レビューテキストを用いた宿泊施設比較のための依存構造解析を用いたアスペクト階層の構築

山口　創也† 　山田　剛一†† 　増田　英孝†

† 東京電機大学大学院未来科学研究科　〒 120–8551 東京都足立区千住旭町 5
†† 東京電機大学未来科学部　〒 120–8551 東京都足立区千住旭町 5
E-mail: †yamaguchi@cdl.im.dandai.ac.jp, ††{yamada,masuda}@mail.dendai.ac.jp

あらまし　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．宿泊施設を地域や施設の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，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ことで，ある要素に関する特徴を宿泊施設ごとに把握できる．既提案手法では，上位下位関係としてふさわしくないアスペクトペアが出現する問題があり，それを改善するために本手法では並列関係を考慮した．本稿では，依存構造を対象としたパターンマッチングを用いて宿泊施設検索サイトのレビューテキストから宿泊施設情報を抽出し，構造化する手法を提案した．また，その構造に比較したい宿泊施設ごとの単語の出現回数で重み付けを行い，宿泊施設間の相違点と共通点を表す比較表の可視化を提案した．

キーワード　宿泊施設, アスペクト抽出，観光，可視化・データ分類，検索モデル

1 はじめに

近年，Web の発達によりオンライン上で宿泊施設情報を収集し，宿泊施設を予約することが可能になった．宿泊施設情報を得られるオンライン上の情報元は，宿泊施設が公式に運営している Web サイトだけでなく，楽天トラベル [1] などの宿泊施設検索予約サイトが主に利用される．特に宿泊施設検索予約サイトでは施設側が編集した公式情報が掲載されている他にも，実際に宿泊施設を利用した宿泊者だけが感想や評価を自由に書き込める宿泊者レビュー用ページを提供している場合が多い．

レビューは施設側がアピールしたいポイントは関係なく，宿泊者視点の感想が率直に書かれる．そのため，宿泊施設がアピールしていない良い点や悪い点，宿泊施設側のアピールポイントと実態の違いなど，宿泊者ごとに多様な視点で書かれる．このように宿泊者レビューの中には，宿泊施設に宿泊しないとわからないことが掲載されている場合もあることから，宿泊施設の検討者にとって貴重な情報源の一つになっている．

また，現在の宿泊施設検索サイトの宿泊施設検索機能では，地域や価格帯,「温泉がよい」「夜景が綺麗」などのこだわり，といった条件を指定して宿泊施設を検索できる．この機能を活用することで，宿泊施設の検討者は自分の好みや用途に合った宿泊施設を絞り込むことができる．

しかし，宿泊施設検索機能を利用しても，検索結果として出力された複数の宿泊施設間の特徴が似ているため，利用する宿泊施設を決める際に参考となる宿泊施設間の違いが分かりにくい問題がある．例えば，箱根や草津の温泉街では，どの旅館も温泉が入れること，都内のシティホテルでは，どのホテルも客室からの夜景が見えることなどが，似通った特徴となる．このように一定の条件で絞られた複数の宿泊施設は，同じ条件で出力されているため，施設間の類似点が多く，特徴的な相違点が分かりにくい問題がある．そのため，宿泊施設間の違いを踏まえより自分の好み・用途に合った宿泊施設を探し，満足のいく宿泊施設を決定するためには多くの時間がかかってしまう．

そこで本研究では，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．また，その構造に比較したい宿泊施設ごとの単語の出現回数で重み付けを行うことで，宿泊施設間の相違点と共通点を表す比較表の生成を行う．

2 関連研究

アスペクトとはあるアイテムについての評価視点のことである．

著者らは，宿泊施設レビューに現れるアスペクトの出現施設数と 1 文内の共起回数から，上下関係を持つアスペクトペアを抽出した．そして，そのアスペクトペアを出現回数順につなぎ合わせるアスペクトの階層構築手法を提案した [2].

Łukasz 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．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].

