

# 階層的トピックモデルに基づく宿泊レビューの時間変化分析

## Time-Variable Analysis of Accommodation Reviews Based on Hierarchical Topic Model

佐藤 裕次郎 山西 良典 西原 陽子 \*  
Yujiro Sato<sup>1</sup> Ryosuke Yamanishi<sup>2</sup> Yoko Nishihara<sup>2</sup>

<sup>1</sup> 立命館大学大学院情報理工学研究科

<sup>1</sup> Graduate School of Information Science and Engineering, Ritsumeikan University

<sup>2</sup> 立命館大学情報理工学部

<sup>2</sup> College of Information Science and Engineering, Ritsumeikan University

### Abstract:

Accommodation review is a valuable resource for future users to know the voices from the users who have already stayed there. However, it is difficult for users to extract the information specific to each topic such as facilities, access, and breakfast. We consider that seasonal features of accommodation are especially important for comfortable and enjoyable staying. This paper proposes the hierarchical topic analysis with time variation to extract seasonal features for accommodations. The proposed method extracts seasonally important words and shows the similarity of topics that the important words belong to between seasons. In this paper, we discuss the effectiveness of the extracted features for references to choose accommodations.

## 1 はじめに

宿泊予約の主流がネットになったことで、インターネット上の情報をもとに宿泊施設を決定する能力が宿泊検討者に必要とされるようになった。宿泊施設の決定要因となる要素は様々であり、ユーザによって基準は異なるものの、Dickingerらの研究によると、友人からの推薦とオンラインレビューがネット予約において最も重要な影響を与える要因であると示されている [1]。その中でも宿泊レビューは幅広く研究されており、テキストマイニングによる分析結果を意思決定に役立てる研究が進んでいる。Vermeulenらの研究によると、宿泊レビューのポジティブな意見、あるいは、ネガティブな意見でさえも宿泊施設へ注目を集める効果があり、消費者の意思決定を促進するとされている [2]。加えて同じ研究では、ポジティブな意見は消費者の宿泊施設への態度や行動を改善するために役立つとされており、宿泊レビューは消費者にも提供側にも貴重なリソースと言える。

宿泊レビューにおける消費者の意思決定要因を考える際に、本研究では宿泊レビューで時間的に変化する特徴に着目した。季節によって大きく気候が変わる地

域では、各季節で衣食住の様子や人気の旅行先は異なる。また、近隣で開催される季節ごとのイベントなどによっても各宿泊施設の価値は大きく変化する。宿泊レビューに含まれる季節的な特徴を抽出して提示することで、宿泊検討者は宿泊する予定の季節の情報を元に宿泊先を決定することができる。

本稿では、宿泊施設の各月のレビューから重要度の高い単語を抽出するとともに、その重要単語を含む階層型のトピックを、月間で比較することによって、宿泊施設の季節的な特徴を分析する手法を提案する。テキストマイニングにおいて文書内の名詞に与える代表的な特徴量として  $tf-idf$  法による値があり、文書分類や機械学習の素性として用いられる [3]。本研究では名詞に月別の  $tf-idf$  値を付与することで、季節の特徴語の抽出を行う。宿泊施設の特徴抽出では、消費者の欲しい情報を的確に明示するためには“食事”や“風呂”など、カテゴリーの存在は考慮しなくてはならない [4]。カテゴリーによっては季節に応じて特徴が変化し、消費者の決定要因となるかどうかとも変化する。提案手法では、カテゴリーの取得にトピックモデル（潜在的意味解析）を用いる。トピックモデルの一つに LDA (Latent Dirichlet Allocation) [5] という教師なし学習手法がある。この手法では、ある文章は複数のトピックを持つと仮定し、そのトピックを抽出することを目

\*連絡先: 〒 525-8577 滋賀県草津市野路東 1-1-1  
E-mail: {is0309he@ed, ryama@media, nisi-hara@fc}.ritsume.ac.jp

