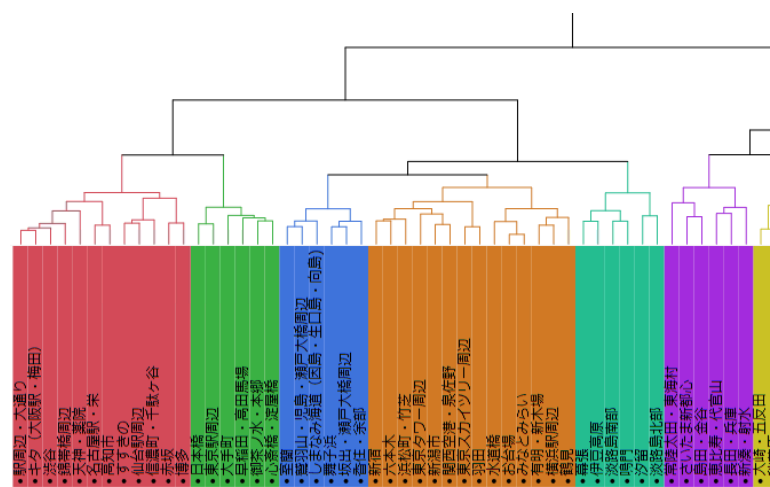
【分析結果の活用】

地域のテーマ性を把握することで競争戦略・連携戦略を検討できます

地域のテーマ性の分析の活用



類似テーマ地域の分析の活用



周辺地域の中で自分の地域の特徴的なテーマとは

- 差別化してアピールすべきテーマを把握できる
- 周辺地域と同じテーマが特徴となる場合は、連携することで効率的なプロモーション活動に取り組むことができる
- 周辺地域の特徴テーマを組み合わせることで広域での観光ルートを設計できる

自分の地域とテーマ性が類似している地域はどこか

- テーマ性が類似する地域と連携して、その場所で互いのプロモーションを行うことで、離れていても同様の価値を提供できる地域として相互送客の効果が期待できる
- 観光地として成功している類似地域を把握すれば、自分の地域にも該当する成功パターンやマーケティング活動を勉強できる

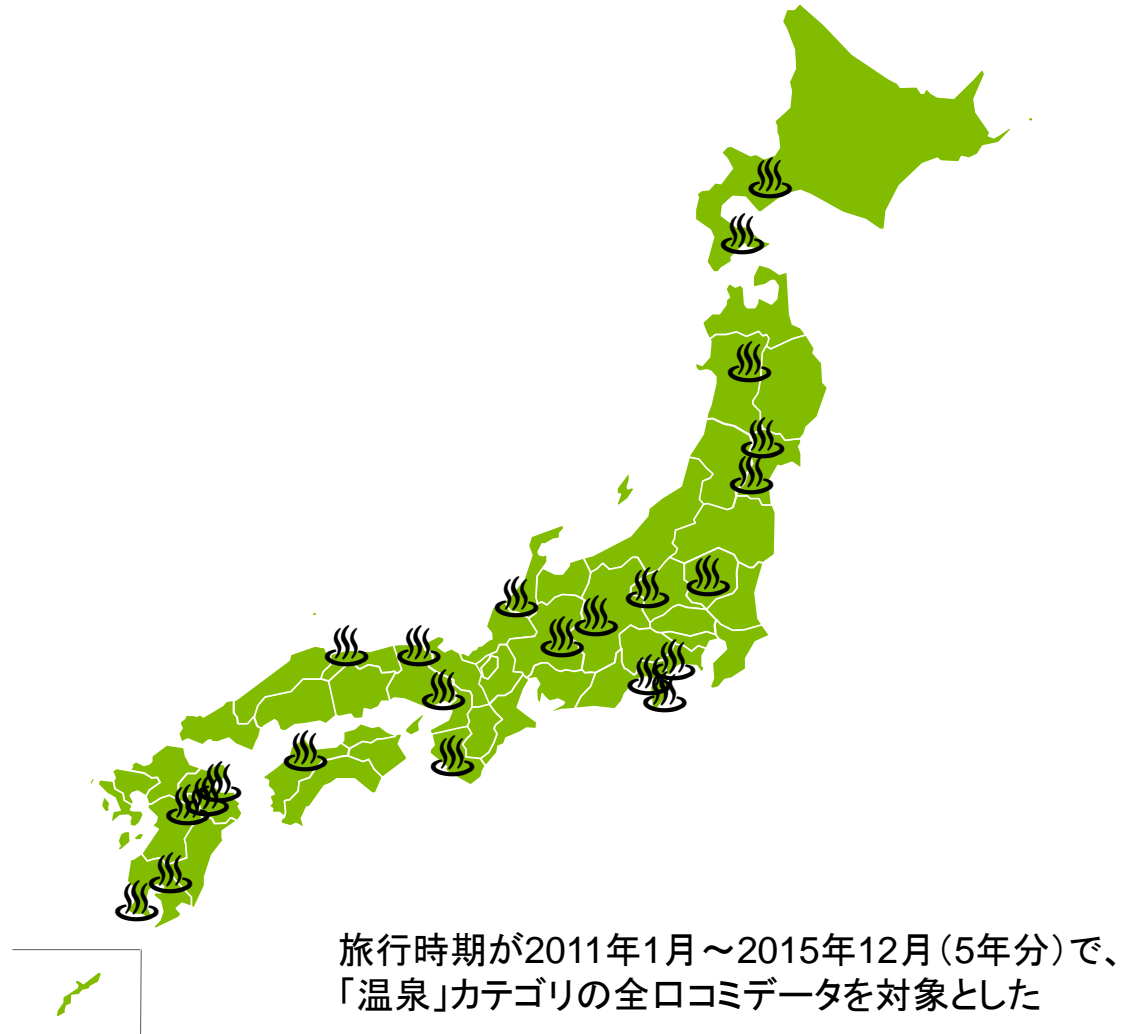
＜事例②＞

温泉地におけるマーケティング分析

【対象データ】

全国2,562スポットの温泉の口コミデータ12,564件を分析します

分析の対象



【ロコミトピックの抽出方法】 テキストマイニングとPLSAを応用してロコミのトピックを抽出します

ロコミのトピック抽出のアプローチ

テキストマイニングの実行



係り受け分析

(名詞⇄動詞・形容詞)

