

宿泊レビューの数値評価を考慮した感想コメント分析

User comments analysis in consideration of numerical evaluation of the Guests' review

辻井 康一
Koichi, Tsujii

津田 和彦
Kazuhiko, Tsuda

筑波大学 システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba

Online hotel reservation is widely used for planning travel today. Users of reservation usually refer to the user reviews before making their reservations. Numerical evaluation and an impression comment exist in a Guests' review. In this paper, we analyze the relevance of numerical evaluation and an impression comment using text mining. And we show that an impression comment complements the content of numerical evaluation.

1. 宿泊レビューの利用状況

インターネット宿泊予約サイトの登場で、宿泊検討者は 24 時間いつでも宿泊施設の予約ができるようになり、利用者数はこの 10 年で 10 倍を超える成長を示している[JATA 2013]。宿泊検討者は最も良い条件の宿泊施設を選ぶために宿泊予約サイトが提供する様々な情報を活用し情報を収集する。昨今の宿泊検討者は、宿泊施設の提供する広告や公式情報だけを一方的に鵜呑みにせず、サービス提供を受けた宿泊利用者が残す宿泊レビューを閲覧し、その評価を宿泊施設選択の参考情報として重要視している[Nielsen 2012]。宿泊レビューは宿泊利用者のプロフィール、数値評価、感想コメントの 3 つ要素から構成され、感想コメントには宿泊利用者の実際の体験に基づく生の声が投稿されている。宿泊レビューは有益な情報であるが、その情報提供方法は宿泊利用者の評価情報を並べて掲載しているだけであるため、人気宿泊施設に膨大に集まる宿泊レビューは、通読することが困難である。また、数値評価情報は個人の主観による評価となるが、一般的に高い評価をつける宿泊利用者が多い。またすべての数値評価を細かく評価せず、オール 5 といった様に全ての数値評価項目を同一にする宿泊利用者が多く存在する。そのため、数値評価を集計するだけでは、宿泊施設の具体的な特徴を知ることは難しい。

本研究では、宿泊利用者から寄せられる宿泊レビューの要素である感想コメントと数値評価を用いて、利用者が数値で評価した内容と感想コメントの評価を示す特徴表現の関連性を分析する方法を検討する。具体的には数値評価において宿泊利用者が注目した評価項目に対する具体的な評価内容を、感想コメントから対象の特徴表現として抽出する方法を検討し、特徴表現を数値評価と対応させることで宿泊レビューの評価項目のさらなる活用と、インターネット宿泊予約の利便性を向上させることを目的としている。

2. 宿泊施設の評価情報

インターネット宿泊予約サイトには、ほぼ必ずと言ってよいほど宿泊レビューの機能を有している。この宿泊レビューを調査した研究として、辻井らは、宿泊レビューの感想コメントから、肯定評価と否定評価の特徴表現を抽出し、数値評価項目と対応させることでエリアの特徴、宿泊施設の特徴を表現する方法を示した。また評価表現は宿泊利用者の関心の高い数値評価項目

連絡先: 辻井 康一, 筑波大学 システム情報工学研究科
〒112 - 0012 東京都文京区大塚 3 丁目 29-1,
e-mail: tsujii@gmail.com

に集中することや、人気宿泊施設において肯定的な評価表現の一部に否定的な評価表現も書かれていることを示した[辻井 2012]。この研究では、感想コメントから肯定・否定評価を示す特徴表現を抽出し、それらを数値評価項目と対応させることで、宿泊利用者がどの数値評価項目に注目しているかの検証を行っている。その結果、ビジネス利用エリアでは、立地・部屋・サービスに対する評価を行う場合が多いこと、レジャー利用エリアでは、風呂・料理・サービスに対する評価が多いことを示している。これらの結果から、宿泊利用者は数値評価項目のすべてを限らず評価しているのではなく、利用目的に合わせて評価対象が異なることを示している。一般的にはユーザーは 4 または 5 の評価をつける場合が多く、1 または 2 の評価をつける人は少ない。宿泊施設にとって改善すべき項目は、概ね良いという判断の中に隠れた課題点を指摘するコメントである。これらを発見することが重要となるのであるが、数値評価のすべての項目のうち、1 項目だけの評価を変える宿泊者が存在する。この 1 項目だけ異なる数値評価項目は、宿泊利用者が体験した中から感じたこだわりの数値評価である。この宿泊者のこだわりと考えられる数値評価に対し、感想コメントには数値評価を変えた理由となる具体的な評価内容が記載されていると想定される。テキストマイニングを用いて感想コメントから評価表現を抽出し、その特徴を検証することでより良いサービスを提供するために必要な要素を発見することが出来ると想定される。

本研究では、ユーザーは意図的に数値評価項目と感想コメントを関連付けて記載しているかについての検証を行う。そして関連付けて記載している場合の特徴を分析し、宿泊検討者や宿泊施設に向けて有用な情報としての情報提供方法について検討を行う。

3. 数値評価の特徴

宿泊利用者の数値評価の採点について分析を行う。表 1 は、宿泊予約サイトじゃらん.net(<http://www.jalan.net>)における数値評価の採点 10,000 件を集計した結果である。

表 1. 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・サービス	清潔感
5	4,316	3,871	3,891	3,703	4,639	4,597
4	4,086	3,908	3,571	2,367	3,340	3,476
3	1,122	1,726	1,925	936	1,521	1,486
2	367	352	396	263	330	329
1	109	143	106	85	170	112
総計	10,000	10,000	9,889	7,354	10,000	10,000

