

口コミ web サイトを用いた道の駅の施設評価

長岡技術科学大学	環境社会基盤工学課程	非会員	○梶笑璃
長岡技術科学大学大学院	環境社会基盤工学専攻	非会員	大沼薫
長岡技術科学大学大学院	環境社会基盤工学専攻	正会員	佐野可寸志
長岡技術科学大学大学院	環境社会基盤工学専攻	正会員	高橋貴生
長岡技術科学大学大学院	環境社会基盤工学専攻	正会員	松田曜子

1. はじめに

インターネットが多くの人に普及した現在、出掛ける際に web 上の情報から目的地を決定することも多くなってきている。特に web 口コミサイトには、施設の充実さや特徴、サービスの内容、サービス内容に対する評価等の声が多く寄せられており、こういった情報を参考の一つとするインターネットユーザーも少なくない。

道の駅は、「休憩機能」、「情報発信機能」、「地域連携機能」を併せ持つ休憩施設である¹⁾。この3機能のほかに、人々の価値観の多様化により、個性的でおもしろい空間が望まれており、道の駅では、沿道地域の文化、歴史、名所、特産物などの情報を活用し個性豊かなサービスが提供されており、休憩施設としてだけでなく目的地として道の駅が利用されることも多い。

本研究では、関東・新潟の道の駅を対象とし、web 口コミサイトから得られた口コミ情報を用いて感性分析を行うことで、各道の駅の施設評価を行うとともに、傾向を把握する。また、評価点数の確認を行い、道の駅の口コミコメントと評価点数の傾向を確認する。

2. 分析方法

2.1. 使用データ

(1) 口コミ情報

使用した口コミ情報のデータ元は、「じゃらんnet」・「トリップアドバイザー」から取得したものである。各サイトより、訪問日時・コメントを抽出し、訪問日時が2015年～2019年の口コミ情報を用いた。評価点数については、じゃらんnetからのみ取得した。理由として、口コミサイトの違いにより評価点数にばらつきがみられることが確認できたためである。また、じゃらんnetの評価点数を使用する理由については、2つのサイトの口コミ数を比較した際、じゃらんnetの方が多く口コミが投稿されており、口コミ数の大きい方が評価点数も安定するとしたためである。各サイトの口コミ数の平均値は表1に示すとおりである。

本研究では道の駅数が184駅と非常に多くなるため、傾向を確認するために道の駅の分類分けを行なった。レジカウント数との相関がある変数を含む重

表1 道の駅1駅当たりの口コミ数平均値

じゃらんnet	トリップアドバイザー	全体
79.8件	42.9件	122.7件

表2 重回帰式に含まれる変数

変数	偏回帰係数	標準誤差	標準偏回帰係数	F値	t値	P値
会議室の有無	113960.212	42742.043	0.158	7.109	2.666	0.009
算出店舗数	41750.326	6998.407	0.353	35.589	5.966	0.000
納入農家登録数	765.934	189.852	0.253	16.276	4.034	0.000
体験施設有無	148386.586	55189.481	0.123	7.229	2.689	0.009
イベント情報発信媒体数	-105187.696	37311.233	-0.240	7.948	-2.819	0.006
駐車場面積	19.067	5.047	0.395	14.273	3.778	0.000
前面道路交通量	6.710	2.281	0.237	8.654	2.942	0.004
施設面積	-1.243	0.467	-0.196	7.075	-2.660	0.009
R:0.923 修正R:0.916 R2乗:0.852 修正R2乗:0.839						

