

## レビューテキストを用いた宿泊施設比較のためのアスペクト階層の構築 Constructing Aspects Hierarchies for Accommodation Comparison using Review Text

山口 創也<sup>†</sup> 山田 剛一<sup>†</sup> 増田 英孝<sup>†</sup>

Soya Yamaguchi Koichi Yamada Hidetaka Masuda

### 1. はじめに

消費者はオンラインで宿泊施設を予約する際に、Web上の消費者レビューやプラン情報をもとに吟味し、宿泊施設を選択する。

宿泊施設を地域や宿泊施設形態で絞り込む場合、施設間における類似点が多く、特徴的な相違点がわかりにくいという問題がある。

そこで本研究では、複数の宿泊施設間の特徴比較を支援するために、宿泊施設に関するアスペクトを階層構造化して扱う。本稿では、宿泊施設検索サイトのレビューテキストから抽出した宿泊施設情報を構造化する手法について述べる。

### 2. 関連研究

Lukasz らは修辞構造理論とグラフ分析を使用した教師なし手法によるアスペクト階層の抽出を行なった[1]。

レビューテキストに対し修辞構造解析を行なった結果から、核となるテキストに出現する名詞を上位アスペクト、

衛星となるテキストに出現する名詞を下位アスペクトとし、大量の文書からそれらの関係を集計しグラフ分析することで、ある商品に関する一つのアスペクト階層構造を作った。

### 3. 研究目的

宿泊施設間の違いが分かりにくいいため、宿泊施設決定に多くの時間がかかってしまう。この問題を解決するため、宿泊施設間の違いを効果的に表すアスペクトの可視化を行い、複数の宿泊施設間の特徴比較を支援することを最終目的とする。また可視化を実現するため、アスペクト階層構造を表すツリー構築を本研究の目的とする。なお、本研究で構築する階層の上位下位が概念の上位下位とは限らない。

### 4. 提案手法

本研究では、宿泊施設レビューに現れる名詞をアスペクトとし、アスペクトの出現施設数とアスペクト間の共起回数に着目し、アスペクトの階層を構築する。

#### 4.1 アスペクトの出現施設数の算出

アスペクトに付与する情報として、アスペクトがレビュー中に現れる施設の数（出現施設数）を用いる。データセットからアスペクトごとに出現する出現施設数を集計する。

#### 4.2 文ごとの上位下位アスペクト候補の抽出

以下の手順で上位下位アスペクト候補の抽出を行う。

- (1) 全てのレビューテキストに対して文区切りを行う。
- (2) 各文に対して、品詞が名詞である単語をアスペクトとして抽出する。

- (3) 1文内で、(2)で抽出した全てのアスペクトの組み合わせに対して、4. 1 で求めた出現施設数を元に出現施設数の高いアスペクトを上位候補、出現施設数の低いアスペクトを下位候補として(上位候補、下位候補)のペアを抽出する。なお、2 単語間で出現施設数が同じ場合は、上位下位候補を入れ替えた2ペアを抽出する。

図 1 の場合、出現施設数を考慮したアスペクトについて、朝食:8>パン:3=卵焼き:3 となる。出現施設数の大小関係より、朝食→パン、朝食→卵焼きが抽出され、出現施設数が等しいため、パン→卵焼き、卵焼き→パンが抽出される。

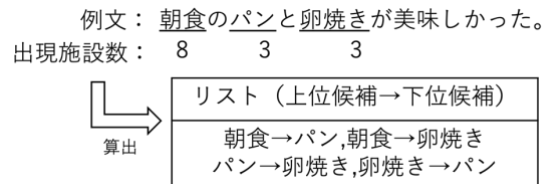


図 1 上位下位候補の抽出

#### 4.3 最上位アスペクトとストップワードの指定

4.1 及び 4.2 で抽出したアスペクトから、あらかじめ指定した 388 単語のストップワードを除く。

階層構築をする際、あらかじめ複数のアスペクトを最上位アスペクトとして指定する。最上位アスペクトは他のアスペクトの子にはならない。

#### 4.4 共起頻度を利用した階層構築

上位下位候補の出現回数を用いて、階層構築を行う。

このとき全ての上位下位候補ペア(上位候補、下位候補)の出現回数を集計する。出現回数が大きい上位下位候補ペアから上位階層として階層を繋ぎ合わせることで、アスペクトの全体階層を構築する。

なお、下位アスペクトが持つ上位アスペクトは出現回数が最大の 1 つに限定する。また上位下位候補ペアの出現回数が同じかつ、同一の下位候補が複数ある場合、上位候補の全文内の出現回数が最大の上位下位候補ペアを採用する。

最上位アスペクトに指定されているアスペクトは出現回数に関わらず最上位アスペクトとなる。

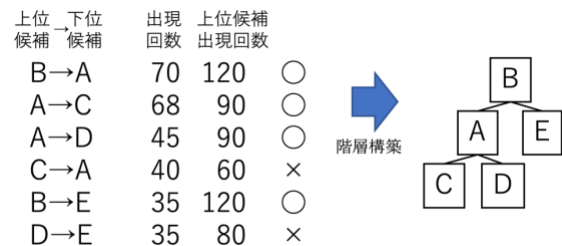


図 2 階層構築の流れ

<sup>†</sup> 東京電機大学 Tokyo Denki University

図 2 では、上位下位候補ペアから階層構築の流れを示す。上位下位候補ペアの出現回数の多い順に階層構築する。B → A, A → C, A → D の順で階層を繋げる。C → A では A はすでに上位に C を持ち、D → E では B → E の方が、上位候補出現回数が多いため階層構築に用いない。