係り受け表現	頻度
泉質⇒良い	318
良い⇒温泉	279
雰囲気⇒良い	263
人⇒多い	221
湯⇒浸かる	159
硫黄⇒匂う	137
景色⇒良い	126
駐車場⇒広い	122

共起行列の作成

抽出した係り受け表現とそれを構成する単語に基づいて、「**単語×係り受け表現**」の共起行列を作成する
(文章単位の共起頻度を集計)

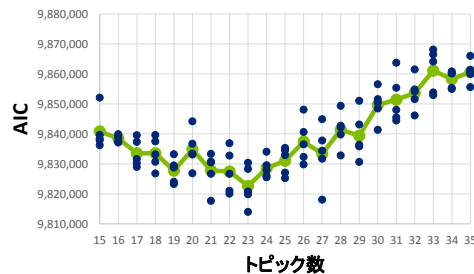
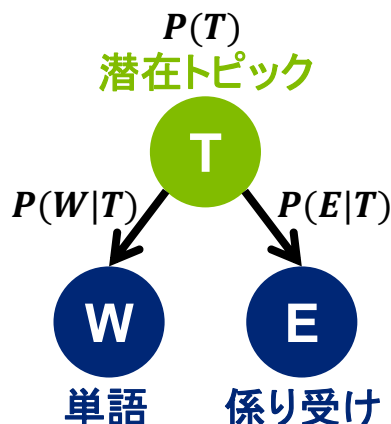
		係り受け表現				
		泉質↓良い	人↓多い	硫黄↓匂う	景色↓良い	⋮
単語	泉質	337	4	12	1	
	良い	337	13	16	135	
	多い	15	235	1	1	
	景色	3	3	1	135	
	⋮					

単語: 1,648語
係り受け: 4,808表現

※係り受けは頻度5件以上を対象

PLSAの実行

共起行列にPLSAを適用する



情報量基準AICを用いて
最適なトピック数を選択する

トピックの抽出

各トピックについて以下の3つの確率が計算される

- ① $P(T)$
潜在トピックの割合
- ② $P(W|T)$
各トピックと単語との関係の強さ
- ③ $P(E|T)$
各トピックと係り受けの関係の強さ

$P(T_4) = 5.1\%$			
$P(W T)$	単語	$P(E T)$	係り受け
12.5%	雰囲気	3.0%	雰囲気⇒良い
6.8%	建物	1.1%	良い⇒雰囲気
6.5%	温泉地	1.0%	落ち着く⇒雰囲気
4.1%	きれい	1.0%	レトロ⇒雰囲気
3.0%	古い	0.9%	雰囲気⇒味わう+できる
2.5%	共同浴場	0.8%	清掃⇒行き届く
1.7%	新しい	0.8%	建物⇒古い
1.7%	風情	0.7%	温泉地⇒ある
1.4%	銭湯	0.7%	古い⇒建物
1.2%	旅館	0.7%	野趣⇒あふれる

$P(W|T)$ と $P(E|T)$ という2つの軸の要素の重みからトピックの意味を解釈する

【ロコミトピックの抽出】

12,564件の温泉のロコミは23個のトピックに集約されました

1 露天風呂からの眺め

景色・素晴らしい 景色・最高 温泉・湧かる 露天風呂・ある 富士山・眺める
海・眺める 露天風呂・良い 富士山・眺める
 景色・見る 景色・良い 景色・見+できる 露天風呂・最高
気持ち・良い 景色・眺める **露天風呂・眺める**
 高台・ある **露天風呂・見る+できる** 露天風呂・入る
 露天風呂・気持ちいい 富士山・見る+できる 海・見る
 夕方 **天気・良い** 広い 露天風呂・景色・見+できる
 景色・楽しむ+できる 山々・見る+できる 露天風呂
 目の前・広がる 朝・霧・かぶる

2 湯船の特徴、源泉湯

湯船・湧かる 天井・高い
 源泉・かけ流す ジェット・並ぶ ぬるい・温泉 湯船・利用
 源泉湯・並ぶ ロッカー・並ぶ
 源泉湯・掛け流し 流す 水風呂・往復 温泉・利用
露天・出る 浴室・入る 広い・湯船
 源泉・利用 入る 湯船 冷たい 水風呂 一段 高い
 露天・ある 広い 岩風呂 湯船 並ぶ スーパー銭湯 作る
 奥・ある 明るい 雰囲気 湯船 入る ぬるい 湯
 湯船・ある 内湯・ある

3 入浴受付の説明

リストバンド 渡す フロント 渡す 交換 渡す
 靴 入る 脱衣所 向かう
 入口 入る フロント 見せる **ロッカー・並ぶ**
券売機・買う チケット・買う 先・ある
鍵・渡す 入口・入る 入る 券売機
 源泉・利用 入る 湯船 浴室 向かう
 1階・見せる
受け付け 見せる 見せる・無料
 中 入る 入る 靴箱 料金 払う 受け付け 渡す
 脱衣所 入る 入る フロント エレベーター 上る 靴 並ぶ

