

# レビューテキストの書き手の評価視点に対する評価点の推定

ZHANG Bo

白井 清昭

北陸先端科学技術大学院大学 先端科学技術研究科

{hiroshichang,kshirai}@jaist.ac.jp

## 1 はじめに

オピニオンマイニングは、製品やサービスといった特定の対象に対するユーザのコメントを分析し、その対象の評判を明らかにする技術である。オピニオンマイニングの結果をユーザに提示する方法のひとつとして、ユーザの対象に対する評価を点数で表わした評価点を示すことが挙げられる。評価点の提示は、ユーザの評価を素早く直観的に把握できるという利点がある。また、複数のユーザの評価点の平均や分布を示すことにより世間の評判を知ることできる。

一般に、評価対象には様々な評価視点が存在する。例えば、ホテルを評価の対象とするとき、サービス、部屋の広さ、スタッフの対応、といった評価視点がある。評価点をユーザに提示する際には、対象そのものの評価点ではなく、評価視点毎に評価点を提示する方が望ましい。

本研究では、ユーザのレビューテキストに対して、評価視点毎にユーザの評価点を推測する手法を提案する [5]。一般に、レビューテキストには、対象に対する全ての評価視点に関するコメントが含まれているわけではなく、ユーザが言及していない評価視点も存在する。提案手法では、まず評価視点に関する文をレビューテキストから抽出し、その文から評価点を推測する点、また評価視点に関する文がレビューテキストに存在しないときは評価点の推測を行わない点に特徴がある。

## 2 関連研究

小林らは、ポジティブ・ネガティブ分析機能を備えたユーザレビュー集約管理システムを提案した [2]。大手ECサイトにおけるレビューテキストに対し、ナイーブベイズ分類器を用いて「ポジティブ」「ネガティブ」「どちらにも属さない」のいずれかに属する確率を推定した。この確率は評価点とみなすこともできる。ただし、確率を推定するのはレビュー全体に対してであり、評価視点毎の評価点を推定しているわけではない。

外山らは、与えられたレビュー文書集合に対し、文書集合全体ならびに各レビューの分散表現を生成し、レビューにおける複数の評価視点に対する評価点を同時に

推測するニューラルネットワークを学習する手法を提案した [4]。楽天トラベル<sup>1</sup>のレビューを用いた評価実験の結果、評価点の正答率は0.5021となり、従来手法と比べて0.0189改善した。Phamらは、least square methodを適用し、ホテルのレビューを対象に、評価視点毎に評価点を推測する手法、ならび最も重要な評価視点を推測する手法を提案した [3]。評価点は、レビューの特徴ベクトルと、それと同じ次元を持つ重みベクトルの内積で推定する。重みベクトルは訓練データにおける正解の評価点との誤差が最小になるように学習する。これらの手法では、全ての評価視点に対する評価点を常に推測し、レビュー内に評価視点に関するユーザの意見が書かれているか否かを考慮していない。しかしながら、レビューサイトにおいては、レビューに評価視点に関する意見が書かれていないのにも関わらず、その評価視点の評価点が与えられていることがある。評価視点に関する具体的な意見が存在しないレビューからその評価点を推定するのは困難である。また、このようなレビューを訓練データとして評価点を推測するモデルを学習するのは適切ではない。本研究では、まず評価視点に関する意見が存在するかを判定し、それが存在するときのみ評価点を推測する。

## 3 提案手法

### 3.1 概要

本研究ではホテルのレビューを対象とする。また、評価点を推測するためのデータとして楽天トラベルに投稿されたユーザレビューを用いる。楽天トラベルでは、ユーザは、ホテルに関するコメントを投稿し、「サービス」「立地」「部屋」「設備・アメニティ」「風呂」「食事」および「総合」のそれぞれの観点から評価点を付けることができる。評価点は1から5までの整数である。本研究では「総合」を除く6つをホテルの評価視点と定義し、それぞれに対する評価点を推測するモデルを学習する。提案手法の処理の流れを図1に示す。

