

# レビューデータを用いたネットユーザの推奨行動に関する予測モデルの構築研究

針尾大嗣 撰南大学経営学部経営情報学科 准教授

## 1 研究の目的

インターネット上の口コミは、現代消費者が商品・サービスの購買意思決定を行う上で最も参考とする情報である。消費者は、口コミ内の様々なレビューを探索、比較、評価し、購買に至る。その選択過程においては、不確実性下における人間の意思決定モデルとして知られるプロスペクト理論で説明できるように口コミ内のネガティブなレビューを重視する傾向が強くみられる (Ren・Zhao・Zhang 2014)。事業者がいかに優れた商品・サービスを提供し、多数の購買経験顧客によりポジティブなレビューが投稿されたとしても、消費者は問題指摘された少数のネガティブなレビューをリスクとして認知し、その損失回避のための不買の意思決定を下す。購買経験顧客により発信される情報がソーシャルメディア上で一瞬にして他の消費者の間に伝搬する現在、これまでリテンション（既存顧客維持）のためのマーケティング施策として取り組まれてきた“顧客の声”いわゆる評判への対応は、アキュイジション（新規顧客獲得）に直結する施策課題となってきた。このインターネット上の口コミへの評判対応は、オンライン・レピュテーション・マネジメント (ORM: Online Reputation Management) と呼ばれており、その施策は、事業者側が積極的な情報発信を行い自社のポジティブな評判を形成していく広報・PR型と誹謗中傷、風評、悪評、クレームなど消費者により発信されたネガティブな情報への対応を行うリスク管理型に大別することができる。近年、消費者の情報探索活動の始点が、事業者側が一次情報を直接発信し管理する自社のウェブサイトや広告サイトではなく、消費者が主体となり情報の選択および発信、共有を行うソーシャルメディアとなりつつある現状を鑑みると、今後、リスク管理型 ORM の重要性が増していくことであろう。

インターネットが消費者の購買行動に強く影響を与えている分野のひとつとして、旅行・観光サービス分野がある。旅行者が旅行関連会社および施設について知る機会は、全体の8割がインターネットであり、およそ6割が旅行前後に旅行会社、航空会社、宿泊施設に関して他者に何らかの情報を伝達していることがわかっている (ADK2013)。また日本政府観光局の調査によると、訪日外国人旅行者が宿泊施設を選定する際には、全体の約6割以上が口コミを扱う宿泊予約サイトを利用すると報告されている (JNT02014)。このように旅行・観光事業者、特に宿泊施設事業者にとって、今やインターネット上の口コミが新規顧客との主要な顧客接点となっていることは明らかである。また我が国では、2013年に閣議決定した観光立国推進基本計画に基づいて、国内旅行消費額と訪日外国人旅行者の拡大、旅行者満足度の向上を基本目標に2020年東京オリンピックに向けて急ピッチでその整備・拡充を進めているが、このなかで観光基本法を改定した観光立国推進基本法の第二十一条においては、情報通信技術を活用した観光に関する情報提供等に必要の施策を講ずるとしている。こうした動向のなかで、今後、旅行・観光事業によるリスク管理型 ORM への関心も高まっていくものと思われる。

そこで研究代表者は、2013年から旅行・観光事業のうち宿泊施設サービス事業を対象としたリスク管理型 ORM に用いる口コミデータの分析アルゴリズム研究を進めてきた (逢坂・針尾 2014)。この度、研究助成を受け、口コミサイトに投稿された宿泊施設サービスの利用経験顧客のレビューを構成するワード情報と顧客態度との関連性を調査した。本報告書では、宿泊施設サービスの利用継続および他者への推奨意思を示した推奨態度と非推奨態度をレビュー中のキーとなるワード（特徴語）を用いて判別する予測モデルと顧客態度への影響要因の分析結果について報告する。

