

# 宿泊施設の属性を利用した

# 楽天トラベルの苦情データのテキストマイニング

成川 健太郎, 三道 弘明(関西学院大学)

## 背景と目的

### 背景

- データマイニングが**ビジネス面で主流**になりつつある  
日々大量のデータが蓄積(=ビッグデータ)

- 宿泊施設の苦情データに対する研究は多くない

### 目的

- ビジネスにおける有益な知見を獲得すること**  
情報技術の開発ではなくビジネスに役立つ苦情データ分析
- ビジネスにおけるデータマイニングの有用性を明らかにすること**

## アプローチ

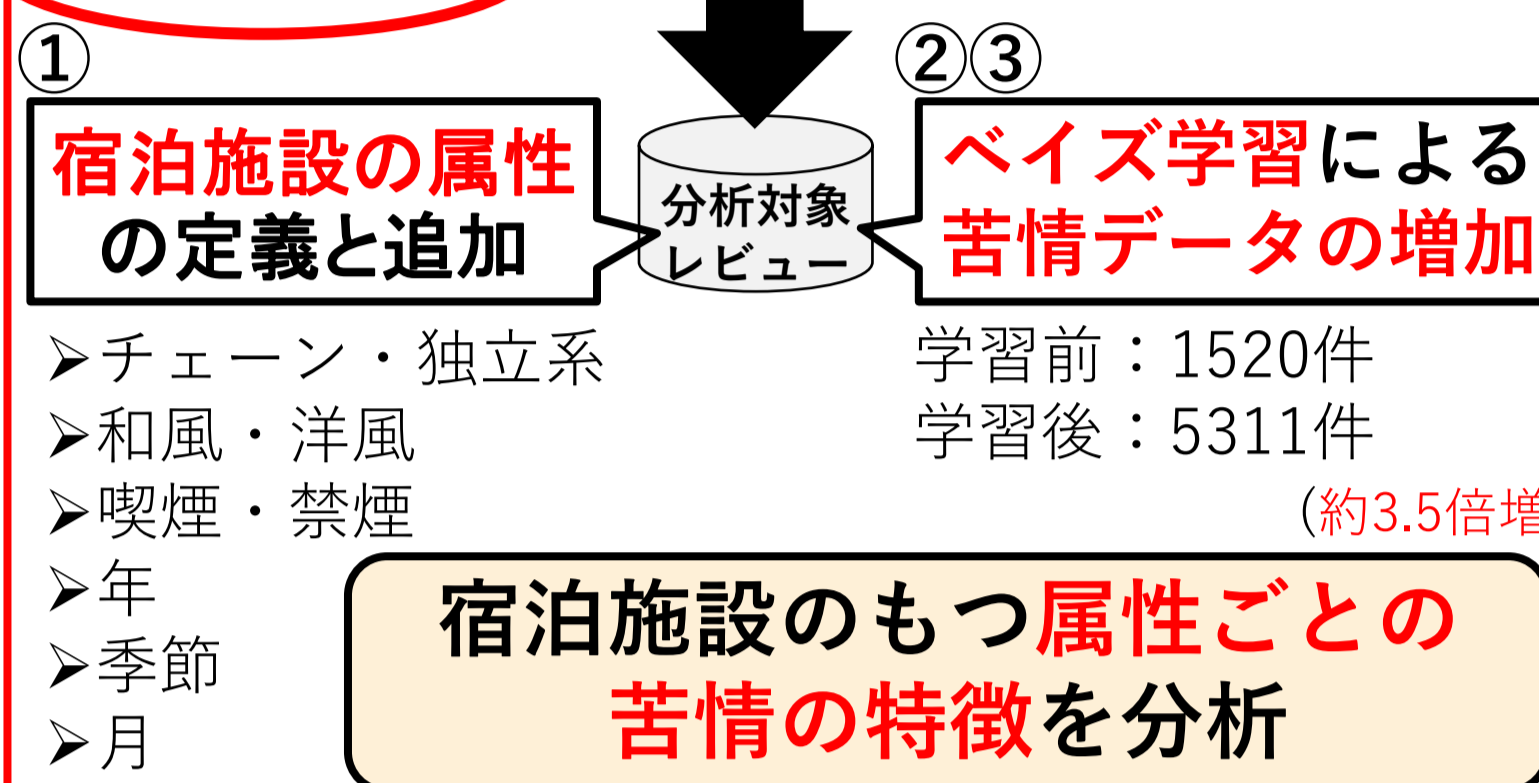
### 本研究で利用する提供されたデータ

楽天トラベル-ユーザーレビュー  
(「感想・情報」,「苦情」カテゴリ)  
・3,202,117レコード, 10フィールド  
・2005年~2015年に集計されたデータ

