訪日中国人観光客のオンラインホテルレビューの感情分析と評価点の関係性分析
Relation Analysis between Hotel Review Rating Scores and Sentiment Analysis of Reviews by Chinese Tourists Visiting Japan

Elisa Claire Alemán Carreón[a,*], Hirofumi Nonaka[a], Toru Hiraoka[b]

[a] *Nagaoka University of Technology, Nagaoka, Japan*
[b] *University of Nagasaki, Nagasaki, Japan*

Abstract

In current times, the importance of online hotel review sites has become more and more apparent. Users of these sites reference of reviews strongly influences their purchase behavior and as such, reviews are important to companies and researchers alike. The majority of review sites offer both text reviews and numerical hotel ratings, and both information sources are widely used by researchers as a representation of a customer's sentiment and opinion. However, an opinion is a difficult concept to measure, and as such, depending on the relation these two sources have, it would be apparent whether or not it is safe to consider them equally in research. In this study we utilize an entropy-based Support Vector Machine to classify positive and negative sentiments in hotel reviews from the site *Ctrip*, then calculating the ratio of positive and negative sentiment in each review and examine their correlation with said review's rating score using Spearman and Kendall Correlation coefficients and Maximal Information Coefficient (MIC).

Keywords:
感情分析, 観光, SVM, 機械学習, 中国語

*Corresponding author
Email addresses: s153400@stn.nagaokaut.ac.jp (Elisa Claire Alemán Carreón), nonaka@kjs.nagaokaut.ac.jp (Hirofumi Nonaka), hiraoka@sun.ac.jp (Toru Hiraoka)

1. はじめに

訪日中国人観光客が増加する中，ホテル産業においては宿泊客のニーズを分析するための市場調査を行うことが重要となっている。このような中で，アンケートやヒアリングなどでのニーズ把握が市場調査の中心になっている。しかしながら，このようなアンケートやヒアリングを用いた調査はコストやリアルタイム性の点で問題があった。一方で，インターネットの普及に伴い，オンラインレビューを通じた口コミが数多く存在する。ユーザが他人の意見を参考にし，購入行動を行うなどの影響がある[1, 2]ため，産業界においても活発に利用されている。多くのオンラインレビューサイトにおけるユーザの評価は大きく，テキスト情報としてのコメント本文と数値で表現される評価点に分かれる。評価点については，構造化された数値情報であり解析に使うことが容易であるため，ユーザの商品・サービスに関する評価指標として利用されている[3, 4, 5]。一方，自然言語処理に基づいたコメント本文の分析も行われている。例えば，情報理論を利用して口コミの感情分析を行う研究[6]や感情分析に基づいて商品の売り上げの予測[7]やランキング[8]を行うなどの研究が行われている。

顧客動向分析においては，何を分析指標として使用するか は極めて重要な観点となる。前述の通り，先行研究では感情分析のみ[6, 8]，もしくは，評価点のみ[3, 4, 5]を使用して顧客動 向を解析した研究が多数ある。このため，これら評価点とコメントの感情評価の関係性を検討する必要がある。仮に，評価点と文書の感情評価の関係性が高いとすれば，数値情報に基づく解析が容易な評価点のみを利用することや，数値情報を教師データに使用することでテキスト情報に基づいた感情分析の学習にも役立つなどメリットは大きい。一方で関連性が低い場合には，評価点とテキスト情報を併用して総合的にレビューを評価する必要がある。あるいは，よりユーザの感情・意見が反映されている方法を選別して評価指標とする必要がある。このように，評価点とコメント本文の感情分析の関係性を調べることはレビューの分析にあたって極めて重要となる。

