

## テキスト分析は現状把握からシミュレーションへ

クチコミデータにTMSとBAYONETを適用した顧客ニーズモデリング

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

2014年11月21日

# トーマツ.

本発表の内容は執筆者の私見であり、有限責任監査法人トーマツの公式見解ではありません。  
また内容に含まれる企業等とは利害関係等一切の関係はありません。



# デロイトアナリティクスとは、 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？

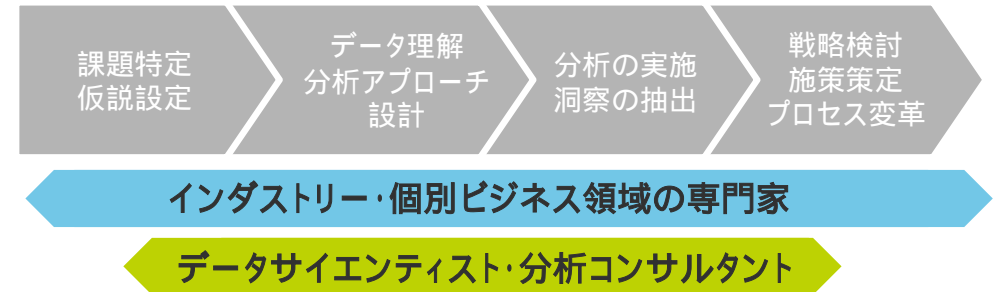
トーマツグループの各種事業分野と連携した  
包括的なコンサルティングサービスです。



- Deloitte(デロイト)は、デロイトトウシュートーマツ リミテッドのネットワーク組織で、監査、税務、コンサルティングおよびファイナンシャルアドバイザー分野等において、世界150カ国以上の拠点でサービスを提供している
- トーマツグループは、日本におけるデロイトのメンバーファームの総称

## サービスアプローチ

インダストリー・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。



## デロイトアナリティクスの特徴

経営の言葉で  
分析を語る

ビジネスとアナリティクスの両面の知見を持ち、経営観点からサービスを提供する。

経営層との  
信頼関係

トーマツの各種事業を通じて築いてきた、経営層との強固な信頼関係がある。

各業界の  
深い理解

各業界で深い知見を持つ専門家が豊富であり、より実態に即した施策を立案する。

豊富なアナリ  
ティクス人材

世界で約9,000人、国内で約100人がアナリティクス業務に従事する。

高度な  
分析技術

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・開発も行っている。

# 本日のアジェンダ

1. 複数の分析技術を統合した新たなテキスト分析アプローチ

---

2. 事例 宿泊施設のクチコミ分析

---

3. 事例 観光地のクチコミ分析

---

# 複数の分析技術を統合した 新たなテキスト分析アプローチ

# クチコミは顧客ニーズを把握し、様々な施策の立案に活用できます

## クチコミデータ活用の可能性

クチコミデータ

顧客ニーズの分析

施策の立案

ホテル

映画

レストラン

病院

化粧品

家電

観光地

大量の“顧客の生の声”

### 強み・弱み分析

- 自社の商品・サービスの好評と不評とは
- 他社の商品・サービスの好評と不評とは

### 顧客の行動理解

- そもそも顧客はどんな関心を持っているのか
- どんな価値観を持っているのか
- どんな要因でどんな反応(行動)を示すのか

- 顧客目線の魅力的な商品やサービスの企画
- 既存の商品やサービスの改善
- 顧客のターゲティング

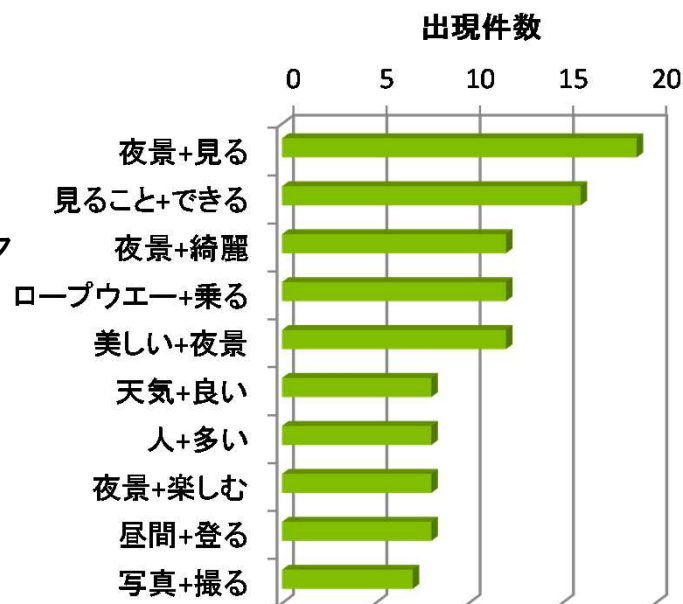


# これまではテキストマイニングをして現状把握をすることが定石でした

## 従来の分析

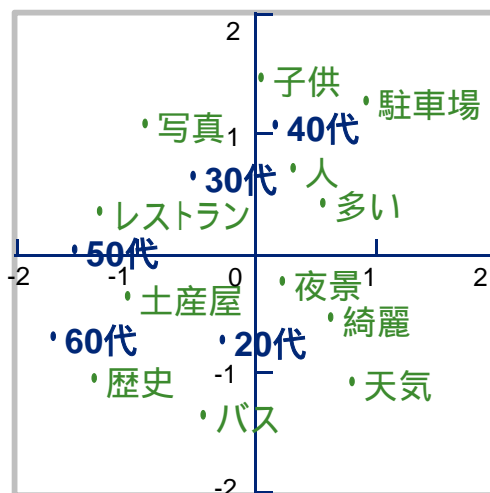
### ■ 頻度の集計

- 単語
- 係り受けペア
- ポジネガ



### ■ クロスの分析

- コレスポネンス分析
- 顧客属性と単語の対応マップ



## 課題

### ■ 解釈がしづらい

- 単語レベルでは意味の解釈が難しい
- 表現のみの抽出でその状況や背景のイメージがしづらい



トピック(話題)を抽出

### ■ 現状把握に留まる

- どのような表現が多いかは把握できる
- その結果を踏まえて何か改善を実施したときなど、条件が変わったときに結果がどのように変化するのは分からない



結果をシミュレーション

# 他の分析技術を組み合わせることでテキスト分析が進化します

## トピックの抽出

### クラスタリング技術

#### の適用

- クチコミを分類
  - 対象によって書き込み内容が大きく異なるクチコミはその表現に基づいて分類する
- トピックを抽出
  - いくつかの表現の集合体を抽出する

- 顧客の関心事や評価の観点を把握できる
- 顧客の属性ごとの価値観を把握できる
- 商品・サービスの顧客目線の特徴を把握できる

## 結果のシミュレーション

### モデリング技術

#### の適用

- クチコミ情報内の定量的な関係を分析
  - クチコミの話題や、顧客属性、商品・サービス特性、クチコミ得点などの関係を分析する
- ある条件下での結果を推論
  - 話題が変化したときの満足度得点を推論する
  - 満足度を向上させるためのサービス要因を把握する

- 顧客満足度向上のための施策の優先順位づけができる
- 施策を実施したときの効果を推論できる
- 一般化された知識として、類似する他の商品・サービスでも適用できる

# 新たなテキスト分析のアプローチで有用な知識を抽出します

## 3つの分析技術を統合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

単語抽出

### 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
- 潜在要因に基づいて、データやその変数をクラスタリングできる

トピック抽出

### ベイジアンネットワーク

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

BAYONET  
ベイジアンネットワーク構築支援システム

モデリング

テキストデータから現状把握だけでなく  
条件を変化させたときのシミュレーションが可能となる

特許出願中(特願2014-174500)



# 事例

## 宿泊施設のクチコミ分析

# 宿泊施設のサービスには様々な評価要素があります

部屋が広い

朝食が美味しい

スタッフの接客

ベッドが大きい

部屋が清潔

駅から近い

スタッフの清掃

朝食の数が豊富

チェックインが早い

禁煙ルームがある

アメニティが充実

風呂が大きい

夕食が美味しい

水圧が高い

ユニットバスのスペース

空気が綺麗

設備が充実

空調の効きが良い

観光地から近い

建物が新しい

価格が安い

タオルが新しい

駐車場がある

嫌な臭いがしない

騒音が少ない

近くにコンビニがある

チェックアウトが遅い

全室ネット対応

サービスが充実

温泉がある





顧客満足に関する評価の観点はこちらです

部屋や風呂が綺麗であること

朝食が美味しいこと

宿の雰囲気良く、楽しい旅となること

スタッフの対応が丁寧であること

臭いや空調など部屋の環境が悪くないこと

風呂の大きさや排水が悪くないこと

# 宿泊予約サイトの宿泊施設のクチコミデータを使用します

分析対象データ: じゃらんnet (<http://www.jalan.net/>) の宿泊施設のクチコミデータ



## 対象

- 京都府にある宿泊施設のクチコミの情報  
(「京都駅周辺」「河原町・烏丸・大宮周辺」)
- 投稿日: 2012年5月16日 ~ 2013年5月16日

## データ数

- 宿泊施設: 169件
- クチコミ件数: 11,535件

## 取得情報

- ユーザ属性  
(性別・年代)
- 宿泊内容  
(旅行目的、価格、部屋タイプ、食事有無)
- 項目得点  
(総合、部屋、風呂、接客サービス、清潔感、朝食、夕食)
- クチコミ内容(テキスト)
- その他宿泊施設情報