### 問題点

- ① 宿泊施設の属性データがあまりない
- ② 提供されたデータに「苦情」データが僅少
- ③ 「感想・情報」に明らかな苦情内容が混在

### 取り組み



### KH-Coderを用いた統計分析

- レビュー本文の形態素解析
- 特徴語(=頻出語)の抽出
- コーディング・ルール作成
- コーディング・ルールを適用し  
分割表による独立性の検定
- コレスポンデンス分析  
(数量化III類=主成分分析)
- 結果の考察

経営戦略に反映可能な分析結果を提示

## 分析のための前処理と分割表による独立性の検定

### 分析のためのデータの再構成

- 全施設からレビュー数の多い20施設を選定
- 個々の施設に属性(=外部変数)を追加  
「和風・洋風」,「チェーン・独立系」,  
「喫煙・禁煙」,「月」,「季節」,「年」
- 16,866レコード,16フィールドにデータを再構成

### 問題点

- 「苦情」レビューは1520件と僅少
- 「感想・情報」レビュー中に苦情内容が混在

### ナイーブベイズ分類

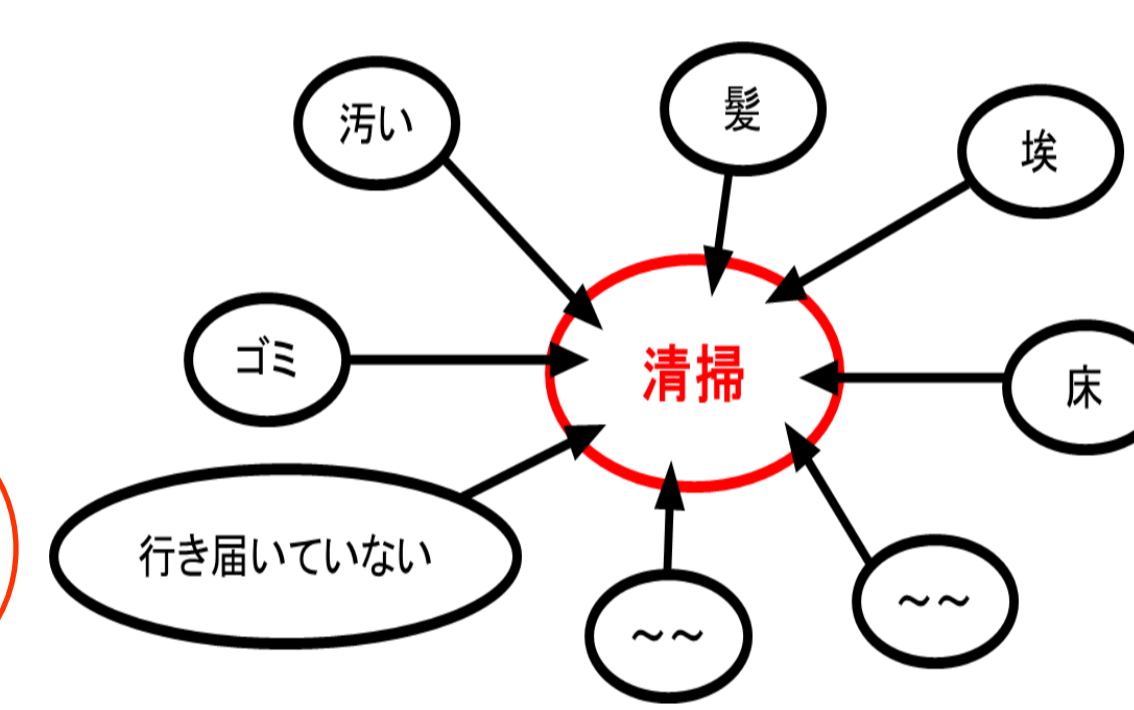
- 教師データ:任意の400件のレビュー
- 「感想・情報」,「苦情」の全レビューを学習  
分類結果→「満足」,「不満足」と呼ぶ