そこで，本研究では，評価点とコメント本文の感情分析の関係性につい

て調査を行った。本研究では，まず，中国人観光客向けのオンラインホテルレビューサイト*Ctrip*から日本のホテルについて書かれた大量のレビュー文書とその評価点の収集を行った。次に，本研究者らが開発したエントロピーをベースとする素性選択手法を用いてSVMで学習し，Positiveな感情を表す文書とNegativeな感情を表す文書の分類を行った。この感情分類に特徴的な素性ベクトルをエントロピーベースでのキーワード抽出により構築したため，統計量に基づいた言語に依存しない手法を利用して分類を行った。さらに各レビューの各文の感情分類に基づき，感情の数値化をするために満足を表す文の比率と不満を表す比率を算出した。最後に評価点とレビューの感情評価の相互関係性を，スピアマンの順位相関係数，ケンドールの順位相関係数MICより分析した。以下，詳細に説明する。

2. 従来研究

従来のレビュー解析においては商品レビューを対象としたものが主流である。例としてはWord Cloudを適用し消費者が多く使われる単語の抽出を行なった研究[9]や，HowNetという感情辞書を利用して商品のレビューの感情分析を行なった上でマーケット調査に適用した研究[10]がある。一方でホテルのオンラインレビューの影響を定量的に評価しているものとしては，ホテルオンラインレビューが顧客の意欲に特徴的な影響を与えられることを証明した研究がある[1]。他に，売り上げと評価されるレビューのテキスト情報との関係性を示した研究がある[11]。以上の研究では，評価点は考慮したものではなく，評価点とテキスト情報の関係性に着目したものではなかった。

3. 手法

3.1. 前処理

クローリングの段階でCtripの各ホテルの特徴としてURLの構造がID番号で決められていることを利用し，自動的に一つ一つのホテルのページを読み込むことができる。次にスクレイピングを行い，HTMLの特徴を利用して各レビューの文書，ID，評価点など取得し，データベースに保存した。レビューに対しては形態素解析も行った。中国語の形態素解析のツールとして

は，スタンフォード大学のThe Stanford NLP Groupが提供しているStanford Word Segmenter[12]を利用した。

3.2. 感情分析

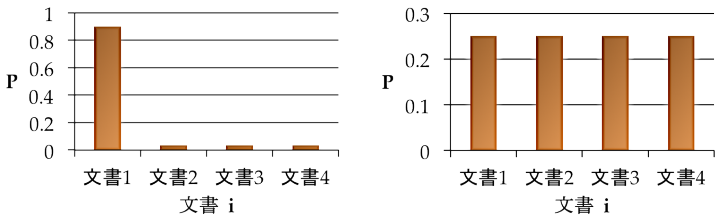
感情分析を行うためにオンラインホテルレビューの収集したデータから標本を抽出し、中国人の研究生３人の協力をいただき、各レビューの各文は満足を表すか、不満店を表現しているかによってタグ付けのPositiveとNegativeの手動分類を行い、教師データを作成した。それから標本のレビューに含まれる語と感情分類タグをSVMに学習させ、母集団となる全データの感情分類を行った。SVMによった感情分類に特徴的な素性ベクトルを後述のエントロピーベースでのキーワード抽出により構築した。以下，その詳細を説明する。

3.2.1. エントロピーベースキーワード抽出

本研究では統計的にもっとも感情と関連しているキーワードを抽出するためにシャノンのエントロピー[13]を利用した。情報理論の分野では，シャノンのエントロピーは信号の情報量の期待値である。すなわち，ある事象の予測不可能性を表す。この概念に基づいて，事象の確率分布の偏在性を表すことができる。例えば，コーパス内のほとんどの文書に含まれる語がどの特定の文書に現れるかの予測は困難になり，その語のエントロピーは高くなる。逆に，特定の文書によく現れ，他の文書にほぼ含まれない大きな文書偏在性を持つ単語を考えると，どの文書に含まれるかという予測の曖昧さは減少し，エントロピーは0に近づく。Fig.1にこの概念を示す。

以上のエントロピーの物理的意味に基づくと，Positiveに関連するキーワードはPositiveが出現する文書セットでエントロピーが大きく，Negativeが出現する文書セットではエントロピーは小さいことが言える。これはNegativeに関連するキーワードでも同様である。そこで本研究ではエントロピーを利用した感情分類における素性選択を行う。まず，レビューに2値の感情分類をタグ付けし，各ワードj が各文書iに含まれる回数を，Positiveな文書の場合にN_{ijP}，Negativeな文書の場合にN_{ijN}を算出する。次に以下に示