## 2 旅行・観光系国内口コミサイトに関する事前調査

国内大手旅行・観光系の口コミサイト4社についての調査を行った。表1は、各社サイトの登録宿泊施設

数、レビュー投稿数、評価指標、収益モデル等をまとめたものである。

宿泊施設の登録件数順位では、R社が3万件弱とトップ、次いで、J社となり、I社は、ハイクラスの宿・施設に特化していることから他の3サイトと比べると少ない。レビューアが行う宿泊施設の評価は、各社が設定した指標（4社いずれも6つの指標を設定）についてレビューアが得点を付ける数値評価と自由記述による定性評価により行われている。登録宿泊施設に関する情報は、エリア検索、評価得点を用いたソート検索により表示される仕組みとなっている。レビューアにより投稿されるレビュー投稿量については、全体としてR社が他3社よりも多い。4社の収益モデルについては、Y社が広告料による広告収益モデル、他3社が登録事業者との取引手数料によるコミッションモデルを採用、レビューアを含むサイト利用者への課金は行っていない。また、4社はいずれもサイト利用会員へのポイント還元を行っているが、これについてはポイント決済カードの普及率が高いR社、Y社の優位性が高いと考えられる。

表 1. 旅行・観光系口コミサイトの比較

観光系口コミサイト	I社	J社	R社	Y社
サービス開始月	2000年5月	2000年11月	1996年1月	確認できず
登録宿泊施設数	1,621件	2万4,336件	2万9,839件	非公開
レビュー投稿数 例: ホテルグランヴィア京都	569件	214件	2,804件	90件
評価指標	6指標(客室・アメニティ/ 接客サービス/バス・お 風呂/施設・設備/お食事 /満足度)	6指標(部屋/風呂/朝食/ 夕食/接客・サービス/清 潔感)	6指標(サービス/立地/ 部屋/施設・アメニティ/ 風呂/食事)	6指標(料金/部屋/設備/ サービス/食事/風呂)
事業者(情報掲載側)システ ム登録料率	10%	10%	7%	0%
収益モデル	コミッション型	コミッション型	コミッション型	広告収益型
会員登録	必要なし	必要なし	必要なし	必要あり
会員ポイント還元率	2%	2%	1%	5~14%
サイトの対応言語	日本語	日本語、英語、韓国 語、中国語	日本語、英語、韓国 語、中国語	日本語
スマートフォン・アプリ	対応	対応	対応	対応

データは2015年8月末時点

旅行・観光口コミサイトのビジネスモデル及び旅行者の口コミ行動に関する調査(針尾・井野戸2015)より

・I社

高級旅館、施設に特化した旅行予約サイト。1999年に宿泊と飲食に特化したオークションサービスサイトとしてスタートした。翌年、国内の高級旅館・施設に特化した宿泊予約サービスを開始。

・J社

某社の旅行部門が格安航空券と海外旅行情報誌の創刊を1984年に開始。1990年から国内旅行情報誌を創刊、2000年11月からインターネットを使った宿泊施設の予約をはじめ。国内・海外の観光地情報など総合的な観光情報サービスの提供も行っている。

・R社

インターネット大手物販サイトを運営する某社の旅行予約サイト。同社が2001年春に旅行事業を立ち上げ、翌年に分社化し設立した。2003年に「旅の窓口」を運営するマイトリップ・ネット株式会社の株式の100%を取得し、2004年に合併し、システム統合を行った。2005年8月には第1種旅行業者となり総合旅行会社として国内・海外のパッケージ旅行商品の取り扱いが可能となった。物販サイト会員約9,700万人、同サイトへの登録施設数は2万9,839施設の国内最大規模の旅行予約サイトである。

・Y社

国内大手ポータルサイトが、2000年に旅行業界大手 JTB と提携し合併会社「たびゲーター」を設立。2015年にそれまで有料であったサービスを無料化、同サイトを介して予約した際、宿泊料金の一部をTポイントに還元するサービスを開始した。2015年末、旅行予約サイトI社の買収を発表。

### 3 顧客態度予測モデルの構築手続き

#### 3.1 サンプル

予測モデル構築に用いるレビューデータの収集は、事前調査の結果をふまえて、宿泊施設登録件数、レビューア数、レビュー投稿量が最も多いR社運営の旅行予約サイトから行った。同サイトで情報提供されている宿泊施設の中から、客室数、レビューの投稿量と投稿頻度を参考とし12施設選んだ。調査対象期間は、2013年4月1日から2015年7月31日とし、この間に投稿されたレビュー（総数2,345件）を収集した。