分析対象:「不満足」レビューに再設定  
「苦情」:1520件 → 「不満足」:5311件  
(約3.5倍増)

5311件の「不満足」レビューを利用し  
宿泊施設のもつ属性ごとの特徴を分析

### コーディング・ルール

同じような概念を持つ抽出語  
をひとつのキーワードに集約  
論理演算子を用いて記述可能



### 分割表による独立性の検定

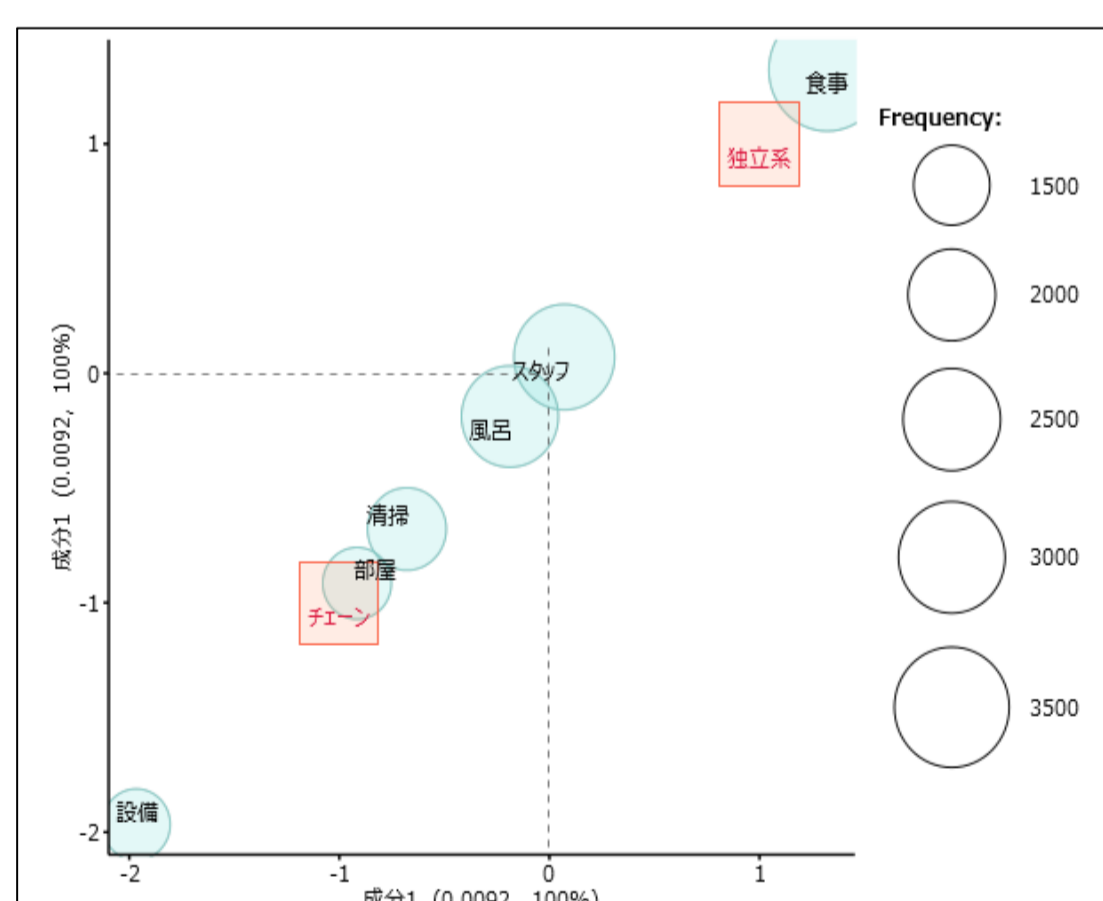
帰無仮説:各属性で宿泊施設間の不満に差がない

#### 「チェーン・独立系」の違いによる全年度での不満の特徴分析

	*スタッフ	*風呂	*部屋	*食事	*設備	*清掃	ケース数
チェーン	1325(8.17%)	1253(7.72%)	654(4.03%)	1579(9.73%)	701(4.32%)	855(5.27%)	16226
独立系	1349(8.19%)	1214(7.37%)	551(3.34%)	2046(12.42%)	481(2.92%)	754(4.58%)	16473
