

テキストマイニングを用いた宿泊レビューからの注目情報抽出方法

辻井 康一（株式会社日本旅行） 津田 和彦（筑波大学大学院）

概要 宿泊検討者がインターネットより宿泊予約を行う際、宿泊利用者が投稿した宿泊レビューが重要視されている。宿泊レビューは、主に宿泊利用者のプロフィール、数項目の数値評価および感想コメントの3要素からなる。多くの宿泊検討者は、容易に情報を取得できる数値評価を重要視するが、数値評価のみでは当該宿泊施設の情報を十分に把握できるとは言い難い。そこで本論文では、テキストマイニングを用い感想コメントを分析することで、宿泊検討者にとって有意義な情報を抽出する手法を提案する。さらに、提案手法を用いて感想コメントを分析した結果、ビジネスエリアとレジャーエリアでは同じ数値評価の項目でも全く異なる指標で採点されていること、エリア間で評価の差異があること、数値評価の項目にない事項が重要視されていることなどを導くことができた。

1. はじめに

インターネットの普及に伴い、宿泊予約の主流は旅行代理店での予約からネット予約へと移行している。宿泊検討者は、旅行代理店での予約においては、専門家と相談しつつ意図する宿を探していたが、ネット予約ではそのようなサービスを楽しむことは困難である[1,2]。そのため、宿泊検討者の判断材料として宿泊利用者から投稿される施設評価（以下、宿泊レビューと示す）が重要視されている[3,4]。

宿泊レビューは、大半の宿泊予約サイトで提供されており、主に宿泊利用者のプロフィール、数項目の数値評価および感想コメントの3要素から構成されている。宿泊レビューの重要性は、宿泊施設、宿泊予約サイト運営側でも認識しており[5]、宿泊レビューを集めるために宿泊費用の一部として使用できるポイントを付与している宿泊予約サイトも存在する。

けれども、宿泊レビューの表示方法は、固定的な表示形式であるため、各々が異なる利用目的や価値観を持っている宿泊検討者の欲する情報を、効率よく取得するのは困難である[6]。さらには、数項目の数値評価を分析してみると、多くの宿泊利用者は「オール4」などと全ての項目で同じ点数を付けていたり、1項目のみ点数が異なったりするなど、各項目を十分に評価していない傾向がみられる。

宿泊施設の評価方法に関する研究として、大内ら[7]は、宿泊施設によるアンケート回答形式での宿泊レビュー手法の研究を行い、宿泊検討者への情報提供方法を検討している。森川[8]は、宿泊者へアンケートを行い、レジャー目的の宿泊利用者の満足度に影響のある項目について

検討している。また田邊ら[9]は、宿泊レビューの感想コメントに対してテキストマイニングを行い、施設特有の特徴項目の分類を行い、宿泊施設側からの戦略検討への活用方法について提案している。

しかし、これらの研究成果では、宿泊レビューにより宿泊検討者へ、十分な情報を提供できるとは言い難い。

そこで本研究では、これらの研究結果を参考としつつ、宿泊レビューから宿泊検討者が施設を選択する際に、重要な判断材料となる情報を抽出する手法と、その情報の提示方法について提案する。具体的には、感想コメントに対して利用目的を考慮したテキストマイニングを行うことで、宿泊利用者が重要視した項目を抽出し、数項目の数値評価項目との関連性に注目した評価方法を提示する。

2. 数値評価の分析

本章では、宿泊レビューのうち数値評価について議論する。数値評価項目は、宿泊予約サイトごとに表示方法は異なるが、その項目に差異はなく、立地、スタッフの対応、快適性、価格などの項目と、それらの項目を総合的に各ユーザが評価した項目である総合評価で構成されている[10]。

2.1 数値評価項目

本節では、宿泊レビューの数値評価項目についてまとめる。数値評価とは、宿泊予約サイトが独自に決めた宿泊検討者にとって重要と思われる項目を列挙したもので、宿泊利用者がその項目に対して採点を行う形式で運営されている。

表1に国内の一部の宿泊予約サイトの数値評価項目

[11-15]をまとめたものを示す。宿泊予約サイトが独自に決めた数値評価項目には、表記方法や順序に差異はあるが、その内容を大まかに分類すると「施設」、「接客」、「立地」、「価格」、「食事」の5つから構成される。