Maria らは or や include などの単語の出現パターンや，単語の出現パターンや，文型 (svo，svoc など)，単語間の分布類似

度に着目して概念間の関係を定義づけるコーパスからドメインオントロジーを自動で構築する教師なしフレームワークの提案をした [4].

3　宿泊施設比較法とアスペクト階層化手法

宿泊施設間の違いが分かりにくいため，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，宿泊施設間の特徴比較を支援することを目指す．

図 1 のような表形式の可視化により，宿泊施設間のある対象に関する特徴を比較する．1 行目では，風呂について，ホテル A ではサウナや足湯が，ホテル B では，露天風呂や貸切が特徴的であり，どちらのホテルも大浴場が特徴的であることが表現できる．この可視化法を実現するには，各宿泊施設の特徴を抽出する必要があり，宿泊施設間で共通の階層構造を持つツリーにアスペクトの出現回数を重み付けすることで，各宿泊施設の特徴を抽出する．検索者は宿泊施設間の特徴を明確に知ることができれば，宿泊施設の取捨選択の参考になる．また，共通の特徴も知ることができれば，共通の特徴は比較対象外となるので，検討時間の短縮につながる．

このような可視化を実現するには，宿泊施設間で共通の階層構造を持つツリーに重み付けし，比較するための特徴を抽出する必要がある．1 行目では，階層構造を表すツリー内の風呂の 1 階層下にサウナや足湯，露天風呂などのアスペクトが存在しそれらを分類することで実現できる．また，階層を構成するアスペクト間の関係は，上位アスペクトのより詳細な状況や視点を表現している下位アスペクトで構成されていることや，上位アスペクトについて比較する際に，参考にしたい情報を持つ下位アスペクトで構成されている必要がある．

差が現れるアスペクト ＼ 選択宿泊施設		ホテルA	ホテルB	共通
1	風呂	サウナ・足湯	露天風呂・貸切	大浴場
2	バイキング	海鮮・地元食材・日本酒	ステーキ・ワイン	カレー
3	アメニティ	髭剃り	ケトル・加湿器	ドライヤー・グラス
4	朝食	御膳・和食	バイキング	フルーツ
	…	…	…	…

図 1　表形式の可視化法

3.1　全体の処理の流れ

全体の処理の流れは以下の通りである (図 2).

（1） レビューテキストから，アスペクトペア候補抽出処理を行い，アスペクトペア候補を抽出する．

（2） 抽出されたアスペクトペア候補を用いて，アスペクトツリー構築処理を行い，アスペクトツリーを構築する．

（3） アスペクトツリーを用いて，比較表生成処理を行い，宿泊施設間の比較表を生成する．

アスペクトペアとは，最終的な可視化に使われる際の上位アスペクトと下位アスペクトの組のことである．

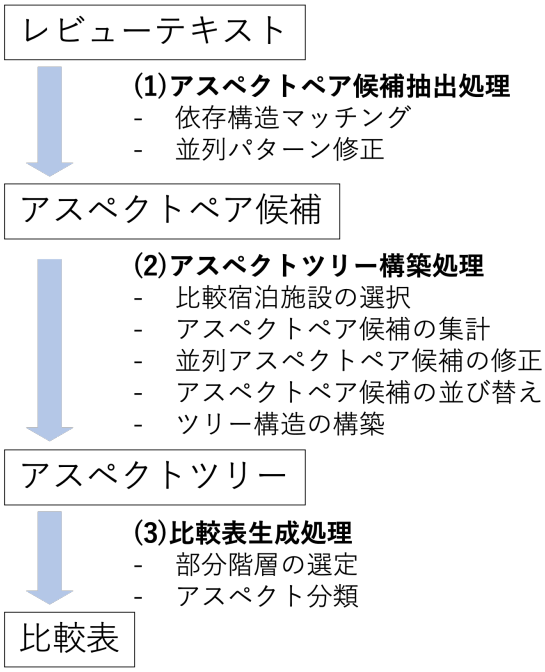

図 2　全体の処理の流れ

3.2　アスペクトの条件

一般的にテキストデータに出現するアスペクトは，名詞や名詞句で表現される [5]．よって本研究においては，レビューテキストに出現する品詞が名詞の単語と，名詞を中心に構成される複合名詞の単語をアスペクトとして抽出する．