的としている。宿泊レビューに対し LDA を使った分析に、Han らの先行研究がある [6]。この研究では、宿泊施設レビューのトピック分析により感情と評価の関係を抽出することに成功している。また、LDA の拡張手法に hLDA(hierarchical Latent Dirichlet Allocation) がある [7, 8]。hLDA では文書に含まれるトピックに階層構造が存在すると仮定し、トピックを確率的に推定する技術である。宿泊レビューの特徴について、Wang らの先行研究 [9] では、レビューのトピック分析から潜在的な意見・評価を得る Latent Aspect Rating Analysis (LARA) という問題が定義されている。LARA から、観点の多さは宿泊レビュー特有の特徴とも言える。観点や評価属性を、カテゴリーとして考えると、宿泊レビューのカテゴリーの中には包含関係が存在する。例えば“食事”に関しての評価の中には、“食事の値段”や“食事の質”などより詳細な情報が含まれる。本稿では、hLDA を用いた分析によって宿泊レビューのトピックの包含関係が抽出できると仮説を立て、その結果と  $tf-idf$  法により抽出した重要単語を用いた宿泊レビューの時間変化分析を行う。

## 2 分析データ

本稿では、楽天トラベルのレビュー 5,082,427 件のユーザの声をを用いる<sup>1</sup>。使用するレビューデータは 1996 年から 2016 年の間に収録されたものであり、29,400 件の宿泊施設に対してのレビューが存在する。本研究の分析では統計的な見地を導出するため、宿泊施設の中でもレビュー数上位 10% の宿泊施設を分析に使用する。本稿では、このうちレビュー数上位 10% の宿泊施設からランダムに選出した以下の 2 件の宿泊施設を今回の分析対象とした。

1. 粟津温泉 旅亭懐石 のとや
2. シーサイドホテル舞子ピラ神戸

これらの宿泊施設のレビューを月別に分割し、一つの宿泊施設につき 12 個の文書集合を入力データとして扱う。

### 2.1 分析対象とする単語

自然言語処理の前処理の重要性は広く知られている。ストップワードや分析する内容に応じた品詞の絞り込みは、分析の精度向上につながる。

本稿では、宿泊施設の特徴の抽出をメインとするため、名詞のみを分析対象とする、トピック抽出を目的と

<sup>1</sup>楽天株式会社 (2016): 楽天トラベルデータ。国立情報学研究所情報学研究データリポジトリ。(データセット)。  
<https://doi.org/10.32130/idr.2.2>

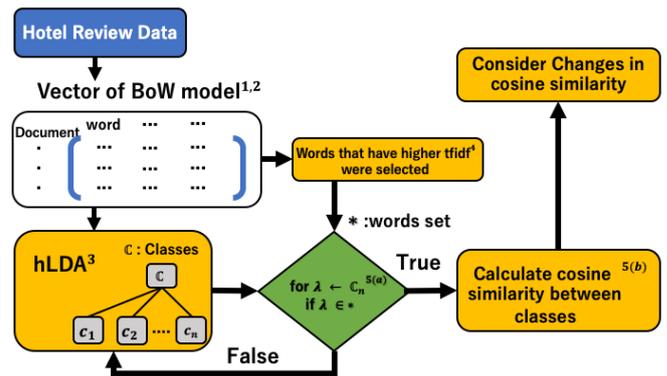


図 1: 分析手順のフレームワーク。図中、3 章に示した分析手順番号を肩文字として示す。

しているため、それ単体で意味のわからない名詞、日にちや記号に関してはストップワードとして扱い、分析から除外する。さらに、hLDA の結果の収束を良くするために、出現回数 3 回以上の名詞のみを使用する。

## 3 hLDA と $tf-idf$ 法を用いた季節に存在する潜在的なトピック分析

提案手法では、宿泊施設の潜在的なトピック分析として hLDA、季節の特徴抽出に  $tf-idf$  法を用いる。この 2 種類の手法により宿泊レビューに存在する季節の潜在的トピックの階層構造での抽出を行う。時間変化分析として、抽出された季節の特徴が他の月でどのようにクラスタリングされているかを考察し、季節の特徴語の変化、トピックの移り変わりを分析する。抽出と考察は、以下の手順 (図 1 の添字番号を参照) に沿って行う、