表 1. 国内宿泊予約サイトの数値評価項目

評価項目	じゃらん net	楽天トラベル	るるぶ	一休.com	宿ぷらざ
施設	清潔感				清潔度
	部屋	部屋設備・アメニティ	部屋設備	客室・アメニティ	設備
				ひと休み度(満足度)	
	風呂	風呂	風呂	バスルーム・お風呂	お風呂
接客	接客・サービス	サービス	サービス	接客・サービス	サービス
立地		立地	立地		立地
価格					料金
食事	料理	食事	食事	お食事	お食事

表 1 に示す国内宿泊予約サイトと、Booking, Agoda, HotelClub, Expedia などの海外宿泊予約サイト[16-19]では、数値評価項目に差異がある。国内宿泊予約サイトには、施設における「風呂」や「食事」は必ず存在するが、海外宿泊予約サイトでは、ほぼ存在しない。この理由として、国内宿泊予約サイトではレジャー目的の温泉旅館などを考慮したことが考えられる。

しかし、ビジネスホテルの宿泊検討者で食事の良し悪しを重要視する人は多数存在するであろうか。少なくともビジネスホテルの感想コメントでは、食事に関するコメントは極めて少ない。また、同じ「立地」という数値評価項目に対する感想コメントでも、ビジネス利用においては駅からの距離や近隣にコンビニがあるかなど利便性が重要視されているが、レジャー利用において景観を楽しむ地域では部屋から見える眺望などが重要視されている。

このように、利用目的やエリアにより重要視する情報が異なるため、同じ数値評価項目でも評価する事項は多種多様となっている。

2.2 宿泊利用者の採点

本節では、宿泊利用者の数値評価採点について分析を行う。表 2 は、宿泊予約サイト「じゃらん net」[11]における数値評価の採点を集計した結果である。集計に用いたデータは、ビジネスエリア 5 地区（東京、大阪、名古屋、博多、札幌）、レジャーエリア 5 地区（箱根、湯布院、草津、登別、道後）の計 10 地区を選択、各地区で書

き込みの多い 10 の宿泊施設を選択、各宿泊施設の宿泊レビュー中の 2011 年 11 月時点で最新のものからそれぞれ 100 件を抽出したものである。

表 2. 数値評価における宿泊利用者の採点

	評価	総合	部屋	風呂	料理	接客	清潔感
分布頻度	1	109	143	105	85	170	112
	2	367	352	396	263	330	329
	3	1,121	1,725	1,922	935	1,521	1,485
	4	4,089	3,910	3,575	2,371	3,342	3,478
	5	4,314	3,870	3,891	3,700	4,637	4,596
分布割合	1	1%	1%	1%	1%	2%	1%
	2	4%	4%	4%	4%	3%	3%
	3	11%	17%	19%	13%	15%	15%
	4	41%	39%	36%	32%	33%	35%
	5	43%	39%	39%	50%	46%	46%
(件数)		10,000	10,000	9,889	7,354	10,000	10,000

表 2 から明らかなように、宿泊利用者の採点の大半は 4, 5 の評価であり、1, 2 という低評価は全体の 10% に満たない。このように偏りのある数値評価では当該宿泊施設の特徴を明確に表すのは難しい。

同データで、全ての数値評価項目を採点している 7,287 件について、その採点傾向を集計した結果を表 3 に示す。全項目で同一の採点が 18.7%、1 項目のみ異なる採点が 26.3%であった。ここから、宿泊利用者の採点の多くは、特に気になった項目以外は適当に 4 や 5 を付与していることが読み取れる。逆に言えば、1 項目のみ異なる採点をした項目は、宿泊利用者にとって強く印象を持った採点項目であると言える。

表 3. 数値評価における宿泊利用者の採点傾向

	件数	割合
全項目で同一の採点	1,361	18.7%
1 項目のみ異なる採点	1,913	26.3%
2 項目以上で異なる採点	4,013	55.0%
合計	7,287	100.0%

3. 感想コメントの分析

感想コメントは、宿泊利用者が実際に当該宿泊施設を利用して感じたことを文章記述したものである。それゆえ、数値評価と異なり入力に対する手間も大きい。にもかかわらず記入した項目なので、感想コメントは宿泊利用者が強い印象を持った事項に対して率直な意見であると考えられる。

しかし、宿泊利用者の生の声であるがゆえ、分析には多くの問題が生じる。まず一番に解決すべき課題は、表記の揺れである。「風呂」と「バス」、「スタッフ」と「係員」など同義語の表記の揺れが多い。また、意図の解釈においても、その係り受けを見なければ判定はできない。たとえば、「大きい」についても、「騒音が大き

い」はネガティブな意見であり、「TV が大きい」はポジティブな意見である。さらには、携帯電話などからの感想コメントの書き込みができるサイトも多く、それゆえ口語調の書き込みも散在しており、自然言語処理技術による解析が困難である。

以降、本章ではこれらの課題を解決し、感想コメントから宿泊検討者にとって有意義な情報を抽出するため我々が提案する手法について説明する。

3.1 テキストマイニング

テキストマイニングとは、テキストを自然言語処理したのち、データマイニングの技術を用い、テキストの記載内容の傾向や特徴を抽出し、新たな知見を見いだす技術である[20,21].