3.3　アスペクトペア候補抽出

アスペクトペア候補抽出処理では，レビューテキストに依存構造マッチングと，並列パターン修正を行うことで，アスペクトペア候補を抽出する．出力されるアスペクトペア候補は，(上位アスペクト候補，下位アスペクト候補，抽出パターンの種類，ペアの種類）の形式で，1 文に対して 0 個以上出力される．

3.3.1　依存構造マッチング

依存構造マッチングでは，レビューテキストを依存構造解析し，あらかじめ規定した依存構造パターンにマッチしている文構造を抽出する．その後，その文構造から依存構造パターンに応じて，上位アスペクト候補，下位アスペクト候補を決定する．依存構造マッチングには，さまざまな言語間で文法や品詞，構文の依存関係などの一貫したラベリングを行うフレームワークである Universal Dependencies(UD) に従う依存構造解析の結果を用いる．

a） パターン「が・は」

パターン「が・は」では，あるアスペクトに関するより詳細な情報を捉えることを意図し，アスペクトペア候補を生成する．品詞が名詞である「wordA」から品詞が助詞である「が・は」に係り，かつ,「wordB」が「wordA」に依存構造ラベル nsubj を持って係るとき，パターン「が」「は」を抽出する．出力される

アスペクトペアは,（wordA,wordB, が, 上下）となる (図 3)．例えば「朝食がパンでした.」からは（朝食, パン, が, 上下）と出力される．これ以降に，紹介するパターンについての図も同様の説明を表す．

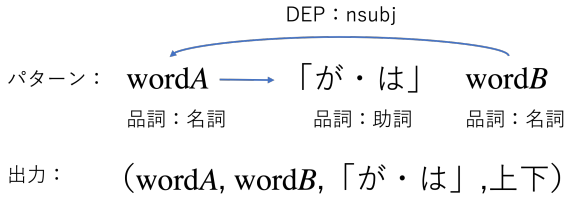

図 3　パターン「が・は」の抽出

b）　パターン「の」

パターン「の」では，包含関係や所有関係を捉えることを意図し，アスペクトペア候補を生成する (図 4)．例えば「バイキングの品揃えが良かったです.」からは（バイキング, 品揃え, の, 上下）と出力される．

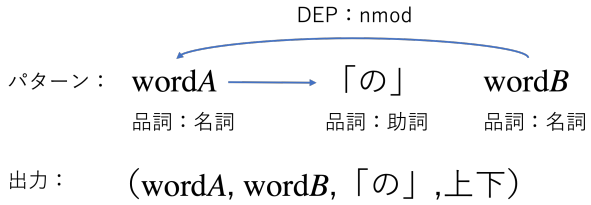

図 4　パターン「の」の抽出

c）　パターン「名詞節修飾」

パターン「名詞節修飾」では，あるアスペクトに関するより詳細な情報を捉えることを意図し，アスペクトペア候補を生成する (図 5)．例えば,「湖が見える部屋で良かったです.」からは（部屋, 湖, 名詞節修飾, 上下）と出力される.「[]」は品詞を指定しないことを表す.

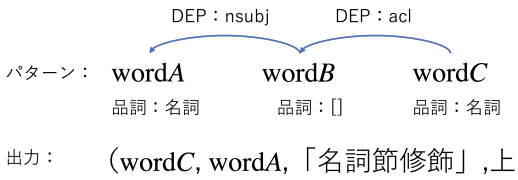

図 5　パターン「名詞節修飾」の抽出

d）　パターン「と・や」

パターン「と・や」では，並列関係を捉えることを意図し，アスペクトペア候補を生成する (図 6)．例えば,「寿司と味噌汁が美味しかった.」からは（寿司, 味噌汁, と, 並列）と出力される．

