



Analytics Design Lab

NomolyticsによるVOCデータの分析

株式会社アナリティクスデザインラボ

1. Nomolyticsについて	3
2. Nomolyticsを適用したロコミ分析事例	10

1. Nomolyticsについて

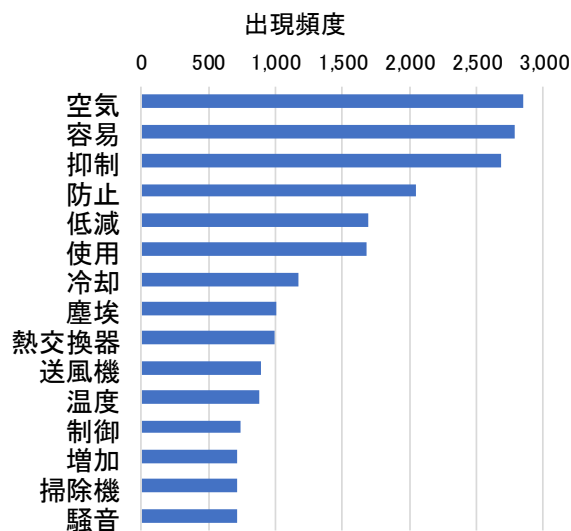
テキストマイニングでよくある可視化のアウトプット

テキストマイニングの分析では、文章に含まれる単語や係り受け表現をベースとした集計・統計分析を実行することで、文章の特徴を可視化して全体像を把握します

テキストマイニングの可視化例

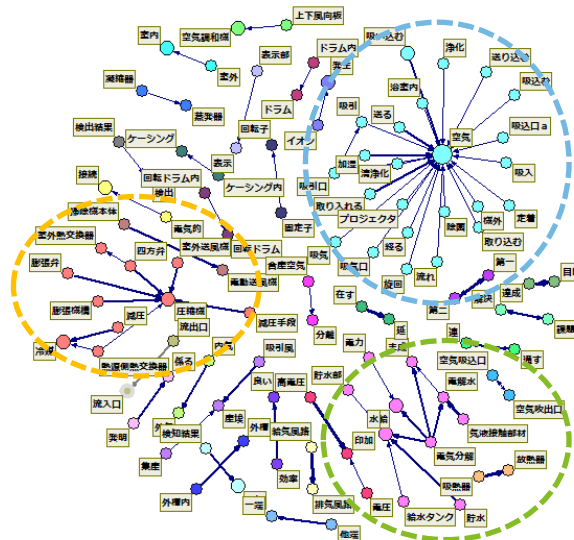
頻度集計

単語や係り受け表現の出現頻度を集計して、どのような記述が多いのか、おおまかな全体像を把握する



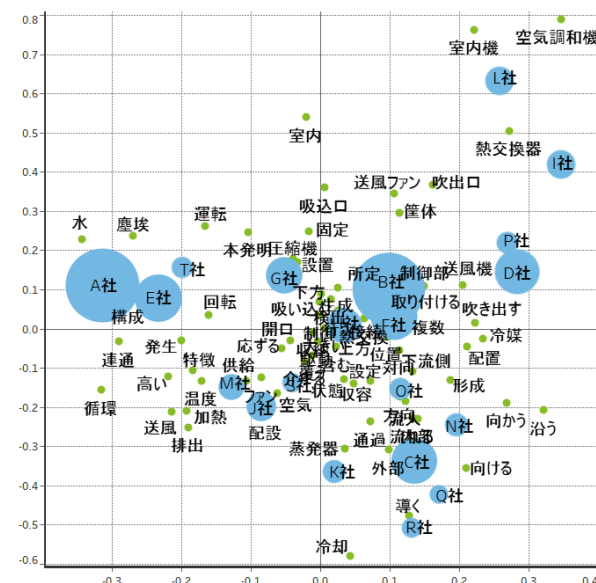
共起ネットワーク

同時に出現しやすい単語同士をネットワークでつなぎ、そのかたまりからどのような話題があるか考察する



コレスポンデンス分析

属性情報と出現単語との対応関係を同じ平面上にマッピングし、その位置関係から属性の傾向を把握する



※特許データをテキストマイニングした例を掲載

テキストマイニングの分析は単語ベースの結果で解釈が難しく、特徴がモデル化されていないので状況の変化に応じたシミュレーションができませんが、これをAI技術で解決します

課題① 解釈がしづらい

- ◆ 単語をベースにした特徴の可視化は複雑であり、結果の解釈が困難になる
- ◆ 単語を人がある程度グルーピングして分析することもあるが、その作業は主観的で負荷が大きい

解決技術

単語をクラスタリングする
AI技術

PLSA
確率的潜在意味解析

- ◆ 単語の出現状況を学習することで、使われ方の似ている単語をまとめ上げ、トピックを抽出する
- ◆ 膨大な「単語」ではなく、いくつかの「トピック」をベースに特徴を可視化し、解釈を容易にする

課題② 現状把握に留まる

- ◆ 記述傾向の現状把握はできるが、その特徴が一般化(モデル化)されていない
- ◆ 現状から状況が変化したときに、それに伴って結果がどう影響を受けるのかシミュレーションできない

解決技術

現象をモデリングする
AI技術

ベイジアンネットワーク

- ◆ PLSAで抽出されたトピックや、その他属性情報などの要因間に潜む要因関係をモデル化する
- ◆ 各要因の条件を変化させると他の要因がどのような挙動をとるのか確率的にシミュレーションする