前章に示したように、宿泊検討者の利用目的により重要視する事項は大きく異なる。これらを一括して分析することは困難である。そこで本論文では、図1に示すように利用目的を大きく「ビジネス」と「レジャー」の2つに分割して、各々に対して分析することとした。さらに詳細な分析を試みるため、エリア間の地区別分析も実施することとした。

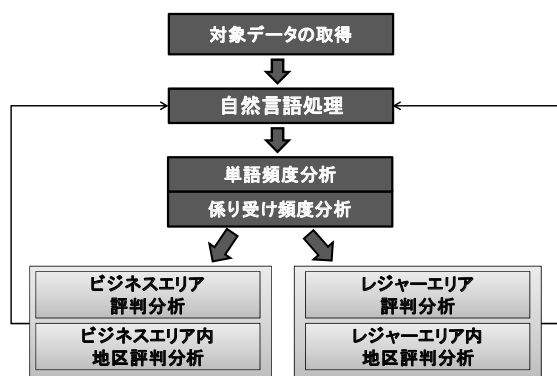


図1. テキストマイニングによる分析フロー

自然言語処理では、まず形態素解析を行い、文を形態素に分割すると共に、各形態素に品詞情報を付与する。このとき、旅行に関する専門用語は形態素辞書に登録することで対処した。

感想コメントには、口語調の表記が多く存在する。形態素解析において、口語調文書の処理は文語調の処理と比較して解析精度が低下する傾向にある。しかし、本研究では、当該宿泊施設やその施設が立地する地域の何に対して満足や不満を感じたかを示す特徴表現を抽出できれば、その後の処理に影響を及ぼさない。さらに、たとえば良いことを示す「グ～～」など口語調で記載された評価単語を形態素辞書に登録することで対処した。また、顔文字などの書き込みも見受けられる。これに対しては、

鈴木ら[22]の研究成果を活用することで対処した。

感想コメントを解析するにあたり、「バス」と「風呂」などの異表記の同義語もしくは類義語が多々存在しているという課題がある。この課題に対しても、異表記の同義語もしくは類義語を類義語辞書としてグルーピングすることで対処した。

3.2 感想コメントの特徴分析

宿泊検討者が感想コメントから欲する情報とは、当該宿泊施設の情報であり、宿泊施設が提供する施設情報の信頼性や宿泊利用者が実際に利用して感じたこと、施設情報に掲載されていない宿泊施設周辺の情報などである。これら宿泊検討者の要望を満たすためには、宿泊利用者が当該宿泊施設やその施設が立地する地域の何に対して満足や不満を感じたかを明らかにする必要がある。

3.2.1 感想コメントの肯定・否定の判定

感想コメントの特徴表現には、「良い」、「悪い」などという明瞭な肯定、否定を表す評価単語を用いた表記と、たとえば「部屋が広がった」のように事実の記載であるが、明らかに肯定と判断できる表記がある。本論文では、前者を感性表現と呼び、後者を事実表現と呼ぶこととする。

感性表現の処理は、感性表現の対象物を明確にすればよい。よって、自然言語処理における構文解析により、感性表現の係り受け表現を抽出し対象物を明確にする。

事実表現では、対象物と評価の係り方により、評価が異なる単語が存在する。たとえば、「料金ー高い」は否定

表4. 評価属性辞書

感性表現				
<対象>	<属性>	表記	<評価>	極性
接客	従業員 (類)	スタッフ	良い	肯定
料理	夕食 (類)	ディナー	美味しい	肯定
風呂	特徴風呂 (類)	露天風呂	良い	肯定
料理	朝食 (類)	朝食バイキング	残念	否定
風呂	お風呂 (類)	大浴場	残念	否定
部屋	部屋 (類)	部屋	残念	否定
事実表現				
<対象>	<属性>	表記	<評価>	極性
部屋	部屋 (類)	部屋	広い	肯定
立地	最寄駅 (類)	大阪駅	近い	肯定
風呂	風呂設備 (類)	バスタブ	広い	肯定
部屋	部屋 (類)	部屋	狭い	否定
部屋	客室寝具 (類)	ベッド	堅い	否定
風呂	お風呂 (類)	大浴場	遠い	否定