1. レビューの分かち書きの際に形態素解析器 MeCab と辞書に NEologd を用いて出現した名詞を取得する。
2. Bag-of-Words モデルに基づき出現した名詞のベクトル化を行う。
3. 作成したベクトルを入力し、hLDA で潜在的なトピックを階層構造でクラスタリングする。
4. 分析する月単位のレビューを一つの文書、他の月のレビューも含め 12 個の文書集合とし、出現した名詞に  $tf-idf$  値を付与する。
5. ストップワードを除く  $tf-idf$  値の上位 10% の名詞について、次の (a) と (b) の処理を行う：
  - (a) 同じクラスに抽出された名詞の傾向を分析する。

- (b) 他の月の（その名詞を含む）クラスタと類似度を計算する。

### 3.1 階層的トピックモデル

入力データとして用いる宿泊レビューには、大カテゴリと小カテゴリが存在するように、トピックとしても抽出可能であると仮定し、このトピックの階層関係を抽出する。LDAモデルの拡張として、トピックの階層関係を解析できる hierarchical Latent Dirichlet Allocation (hLDA) に着目し、宿泊レビューの季節的なトピックの変化を評価する。

#### 3.1.1 nested Chinese Restaurant Process

nested Chinese Restaurant Process (nCRP) とは木構造上の確率過程である。この確率過程は hLDA のために用いられ、CRP を利用して以下の比喻で表される [10]。

CRP は  $N$  人の客が無限のテーブルを持つ中華レストランに座るプロセスを仮定して得られる分布のことである。レストランに入ってきた順に、 $1, 2, \dots, N$  とラベル付けできる客が、レストランのテーブルを選択して座っていくとする。最初の客は最初のテーブルに座り、 $n$  人目の客が座るテーブルを  $c_n$  とし、 $i$  番目のテーブルに座る確率は式 (1) によって決定される。

$$p(c_n = i | c_{n-1}) = \begin{cases} \frac{n_i}{\gamma + n - 1} & (\text{occupied table } i) \\ \frac{\gamma}{\gamma + n - 1} & (\text{next unoccupied table}) \end{cases} \quad (1)$$

ここでの  $n_i$  は現在テーブル  $i$  に座っている顧客の数であり、 $\gamma$  はレストランの顧客の数と比較して、顧客が新しいテーブルを選択する頻度を制御する実際のパラメータである。

nCRP において、客は CRP に基づき木構造を形成する。そこで nCRP は以下の比喻で表される。街に無限個のレストランが存在し、各レストランは無限個のテーブルを持つと仮定する。また階層の根となるレストランが存在し、各レストランのテーブルは他のレストランを指定している。まず最初の客が根のレストランに入り、CRP に従いテーブルを選択し、次のレストランへのルートを得る。再び CRP に従いテーブルを選択し、これを無限回繰り返す、木構造におけるパスを決定する。これをすべての客について行うと、無限に枝分かれした無限深さの木の部分木を得る。本研究では客が単語、レストランが階層、テーブルの席がトピックを表し、この部分木の決定に用いられる。

#### 3.1.2 hLDA の生成過程

hLDA の生成過程 [10] において、木構造は nCRP によって生成され、nCRP における客が文書、テーブル（またはレストラン）がトピックを表している。ここからは、先行研究 [7, 8] にならい、hLDA の生成過程を以下に記す。