合計	2674(8.18%)	2467(7.54%)	1205(3.69%)	3625(11.09%)	1182(3.61%)	1609(4.92%)	32699

カイ2乗値: 0.003, 1.406, 10.636\*\*, 59.692\*\*, 45.603\*\*, 8.223\*\*

部屋, 食事, 設備, 清掃に対する不満は  
チェーン・独立系間に統計的有意差が認められる( $p \leq 0.01$ )



- チェーンよりも独立系のほうが**食事**に対する不満が相対的に多い
- 独立系よりもチェーンのほうが**部屋, 設備, 清掃**に対する不満が相対的に多い

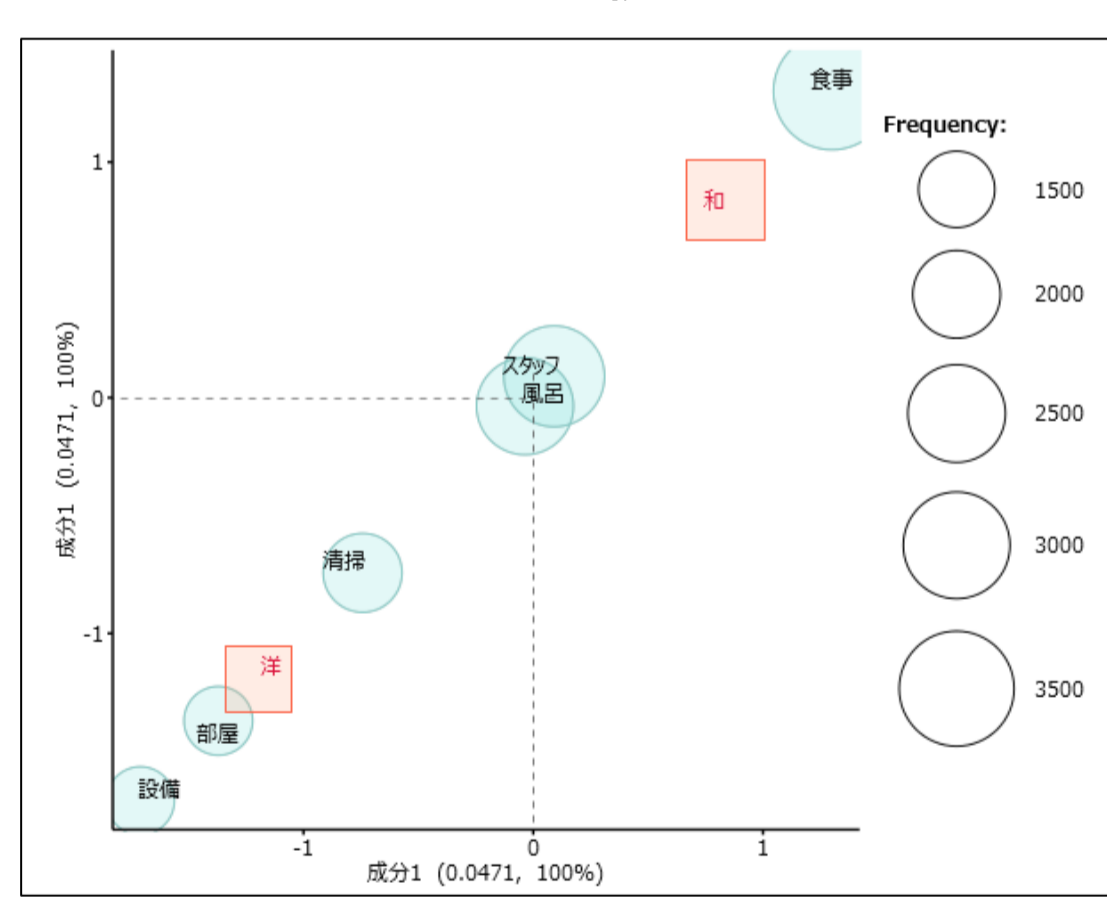
- 独立系施設(リゾートホテルなど)  
食事に対する不満が多い
- チェーンの施設(ビジネスホテルなど)  
設備に対する不満が多い