#### 3.2 予測モデルと変数

予測モデルは、レビューアが旅行・観光分野の口コミサイトに投稿したレビュー中の自由記述文書で表明した顧客態度を分類したカテゴリーデータを従属変数、文書内での発言に使用されたキーとなる特徴語および特徴語グループ（ファセット）の出現を2値化したカテゴリーデータおよび施設属性を独立変数とする累積ロジットモデルである。

##### (1) 顧客態度（従属変数）

宿泊施設サービスに対するレビューアの顧客態度を従属変数とする（図1）。自由記述で投稿されたレビューアのレビュー内容を読解し、次回以降の利用を表明する、もしくは他者に当該施設の利用を薦める「推奨態度」の優良顧客、明確に次回以降の利用を行わないと表明する、もしくは他者に当該施設の利用を薦めない、ないしは他の施設の利用を薦める「非推奨態度」の離脱顧客、そのいずれにもあてはまらない、たとえば、サービスに関しての感想のみ言及している「中立態度」については次回利用が定まらない浮動顧客として、3分類する。

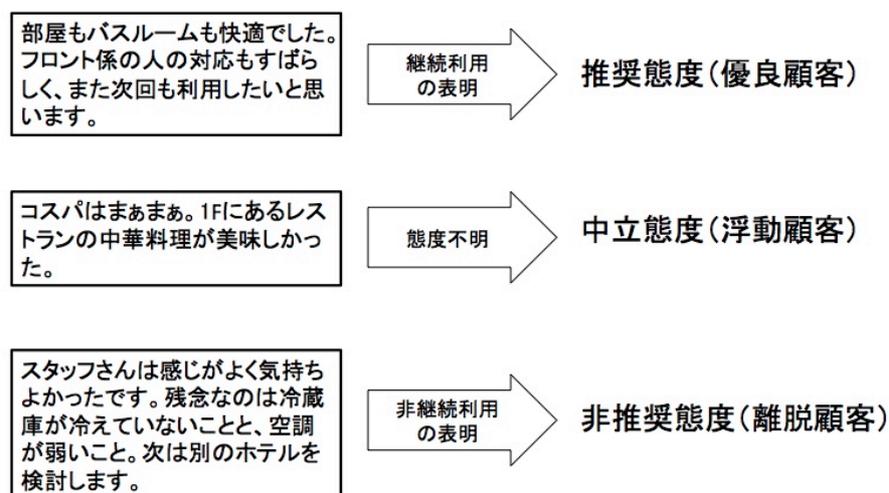


図1: 自由記述の読解判別による顧客態度の分類

##### (2) 特徴語（独立変数）

当該施設のレビュー評価の内容傾向を表す特徴語を独立変数とする。特徴語は、レビューアが投稿したレビュー文書データを形態素解析し抽出する。一般的に意味のある最小の文字列単位である単語を形態素

(morpheme) と呼び、形態素解析は、文章中の形態素の取り出しと品詞情報などを付け加える作業のことである。この形態素解析により抽出した形態素の出現頻度、順位および共起パターンをもとに特徴語を明らかにする。