表現であり、「品質—高い」は肯定表現である。そこで、本研究では、小林ら[23]の評判分析に関する研究を、宿泊レビューに適用すると共に拡張することで対応した。

小林らは、評判分析に用いる肯定・否定表現の評価は、〈対象〉、評価対象の具体的な項目である〈属性〉、および〈評価〉の3要素から構成されるとしている。本研究では、数値評価との連携を考慮し〈対象〉は数値評価の項目とし、〈属性〉を“大浴場”や“ベッド”など具体的な評価物、〈評価〉を評価物の状態とした。

そして、〈属性〉と〈評価〉の係り受けに対して、肯定極性・否定極性を付与した評価属性辞書を構築した。構築した辞書の一例を表4に示す。表中の〈属性〉に示す(類)が付いた単語は、類義語辞書に登録している代表表記であることを表している。感想コメントには表記の揺れが多いため、類義語辞書に登録している語に対しては、その代表表記を〈属性〉として登録することで評価属性辞書への登録数を削減した。

3.2.2 感想コメントにおける属性の確定

日本語の特徴として、主語を省略する傾向がある。感想コメントにおいても同様の傾向がみられ、〈属性〉が曖昧になったり、抽出できなかつたりするという課題が生じる。また、1つの〈属性〉に対して〈評価〉が2つあるものや、同時に複数の〈属性〉に対して〈評価〉をしているものがある。

このような課題に対して、本研究では属性継承の手法を用いて対応した。属性継承の基本的な考え方は、文を前方から解析していき、一度〈属性〉を抽出すると次の〈属性〉を抽出するまで、抽出している〈属性〉を継承し続けるというものである。属性継承は、〈属性〉が属する

例文 No.	例文
〈対象〉	〈属性〉継承の状況
例文 1	風呂は狭いけど、ベッドは広かった。
施設	風呂 → ベッド →
例文 2	風呂もベッドも狭かった。
施設	風呂 → & ベッド
例文 3	風呂は狭いけど、予想通りであった。
施設	風呂 →
例文 4	風呂は狭いけど、値段が安いので許す。
施設	風呂 →
価格	値段 →

図2. 属性継承の事例

〈対象〉ごととする。

以下、図2に示す事例を用いて属性継承の基本的なルールを説明する。ここでは〈対象〉を表1に示す数値評価項目とし、例文中の一重下線は〈属性〉を、二重下線は〈評価〉を表す。

例文1と2は、施設の〈属性〉“風呂”，“ベッド”が出現する。例文1では、最初に出た“風呂”は“ベッド”が出現するまで継承し，“ベッド”は出現位置から文末まで継承する。例文2では，“風呂”，“ベッド”は並列句と判断されるため、この2つを文末まで継承し、「風呂—狭い」と「ベッド—狭い」という2つの評価属性を抽出する。例文3では、施設の〈属性〉として“風呂”が抽出され、文末まで継承する。これにより「風呂—狭い」と「風呂—予想通り」の2つの属性評価を抽出する。

例文4では、まず“風呂”が施設の〈属性〉として抽出され、9文字目で“値段”が抽出される。ここで、“値段”は価格の〈属性〉であるので、施設〈属性〉の“風呂”は文末まで継承する。ここまでの「風呂—狭い」と「値段—安い」が抽出される。文末の“許す”に対しては「風呂」か“値段”かが曖昧になる。しかし，“許す”は否定の許容なので否定極性を示す「風呂—狭い」への対応と判断し、「風呂—許す」を抽出する。このような処理を実現するため、“許す”や“我慢”などの一部の〈評価〉に対しては、辞書にどちらの極性を優先するかのフラグを付けることで対処した。

4. テキストマイニングによる分析

本章では、前章で提案した手法を用いて、感想コメントを分析した結果を示す。はじめにビジネスエリアおよびレジャーエリアについての特徴の違いを示す。次にビジネスエリア内、レジャーエリア内のそれぞれ地区による特徴の違いを示す。

この分析における〈対象〉は、「じゃらん net」の数値評価項目と対応を取るため〈部屋〉、〈風呂〉、〈料理〉、〈接客〉、〈清潔感〉の5項目に感想コメント中に書き込みが多い〈立地〉を加えた6項目とした。