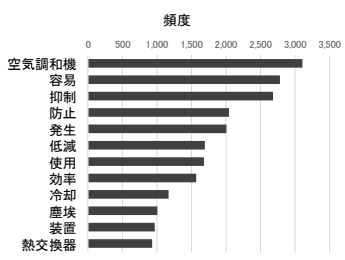
膨大なテキストデータをトピックに変換して解釈を容易にし、テキスト情報内に潜む要因関係をモデル化して、ビジネスアクションに有用な特徴を把握可能にします

Nomolytics : Narrative Orchestration Modeling Analytics

テキストマイニング

文章に含まれる単語を抽出し、その出現頻度を集計する

単語抽出



PLSA
確率的潜在意味解析

単語が出現する特徴を学習し、膨大な単語を複数のトピックにまとめる

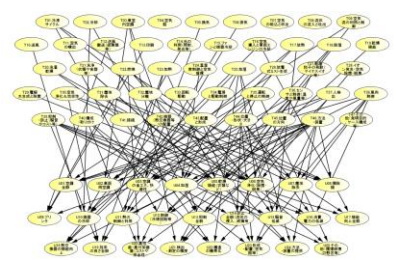
トピック化



ベイジアンネットワーク

トピックやその他属性情報など、テキスト情報内の要因関係をモデル化する

因果分析



Nomolyticsのメリット

膨大なテキストデータをいくつかのトピックという人間が理解しやすい形に整理し類型化できる

テキスト情報に潜む要因関係を構造化し、特徴を見たいターゲットのキーワードを発見できる

条件を変化させたときの効果を確率的にシミュレーションでき、有効なアクションを検討できる

Nomolyticsは様々なテキストデータに適用できますが、口コミ、アンケート、コールセンター等VOCの分析では、顧客の声をトピックに類型化し、顧客満足度の要因を発見できます



口コミ

- 顧客ターゲット別の関心トピックを把握
- 製品・サービス別のトピックを把握
- 口コミ得点に寄与するトピックを把握
- ニーズに応じたマーケティングを検討



アンケート

- 自由記述回答の内容をトピックで把握
- トピック化された自由記述回答と通常の定型設問回答の関係を統計分析
- 顧客満足度に寄与するトピックを把握



コールセンター履歴

- 問い合わせ内容をトピックで把握
- 製品別・顧客別のトピック傾向を把握
- 解約・退会に寄与するトピックを把握
- 満足度向上、顧客離反抑制の施策検討



特許文書

- 特許文書の内容をトピックで把握
- トrendや競合他社の動向を把握
- 用途と技術の関係分析から用途実現の技術戦略や保有技術の新規用途を検討



営業日報

- 営業活動内容をトピックで把握
- 営業属性別のトピック傾向を把握
- 成約に寄与するトピックを把握
- 成約のための効果的な営業教育を検討



有価証券報告書

- 企業・業界の事業内容をトピックで把握
- 事業内容トピックのトレンドを把握
- 好業績に寄与する事業トピックを把握
- 定性情報から行う企業分析・業界分析



エントリーシート

- 志望動機やPR文のトピックを把握
- 記述トピックに基づいて学生を分類
- トピック傾向から面接の質問内容を検討
- 選考通過に寄与するトピックを把握



診療・看護記録

- 診療記録、看護記録をトピックで把握
- 患者の属性別のトピック傾向を把握
- 検査指標に寄与する定性情報を把握
- 定性情報も用いた診療・助言を検討



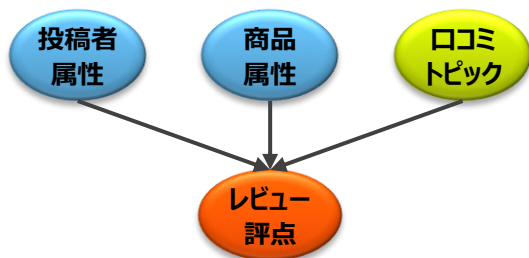
問題発生レポート

- 不具合やヒヤリハットをトピックで整理
- 作業環境別のトピック傾向を把握
- 重大問題に寄与するトピックを把握
- 問題を抑制する作業・環境改善を検討

VOCデータにNomolyticsを適用することで、VOCのトピックを軸に、顧客の行動の要因構造をベイジアンネットワークでモデル化でき、そのキードライバーを定量的に把握できます