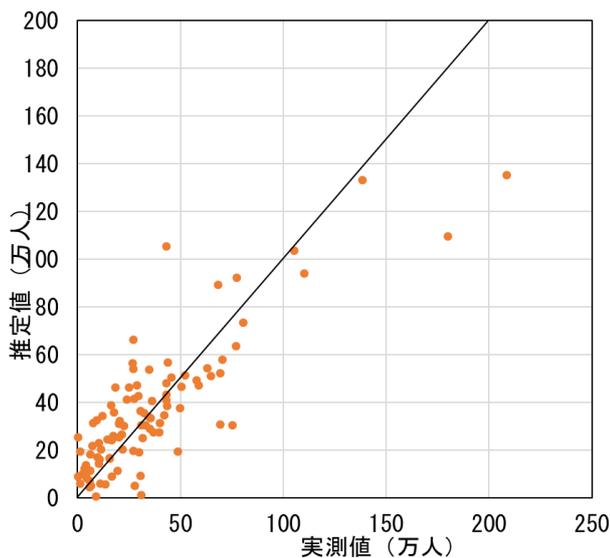


図 1 重回帰分析によるレジカウント人数の
予測値と実測値

回帰モデルを作成した。そのモデルで用いられた変数に欠損値のない道の駅と欠損値のある道の駅で分類し、また欠損値のない道の駅の中で、重回帰モデルとの誤差率別に 50%以上・10%~50%・-10%~10%・-50%~-10%・-50%未満の 5 つに分類を行なった。欠損値のない道の駅は計 98 駅である。作成した重回帰式は式 (1)、使用変数及び重相関係数は表 2、分布図は図 1 に示すとおりである。

$$\hat{y}_i = a_1x_{1i} + a_2x_{2i} + a_3x_{3i} + a_4x_{7i} + a_5x_{5i} + a_6x_{6i} + a_7x_{7i} + a_8x_{8i} \quad (1)$$

x_{1i} :会議室の有無 x_{2i} :算出店舗数
 x_{3i} :納入農家登録数 x_{4i} :体験施設有無
 x_{5i} :イベント情報発信媒体数
 x_{6i} :駐車場面積 x_{7i} :前面道路交通量
 x_{8i} :施設面積
 $a_1 \sim a_8$:パラメータ (偏回帰係数)

(2) 極性辞書使用について

取得した口コミ情報の分析を行なう。2.4. で述べるが、形態素解析によって抽出された語句を、極性辞書を用いポジティブ・ネガティブに分類する。本研究では、日本語評価極性辞書(用言編・名詞編)^{2), 3)}を用いた。日本語評価極性辞書(用言編)の特徴として、評価極性タグがポジティブ(経験)・ポジティブ(評価)・

ネガティブ(経験)・ネガティブ(評価)の 4 分類に分けられる。経験は“客観的”，評価は“主観的”な評価語となる。分析対象として、主観的なコメントと客観的なコメントの両方を分析対象とするため、それぞれをまとめてポジティブ・ネガティブと分類した。また、形容動詞については用言編のみでなく、名詞編でも参照し、サンプル数の増加を図った。

2.2. 評価点の集計

取得した評価点数を用いて、評価点の集計・比較を行う。集計した評価点の平均をコメント数で割ることで、レジカウント数誤差率別に検討する。

2.3. 分析方法

分析方法は、山田ら (2018)⁴⁾の感性評価モデルの提案手法を参考にした。口コミの分析は以下の通りである。

(1) 前処理

分析の前処理として、コメントの形態素解析と表記ゆれの統一を行う。この処理には KH Coder を用いた。抽出したコメントを一つのデータとして捉え、多数のインターネットユーザーから寄せられたコメントであるため、表現方法の揺れが多く発生している。このため、多く出現している語句であっても少なくカウントされてしまう可能性があったため、表記ゆれの統一を行なった。

(2) 評価語の収集及び選定

形態素解析によっていくつかの品詞に分類された語句と、語句の評価を行う極性辞書に記載されている評価語から、分析に用いる評価語の選定を行う。評価語の条件は、①品詞が形容詞自立、動詞、形容動詞であること、②日本語評価極性辞書(用言編・名詞編)に含まれている事、とした。また、極端に出現頻度の低い語句は分析のノイズとなりうるため、全コメント数に対し、出現率が 0.1%未満のものは除くこととする。

(3) 評価語の分類・分析

選定した評価語を、極性辞書を用い、ポジティブ・ネガティブに分類する。また、各評価語の出現回数・出現率を集計し、レジカウント数誤差率別に道の駅の傾向を検討する。

表3 データ数詳細と評価点数

レジカウント誤差率	50%~	10%~50%	-10~10%	-50%~-10%	~-50%
ロコミ数	2438	4173	1641	2283	1287
道の駅数	13	18	13	27	18
道の駅1件当たりのロコミ数	187.54	231.83	126.23	84.56	71.50
平均評価点数	4.17	4.00	3.99	3.92	4.00