4.1 分析データと分析システムの概要

分析に用いた感想コメントは、2.2節に示す「じゃらん net」における宿泊レビューのデータと同じで、その平均文字数と単語数を表5に示す。

形態素解析は、オープンソースの形態素解析エンジン MeCab[24]を用い、形態素辞書にはIPA辞書に登録した。

本分析のために、形態素辞書には新たに2,454語を、類義語辞書には158組1,505語を登録した。また、評価

属性辞書には類義語辞書に登録した1組を〈属性〉や〈評価〉における1語とみなして1,825組を登録した。

表 5. 分析対象感想コメントの規模

エリア・地域	平均文字数	平均単語数
東京	147.7	87.2
大阪	146.4	86.4
名古屋	139.8	83.0
博多	140.5	83.6
札幌	148.6	87.8
ビジネスエリア計	144.6	85.6
箱根	222.6	134.4
湯布院	226.5	136.1
草津	217.4	131.4
登別	191.0	114.7
道後	195.3	118.5
レジャーエリア計	210.6	127.0
全体計	177.6	106.3

4.2 エリア別特徴表現

感想コメント全体をビジネスエリア、レジャーエリアに二分し、係り受け表現を集計し、それぞれのエリアで注目される特徴表現の頻度を集計した結果を表6に示す。

表 6. エリア別の特徴表現

ビジネスエリア			レジャーエリア		
〈属性〉	〈評価〉	頻度	〈属性〉	〈評価〉	頻度
接客(類)	良い	336	お食事(類)	美味しい	462
最寄駅(類)	近い	244	接客(類)	良い	408
部屋(類)	広い	217	感じ(類)	良い	266
立地(類)	良い	217	お風呂(類)	良い	237
部屋(類)	狭い	198	気持ち(類)	良い	198
清潔(類)	良い	194	部屋(類)	広い	189
朝食(類)	美味しい	179	お食事(類)	良い	158
感じ(類)	良い	140	清潔(類)	良い	120
交通アクセス	良い	124	眺め(類)	良い	104
最寄駅(類)	便利	110	温泉(類)	良い	103

出張利用者の多いビジネスエリアでの集計では、「最寄駅-近い」、「部屋-広い」、「部屋-狭い」といった事実表現と、「接客-良い」、「立地-良い」、「交通アクセス-良い」、「朝食-美味しい」、と接客、駅からの立地、朝食に対する感性表現が多く出現した。

一方、レジャーエリアでは「お食事-美味しい」、「夕食-美味しい」、「お食事-良い」、「お風呂-良い」等の感性表現がみられ、事実表現では「部屋-広い」がみられる。レジャーエリアにおいては、「良い」という表現を用いた感性表現により、滞在中の満足度を高める要素で

ある食事と風呂に対する肯定的な感性表現が上位を占めた。

すなわち、ビジネスエリアでは、立地と部屋に関して、具体的な事実を評価する事実表現の投稿が多くみられることから、旅先で普段通りの生活を維持するため、宿泊本来の目的に対する利便性が重要視されていることが読み取れる。一方、レジャーエリアでは、客室や立地の良さではなく、食事、風呂といった旅先の宿泊施設で受けた非日常的な体験から発する要素への評価が重要視されていることが分かる。

以上から、宿泊検討者の利用目的により、必要とする数値評価項目とその内容に違いがあることを表していると考えられる。

4.3 エリア別数値評価との対応

本節では、ビジネスエリア、レジャーエリア別で、感想コメント中に出現する〈属性〉と、数値評価項目の対応について明らかにする。

まず、感想コメントから評価属性辞書を用いて抽出した肯定・否定極性を持つ特徴表現を、〈対象〉の情報を用いて宿泊レビューの数値評価項目へ割り当て、その頻度を集計した。ビジネスエリアの集計結果を表7に、レジャーエリアの集計結果を表8に示す。

表 7. ビジネスエリア数値評価平均と特徴表現分析表

数値評価	部屋	風呂	料理	接客	清潔感	
平均点	4.1	3.9	4.1	4.1	4.2	
感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	813	439	694	916	350	1,404
否定表現	406	91	78	47	34	40
中立・他	48	32	22	23	1	11
総計	1,267	562	794	986	385	1,455

表 8. レジャーエリア数値評価平均と特徴表現分析表

数値評価	部屋	風呂	料理	接客	清潔感	
平均点	4.1	4.3	4.4	4.3	4.2	
感想評価	部屋	風呂	料理	接客	清潔感	立地
肯定表現	526	1,118	1,900	1,309	274	471
否定表現	164	233	184	74	29	19
中立・他	35	65	67	33	2	5
総計	725	1,416	2,151	1,416	305	495