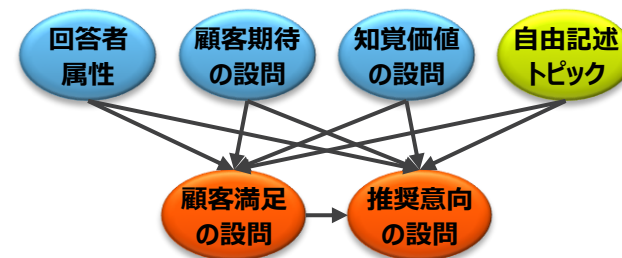
口コミ

口コミのレビュー評点の要因として、投稿者の属性や商品の属性、口コミトピックの効果把握する



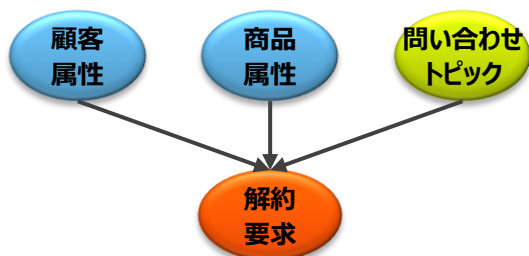
アンケート

顧客満足度の要因として、回答者の属性や選択式設問の回答結果、自由記述トピックの効果把握する



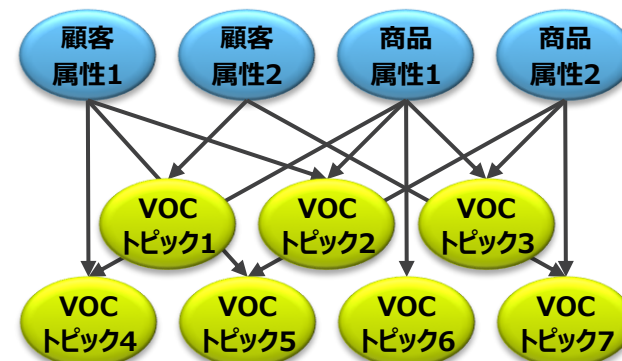
コールセンター問い合わせ履歴

解約などの要求行動の要因として、顧客の属性や商品の属性、問い合わせトピックの効果把握する



VOC共通

どのような顧客の属性や商品の属性で、どのような顧客の声が寄せられるか把握する



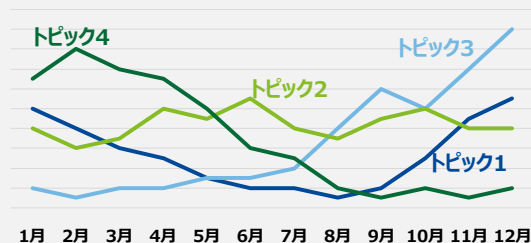
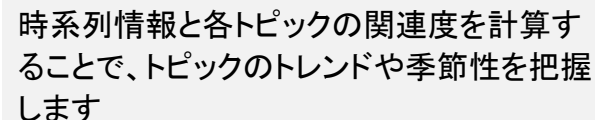
VOCのテキストデータにテキストマイニング
×PLSAを適用することで、コメントの内容を
複数のトピックに機械的に類型化し、大量の
コメントの全体像を把握します



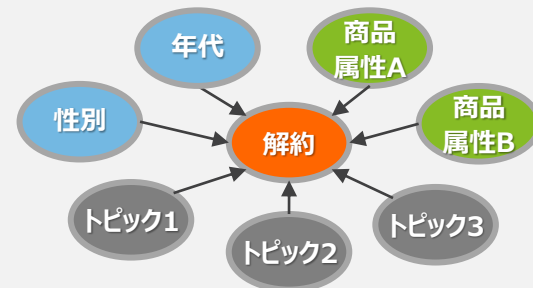
顧客属性別、商品属性別に各トピックの関連度を計算することで、各属性におけるコメントの傾向を把握します

	トピック1	トピック2	トピック3	トピック4
20代30代	0.6	1.2	4.3	0.9
40代50代	2.4	1.5	1.6	2.3
60代70代	3.8	1.1	0.4	1.2
商品属性1	0.5	3.4	1.2	2.2
商品属性2	4.1	0.8	0.3	1.0

各トピックに影響を与える要因を各属性情報（顧客属性、商品属性、時間属性など）から探索し、そのトピックが発せられやすい要因条件を把握します



データに満足度や解約などの行動情報があれば、その行動に影響を与える要因をトピックや属性から探索し、行動を確率的に予測したり、行動を制御する条件を把握します



2. Nomolyticsを適用した口コミ分析事例

旅行口コミデータの分析による地域の観光マーケティングの検討

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

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

わさんぽんさん

男性 / 函館のクチコミ: 1件

旅行時期: 2013/08(約7ヶ月前)

同行者 一人旅

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、早めの場所取りをオススメします。

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: 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

[クチコミ 41,669件](#)