**サトちゃんさん**  
男性 / 50代  
投稿日: 2014/08/24

**駅近便利**

京都の駅から歩いて数分という立地の良さは、非常に便利で価値のあるものでした。コンシェルジュの方が非常に詳しく観光案内をして下さいました。荷物を運んで下さった若い男性の物腰も、よい印象が残りました。部屋は可もなく不可もなく。窓際の障子が京都らしさを印象づけました。窓から、わずかに新幹線と東海道本線の電車が見えて楽しかったです。

総合 3 ★★★★★ 部屋 3 | 風呂 3 | 料理(朝食) - | 料理(夕食) - | 接客・サービス 4 | 清潔感 3

【2014年08月宿泊 夫婦旅行】  
【宿泊価格帯】5,001~6,000円(大人1人あたり/税込)  
【宿泊プラン】★直前予約★お日にち限定プライス! 16:00チェックインでお得にステイ~食事なし~ [予約](#) [食事なし](#)

**ゆうちんさん**  
女性 / 50代  
投稿日: 2014/08/23

**満足しました。**

毎年夏のお参りにご利用させて頂いております。  
今回は4人一部屋ご泊りでしたが、お部屋がきれいでとても満足しています。  
車で来る方々にとっては駐車場が無料、いつも助かっています。  
今回は、朝食会場がいっぱいということで、二階の大広間が食卓もつけられましたが、却って広々として、都合が良かったと思えました。食事も美味しかったです。  
又、来年も利用したいと思っています。

総合 4 ★★★★★ 部屋 5 | 風呂 3 | 料理(朝食) 5 | 料理(夕食) - | 接客・サービス 4 | 清潔感 3

【2014年08月宿泊 家族旅行】  
【宿泊価格帯】6,001~7,000円(大人1人あたり/税込)  
【宿泊プラン】【早期予約】家族限定! 早期予約でお得に朝食付~ [予約](#) [詳細](#)

**ももさん**  
女性 / 40代

**アメニティ・施設・サービス** [詳細情報の見方・説明](#)

<input type="checkbox"/> ハンドタオル	<input type="checkbox"/> ボディソープ	<input type="checkbox"/> ドライヤー	<input type="checkbox"/> 温水洗浄トイレ
<input type="checkbox"/> 歯ブラシ・歯磨き粉	<input type="checkbox"/> 石けん	<input type="checkbox"/> 羽毛布団	<input type="checkbox"/> くし・ブラシ
<input type="checkbox"/> バスタオル	<input type="checkbox"/> 浴衣	<input type="checkbox"/> 髪剃り	<input checked="" type="checkbox"/> バストイレ別
<input type="checkbox"/> シャンプー (のりすけやが含む)	<input checked="" type="checkbox"/> バジャマ	<input type="checkbox"/> シャワーキャップ	
<input type="checkbox"/> リンス	<input checked="" type="checkbox"/> バスローブ	<input type="checkbox"/> 縁棒	
<input checked="" type="checkbox"/> バジャマ(ズボンつき)	<input checked="" type="checkbox"/> バジャマ(シャツ)	<input checked="" type="checkbox"/> シャワー独立	

総合

【2014年08月宿泊 夫婦旅行】  
【宿泊価格帯】12,001~13,000円  
【宿泊プラン】じゃらん限定!

**施設内容** 屋内プール(有料)・バー(有料)・宴会場(有料)・ラウンジ・禁煙ルーム・製氷機

**サービス&レジャー(手配含む)** ルームサービス・クリーニングサービス・エステ(有料)・マッサージ(有料)・テリカ・パティスリー・チャペル

**現地で利用可能なクレジットカード** JCB・VISA・マスター・AMEX・UC・DC・NICOS・ダイナース・UFJ・セゾン・デビットカード  
※オンラインカード決済で利用可能なクレジットカードとは異なる場合がございます。

**標準的なチェックイン時間** 14:00~ ※宿泊プランごとにチェックイン時間の設定がある場合は、そちらが優先されます。

**標準的なチェックアウト時間** ~11:00 ※宿泊プランごとにチェックアウト時間の設定がある場合は、そちらが優先されます。



# 宿泊施設の顧客ニーズをモデリングします

3つの分析技術を統合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

単語抽出

## 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
- 潜在要因に基づいて、データやその変数をクラスタリングできる

トピック抽出

## ベイジアンネットワーク

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

モデリング

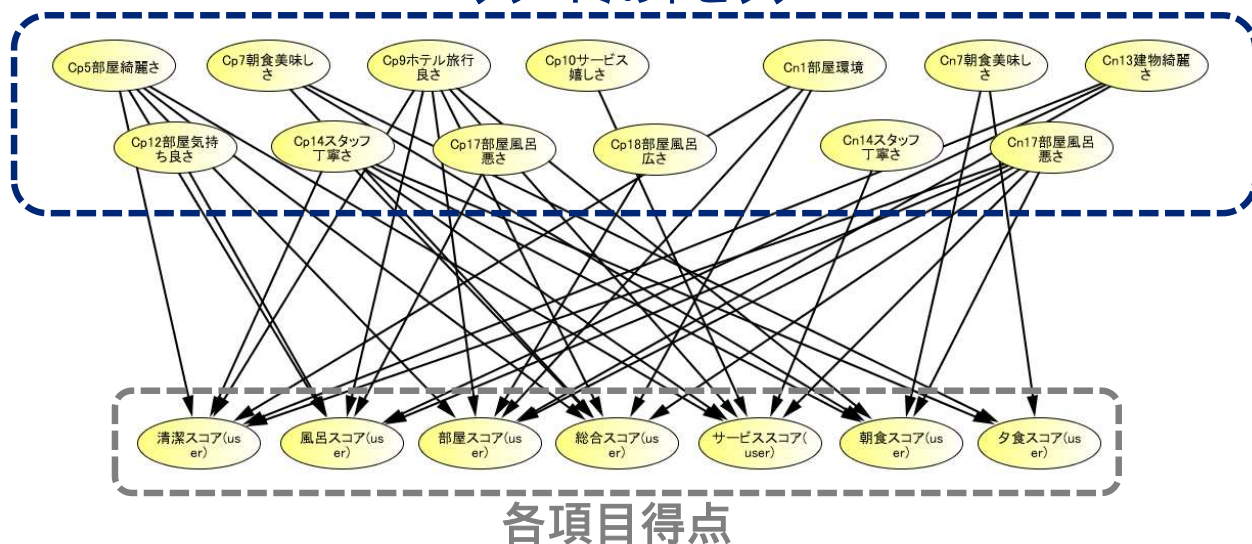
テキストデータから現状把握だけでなく  
条件を変化させたときのシミュレーションが可能となる

特許出願中(特願2014-174500)

# クチコミのトピックから満足度得点を推論するモデルを構築できます

## トピック満足モデルと推論結果例

クチコミのトピック



部屋風呂が悪い	風呂得点: 3点以下の確率	
話題多い	0.53	<div style="width: 53%;"></div>
話題少ない	0.37	<div style="width: 37%;"></div>
部屋環境が悪い	総合得点: 3点以下の確率	
話題多い	0.29	<div style="width: 29%;"></div>
話題少ない	0.16	<div style="width: 16%;"></div>

スタッフが丁寧	サービス得点: 5点満点の確率	
話題多い	0.61	<div style="width: 61%;"></div>
話題少ない	0.33	<div style="width: 33%;"></div>
ホテル旅行が良い	総合得点: 5点満点の確率	
話題多い	0.47	<div style="width: 47%;"></div>
話題少ない	0.33	<div style="width: 33%;"></div>
部屋が綺麗	部屋得点: 5点満点の確率	
話題多い	0.42	<div style="width: 42%;"></div>
話題少ない	0.29	<div style="width: 29%;"></div>
部屋が綺麗	総合得点: 5点満点の確率	
話題多い	0.43	<div style="width: 43%;"></div>
話題少ない	0.31	<div style="width: 31%;"></div>
朝食が美味しい	総合得点: 5点満点の確率	
話題多い	0.43	<div style="width: 43%;"></div>
話題少ない	0.34	<div style="width: 34%;"></div>

# Step1

## テキストマイニングで名詞と形容詞の単語を抽出します

### テキストマイニングの結果

- 「評価」に関わる**名詞**と**形容詞**に着目
- 名詞287 words、形容詞形容動詞111 words  
文章単位で出現件数が30件以上の単語に限定
- Text Mining Studio 4.2 (NTTデータ数理システム) を使用

Text Mining Studio

### 名詞の抽出単語と出現頻度の例

単語	件数	単語	件数
部屋	8,542	場所	2,280
利用	7,926	対応	2,279
ホテル	6,226	フロント	2,066
朝食	3,309	宿泊	1,988
京都	3,231	旅行	1,947
駅	3,191	値段	1,813
満足	2,487	スタッフ	1,627
お風呂	2,423	感じ	1,149