まず、レビューテキストから、それぞれの評価視点からホテルを評価している文を抽出する。以下、これを

<sup>1</sup><http://travel.rakuten.co.jp>

表 1: 評価視点の関連語

サービス	サービス, 接客, フロント, 受付, 対応, スタッフ, 従業員, 荷物, 女将, 応対, うけつけ, 受け付け, 係員, 係+, の, 係り+, の, 担当+, の, 宿+, の+, 方, 施設+, の+, 方
立地	立地, 位置, 駅, 繁華街, コンビニ, 距離, アクセス, 交通, 利便, JR, 新幹線, 地下鉄, タクシー, 場所, ロケーション, 市街地, 街+から, 近く+便利, 近く+で+便利
部屋	部屋, ルーム, 室内, 客室
設備・アメニティ	設備, アメニティ, アメニティー, 空調, エアコン, ベッド, 冷蔵庫, エレベーター, インターネット, Wi-Fi, LAN, コンセント, 便座, ウォシュレット, 暖房, 布団, 枕, シャンプー, リンス, ソープ, 洗浄機, 洗浄器, タオル, ドライヤ, ドライヤー, スリッパ, 洗顔, 浴衣
風呂	風呂, ふろ, シャワー, バス, 温泉, 浴室, 浴場, お湯, 男湯, 女湯, 入浴, 露天風呂, 浴槽, 源泉
食事	食事, ご飯, 御飯, ごはん, 料理, 朝モーニング, 朝食, 昼食, 晩食, 夕食, レストラン, メニュー, パイキング, 朝飯, 昼飯, 夕飯, ディナー, おかず, 【日本語語彙大系における料理名】

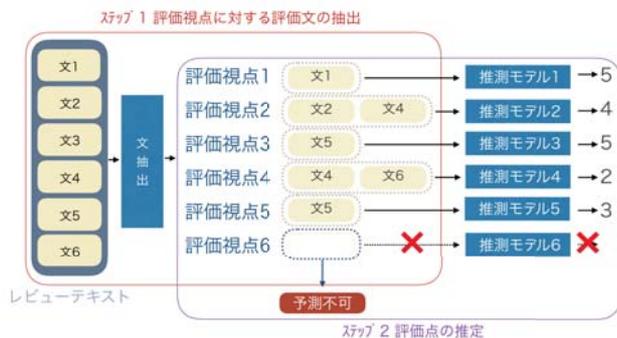


図 1: 提案手法の概要

評価視点の「評価文」と呼ぶ。評価文がひとつも得られなかったとき、その評価視点に対する評価点は予測できないため、「不明」と判定する。次に、抽出した評価文を入力とし、評価点を推測するモデルを適用し、1～5の評価点を出力する。評価点を推測するモデルは、評価視点毎に、楽天トラベルにおけるレビューと評価点を訓練データとして、教師あり機械学習により学習する。学習時には、同様にレビューから評価視点の評価文を抽出し、評価文を抽出できないレビューは訓練データから除外する。

### 3.2 評価文の抽出

まず、それぞれの評価視点について、その関連語のリストを用意する。評価視点の関連語とは、文がそれを含むとき、その文が対象の評価視点について言及していると考えられるキーワードである。レビューテキストから、評価視点の関連語を含む文を評価文として抽出する。

評価視点の関連語は以下の4つの方法で用意する。  
 人手による選別 以下の手続きで関連語を獲得する。(1) 初期の関連語の集合を評価視点を表わす語(「サービス」「立地」など)とする。(2) 関連語を含むレビューを抽出する。(3) レビュー集合に含まれる出現頻度上位100件の名詞を獲得する。(4) その中から関連語にふさわしい

ものを人手で選別する。(5) 上記2,3,4のステップを繰り返す。

単語 **n-gram** 上記の手続きで獲得される関連語はキーワードであるが、複数の単語の並びが評価視点を示唆することがある。そこで、関連語としてふさわしい単語 **n-gram** を人手で追加した。例えば、評価視点「サービス」については「宿+の+方」、評価視点「立地」については「近く+で+便利」などを追加した。

シソーラスの利用 予備調査の結果、「食事」の評価視点については、様々な料理名が関連語とみなせることがわかった。そのため、日本語語彙大系 [1] における「料理 (0847)」というカテゴリに登録されている単語を評価視点「食事」の関連語とした。

並列表現を手がかりとした関連語の獲得 ホテルには様々な設備があり、また顧客に提供するアメニティの種類も多いため、「設備・アメニティ」という評価視点の関連語は多岐にわたる。ここでは、並列表現を手がかりとし、関連語の数を増やすことを試みる。具体的には、式 (1) に示すパターンを用いる。

$$A \text{ (と|や|また|および) } B \quad (1)$$

A(もしくは B) が既に獲得されている「設備・アメニティ」の関連語のとき、B(もしくは A) に該当する単語を人手でチェックし、適切と判定した語を新しい関連語として獲得する。

評価視点の関連語の一覧を表 1 に示す。

### 3.3 評価点の推定

本研究では、評価点を推定するモデルを L2 正則化ロジスティック回帰によって学習する。学習ツールとして LIBLINEAR<sup>2</sup> を用いた。学習のパラメタはデフォルト値を選択した。

評価視点に関する評価文から素性ベクトルを作成する。学習に用いる素性は以下の通りである。

<sup>2</sup><https://www.csie.ntu.edu.tw/~cjlin/liblinear/>

表 2: 評価文の抽出の評価

	精度	再現率	F 値	件数
サ	90.20	77.67	83.46	101,725 (37.3%)
立	94.34	73.10	82.39	80,679 (29.6%)
部	85.17	87.45	86.29	125,783 (46.1%)
設	92.42	61.68	73.99	56,203 (20.6%)
風	93.48	75.41	83.48	100,502 (36.9%)
食	98.19	81.23	88.91	166,017 (60.9%)