数値評価の平均点は、ビジネスエリアでは3.9~4.2、レジャーエリアでは4.1~4.4と、どちらのエリアもその振れ幅は0.3と極めて小さい値であった。このことから、宿泊利用者が投稿する数値評価は、それぞれを平均点にしてしまうと特徴が表れないと言える。

けれども、宿泊利用者が投稿する感想コメントから抽

出した評価を用いると、差異が明らかになる。たとえば、表7に示すビジネスエリアの〈部屋〉と〈接客〉の数値評価項目の平均点数は同じ4.1となっている。しかし、感想コメントの分析結果では、「部屋」においては、1,267コメント中813コメントの肯定極性となり、肯定率は約64%であった。一方「接客」においては、986コメント中916コメントが肯定極性となり、肯定率は約93%であった。

また、数値評価項目に対応する特徴表現の出現数より、宿泊利用者が何に注目したかが明らかになる。ここで、「じゃらん net」の数値評価項目にない「立地」に関する特徴表現が多く出現する。

以降は、それぞれのエリアの結果を分析する。

まず、表7に示すビジネスエリアの結果であるが、抽出した特徴表現の頻度は「立地」、「部屋」、「接客」に関するものの順に多かった。このことから宿泊利用者は「立地」、「部屋」、「接客」に注目し感想コメントを投稿していることがわかった。「じゃらん net」の数値評価項目には存在しない〈立地〉に対する肯定評価の頻度が最も高い。これは、ビジネスエリアを利用する際に、数値評価項目として存在する指標より、立地が重要な要素であることを示しており、特徴的である。

一方、表8に示すレジャーエリアで同様の集計を行った結果では、抽出した特徴表現の頻度は「料理」、「風呂」、「接客」に関するものの順に多かった。このことから、宿泊利用者は数値評価項目のうち、これらの項目に対する注目度が高いことが分かる。特徴表現の出現数から、レジャー目的の宿泊利用者は滞在中に泊まる部屋にあまり注目しておらず、滞在中の非日常的な満足を得るための項目である「料理」、「風呂」、「接客」に注目していることは興味深い結果である。

また、ビジネスエリア、レジャーエリアいずれの場合も、「部屋」と「風呂」に関しては他の数値評価項目と比較して、否定極性を持つ特徴表現の比率が高い結果となった。この2つの評価項目に対して宿泊利用者は、数値評価が高いにもかかわらず、不満を持っていることを示していることは興味深い結果である。

ただし、3.2.1に示した属性継承の概念を用いて、部屋の否定極性を持つ特徴表現について検証したところ、宿泊利用者は部屋の狭さに対して否定極性の評価を行うが、料金や、清潔感、立地といった別の属性との総合的な評価で、相応の満足をしているということも明らかとなった。

以上、感想コメントを分析することで、数値評価の高低とは別に、エリアにより注目される数値評価項目が異なること、数値評価項目の優先順位が存在すること、数

値評価の点数だけでは読み取ることできない差異があることが明らかになった。

4.4 地区別分析

次にビジネスエリア、レジャーエリア内の各地区の特徴について分析を行った。いずれの地区においても、立地に関する表現が多いことが特徴的であった。特にビジネスエリアの地区における特徴表現には、当該地区の最寄り駅名が多くみられ、レジャーエリアの地区においては最寄りの観光地名やランドマーク的名所の単語が多くみられた。

また、ビジネスエリアの各地区において、「部屋の狭さ」という否定極性を持つ特徴表現の頻度は、東京、大阪、名古屋では多く、博多、札幌では少ないという傾向の違いがみられた。

一方、レジャーエリアの各地区は、それぞれ温泉地であるにも関わらず、最も重要視される数値評価項目は、「風呂」ではなく「料理」に関するものであることがわかった。次に重視される数値評価項目は、湯布院、道後の各地区では「接客」、箱根、草津、登別の各地区では「風呂」に対する特徴表現の頻度が高く、宿泊利用者が重要視する数値評価項目は地区により順序が異なることが明らかとなった。

また、宿泊利用者の感想テキストに具体的な地名が書かれることが多かった。けれども、温泉の効能や泉質、郷土料理の内容の詳細など、より地区の特徴を示す内容の感想コメントは少なかった。

5. 活用方法の提案

本章では、本論文の研究の結果を活用した宿泊レビューの活用方法を提案する。宿泊予約サイトの多くが提供する宿泊レビューは、宿泊施設の数値評価の平均値集計を行った表と、宿泊利用者から投稿された感想コメントを並べて表示するだけであった。それゆえ、数値評価項目にない項目を調べたい場合は、感想コメントを読まなければならなかった。