### 形容詞の抽出単語と出現頻度の例

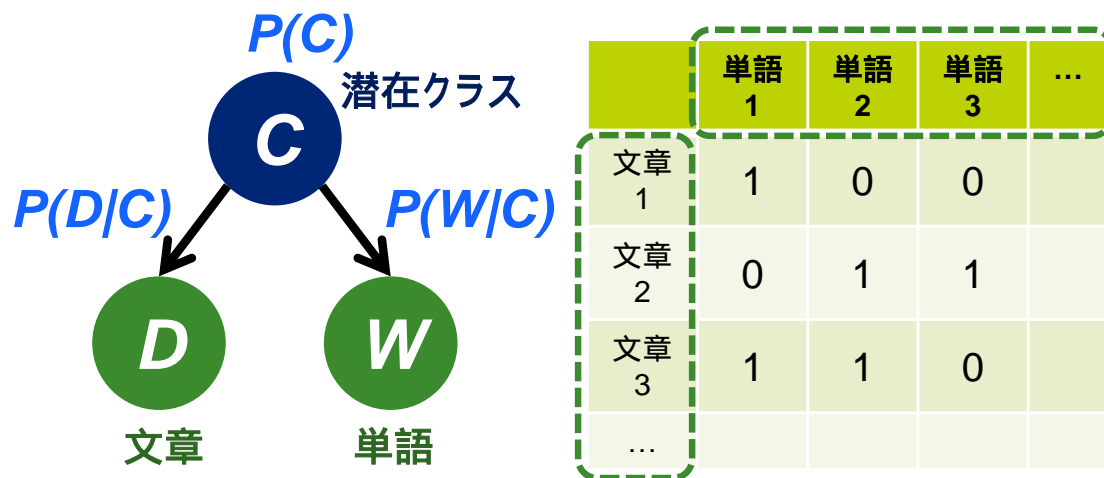
単語	件数	単語	件数
良い	7,697	快適	1,247
便利	2,413	安い	1,164
広い	2,311	狭い	1,106
美味しい	2,181	気持ち良い	1,033
綺麗	2,160	嬉しい	1,032
清潔	2,133	丁寧	930
近い	1,712	大変	843
残念	1,292	多い	784

## Step2

# 名詞vs形容詞のクロス集計にPLSAを適用して評価トピックを抽出します

### PLSA (確率的潜在意味解析)

- 文章分類に用いられるクラスタリング手法
- 文章とそこに出現する単語の間には潜在的な意味クラスがあることを想定



### ■ 行と列を同時にクラスタリング

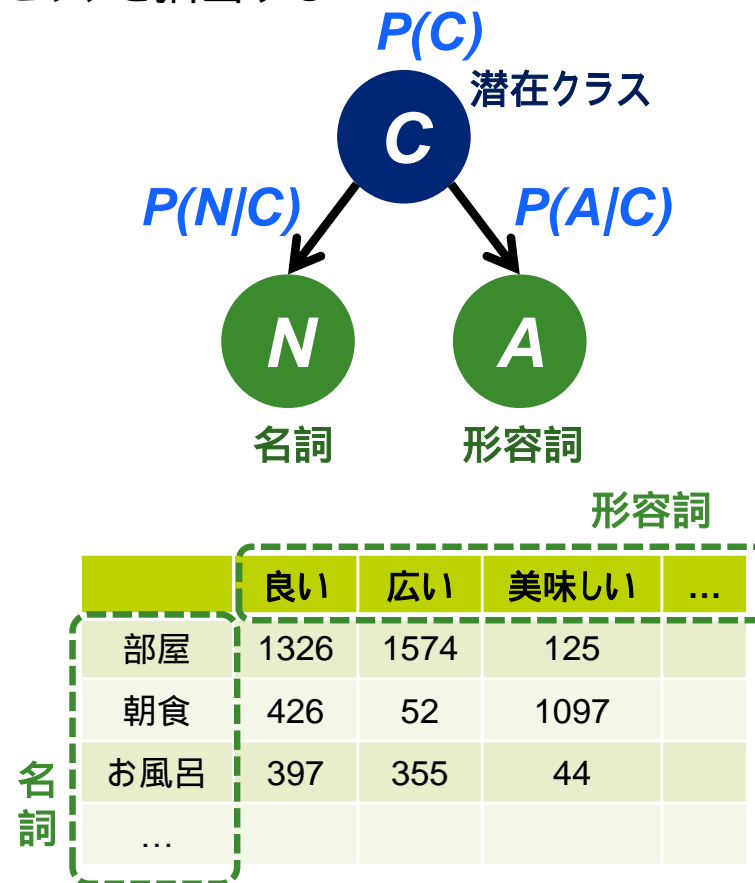
- 共起行列の行にある変数(文章)と列にある変数(単語)が共通のクラスに所属する

### ■ ソフトクラスタリング

- 全ての変数は各クラスにまたがって所属し、その所属確率( $P(D|C)$ ,  $P(W|C)$ )が与えられる

### 今回のPLSAの使い方

「名詞×形容詞」の共起行列(クロス集計)にPLSAを適用することで、口コミの評価に関するトピックを抽出する



APOSTOOL2 (産業技術総合研究所) を使用



# Step2

## PLSAで18個の評価トピック(評価の観点)が抽出されました



# Step2

## クチコミごとに18個のトピックのスコア(その話題の多さ)を計算します

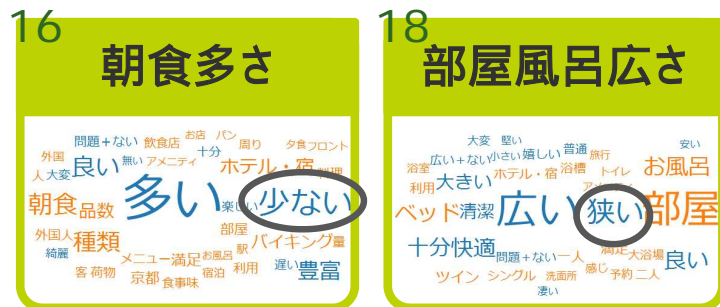
### クチコミごとのトピックスコア

クチコミID	C1	C2	C3	...	C18
1	3.1	0.9	5.5		2.4
2	1.2	5.8	2.6		1.6
...					

クチコミに含まれる単語におけるトピックの所属確率 $P(N|C)$ ,  $P(A|C)$ に基づいて計算

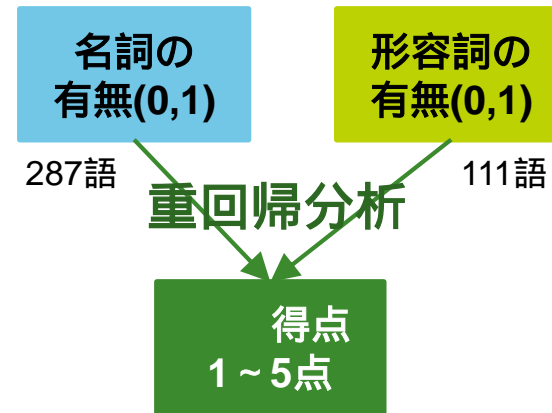
### 好評と不評が混在するトピック

- PLSAで抽出したものはあくまで評価の“観点”
- 同じトピックでも好評と不評の意味が混在している



18個のトピックを好評トピックと不評トピックに分け、クチコミの文章の好評と不評に応じてどちらか一方のトピックにスコアをつける

### クチコミ文章の好評と不評の判定方法



総合得点、部屋得点、風呂得点  
清潔得点、サービス得点、朝食得点

### ポジネガポイントの例

単語	品詞	ポジネガポイント
完璧	形容詞	0.41
駅直結	名詞	0.26
素敵	形容詞	0.21
カビ	名詞	-0.38
悪い	形容詞	-0.41
髪の毛	名詞	-0.60

- クチコミに含まれる単語を説明変数に、各項目得点を目的変数に重回帰分析する
- 6つの重回帰モデルの偏回帰係数の平均値をその単語のポジネガポイントとする
- 文章に含まれる単語のポジネガポイントの合計が、「0以上 好評文章」「0未満 不評文章」
- 判別精度(500件のサンプルで再現率)は、好評文章で約85%  
不評文章で約70%

このデータだけに特化した好評語と不評語の辞書を自動で構築可能に

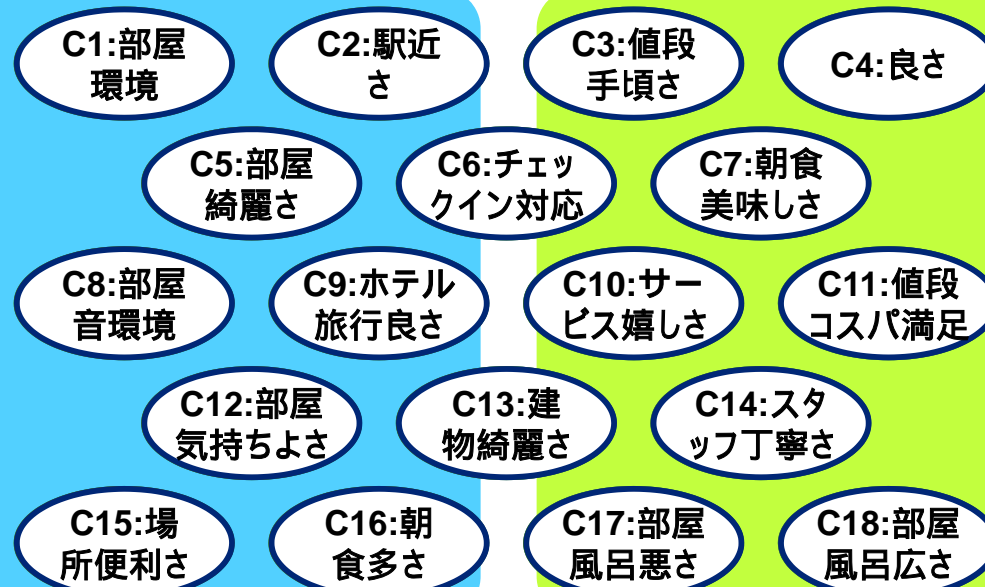
# Step3

## ベイジアンネットワークで評価トピックと項目得点の関係をモデル化します

### トピック満足モデルの構築

#### 好評トピック

#### 不評トピック



#### ベイジアンネットワーク

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下での他の変数の確率を推論する



#### ベイジアンネットワーク



トピック変数のカテゴリ設定  
(1) 話題多い (スコア3超)  
(2) 話題少ない (スコア3以下)