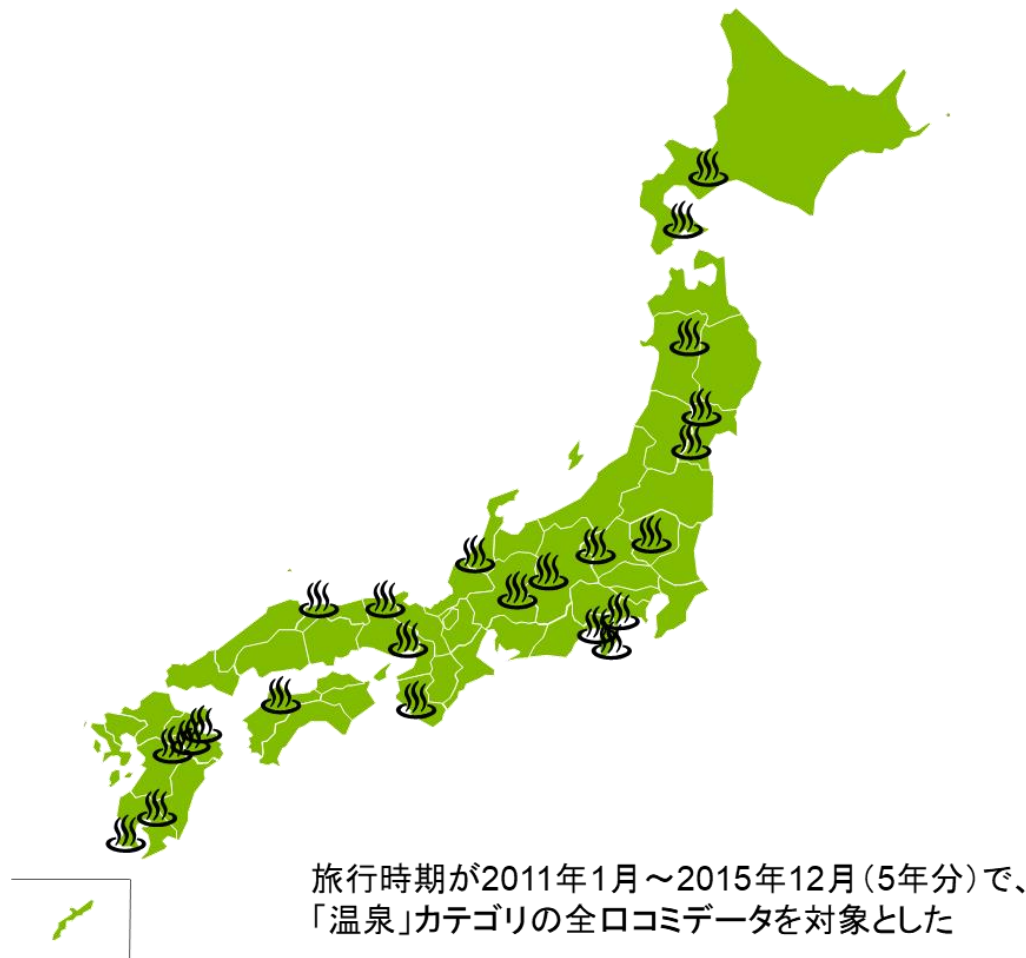
みんなの満足度: ★★★★★ 4.54

アクセス: 4.24

人混みの少なさ: 3.40

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

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



※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

口コミのトピックを抽出し、そのトピックをベースに地域や観光客の特徴、満足度の要因を分析することで、地域観光におけるマーケティング戦略を観光客目線で検討します

トピックの抽出

口コミのコメントに**テキストマイニング**を実行して単語や係り受け表現を抽出

PLSAの適用

係り受け	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
...	...

露天風呂の眺め

露天風呂の良さ

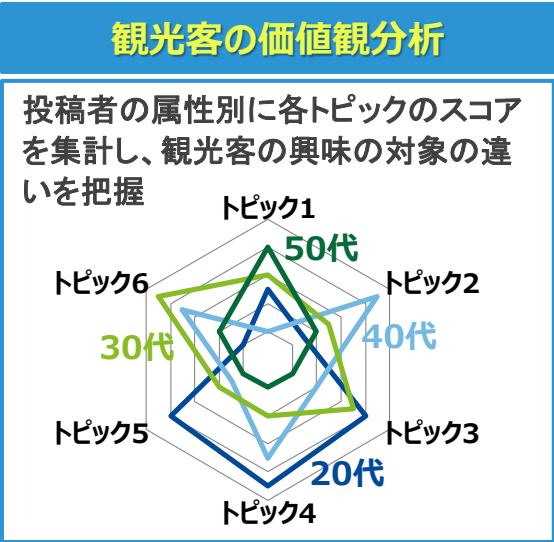
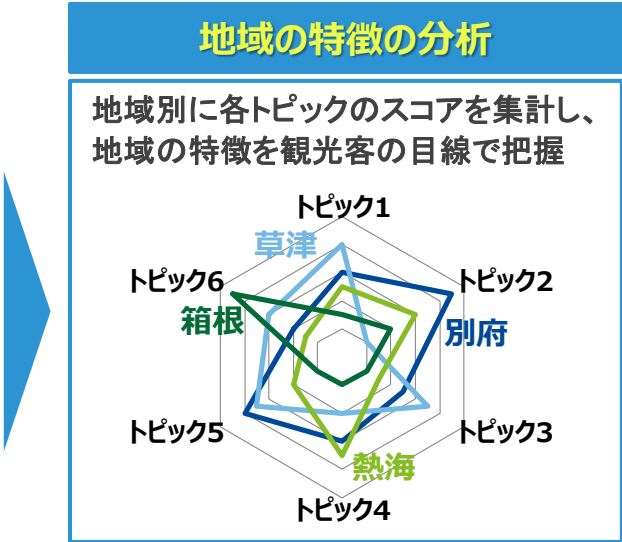
泉質・湯量

アメニティの有無

トピックのスコアリング

全口コミデータに対する各トピックの該当度を計算

口コミID	施設	地域	評点	性別	年代	同行者	トピック1	トピック2	トピック3
1	おかだ	箱根	5.0	女性	30代	友人	5.2	0.9	...
2	西の河原	草津	4.0	男性	40代	一人	3.6	4.8	...
3	吉祥の湯	別府	3.5	男性	50代	家族	1.2	2.3	...
n	...	...	...	...	...	...	...	...	...



トピック抽出のアプローチ

テキストマイニングで単語と係り受け表現を抽出し、「単語 × 係り受け」で構成される共起行列にPLSAを適用することで、単語と係り受けが集約された口コミのトピックを抽出します

テキストマイニングの実行

共起行列の作成

PLSAの実行

トピックの抽出



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

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	良い ↓ 温泉	雰囲気 ↓ 良い	人 ↓ 多い	⋮
単語					
泉質	318	166	81	22	
良い	318	279	263	36	
多い	45	32	20	221	
景色	85	98	196	14	
...					