4 雰囲気の良さ

新しい建物 中心 有る
 建物 新し い 建物 きれい 鄙びる 雰囲気
良い 建物 雰囲気 雰囲気 湧き 湧き 湧き
 野趣 有る **雰囲気 良い** 落ち着いた 雰囲気
 清掃 行ける 廊下 温泉地 有る レトロ 雰囲気
 建物 古い 温泉文化 指定 並立 建物 古い 建物
 静かな 温泉地 音 程 が良い

5 訪問した時間帯

夜 行く 湯 行く
 温泉 入る 平日 行く 時間 行く 行く 価値
 一度 行く したい 車 行く 行く おすすめ **朝 早い**
 温泉 入る 入る 見つけ 通観 一度 行く してほしい
 早い 時間 温泉 行く 今回 行く **通観 掲載**
 一度 行く 昼間 行く 夕方 行く 行く 温泉 夜 遅い

6 アクセス

駅 離れる 駐車場 有る バス 乗る
 距離 有る **車 行く 駅 歩く** 駐車場 止める アクセス 便利
駅 近い バス 行く 送迎 バス 乗る 近い 温泉
 駐車場 停める **場所 有る** 駅 出る **アクセス 良い**
 シャトル バス 出る 車 有る バス 出る 駐車場 広い
 徒歩 行く できる 車 有る バス 出る 車 行く できる
 駅 有る 本数 少ない 10分 有る 車 停める 6分 有る
 バス 有る 時間 かかる 車 止める

7 良い温泉、疲れの癒し

良い 雰囲気
 良い 癒される +できる 体 癒す 温泉 有る
 癒れる 取る +できる 風呂 有る **雰囲気 良い**
良い 湯 アクセス 良い **良い 温泉**
 景色 有る **泉質 良い** 湯 有る
疲れる 癒す 肌 有る 気持ち 有る
 良い 泉質 温泉 有る
 気持ち 癒される 露天風呂 有る 利用 有る 肌 癒す

8 設備の広さ、充実性

風呂 充実
 気持ち 有る 風呂 有る 広い
 露天風呂 有る 風呂 有る 普通 銭湯
 露天風呂 有る 風呂 有る 休憩所 有る
 駐車場 有る 広い サウナ 有る **露天風呂 有る**
 風呂 有る 風呂 有る 楽しむ +できる 風呂 有る
 広い 露天風呂 露天風呂 楽しむ +できる サウナ 有る
 設備 充実 種類 多い サウナ 有る
 洗い場 有る 風呂 有る 種類 豊富
 風呂 有る 施設 充実

9 安価、無料入浴

湯 入る 有る 値段 安い 料金 高い
 無料 有る 風呂 有る 有る 銭湯 有る 入る 温泉
 露天風呂 有る 風呂 有る 温泉 有る 楽しむ +できる
無料 入浴 有る 無料 有る 100円 有る
無料 利用 +できる 温泉 有る
料金 安い 気軽 利用 +できる 入浴 料金 安い
 500円 有る 24時間 営業 200円 有る
 外湯 管理 気持ち 有る 500円 有る
 新 有る

10 目的地までのルート

通 有る 有る 有る 山 有る
 住宅街 有る 住宅街 有る 急 坂 有る
 川 有る 向かう 途中 道 有る
 看板 有る 有る 有る 橋 有る 有る 有る
 場所 有る 階段 有る 有る 有る 有る 有る 有る
 沿 有る 道 有る 階段 有る 有る 階段 有る 有る
 歩 有る 有る 坂 有る 有る 5分 有る 有る
 色 有る 有る 白湯 有る 湯 有る 湯 有る
 狭い 道 20分 有る 有る 湯 有る 湯 有る 湯 有る

11 泉質・湯量

湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

12 温泉に入ること

サウナ 有る
 浴室 有る 風呂 有る 有る 湯 有る
 入る 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

13 自己・他者の利用

人 有る 有る 有る 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

14 湯の温度

湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

15 足湯

足湯 有る 足湯 有る 足湯 有る 足湯 有る
 足湯 有る 足湯 有る 足湯 有る 足湯 有る
 足湯 有る 足湯 有る 足湯 有る 足湯 有る
 足湯 有る 足湯 有る 足湯 有る 足湯 有る
 足湯 有る 足湯 有る 足湯 有る 足湯 有る
 足湯 有る 足湯 有る 足湯 有る 足湯 有る

16 宿泊でも日帰りでも楽しめる

日 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

17 体を温めること

湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

18 砂湯、有名な温泉

湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

19 充実した食事・休憩所

一日 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

20 綺麗な施設、家族で楽しめる

施設 有る 施設 有る 施設 有る 施設 有る
 施設 有る 施設 有る 施設 有る 施設 有る
 施設 有る 施設 有る 施設 有る 施設 有る
 施設 有る 施設 有る 施設 有る 施設 有る
 施設 有る 施設 有る 施設 有る 施設 有る
 施設 有る 施設 有る 施設 有る 施設 有る

21 人の多さ

湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

22 アメニティの有無

タオル 有る タオル 有る タオル 有る タオル 有る
 タオル 有る タオル 有る タオル 有る タオル 有る
 タオル 有る タオル 有る タオル 有る タオル 有る
 タオル 有る タオル 有る タオル 有る タオル 有る
 タオル 有る タオル 有る タオル 有る タオル 有る
 タオル 有る タオル 有る タオル 有る タオル 有る