本論文で提案した手法を用いれば、感想コメントに書き込まれた特徴表現を、数値評価項目ごとに肯定極性と否定極性に分類して抽出することが可能である。この情報を宿泊レビュー中にわかり易く追加表示すれば、宿泊検討者にとって利便性が向上するであろう。そこで、本研究成果を活用した宿泊レビュー画面デザインの一例を図3に示す。

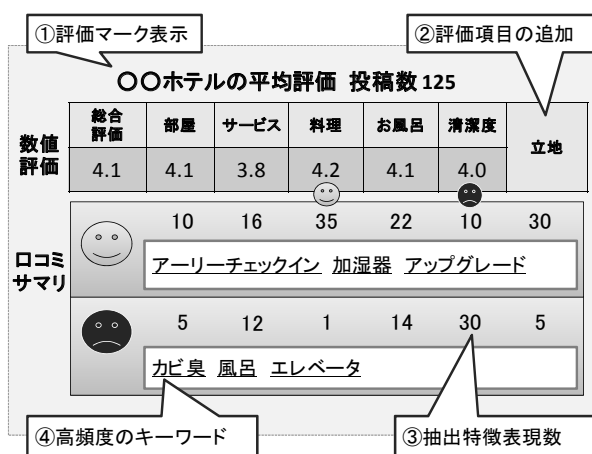


図 3. 宿泊レビュー画面デザインの一例

これまでの宿泊予約サイトが提供する宿泊レビューに追加する提供情報は下記の4つである。

- ① 数値評価項目へ評価マークの表示
感想コメントから抽出した肯定極性と否定極性の書き込みの頻度、もしくは、比率により評価が良い項目と評価が悪い項目を判定し、当該項目に宿泊検討者が一目で判断できるアイコンを表示する。
- ② 書き込みの多い評価項目の追加表示
数値評価項目として定義されていない項目だが、感想コメント中に高頻度で出現する項目を表示する。
- ③ 抽出した特徴表現数の表示
感想コメントから抽出した特徴表現を、数値評価項目ごとに肯定極性と否定極性に分類して、その数をカウントして表示する。
- ④ 高頻度で出現した特徴表現<属性>の表示
感想コメントから抽出した特徴表現中の<属性>を、肯定極性と否定極性ごとに集計し、高頻度で出現した<属性>をキーワードとして表示する。さらには、表示キーワードにハイパーリンクを付加して、表示キーワードを含む感想コメントの原文を容易に表示できるようにする。

例示した図3において①～④を表示することで、数値評価だけでは宿泊検討者が把握できなかった宿泊施設の特徴が見えてくる。

立地は数値評価の項目にないが多くの感想コメントが存在し好評であることがわかる。また、サービスの数値評価は3.8と最低点であるが、感想コメントでは良い評価としてアアリーチェックインやアップグレードなどがあり、運用上は良いサービスを提供していることがわかる。さらに、部屋の数値評価は4.1であるが、感想コメントではカビ臭やエレベータなど施設全般に悪い評価が

あり、施設が古いもしくは整備が行き届いていないことがわかる。

ここに示す表示方法は一例であるが、宿泊検討者が当該宿泊施設の感想コメントの概要を容易に理解できる表示ができれば、利用する宿泊施設を選択する作業が効率的になるであろう。

6. おわりに

本論文では、宿泊レビューの感想コメントから、テキストマイニングを用いて宿泊検討者にとって有益となる情報を抽出する方法を提案した。また提案方法を用いて感想コメントの分析を行った。

その分析結果を数値評価と比較すると、宿泊レビューの数値評価において同じ点数になっている項目でも、宿泊利用者の満足度に差異があることがわかった。また、数値評価にない項目でも宿泊利用者が重要視した項目があることがわかった。

さらに感想コメントの内容を分析すると、同じ数値評価項目でも、ビジネスエリアとレジャーエリアの宿泊利用者では、重要視している事柄が異なることがわかった。また、同じ利用目的であっても、地区ごとに注目される数値評価の項目が異なること、駅名や観光地名などが多く出現することなどから、地域性が表れていることがわかった。

分析におけるプラクティスには、エリアや地域ごとに顧客の要望は異なるが、個々の要素に分割すれば、同一の評価基準で評価できるのではないかとこの想定があった。そのプラクティスから、顧客ニーズの対象を本研究で示した具体的な評価物である<属性>まで細分化すれば、顧客の要望はエリアや地域に影響されないという知見が得られた。その結果、全ての分析を同一の評価属性辞書で処理できた。