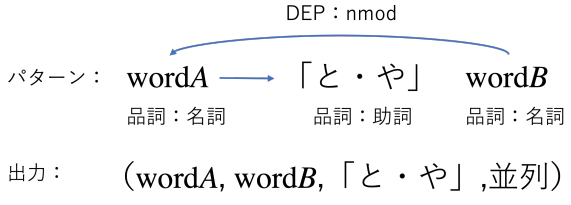

図 6　パターン「と・や」の抽出

e）　パターン「も」

パターン「も」では，並列関係を捉えることを意図し，アスペクトペア候補を生成する (図 7)．例えば,「夕食も朝食も良かったです.」からは（夕食, 朝食, も, 並列）と出力される．

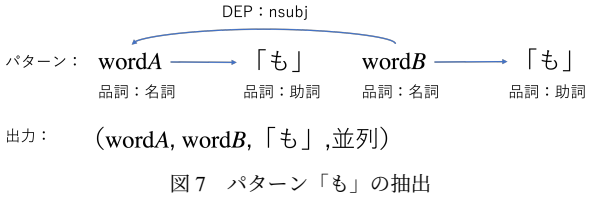

図 7　パターン「も」の抽出

f）　パターン「複合語」

パターン「複合語」では，あるアスペクトのより詳細な情報を抽出することを意図し，アスペクトペア候補を生成する (図 8)．例えば,「ファミリールームが広かったです.」からは（ルーム, ファミリールーム, 複合語, 複合語）と出力される．

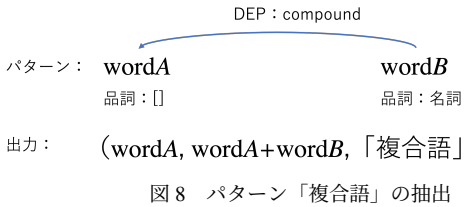

図 8　パターン「複合語」の抽出

3.3.2　並列パターン修正

アスペクトペア候補抽出時に，1 文内で上下アスペクトペア候補の下位アスペクトが，並列アスペクトペア候補としても同時に存在する時，下位アスペクトをもう一方の並列アスペクトペアのアスペクトに入れ替えて上下アスペクトペア候補を生成する (図 9)．

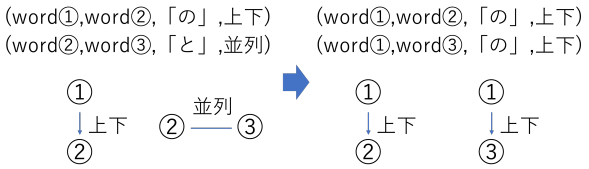

図 9　並列パターン修正処理

3.4　アスペクトツリー構築処理

比較する宿泊施設のアスペクトペア候補と，全ての宿泊施設のアスペクトペア候補を抽出し，それぞれ集計する．その後，アスペクトペア候補抽出パターンや，アスペクトの出現回数などでアスペクトペア候補を並び替えし，アスペクトツリーを構築する．

3.4.1　アスペクトペア候補の並び替え

アスペクトペア候補の集計結果を用いてアスペクトペア候補を表 1 の順に並び替える．上下パターンとは，先述したパターン「が・は・の・名詞節修飾」のことであり，全パターンとは，上下パターンと並列パターンを合わせた先述した全てのパターンをことである．上下パターン出現回数 (選択施設合計) や，全パターン出現回数 (選択施設合計) など，選択宿泊施設のレビューに出現する特徴を優先するように並び替えることで，