1. 各トピック  $k \in T$  に対して
  - (a) 単語上の多項分布パラメータ  $\beta_k \sim Dir(\eta)$  を選択
2. 各文書  $d \in \{1, \dots, D\}$  に対して
  - (a) レストランプロセスに基づきノードを選択  $c_d \sim nCRP(\gamma)$
  - (b) レベル上の多項分布パラメータ  $\theta_d \sim Dir(\alpha)$  を選択
  - (c) 各単語  $w_{d,n} (n \in \{1, \dots, N_d\})$  に対して
    - i. レベル  $z_{d,n} \sim Mult(\theta_d)$  を選択
    - ii. 単語  $w_{d,n} \sim Mult(\beta_{c_d, z_{d,n}})$  を選択

hLDA の実装では、あらかじめパラメータ ( $\alpha, \gamma, \eta$ , 階層数) を設定する必要がある。このハイパーパラメータ問題に対し、分析対象の宿泊施設によって抽出の「適切さ」が異なる。このことから、同一の条件で抽出と「適切さ」を得るために先行研究 [7] にならい、より多くのトピックを抽出するため、確率分布の中レベルの層が抽出されやすいように、 $\gamma = 1.0, \eta = 1.0$  に定めた。

本稿では、宿泊施設の特徴の抽出を大きな目標としているため、そのトピックが何に基づいて分類されたかがわかる方が望ましい。そこで本研究では宿泊施設の特徴のトピックの階層数を、よりトピックの内容が特徴的である階層数 3 に設定し、収束した名詞のみを分析に用いた。

### 3.2 $tf - idf$ 法に基づく季節の特徴分析

$tf - idf$  値は、その文書内で使われている単語の重要度を表す [3]。対象の文書内での出現頻度が高い単語ほど重要であるという考えに基づく  $tf$  値と、いくつもの文書で横断的に出現する単語は重要でないという考えに基づく逆文書頻度、 $idf$  値の掛け合わせで  $tf - idf$  値を計算する。月毎のレビューを文書集合として入力するため、 $tf - idf$  値の高い名詞は、他の月に出現していない季節を表す特徴的な名詞と言える。 $tf - idf$  値の特徴量付与は  $tf$  値を式 (2)、 $idf$  値を式 (3) で行う。

$$tf(t, d, month) = \frac{n_{t,d,month}}{\sum_{s \in d} n_{s,d,month}}, \quad (2)$$

$$idf(t, month) = \log \frac{N}{df(t, month)}. \quad (3)$$

### 3.3 他クラスタ間コサイン類似度

一般的に自然言語処理で使われるコサイン類似度は、ベクトル化した文書間の類似性を測ることによく使われる。本稿では、hLDAによって分類された月別のクラスタの周辺単語の違いから、その季節を特徴付ける名詞や特徴の移り変わりを抽出する。そこで、hLDAで獲得された各月のクラスタ間でコサイン類似度を計算する。

クラスタは複数の名詞の集まりで形成され、それぞれがBoWモデルに従いベクトル化した値を計算に用いる。m1月とm2月のクラスタC<sub>m1</sub>とC<sub>m2</sub>の集合を、式(4)と式(5)に記す。

$$C_{m1}(1, \dots, n) = \{W_1, W_2, \dots, W_n\}, \quad (4)$$

$$C_{m2}(1, \dots, n) = \{W_1, W_1, \dots, W_n\}. \quad (5)$$

この2つの集合に対し、コサイン類似度の計算は式(6)で行う。

$$\begin{aligned} \cos(C_{m1}, C_{m2}) \\ = \frac{\sum_{k=1}^n C_{m1}(k) \cdot C_{m2}(k)}{\sqrt{\sum_{k=1}^n (C_{m1}(k))^2} \cdot \sqrt{\sum_{k=1}^n (C_{m2}(k))^2}}. \end{aligned} \quad (6)$$

## 4 分析結果

ユーザのニーズの多さという観点から分析対象を宿泊施設の繁茂期に絞り、1月、5月、8月、12月の結果を出力した。2件の宿泊レビューの名詞のうち、tf-idf値の高かった上位5件を表1に記す。

図2、図3はhLDAの取束しない名詞を除いた分析結果の一部である。次に表1から、季節そのものや時期を表す名詞以外の“食事”や“イベント”など評価の対象として相応しいものを選択する。本稿では「粟津温泉 旅亭懐石 のとや」の“鯛”, 「シーサイドホテル舞子ピラ神戸」の“プール”に着目する。また分析するクラスタを、以下のように定義する。