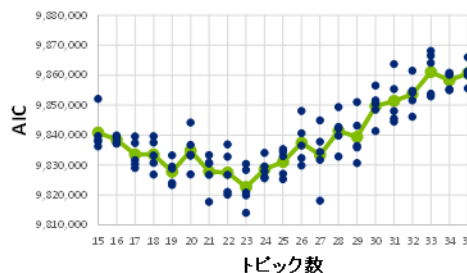
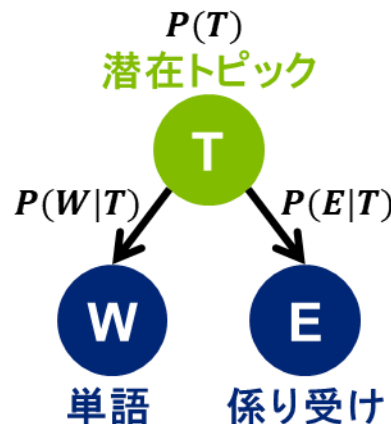
単語

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

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

各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

$P(T_4) = 5.1\%$

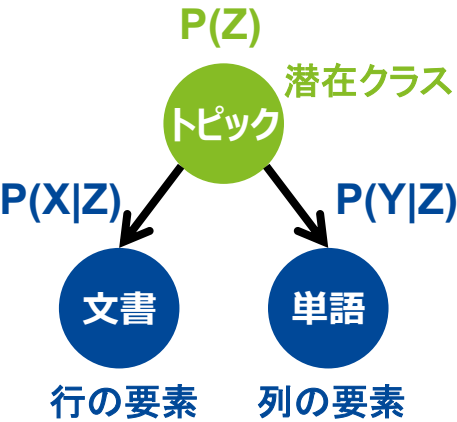
$P(W T)$	単語	$P(E T)$	係り受け
12.5%	雰囲気	3.0%	雰囲気⇒良い
6.8%	建物	1.1%	良い⇒雰囲気
6.5%	温泉地	1.0%	落ち着く⇒雰囲気
4.1%	きれい	1.0%	レトロ⇒雰囲気
3.0%	古い	0.9%	雰囲気⇒味わう+できる
2.5%	共同浴場	0.8%	清掃⇒行き届く
1.7%	新しい	0.8%	建物⇒古い
1.7%	風情	0.7%	温泉地⇒ある
1.4%	銭湯	0.7%	古い⇒建物
1.2%	旅館	0.7%	野趣⇒あふれる

$P(W|T)$ 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

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

PLSAのインプットとする共起行列の構成を「文書 × 単語」ではなく「単語 × 係り受け」とすることで、要素間の違いが出やすくなり、解釈のしやすいトピックを抽出できます

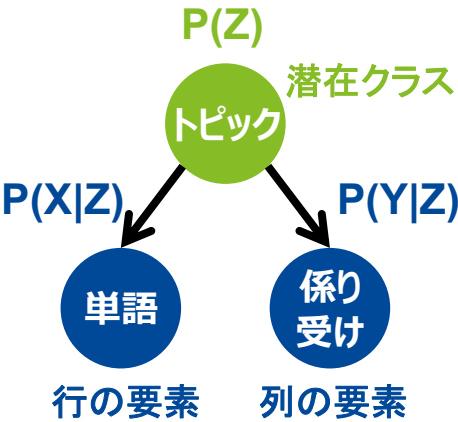
一般的なPLSAの共起行列



	単語1	単語2	単語3	単語4	...
文書ID:1	1	1	0	0	
文書ID:2	0	0	2	0	
文書ID:3	0	0	0	1	
文書ID:4	2	0	1	0	
...					

- 共起行列はBag-of-Wordsによる単語の頻度で構成され、ほとんどが“0”となる疎なデータであるため、要素間の違いが現れにくく、クリアなトピックを抽出しにくい
- PLSAのトピックには行の要素と列の要素が同時に所属し、両方の情報軸からトピックの意味を解釈できるが、一方の軸(行)は文書IDという意味性の低い情報で、トピックの解釈に使用しにくい

NomolyticsでのPLSAの共起行列



	係り受けa	係り受けb	係り受けc	係り受けd	...
単語1	325	264	11	20	
単語2	241	201	6	8	
単語3	28	41	288	14	
単語4	9	15	4	172	
...					

- 共起行列はクロス集計型の行列で、単語と係り受けの共起頻度が入った密なデータであるため、要素間での違いが現れやすく、クリアなトピックを抽出しやすい
- 行と列が単語と係り受けで構成されている共起行列では、どちらもそれぞれ単独で意味を持つ情報となるため、両方の情報軸からトピックの意味を解釈することができ、解釈の容易性が高まる

1

露天風呂
からの眺め

景色・素晴らしい 景観・最高
露天風呂・ある 温泉・遠かる
景色・見る 露天風呂・良い 富士山・眺める
景色・見る 景色・良い 景色・見る・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・ある
露天風呂・気持ち・良い 富士山・見る・できる 海・見る
見る・できる 露天風呂・見る 海・見る・できる
夕方 天気・良い 広い 露天風呂・見る・できる
景色・楽しむ・できる 山々・見る・できる 眺める 露天風呂
目の前・広がり 眺める・良

2

湯船の特徴、
源泉湯

温泉・遠かる 天井・高い ゆめい・源泉
源泉・かけ流す ジェット・並ぶ 温泉・利用
座風呂・並ぶ ロッカー・並ぶ 温泉・利用
源泉掛け流し・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・林道・作る
露天・ある 明らに 露天風呂・湯船・入る ゆめい・湯
湯船・ある 内湯・ある

3

入浴受付
の説明

リストバンド着付フロント渡す 交換・渡す
靴・入れる 脱衣所・向かう ロッカー・並ぶ
入浴・お風呂フロント見せる 温泉・向かう
券売機・買う チケット・買う 先・ある
鍵・渡す 入口・入る 浴室・向かう
鍵・ロッカー 入る 温泉・利用
受付付け・見る 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 靴箱・並ぶ