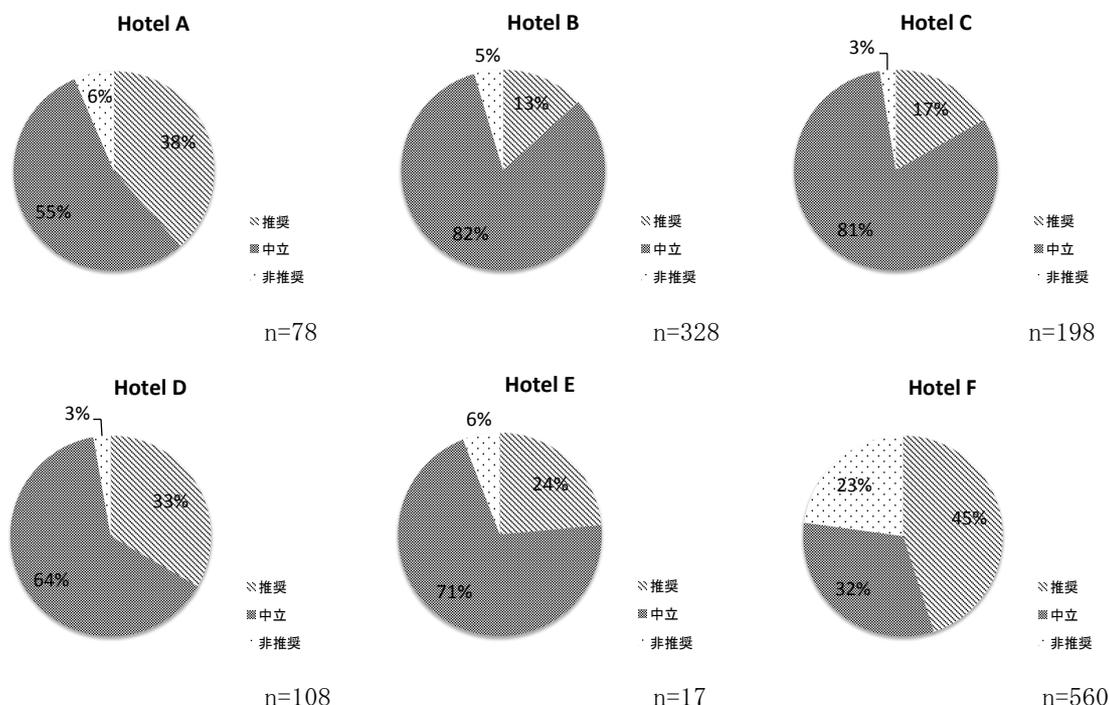
## 4 結果

### 4.1 顧客態度比

図2は、12施設についての調査期間中にレビュー投稿を行っているレビューアの顧客態度比を表したものである。総レビューア数は、のべ2,345人(n=2,345)あり、全体の顧客態度比は、推奨態度(優良顧客)30%、中立態度(浮動顧客)64%、非推奨態度(離脱顧客)6%であった。中立態度の比率が他2つの顧客態度より高い理由は、レビュー内容が、施設サービスについての単純な言及、お礼となっているものが多く、それらを中立態度として分類した結果であると言える。

最もレビューア数が多い施設は、のべ560人(n=560)の施設F、次いで387人(n=387)の施設Lであった。両施設共に宿泊利用者の多い主要駅に隣接する大型観光施設である。レビューアの顧客態度比は、施設Fが、推奨45%(n=252)、中立32%(n=180)、非推奨23%(n=168)、施設Lが、推奨34%(n=132)、中立60%(n=232)、非推奨6%(n=23)となっている。いずれも推奨態度を示す優良顧客の比率が全施設平均30%よりも高いが、施設Fについては、非推奨態度を示す離脱顧客が全国平均よりも17%高く存在している。施設Fは優良顧客、離脱顧客の割合が他の12施設より明らかに高く、宿泊利用客によるインターネット上での評価が明確に分かれる施設であると言える。一方、レビューア数が少ない施設は、17人(n=17)の施設E、次いで29人(n=29)の施設Hであった。両施設共に郊外型の小規模型ビジネス施設である。レビューアの顧客態度比は、施設Eが、推奨24%、中立71%、非推奨6%、施設Hが、推奨17%、中立79%、非推奨3%となっている。いずれも優良顧客の比率が全施設平均30%より低い。