23 立ち寄る温泉

湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る
 湯 有る 湯 有る 湯 有る 湯 有る

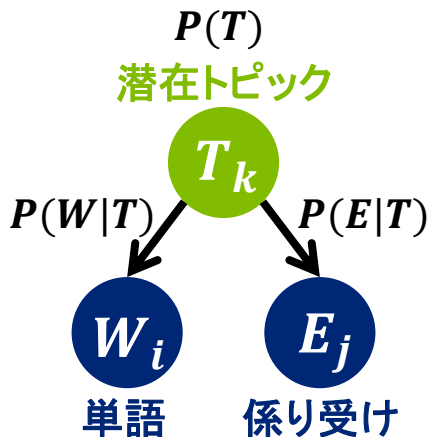
※P(E|T) > 0.05
 の係り受け表現を
 表示している

※係り受けの文字の大きさはそのトピックに対する関係の強さを表現している（緑色は上位5つの係り受け表現）

【ロコミトピックの変数化】

ロコミごとに23個のトピックのスコアを計算して変数化します

各ロコミデータのトピックスコアの算出



文章単位に各トピックのスコア(関連度)を計算し、ロコミ単位にそれを集約する

文章 S_h のトピック T_k のスコア
(事後確率 ÷ 事前確率)

$$\frac{P(S_h|T_k)}{P(S_h)}$$

$S_h = \{E_1, E_2, \dots, E_J\}$	文章 S_h を構成する係り受け表現 E_j
$P(S_h T_k) = \sum_j P(S_h E_j)P(E_j T_k)$	トピック T_k を条件とした文章 S_h の出現確率 (トピック T_k と文章 S_h の関係の強さ)
$P(S_h E_j) = \frac{1}{n(E_j)}$	係り受け E_j が出現する中で文章 S_h が出現する確率 (E_j の出現文章数の逆数)
$P(S_h) = \sum_k P(S_h T_k)P(T_k)$	文章 S_h の出現確率

トピックスコア算出プロセス

①文章ごとにスコアを計算

ロコミID	文章ID	T01	T02	T03	...	T23
1	1	3.1	0.9	2.0		1.1
1	2	1.4	0.2	5.5		2.4
2	1	0.8	5.8	1.3		0.9
2	2	1.2	3.2	1.7		1.0
2	3	0.6	1.8	2.6		1.6
...						

※文章数: 99,938件

②ロコミごとに文章スコアを集約(最大値を採用)

ロコミID	T01	T02	T03	...	T23
1	3.1	0.9	5.5		2.4
2	1.2	5.8	2.6		1.6
...					

※ロコミ数: 12,564件

③スコア3以上にフラグを立てる

ロコミID	T01	T02	T03	...	T23
1	1	0	1		0
2	0	1	0		0
...					

※実際のコメントを確認して閾値を設定

【地域の特徴分析】

各温泉地はどのトピックが多く話題にされているか定量的に把握できます

温泉地別のトピック該当割合

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

全体 (12,564)	トピック	別府 (599)	道後 (449)	城崎 (233)	指宿 (198)	草津 (158)	熱海 (154)	有馬 (124)	山梨 (119)	箱根 (111)	諏訪 (110)	黒川 (109)	由布院 (104)
23%	T01 露天風呂からの眺め	14%	8%	16%	30%	17%	14%	2%	71%	37%	13%	19%	19%
16%	T02 湯船の特徴	14%	12%	17%	5%	32%	5%	13%	18%	19%	28%	8%	12%
10%	T03 入浴受付の説明	12%	12%	11%	11%	8%	5%	8%	16%	11%	11%	6%	12%
26%	T04 雰囲気の良い	31%	42%	48%	14%	32%	26%	23%	17%	39%	31%	53%	37%
26%	T05 訪問した時間帯	21%	42%	23%	30%	25%	16%	29%	50%	38%	15%	33%	9%
25%	T06 アクセス	21%	12%	24%	23%	21%	30%	16%	20%	38%	31%	15%	21%
28%	T07 良い温泉、疲れの癒し	19%	20%	20%	24%	25%	19%	20%	38%	43%	31%	15%	21%
23%	T08 設備の広さ・充実性	14%	18%	26%	15%	19%	6%	22%	20%	41%	31%	15%	21%
21%	T09 安価、無料入浴	20%	22%	20%	9%	27%	13%	23%	18%	15%	31%	15%	21%
18%	T10 目的地までのルート	18%	12%	9%	15%	16%	28%	10%	20%	25%	12%	20%	23%
23%	T11 泉質・湯量	24%	10%	15%	10%	34%	22%	39%	14%	19%	36%	21%	19%
19%	T12 温泉に入ること	22%	32%	21%	22%	28%	5%	25%	29%	29%	29%	21%	21%
23%	T13 自己・他社の利用	23%	31%	14%	18%	23%	17%	35%	21%	21%	21%	18%	18%
21%	T14 湯の温度	20%	20%	20%	17%	37%	19%	23%	30%	30%	30%	23%	23%
23%	T15 足湯	20%	20%	20%	18%	37%	34%	27%	13%	20%	40%	31%	27%
22%	T16 宿泊でも日帰りでも楽しめる	20%	20%	20%	14%	20%	12%	39%	13%	31%	15%	40%	25%
19%	T17 体を温めること	18%	14%	13%	43%	18%	8%	19%	33%	18%	18%	12%	12%
22%	T18 砂湯、有名な温泉	41%	41%	20%	73%	22%	22%	22%	27%	26%	26%	19%	19%
19%	T19 充実した食事・休憩所	18%	15%	6%	13%	8%	7%	21%	19%	40%	13%	13%	13%
22%	T20 綺麗な施設、家族で楽しめる	15%	11%	20%	9%	29%	7%	24%	18%	36%	23%	23%	13%
18%	T21 人の多さ	20%	26%	20%	16%	20%	16%	29%	22%	14%	18%	24%	26%
17%	T22 アメニティの有無	19%	31%	12%	44%	17%	7%	11%	14%	19%	16%	10%	20%
19%	T23 立ち寄る温泉	16%	11%	15%	27%	15%	12%	15%	18%	30%	18%	14%	14%

