

評価分析における評価対象のカテゴリ分類

- 楽天トラベル「お客様の声」の分析を例として -

Nguyen.Pham Thanh.Thao[†] 林 貴宏^{††} 尾内 理紀夫^{††} 西岡 悠平^{†††} 竹中 孝真^{†††}

森 正弥^{†††}

[†] 電気通信大学大学院 電気通信学研究科 情報工学専攻 & 楽天技術研究所 客員研究助手

^{†††} 楽天技術研究所

^{††} 電気通信大学

あらまし 近年、レビューサイト、お客様アンケート、個人ブログ内の意見など、評価情報を扱う研究が活発に行われている。本研究は相互情報量を用いた評価対象のカテゴリ分類手法を提案する。評価対象がカテゴリ分類されることで、ユーザは目的に応じて効率よく評価情報から評価対象をブラウジングできる。本手法は、学習データを必要としない学習不要な分類手法であるため、事前知識を必要とせず様々な種類の評価情報を扱うことができる。本手法は、すべての名詞句を評価対象候補とするため、出現頻度が低い単語も分類可能である。Web で公開されている楽天トラベル「お客様の声」に対し、カテゴリ分類実験を行った。各カテゴリの上位 150 語を評価した結果、上位の単語が適切に分類され、かつ、出現頻度が低い単語でも評価対象として抽出され、適切に分類されることを確認した。さらに、ユーザが設定した各カテゴリ以外に、ゴミカテゴリを設定することにより、評価対象とならない単語をそこに分類することができた。

キーワード 評判分析、評判要約、評価対象のカテゴリ分類、ゴミカテゴリ

Opinion Features Categorization in the Opinion Analysis

- Empirical Result as Applied to the “Customer Voice” from “Rakuten Travel” -

Thanh.Thao NGUYEN.PHAM[†], Takahiro HAYASHI^{††}, Rikio ONAI^{††}, Yuhei NISHIOKA^{†††},

Takamasa TAKENAKA^{†††}, and Masaya MORI^{†††}

[†] Department of Computer Science, Graduate School of Electro-Communications & Research Assistant at Rakuten Institute of Technology

^{†††} Rakuten Institute of Technology

^{††} University of Electro-Communications

Abstract Mining or summarizing opinion expressed in blogs, review sites etc. is a new and potential research trend. In this paper, we propose a mutual-information-based technique to organize opinion features into categories. Organizing opinion features into categories helps users browse and find what they are interested in more easily. Our technique does not require neither learning data nor learning phase; and it can be applied to many kinds of review information without requirement for the prior domain knowledge. We treat all noun phrases as candidate of opinion features, so that even words with low frequency can be categorized. Our experiment was conducted for data crawled from the “Rakuten Travel” page, a review site for hotel guests. The results from evaluating the top 150 words of each category showed that our technique is effective. Also, beside the purpose categories set up by users, we could gather non-feature words into a “trash category” properly.

Key words opinion analysis, opinion summarization, opinion features categorization, trash category

1. はじめに

近年、レビューサイト、お客様アンケート、個人ブログ内の意見など、評価情報を扱う研究が活発に行われている。製品又はサービスの評判を抽出し、整理し、提示することは企業にとって、一般ユーザにとって利点となる。様々な研究目的があるが、評価情報から、評価対象が何か（何が評価されたか）を抽出し、さらに、それらをどう整理するかは重要な問題である。

既存研究は、予め作成しておいた評価対象辞書を用いて評価対象を抽出する場合や、高い出現頻度の名詞を評価対象として抽出する場合が多かった[1]～[8]。しかし、これら方法は、本来、評価対象として抽出されるべき単語が抽出されないことがある。また、多くの既存研究では、評価対象の抽出に主眼が置かれており、評価対象の整理法、提示法はほとんど触れられていなかった。

本研究は相互情報量を用いた評価対象のカテゴリ分類手法を提案する。評価対象がカテゴリ分類されることで、ユーザは目的に応じて効率よく評価情報から評価対象をブラウジングできる。本手法は、学習データを必要としない学習不要な分類手法であるため、事前知識を必要とせず様々なジャンルの評価情報を扱うことができる。

以下、本稿では、2章に、まず本手法の概要を説明する。3章では、前処理である複合語句の抽出と係り受け関係抽出について述べる。4章は提案手法の分類手法を説明する。5章は実験結果と考察を述べる。6章は関連研究を挙げ、7章でまとめる。

2. 提案手法の概要

本手法の基本的な考え方は同じ特徴的な単語群と共起する名詞句を同じカテゴリに分類することにある。例えば、“食べる”、“おいしい”、“飲む”などと共起する名詞句はカテゴリ「食事」のメンバーである可能性が高いと考えられる。ここでいう特徴的な単語群はそのカテゴリを特徴づけることから特徴語と呼ぶ。本手法はすべての名詞句を評価対象候補とするため、1つのカテゴリがそのカテゴリの中身となる名詞句集合と特徴語集合からなる。