本研究では、3つの顧客態度を示す顧客が一定比率存在し、形態素解析を行ううえで必要なレビューデータ量が得られる施設Fを調査対象として選択した。



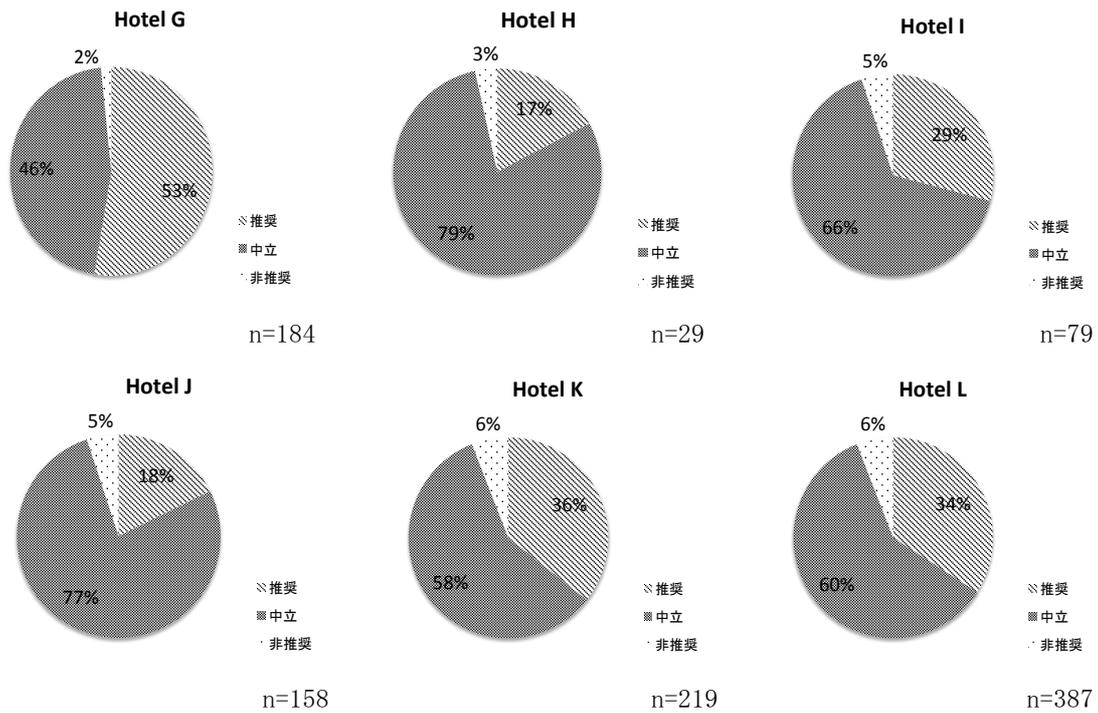


図 2：12施設についてのレビュー内顧客態度比

## 4.2 特徴語

### (1) 語の出現頻度・順位

施設 F(n=560)に関するレビューで使用された語の頻度を集計した。このうち助詞、助動詞、記号等の不要語を除いた語の出現頻度を縦軸、順位を横軸としたグラフを図 3 a、それぞれの対数をとったグラフを図 3 b に示す。単語の使用頻度と順位との間にジップの法則が成り立っていることが確認できる（金 2009）。

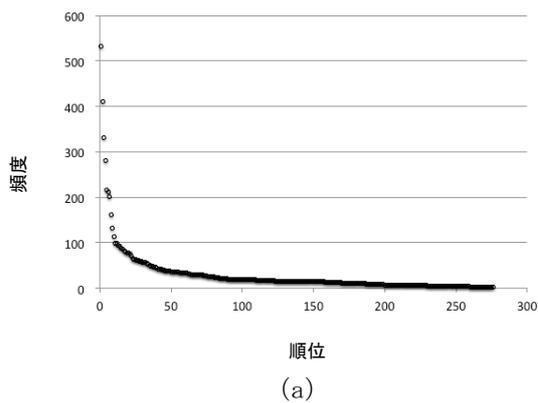


図 3 a：順位と頻度の散布図

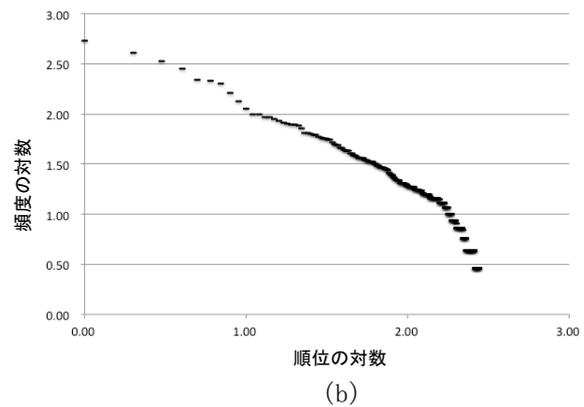


図 3 b：順位と頻度の対数値の散布図

次に、顧客態度別に使用された語の頻度を集計した。優良顧客と離脱顧客が使用する語の出現頻度をそれぞれのレビュー数で割った相対頻度を縦軸、順位（上位 30 まで）を横軸としたグラフを、それぞれ図 4 a、図 4 b に示す。

