

地域特性を表現する概念と地理的特徴の関係性分析 Comparative Analysis between Description of Regional Characteristics and Geographical Features

福地 湧*
Yu Fukuchi

牛尼 剛聡†
Taketoshi Ushiyama

角谷 和俊*
Kazutoshi Sumiya

1. はじめに

近年、ホテルなどの宿泊施設に対するテキストは多数存在し、楽天トラベル [1] などの宿泊サイト上には利用者による多くのレビューが掲載されている。ホテルのレビューには、部屋数や清潔感などホテル自体を評価するものもあるが、「駅から徒歩5分」や「コンビニが目の前」のように、他の地物との関係性から目的のホテルを評価するレビューも多い。宿泊客は周辺の地物との関係性からもホテルを評価する場合がある。レビューからホテルの地域特性を抽出できれば、ユーザがホテルを選択するときにも有用であると考えられる。そこで本研究は、レビュー中に出現する周辺地物との関係性を示す「地域特性を表現する概念」に基づいてホテルと周辺地物の地理的関係性を抽出することを目的とする。

以下、2章では関連研究との比較について述べる。3章では本研究のアプローチについて述べる。4章では地域特性を表す概念からホテルと周辺地物の関係性を読み取る手法について述べる。5章では本手法の有用性を実験結果に基づいて述べる。最後に、6章でまとめと今後の検討について述べる。

2. 関連研究

地物間の関係性に関して、嬉野 [2] らは地物同士の主観的な距離感覚を分析する手法を提案しているが、本研究は地域特性を表現する概念に着目することで、距離的關係だけでなく位置的關係も分析している点で異なる。また、大塚 [3] らは異なる地物カテゴリ間の関係性を地物間の距離から分析しているが、本研究は地域特性を表す概念から地物間の関係性を読み取る点で異なる。

本研究は、レビューからホテルの特徴を読み取る点に関しては、安部 [4] らの研究と類似している。しかし、安部らはホテル自体の評価値を予測するが、本研究はレビュー上の地域特性を表す概念から、他の地物との関係性を読み取る点で異なる。

3. アプローチ

本論文では、地理的関係性を表現するテキストに基づいてホテルの特徴を抽出する手法を提案する。提案手法の概要を図1に示す。

本手法では、まず、対象とするホテル h と、そのホテルに対するオンラインレビューを入力する。次に地域特性を表現する概念に着目する。例えば、ホテル h のオンラインレビュー上に「駅チカ」という地理的概念が頻出していればホテル h には「駅から近い」という特徴

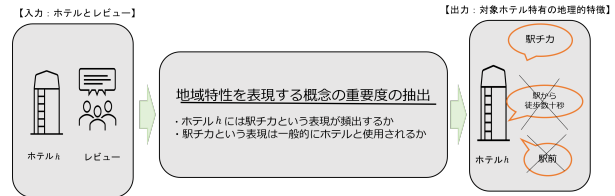


図 1: 提案手法の流れ

があるものとみなす。さらに、駅チカという表現を含むオンラインレビューがレビュー全体を通して少ないのであれば、「駅チカ」はホテル h の特徴的な地理的特性だと考えることが出来る。こうした観点からホテル h における地域特性を表す概念の重要度を測る。最後に、重要度が高い概念を対象ホテルの特徴として出力する。この場合、「駅チカ」がホテル h の特徴として出力されている。この処理によって、他の地物との地理的関係性から各ホテル特有の特徴を読み取ることが可能になる。

4. 提案手法

4.1. ホテルに対する比較地物の選定方法

ホテルと周辺地物の地理的関係性を調べるために、事前準備として、ホテルが有する各特徴に影響を与える地物を選定した。本研究では、ホテルの特徴に影響を与える地物をホテルに対する周辺地物と呼ぶ。今回は、周辺地物から影響を受けるホテルの特徴として「アクセス」「食事」「くつろぎ」の3つを考えた。また、各特徴に関するレビュー中で出現頻度が高い上位5件の地物を周辺地物と仮定した。例えば、ホテルの「アクセス」という特徴に影響する地物はアクセスに関するレビュー中で出現頻度が高い上位5件の地物となる。データセットは楽天グループ株式会社が国立情報学研究所で公開している楽天トラベルのレビューデータのうち、任意に抽出した1000件を使用した。結果を表1に示す。以後、表1に示した周辺地物を対象にホテルとの地理的関係性を考察する。