図1は本手法の処理手順を示している。まず、複合語と係り受け関係が抽出される。そして、人手によるカテゴリの初期設定が行なわれる。その次は自動分類プロセスである。

初期設定として、人手により、出現頻度が高い名詞句を各カテゴリの種語として、少数選択すると、以降は、システムが自動的に各カテゴリの名詞句集合と特徴語集合を同時に拡張しながら、未分類の名詞句を分類していく。具体的には、ある名詞句Nの分類を考える時に、名詞句Nの各カテゴリに対する配属スコアが計算され、一番高いスコアが出すカテゴリに、名詞句Nが分類される。名詞句NとカテゴリCとの間の配属スコアは名詞句NがカテゴリCの要素となる相応しさを表し、このスコアが名詞句NとカテゴリCの特徴語集合との相互情報量で計算される。

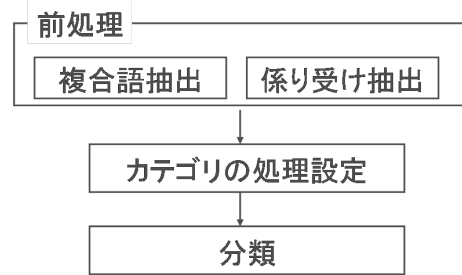


図1 処理の手順

3. 前処理

3.1 複合語の抽出

本研究は評価対象のほとんどが名詞句であることを想定し、各名詞句と係り受け関係を持つ単語（名詞句、動詞句、形容詞）を特徴語として扱う（既存研究では、動詞のみか形容詞のみかを特徴語として使っていることが多い[8]）。形容詞と単一の品詞情報からなる動詞（例：“食べる”、“行く”など）は普通の形態素解析ツールで容易に抽出できるが、名詞句と動詞句の抽出は少し工夫が必要である。本研究は形態素解析器 Mecab^(注1)を使って、各品詞単語を以下のように抽出した（本稿に使用する品詞情報はAPI品詞体系^(注2)を参考している）。

- 形容詞としては「自立-形容詞」と「非自立-形容詞」の両方が特徴語として抽出される。例えば、

例1：リンゴが 赤い

例2：リンゴが 赤く ておいしい

Mecabは例1と例2の「赤い」をそれぞれ「自立-形容詞」、「非自立-形容詞」として解析している。

- 動詞句としては「名詞-サ変連続」に続き、「自立-動詞」が出現するパターンを抽出する。

例えば、“更新できる”、“利用する”などがある。

- 名詞句は3つの品詞「名詞-一般」、「名詞-サ変接続」と「名詞-接尾 一般」の組合せで抽出される。

表1に抽出されたパターンの中から例として6パターンを示している。

3.2 係り受け関係抽出

本研究は各名詞句と係り受け関係を持つ単語（名詞句、動詞句、形容詞）を特徴語として扱うので、特徴語を生成するために、係り受け関係の抽出が必要である。係り受け関係は文単位で行われ、名詞句とその名詞句に係る前文脈、又は、後文脈の特徴語をペアとして抽出する。本手法では、係り受け分析器 Cabocha^(注3)を利用して、出現する頻度2以上の全て係り受けペア（名詞句-特徴語）を抽出した。

(注1)：形態素解析器 Mecab

<http://mecab.sourceforge.net/>

(注2)：IPA 品詞体系

<http://www.unixuser.org/~euske/doc/postag/index.html>

(注3)：日本語係り受け解析器 Cabocha

<http://chasen.org/~taku/software/cabocha/>

表 1 名詞句の抽出パターンの例

パターン と例の名詞句
名詞-一般・名詞-一般: 温水・プール, ダブル・ルーム, ホテル・周辺, レジャー・チケット, ユニット・バス, 無料・インターネット, 格安・プラン, ビジネス・ホテル, 立地・条件, 電子・レンジ, ミニ・キッチン, ルーム・キー, サラダ・バー, 大通り・公園, ...
名詞-サ変接続・名詞-接尾-一般: 従業・員, 乾燥・室, 駐車・場, 安心・感, 調理・法, 宿泊・者, 喫煙・室, 宿泊・費, 飲食・店, 利用・者, 禁煙・室, 食事・面, 脱衣・場, 開放・感, 宿泊・代, 仮眠・室, できあい・物, 案内・板, 洗濯・機, ...
名詞-一般・名詞-接尾-一般: 繁華・街, ホテル・代, 映画・祭, イオン・水, 清潔・感, ローカル・線, タクシー・代, 置き場・所, 中心・部, タオル・類, 最終・日, 旅館・側, ホテル・代, ファミリー・層, 家族・連れ, ホテル・内, 餅・つき, 温度・差, 中心・街, 幼稚園・生, ...
名詞-一般・名詞-サ変接続: 温度・設定, コーヒー・サービス, 新聞・サービス, 温度・調節, 無料・送迎, 国籍・料理部屋・担当, 価格・重視, ネット・接続, サイズ・設定, 無料・サービス, LAN・サービス, 手作り・料理, 無料・朝食, 団体・予約, 旅館・設置, 値段・対比, ルーム・サービス, ...
名詞-一般・名詞-一般・名詞-一般: 足・もみ・マッサージ, コンタクト・ケア・用品, レディース・プラン・特典, マイナス・イオン・効果ビジネス・ホテル・感覚, 椅子・部屋・面積, 冬・期間・料金, 公衆・温泉・銭湯, 和食・膳・プラン, 春・スキー・状態, ミニ・お菓子・セット, 大型・温泉・ホテル, 死・海風・呂 ...
名詞-一般・名詞-サ変接続・名詞-接尾-一般: Drink・サービス・券, 空気・清浄・機, 自動・精算・機, 自動・販売・機, コーヒー・割引・券, 靴・脱・場, 男性・従業員・員, 海鮮・料理・店, 青空・駐車・場, 中華・料理・店, 地下・駐車・場, 立体・駐車・場, メガネ・洗浄・機, 風呂・利用・券, ...