選択宿泊施設にあった階層構造を構築する．また，並び替えには，固有名詞を含むアスペクトペアの処理も行うことで，宿泊施設の状況を把握しやすくする．

表 1　アスペクトペア候補の集計項目
並び替え基準の優先順位

上下パターン出現回数（選択施設合計）
全パターン出現回数（選択施設合計）
上下パターン出現回数（全施設）
全パターン出現回数（全施設）
上位アスペクト候補出現回数（選択施設合計）
下位アスペクト候補出現回数（選択施設合計）

a）　固有名詞を含むアスペクトペアの処理

固有名詞以外のアスペクトの上位に固有名詞のアスペクトが存在すると，階層の状況の把握が困難になる恐れがあるため，固有名詞を含むアスペクトは子に固有名詞以外のアスペクトを持たない制約を設ける．これにより，固有名詞以外の語で中間層以上のアスペクトツリーが構築され，下位階層で固有名詞のアスペクトが出現するようになる．

3.4.2　ツリー構造の構築

並び替え後アスペクトペア候補の上位のアスペクトペア候補から繋ぎ合わせていくことでツリーを構築していく．このとき，あるアスペクトが持つ上位アスペクトを 1 つに限定することで，木構造を持つようにする．

図 10 の例では，B → A，A → C，A → D の順で構築していく．D → C は，C はすでに上位に A を持っているため構築には使用しない．その後，B → E を構築するようにして，ツリー構造を構築する．

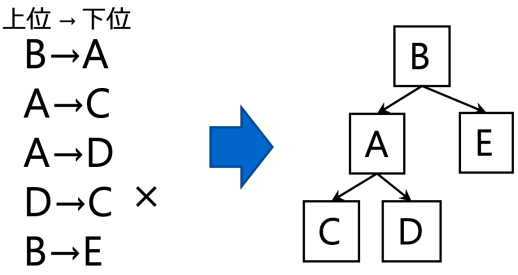

図 10　ツリー構造の構築

3.5　比較表生成処理

比較表生成処理では，構築されたアスペクトツリーから可視化する部分階層 (2 階層) を抽出する．その後，その部分階層の下位アスペクトを，各宿泊施設の特徴と，宿泊施設間で共通の特徴に分類し，比較表を生成する．

3.5.1　部分階層の選定

ここでは，宿泊施設間の違いが大きく現れる下位アスペクトをより多く持つアスペクト階層を可視化するアスペクト階層として抽出する (図 11)．まず部分階層の下位アスペクトごとに，出現回数で宿泊施設間で差を取る．その後，下位アスペクトを，宿泊施設間の出現回数の差の昇順に並び替え，差の値の最大・最小から，両端 1/3 未満の個数ずつ抽出する．それらの分散の値が大きい階層ほど，可視化すべき部分階層とする．

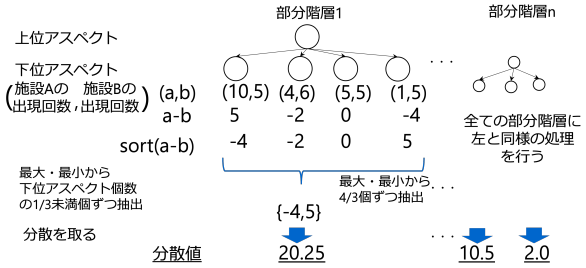

図 11　部分階層の選定

3.5.2　アスペクト分類

可視化する部分階層の下位アスペクトを，アスペクト出現頻度の出現施設間の割合から各宿泊施設の特徴と共通の特徴に分類する．アスペクトの出現頻度の宿泊施設間の割合をとり，両端 1/3 を各ホテルの特徴，中央 1/3 を共通の特徴とする．この分類は式 (1) の f(a,b) で表される．a,b はそれぞれあるアスペクトの各宿泊施設の出現回数である．

$$f(a,b) = \begin{cases} \text{ホテル A} & \frac{2}{3} < \frac{a}{a+b} \text{のとき} \\ \text{共通} & \frac{1}{3} < \frac{a}{a+b} \leqq \frac{2}{3} \text{のとき} \\ \text{ホテル B} & \frac{a}{a+b} \leqq \frac{1}{3} \text{のとき} \end{cases} \quad (1)$$