す式を利用し，各ワードのPositiveとNegativeの文書に含まれる確率，P_{ijP} (1)とP_{ijN} (2)を算出する。



(a) エントロピーが0に近い　　(b) エントロピーが高い

Figure 1: ワードjが文書iに含まれる確率Pを棒グラフで表示

$$P_{ijP} = \frac{N_{ijP}}{\sum_{i=1}^{M} N_{ijP}} \tag{1}$$

$$P_{ijN} = \frac{N_{ijN}}{\sum_{i=1}^{M} N_{ijN}} \tag{2}$$

以上の値を次の式に代入し，各ワードjのPositiveに関するエントロピーH_{Pj} (3)，各ワードjのNegativeに関するエントロピーH_{Nj} (5)を算出した。ただし，確率が0となる場合については式(4)と(6)に示す極限を利用して(3)と(5)の和に代入する。

$$H_{Pj} = -\sum_{i=1}^{M} [P_{ijP} \log_2 P_{ijP}] \tag{3}$$

$$\lim_{P_{ijP} \to 0+} P_{ijP} \log_2 P_{ijP} = 0 \tag{4}$$

$$H_{Nj} = -\sum_{i=1}^{M} [P_{ijN} \log_2 P_{ijN}] \tag{5}$$

$$\lim_{P_{ijN} \to 0+} P_{ijN} \log_2 P_{ijN} = 0 \tag{6}$$

各語の感情分類毎のエントロピーを算出した後，α の調整を行った。αは評価データのF値がもっとも高くなるものを選択する。最適なαのもとで以

下の(7),(8)を計算し，(7)を満たす語をPositiveのキーワード，(8)を満たす語をNegativeのキーワードと分類した。これを素性としてSVM[14]による学習を行った。評価はk-fold cross validationを用いてF_1[15]により行なった。

$$H_{Pj} > \alpha H_{Nj} \tag{7}$$

$$H_{Nj} > \alpha' H_{Pj} \tag{8}$$

3.3. 相関分析

各レビューに含まれるポシティブとして判定された文を全体で割った率xが評価点yの関連性を測るために次に述べる手法を実験的に利用した。

3.3.1. ピアソン相関係数 r

各レビューに含まれるポシティブとして判定された文i_Pを全体の文の数i_Tで割った率xが評価点yの関連性を測るためその一つとしてピアソンの相関係数r(10)を利用した。以下に式を示す。式(10)に式(9)の値を代入する。ネガティブ率を計算する際には式(9)にネガティブと判定された文i_Nを代入する。

$$x = \frac{i_P}{i_T} \tag{9}$$

$$r = \frac{\sum_{i=1}^{M}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{M}(x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{10}$$

3.3.2. スピアマンの順位相関係数 ρ

各レビューに含まれるポシティブとして判定された文i_Pを全体の文の数i_Tで割った率xが評価点yの関連性を測るため，評価点が順位を表すものだと考える上で順位変数にピアソンの相関係数に基づいたスピアマンの順位相関係数ρ(11)を利用した。以下に式を示す。式(11)に式(9)の値を代入する。

$$r_s = \rho_{rg_X, rg_Y} = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}} \tag{11}$$

3.3.3. ケンダルの順位相関係数τ

スピアマンの順位相関係数のように，ケンダルの順位相関係数は順位を表す値の関係性を調べるために用いるものである。以下に式(12)を示す。ただし，式(12)に式(13)と式(14)を代入する。それぞれに式(9)を代入する。

$$\tau = (K - L)/\binom{n}{2} \tag{12}$$

$$L = \#\Big\{\{i,j\} \in \binom{[n]}{2} \mid \neg(x_i \leqq x_j, y_i \leqq y_j)\Big\} \tag{13}$$

$$K = \#\Big\{\{i,j\} \in \binom{[n]}{2} \mid (x_i \leqq x_j, y_i \leqq y_j)\Big\} \tag{14}$$