得点変数のカテゴリ設定  
(1) 3点以下 (低満足)  
(2) 4点 (中満足)  
(3) 5点 (高満足)

#### 各項目得点

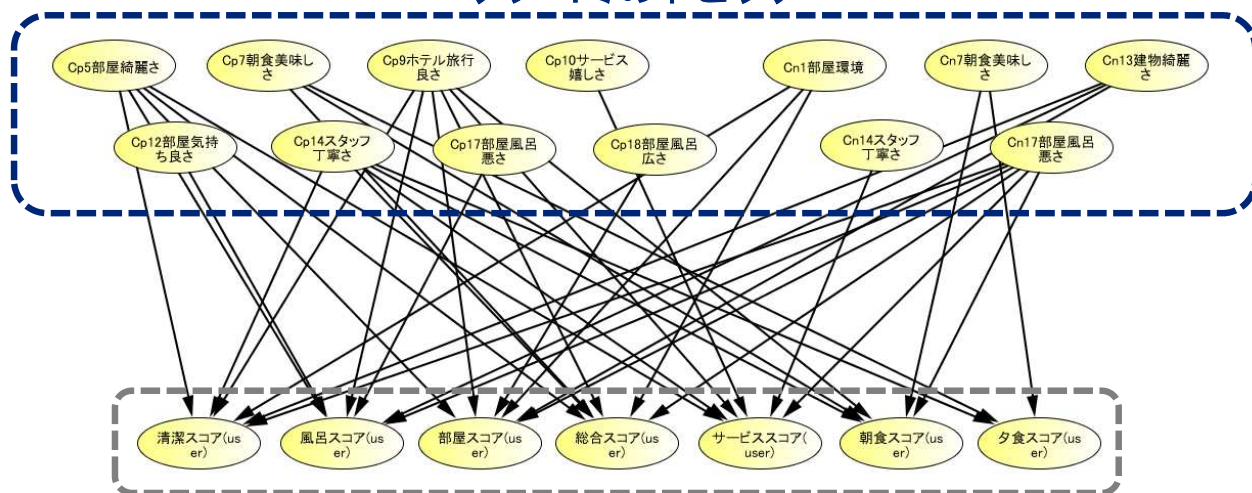
BAYONET 6.1 (NTTデータ数理システム) を使用

# Step3

## このモデルによりクチコミのトピックから満足度得点を推論できます

### トピック満足モデルと推論結果例

クチコミのトピック



各項目得点

部屋風呂が悪い	風呂得点:3点以下の確率	
話題多い	0.53	<div style="width: 53%;"></div>
話題少ない	0.37	<div style="width: 37%;"></div>
部屋環境が悪い	総合得点:3点以下の確率	
話題多い	0.29	<div style="width: 29%;"></div>
話題少ない	0.16	<div style="width: 16%;"></div>

スタッフが丁寧	サービス得点:5点満点の確率	
話題多い	0.61	<div style="width: 61%;"></div>
話題少ない	0.33	<div style="width: 33%;"></div>
ホテル旅行が良い	総合得点:5点満点の確率	
話題多い	0.47	<div style="width: 47%;"></div>
話題少ない	0.33	<div style="width: 33%;"></div>
部屋が綺麗	部屋得点:5点満点の確率	
話題多い	0.42	<div style="width: 42%;"></div>
話題少ない	0.29	<div style="width: 29%;"></div>
部屋が綺麗	総合得点:5点満点の確率	
話題多い	0.43	<div style="width: 43%;"></div>
話題少ない	0.31	<div style="width: 31%;"></div>
朝食が美味しい	総合得点:5点満点の確率	
話題多い	0.43	<div style="width: 43%;"></div>
話題少ない	0.34	<div style="width: 34%;"></div>

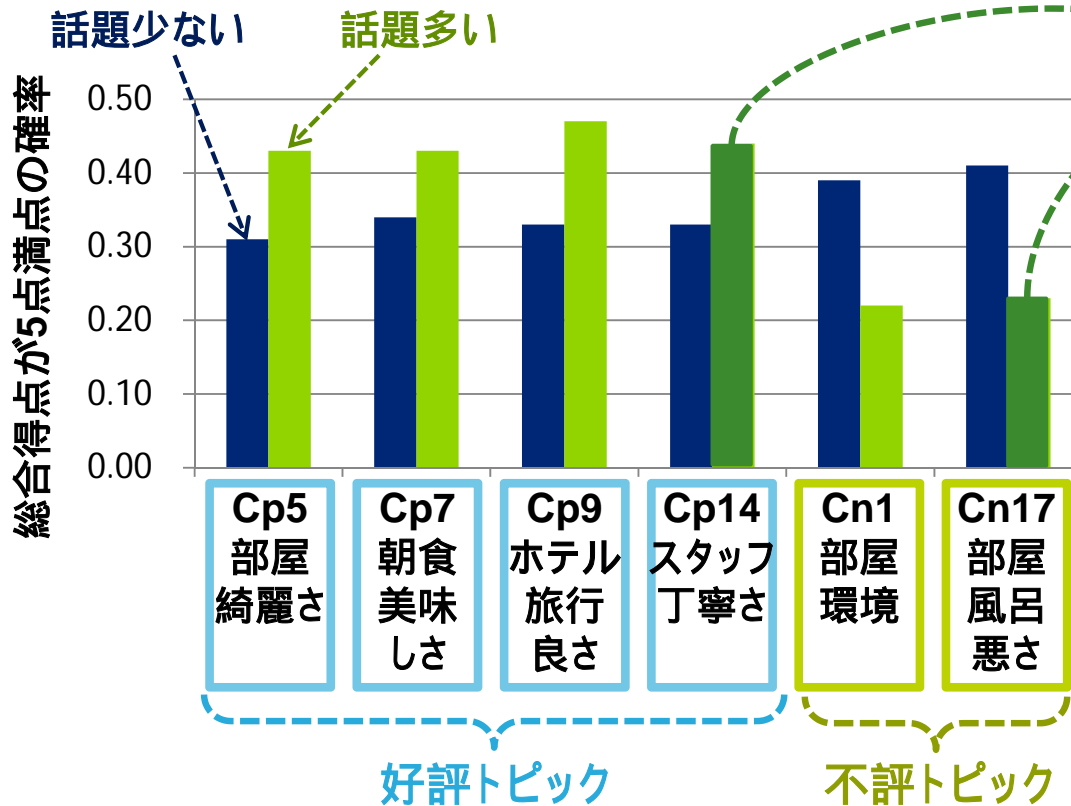


# Step3

## 宿泊施設のスタッフは笑顔で元気良く、水まわりは排水の良さも重要です

総合得点が5点満点となる確率

推論結果の具体化



再びテキストマイニングに立ち返る

Cp14: スタッフ丁寧さ

Cn17: 部屋風呂悪さ

係り受け表現	頻度	リフト
対応+良い	457	13.2
丁寧な+対応	209	14.1
親切な+対応	126	14.2
接客+丁寧な	80	14.0
フロント+親切な	46	13.4
気持ち良い+対応	42	14.2
笑顔+素敵な	26	13.5
丁寧な+説明	22	10.7
迅速な+対応	21	12.2
挨拶+気持ち良い	11	14.5
気遣い+素晴らしい	11	14.5
嫌な+顔+ない	10	14.5

係り受け表現	頻度	リフト
部屋+狭い	222	13.8
お風呂+狭い	153	14.6
気持ち+悪い	45	14.9
狭い+感じ	31	13.5
排水+悪い	26	15.2
トイレ+狭い	22	14.5
浴槽+狭い	14	15.2
流れ+悪い	13	15.2
脱衣所+狭い	12	15.2
冷蔵庫+小さい	12	13.0
駐車場+狭い	11	15.2
テレビ+小さい	9	13.7

宿泊客はスタッフの表情や挨拶をよく見ている

狭さだけでなく、排水の悪さは大きなマイナス

どのような観点のサービスが満足度をどの程度押し上げ、押し下げるのか定量的に把握することで、サービス充実化の優先順位をつけられる

推論結果から改めて口コミの内容を確認することで具体的な改善施策を検討できる

# 宿泊施設のクチコミから顧客ニーズをモデル化できました

3つの分析技術を統合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

宿泊施設の評価にはどのような表現が多いのか？

単語抽出



PLSA  
確率的潜在意味解析

お客様はどのような観点の評価軸を持っているのか？

トピック抽出



ベイジアンネットワーク

満足度アップにはどんな観点のサービスから優先的に充実化すべきか？

モデリング

宿泊施設の現状評価を把握するだけでなく、  
施策を実施したと仮定したときの効果を  
定量的にシミュレーションすることができる

# 事例

## 観光地のクチコミ分析

# 旅行サイトの観光地(北海道)のクチコミデータを使用します

旅行クチコミサイト フォートラベル (<http://4travel.jp>)



**観光地**

函館山のクチコミ 172件

**タイトル(テキスト)**

絶景は日没20分後

満足度: ★★★★★ 5.0

**総合得点**

わさんぼんさん

男性 / 函館のクチコミ: 1件

**性別**

**旅行時期**

旅行時期: 2013/08(約7ヶ月前)

**同行者**

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美しい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。

**クチコミ(テキスト)**

観光地のクチコミは観光ジャンルによって  
クチコミする話題が大きく異なる特徴がある

- 北海道の観光地: 376スポット(クチコミ10件以上)
- クチコミ件数: 12,303件
- 対象期間: 旅行時期が2014年1月以前の全投稿データ(古いものでは旅行時期が20年前のものもある)