これら研究成果の活用方法として、数値評価の項目へ評価マークの表示、書き込みの多い評価項目の追加表示、抽出した特徴表現数の表示、高頻度で出現した特徴表現<属性>の表示と、宿泊検討者にとって有益となる情報を宿泊レビューに掲載する手法を提案した。

今後は、本論文で検討した手法の精度をより高めると共に、数値評価には現れない当該宿泊施設や地区の情報を、宿泊検討者へ効果的に伝える手法について検討する。さらには、提案手法をリアルタイムで実行し、宿泊検討者へタイムラグなしで情報提供できる動的情報更新システムの実用化を検討する。

参考文献

- 1) 日本交通公社: 旅行者動向 2007, 財団法人日本交通公社 (2007).
- 2) 宿泊旅行統計 平成 22 年版, 観光庁 (2011).
- 3) Gretzel, Kyung: Use and Impact of Online Travel Reviews, Information and Communication Technologies in Tourism 2008, pp.35-46 (2008).
- 4) 長島 健太, “サービス商品の選択・評価におけるクチコミの機能—ホテルサービスの場合—”, 東洋大学経営学部 経営論集 74 号 pp.91-108 (2009).
- 5) 植竹 朋文, 青木 章通, “繁閑格差の大きい地区のホテルにおけるネット直販に関する意識調査”, 専修大学 情報科学研究所報 No.76 pp.1-26 (2011).
- 6) 辻井 康一, 津田 和彦: ユーザーレビューを用いた宿泊施設の情報補完手法, 第 10 回情報科学技術フォーラム FIT2011, E-003 (2011).
- 7) 大内 東, 大柳 幸彦: 日本のホテル評価システムに関する研究, 開発こうほう, (539), pp.27-31 (2008).
- 8) 森川 毅: ホテル・旅館の評価: 心理学的側面からの分析, 長崎国際大学論叢 1, pp.273-280 (2001).
- 9) 田邊 亘, 後藤 正幸: 宿泊施設の戦略構築を支援するユーザーレビュー分析に関する一考察, 武蔵工業大学環境情報学部情報メディアセンタージャーナル, Vol.9 (2008).
- 10) 槻木 公一: 旅行者中心の旅行支援システムに関する一考察, 新潟国際情報大学情報文化学部紀要, pp.205-219 (2005).
- 11) じゃらん net: <http://www.jalan.net/>
- 12) 楽天トラベル: <http://travel.rakuten.co.jp/>
- 13) るるぶトラベル: <http://rurubu.travel/>
- 14) 一休.com : <http://www.ikyuu.com/>
- 15) 旅ぷらざ: <http://www.tabiplaza.net/>
- 16) Booking : <http://www.booking.com/index.ja.html>
- 17) Agoda : <http://www.agoda.jp/>
- 18) HotelClub : <http://www.hotelclub.co.jp/>
- 19) Expedia : <http://www.expedia.co.jp/>
- 20) 那須川 哲也, 『テキストマイニングを使う技術/作る技術』, 東京電機大学出版局 (2006).
- 21) 乾 孝司, 奥村 学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201-241 (2006).
- 22) Nobuo Suzuki, Kazuhiko Tsuda: An efficient method of intention understanding and learning for the ambiguous expressions of sentence end. IJCAT 33(2/3), pp.99-108 (2008).
- 23) 小林 のぞみ, 飯田 流, 乾 健太郎, 松本 裕治, 照応解析手法を利用した属性-評価対および意見性情報の抽出, 言語処理学会第 11 回年次大会, C2-6 (2005).
- 24) <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

辻井 康一 (正会員)

E-mail: tsujii@gmail.com

1999 年大阪市立大学文学部史学地理学科卒業。同年西日本旅客鉄道 (株) 入社。2001 年 (株) 日本旅行入社。2012 年筑波大学大学院ビジネス科学研究科経営システム科学専攻修了。修士 (経営学)。インターネット上での旅行者の行動, 満足度評価に興味を持つ。情報処理学会会員。

津田 和彦 (正会員)

E-mail: tsuda@gssm.otsuka.tsukuba.ac.jp

1986 年徳島大学工学部情報工学科卒業。同年三菱電機 (株) 入社, 1991 年住友金属工業 (株) 入社。1994 年徳島大学大学院工学研究科システム工学専攻修了。博士 (工学)。1998 年筑波大学大学院助教授。2005 年同大学院教授。情報検索, 自然言語処理に興味を持つ。情報処理学会, 電子情報通信学会, 人工知能学会等会員

投稿受付: 2012 年 03 月 05 日

採録決定: 2012 年 05 月 25 日

編集担当: 大場みち子 (公立はこだて未来大学)