3.3.4. MIC

ピアソンの相関係数は線形の関係性のみ抽出できる。一方で，非線形性も含めた二変数間の関連性を分析する手法としてMIC（Maximal Information Coefficient）がある[16]。関連性を分析したい二変数を確率変数と捉えて，その相互情報量に基づき非線形性も含めた変数間の関連性を分析するための指標である。MICとピアソンの係数を比較する複数の例をFig.2に示す。ピアソンの係数は線型的な関係性で0から1までの値となり、傾きの方向によって正の値か負の値か決定される。一方で、MICの場合には非線形的な関係性でも、関係性がある限り0から1の値を利用して総合関係を表すことはFig.2の各例に比較が示される。

MICの算出手順について述べておく。まず，分析対象である変数X,Yに対して，二変数を座標空間上にプロットしたあとで，$a * b$(X方向にa分割，Y方向にb分割)のセルへ分割を行う。その上で二変数それぞれについて，各セルに所属するサンプル点の数を全サンプル数で除すことでセルの存在確率を算出することができる。すなわち，X,Yをセルにおける存在確率をベースとした確率変数と捉える。これによりX,Yに関する相互情報量を計算することができる。

このとき，もとの二変数が線形の関係に限らず非線形の関係にある場合でも，確率変数間の依存性は強くなるため，相互情報量は大きな値をとる。

よってピアソンの相関係数と異なり非線形の関係性を抽出することが可能となる。なお，MICでは，各セルの幅と長さは非等間隔とするため，分割方法は無数に存在する（なお，セルには，最大解像度が存在するものとする）。非線形性の抽出という目的のためにはなるべく，相互情報量を最大化するような分割グリッドを見つける必要がある。MICでは，今，Xを任意のa個，Yを任意のb個に分割したと仮定する。このとき総当たりで$a*b$の分割における相互情報量を最大化する分割を見出す。相互情報量の式(15)を以下に示す。

$$I(X;Y) = \int_Y \int_X p(x,y) \log(\frac{p(x,y)}{p(x)p(y)})\, dx\, dy \tag{15}$$

本研究ではPythonのライブラリminepy[17]を利用してMICを算出した。以下にピアソンの相関係数とMICを比較するFig.2をminepyのAPIのサイトから取得した。

4. 実験結果

以前に述べた手法を利用し，具体的に本研究で行われたことを以下に述べる。

4.1. 前処理

まず，*Ctrip*より，2016年5月から2016年9月までの154万1424件のHTMLファ

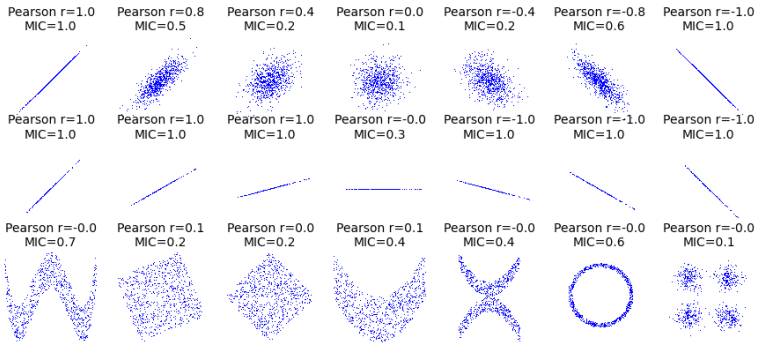


Figure 2: minepyのMICとピアソンのrを様々な場合での比較

イルを収集した。154万1424個HTMLファイルのうち，5938件の日本に存在するホテルのデータを収集した。その中に存在した 4万4912件のレビューのコメント本文をスクレイピングした。これらについて文単位に分割した28万6109文を解析対象とする。また，あわせてレビューの評価点も収集した。

4.2. 感情分類の評価実験

教師データを作成するためには全データからレビューの標本をランダムに抽出し、満足か不満を表すことにより文ごとに分割し、159文のタグ付け作業を手動で行い、教師データとなる文からエントロピーの計算を行った。