4. 分類手法

本章では、カテゴリの初期設定、特徴語の重み係数、配属スコア、具体的なカテゴリ分類手順の順に述べていく。

4.1 カテゴリの初期設定

人手によって、各カテゴリに対して、出現頻度の高い名詞句を種語として 10 個程度選択する。以降は、システムが自動的に、各カテゴリの特徴語集合を生成してから、各特徴語の重み係数を計算する（計算式は 4.2 節で説明する）。

種語の設定はカテゴリごとに思いついた単語を設定することでも考えられるが、本研究では、少数の種語で、分類の最初のステップでも、多くの特徴語を生成するため以下の方針で種語を設定した。降順にソートされた名詞句リストの中から、一個一個を各目的カテゴリに設定する。目的カテゴリに設定できない単語はゴミカテゴリに設定することになっている。表 2 は実験で設定したカテゴリの種語を示している。

4.2 特徴語の重み係数

一つの特徴語が同時に多数のカテゴリの特徴語となり得るので、その特徴語の重み係数をカテゴリごとに設ける。カテゴリの多くの名詞句と共起する、そして、その共起頻度が高いほど、

重み係数が大きく設定される。

ここでは、カテゴリ C_k を考え、 C_k の特徴語集合を $F_k = \{f_1, f_2, f_3, \dots, f_h\}$ とし、 C_k の名詞句集合を $N_k = \{n_1, n_2, n_3, \dots, n_l\}$ とする。この時、特徴語 $f_i \in F_k$ のカテゴリ C_k に対する重み係数は基本的に語 f_i と集合 N_k の間の相互情報量で計算されるが、多くのカテゴリの特徴語となっている単語の影響力を減らすために、さらに、上の相互情報量を F_i^{cat} で割る。 F_i^{cat} は特徴語 f_i が特徴語として含まれているカテゴリの数である。

$$coef_k(f_i) = \frac{1}{F_i^{cat}} \sum_{n_j \in N_k} P(f_i, n_j) \log \frac{P(f_i, n_j)}{P(f_i)P(n_j)}$$

但し、

$$P(f_i, n_j) = \frac{freq(f_i, n_j)}{F_{pair}}$$

$$P(f_i) = \frac{freq(f_i)}{F_{word}}$$

$$P(n_j) = \frac{freq(n_j)}{F_{word}}$$

ここで、 F_i^{cat} は f_i を特徴語として含まれるカテゴリの数、 F_{word} は単語の出現頻度の総数、 F_{pair} は（特徴語、名詞句）ペアの出現頻度の総数、 $freq(f_i)$ 、 $freq(n_j)$ はそれぞれ f_i と n_j の出現頻度、 $freq(f_i, n_j)$ は n_j に係った f_i の出現頻度である。

4.3 配属スコア

ある名詞句 N とカテゴリ C に対する配属スコアは N がカテゴリ C に分類されるに相応しい度合いを表す数値である。実際に、名詞句 N を分類する際、 N の各カテゴリに対する配属スコアを計算し、一番高い配属スコアが算出されたカテゴリに N が分類される。配属スコアの計算は特徴語の重み係数を考慮した相互情報量を用いる。

本手法では、分類過程中、カテゴリの特徴語が拡張されていく。カテゴリ C_k のある時点における特徴語集合を F_k とすると、評価対象 n_i のカテゴリ C_k に対する配属スコアは次の式で計算される。

$$As_k(n_i) = \sum_{f_j \in F_k} coef_k(f_j) P(f_j, n_i) \log \frac{P(f_j, n_i)}{P(f_j)P(n_i)}$$

4.4 分類アルゴリズム

本手法の分類手順を以下に述べる。図 3 ～図 6 は例として 3 つのカテゴリの初期設定と分類過程を示している。

ステップ 1: カテゴリの初期設定

- 各カテゴリの種語が人手によって設定される。図 3 は 3 つのカテゴリの初期設定の例を挙げている。

- 各カテゴリに対し、特徴語集合を生成してから、各特徴語の重み係数を計算する。図 4 は特徴語生成と重み計算の例を示している。

ステップ 2: すべての名詞句が分類されるまで、このステップ 2 が繰り返される。

- 未分類の名詞句 n_i を取得する。
- n_i の各カテゴリに対する配属スコアを計算する。

• n_i に対して、一番大きな配属スコアが算出するカテゴリを C_k とすると、 n_i は C_k に分類される。図 5 は名詞「パン」の分類を例として示している。