#### 「和風・洋風」の違いによる全年度での不満の特徴分析

	*スタッフ	*風呂	*部屋	*食事	*設備	*清掃	ケース数
和	1596(8.23%)	1439(7.42%)	531(2.74%)	2632(13.57%)	478(2.47%)	817(4.21%)	19389
洋	1078(8.10%)	1028(7.72%)	674(5.06%)	993(7.46%)	704(5.29%)	792(5.95%)	13310
合計	2674(8.18%)	2467(7.54%)	1205(3.69%)	3625(11.09%)	1182(3.61%)	1609(4.92%)	32699

カイ2乗値: 0.167, 0.988, 119.564\*\*, 298.694\*\*, 179.832\*\*, 50.508\*\*

部屋, 食事, 設備, 清掃に対する不満は  
和・洋間に統計的有意差が認められる( $p \leq 0.01$ )



- 和風の施設のほうが**洋風の施設**よりも**食事**に対する不満が相対的に多い
- 洋風の施設のほうが**和風の施設**よりも**部屋, 設備, 清掃**に対する不満が相対的に多い

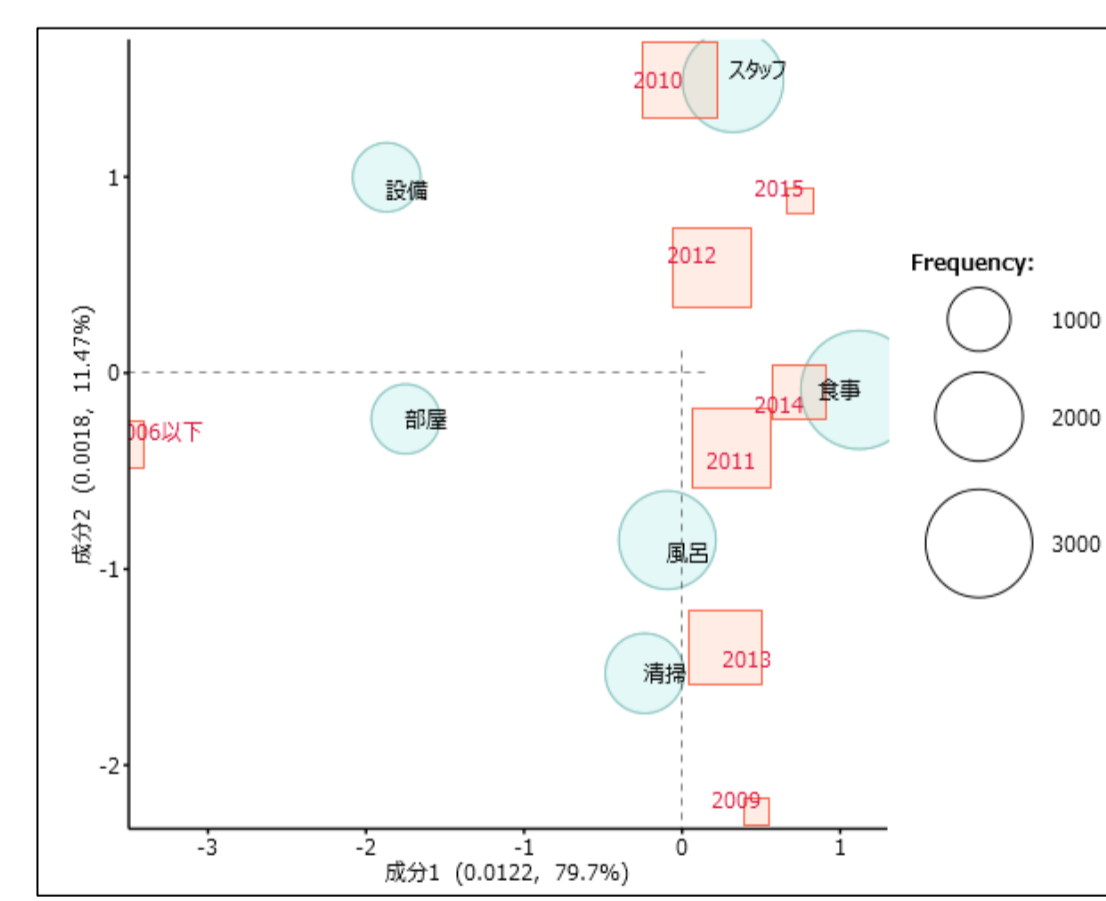
- 和風の施設(旅館など)  
食事に対する不満が多い
- 洋風の施設(ビジネスホテルやリゾートホテルなど)  
部屋・設備に対する不満が多い