## 5. 実験

実際の宿泊施設レビューを用いて、提案手法を用いて階層を構築した。

### 5.1 使用したデータセット

本研究では、「楽天データ公開」より公開されている楽天データセット[2]を利用する。本実験では、2018 年のデータとして収録されている、レビュー件数上位 40 施設より、各施設 300 件、計 12,000 件のレビューを取得し、階層構築に用いる。

### 5.2 結果と考察

宿泊施設の比較の有用性を評価するため、提案手法を用いた構築した階層構造について、結果と考察を述べる。

本研究では可視化を実現するため、アスペクト階層構造を表すツリー構築を目標とした。よって比較に用いる際に有用であるかという観点から評価考察した。

図 3 は最上位アスペクトが朝食の階層部分を、可視化ライブラリ pyvis を用いて可視化したものである。

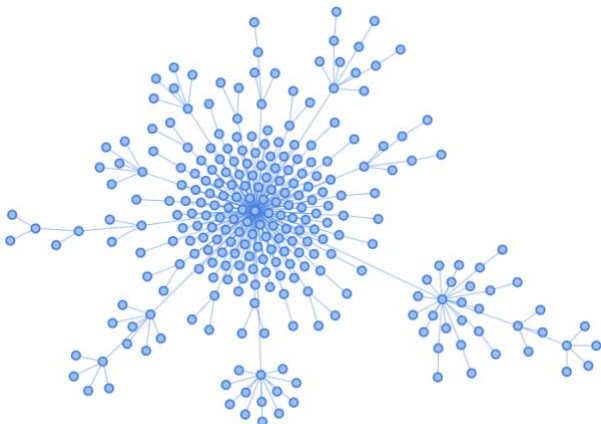


図 3 朝食を中心とした階層構造

#### 5.2.1 出力した階層関係を用いた比較手法

本手法で出力した階層関係を用いた宿泊施設間の比較手法について示す。表形式で特定のアスペクトに関する指標を同指標で比較する。なお、差が現れるアスペクト、及び各マスに出現するアスペクトの関係性は提案手法によって自動的に出力されたものであるが、ホテルごとの特徴は、著者が説明用に振り分けたものである(図 4)。

1 つ目のアスペクトである朝食のパンについて、ホテル A では焼きたて、デニッシュが、ホテル B では、トースターのブランドの一つであるバルミューダが特徴的であることが表現できる。

3 つ目のアスペクトである風呂について、ホテル A では足湯や岩風呂が、ホテル B では岩盤浴や手すりが特徴的であり、どちらのホテルも掛け流しであることが表現できる。

選択宿泊施設 差が現れる アスペクト	ホテルA	ホテルB	共通
1 朝食/パン	焼きたて・黒糖・デニッシュ	トースター・バルミューダ	マーガリン・パンケーキ
2 ホテル/駐車場	地下・提携・契約	立体・有料・屋外	敷地内
3 風呂	足湯・岩風呂・ジャグジー	岩盤浴・手すり・洞窟	掛け流し・数種類
4 部屋/音	高速道路	波	隣室
...	...	...	...

図 4 表形式の可視化法

以上のような可視化手法が実現できる階層を持つことから本階層構築手法は有効である。

#### 5.2.2 不適当な上位下位関係

今回の階層構築手法には不適当な階層関係がいくつかみられた。

例えば、[部屋→チェックイン→チェックアウト]の階層構造が見られ、チェックインの下の階層にチェックアウトが存在するが、これら 2 つのアスペクトは対になる行為を表しているため同じ階層に位置するのが適切である。

これは チェックインとチェックアウトの共起頻度が部屋とチェックアウトの共起頻度より高かったからである。

#### 5.2.3 アスペクトが持つ子の数

アスペクトが持つ子の数にはアスペクトによって大きく異なる。あるアスペクトが持つ子の数は部屋 491 個、朝食 147 個、風呂 55 個であった。このようなアスペクトごとに持つ子の数のばらつきはアスペクトの種類によって関係するアスペクトの数が異なるため適している。しかし、アスペクト階層に注目すると、部屋→ボール、部屋→余裕、部屋→港など関連性を解釈するのが難しい階層を持っている。これは下位アスペクトが出現頻度の高いアスペクトばかり共起しているため起こり、出現頻度が高いアスペクトは必要以上に多くの子を持ってしまったためである。

## 6. おわりに

本研究では宿泊施設に関するアスペクトの階層構築手法を提案した。宿泊施設レビューに現れるアスペクトの出現施設数と上位下位候補の出現回数に着目し、アスペクトの階層構築手法を試みた。

### 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グループ株式会社から提供を受けた「楽天データセット」 ([https://rit.rakuten.com/data\\_release/](https://rit.rakuten.com/data_release/)) を利用した。

### 参考文献

- [1] Augustyniak, Lukasz, Tomasz Kajdanowicz and Przemyslaw Kazienko. "Extracting Aspects Hierarchies using Rhetorical Structure Theory." *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence* (2018)
- [2] 楽天グループ株式会社 (2020): 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.2.2s>