

学習不要な名詞句のカテゴリ分類手法

NGUYEN P. T. Thao^{†1} 林 貴 宏^{†3} 尾内 理 紀 夫^{†4}
西 岡 悠 平^{†2} 竹 中 孝 真^{†2} 森 正 弥^{†2}

本研究は係り受け関係と相互情報量を基づき、学習不要な名詞句のカテゴリ分類手法を提案する。分類システムの入力は、各カテゴリに対する少数の種語群と分析用のコーパスのみで、学習データを必要としない。本手法は、すべての名詞句を扱うため、出現頻度が低い名詞句も分類可能である。そして、ユーザが設定した各カテゴリ（目的カテゴリと呼ぶ）以外に、ゴミカテゴリを設定する。これにより、目的カテゴリに分類すべきでない名詞句がゴミカテゴリに分類され、誤分類防止の効果があることを実験で確認した。

A Noun Phrase Categorization Method without Requirement for a Learning Phase

NGUYEN PHAM THANH THAO,^{†1}
TAKAHIRO HAYASHI,^{†3} RIKIO ONAI,^{†4}
YUHEI NISHIOKA,^{†2} TAKAMASA TAKENAKA^{†2}
and MASAYA MORI^{†2}

We propose a noun phrase categorization method without requirement for a learning phase. Our method bases on the combination of predicate-argument relations and mutual information measure⁸). The system input requires only a small set of seed words for each category and a text corpus for analysis, but not any learning data. We treat all noun phrases as category candidates; therefore even words with low frequency can be categorized. Also, beside the purpose categories set up by users, by setting an extra “trash category”, we could gather unexpected words into this “trash category” properly (unexpected words refer to words should not be categorized into one of the purpose categories). The experiment results showed that “trash category” is effective at preventing unexpected words from being mis-categorized into purpose categories.

1. はじめに

単語のカテゴリ分類は分野オントロジ構築や、検索における単語の整理法（効率的な検索機能を提供するための単語の意味関係を含んだ単語の整理法）などに必要不可欠な技術であるが、これらの用途に分野特有性があるため、一般性を持つ汎用辞書が利用できない。既存研究は、簡易な共起パターンによりカテゴリのメンバーを抽出する手法^{1),2)}と学習を駆使する手法³⁾⁻⁷⁾に分けられる。前者は分類精度に問題がある一方、後者は大量の学習データの作成コストが高いという問題がある。そこで、本研究は、適切な分類が可能かつ学習不要なカテゴリ分類手法を提案する。分類システムの入力は、各カテゴリに対する少数の種語群と分析用のコーパスのみで、学習データを必要としない。コーパスに出現する名詞句が分類されるため、分野特有な分類結果が得られる。

提案した分類手法の一部については DEIM フォーラム 2009⁹⁾ で発表した。本稿と論文⁹⁾の差分はゴミカテゴリ設定の詳細化（5章）と重み係数の効果（6.2節）、ゴミカテゴリの効果（6.3節）の実験による検証である。

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章はゴミカテゴリの設定について述べる。6章は重み係数の効果、ゴミカテゴリの効果、実験結果と考察を述べる。7章は関連研究を挙げ、8章でまとめる。

2. 提案手法の概要

本研究の基本的な考え方は特徴が類似の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”などとよく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。この特徴的な単語は特徴語と呼び、名詞句との係り受け関係により抽出される（詳しくは 3.2 節で述べる）。

†1 電気通信大学大学院 電気通信学研究科 情報工学専攻

Department of Computer Science, Graduate School of Electro-Communications

†2 楽天技術研究所

Rakuten Institute of Technology

†3 新潟大学

Niigata University

†4 電気通信大学

University of Electro-Communications

図1は本手法の処理手順を示している。前処理では、コーパスから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。処理の詳細は3章と4章で説明する。

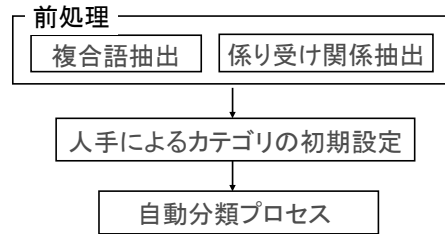


図1 処理の手順⁹⁾

3. 前処理

3.1 複合語抽出

本研究は形態素解析器 Mecab^{*1} を使って、コーパスに出現する全ての形容詞、動詞句、名詞句を抽出した。単語は品詞を、複合語は品詞のパターンを指定し抽出する。