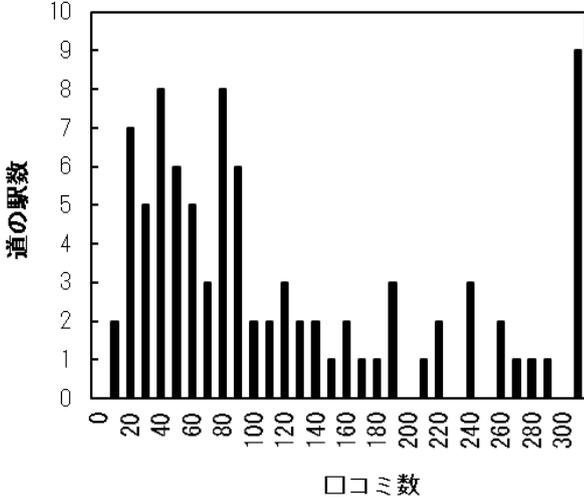


図2 ロコミ数分布

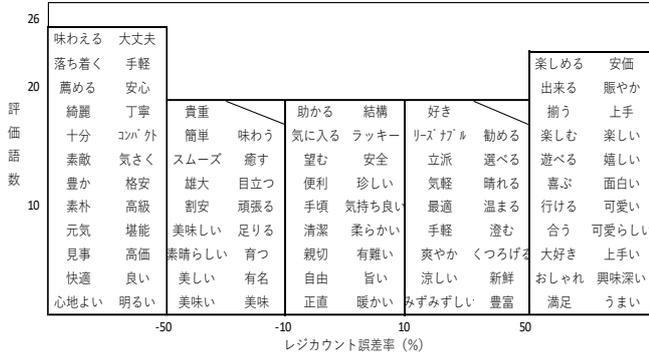
3. 分析結果及び考察

3.1. データ数の集計

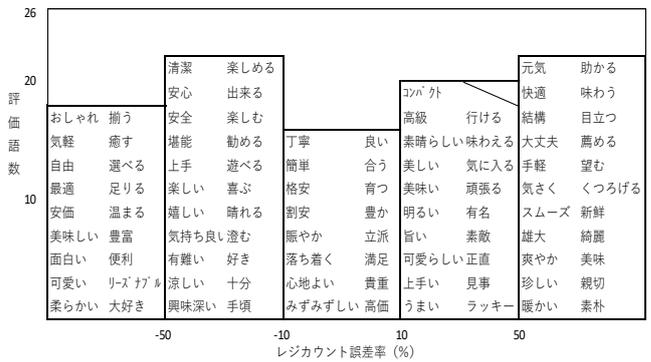
各道の駅でのロコミデータ数の分布は図2, データ数の詳細は表3の通りである. 図2より, 0から100件までにロコミ数が集中していることが確認できる. また, 100件以上のロコミ数の道の駅も多く分布しており, 300以上ある道の駅も全体の約1割程度あることが確認できる. 表2より, レジカウント数誤差が+側に多い方が道の駅1件あたりに対するコメント数が多くなっていることが確認できる.

3.2. 評価点数の集計

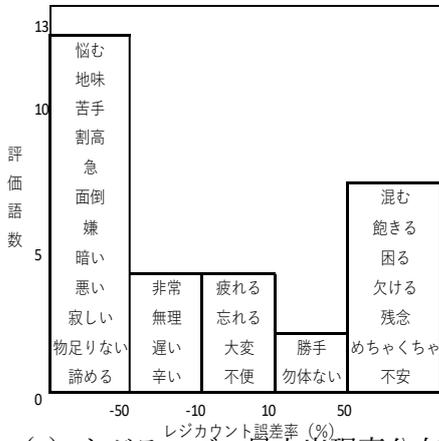
各レジカウント数誤差率での評価点数の平均は表4の通りとなる. 誤差が+50%以上で評価点が大きくなっていることがわかる. そのほかではあまり大きな違いはみられなかった. また, 全体の評価点数の平均点は4.01であった.