北海道 観光 人気ランキング 4,286件

	<p><b>1位 旭川市旭山動物園</b></p> <p>カテゴリ: 観光・遊ぶ / 動物園</p> <p>エリア: 旭川 [地図]</p> <p>🚗 駐車場に注意</p> <p>🐼 動物の魅せ方が上手です</p> <p>みんなの満足度: ★★★★★ 4.65</p> <p>アクセス: 3.31</p> <p>コスト: 4.01</p> <p>動物・展示物の充実度: 4.43</p> <p>施設の快適度: 3.88</p> <p>人混みの少なさ: 2.98</p> <p>クチコミ 525件 &gt;&gt;</p>
	<p><b>2位 小樽運河</b></p> <p>カテゴリ: 観光・遊ぶ / 名所・史跡</p> <p>エリア: 小樽 [地図]</p> <p>🌃 夜がオススメ</p> <p>🌀 混雑</p> <p>みんなの満足度: ★★★★★ 4.54</p> <p>アクセス: 3.65</p> <p>人混みの少なさ: 3.10</p> <p>見ごたえ: 3.67</p> <p>バリアフリー: 3.14</p> <p>クチコミ 321件 &gt;&gt;</p>
	<p><b>3位 大通公園</b></p> <p>カテゴリ: 観光・遊ぶ / 公園・植物園</p> <p>エリア: 駅周辺・大通り [地図]</p> <p>🌞 きれいな公園</p> <p>みんなの満足度: ★★★★★ 4.54</p> <p>アクセス: 4.24</p> <p>人混みの少なさ: 3.40</p>







# 北海道の観光地は7つグループに分類されることが分かりました

## 1 車で行くような迫力のある景勝地

オシンコシンの滝



神威岬



硫黄山



知床峠



宗谷岬



立待岬



## 2 動物や展示、体験等家族で楽しめる施設

旭山動物園



白い恋人パーク



円山動物園



網走監獄



札幌ビール博物館



摩周丸



## 3 屋内の見学も可能な歴史的建造物

北海道庁旧本庁舎



時計台



青い池



旧函館区公会堂



カトリック元町教会



ハリストス正教会



## 4 写真撮影の人気スポット

小樽運河



ケンとメリーの木



セブンスターの木



ホワイトイルミネーション



雪まつり



八幡坂



## 5 買物や飲食等を楽しめる観光スポット

金森赤レンガ倉庫



函館朝市



狸小路商店街



すすきの



ら~めん共和国



阿寒湖アイヌコタン



## 6 展望台等からの景色が綺麗な場所

函館山



摩周湖



五稜郭タワー



テレビ塔



知床五湖



洞爺湖



## 7 花や緑などが豊かな公園・花園

ファーム富田



大通公園



四季彩の丘



モエレ沼公園



中島公園



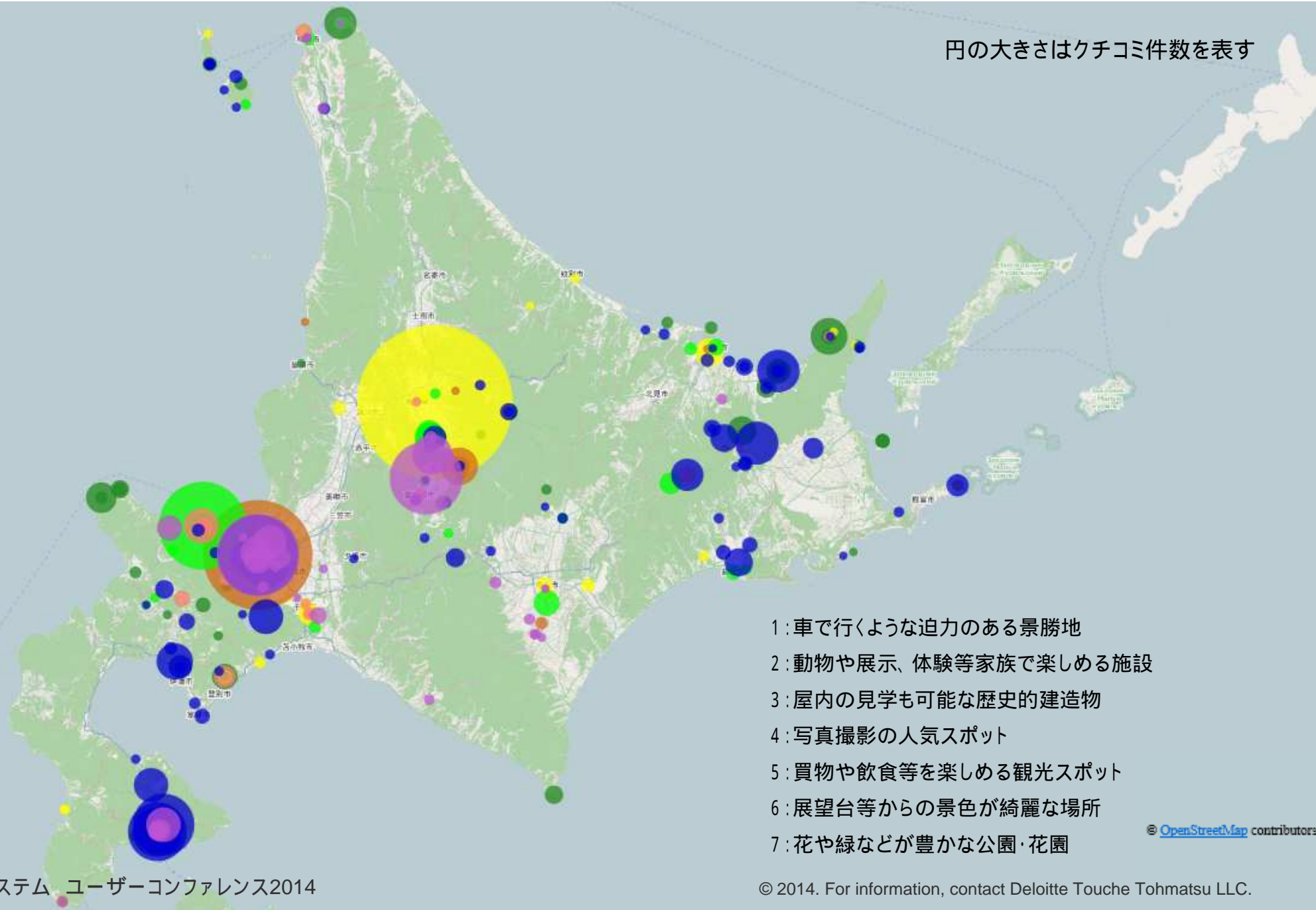
五稜郭公園



クチコミに含まれる係り受け表現と観光地のクロス集計にPLSAを実行して観光地を分類

# グループ別にマッピングしますと北海道の観光特徴の分布が分かります

円の大きさはクチコミ件数を表す



- 1: 車で行くような迫力のある景勝地
- 2: 動物や展示、体験等家族で楽しめる施設
- 3: 屋内の見学も可能な歴史的建造物
- 4: 写真撮影の人気スポット
- 5: 買物や飲食等を楽しめる観光スポット
- 6: 展望台等からの景色が綺麗な場所
- 7: 花や緑などが豊かな公園・花園

© OpenStreetMap contributors





# Group6「展望台等からの景色が綺麗な場所」について

## 代表的な観光地

函館山



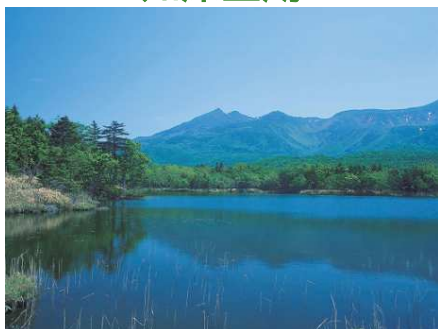
摩周湖



五稜郭タワー テレビ塔



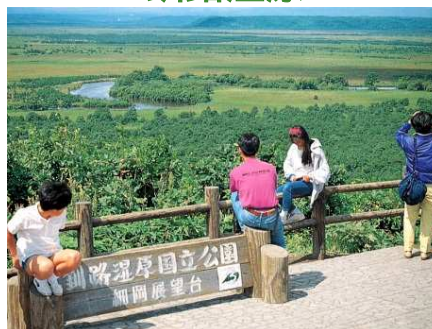
知床五湖



洞爺湖



釧路湿原



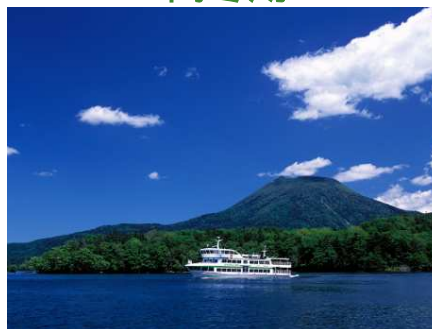
もいわ山展望台



さっぽろ羊ヶ丘展望台



阿寒湖



## 代表的な表現

見ること+できる

展望台+ある

天気+良い

夜景+見る

晴れる+日

ある+展望台

すばらしい+景色

楽しむこと+できる

景色+楽しむ

## データ情報


■ クチコミ: 3,067件

■ 観光地: 83件



# Group6の観光地では、観光客は14種の体験を話題にしています

1 高い場所に登ること




新しい 高い のぼる 上 ここ 遠く 形  
高き 出来る 高い 大道公園 山 上がる  
景観 ロープウェイ テレビ塔 ジャンプ台 綺麗だ  
見る ある 展望台 タワー 五稜郭  
眺める 分かる 見下ろす なる 見える 作る  
星型 登る 思う 上る 行く ない すぐ 建物  
見渡す 夜景 する 最高 歴史