教師データに関してエントロピーを計算した後，最適なαの値を求めるため1.0から3.75まで，0.25の刻み幅で評価し，最大のF_1値を持つ素性を選択した。SVMでの学習後，評価データに対して5-Fold Cross Validation($k = 5$)の結果，F_1値が最大となったαに基づきキーワードリストを選択した。さらに，両方のリストを組み合わせ，新たなリストを作成し，同様に5-Fold Cross Validation を行った。それらの評価結果をTable 1に示す。両方を組み合わせた素性が最もF_1値が高くなり、F_1は0.95の高い精度を持ったSVMで全データの分類を行った。Positiveなキーワードは例えば、「情」、「景色」（それぞれ「人懐っこい」、「（良い）景色」を示す）で、Negativeな場合「价格」（高いため不満を示す）の例があった。

Table 1: 5-fold Cross Validation 精度結果

キーワードリスト	C	$F_1\mu$	$F_1\sigma$
Positiveキーワード ($\alpha = 2.75$)	2.5	0.91	0.01
Negativeキーワード ($\alpha' = 3.75$)	0.5	0.67	0.11
組み合わせ	**0.5**	**0.95**	**0.01**

最適なモデルでの学習後，未知データに関して感情分析を行い，前述の「Positive率」と「Negative率」を算出した。これらの値は満足を表す文章と不満足を表現する文章の比率であるため、各文章の感情を表す数値的な係数として扱い、相関分析を行った。

Table 2: 感情分析と評価点の相関分析結果

文書の率	スピアマンのρ	ケンドールのτ	MIC
Positive率	0.161	0.125	0.049
Negative率	-0.149	-0.122	0.0447

4.3. 相関分析

全データにある各レビューのPositive率とNegative率とそのレビューの評価点についてスピアマンの順位相関係数，ケンドールの順位相関係数，MICを利用して関連性分析を行った。その結果をTable 2で示す。

5. 考察

全ての指標において，Positive率，Negative率ともに評価点との関連性が非常に低いことを示した。

よって、コメント本文における感情分析の結果と評価点の関係性は低いことからユーザの意見をレビューより分析する際には，テキストの内容と数値的な評価点の双方についてそしてそれらの違いを検討することが重要である。

従来、この関係性は示されていないにも関わらず、満足度や感情評価の指標として評価点のみが使用されることは多くある。例えば、XieらとBulchand-Gidumalらの研究にはホテル全体の評価点を満足度の代理であると出張しており[3, 4]、Zhouらは満足の要因を調査するために多変数分析を利用したが、従属変数は評価点であった[5]。 このような研究に対して、本研究の結果を踏まえて、評価点のみ、または、感情分析のみを観光客の満足度を測る指標として利用することは適切ではないことが示唆された。

6. おわりに

本研究ではオンラインホテルレビューのテキスト本文の感情分析の結果と評価点の関係性を調べた。感情分析に当たって分類性能が高い（$F_1 = 0.95$)手法を構築し，各口コミのPositiveと分類された文の比率とNegativeと分類さ

れた文の比率を算出した。関係性についてはスピアマンの順位相関係数，ケンドールの順位相関係数とMICを利用した。その結果，いずれも低い値を示した。よって、コメント本文の感情分析の結果と数値情報である評価点の関係性は低いことが明らかになった。よって、より全面的な評価が重要であると考察された。今後はこれらの結果を踏まえて、多言語の情報を利用した解析と比較をした上で感情分析の結果と評価点のどちらがよりユーザの意見を表現しているか調査を行い，レビューの総合的な分析手法の開発を進めていく。

謝辞

中国語での教師データ作成等で支援をいただいた周良遠氏，エルデェンチチグ氏に感謝する。また、本論文は「財団法人日本建設情報総合センター」の支援で行なわれた。

References

[1] I. E. Vermeulen, D. Seegers, Tried and tested: The impact of online hotel reviews on consumer consideration, Tourism Management 30 (1) (2009) 123–127. doi:10.1016/j.tourman.2008.04.008.
URL https://www.sciencedirect.com/science/article/pii/S0261517708000824