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 地区を選択し, 各 10 の宿泊施設を選択, 宿泊レビューを各 100 件抽出したものである。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%に満たず数値評価には偏りが大きいことがわかる。同データで, 全ての数値評価項目を採点している 7,286 件について, その採点傾向を集計した結果を表 2 に示す。

表 2. 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	1,341	18.4%
1項目のみ異なる採点	1,682	23.1%
2項目以上で異なる採点	4,263	58.5%
合計	7,286	100.0%

全項目で同一の採点が 18.4%, 1項目のみ異なる採点が 23.1%であった。約 4 割の宿泊利用者は, 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, 強く印象を持った採点項目であると言える。1 項目だけ数値評価を変える方法には 1 項目だけ評価を上げる場合と, 下げる場合の 2 種類が想定される。表 3 は 1 項目のみ評価点数を変えた宿泊利用者が, どの数値評価項目に対し, 評価を上げ, また下げたのかを集計し, 頻度で示したものである。集計の結果は 1 項目だけ評価を下げる場合の方が多くみられた。そして評価を下げる対象として風呂, 部屋, 料理に対するものが多く, 1 項目だけ評価を上げる場合には, 料理, 風呂, サービスの順に頻度が高いことがわかった。

表 3. 1項目のみ異なる採点の傾向

	総合	部屋	風呂	料理	接客・サービス	清潔感	合計
上げて評価	6	57	99	110	74	47	393
下げて評価	9	326	476	234	136	108	1,289
合計	15	383	575	344	210	155	1,682

4. 評価に関する特徴表現の抽出

宿泊利用者の数値評価における行動の特徴について, 感想コメントのテキスト情報を利用し, その関連性を分析する。数値評価項目ごとにみられる特徴表現をテキストマイニングの手法により抽出する方法を検討する。感想コメントは宿泊利用者が自由に記述する情報のため, 表現の揺れが大きい。そのため類義語辞書を構築し, 表現を揃える処理を行う。次に評価表現の肯定, 否定極性を判定するために評価表現辞書を構築する。その際, 本研究では小林らの評判分析に用いる肯定・否定表現の対象, 属性, 評価の3つ組構成を活用した[小林 2005]。数値評価との連携を考慮し<対象>は数値評価項目とし, <属性>を“大浴場”や“ベッド”など具体的な評価物, <評価>を評価物とする構造とした。これらの辞書をもとに, 感想コメントの係り受け解析を行い, 肯定および否定評価を示す特徴表現を抽出する。

5. 数値評価を考慮した特徴表現分析

前章で検討した評価表現の抽出方法を利用し, 1 項目のみ異なる採点を行った感想コメントの分析を行う。数値評価の採点には 1 項目だけ数値評価を上げる場合と, 評価を下げる場合が存在する。3 章で示した数値評価と同時に記載された感想コ

メントを用いて分析をおこない, 評価を示す特徴表現の出現頻度を集計した。その結果を表 4, 表 5 に示す。

表 4. 1 項目のみ数値評価を上げた場合の特徴表現分析表

	部屋	風呂	料理	接客・サービス	清潔感	合計
肯定表現	34	54	99	39	21	247
否定表現	11	16	11	4	4	46
合計	45	70	110	43	25	293

表 5. 1 項目のみ数値評価を下げた場合の特徴表現分析表

	部屋	風呂	料理	接客・サービス	清潔感	合計
肯定表現	138	181	427	299	63	1108
否定表現	30	44	24	2	2	102
合計	168	225	451	301	65	1210

表中の肯定表現と否定表現は 4 章で示した評価に関する特徴表現の抽出方法を用いて感想コメントから特徴表現を抽出し, それらを数値評価項目と対応させ頻度で示したものである。表 4 のように数値評価を 1 項目のみ上げた場合, 高く評価される対象となった数値評価項目に対し, 肯定表現が多く集まる結果となった。たとえば料理の場合では, 料理メニューが良かったことへの評価が示されている。一方, 表 5 のように 1 項目のみ数値評価を下げた場合には, その数値評価項目に対して感想コメントには否定評価が書かれる結果となった。具体的には, 風呂の場合, 客室風呂や露天風呂が狭かったことへの不満が記載されている。これらの内容は数値評価を 1 項目だけ上下させた理由を示すものであり, 表 3 で提示した異なる採点の数値評価項目を見ただけでは理解できない宿泊利用者のこだわりの採点理由を感想コメントにおいて示すものである。

6. おわりに

本研究では, 宿泊レビューの数値評価と感想コメントを用いて数値評価の内容と感想コメントの評価を示す特徴表現との関連性を分析した。宿泊レビューの中から数値評価を 1 項目だけ変える宿泊利用者の数値評価情報と感想コメントの特徴表現を対象に検証したところ, 1 項目だけ数値評価を変化させた数値評価と特徴表現とに連動性が見られた。またその際に出現する特徴表現には, 1 項目だけ数値評価を変える理由となった事象が見られることを示した。

今後の取り組みとして, 評価情報収集の効率化を行うとともに, 宿泊検討者への情報提示方法の検討を行う。

参考文献

- [JATA 2013] “数字が語る旅行業 2013”, 日本旅行業協会 2013
- [Nielsen 2012] “Consumer Trust in Online, Social and Mobile Advertising Grows”, Nielsen, 2012
<http://www.nielsen.com/us/en/newswire/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html> (2013/11/25 accessed)
- [辻井 2012] 辻井 康一, 津田 和彦, 2012. “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”, 情報処理学会デジタルプラクティス, 3(4), pp.289-296.
- [小林 2005] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一 “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.2 2005