2 土産屋や飲食店の存在




景観 ない 見る すぐ 1 購買 いる ここ 中  
いろいろ 多い 楽しむ 土産 店 売る レストラン  
タワー 土産屋 ある レストラン 売店 像  
おいしい 夜景 ある 周辺 ソフトクリーム  
眺める 出来る 有名 だ 展望台 展示  
思う テレビ塔 食べる 宿泊施設 する 良い 近く お勧め  
写真 なる いう 駐車場  
そこ

3 車でのアクセス




上る そこ 寄る 停める  
市電 時間 アクセス 良い 思う 遊歩道 ない 道 登る  
降りる 行く 楽になる 着く レンタカー  
歩く 向かう ある 展望台 駐車場  
すぐ だ 途中 かかる かなり だ 走る 駅 ロープウェイ  
進む 国道 無料 だ バス 便利 だ 車 湖 階段 広い  
通る 景観

4 湖の神秘性



呼ぶ 中心 景観 周辺 原生林 湖 咲く 湖 水  
残る 最高 だ 雄大 だ 透明度 良い 静か だ 高い する 感じる 湖 島  
自然 だ 眺める 綺麗 だ 海 ある 姿 見える  
雪 神秘的 だ 出来る そびえる 雰囲気 色 青い 湖面  
森 浮かぶ 囲む 異なる 晴れる 山 天気 空 映る よる  
湖畔 癒す 誇る 美しい 見る 中 いる 言う

5 ガイドツアー




熊 受ける かける 聞く 散策 説明 後 歩く 私  
中 コース 思う いう 自分 分かる 必要 だ みる  
する まわる バス ある 行く なる 案内 見る  
良い ない 時間 遊歩道 いる 周辺 話 出る 出来る  
形 多い よる 湖 子供

6 鮮やかな草花



他  
いう 紅葉 景観 高い かける 多い 散策 高 満 咲 車 時期  
島 花 一番 ここ 展望台 コース 咲く これ くる ない  
自然 だ 感じる 夏 ある 低い 楽しむ 私 思う  
咲き 誇る スポット 綺麗 だ 良い 言う 十分 だ 好き だ 湖  
自分 眺める 比べる 季節 価値 見る 出来る 近い 早い  
周辺 第一 桜 山

7 温泉やスキー等のレジャー



名前  
観光 すぐ だ 温泉街 ここ 湖畔 山 砂  
入る 異なる 大きい テレビ塔 冬 ない これ 楽しむ 霧  
感じる 観光地 ある 有名 だ 像 いう  
湖 丘 温泉 なる 出る する 良い 寄る 思う 雰囲気  
近い 近く 言う 面白い 中心 高さ 観光客 規模 一つ

8 落ち着いて楽しむ景色




雄大 だ 人  
登ること 駐車場 ある こと ない 遊歩道 良い  
眺望 中 眺望 出来る 湖 歩く ある 楽しむ 事  
行く 事 無料 だ する すぐ だ 少ない 整備 車 広い 見る 事  
自然 だ

9 天気の良い悪し




あまり だ 時間 恵む 見る 雄大 だ 景観  
眺望 アクセス する ない 時間 いう 行く 景観  
ここ 楽しむ 眺める 良い 見晴らし 最高 だ  
ただ 異なる 展望台 悪い 出来る 今回 ある  
思う 夜景 見る 事 日 後 群 だ よる

10 乗物 & 写真撮影




やる とる 自分 上がる いる 今回 使う 子供  
バス 見る 展望台 登る 上 同じ だ 降りる 写真  
ここ エレベーター 宿泊施設 借りる する 前  
楽しむ 思う ロープウェイ 泊まる 車 乗る  
私 撮る ゆっくり する リフト 行く レンタカー  
夜景 時間 かかる 遊覧船 いう くる なる  
待つ 入 料 私 たち 景観 混む 観光  
並ぶ

11 時期で異なる綺麗な景色




登る 天気 夜 寒い 日 山  
昼 澄む 見る 事 楽しむ 訪れる 姿 ない ここ 行く  
冬 素敵 だ イルミネーション 見える 眺める  
景観 なる 綺麗 だ 夜景 見る  
いう ライトアップ 思う 出来る 夕日 最高 だ 湖  
季節 本 当然 だ お勧め 時期 する 空気 中 紅葉 桜  
よる 途中

12 雲霧雨雪等の天候と視界




眺める 自分 言う 多い 降る 湖面 絶景 雪 出る なる  
ある 見る 見る 事 ない この 日 夜 景観 真っ 白 だ  
あいに くだ 思う 見える 綺麗 だ 晴れる  
雨 残念 だ 天気 展望台 出来る 日霧 私  
雨 かかる 訪れる 行く いう 頂上 期待 山  
有名 だ 冬 覆う 訪れる 行く いう 頂上 期待 山  
遠く 登る 雲

13 観光客の多さ



する 混む 賑わう 雪 寄る 平日  
展望台 ある 時間 くる 何 度 時期 冬 バス 見る  
強い ゆっくり する かなり だ 早い 駐車場 夕方 あまり だ  
多い 車 いる 行く 風 観光客 少ない  
多数 ロープウェイ 寒い 感じる お勧め 訪れる 人  
日 思う ない いう なる ただ 良い 観光地 夏  
非常 だ ここ

14 眼下に広がる大自然



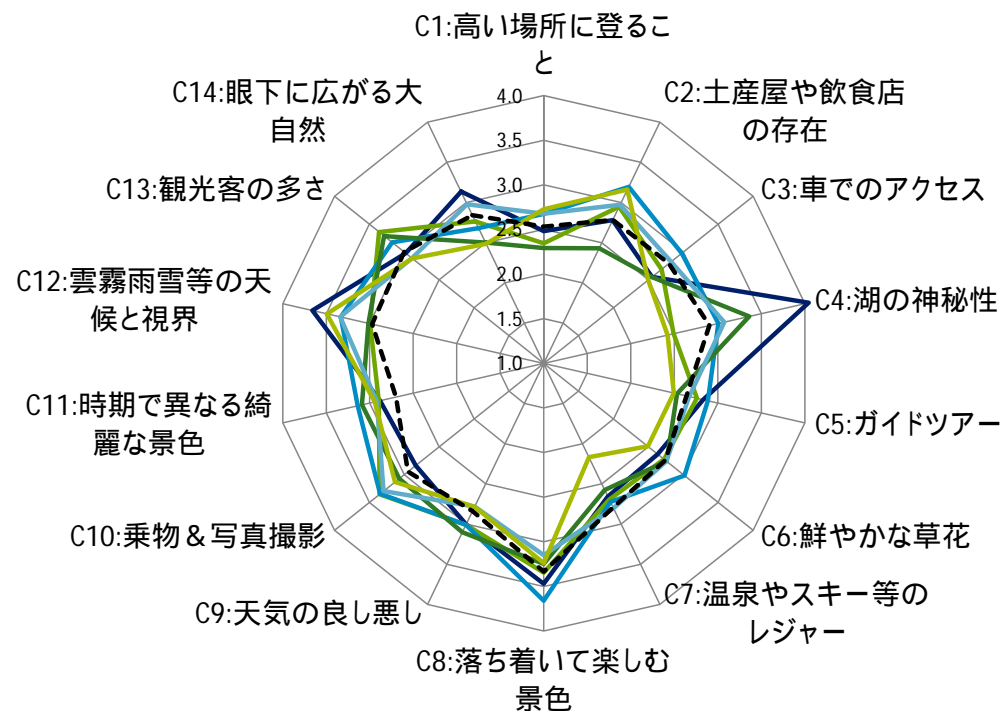
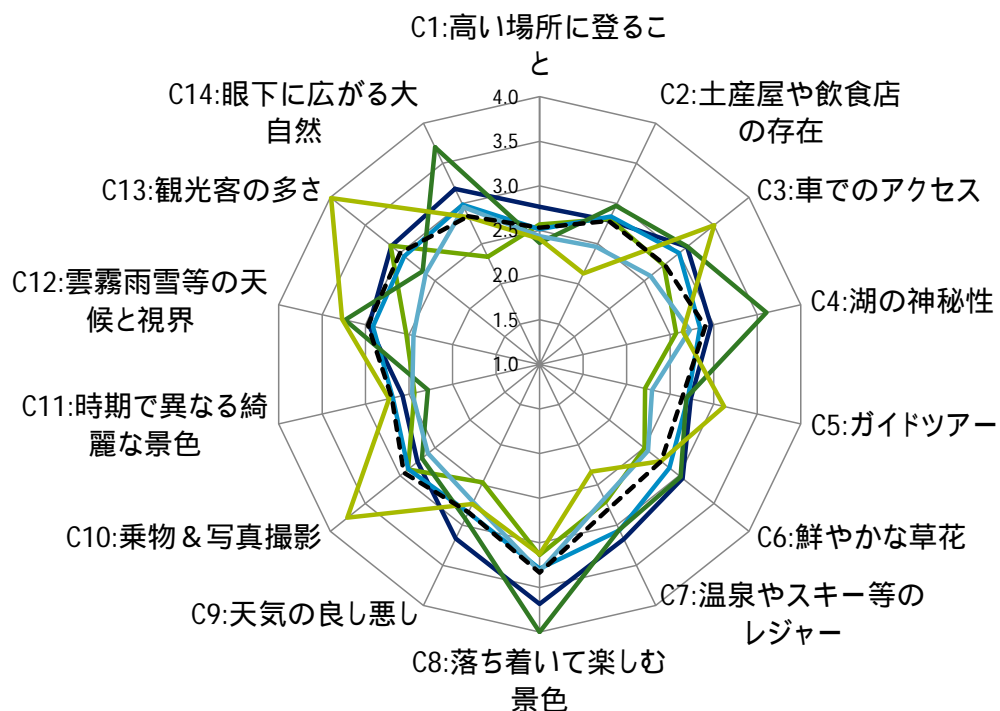
海 地球 眼下 思う 視界 中 いう 感動  
入る 釧路 湖原 360度 絶景 望む 眺める 上  
出来る 感じる 最高 だ パノラマ 丸い  
景観 見る 見える 湖 湖原 丘 広がる  
広い 見渡す ある 広大 だ 展望台 雄大 だ  
何処 ない ここ 反対側 綺麗 だ 続く 遠く 楽しむ  
周辺 走る