[2] B. A. Sparks, V. Browning, The impact of online reviews on hotel booking intentions and perception of trust, Tourism Management 32 (6) (2011) 1310–1323. doi:10.1016/j.tourman.2010.12.011.
URL https://www.sciencedirect.com/science/article/pii/S0261517711000033

[3] K. L. Xie, Z. Zhang, Z. Zhang, The business value of online consumer reviews and management response to hotel performance, International Journal of Hospitality Management 43 (2014) 1–12. doi:10.1016/j.ijhm.2014.07.007.
URL https://www.sciencedirect.com/science/article/pii/S027843191400125X

[4] J. Bulchand-Gidumal, S. Melián-González, B. González Lopez-Valcarcel, A social media analysis of the contribution of destinations to client satisfaction with hotels, International Journal of Hospitality Management 35 (2013) 44–47. doi:10.1016/j.ijhm.2013.05.003.
URL https://www.sciencedirect.com/science/article/pii/S0278431913000728

[5] L. Zhou, S. Ye, P. L. Pearce, M.-Y. Wu, Refreshing hotel satisfaction studies by reconfiguring customer review data, International Journal of Hospitality Management 38 (2014) 1–10. doi:10.1016/j.ijhm.2013.12.004.
URL https://www.sciencedirect.com/science/article/pii/S0278431913001801

[6] R. K. Amplayo, M. Song, An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews, Data & Knowledge Engineering 110 (2017) 54–67. doi:10.1016/j.datak.2017.03.009.
URL https://www.sciencedirect.com/science/article/pii/S0169023X16301525

[7] Z.-P. Fan, Y.-J. Che, Z.-Y. Chen, Product sales forecasting using online reviews and historical sales data: A method combining the bass model and sentiment analysis, Journal of Business Research 74 (2017) 90–100. doi:10.1016/j.jbusres.2017.01.010.
URL https://www.sciencedirect.com/science/article/pii/S0148296317300231

[8] Y. Liu, J.-W. Bi, Z.-P. Fan, Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory, Information Fusion 36 (2017) 149–161. doi:10.1016/j.inffus.2016.11.012.
URL https://www.sciencedirect.com/science/article/pii/S1566253516301580

[9] C. Hargreaves, Analysis of hotel guest satisfaction ratings and reviews: An application in Singapore, American Journal Of Marketing Research 1 (4) (2015) 208–214.

[10] H. Zhang, Z. Yu, M. Xu, Y. Shi, Feature-level sentiment analysis for chinese product reviews, in: 2011 3rd International Conference on Computer Research and Development, Vol. 2, IEEE, 2011, pp. 135–140. doi:10.1109/ICCRD.2011.5764099.

[11] S. Basuroy, S. Chatterjee, S. Ravid, How critical are critical reviews? the box office effects of film critics, star power, and budgets, Journal Of Marketing 67 (4) (2003) 103–117. doi:10.1509/jmkg.67.4.103.18692.

[12] P. Chang, M. Galley, C. Manning, Optimizing Chinese word segmentation

for machine translation performance, in: Proceedings of the Third Workshop On Statistical Machine (Statmt '08), Columbus,Ohio,USA, 2008, pp. 224–232.
URL http://nlp.stanford.edu/pubs/acl-wmt08-cws.pdf

[13] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (3) (1948) 279–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

[14] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/bf00994018.

[15] D. Powers, Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation, Journal Of Machine Learning Technologies 2 (1) (2011) 37–63.
URL http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf

[16] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, P. C. Sabeti, Detecting novel associations in large data sets, Science 334 (6062) (2011) 1518–1524. arXiv:https://www.science.org/doi/pdf/10.1126/science.1205438, doi:10.1126/science.1205438.
URL https://www.science.org/doi/abs/10.1126/science.1205438

[17] D. Albanese, M. Filosi, R. Visintainer, S. Riccadonna, G. Jurman, C. Furlanello, minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers, Bioinformatics 29 (3) (2012) 407–408. arXiv:https://academic.oup.com/bioinformatics/article-pdf/29/3/407/17105876/bts707.pdf, doi:10.1093/bioinformatics/bts707.
URL https://doi.org/10.1093/bioinformatics/bts707