表 1: ホテルに影響を与える地物

各特徴に影響を与える地物	
アクセス	駅, コンビニ, 観光地, バス, 繁華街
食事	コンビニ, 駅, 居酒屋, スーパー, 繁華街
くつろぎ	駅, 山(緑), 大通り, 繁華街, 居酒屋

*関西学院大学総合政策学部メディア情報学科, School of Policy Studies, Kwansai Gakuin University.

†九州大学大学院芸術工学研究院, Faculty of Design, Kyushu University.

4.2. 地理的關係性の抽出方法

4.2.1. 地域的特性を表現する概念

本研究では対象ホテルの地理的關係性を読み取るために、地域特性を表現する概念に着目する。地域特性を表現する概念とは、他の地物との距離関係や位置関係を表すものとする。今回は全てホテルを始点としており、表 2 に示す概念を対象とする。距離的關係に関してはホテルからの近さに応じてグループ化している。

表 2: 地域特性を表現する概念

距離的關係性を表す概念	位置的關係性を表す概念
すぐ・徒歩 1 分以内/間近/直結	前
近い・徒歩 5 分以内/徒歩圏内	隣・横
少し離れた/距離がある/徒歩 10 分	中心/周りに

4.2.2. 地域特徴を表す概念の頻出度

本研究では、対象とするホテルのレビュー上に頻繁に現れる表現ほど、対象ホテルの特徴であると仮定する。例えば、対象ホテルの周辺地物コンビニに関するレビュー中に「コンビニ」と地理的概念「前」が共起するレビューが多ければ、対象ホテルには「目の前がコンビニ」という地理的な特性があると考えられる。ホテル h に関するレビューにおける周辺地物 i と地理的概念 c の共起度 $rf(i, c, h)$ を式 (1) として定義する。

$$rf(i, c, h) = \frac{|R(i, c, h)|}{|\bigcup_{x \in C} R(i, x, h)|} \quad (1)$$

ここで C はすべての地理的概念の集合を表し、 $R(i, c, h)$ は周辺地物 i と地理的概念 c が共起するホテル h のレビュー集合を指す。この値からホテル h は周辺地物 i とどのような地理的關係性にあるかを予測できる。

4.2.3. 地域の特徴を表す概念の重要度

地域特性を表現する概念と周辺地物の関連性の重要度 $w(i, c)$ を次式より算出する。

$$w(i, c) = \log \frac{|\bigcup_{x \in C} \bigcup_{y \in H} R(i, x, y)|}{|\bigcup_{y \in H} R(i, c, y)| + 1} \quad (2)$$

ここで H はすべてのホテルから成る集合を表す。式 (2) から周辺地物 i とホテルの地理的概念 c の関係が稀であるかを分析できる。この値が大きいつき、ホテルと地物 i が地理的概念 c の関係にあることは稀であるため、 i と c の関係はホテルにとって重要な地理的關係性であると考えられる。最後に $w(i, c)$ を式 (1) に重みづけすることで、ホテル h における周辺地物 i と概念 c の地理的關係性を示すスコア $Score(i, c, h)$ を算出する。 $Score(i, c, h)$ が高いとき、ホテルと地物 i が c の関係にあることがホテル h にとって地理的な特徴だと判定できる。

$$Score(i, c, h) = rf(i, c, h) \times w(i, c) \quad (3)$$

$Score(i, c, h)$ の一覧を表 3 に示す。国立情報学研究所の IDR データセット提供サービスで楽天グループ株式会社が公開している楽天トラベルのレビューデータのうち、周辺地物を含む任意のレビューを基に作成した。地域特性を表現する概念は距離的概念と位置的概念に区別して表を作成した。値が 1 より大きいものは稀な地理的關係性であるため、ホテル特有の特徴になり得るが、値が 1 より小さいものは頻出度が高い地理的關係性であるため、ホテルの特有の特徴ではないと考えられる。駅やコンビニに関してはデータを多く得られたが、それ以外の地物、特に、「大通り」に関してはレビューが少なくデータを得ることが困難だった。また、一部の周辺地物に関してはレビュー数が少なかったために $R(i, c, y)$ の要素数が 0 になることが多かった。