#### 全年度での不満の時系列的な特徴分析

	*スタッフ	*風呂	*部屋	*食事	*設備	*清掃	ケース数
~2006	160(6.97%)	184(8.01%)	143(6.23%)	144(6.27%)	144(6.27%)	126(5.49%)	2296
2009	49(6.91%)	61(8.60%)	29(4.09%)	78(11.00%)	14(1.97%)	30(4.23%)	709
2010	540(8.66%)	434(6.96%)	226(3.62%)	686(11.00%)	240(3.85%)	269(4.31%)	6238
2011	523(7.45%)	488(6.95%)	239(3.41%)	778(11.09%)	220(3.13%)	327(4.66%)	7018
2012	627(9.13%)	533(7.76%)	248(3.61%)	769(11.19%)	236(3.44%)	345(5.02%)	6870
2013	452(7.81%)	469(8.10%)	205(3.54%)	697(12.04%)	201(3.47%)	329(5.68%)	5788
2014	265(8.61%)	243(7.89%)	90(2.92%)	374(12.15%)	100(3.25%)	156(5.07%)	3078
2015	58(8.26%)	55(7.83%)	25(3.56%)	99(14.10%)	27(3.85%)	27(3.85%)	702
合計	2674(8.18%)	2467(7.54%)	1205(3.69%)	3625(11.09%)	1182(3.61%)	1609(4.92%)	32699

カイ2乗値: 22.864\*\*, 12.127, 49.277\*\*, 69.511\*\*, 59.868\*\*, 17.495\*

スタッフ, 部屋, 食事, 設備 ( $p \leq 0.01$ ), 清掃 ( $p \leq 0.05$ ) に対する不満は  
各年度間に統計的有意差が認められる



- ケース数をみると,  
2011年~2015年にかけて**レビュー数**が減少している
- 食事に対する不満は次第に増加している
- 部屋に対する不満は次第に減少している

- レビューの件数が年々減少
- 食事に対する不満は年々増加
- 部屋に対する不満は年々減少

## 結果と考察

- 和風の施設は**食事**に対する不満が多い  
特に独立系の施設で多い→改善が急務
- 洋風の施設では**部屋, 設備, 清掃**に対する不満が多い  
特にチェーンの施設で多い→改善が急務
- 全体的に**食事**に対する不満は年々増加, **部屋**に対する不満は減少  
原材料の高騰による食事の品質低下? 部屋は快適さを求める工夫の実践か?
- レビュー数**が年々減少  
投稿者に対する報酬やコンスタントにユーザの評価を得る仕組み作りが急務

これらの結果を宿泊施設に提示

宿泊施設自身が改善, 経営戦略にフィードバックするかどうか判断

## 今後の課題

- コーディング・ルールの妥当性評価  
どのようなルールが正確なのか一切不明
  - 緻密にして網羅性を高める  
→元のレビューと同等
  - カバレッジを広くする  
→ノイズが多くなる
- 宿泊施設における**苦情の地域差**分析
  - 八地方区分における苦情分析
  - 地方ごとの苦情の特性を発掘