• C_k の特徴語集合を F_k とする。 F_k に属していない、 n_i と係り受け関係ともつ単語を F_k に追加する。

• C_k の名詞句集合が増えたら、 F_k の各特徴語の重み係数を更新する。図 6 は特徴語集合と重み係数の更新を示す例である。

このように、分類プロセス中に各カテゴリの名詞句集合と特徴語集合が拡張されていく。

5. 評価実験

5.1 実験データと評価方法

楽天トラベル「お客様の声」の 10,000 感想文書（全部で 40,007 文が含まれている）を実験データとし、「部屋」、「食事」、「風呂」、「サービス」、「設備」、「立地」、「料金」、「ゴミ」の計 8 カテゴリに評価対象を分類した。ゴミカテゴリを設定することで、評価対象となりえない単語がフィルタリングされるかどうか確認してみた。

評価方法として、分類結果を被験者による 3 値評価（「正」、「負」、「疑」）してもらうことにした。「正」、「負」、「疑」はそれぞれ「正しく分類された」、「他のカテゴリに分類されるべき」、「場合によってこのカテゴリに分類されることも考えられる」と設定している。

今回は分類された単語のランクの適切さを評価するために、各カテゴリの上位の単語を段分けして、上位 10、上位 20、上位 40、上位 60、上位 80、上位 100、上位 150 の適合率を集計する。

5.2 実験結果と考察

表 4 は分類結果のカテゴリごとの単語数を示している。これを見ると、ゴミカテゴリは単語全体の大半を占めている。これはゴミカテゴリの種語設定に関係していると考えられる。4.1 節の種語の設定方針により、“ホテル”、“宿”、“旅館”、“建物”、“景色”など、評価対象となりえると考えられるにもかかわらず、今回設定されたカテゴリのどこにも入らない、しかも、出現頻度が高い単語がゴミカテゴリの種語に設定されている。このため、ゴミカテゴリの中には、「ゴミ」以外に様々な「話題」が入っていると考えられる。目的カテゴリ（ゴミカテゴリでないもの）の中では、単語のほとんどが「部屋」カテゴリに分類されている。これは、他の目的カテゴリに比べて、「部屋」カテゴリの中に「部屋のサイズ」、「部屋の中の設備」、「部屋の位置」（窓から見える景色、線路の横とか）など多数のサブカテゴリが含まれていることが考えられる。

図 3 は各カテゴリの上位最高 150 語を評価した適合率を示すグラフである。左の図は「正」のみを正解、右の図は「正」と「疑」を正解とする場合の結果である。上位に行くほど、適合率が高いため、上位の単語が適切に分類されていることが分かる。また、ゴミカテゴリの適合率が平均で 80% となっているので、ゴミカテゴリも適切に分類された。

表 5 は各カテゴリの上位 50 語と出現頻度を挙げている。但

し、単語数が 50 を満たないカテゴリも含んでいる。「食事」カテゴリの上位 50 を見ると“定食”、“牛刺し”、“釜飯”、“ディナー”のような出現頻度が低い（2 回）単語でも上位に分類されている。この結果は、既存研究の方法のように出現頻度が高い単語を分類対象とする方法では得られないものである。

表 3 は各カテゴリに対し、上位 25 の特徴語を挙げている。「食事」カテゴリに着目すると、カテゴリに固有の特徴語（“おいしい”、“食べる”、“飲む”、“頂く”など）が上位に表れている。これにより、「食事」カテゴリの高い適合率（平均で 90%）が説明できる。「サービス」のカテゴリにも同じことが見える。固有な特徴語は“フロント”、“従業員”、“スタッフ”などホテルのスタッフを表す単語と“対応”、“丁寧”、“親切”など接客態度を表す単語であることが分かる。つまり、本手法により、意味範囲が狭いカテゴリほど分類精度がよい。これは本研究の仮定（2 章に説明した。「同じ特徴的な単語群と共起する名詞句を同じカテゴリに分類する」ということ）に合致している。

しかし、「部屋」、「風呂」、「設備」、「料金」の方は、“広い”、“きれい”、“よい”、“高い”、“安い”などの一般性が高い特徴語（つまり、様々なカテゴリの特徴語となりうる単語）が上位に上がっている。このため、これらカテゴリの適合率が低下する傾向を見せている。具体的な例を挙げると、「料金」カテゴリは“安い”、“高い”が 1 位と 2 位の特徴語になっている。しかし、“満足度が高い”、“天井が高い”、“ポイントが高い”、“レベルが高い”、“評価が高い”、“飲み物が高い”、“身長が高い”のような文脈もあるため、これらの単語が誤って分類されていることが分かる。4.4 節で述べたように、このような一般性が高い特徴語の影響力を減らすため、特徴語の重み係数はこれらの特徴語が含まれているカテゴリの数で割っているが、今後は更なる対策が必要と考えられる。