サ=サービス, 立=立地, 部=部屋,  
設=設備・アメニティ, 風=風呂, 食=食事

● 評価語

文中に含まれる評価語を素性とする。本研究では、日本語評価極性辞書(用言編と名詞編)<sup>3</sup>に登録されている単語を評価語と定義する。ただし、評価語に否定表現が続くとき、「(評価語)+否定」を素性とする。

● 自立語

評価文を MeCab<sup>4</sup>で形態素解析し、名詞、動詞、形容詞、副詞を抽出して素性とした。ただし、評価語として抽出された自立語は除く。また、評価語と同様に、否定表現が続くときは「(自立語)+否定」を素性とする。

### 4 評価実験

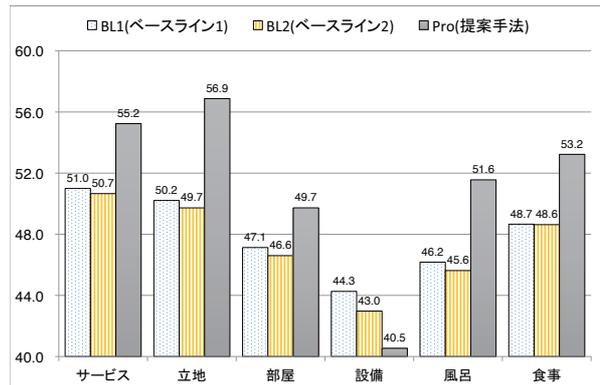
楽天トラベルから、ユーザレビューと6つの評価視点に対する評価点の組を収集した。レビューの中には6つの評価視点の全てに評価点が付与されていないものもある。このようなレビューを除外し、272,665件のレビューを得た。これを実験データとして使用する。

まず、評価視点に関する評価文の抽出手法を評価する。ランダムに選択した500件のレビューに対し、6つの評価視点毎に、評価文とみなせる文を手で付与した。これを正解とみなしたときの評価文抽出の精度、再現率、F値を測った。結果を表2に示す。精度は、「部屋」の85%を除いて90%を超えており、十分に高いことがわかる。一方、再現率は61%から87%となり、改善の余地がある。これは、評価視点の関連語が不足していたためと考えられる。特に、「設備・アメニティ」の再現率が低い。設備やアメニティには様々なものがあり、これを示唆する関連語の数も多いが、提案手法ではその全てを網羅的に収集できていない。

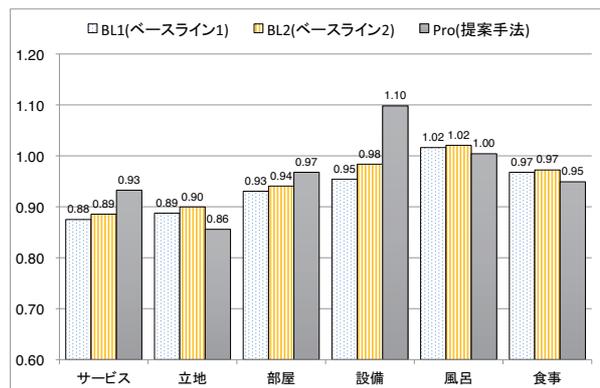
表2の「件数」は、評価文を1つ以上取り出すこと

<sup>3</sup><http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>

<sup>4</sup><http://taku910.github.io/mecab/>



(a) 正解率



(b) RMSE

図 2: 評価点推測の評価

のできたレビューの件数、およびその全体(272,665件)に対する割合である。「設備・アメニティ」のレビュー件数が最も少ないが、評価文抽出の再現率が低いことが原因のひとつと考えられる。ただし、レビューの中には評価視点に関する意見が書かれていない場合もあるため、全てのレビューから評価文を取り出す必要はない(抽出率を100%に近づける必要はない)ことに注意していただきたい。

次に、評価点の推定手法を評価する。5分割交差検定により、評価点を推定する提案手法を評価した。ただし、1つ以上の評価文を抽出できたレビューのみを評価の対象とした。評価基準は、評価点の正解率および平均二乗誤差(Root Mean Square Error; RMSE)とした。RMSEは真の評価点と推定した評価点の誤差であり、値が低いほど良い結果が得られたことを表わす。結果を図2に示す。

ベースライン1 (BL1)は、評価視点に関する評価文を抽出する処理を省略し、レビューに含まれる全ての文から評価点を推測したときの結果である。すなわち、評価視点に関する意見がレビュー文に含まれていない場合でも評価点の推測を試みる。ベースライン2 (BL2)は、BL1と同様に評価視点に関する評価文を抽出する処理