雰囲気の良い

 雰囲気の良い

宿泊でも日帰りでも楽しめる

砂湯、有名な温泉

人の多さ

【観光客の価値観分析】

各観光客層はどのトピックに関心が強いのか定量的に把握できます

観光客属性別のトピック該当割合

全体の1.1倍以上
 全体の1.2倍以上

全体 (12,564)	トピック	女性					男性					同行者					
		20代 (577)	30代 (1,733)	40代 (1,554)	50代 (762)	60代 (192)	20代 (337)	30代 (1,623)	40代 (2,509)	50代 (2,172)	60代 (903)	乳幼児 連れ 家族 旅行 (349)	家族 旅行 (1,817)	カッ プル ・ 夫 婦 (2,436)	カッ プル ・ 夫 婦 (シ ニア) (543)	友人 (1,408)	一人 旅 (3,548)
23%	T01 露天風呂からの眺め	24%	25%	26%	25%	28%	20%	23%	20%	20%	24%	26%	25%	24%	28%	27%	20%
16%	T02 湯船の特徴	12%	14%	17%	13%	18%	9%	18%	20%	15%	17%	14%	13%	17%	16%	14%	19%
10%	T03 入浴受付の説明	9%	10%	10%	11%	6%	7%	11%	13%	8%	10%	9%	8%	10%	12%	11%	11%
26%	雰囲気良さ	27%	27%	26%	26%	31%	24%	27%	24%	24%	25%	30%	27%	28%	33%	25%	25%
26%	T05 訪問した時間帯	32%	30%	28%	29%	21%	31%	24%	24%	24%	18%	21%	27%	27%	23%	29%	23%
25%	T06 アクセス	23%	25%	21%			28%	30%	26%							26%	30%
28%	T07 良い温泉、疲れの癒し	26%	32%	31%			28%	28%	26%							28%	25%
23%	T08 設備の広さ・充実性	27%	29%	23%			22%	23%	20%							26%	21%
21%	T09 安価、無料入浴	20%	22%	22%	20%	19%	18%	21%	23%	17%	23%	23%	18%	20%	27%	21%	22%
18%	T10 目的地までのルート	14%	14%	17%	18%	18%	16%	20%	20%	19%	21%	13%	17%	17%	18%	16%	21%
23%	T11 泉質・湯量	14%	21%	23%	21%	31%	14%	24%	25%	25%	30%	20%	22%	26%	29%	21%	25%
19%	T12 温泉に入ること	18%	20%	21%			14%	19%	20%						21%	18%	18%
23%	T13 自己・他社の利用	26%	27%	24%			21%	23%	21%						23%	23%	23%
21%	T14 湯の温度	24%	25%	23%			17%	22%	21%					23%	23%	21%	21%
23%	T15 足湯	21%	24%	24%	25%	28%	19%	27%	23%	23%	19%	22%	26%	24%	24%	23%	24%
22%	宿泊でも日帰りでも楽しめる	24%	24%	22%	21%	29%	17%	23%	20%	20%	25%	26%	25%	24%	32%	22%	19%
10%	T17 休を温めること	22%	22%	21%	22%	18%	15%	18%	20%	17%	16%	20%	20%	21%	21%	21%	18%
22%	砂湯、有名な温泉	27%	25%	25%	24%	24%	23%	21%	21%	19%	18%	23%	24%	23%	27%	23%	20%
19%	T19 充実した食事・休憩所	23%	23%	22%	21%	13%	15%	18%	19%	16%	18%	28%	19%	21%	20%	21%	18%
22%	T20 綺麗な施設、家族で楽しめる	26%			23%	13%	18%	21%	23%	19%							21%
18%	T21 人の多さ	20%			19%	20%	17%	19%	15%	20%							18%
17%	T22 アメニティの有無	21%			20%	13%	14%	13%	16%	13%							15%
19%	T23 立ち寄る温泉	22%	21%	18%	20%	17%	15%	19%	19%	18%	18%	18%	20%	19%	19%	22%	18%

60代女性

乳幼児連れ
家族旅行

シニア夫婦

60代女性

60代男性

家族旅行
(乳幼児連れ含む)

シニア夫婦

女性全般
(特に20代)

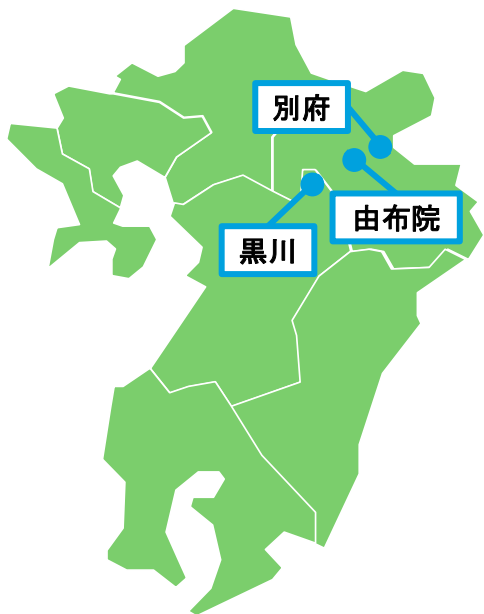
家族旅行

シニア夫婦

【マーケティングの検討】

温泉地の特徴トピックに関心を持つ観光客層をターゲットに設定できます

別府・由布院・黒川の温泉地の特徴と関心のある観光客層



別府温泉



由布院温泉



黒川温泉



温泉地の特徴

ターゲット層

- T18: 砂湯、有名な温泉

- 女性(特に20代)
- カップル・夫婦(シニア)
- 家族旅行

- T04: 雰囲気の良い
- T21: 人の多さ

- 60代女性
- カップル・夫婦(シニア)
- 乳幼児連れ家族旅行

- T04: 雰囲気の良い
- T16: 宿泊でも日帰りでも楽しめる

- 60代
- カップル・夫婦(シニア)
- 家族旅行(乳幼児連れ含む)

