

# 宿泊施設レビューにおける評判の要約と評価基準の重要度を用いたランキング

小宮山 博之<sup>†</sup> 太原 育夫<sup>‡</sup>

東京理科大学大学院 理工学研究科 情報科学専攻<sup>†</sup>

東京理科大学 理工学部 情報科学科<sup>‡</sup>

## 1. はじめに

旅行を計画するとき考えるべき要素の一つとして宿泊施設の選択がある。その際、私たちはレビューを利用することがある。それらは他の利用者が記述したままの文章で提示されていることが多いので、レビュー中から欲しい情報を的確に得るために必要とする時間と労力は少なくない。

一方、意見を含んだテキストから評判を取得する研究が盛んである [1]。その多くにおいて、評判は小林ら [2] が提案した評判情報と呼ばれる (対象, 項目, 評価) の 3 つ組の形で取得される。杉木らは入力したクエリから、レビュー文書を対象に評判情報を抽出し提示するシステムを提案している [3][4]。しかし、このシステムは出力内容をクエリに大きく依存しているため、クエリに含まれなかった条件についてユーザーが不満を持つ可能性がある。また、宿泊施設群のランキングはユーザーの要求とレビュー文書中の評判情報との形態素一致数のみに基づいており、各要求間の重要度の差を考慮していない。

本稿では、ユーザーがレビュー文書中から必要な情報を取得する時間と労力を軽減することを目的とした手法について述べる。ただし、対象とするレビュー文書は杉木らと同様楽天トラベル<sup>\*1</sup>のものを用いる。

## 2. システムの概要

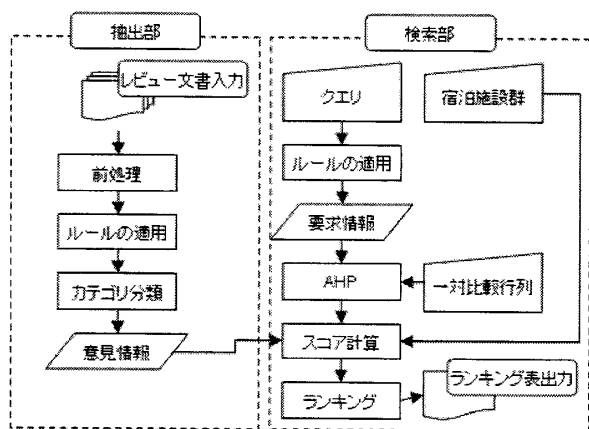


図 1: システムのフローチャート

システムの概要を図 1 に示す。システムは大きく抽

Summarizing of Accommodation Reviews by Opinion Extraction and Ranking Based on Importance of Criteria

<sup>†</sup>Hiroyuki KOMIYAMA

Graduate School of Tokyo University of Science

<sup>‡</sup>Ikuo TAHARA

Tokyo University of Science

出部と検索部の 2 つに分けられる。抽出部はレビュー文書中から評判情報 (この場合、意見情報) を抽出し、出現回数を数えてカテゴリに分ける。検索部はクエリ中の評判情報 (この場合、要求情報) の抽出およびスコアに基づく各宿泊施設のランキングを行う。最後にランキングされた宿泊施設群とそれぞれの宿泊施設における意見情報をまとめたものを出力する。

## 3. 評判情報の抽出

評判情報は各文の係り受け関係 (「主述の関係」または「修飾・被修飾の関係」) にある 2 要素に対して、次の 6 つのルール [4] を適用することで抽出する。ただし、 $O, X, Y$  はそれぞれ対象, 項目, 評価にあたり、「 $\rightarrow$ 」は係り受け関係を表す。「 $\Rightarrow$ 」の左辺のパターンから右辺の評判情報が生成される。

- (1) (名詞  $x$ +は/が/も) $\rightarrow$ (動詞/形容詞/サ変名詞) $_Y$   
 $\Rightarrow (O, X, Y)$
- (2) (形容詞  $_Y$ (接続助詞を含まない)) $\rightarrow$ (名詞  $x$ +は/が/も/を/に/だ/です) $\Rightarrow (O, X, Y)$
- (3) (名詞  $x_1$ +の (助詞)) $\rightarrow$ (名詞  $x_2$ +は/が/も) $\rightarrow$ (動詞/形容詞/サ変名詞) $_Y \Rightarrow (O, X_1-X_2, Y)$
- (4) (名詞  $x$ +は/が/も) $\rightarrow$ (動詞/形容詞/サ変名詞) $_{Y_1}$   
 $\rightarrow$ (動詞/形容詞/サ変名詞) $_{Y_2}$   
 $\Rightarrow (O, X, Y_1), (O, X, Y_2)$
- (5) (名詞  $x_1$ +の (助詞)) $\rightarrow$ (名詞  $x_2$ +は/が/も) $\rightarrow$ (動詞/形容詞/サ変名詞) $_{Y_1}$   
 $\rightarrow$ (動詞/形容詞/サ変名詞) $_{Y_2} \Rightarrow (O, X_1-X_2, Y_1), (O, X_1-X_2, Y_2)$
- (6) (動詞/形容詞/サ変名詞) $_Y \Rightarrow (O, -, Y)$

## 4. 意見情報のカテゴリ分類

抽出された意見情報は、あらかじめ作成した辞書によって「項目」をもとにカテゴリに分類する。例えば、意見情報の「項目」が「コンビニ」や「地下鉄」などのときは「立地」のカテゴリに分類し、「項目」が「ベッド」や「洗面台」のときは「部屋」のカテゴリに分類する。カテゴリは楽天トラベルの評価項目にしたがって、サービス、立地、部屋、設備・アメニティ、風呂、食事、その他の 7 つと総合 (「項目」がない場合) の計 8 つとした。