を省略するが、訓練データとして提案手法と同じ数のレビュー文を用いたときの結果である。表2に示すように、提案手法では訓練データの数60%以下に減少するが、一般に機械学習では訓練データの量が多いほど正解率は高くなる。評価文を抽出することの効果厳密に調べるため、同じ量の訓練データを用いた手法をBL2とした。なお、BL2の学習に用いるデータはランダムに選択した。

提案手法の正解率は、「設備・アメニティ」以外の5つの評価観点についてベースラインを上回った。評価観点についての評価文が存在しないレビューを学習やテストの対象としないことにより、評価点を予測するモデルの性能が向上したことがわかる。「設備・アメニティ」については正解率が低下したが、学習に用いたレビューの件数が少ないことが原因のひとつと考えられる。また、訓練データ量が同じBL2よりも劣るが、BL2ではレビュー中の全ての文から素性ベクトルを作成するのに対し、提案手法では評価文のみから素性ベクトルを作るため、評価点を推測するのに十分な情報が得られなかったためと考えられる。「設備・アメニティ」の評価文ではない文も、総合評価の評価点の推測には有効であり、かつ総合点と「設備・アメニティ」の評価点には正の相関関係(総合点が高いと設備の評価も高い)があると思われるため、有効な情報を含むと考えられる。提案手法では評価観点の評価文以外の文を利用しないが、これを評価点推定モデルで利用することは今後の課題である。なお、BL1はBL2と比べて正解率が高いことから、訓練データの量は正解率に影響を与えることが確認された。

RMSEの結果を見ると、正解率とRMSEとで手法の優劣が一致しない場合がある。具体的には、「サービス」と「部屋」については、正解率については提案手法はベースラインを上回るが、RMSEについては逆に高く(結果が悪く)なっている。つまり、真の評価点よりも大きく異なる点を予測する場合が多い。この原因を調べるため、真の評価点が5でシステムが推測した評価点が1のとき、あるいは真の評価点が1でシステムが推測した評価点が5のときのレビューをランダムに50件選択し、誤りの要因を分析した。誤りの要因とそれに該当するレビューの件数を表3に示す。評価点の推定を大きく誤る要因は多岐にわたることがわかる。特に多い誤りの原因とその対応策を述べる。評価文が短いため評価点を推定するための情報が少ないという問題については、前述のように評価文以外の文を利用することが考えられる。逆に、評価文が長くて複雑な場合や、評価語が評価観点以外の対象に対する意見を示している場合については、評価文に含まれる全ての単語を素性とするのではなく、

表 3: 誤り分析

	サービス	部屋
評価文が短く、評価点を推測するための情報が不足している	5	4
評価文が長く複雑で、ノイズとなるような素性が抽出されている	4	8
評価文抽出の誤り	2	5
レビューの内容とユーザが付与した評価点が矛盾している	3	0
評価語が辞書に登録されていない	8	3
評価語が別の対象もしくは他の評価観点に対する意見を示している	5	6
否定の判定の誤り	5	5
評価文が肯定的・否定的意見の両方を含む	2	3
形態素解析の誤り	0	1
原因不明	16	15

評価観点の評価に関連のある単語のみを素性とする工夫が必要である。評価語辞書に登録されている単語が不足しているという問題については、評価観点に特有の評価語を自動的に獲得し、評価辞書を拡充する手法が有効であろう。また、否定の判定手法も洗練する必要があることがわかった。

## 5 おわりに

本論文では、ホテルのレビューを対象に、まず評価観点に関する意見文を抽出し、その意見文を入力として評価観点に関する評価点を自動的に推定する手法を提案した。評価実験の結果、提案手法のアプローチが有望であることを確認した。今後は、誤り分析の結果を踏まえて、評価文を抽出するアルゴリズムを洗練したり、評価観点に関連のある学習素性を選別することが課題となる。

## 参考文献

- [1] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系 CD-ROM 版. 岩波書店, 1999.
- [2] 小林亮, 鈴木浩, 服部哲, 速水治夫. ポジティブ・ネガティブ分析を備えたユーザレビュー集約管理システムの提案. 情報処理学会研究報告, Vol. 2013-GN-87, No. 1, 2013.
- [3] Duc-Hong Pham, Anh-Cuong Le, and Thi-Kim-Chung Le. A least square based model for rating aspects and identifying important aspects on review text data. In *Proceedings of the 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science*, pp. 265–270, 2015.
- [4] 外山洋太, 三輪誠, 佐々木裕. 文書・文間及びカテゴリ間の関係を考慮したレーティング予測. 言語処理学会第22回年次大会発表論文集, pp. 158–161, 2016.
- [5] Bo Zhang. レビューテキストの書き手の評価観点に対する評価点の推定. 修士論文, 北陸先端科学技術大学院大学, 2017.