**小クラスタ** 該当する名詞を含む階層2の単一のクラスタ、図2、図3に表す所のC<sub>a1</sub>やC<sub>a2</sub>に当たる。

**大クラスタ** 該当する名詞を含む階層3のクラスタと同じ親を持つ複数のクラスタ、図2、図3に表す所のC<sub>a</sub>やC<sub>b</sub>と、それぞれの子クラスタ全ての集合。

表 1: 2件の宿泊施設のレビューから抽出したtf-idf値が高かった名詞, 上位5件

	粟津温泉 旅亭懐石 のとや	シーサイドホテル舞子ピラ神戸
1月	お正月, 年越し, 年始年末, 餅搗ぎ, 最上級	お正月, お年玉, お菓子, スペシャル, 周年
5月	ゴールデンウィーク, 鯛, 春の, 愛, 連休	室温, ゴールデンウィーク, 週末, グレード, エステ
8月	お盆, 合掌, 道場六三郎, 鮎, 納涼	プール, 夏休み, 海水浴場, 海水浴, バーベキュー
12月	クリスマス, めし, 突破, 名物, 最上級	ルミナリエ, クリスマス, 冬, おもてなし, スペシャル

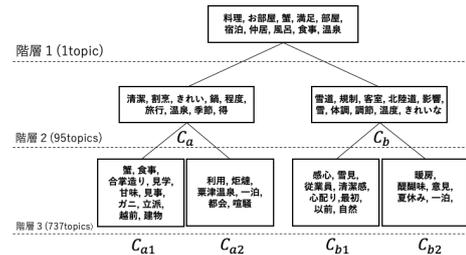


図 2: hLDA, 1月, 粟津温泉 旅亭懐石 のとや

このように大クラスタと小クラスタの類似度の推移を分析することにより、トピックの時間的な大規模な変化と小規模な変化を比較することが可能になる。この結果を表2、表3にそれぞれ記す。また注目する名詞を含む他の月の階層構造を図4、図5に示す。

## 5 考察

### 5.1 hLDAの抽出結果

図2のC<sub>a</sub>は“清潔”や“温泉”, “鯛”の一般的なサービスや、施設への評価がトピックとして存在することがわかる。対症的にC<sub>b</sub>には“雪道”や“雪”, “温度”など天候や環境に関するトピックが抽出された。またC<sub>a1</sub>には“蟹”に関する名詞や“甘味”といった食事に関する評価が抽出され、C<sub>a2</sub>~C<sub>b2</sub>まで比較しても明らかに食事関連のトピックが分類されていることがわかる。またC<sub>b1</sub>はサービスや環境に対してのトピック、C<sub>b2</sub>は温度に対してのトピックであり、明確な差が見取れる。ここからC<sub>b</sub>とC<sub>b1</sub>, C<sub>b2</sub>の階層関係が見取れる。

図3のC<sub>a</sub>は価格、立地に関するトピックが存在することが見てわかる。他にC<sub>b</sub>には“海水浴場”や“眺め”から外のレジャー、あるいは施設の特徴がトピックとして存在することがわかる。C<sub>a1</sub>, C<sub>b1</sub>に、“家族連れ”や“海水浴場”が分類されており、家族旅行が楽しめ、海水浴ができることが想像される。実際にそのようなレ

表 2: 粟津温泉 旅亭懐石 のとや :  $tf-idf$  法で抽出された 5 月の “鯛” を含む他クラスタ間のコサイン類似度. 表中, 同クラスタ間の類似度を下線部に示す.

	1月	2月	3月	4月	<u>5月</u>	6月	7月	8月	9月	10月	11月	12月
大クラスタ	0.0	0.0	0.0	0.029	<u>1.0</u>	0.046	0.018	0.039	0.021	0.028	0.0	0.0
小クラスタ	0.0	0.0	0.0	0.099	<u>1.0</u>	0.199	0.099	0.099	0.099	0.105	0.0	0.0

表 3: シーサイドホテル舞子ピラ神戸 :  $tf-idf$  法で抽出された 8 月の “プール” を含む他クラスタ間のコサイン類似度. 表中, 同クラスタ間の類似度を下線部に示す.