但し、以下に使用する品詞情報は API 品詞体系^{*2} を参考している。

- 形容詞は以下の品詞によって抽出される。
 - 「自立-形容詞」(例：リンゴが 赤い)
 - 「非自立-形容詞」(例：リンゴが 赤く ておいしい^{*3})
- 動詞句は以下の品詞とパターンによって抽出される。
 - 「自立-動詞」(例えば、“食べる”、“行く”など)
 - 「名詞-サ変連続・自立-動詞」(例えば、“更新・できる”、“利用・する”など)

*1 形態素解析器 Mecab
<http://mecab.sourceforge.net/>

*2 IPA 品詞体系
<http://www.unixuser.org/euske/doc/postag/index.html>

*3 “おいしい”も「自立-形容詞」として抽出される。

- 名詞句は以下の品詞とパターンによって抽出される。
 - 「名詞-一般」(例えば、“風呂”、“部屋”など)
 - 「名詞-サ変接続」(例えば、“対応”、“食事”など)
 - 3つの品詞「名詞-一般」、「名詞-サ変接続」と「名詞-接尾一般」を組合せたパターンにより抽出(表1に例として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は名詞句を特徴づける役割があると考え、特徴語と呼ぶ。特徴語として扱われる単語類は名詞句、動詞句、形容詞である。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。例えば、例文「おいしいジュースを飲みました」から以下の2つの係り受けペアが抽出される。

- (ジュース、おいしい)：“おいしい”が前文脈の特徴語
- (ジュース、飲む)：“飲む”が後文脈の特徴語

本研究では、係り受け分析器 Cabocha^{*4} 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。

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、具体的な分類アルゴリズムの順に述べていく。

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)。

種語の設定の仕方はカテゴリごとに思いついた単語を設定することも考えられるが、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。

- 名詞句を出現頻度の降順にソートする。
- 上位から一個一個を各目的カテゴリに設定する。

*4 日本語係り受け解析器 Cabocha
<http://chasen.org/taku/software/cabocha/>

- 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 .

表 2 は実験で設定したカテゴリの種語群を示している .

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 . あるカテゴリの重み係数は特徴語のそのカテゴリに対する貢献度を表す . カテゴリの多くの名詞句と共に起る、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 と名詞句集合 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 . 下の式はカテゴリ C_k における特徴語 f_i の重み係数の計算式である .

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

$$\text{但し、} P(f_i, n_j) = \frac{freq(f_i, n_j)}{F_{pair}}, \quad P(f_i) = \frac{freq(f_i)}{F_{word}}, \quad P(n_j) = \frac{freq(n_j)}{F_{word}}$$

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} は係り受け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_k に対する配属スコアは N がカテゴリ C_k に分類されるに相応しい度合いを表す数値である . 実際、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 .

本研究では、分類プロセス中、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、名詞句 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

以下に分類アルゴリズムを述べる . 図 3(a) ~ 図 3(d) は例として 3 つのカテゴリの初期設定と分類プロセスを示している .

- ステップ 1: カテゴリの初期設定
 - 各カテゴリの種語群が人手によって設定される . 図 3(a) 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 3(b) は特徴語生成と重み計算の例を示している (“うれしい” は一般性が高い単語の例) .
- 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 3(c) 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 (図 3(d) では、“香ばしい” がこの単語に相当する) .
 - C_k の名詞句集合が一個増えたから、 F_k の各特徴語の重み係数を更新する . 図 3(d) 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. ゴミカテゴリの設定

4 章で説明した分類アルゴリズムは、一番高い配属スコアを出すカテゴリに名詞句が分類されるため、誤分類が起こりかねない . この問題を対応するために、本研究は、ユーザが設定した各カテゴリ (目的カテゴリと呼ぶ) 以外に、ゴミカテゴリを設定することにした . ゴミカテゴリを設定することにより、目的カテゴリに分類すべきでない名詞句がゴミカテゴリに分類されることを狙う . 図 2 は名詞 “量” の誤分類例の防止の様子を示している . ゴミカテゴリも目的カテゴリと同様、種語群が人手により設定され、名詞句集合と特徴語集合も分