4　評価実験

提案手法により生成された比較表が宿泊施設間の比較に有用であるかどうかを検証するため，評価実験を行った．依存構造解析には，日本語自然言語処理ライブラリである GiNZA を使用した．これ以降に，出力されたアスペクトツリーの観察，比較表の可視化に対する評価，比較表を構成するアスペクトペアに対する評価について述べる．

4.1　使用するデータセット

本研究で使用する宿泊施設レビューは，楽天トラベルデータセットに含まれるものである．本実験では，2019 2015 年のデータとして収録されている，レビュー件数上位 100 施設より，各施設 1,000 件，計 100,000 件のレビューを取得し，階層構築に用いる．今回は，宮城県の観光地にある二つの宿泊施設 (ホテル A：秋保温泉 秋保グランドホテル，ホテル B:仙台 秋保温泉 ホテル瑞鳳) を比較に用いた．

4.2　出力されたアスペクトツリー

図 12 は最上位アスペクトが朝食のアスペクトツリーを，可視化ライブラリ pyvis を用いて可視化したものである．ノードがアスペクトを表し，エッジで繋がれている 2 つのアスペクトの組をアスペクトペアとして表している．この図から，アスペクトによって，そのアスペクトを上位に持つアスペクトペアの数が異なることがわかる．また，アスペクトによって，そのアスペクトの下位のツリーの深さも異なることがわかる．

実際には最上位アスペクトが朝食以外の部屋や風呂といった

アスペクトツリーも存在しており，これらは比較する宿泊施設によって全く別の階層を構築する．

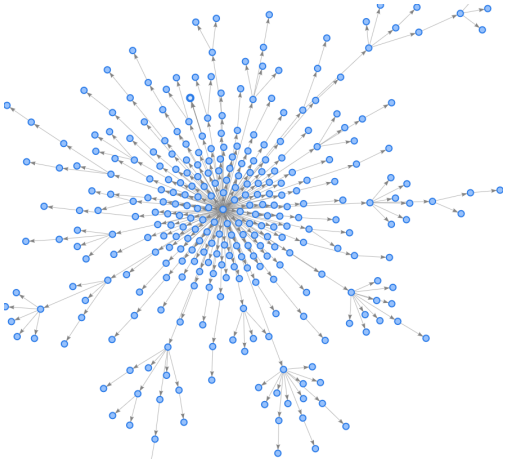

図 12　出力されたアスペクトツリー

図 13 は，出力されたアスペクトツリーの「部屋→窓」の下位の部分階層を抜き出したものである．ノードの色はアスペクトの施設ごとの出現回数の割合を表しており，ホテル A の割合が大きいほど赤色，ホテル B の割合が大きいほど青色となるグラデーションで表している．この部分階層からは，ホテル A で部屋の窓から渓谷や紅葉が見えることが想像できる．

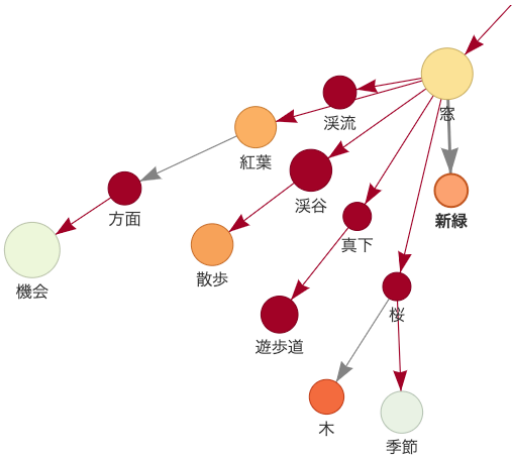

図 13　部分階層の例

4.3　実験方法

提案手法で生成した 4 つの比較表を用いて，10 人の被験者に，比較表の可視化に対する評価と，比較表を構成するアスペクトペアに対する評価をしてもらった．

4.4　結　　果

表 2 は比較表の可視化に対する 4 段階評価の平均である．
質問 a：比較表は宿泊施設間の差を把握しやすいか．
質問 b：比較表を宿泊施設を選択する際に参考にしたいか．