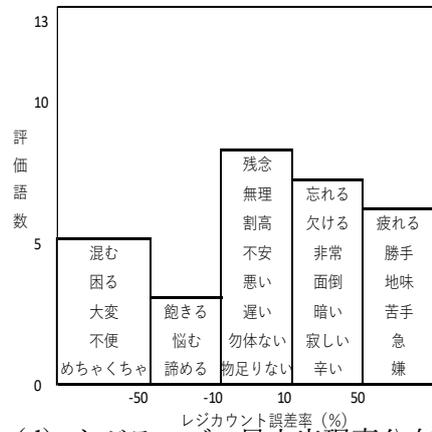
(a) ポジティブ・最大出現率分布



(b) ポジティブ・最小出現率分布



(c) ネガティブ・最大出現率分布



(d) ネガティブ・最小出現率分布

図3 ポジティブ・ネガティブ 出現率分布図

3.3. 評価語の分類

形態素解析を行ない品詞別に分類された語句を、極性辞書を用いポジティブ・ネガティブに分類した結果、ポジティブな語句が 97 表現、ネガティブな額が 29 表現抽出することができた。

評価点数と比較すると、評価点数平均が 4.01 と評価が全体的に高かったと言える。その為コメントもポジティブな語句が多くなったと考えられ、評価点数との傾向があるといえる。ネガティブについては、ノイズ除去のために、全コメント数に対して出現率の 0.1%未満を除いた為、そもそも出現回数の少なかったネガティブ語句が多く除かれたと考えられる。

3.4. 評価語の分析

分類分けを行なった表現語をレジカウント数誤差率ごとに出現回数・出現率を集計し、各表現語を集計した結果が図 3 となる。最大出現率分布では、各表現語がどの誤差率で一番出現率が高いのかを集計しており、最小出現率分布では一番出現率が少なかった評価語を各誤差率で集計したものとなっている。

(a) では、レジカウント数誤差率の+側で評価語数が大きくなると、レジカウント数の推定値より実測値が多い理由が口コミ内に現れていると考えることができ、良い傾向があるといえることができる。(c) では、レジカウント数誤差率の-側で評価語数が大きくなると、レジカウント数の実測値が予測値より少なくなる理由が口コミ内に現れていると考えることができる。(b)、(d) ではそれぞれ (a)、(c) の反対のことがいえる。

まず (a) (b) を見ると、どのレジカウント数誤差率でも評価語数に大きな違いが見られない。よって、ポジティブな表現からのみでは傾向をつかむことが難しい。(c) を見ると、レジカウント数誤差率-50%以下が他に比べ多くなっていることが確認できる。また、(d) では、比較的レジカウント数誤差率が+側の方が多くなっていることが確認できる。よって、ネガティブではレジカウント数誤差率が+側と-側での傾向が見られるといえる。

ネガティブな表現について、評価点数と比較してみる。評価点数ではレジカウント数誤差率 50%以上のとき、表現語ではレジカウント数誤差率が+側と-側でそれぞれ違いが見られた。これらから、レジカ

ウント数誤差率が 50%以上と-50%以下のときに評価語ネガティブな表現で明確な差を見ることができるといえる。しかし、-50%から 50%では評価点数・表現語ともに有意な差が見られないため、別の観点からの分析も必要であると考えられる。

4. まとめ

分析結果では、ネガティブな表現では各誤差率で違いを確認することができた。しかし、評価点数との関連や、ポジティブな表現での違いを確認することができなかった。本研究では、語句をポジティブ・ネガティブの 2 つに分類し、5 つの大きな分類に分けて分析を行なったが、道の駅ごとでの分析を行なうことでより詳細に傾向を確認することができると考えている。また、本研究では評価語の分析で出現率の大きさのみ見ていたが、ポジティブ・ネガティブの中でもどのような傾向のある評価語であるのか等検討することでより精度の高い分析を行うことができると考える。この分析での最終的な目的としては、口コミから各道の駅での傾向を検討しレジカウント数との相関等を見ることで、分類分けで用いた重回帰モデルの変数の一つとして質的な情報として取り込むことである。そのためポジティブ・ネガティブだけでなく、道の駅の特徴となる語句等での分析を行なうことで、誤差率での分類分けでも違いを見ることができるのではないかと思う。

5. 参考文献

- 1) 国土交通省、道の駅案内、<https://www.mlit.go.jp/road/Michi-no-Eki/index.html>, 最終閲覧日 2020.10.13.
- 2) 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.3, pp.203-222, 2005.
- 3) 東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, pp.584-587, 2008.
- 4) 山田篤拓, 橋本翔, 長田典子, テキストマイニングを用いた感性評価モデルの構築, 第 13 回日本感性工学会春季大会予稿集, 2018.