類プロセス中に拡張されていく。

なお、ゴミカテゴリと名付けているが、このカテゴリに分類される名詞句が全てゴミ（不適切な単語）ではない。誤分類の防止を目的とするため、目的カテゴリと違う話題も含む可能性がある。考察は 6.5 節の潜在的な話題において記述した。

6. 評価実験と考察

6.1 実験データと評価方法

実験データは Web で公開されている楽天トラベル「お客様の声」を利用した。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）を分析し、「部屋」、「食事」、「風呂」、「サービス」、「設備」、「立地」、「料金」、「ゴミ」の計 8 カテゴリに名詞句を分類した。そして、ゴミカテゴリを設定することで、分類対象となりえない名詞句がフィルタリングされるかどうか確認した。

評価方法として、分類結果に対して被験者による 3 値評価（「正」、「負」、「疑」）を行う。「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。

今回は分類された名詞句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 の適合率を集計した。

6.2 重み係数の効果

本節では、重み係数は分類精度を向上させる効果があるかどうかを確かめる。

図 4 の左側は配属スコアの降順にソートされた食事カテゴリ分類結果の上位 100 語における適切に分類された単語の分布を示している。左の列は重み係数をかけない場合、右の列は重み係数をかけた場合である。色でマークされた単語が適切に分類された単語を示す。この通り、重み係数は上位の分類精度を向上させる効果があると分かる。

図 4 の真ん中は食事カテゴリの上位 N 適合率を示している。上位 150 語に対しては、重み係数なし (un-weighted) の場合は適合率がおおよそ 70% からやく 25% に落ちる。一方、重み係数あり (weighted) の場合は適合率が 90% から 70% までという精度の向上を見せている。この通り、重み係数をかけるとにより、分類精度が大幅改善されたことが分かる。

6.3 ゴミカテゴリの効果

本節は、ゴミカテゴリは分類精度を向上させる効果があるかどうかを確かめる。図 4 の右側は食事カテゴリの分類結果の上位 N 適合率を示している。これにより、ゴミカテゴリの設定は、分類精度を良くする効果があることが分かる。図 4 の右側は分類結果の上位 150 語への効果のみを示しているが、図 5 のゴミカテゴリ (trash) の適合率を見るとゴミカテゴリの分類精度の平均がおおよそ 80% であることが分かり、ゴミカテゴリが高いフィルタリング効果を果たしていると言える。

6.4 分類精度

図 5 は各カテゴリの上位 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。

表 5 は各カテゴリの上位 30 語と出現頻度を挙げている。但し、単語数が 40 を満たないカテゴリも含んでいる。「食事」カテゴリの上位 30 語を見ると「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、既存研究の手法のように出現頻度が高い単語を分類対象とする手法では得られないものである。

6.5 結果説明可能

表 4 は各カテゴリに対し、重み係数が降順にソートされた上位 25 の特徴語を挙げている。「食事」カテゴリに着目すると、カテゴリに固有な特徴語（「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言える。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、本研究により、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られているが、今後は更なる対策が必要と考えられる。

6.6 潜在的な話題

表3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、分類対象(適切な名詞句)となりえると考えられるにもかかわらず、今回設定された目的カテゴリのどこにも入らない、しかも、出現頻度が高い名詞句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」(不適切な名詞句、例えば“他”、“次”、“割”)以外に色々な「話題」が入っていると考えられる。

目的カテゴリの中では、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の話題が含まれていることが考えられる。

7. 関連研究

ドメイン特有の辞書(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、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、本研究は学習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る)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、この研究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。

8. おわり

本稿では、係り受け関係と相互情報量を用いた名詞句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」に含まれる名詞句を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、出現頻度が低い名詞句も適切に分類された。さらに、ゴミカテゴリを設定することにより、目的カテゴリに分類すべきでない単語がゴミカテゴリに分類された。