6. 関連研究

Domain-specific lexicon の自動構築を目的とした研究が多くある [1]~[7]。これらの研究は SVM や Naive Bayes や Adaboost など、教師あり学習手法を駆使して名詞のカテゴリ分類を行っている。これに対して、本研究は教師データなしで学習不要な手法を提案している。

また、Riloff らの研究では、学習不要な手法で、名詞のカテゴリ分類を行う [1], [2]。この研究の基本的な考えは、同じ名詞群と共起する名詞を同じカテゴリに分類することにある。本研究と同じように、名詞を分類する際、名詞のカテゴリに対する配属スコア（論文ではカテゴリスコアと読んでいた）を計算する設定であるが、この配属スコアは分類したい名詞がカテゴリの名詞群と共起する確率で計算されている。つまり、この手法は配属スコアが頻度のみを考慮している点、特徴語が名詞のみを扱っている点、名詞間の共起関係（ここでは、同一文に、他品詞の単語を飛ばした時に、隣に出現する 2 つの名詞が共起関係にあると呼んでいる。）が係り受け関係ではない点で本研究と異なる。

7. お わ り

本研究では、相互情報量を用いた評価対象のカテゴリ分類手法を提案した。評価実験として、Web で公開されている楽天トラベル「お客様の声」に含まれる評価対象を8つのカテゴリ（部屋、食事、風呂、サービス、設備、立地、料金、ゴミ）に分類した。各カテゴリの上位語を評価した結果、上位の単語が適切に分類された。しかも、出現頻度が低い単語でも評価対象として抽出され、適切に分類された。さらに、ゴミカテゴリを設定することにより、評価対象とならない単語がゴミカテゴリに分類された。

今後は様々な評価情報（日本語、英語）に適用して、この手法が分野独立 (Domain-independent) かどうかを検証する必要がある。そして、種語の数や名詞句を分類する順番などが分類結果にどう影響することと、一般性が高い特徴語の対策は検討課題だと考えられる。

謝 辞

本研究は楽天技術研究所の支援を受けた。記して深謝する。

文 献

- [1] E. Riloff and J. Shepherd, 'A Corpus-Based Bootstrapping Algorithm for Semi-Automated Semantic Lexicon Construction', *Journal of Natural Language Engineering*, 1999
- [2] E. Riloff and J. Shepherd, A Corpus-Based Approach for Building Semantic Lexicons, *Proceedings of the second conference on Empirical Methods in Natural Language Processing (EMNLP-2)*, 1997
- [3] H. Avancini, A. Lavelli, B. Magnini, F. Sebastiani, and R. Zanolini, 'Expanding domain-specific lexicons by term categorization', *Proceedings of 18th ACM Symposium on Applied Computing (SAC)*, 2003
- [4] L. Rigutini, E. Di Iorio, M. Ernandes and M. Maggini, 'Automatic term categorization by extracting knowledge from web', *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI)*, 2006
- [5] L. Rigutini, E. Iorio, M. Ernandes and M. Maggini, Semantic labeling of data by using the Web, *Proceedings of the Web Intelligence Conference (WIC)*, 2006
- [6] V. Pekar and S. Staab, 'Word classification based on combined measures of distributional and semantic similarity', *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics (EACL)*, 2003
- [7] V. Pekar, M. Krkoska and S. Staab, 'Feature Weighting for Co-occurrence-based Classification of Words', *Proceedings of the 20th international conference on Computational Linguistics (COLING)*, 2004
- [8] A. Aizawa, 'A method of Cluster-based indexing of textual data', *Proceedings of the 19th international conference on Computational linguistics (COLING)*, 2002

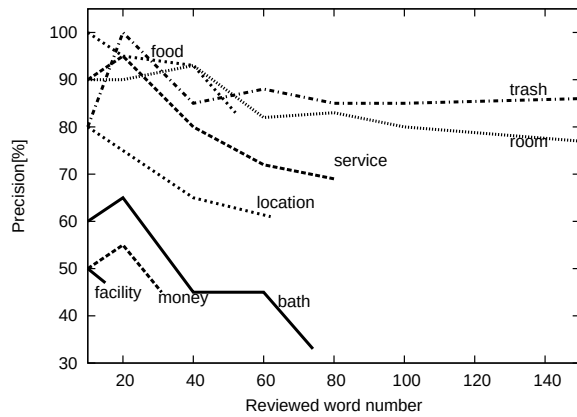
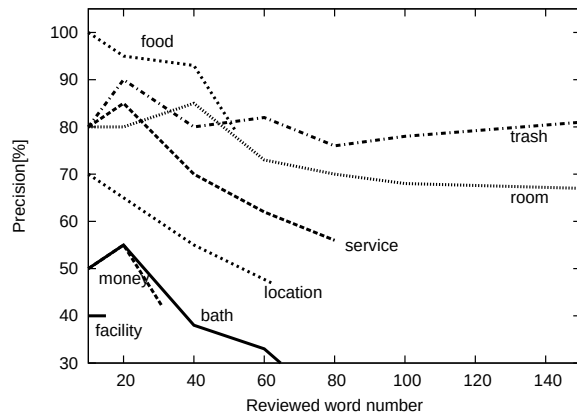