	1月	2月	3月	4月	5月	6月	7月	<u>8月</u>	9月	10月	11月	12月
大クラスタ	0.0	0.0	0.0	0.0	0.0	0.110	0.042	<u>1.0</u>	0.164	0.0	0.0	0.0
小クラスタ	0.0	0.0	0.0	0.0	0.0	0.099	0.099	<u>1.0</u>	0.099	0.0	0.0	0.0

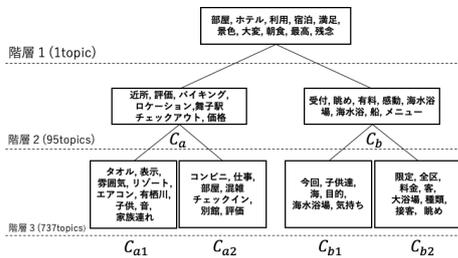


図 3: hLDA, 8 月, シーサイドホテル舞子ピラ神戸

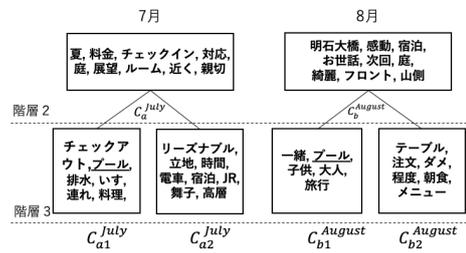


図 5: “プール” を含む 7 月, 8 月の階層構造

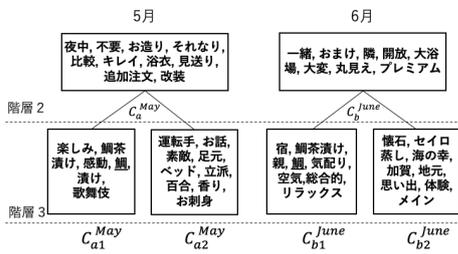


図 4: “鯛” を含む 5 月, 6 月の階層構造

ビューも存在した. これも hLDA の結果の特徴である  
 と考える.

結果として, [粟津温泉 旅亭懐石 のとや] の宿泊レ  
 ビューから, 1 月に料理, 特に蟹や鍋が楽しめ, また天  
 候では雪が想定されることがわかる. また [シーサイド  
 ホテル舞子ピラ神戸] の 8 月の宿泊レビューから, 家族  
 連れが海水浴に訪れやすい季節であることがわかった.

## 5.2 コサイン類似度による他クラスタ間評価

表 2 の結果から, 大クラスタの推移を見ると, 6 月  
 が一番高く, 次に 8 月, 4 月が高いことがわかった. 次  
 元数が増えることも影響し, 大クラスタでさまざまなト

ピックを含むため, より詳細な変化が見て取れる. ま  
 た, 5 月に対して, 4 月と 6 月の類似度が高いことか  
 ら, 近いトピックが存在することがわかる. このこと  
 から季節の類似性は宿泊レビューのトピックに現れる  
 ことがわかる. 図 4 の “鯛” を含む 5 月, 6 月の階層構  
 造には, “鯛茶漬” がどちらの 3 階層にも含まれてお  
 り, 料理名が共起している. 実際には鯛の旬は 3 月~  
 6 月と 9 月~11 月と言われている, 旬を考慮しなけれ  
 ばオールシーズン食べることができる魚である. しか  
 し, この宿泊施設の場合, 5 月 6 月で注目されている  
 トピックであることが明らかになった.