推奨態度を示す優良顧客のレビューでは、5 件に 1 件（相対度数 0.20 以上）の割合で、「利用」「ホテル」

「部屋（ルーム）」「思う」「また」「京都」「宿泊」「良い」「広い」「朝食」の10語が含まれている。また全体の半数（相対度数0.50以上）に「利用」「ホテル」「部屋（ルーム）」「思う」「また」の5語が含まれている。同様に非推奨態度を示す離脱顧客のレビューでは、5件に1件の割合で含まれている語は「ホテル」「食事（料理、ご飯）」「部屋（ルーム）」の3語のみであり、優良顧客のように全体の半数のレビューで使用されている語は確認できない。

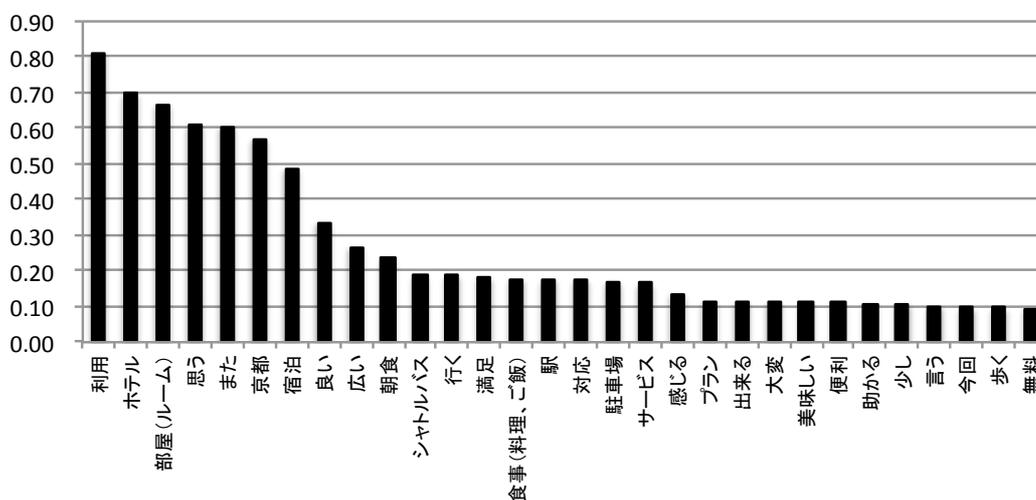


図4a：優良顧客（推奨態度）の使用語上位30

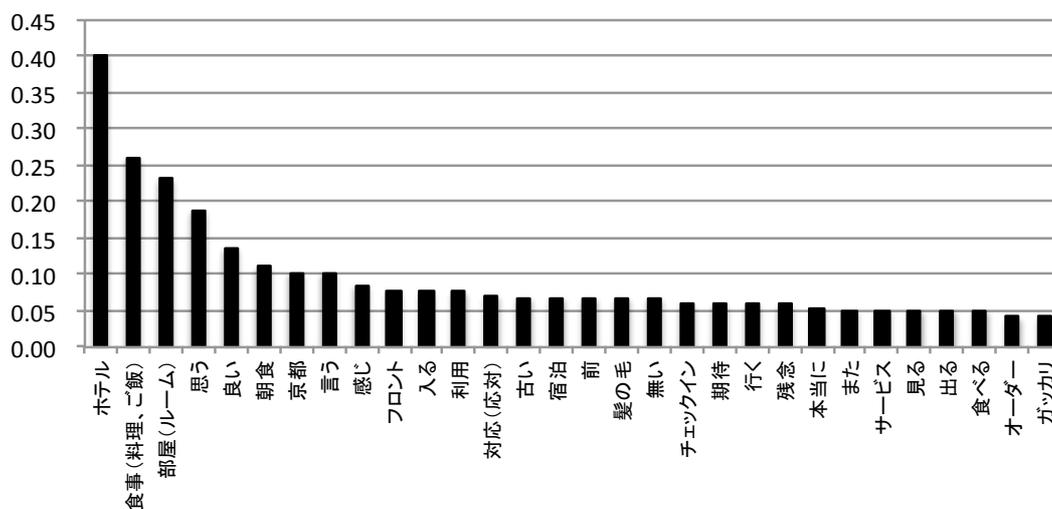


図4b：離脱顧客（非推奨態度）の使用語上位30

## (2) 共起パターン

レビューによる語の使用法および内容を把握することを目的に語と語の共起パターンを明らかにする。そこでレビューデータを形態素解析し、助詞、助動詞および記号等の不要語を除いた上位100語のうち共起頻度が2以上の語について共起ネットワーク分析を行った。優良顧客と離脱顧客がレビュー中で使用する語の共起パターンを表したネットワークマップを図5a、図5bに示す。

優良顧客が使用する語の共起ネットワークマップでは、「対応」と「良さ」がハブとなり、部屋（広さ・利用）、食事（朝食・バイキング）、接客（フロント・案内）の3つのクラスターが結びついていることがわかる。図4aの傾向も踏まえると、F施設の推奨利用を表明する優良顧客は、同施設の部屋、食事、接客の3つのサービスの対応が良いと明確に言及していることがわかる。なお半数は、部屋の良さについて言及してい



(3) 特徴語ファセット

レビューアが施設評価を行う際に用いるレビュー中の語の出現頻度、共起パターン、TF-IDF による語の重み値を参考とし、特徴語のファセットグループを作成した。なお推奨態度と非推奨態度においてそれぞれ肯定的／否定的な言及の際に用いられることが明確である同一カテゴリーの特徴語は、あらかじめ異なるファセットに分類した。

表 2. 特徴語ファセットの一例