クチコミに含まれる名詞、形容詞、動詞のクロス集計にPLSAを適用してクチコミのトピックを抽出した

# 14個の話題のスコアを計算して観光客の価値観を定量分析できます

## 男性 × 同行者の価値観

## 女性 × 同行者の価値観



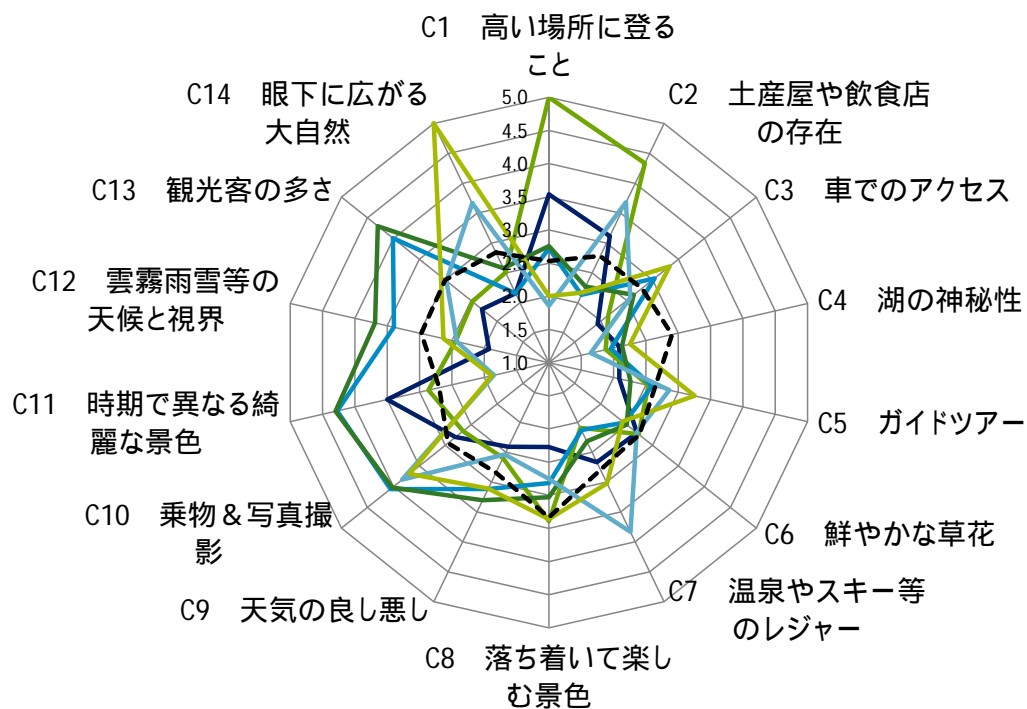
..... 平均

- 一人旅
- カップル・夫婦
- 家族旅行
- 友人
- カップル・夫婦(シニア)
- 乳幼児連れ家族旅行

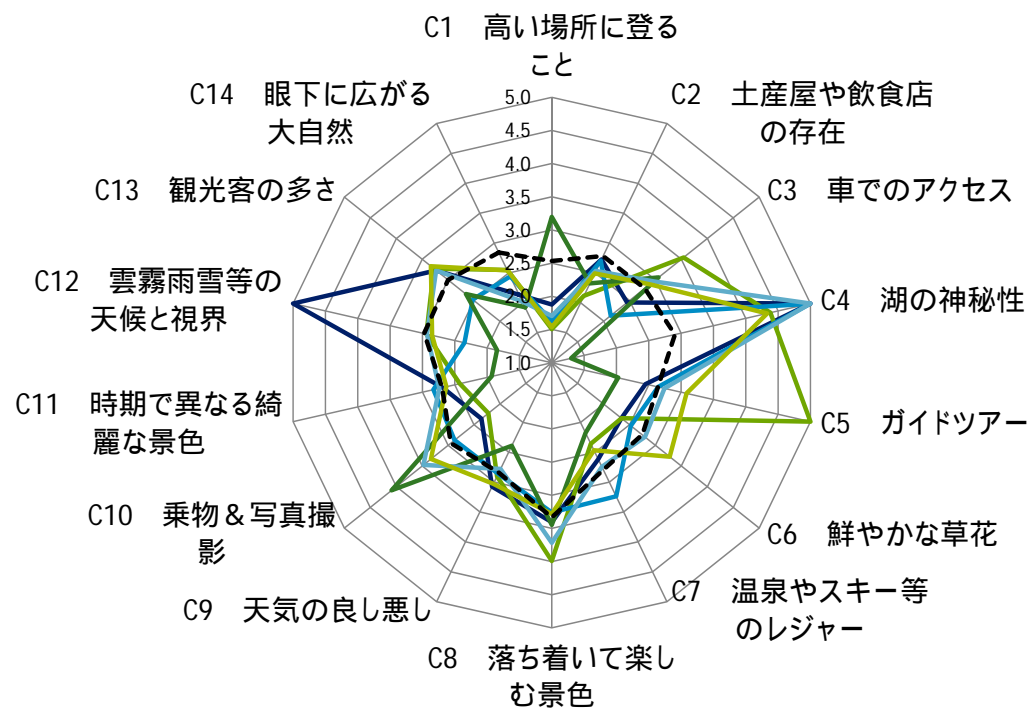
- 一人旅
- カップル・夫婦
- 家族旅行
- 友人
- カップル・夫婦(シニア)
- 乳幼児連れ家族旅行

# 14個の話題のスコアを計算して観光地の特徴を定量分析できます

## 各観光地で話題にされる観光価値の特徴



..... 平均



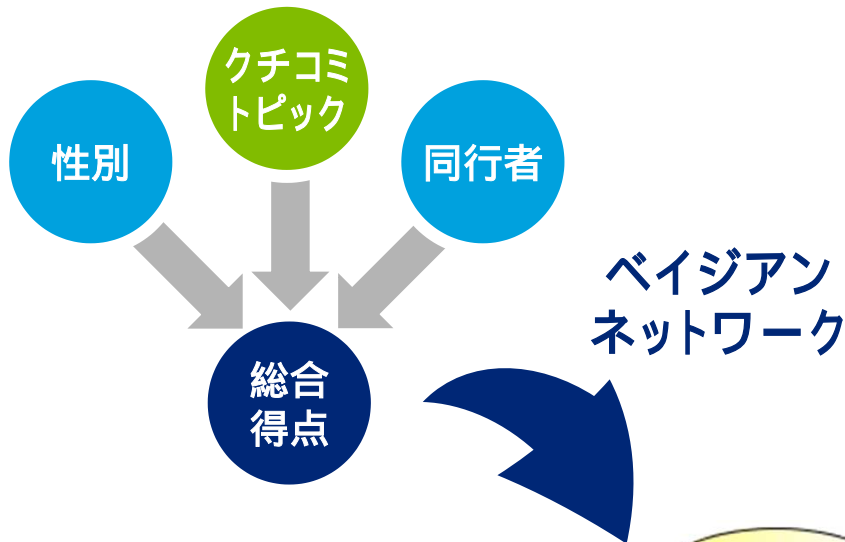
- さっぽろテレビ塔
- 函館山
- さっぽろ羊ヶ丘展望台
- 五稜郭タワー
- 函館山展望台
- 釧路湿原