表 3: 地域特性を表現する概念の特徴スコア一覧

周辺地物 i	地域特性を表す概念 c (距離的概念)	$R(i, c, y)$	$w(i, c)$
駅 (334件)	すぐ・徒歩 1 分以内・間近・直結	40	0.911
	近い・徒歩 5 分以内・徒歩圏内	112	0.471
	少し離れた・距離がある・徒歩 10 分	13	1.378
コンビニ (63件)	すぐ・徒歩 1 分以内・間近・直結	5	1.021
	近い・徒歩 5 分以内・徒歩圏内	22	0.438
	少し離れた・距離がある・徒歩 10 分	3	1.197
観光地 (30件)	すぐ・徒歩 1 分以内・間近・直結	2	1.000
	近い・徒歩 5 分以内・徒歩圏内	15	0.273
	少し離れた・距離がある・徒歩 10 分	4	0.778
バス (13件)	すぐ・徒歩 1 分以内・間近・直結	2	0.637
	近い・徒歩 5 分以内・徒歩圏内	3	0.512
	少し離れた・距離がある・徒歩 10 分	0	1.114
繁華街 (36件)	すぐ・徒歩 1 分以内・間近・直結	5	0.778
	近い・徒歩 5 分以内・徒歩圏内	15	0.352
	少し離れた・距離がある・徒歩 10 分	4	0.857
居酒屋 (14件)	すぐ・徒歩 1 分以内・間近・直結	1	0.845
	近い・徒歩 5 分以内・徒歩圏内	7	0.243
	少し離れた・距離がある・徒歩 10 分	0	1.146
スーパー (13件)	すぐ・徒歩 1 分以内・間近・直結	1	0.813
	近い・徒歩 5 分以内・徒歩圏内	2	0.637
	少し離れた・距離がある・徒歩 10 分	2	0.637
山 (26件)	すぐ・徒歩 1 分以内・間近・直結	1	1.114
	近い・徒歩 5 分以内・徒歩圏内	3	0.813
	少し離れた・距離がある・徒歩 10 分	2	0.938
大通り (3件)	すぐ・徒歩 1 分以内・間近・直結	1	0.176
	近い・徒歩 5 分以内・徒歩圏内	0	0.477
	少し離れた・距離がある・徒歩 10 分	2	0.000

周辺地物 i	地域特性を表す概念 c (位置的概念)	$R(i, c, y)$	$w(i, c)$
駅 (334件)	前	29	1.047
	隣・横	7	1.621
	中心・周りに	4	1.825
コンビニ (63件)	前	1	1.498
	隣・横	5	1.021
	中心・周りに	3	1.197
観光地 (30件)	前	2	1.000
	隣・横	0	1.477
	中心・周りに	10	0.436
バス (13件)	前	5	0.336
	隣・横	0	1.114
	中心・周りに	0	1.114
繁華街 (36件)	前	0	1.556
	隣・横	0	1.556
	中心・周りに	6	0.711
居酒屋 (14件)	前	2	0.669
	隣・横	0	1.146
	中心・周りに	3	0.544
スーパー (13件)	前	1	0.813
	隣・横	4	0.415
	中心・周りに	1	0.813
山 (26件)	前	0	1.415
	隣・横	0	1.415
	中心・周りに	5	0.637
大通り (3件)	前	1	0.176
	隣・横	0	0.477
	中心・周りに	0	0.477

5. 評価実験

5.1. 実験設定

提案手法の有効性を評価するために、「神戸元町東急 REI ホテル」を対象ホテルとして地理的な特徴を抽出した。ここでは表 3 に示す周辺地物の中で、駅とコンビニ

二つの地理的関係性から、当該ホテルの地理的特徴を分析した。使用したデータは楽天トラベルのレビューである。「神戸元町東急 REI ホテル」のレビューの中で「駅」と「コンビニ」を含むそれぞれのレビュー 100 件を対象とした。