図 2 上位の単語を評価した適合率

左は「正」のみを正解とした場合、右は「正」と「疑」を正解とした場合である。

表 2 各カテゴリの種語

カテゴリと種語（出現頻度）
食事: 朝食 (1636), 食事 (1575), 料理 (745), 夕食 (499), バイキング (406), パン (375), レストラン (197), 味 (193), コーヒー (168)
部屋: 部屋 (6127), 音 (439), ベッド (337), トイレ (268), 臭い (233), 窓 (186), テレビ (175), 冷蔵庫 (177), シャワー (172)
風呂: 風呂 (1932), 温泉 (521), 浴場 (483), 露天風呂 (246), お湯 (214)
サービス: 対応 (2055), サービス (1270), フロント (913), スタッフ (379), 従業員 (362), 笑顔 (248), 応対 (247)
立地: 駅 (682), 立地 (686), 場所 (459), コンビニ (261), スーパー (32)
設備: 設備 (497), 駐車場 (390), アメニティ (243), 施設 (143), 空気清浄機 (63), 加湿器 (52)
料金: 値段 (794), 料金 (569), 価格 (394), コストパフォーマンス (134)
ゴミ: ホテル (1846), 気 (1217), 感じ (767), 機会 (743), 人 (568), お世話 (532), 宿 (519), 宿泊 (519), 気持ち (380), 他 (352), 子供 (307), 雰囲気 (291), 種類 (276), 割 (244), 印象 (240), 内容 (240), プラン (238), 利用 (236), 気分 (227), 旅館 (226), 仕事 (221), 申し分 (206), 最高 (203), 建物 (195), 思い出 (191), 景色 (181), 外 (174), 好感 (172), 是非 (167), 不満 (167), 旅行 (164), とも (157)

表 3 各カテゴリの上位 20 特徴語

食事: おいしい, 美味しい, 朝食, 食べる, いただく, 頂く, サービス, バイキング, 夕食, 良い, 食事, パン, 出る, 満足する, つく, 飲む, コーヒー, 無料, 海, 大変, 種類, とる, おにぎり, 風呂, 嬉しい
部屋: 部屋, 広い, 行き届く, 見る, きれい, 綺麗, 風呂, 狭い, 置く, 用意する, 大きい, 聞こえる, 使える, 見える, 入る, 開ける, 臭い, ツイン, 音, 清潔, 沈む, タバコ, ベッド, 隣, 気
風呂: 入れる, 入る, 広い, 取れる, 良い, 癒す, 部屋, トイレ, 熱い, 気持ちよい, 冷たい, とれる, 出る, 弱い, シャワー, 行く, 風呂, 温まる, 食事, 狭い, お湯, 大きい, 振る, 露天風呂, 急
サービス: 対応, 良い, フロント, 方々, 従業員, スタッフ, 丁寧, 親切, 皆さん, こもる, 応対, 女性, 借りる, 皆様, 対応する, 遅い, 笑顔, 素晴らしい, 満足する, 感じ, 貸す, 接客, 受ける, 朝食, うれしい
立地: 近い, 良い, 便利, 遠い, 歩く, 駅, わかる, 離れる, 分かる, 地下鉄, いい, 迷う, 部屋, ホテル, 教える, 繁華街, 面す, 最高, 駅前, 行く, 満足する, 助かる, 送る, 細い, コンビニ
設備: 充実する, 整う, 必要, 古い, 部屋, 遠い, 広い, 揃う, 新しい, 乗る, 降りる, きれい, サービス, 離れる, 狭い, 女性用, 案内, 立地, わかる, 止める, 置く, 充実, 清潔, 十分, 良い
料金: 安い, 高い, 考える, リーズナブル, 割, 良い, 割り, 手頃, 泊まれる, 優れる, 宿泊できる, わり, 部屋, 大変, 見合う, 自販機, 満足する, 格安, 満足, 払う, いい, 手ごろ, 良心, ホテル, 立地
ゴミ: 良い, 行く, 宿泊する, 多い, 泊まる, 言う, 悪い, ホテル, 他, 値段, いい, いる, 持てる, 使う, 喜ぶ, いく, 部屋, 少ない, 出来る, 古い, 来る, 満足, 出る, 降る, 無い