【満足度の要因と効果の分析】

満足度に効く要因の構造をロコミピックや属性情報からモデル化します

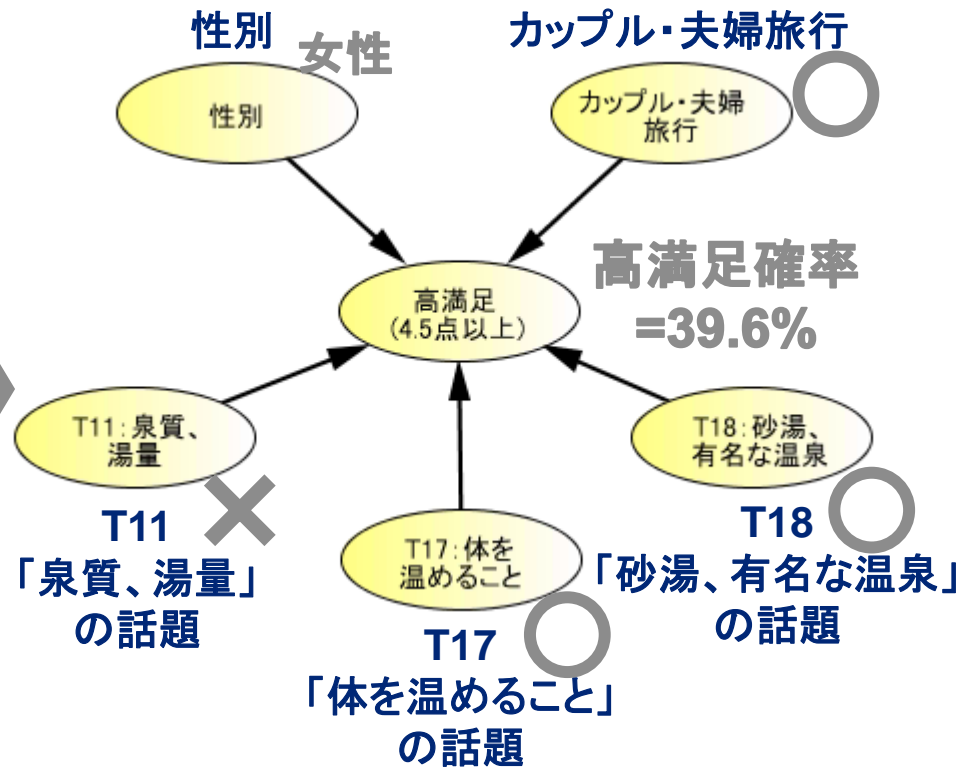
ベイジアンネットワークによる満足度モデルの構築



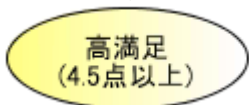
ベイジアンネットワーク

高満足には5つの要因が影響する

※ロコミの評点4.5点以上を高満足とする



これらの要因の中から高満足に影響を与える要因はどれか？



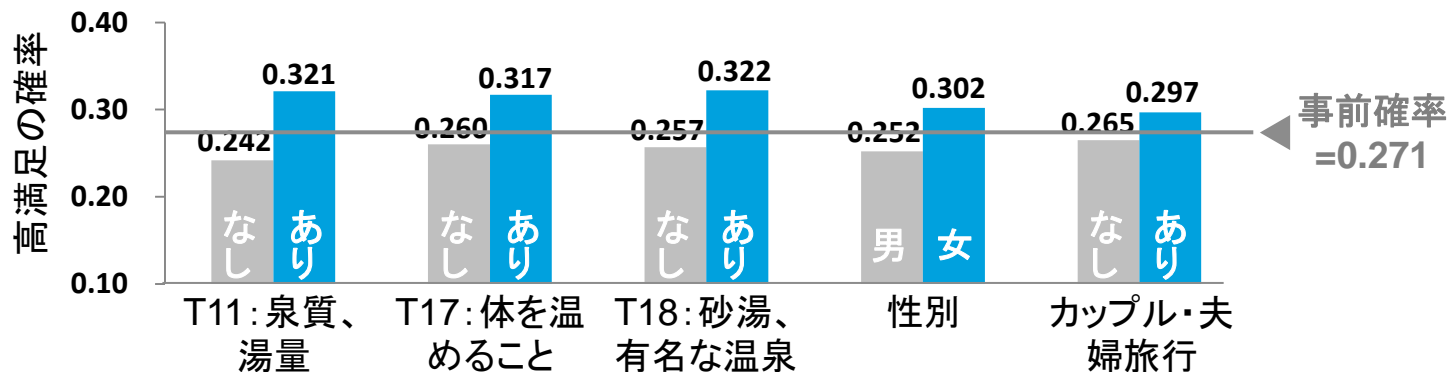
ベイジアンネットワーク

複数の変数の関係をネットワーク構造で表わし、ある変数の状態を指定したとき、その条件下での他の変数の確率を推論できる

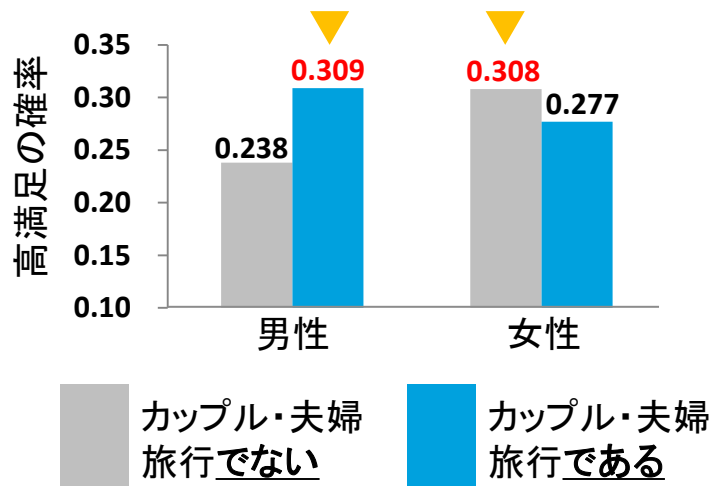
【満足度の要因と効果の分析】

観光客の満足度を効果的に高める条件を定量的に把握できます

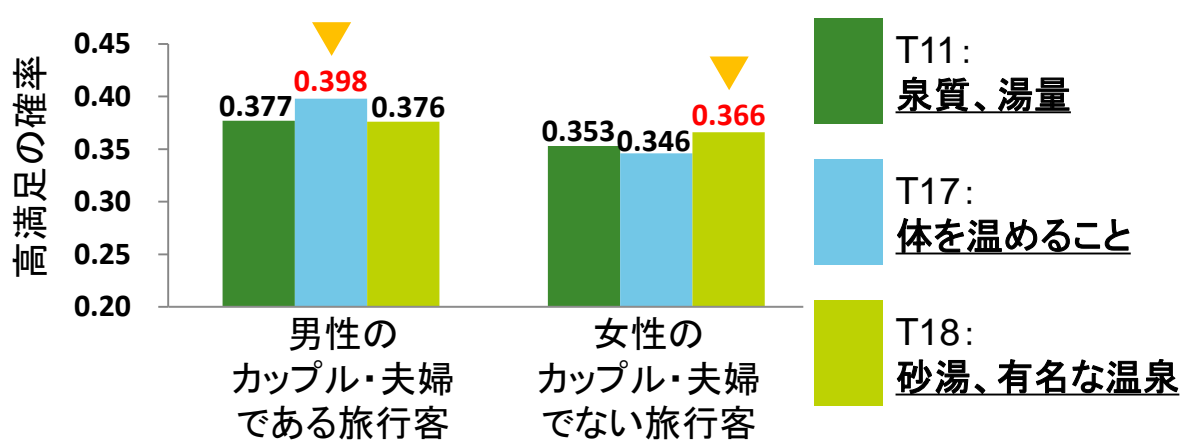
一つの変数を条件にした高満足の確率



属性を条件にした高満足の確率



性別・同行者別でトピックを条件にした高満足の確率



【プロモーションの検討】

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ます

男性向けには・・・

恋人との温泉旅行



寒いなか温泉で温まる

女性向けには・・・

女友達との温泉旅行



砂湯を楽しむ

まとめ

【新たなテキスト分析】

※特許出願中(特開2016-51220)