4

雰囲気
の良さ

新しい建物 中心・ある 鄙び・雰囲気
建物・新しい 建物・向かう 温泉・ある
良い 雰囲気 温泉・ある 温泉・味わい・できる 山・ある
野趣・ある 温泉・ある 温泉・ある 温泉・ある
清潔・行き 温泉・ある 温泉・ある
建物・古い 温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある 温泉・ある

5

訪問した
時間帯

夜・行く 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
一度・行く 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

6

アクセス

駅・離れた 駐車場・ある バス・乗る
距離・ある 車・行く 駅・歩く 駐車場・止める アkses・便利
駅・近い バス・行く 送迎・バス・出る 近い 温泉
駐車場・停める 場所・ある アkses・ある
シャトル・バス・出る 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある 温泉・ある

7

良い温泉、
疲れの癒し

良い 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

8

設備の広さ
・充実性

気持ち・良い 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

9

安価、
無料入浴

湯・入る 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

10

目的地
までのルート

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

11

泉質・湯量

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

12

温泉に
入ること

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

13

自己・他者
の利用

人・行く 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

14

湯の温度

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

15

足湯

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

16

宿泊でも日帰り
でも楽しめる

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

17

体を温めること

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

18

砂湯、
有名な温泉

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

19

充実した食事
・休憩所

一日・過ごす 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

20

綺麗な施設、
家族で楽しめる

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

21

人の多さ

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

22

アメニティ
の有無

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

23

立ち寄る温泉

温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある
温泉・ある 温泉・ある 温泉・ある

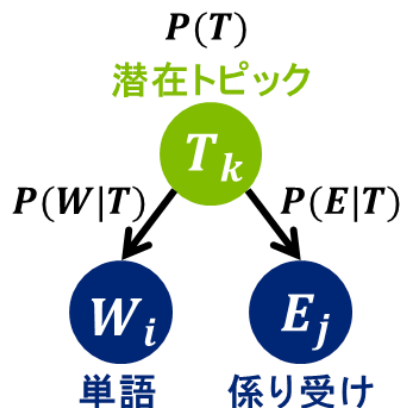
※P(E|T) > 0.05
の係り受け表現を
表示している

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

トピックのスコアリング（該当度の計算）

各口コミに対するトピックの該当度を計算し、「1:該当有」「0:該当無」のフラグを与えます

各 口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、口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

口コミ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口コミごとに文章スコアを集約(最大値を採用)

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

※口コミ数: 12,564件

③スコア3以上にフラグを立てる

口コミ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

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

トピックのフラグデータの作成

全口コミデータに対して各トピックのスコア(該当有無)を計算することで、トピックをベースとした様々な集計・分析を実行することができます

トピックのスコア（フラグ情報）を紐づけた口コミデータ

口コミID	施設名称	施設所在地域	コメント	評点	性別	年代	同行者	旅行時期	トピック1	トピック2	...	トピック23
1	湯の里おかだ	箱根	日帰り温泉で利用しました。広々とした露天風呂は絶景です。	5.0	女性	30代	友人	2012年2月	1	0		1
2	西の河原露天風呂	草津	緑に囲まれた広い湯船の露天風呂で、ゆっくりできます。	4.0	男性	40代	一人旅	2014年8月	1	1		0
3	やまなみの湯	別府	車で立ち寄りやすくて手軽です。別府だけにお湯はいいです。	3.5	男性	50代	家族	2013年6月	0	0		1
4	吉祥の湯	黒川	お風呂は露天風呂が5つ。紅葉の時期でとても綺麗でした。	4.5	女性	40代	カップル・夫婦	2012年11月	1	0		0
...												
12,564	家康の湯	熱海	駅前にある足湯ですが、こんなに広い足湯は初めてです。	3.0	女性	20代	一人旅	2015年9月	0	1		1

①地域の特徴分析

②観光客の価値観分析

③満足度の要因分析

【分析①】 地域の特徴分析

地域×トピックの集計で、各温泉地で多く話題にされるトピックを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 雰囲気の良いさ	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%			
28%	T07 良い温泉、疲れの癒し	19%	20%	20%	24%	25%	19%	20%	38%	43%			
23%	T08 設備の広さ・充実性	14%	18%	26%	15%	19%	6%	22%	20%	41%			
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%				21%
23%	T13 自己・他社の利用	23%	31%	14%	18%	23%	17%	35%	21%				18%
21%	T14 湯の温度	20%	20%	17%	17%	37%	19%	23%	30%				23%
23%	T15 足湯	20%	20%	17%	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%	14%	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%	10%	12%	12%
22%	T18 砂湯、有名な温泉	41%	41%	20%	73%	22%	22%	22%	27%	26%			19%
19%	T19 充実した食事・休憩所	18%	15%	6%	13%	8%	7%	21%	19%	40%			13%
22%	T20 綺麗な施設、家族で楽しめる	15%	11%	20%	9%	29%	7%	24%	18%	36%	25%	25%	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