表 4 各カテゴリの単語数

カテゴリ	食事	部屋	風呂	サービス	立地	設備	料金	ゴミ
単語数	52	435	74	80	62	15	31	734

表 5 各カテゴリの上位 50 語とその出現頻度

<u>食事:</u> ごはん (39), 肉 (51), 酒 (36), 芋煮 (6), 味噌汁 (49), ワイン (30), すき焼き (9), しゃぶしゃぶ (11), 朝食共 (14), 御飯 (5), 土産 (34), 米 (3), 天ぷら (7), 雑煮 (6), 焼きたて (5), 1つ1つ (2), ぶと揚げ (2), 定食 (2), 牛刺し (2), 釜飯 (2), ディナー (2), 蕎麦 (2), メチャクチャ (2), 魚料理 (2), 地酒 (5), スープ (4), サラダ (10), ビール (18), ウェルカムドリンク (3), 返事 (7), 梅酒 (2), 前菜 (6), 無料朝食 (4), 珈琲 (4), 牛乳 (13), ウーロン茶 (8), レンタサイクル (14), 赤ベコ (5), みそ汁 (6), 鍋 (6), 赤 (2), パラエティー (2), パラエティ (2), 塩分 (2), 等 (2), 味噌陶板焼き (4), ソーセージ (2), 跡 (4), 本当 (2), 小鉢 (4)
<u>部屋:</u> 掃除 (199), 清潔 (168), ベット (154), 清掃 (120), ツイン (160), 隣 (163), 快適 (126), 夕日 (146), ユニットバス (69), シングル (125), バスルーム (57), 声 (234), 海 (115), メゾネットタイプ (22), エアコン (104), 室内 (67), 廊下 (133), ダブル (31), ドア (61), 鍵 (81), トリプル (19), 広め (65), 畳 (45), 禁煙 (26), 教育 (43), ロビー (77), インターネット (99), タバコ (108), バスタブ (28), 浴室 (44), スペース (85), 照明 (69), デスク (20), 空調 (78), 最上階 (27), テーブル (69), 浴槽 (32), 客室 (29), 喫煙 (21), 隅々 (17), 別 (77), メゾネット (9), 壁 (63), 館内 (87), 写真 (47), 洗面台 (36), 暖房 (60), 手入れ (11), 布団 (63), 露天風呂付き (6)
<u>風呂:</u> 疲れ (227), 予約 (314), 水 (117), 手 (57), 体 (83), 中 (13), 連絡 (19), 便座 (31), 力 (8), 朝風呂 (7), 事前 (52), 熱め (21), ヒビ (4), 狸小路 (23), 水圧 (22), 源泉 (9), 天然温泉 (5), 貸切状態 (7), 年季 (9), 勢い (16), 低温 (2), 真っ先 (2), エキストラベッド (2), 考慮 (2), TEL (2), 貸し切り (9), 容器 (4), 休み (10), 氷 (4), 麦茶 (6), 真ん中 (21), スイッチ (12), システム (8), ヨーグルト (8), 小路 (2), かご (2), 部類 (2), 段階 (2), 初日 (2), 袋 (4), とこ (4), 具 (15), 口 (11), 芯 (13), タンク (6), 和式 (6), 出方 (6), のど (5), 内風呂 (2), 便器 (4), 方々 (603), 丁寧 (130), 皆さん (291), 皆様 (119), 心 (150), 方達 (41), 男性 (52), 姉さん (20)
<u>サービス:</u> 方たち (24), 受付 (28), フロントスタッフ (8), 説明 (117), 女将 (56), 女性スタッフ (12), 自転車 (27), 対応等 (14), ホテルマン (3), フロント係 (3), 人達 (6), 電話 (67), ホテル側 (11), 女性用 (33), 感謝 (24), 職員 (4), 紳士 (2), フロントマン (2), 係員 (2), ホテル従業員 (2), 人たち (4), 地元 (28), マニュアル (4), 場 (10), おもてなし (36), 領収書 (4), 掲示板 (7), そつ (9), 皆さま (4), お願い (39), レンタカー (4), 言葉使い (4), どの方 (2), 真心 (8), 投稿通り (4), 気遣い (21), 延長コード (2), 反発枕 (2), モデム (2), 即座 (2), 心使い (14), アメニティセット (6)
<u>立地:</u> 繁華街 (136), 地下鉄 (80), 出張先 (9), 距離 (76), 仕事場 (4), 国道 (11), ススキノ (8), ススキ (10), コンビ (25), バス停 (17), 大通り公園 (3), インター (3), スキー場 (9), 道 (38), 難点 (32), 移動 (46), 街 (26), 線路 (2), 飲み屋街 (2), 市役所 (2), 飲み屋 (2), 場外市場 (2), 満席 (2), 目的地 (2), 山 (4), 空港 (8), 満点 (17), 使い方 (5), 範囲 (7), 裸足 (3), 市街地 (4), 中心街 (6), 徒歩圏内 (4), 買い物等 (2), 長期滞在 (2), 飲食 (2), 近所 (10), プラプラ (2), 最寄り駅 (2), 雪道 (2), もさ (2), 町 (10), 送迎 (33), 真横 (6), 反対側 (9), 道順 (2), 状況 (7), 中心 (24), 温泉街 (6), 道のり (2)
<u>設備:</u> アメニティー (25), 売店 (11), ソフト面 (2), 注意 (8), マッサージ (11), 新幹線 (7), 最低限 (4), ケア (2), 相談 (5), エレベーター (3), チェック (4), S L (2), 大型車 (2), 食事場所 (2), コスト (2)
<u>料金:</u> 宿泊料金 (57), 宿泊費 (28), 金額 (52), 満足度 (24), 天井 (47), ポイント (36), レベル (42), 評価 (96), 宿泊代 (9), 宿泊料 (8), 駐車料金 (9), 朝食料金 (4), 宿泊代金 (3), 飲み物 (49), 駐車場代 (2), 我慢 (2), 通常 (11), ジュース (16), 身長 (2), 標高 (2), 鼻 (7), 配置 (15), 商品 (4), 価格設定 (5), ホスピタリティ (8), 理由 (15), 目線 (2), トータル (4), 格安料金 (2), 食事代 (2), ドリンク (6)
<u>ゴミ:</u> 居心地 (75), 量 (375), 泊まり (61), 欲 (71), 眺め (108), 心地 (25), 車 (299), 評判 (36), 家族 (143), 荷物 (161), 寝心地 (42), 品数 (160), 天気 (52), 一緒 (134), ビジネスホテル (196), チェックイン (274), 割り (148), 遊び (29), アクセス (56), 使い勝手 (28), お客 (93), 旅 (142), 雪 (126), 接客態度 (26), 前回 (68), 気持 (16), 予定 (91), 質 (100), 総合 (57), 湯加減 (23), 温度 (146), 久しぶり (53), 夫婦 (39), 本館 (43), バランス (24), 母 (82), メニュー (135), 愛想 (17), 次 (77), チェックアウト (122), 季節 (105), 手際 (8), 客 (59), お客様 (79), 数 (48), 一つ (136), 無料 (208), 具合 (9), 人柄 (37), おかげ (95)