表 3 の結果から, “プール” は 6 月から 9 月まで出現し  
 ていることがわかる. 大クラスタにおいて, 類似度は 9  
 月が一番高くなっており, 近い季節において似たトピッ  
 クが存在することがわかる. しかし, 一般的な “プー  
 ル” の人気が高まる 7 月と 8 月の類似度が低くなった.  
 図 5 から, “プール” と同じクラスタに “大人” や “子供”  
 といった家族旅行の様子がわかる. 実際のレビューに  
 も, 「家族連れ」の旅行形式が, 7 月より 8 月に多く出  
 現する. この結果から, 家族旅行に適した季節, また  
 はそうでない季節が類似度の結果から明らかになった.

結果から, 季節で注目される食事内容や, 家族旅行  
 に適した季節, または家族連れで賑わう季節が明らか

になった。

### 5.3 今後の展望

改善点・展望として、(1)hLDA のハイパーパラメータに起因する精度問題、(2) 意思決定支援に有効な可視化手法、(3) 地域差に現れるトピックの差の抽出、が挙げられる。それぞれ、(1) については分析精度の向上、(2) と (3) については宿泊検討者の意思決定支援に有効なインタラクションモデルに関わる問題と分類される。

hLDA の結果として分類が収束しない名詞も見られ、精度の向上が必要なことは明らかである。解決策としてハイパーパラメータの調節が有効であるが、宿泊施設の各文書によって調整を行なうことは、手動では限界があり、可視化手法の発展が必要である。本稿では、分析と結果の考察までを紹介したが、大きな目標である“消費者の意思決定支援”につなげることの本質を見抜く必要がある。そこで、決定要因になり得る情報の選別などが発展として期待できる。また季節と同じように地域差でも特徴量の違いを抽出できることが予想される。

## 6 おわりに

本稿では、宿泊検討者に対して意思決定支援につながる分析を目的として、季節毎に重要な単語の意味的な変化を抽出する手法を提案した。提案手法では、hLDA の分析によって宿泊施設の大カテゴリーとその中にある小カテゴリーを階層構造で抽出することを試みた。このとき、精度の定量的な評価や、注目すべき名詞の選別として、コサイン類似度と  $tf-idf$  法を用いて分析した。その結果、宿泊施設のトピックを階層関係で抽出でき、季節特有のトピックの情報を抽出できた。また、ハイパーパラメータに起因する精度の問題が明らかになった。今後の展望として、季節と同じように地域差でも特徴量の違いを抽出でき、宿泊検討者への可視化を行う予定である。

## 謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天株式会社から提供を受けた「楽天データセット」を利用した。記して謝意を表す。

## 参考文献

[1] Astrid Dickinger and Josef Mazanec. Consumers' preferred criteria for hotel online booking. *In-*

*formation and communication technologies in tourism 2008*, pp. 244–254, 2008.

- [2] Ivar E Vermeulen and Daphne Seegers. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, Vol. 30, No. 1, pp. 123–127, 2009.
- [3] Juan Ramos, et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242, pp. 133–142. Piscataway, NJ, 2003.
- [4] Katerina Berezina, Anil Bilgihan, Cihan Cobanoglu, and Fevzi Okumus. Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, Vol. 25, No. 1, pp. 1–24, 2016.
- [5] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- [6] Hyun Jeong Han, Shawn Mankad, Nagesh Gavirneni, Rohit Verma, et al. What guests really think of your hotel: Text analytics of online customer reviews. 2016.
- [7] David M Blei, Thomas L Griffiths, and Michael I Jordan. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *Journal of the ACM (JACM)*, Vol. 57, No. 2, p. 7, 2010.
- [8] Thomas L Griffiths, Michael I Jordan, Joshua B Tenenbaum, and David M Blei. Hierarchical topic models and the nested chinese restaurant process. In *Advances in neural information processing systems*, pp. 17–24, 2004.
- [9] Hongning Wang, Yue Lu, and Chengxiang Zhai. Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 783–792, 2010.
- [10] 山本浩平, 江口浩二, 高須淳宏. カテゴリ階層の拡張を目的とした階層的トピックモデル. 第 6 回データ工学と情報マネジメントに関するフォーラム, 2014.