特徴語ファセット		対象となる特徴語(一部)
Staff servicel (SS1)	接客 I	フロント、デスク、窓口、従業員、スタッフ
Staff servicell (SS2)	接客 II	対応、動作、反応、受付、チェックイン、チェックアウト
Location/Access (LA1)	立地アクセス I	駅近、徒歩、バス、電車
Location/Access(LA2)	立地アクセス II	自動車、車
RoomI (RM1)	部屋 I	部屋、景色
RoomII (RM2)	部屋 II	ベッド、シーツ
Facility&AmenityI (FA1)	設備 I	シャワー、お風呂、廊下
Facility&AmenityII (FA2)	設備 II	エレベーター、階段、冷蔵庫、空調、クーラー、暖房、テレビ、トイレ、洗面台、洗い場、バスルーム
Meal (ML)	食事	食事、朝食、夕食、ご飯、バイキング
Noisel (NS1)	音 I	人の声、子供の声、隣の部屋、廊下の音、壁
NoiseII(NS2)	音 II	がやがや、ガチャガチャ、ガタガタ
Cleaness (CL)	清掃	髪の毛、髪、よごれ、ゴミ、ホコリ
EvaluationI (EV1)	評価 I	快適、良い、素晴らしい、便利、美味しい、うまい、嬉しい、満足、キレイ、広い、はじめて
EvaluationII (EV2)	評価 II	不快、悪い、最悪、ダメ、嫌、不便、不味い、残念、汚い、狭い、今までに
Area(AR)	地域 I	京都、繁華街
Adverbl (AV1)	状態 I	また、次回、かなり、なかなか、とても、非常に、必ず、近い、一番
AdverblI(AV2)	状態 II	絶対に、遠い
Behave(BH)	行動 I	泊まる、宿泊する、来る、過ごす、預ける、食べる、出来る
.		
.		
.		

4.3 顧客態度の予測モデルと影響要因

レビューアの顧客態度を予測するため、顧客態度（3＝推奨態度、2＝中立態度、1＝非推奨態度）を従属変数、特徴語ファセット（1＝有り、0＝無し）を独立変数とした回帰分析を実施した。モデルによる判別の対象が多群（3群以上）の際、従属変数は多値データであり、従属変数を3つの顧客態度に方向性を与えた順序尺度として累積ロジスティック回帰モデルを適用した。