表 2　比較表の可視化に対する評価結果

	質問 a	質問 b
夕食	2.9	2.9
プラン	3.2	2.7
バイキング	2.2	2.5
食事	2.5	2.8

表 3 は比較表を構成するアスペクトペアに対する 4 段階評価の平均である．
質問 a
対象と特徴から状況を把握できるか，
質問 b
対象について，宿泊施設間の違いを把握できる特徴か．
質問 c
この形態の施設を検討するとき参考にしたい情報か．

表 3　アスペクトペアに対する評価結果

分類	上位	下位	質問 a	質問 b	質問 c
ホテル A	夕食	せり	2.1	1.8	1.6
ホテル A	夕食	前菜	2.6	2.1	2.0
ホテル A	夕食	コース	3.2	2.9	2.8
ホテル B	夕食	デザート	2.9	2.2	2.2
ホテル B	夕食	噂	1.4	1.2	1.5
ホテル B	夕食	中華	3.6	3.5	3.2
ホテル B	夕食	刺身	3.4	3.0	3.1
ホテル B	夕食	ズワイガニ	3.9	3.6	3.5
共通	夕食	バイキング	3.8	3.6	3.6
共通	夕食	カニ	3.5	3.1	3.1
共通	夕食	ビュッフェ	3.7	3.4	3.6
共通	夕食	経験	1.1	1.2	1.0
共通	夕食	スタイル	1.4	1.5	1.4

4.5　考　　察

表 2 より，質問 a の数値が高いとき，質問 b の数値も高くなることから，宿泊施設間の差を把握できる表であるとき，宿泊施設を選択する際に参考にしたい表であるといえ，比較表の可視化は宿泊施設を選択するときに有用であると考えられる．

表 3 より，「夕食→ズワイガニ」や，「夕食→ビュッフェ」などは宿泊施設間の違いを把握できる特徴であると評価された．これより下位アスペクトが上位アスペクトの内容や形式を表しているものは，宿泊施設間の違いを把握できる特徴であると考えられる．

一方で，「夕食→前菜」，「夕食→デザート」などは，宿泊施設間の違いを把握できない特徴と評価された．これより，カテゴリのような抽象的なアスペクトの下位に，さらに抽象的なアスペクトが存在する場合，具体性に欠けているため，宿泊施設間の違いを把握することには使えないと考えられる．

5 おわりに

本研究では，宿泊施設間の比較の支援を目的とし，宿泊施設間の違いを可視化するためのアスペクトツリーの構築と可視化手法を提案した．提案手法を用いて生成した比較表に対して，評価実験を行った結果，比較表の可視化の有効性を確認した．

今後は，アスペクト抽出時に，アスペクトに係っている評価表現も同時に抽出することで，可視化時にアスペクトの評価情報も同時に可視化したり，評価情報から可視化すべきアスペクトを絞り込むことで，より状況の把握ができる比較表の可視化を実現できると考えられる．また，アスペクトペア抽出の精度評価，同義語の処理なども検討していく必要がある．

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/) を利用した．

文　　献

[1] 楽天グループ株式会社，“楽天トラベル：宿・ホテル予約国内旅行・海外旅行予約サイト，” 入手先〈https://travel.rakuten.co.jp〉，参照 2022-12-20.

[2] 山口創也, 山田剛一, 増田英孝，“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，” 第 21 回情報科学技術フォーラム講演論文集, 第 21 巻, pp.161–162,2022.

[3] Łukasz Augustyniak et al. , “Method for aspect-based sentiment annotation using rhetorical analysis, ”Intelligent Information and Database Systems, pp. 772–781，2017.

[4] Mukherjee, S et al. , “Domain cartridge: Unsupervised framework for shallow domain ontology construction from corpus, ”In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp.929-938，2014.

[5] Maria Pontiki et al. , “SemEval-2015 task 12: Aspect based sentiment analysis，”In Proceedings of the 9th International Workshopon Semantic Evaluation, pp.486–495，2015.