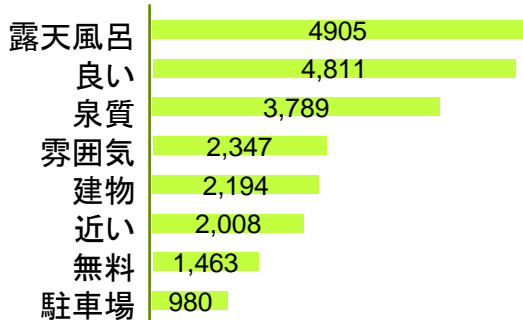
確率モデルの応用でテキストデータの活用が進化します



Text Narrative Orchestration Modeling Analytics: Text Nomolytics



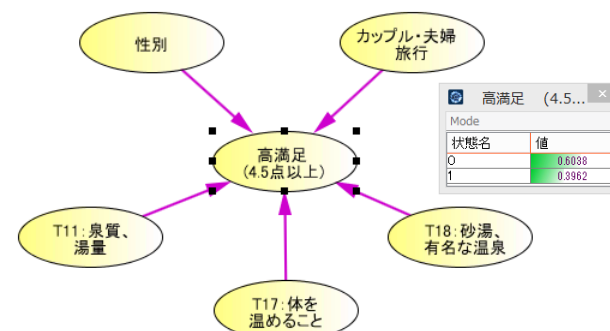
単語抽出



トピック抽出



モデリング



膨大なテキストデータを複数のトピックに変換して理解を支援する

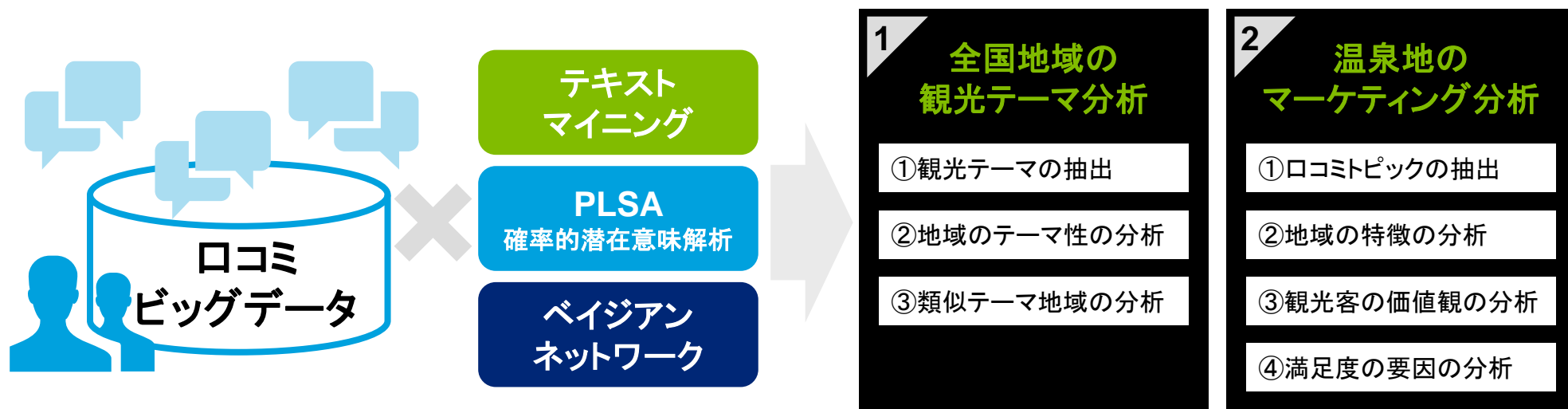
トピックを軸にした集計で属性の特徴を定量的に把握する

複雑な要因関係をモデル化し効果をシミュレーションする

【観光地の口コミへの適用】

地域観光のマーケティング戦略を観光客の目線で科学的に検討できます

観光地の口コミ分析と地域観光マーケティング



差別化してPRすべき観光テーマやトピックとは？



連携することで相互送客が見込める地域とは？



親和性の高い成功パターンを持つ類似地域とは？



特徴に関心のある効果的なターゲット層とは？



ターゲット別の魅力的なPRコンテンツとは？



【適用対象例】

様々な自由記述式データに適用することができます

SNS口コミ

- 顧客の関心トピックの抽出
- 顧客目線での競合比較分析
- 満足度向上の要因抽出
- 顧客目線のマーケティング施策検討

アンケート自由記述

- 回答者の関心トピックの抽出
- 自由記述を含む設問間の関係分析
- 話題を生む商品・サービスの特徴分析
- 評価向上の効果的な条件発見

コールセンター対応履歴

- 問い合わせトピックの抽出
- トピック関連する商品・顧客の特徴分析
- 問い合わせトピックによる解約の発生確率シミュレーションと制御

特許公報

- 技術トピックの抽出
- 特許技術の経年変化の把握
- 技術の競合比較分析
- 課題背景から開発される技術の予測

有価証券報告書

- 企業の事業トピックの抽出
- 事業状況の経年変化の把握
- 各種IR指標と事業トピックの関係分析
- 効率的な業界分析、企業診断

診療記録

- 診療記録、看護記録のトピックの抽出
- 生活習慣と病状の関係、治療内容とその効果(経過)の関係分析
- 効果的な診療支援

営業日報

- 営業活動内容のトピックの抽出
- 営業活動による契約獲得率のシミュレーションとその向上要因の抽出
- 効果的な営業教育支援

エントリーシート

- 志望動機やPR文のトピックの抽出
- 記述内容と入社後の業務成果との関係分析
- 効率的な人材発掘

問題発生レポート

- 問題発生(不具合やヒヤリハット等)のトピックの抽出
- 問題の発生確率の要因抽出
- 効果的な製品や作業環境の改善支援

Deloitte. トーマツ.

デロイトトーマツ