<sup>\*1</sup>楽天トラベル お客様の声

[http://travel.rakuten.co.jp/auto/tabimado\\_bbs\\_top.html](http://travel.rakuten.co.jp/auto/tabimado_bbs_top.html)

## 5. 宿泊施設の検索

ユーザは自然文による要求(宿泊施設の条件)を入力し、比較対象とする宿泊施設群を指定する。本稿では、ユーザが「料理がおいしくて、部屋がきれいな(宿泊施設)」というような項目と評価の組によるクエリを入力することを仮定している。システムは評判情報の抽出と同様にクエリを項目と評価の組に分割し要求情報を作成する。

## 6. 要求を考慮したスコアリング

### 6.1 AHP を用いた重要度算出

各評価基準に重要度を与えるために AHP (Analytic Hierarchy Process, 階層分析法) を利用する。AHP は複数の評価基準に対して、各評価基準同士について相対的な重要度を与えることにより、選好を確実に明示できる [5]。

### 6.2 重要度を用いたスコアリング

評価基準  $i (i = 1, \dots, k)$  の重要度  $w_i$  を要素とするベクトル (重みベクトル)  $w = (w_1, w_2, \dots, w_i, \dots, w_k)$  を用いて、A ホテルのスコア  $S_A$  を次式で計算する。

$$S_A = \frac{1}{\theta_A + 1} \sum_{i=1}^k w_i n_i \quad (1)$$

$$\theta_A = \frac{r_A - r_{\min}}{r_{\max} - r_{\min}} \quad (2)$$

ここで、 $n_i$  は要求情報  $i$  と意見情報との一致数、 $r_A, r_{\max}, r_{\min}$  はそれぞれ A ホテルの宿泊料金、比較対象の全宿泊施設の宿泊料金の最大値、最小値である。 $\theta_A$  は全比較対象の宿泊料金を  $[0, 1]$  の範囲に正規化したときの A ホテルの値である。

ユーザの指定した全宿泊施設に対してスコアを求め、降順にランキングする。

## 7. 評価実験

### 7.1 実験 1: 意見情報の要約

10 件のレビュー文書とその要約を用意し、5 名の被験者にそれぞれどちらを使いたいか 5 段階 (5 に近い方が要約を支持) で評価してもらった。結果は図 2 の通りである。

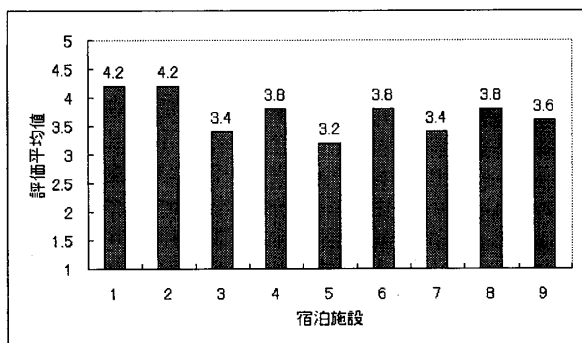


図 2: 被験者 5 人の評価平均

### 7.2 実験 2: ランキングの妥当性

10 件の宿泊施設の情報とレビュー文書を 5 名の被験者に読んでもらい、10 個の検索クエリ (図 3) にそれぞれ一番適していると思う宿泊施設を各被験者に選んでもらう。ベースライン (杉木らのシステム) と提案手法のランキングでそれらの平均順位を比較した。結果は図 4 の通りである。

1. 駅が近くて、部屋が広くて、風呂が広い
2. コンビニが近くて、冷蔵庫がある
3. 朝食がおいしくて、フロントの対応が良い
4. 風呂が広くて、ベッドの寝心地が良い
5. コンビニが近くて、スターバックスが近い
6. 駅が近くて、フロントの対応が良い
7. 冷蔵庫があって、清潔感がある
8. 立地が良くて、チェックアウトが遅い
9. 部屋が明るくて、清潔感がある
10. 部屋が広くて、インターネットが使える

図 3: 検索クエリ

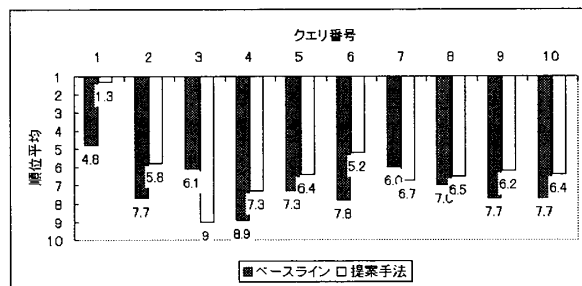


図 4: 各宿泊施設の評価平均

## 8. おわりに

宿泊施設の決定を支援するために、評判の要約とユーザの検索クエリをもとに宿泊施設にランキングを行うシステムを提案し、被験者アンケートを用いた実験により有効性を確認した。

## 参考文献

- [1] 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.
- [2] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “テキストマイニングによる評価表現の収集,” 情報処理学会研究報告, 2003-NL-154, pp.77-84, 2003.
- [3] 杉木健二, 松原茂樹, “クエリの主観性に頑健な商品システム,” 第 6 回情報科学技術フォーラム (FIT2007), No.E-056, pp.271-274, 2007.
- [4] 杉木健二, 松原茂樹, “問合せの主観性にロバストな商品検索システム,” 言語処理学会第 13 回年次大会, pp.748-751, 2007.
- [5] 刀根薫, “ゲーム感覚意思決定法-AHP 入門-,” 日科技連出版社, 1986.