雰囲気
の良さ

宿泊でも日帰り
でも楽しめる

砂湯、
有名な温泉

人の多さ

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

【分析②】観光客の価値観分析

投稿者の属性×トピックの集計で、各トピックはどんな観光客層で関心が高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
1.1倍以上
全体
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%	T04 雰囲気の良さ	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%	31%	27%	27%	23%	29%	23%		
25%	T06 アクセス	23%	25%	21%	60代 女性		28%	30%	26%	乳幼児連れ 家族旅行	シニア夫婦					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%	60代 女性		14%	19%	20%	60代 男性	家族旅行 (乳幼児連れ含む)	シニア夫婦					18%		
23%	T13 自己・他社の利用	26%	27%	24%			21%	23%	21%								13%		23%
21%	T14 湯の温度	24%	25%	23%			17%	22%	21%								13%		21%
23%	T15 足湯	21%	24%	24%	25%	28%	19%	27%	23%	23%	19%	22%	26%	24%	21%	23%	24%		
22%	T16 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%		
19%	T17 体を温めること	22%	22%	21%	23%	18%	15%	18%	20%	17%	16%	20%	20%	21%	21%	21%	18%		
22%	T18 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%	女性全般 (特に20代)		23%	13%	18%	21%	23%	19%	家族旅行	シニア夫婦					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%		

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
 における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

各温泉地の特徴的な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

※写真は"九州旅ネット"フォトギャラリーより転載



別府温泉



温泉地の特徴

- T18: 砂湯、有名な温泉

ターゲット層

- 女性(特に20代)
- カップル・夫婦(シニア)
- 家族旅行

由布院温泉



- 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

黒川温泉



- 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)

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

【分析③】 満足度の要因分析

満足度に効く要因構造をトピックや属性情報からベイジアンネットワークでモデル化します

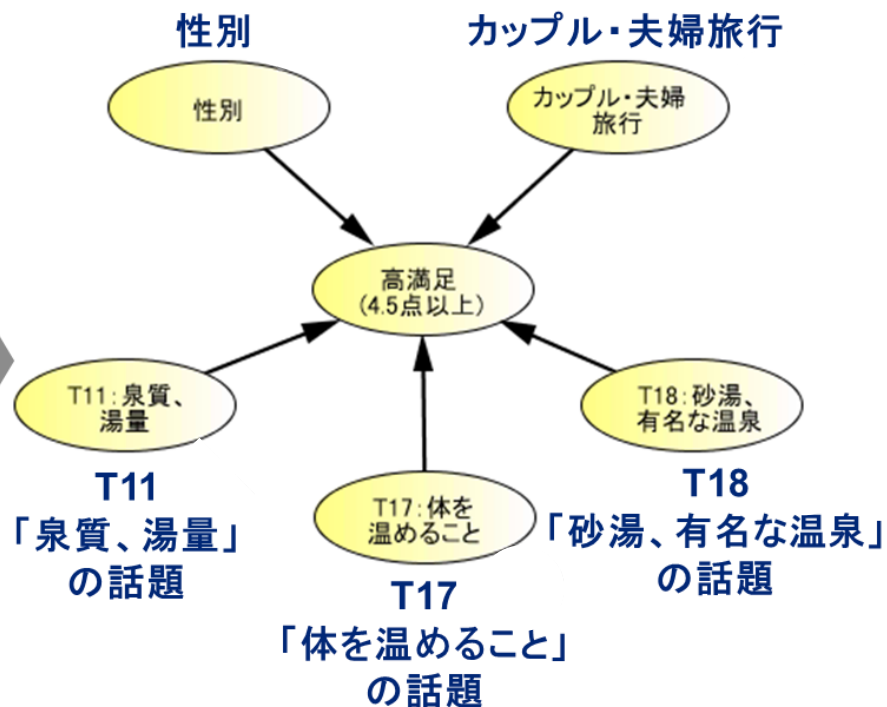
ベイジアンネットワークによる満足度モデルの構築

※ 口コミの評点4.5点以上を高満足とする



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



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？

高満足 (4.5点以上)

各要因の条件を与えたときの結果(高満足の確率)をシミュレーションできる

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	高満足 (4.5点以上)
女性	Yes	No	Yes	Yes	39.6%

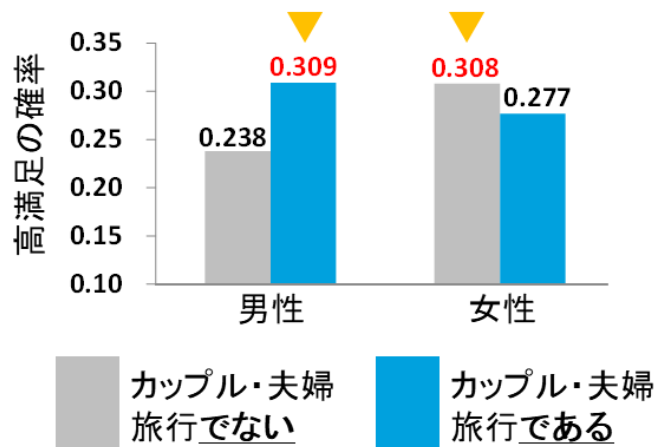
※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『テキストデータを用いた施策検討の切り札となる確率モデルの応用』より転載

【分析③】 各要因の条件による満足度の確率シミュレーション

各要因の条件を指定したときの高満足の確率をベイジアンネットワークでシミュレーションすることで、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