累積ロジスティック回帰分析の結果、回帰式は有意（P<.001）であり、寄与率は、Cox と Snell 44.1%、Nagelkerke 58.3%、McFadden 32.0%であった（表 3）。有意な独立変数、すなわちレビューアの顧客態度に影響を与える要因ファセットは、接客 II（SS2）、立地アクセス I（LA1）、清潔（CL）、状態 I（AV1）である（表 4）。

表 3. 回帰式の適合性

モデル適合情報					擬似 R2 乗	
モデル	-2対数尤度	カイ 2乗	自由度	有意確率	Cox と Snell	
切片のみ	96.214				.441	
最終	50.802	45.558	3	.000	.583	.320

リンク関数: ロジット

適合度			
	カイ 2乗	自由度	有意確率
Pearson	26.508	31	.623
逸脱	29.772	31	.542

表4. レビューアの顧客態度に影響を与える要因ファセット

パラメータ推定値

		B	標準誤差	Wald	自由度	有意確率
しきい値	[顧客態度=1]	6.443	1.834	10.870	1	.001
	[顧客態度=2]	9.508	2.150	10.870	1	.000
位置	SS2	2.324	.382	7.062	1	.008
	LA1	-1.562	.334	8.594	1	.041
	CL	0.334	.134	14.293	1	.000
	AV1	0.832	.466	11.205	1	.042

リンク関数:ロジット

以上から顧客態度の予測式として、推奨態度（優良顧客）式1、非推奨態度（離脱顧客）式2が得られた。

$$\logit\{\Pr(1)\} = 6.443 - (2.324x_1 - 1.562x_2 + 0.334x_3 + 0.832x_4)$$

$$\Pr(1) = 1 / \{1 + \text{Exp}(-6.443 + 2.324x_1 - 1.562x_2 + 0.334x_3 + 0.832x_4)\} \quad \dots \text{式1}$$

$$\logit\{\Pr(1or2)\} = 9.508 - (2.324x_1 - 1.562x_2 + 0.334x_3 + 0.832x_4)$$

$$\Pr(2) = \Pr(1or2) - \Pr(1)$$

$$\Pr(3) = 1 - \Pr(1or2) \quad \dots \text{式2}$$

$$x_1 = \text{SS2}, x_2 = \text{LA2}, x_3 = \text{CL}, x_4 = \text{AV1}$$

## 5 まとめ

本研究では、旅行・観光系口コミサイトのレビュー中で用いられているワード情報を用いて、レビューアによる施設利用の推奨・推奨と非推奨・非推奨の顧客態度を判別する予測モデルを構築し、宿泊施設利用顧客の顧客態度に影響を与える要因を明らかにした。今後は、施設サービスについての口コミ評価に使用されるワード（特徴語）の整理、重要度の測定、形態素解析に用いる辞書作成を行い、モデル改良と予測の検証を進める。

## 【参考文献】

1. Ren・Zhao・Zhang2014; J.Ren,D.M.Zhao,X.H.Zhang,“Online Buying Decision Preference and Prospect Theory”,Advanced Materials Research, vols.926-930,pp.3850-3853.2014
2. ADK2013 ; 株式会社アサツーディ・ケイ,「旅行クチコミ調査」, [https://www.adk.jp/wp/wp-content/uploads/news/20120820\\_1.pdf](https://www.adk.jp/wp/wp-content/uploads/news/20120820_1.pdf), 2013年8月20日
3. JNTO2014; 日本政府観光局事業連携推進部観光情報戦略室,「TIC 利用外国人旅行者調査報告書」, [https://www.jnto.go.jp/jpn/reference/tourism\\_data/pdf/tic\\_report\\_2012.pdf](https://www.jnto.go.jp/jpn/reference/tourism_data/pdf/tic_report_2012.pdf), 2014年3月
4. 逢坂・針尾 2014;「宿泊客のリピート意向に影響を及ぼす宿泊施設サービスに関する一考察」, 情報コミュニケーション学会第13回研究会論文集, pp.8-9、2014年7月
5. 金 2009; 金明哲,「テキストデータの統計科学入門」, pp.52-53、岩波書店、2009年

〈 発 表 資 料 〉

題 名	掲載誌・学会名等	発表年月
Finding the man who is profitable or not : An algorithm to predict online rating behavior	Social Media society 16 予稿集	2016年7月 (予定)