今後は様々な言語(日本語、英語)の多様なコーパスに適用して、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、種語群の数や名詞句を分類する順序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。

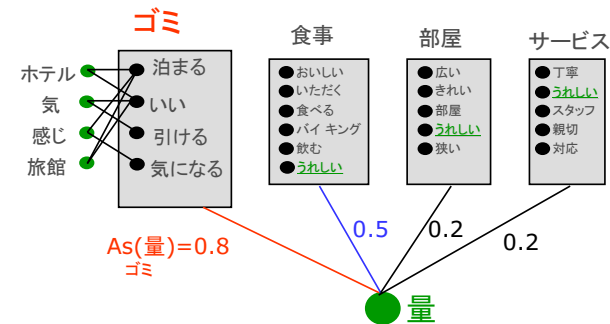


図2 ゴミカテゴリの設定例

参 考 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. Zanoli, 'Expanding domain-specific lexicons by term categorization', Proceedings of 18th ACM Symposium on Applied Computing (SAC), 2003
- 4) L. Rigutini, E. Di Iorio, M. Ernanandes 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. Ernanandes 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) Thomas M. Cover and Joy A. Thomas, E. Riloff and J. Shepherd, 'Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)', p20-22
- 9) Nguyen Pham Thanh Thao, 林貴宏, 尾内理紀夫, 西岡悠平, 竹中孝真, 森正弥「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—」データ工学と情報マネジメントに関するフォーラム (DEIM2009), 2009