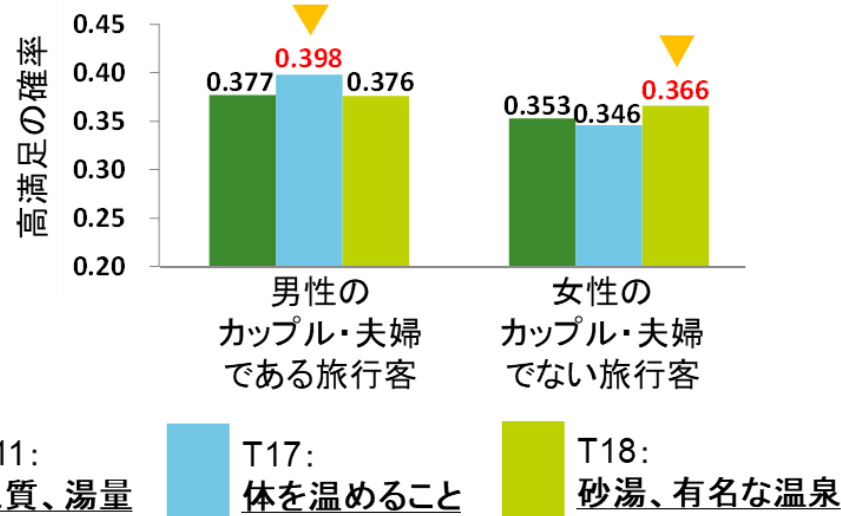
性別×同行者を条件にした高満足の確率

- 男性は、カップル・夫婦で温泉旅行に行く方が高満足の確率が高い
- 女性は、カップル・夫婦でない人と温泉旅行に行く方が高満足の確率が高い



性別×同行者×トピックを条件にした高満足の確率

- 男性がカップル・夫婦で温泉旅行に行く場合、トピックの「T17: 体を温めること」の話題がされていると、高満足の確率がさらに高まる
- 女性がカップル・夫婦でない人と(女友達や家族と)温泉旅行に行く場合、トピックの「T18: 砂湯、有名な温泉」の話題がされていると、高満足の確率がさらに高まる



※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

満足度の要因分析の結果から、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

※一部写真は"九州旅ネット"フォトギャラリーより転載

男性向けのプロモーション

恋人との温泉旅行



寒いなか温泉で温まる

女性向けのプロモーション

女友達との温泉旅行

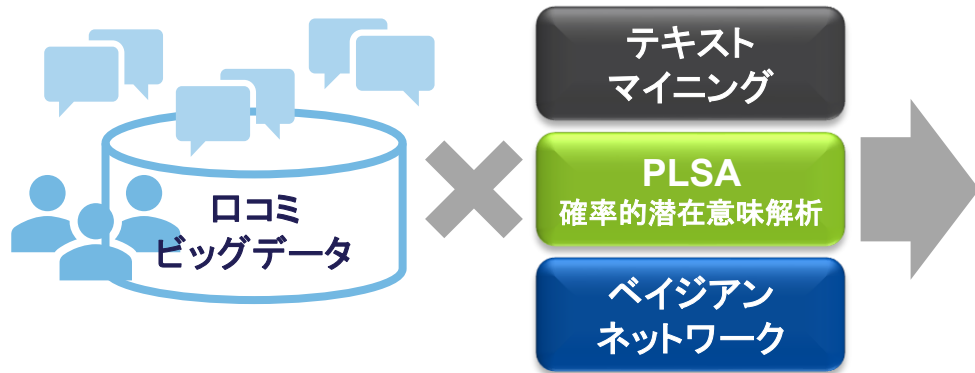


砂湯を楽しむ

※ 株式会社NTTデータ数理システムのユーザーコンファレンス2016 (<https://www.msi.co.jp/event/conference/uc2016/index.html>)
における有限責任監査法人トーマツ(野守耕爾)の講演資料『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』より転載

【まとめ】Nomolyticsで旅行口コミデータを分析するメリット

口コミのトピックを抽出し、そのトピックを軸に地域や観光客の特徴、満足度の要因を分析することで、地域観光におけるマーケティング戦略を観光客の声に基づいて検討できます



各地域が
差別化して
PRすべき
特徴を観光
客目線で把
握できる

各地域の
特徴に関心
のある効果
的なター
ゲット層を
特定できる

ターゲット
に応じた魅
力的なPR
コンテンツ
を把握でき
る

① 口コミのトピックを抽出

口コミで話題にされていることを客観的に把握できる

トピック	露天風呂の眺め	雰囲気の良い	泉質・湯量	アメニティの有無
171	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
172	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
173	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
174	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
175	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
176	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
177	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
178	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
179	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め
180	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め	露天風呂の眺め

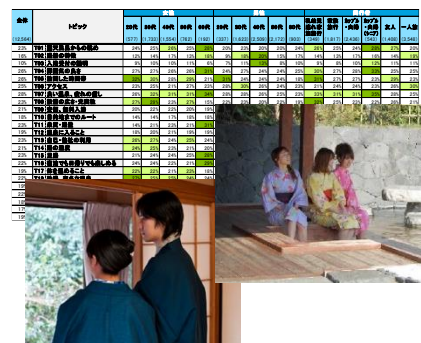
② 地域×トピックの分析

地域の特徴を観光客目線で把握できる



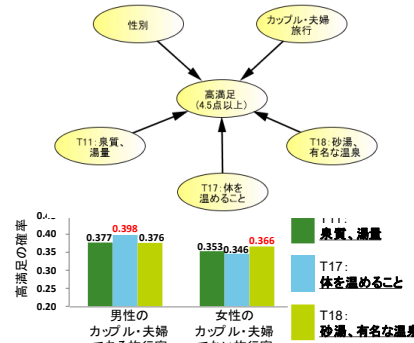
③ 観光客×トピックの分析

観光客の属性別の価値観を把握できる



④ 満足度の要因の分析

ターゲット別に満足度を高めるサービスを把握できる



**資料に関するお問い合わせやコンサルティングの
ご相談は以下までお願いします。**

analytics.office@analyticsdlab.co.jp

**会社ホームページもご参考にしてください。
過去の講演・論文資料や技術解説も掲載しています。**

<https://www.analyticsdlab.co.jp/>

株式会社アナリティクスデザインラボ