カテゴリ	食事	部屋	サービス
種語	● 食事	● 部屋	● 対応
	● バイキング	● ベッド	● フロント
	● 夕食	● トイレ	● サービス
		● テレビ	

図 3 人手による種語の設定例

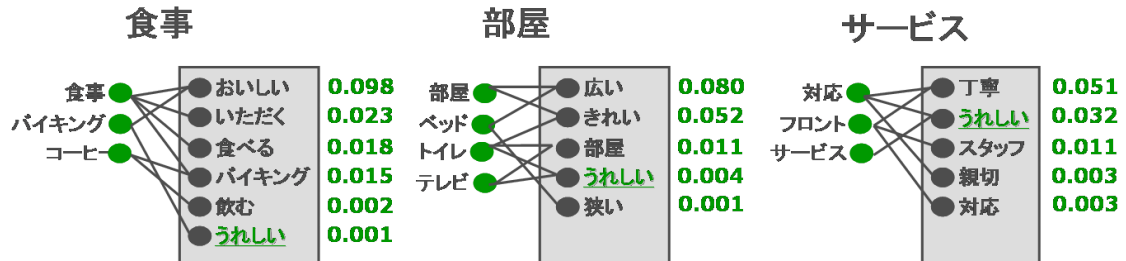


図 4 自動的特徴語生成と重み係数計算例（「うれしい」は一般性が高い単語の例）

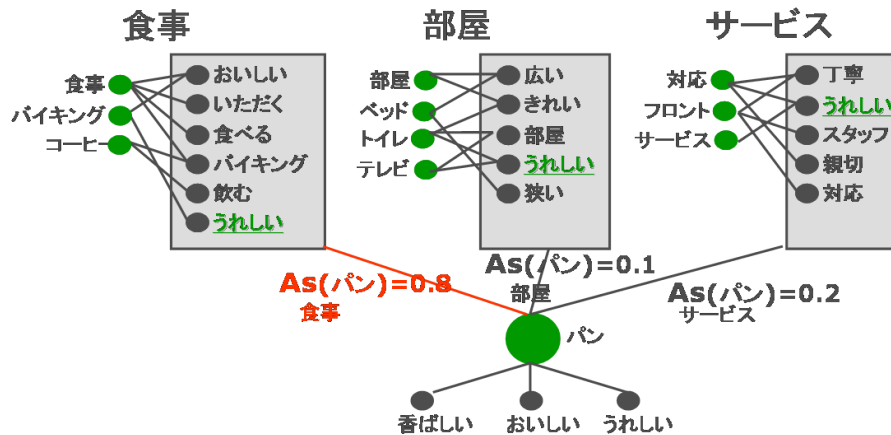


図 5 名詞句の分類例

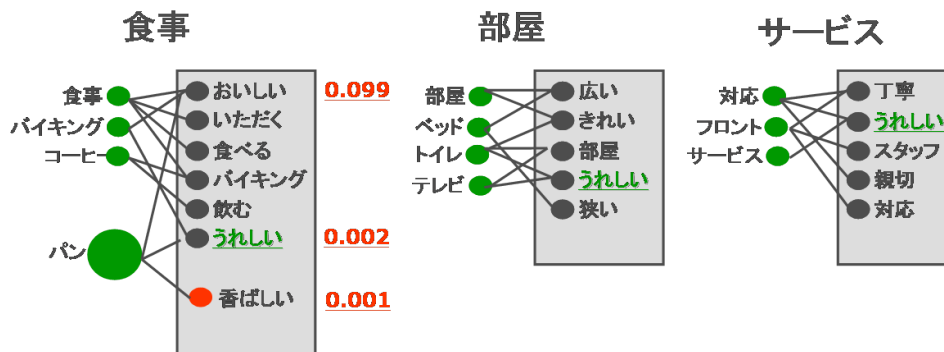


図 6 特徴語集合と重み係数の更新例

アンダーラインの部分は更新された重み係数である。