デロイトトーマツグループは日本におけるデロイトトウシュトーマツ リミテッド(英国の法令に基づく保証有限責任会社)のメンバーファームおよびそのグループ法人(有限責任監査法人 トーマツ、デロイトトーマツ コンサルティング合同会社、デロイトトーマツ ファイナンシャルアドバイザー合同会社、デロイトトーマツ税理士法人およびDT弁護士法人を含む)の総称です。デロイトトーマツグループは日本で最大級のビジネスプロフェッショナルグループのひとつであり、各法人がそれぞれの適用法令に従い、監査、税務、法務、コンサルティング、ファイナンシャルアドバイザー等を提供しています。また、国内約40都市に約8,700名の専門家(公認会計士、税理士、弁護士、コンサルタントなど)を擁し、多国籍企業や主要な日本企業をクライアントとしています。詳細はデロイトトーマツグループWebサイト(www.deloitte.com/jp)をご覧ください。

Deloitte(デロイト)は、監査、コンサルティング、ファイナンシャルアドバイザーサービス、リスクマネジメント、税務およびこれらに関連するサービスを、さまざまな業種にわたる上場・非上場のクライアントに提供しています。全世界150を超える国・地域のメンバーファームのネットワークを通じ、デロイトは、高度に複合化されたビジネスに取り組むクライアントに向けて、深い洞察に基づき、世界最高水準の陣容をもって高品質なサービスをFortune Global 500® の8割の企業に提供しています。“Making an impact that matters”を自らの使命とするデロイトの約225,000名の専門家については、[Facebook](#)、[LinkedIn](#)、[Twitter](#)もご覧ください。

Deloitte(デロイト)とは、英国の法令に基づく保証有限責任会社であるデロイトトウシュトーマツ リミテッド(“DTTL”)ならびにそのネットワーク組織を構成するメンバーファームおよびその関係会社のひとつまたは複数を指します。DTTLおよび各メンバーファームはそれぞれ法的に独立した別個の組織体です。DTTL(または“Deloitte Global”)はクライアントへのサービス提供を行いません。DTTLおよびそのメンバーファームについての詳細は www.deloitte.com/jp/about をご覧ください。

本資料は皆様への情報提供として一般的な情報を掲載するのみであり、その性質上、特定の個人や事業体に具体的に適用される個別の事情に対応するものではありません。また、本資料の作成または発行後に、関連する制度その他の適用の前提となる状況について、変動を生じる可能性もあります。個別の事案に適用するためには、当該時点で有効とされる内容により結論等を異にする可能性があることをご留意いただき、本資料の記載のみに依拠して意思決定・行動をされることなく、適用に関する具体的事案をもとに適切な専門家にご相談ください。