- 摩周湖
- 洞爺湖
- 支笏湖
- 知床五湖
- 大倉山ジャンプ競技場
- 大沼国定公園

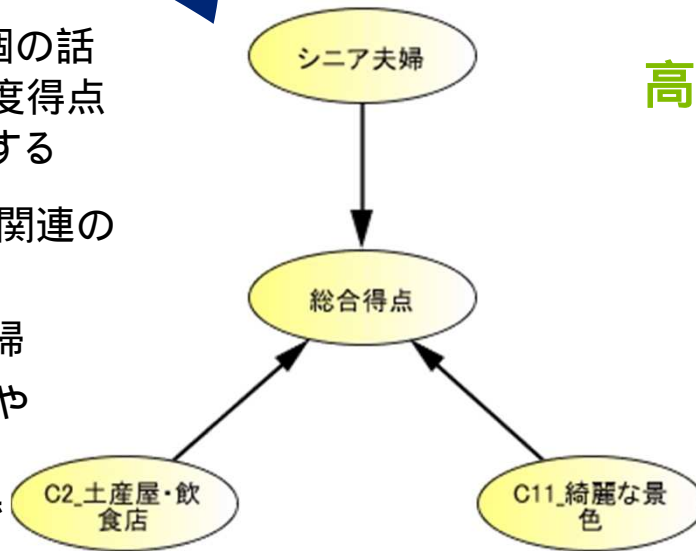


# 観光客の満足度を形成する要因の関係をモデル化します

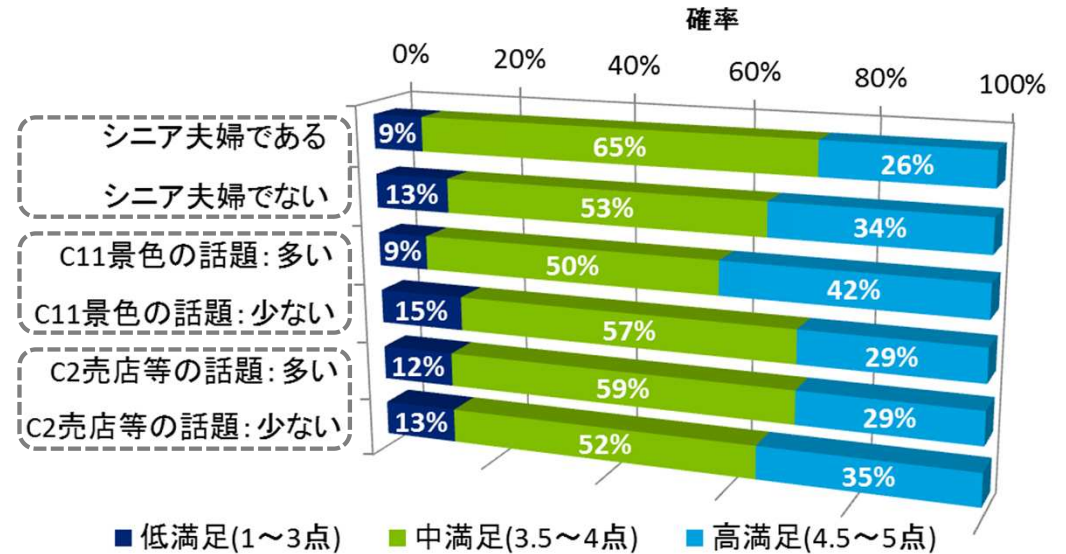
## 観光客の満足度形成モデル



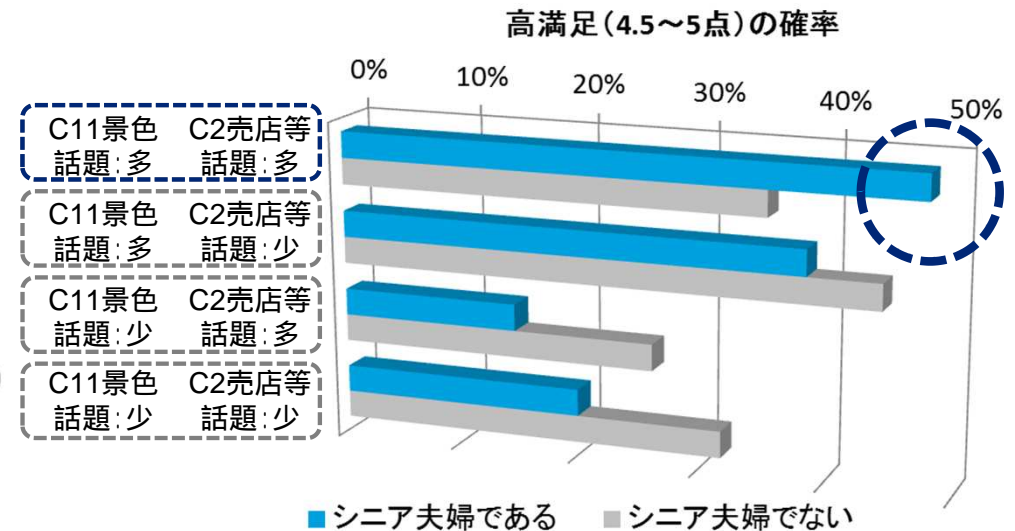
- PLSAで抽出した14個の話題と顧客属性と満足度得点との関係をモデル化する
- 総合得点(満足度)と関連の強い要因は
  - 同行者: シニア夫婦
  - 話題: C2\_土産屋や飲食店の存在
  - 話題: C11\_時期で異なる綺麗な景色



## 各変数を条件とした満足度シミュレーション



## 高満足度を向上させるサービス条件の発見





# 例) 函館山は景色だけの観光地ではありません

充実したお土産ショップ



景色を眺めながらの食事



景色は格別



5.1ch  
サラウンド  
シアターの  
上映



# 観光地のクチコミから顧客ニーズをモデル化できました

## 3つの分析技術を統合した新たなテキスト分析アプローチ

函館山のクチコミ 172件

絶景は日没20分後  
満足度: ★★★★★ 5.0

テキストマイニング  
PLSA

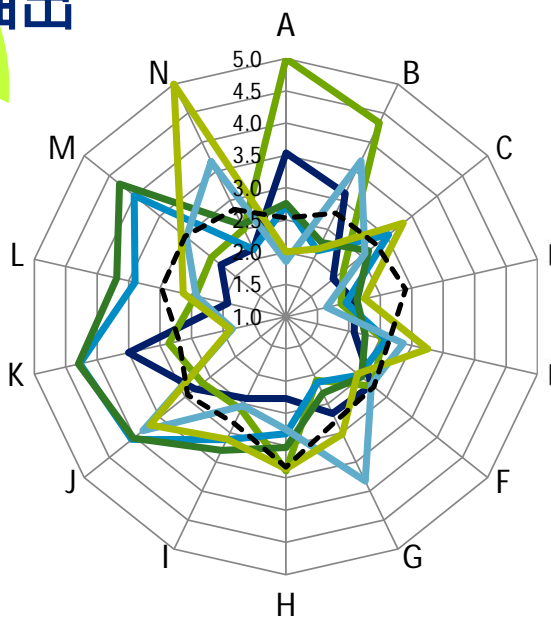
- 高い場所に登ること
- 土産屋や飲食店の存在
- 車でのアクセス
- 湖の神秘性
- ガイドツアー
- 鮮やかな草花
- 温泉やスキー等のレジャー
- 落ち着いて楽しむ景色
- 天気の良い/悪い
- 乗物&写真撮影
- 時期で異なる綺麗な景色
- 雲霧雨雪等の天候と視界
- 観光客の多さ
- 眼下に広がる大自然

面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、星に来ると面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美し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。

トピックを抽出

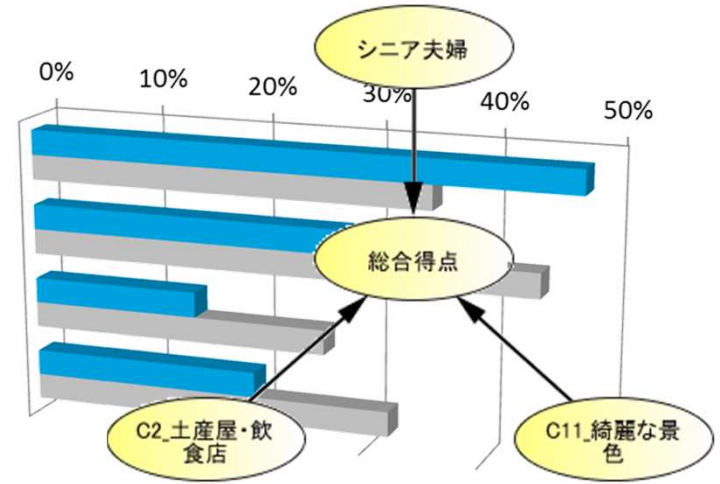


観光客目線から観光地を分類



顧客の価値観やサービスの特徴を定量把握

ベイジアンネットワーク



様々なサービス要因から観光客満足度をシミュレーション

# 新たなテキスト分析のアプローチで有用な知識を抽出できます

## 3つの分析技術を統合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

単語抽出

### 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
- 潜在要因に基づいて、データやその変数をクラスタリングできる

トピック抽出

### ベイジアンネットワーク

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

モデリング

テキストデータから現状把握だけでなく  
条件を変化させたときのシミュレーションが可能となる

特許出願中(特願2014-174500)

# クチコミに限らずテキストデータであれば同様の分析が可能です

## テキストデータの適用例



### コールセンターログ

どのような内容の問合せ・苦情だと  
解約の発生確率を高めてしまう？



### 営業日報

どのような営業サービスが契約の  
獲得率を促進できる？



### 自由記述アンケート

どのような商品・サービス特性が  
どのような話題をつくりだす？



### ヒヤリハットレポート

どのような作業環境に改善すれば  
事故の発生確率を抑制できる？

# Deloitte. トーマツ.

トーマツグループは日本におけるデロイトトウシュートーマツ 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れらの関係会社(有限責任監査法人トーマツ、デロイトトーマツ コンサルティング株式会社、デロイトトーマツ ファイナンシャルアドバイザー株式会社および税理士法人トーマツを含む)の総称です。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、各社がそれぞれの適用法令に従い、監査、税務、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7,600名の専門家(公認会計士、税理士、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。詳細はトーマツグループWebサイト([www.tohmatsu.com](http://www.tohmatsu.com))をご覧ください。

Deloitte(デロイト)は監査、税務、コンサルティングおよびファイナンシャル アドバイザーサービスをさまざまな業種にわたる上場・非上場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提供しています。デロイトの約200,000名を超える人材は、“standard of excellence”となることを目指しています。

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ートーマツ 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.tohmatsu.com/deloitte/](http://www.tohmatsu.com/deloitte/) 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。

Member of  
**Deloitte Touche Tohmatsu Limited**