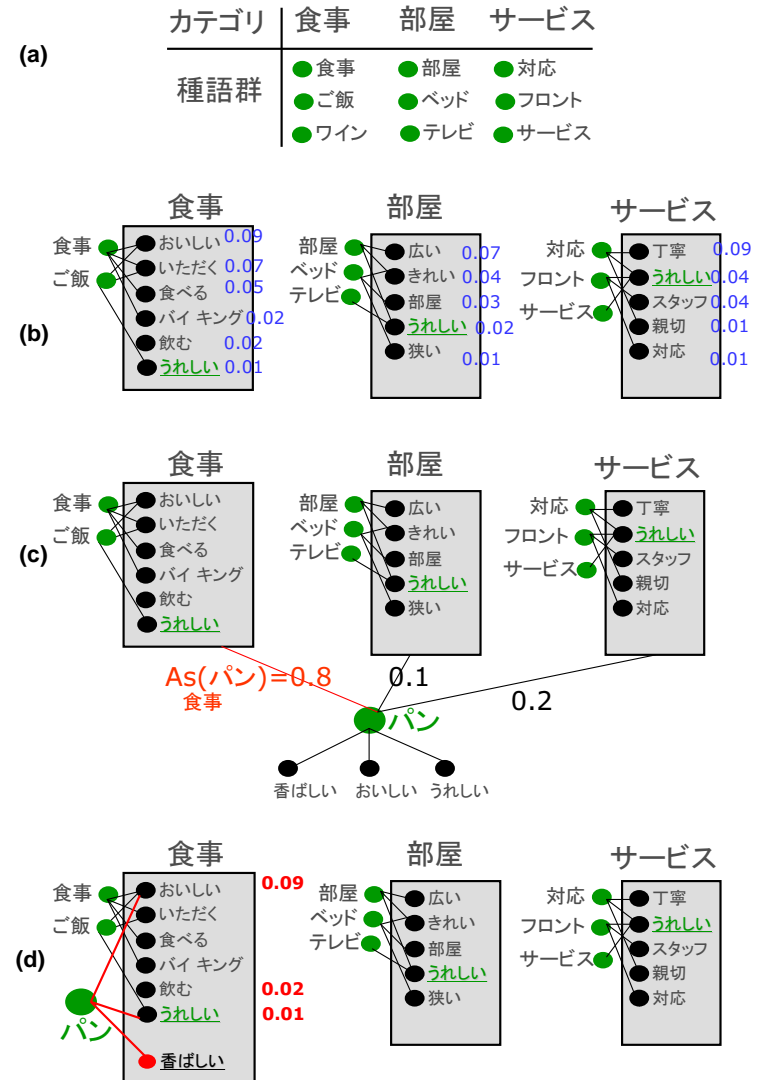


図3 初期設定と分類プロセスの例⁹⁾

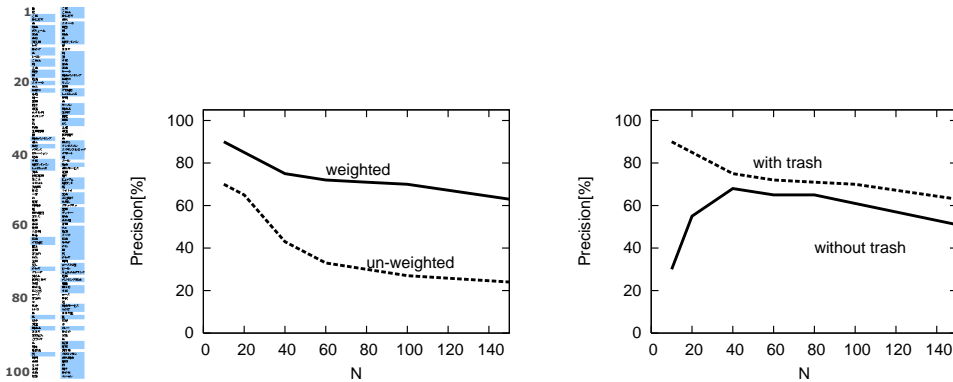


図4 重み係数とゴミカテゴリの効果

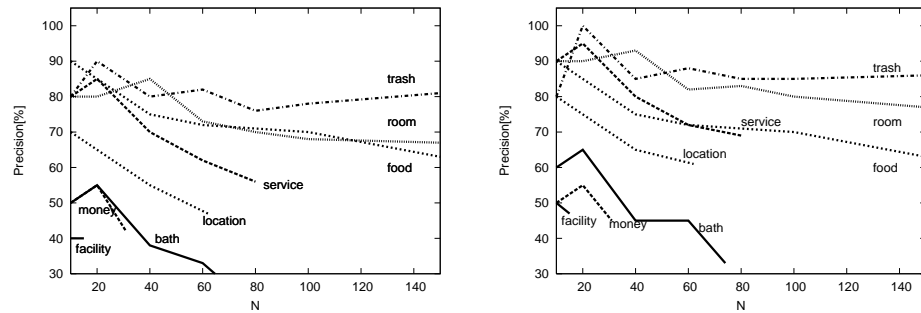


図5 上位 N 単語の適合率
 左図は「正」のみを正解とした場合、右図は「正」と「疑」を正解とした場合

表1 名詞句の抽出パターン例⁹⁾

「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, ...
「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, ...
「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 ...
「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, ...
「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, 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, ...

表2 各カテゴリの種語群⁹⁾

カテゴリと種語(出現頻度)
食事:(計9語) 朝食(1636), 食事(1575), 料理(745), 夕食 (499), パイキング(406), パン(375), レストラン (197), 味(193), コーヒー(168)
部屋:(計9語) 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:(計5語) 風呂(1932), 温泉(521), 浴場(483), 露天風呂 (246), お湯(214)
サービス:(計7語) 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:(計5語) 駅(682), 立地(686), 場所(459), コンビニ (261), スーパー(32)
設備:(計7語) 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:(計4語) 値段(794), 料金(569), 価格(394), コストパ フォーマンス(134)
ゴミ:(計32語) 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 5 各カテゴリの上位 30 語とその出現頻度

表 3 カテゴリ間の単語の分布⁹⁾

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 4 各カテゴリの上位 25 特徴語⁹⁾

<p>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, タ食, 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</p>
<p>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</p>
<p>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</p>
<p>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, 対応, 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</p>
<p>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</p>
<p>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</p>
<p>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</p>
<p>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</p>

<p>食事: 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), デイナー (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), 梅酒 (2)</p>
<p>部屋: 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)</p>
<p>風呂: 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)</p>
<p>サービス: 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)</p>
<p>立地: 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)</p>
<p>設備: 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 (3), チェック (4), S L (2), 大型車 (2), 食事場所 (2), コスト (2)</p>
<p>料金: 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)</p>
<p>ゴミ: 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)</p>