表 4: 神戸元町東急 REI ホテルの地理的スコア

神戸元町東急REIホテルとコンビニの地理的関係性			
地域特性を表す概念c		$rf(i, c, h)$	$Score(i, c, h)$
距離的関係性	すぐ/徒歩1分以内/間近/直結	15	24.94
	近い/徒歩5分以内/徒歩圏内	57	30.47
	少し離れた/距離がある/徒歩10分	0	0.00
位置的関係性	前	5	9.00
	隣・横	1	1.10
	中心/周りに	4	5.23

神戸元町東急REIホテルと駅の地理的関係性			
地域特性を表す概念c		$rf(i, c, h)$	$Score(i, c, h)$
距離的関係性	すぐ/徒歩1分以内/間近/直結	7	6.38
	近い/徒歩5分以内/徒歩圏内	77	36.24
	少し離れた/距離がある/徒歩10分	7	9.64
位置的関係性	前	0	0.00
	隣・横	0	0.00
	中心/周りに	7	12.77

5.2. 結果と考察

結果を表 4 に示す。今回は対象とした駅とコンビニのレビューが共に 100 件と同じであるため、地物と各地理的概念が共起するセンテンス数を $rf(i, c, h)$ とした。この結果より、「神戸元町東急 REI ホテル」と駅は、「近い、徒歩圏内」の距離関係にあり、ホテルを中心に駅がいくつか点在する位置関係にあることが分かる。一方で、コンビニに関しては、ホテルから「すぐ」ではなく「徒歩 5 分程度」の距離にあることが示されているが、位置関係ではホテルの「前」にあると示されていて、矛盾を生む結果となった。しかし、重みづけにより $Score$ (コンビニ, すぐ/徒歩 1 分以内 $\cdot\cdot$, 神戸元町東急 REI ホテル) が大きかった一方で、 $Score$ (コンビニ, 近い/徒歩 5 分以内 $\cdot\cdot$, 神戸元町東急 REI ホテル) は小さな値となった。実際には当該ホテルと最も近くにあるコンビニの距離が 26m であるため、重みづけによって $Score$ (コンビニ, すぐ/徒歩 1 分以内 $\cdot\cdot$, 神戸元町東急 REI ホテル) が大きくなったことは評価できる。

今後の課題として、地理的概念を自動で検出することが挙げられる。今回は人手で地域特性を表す概念を絞って分析したため、調査できていない概念が多数あり、取り扱えたデータ数も少なく、コンビニと駅以外の比較地物では、意味のある地理的スコアを得ることは困難だった。

6. おわりに

本研究では、オンラインレビューに基づき、周辺地物との地理的関係性からホテルの特徴を抽出する手法を提案した。本手法では、ホテルとの周辺地物と地域特性を表現する概念の共起度や概念自体の重要度から対象とするホテル特有の地理的な特徴を抽出する。評価実験においては、特定のホテルの地理的特徴をコンビニ・駅との地理的関係性から分析した。その結果、提案手法の重みづけがホテルの地理的特徴を表すことが出来る可

能性があることが分かった。

今後の課題として、今回得られた地物間の地理的関係性以外の関係性を画像解析を用いて自動的に抽出することが挙げられる。今回は人手で分析したことにより、得られたレビューが少なく、読み取れる地理的関係性に限りがあったため、今後は画像解析により地物間の距離や位置の関係を自動的に抽出する必要がある。また、今回はホテルと特定の比較地物の 1 対 1 の関係性を読み取ったが、今後はホテルと複数の地物の距離的關係・位置的關係を分析することで 1 対多の地理的関係性を読み取れることを視野に入れる。ホテルだけでなく、アパートや駅などの地物も研究の対象とすることで不動産関連のアプリケーションや地域特性の抽出に応用することを検討する。

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。また、本研究の一部は、令和 4 年度科研費基盤研究 (B) (課題番号: 19H04118), 令和 4 年度科研費基盤研究 (B) (課題番号: 22H03905) の助成を受けたものである。

参考文献

- [1] 楽天トラベル, <https://travel.rakuten.co.jp/>
- [2] 嬉野 孝延, 北山 大輔, 角谷 和俊, "地物カテゴリ間の距離の集合知に基づく地物表示制御", 第 142 回情報基礎とアクセス技術, Vol.2021-IFAT-142, No.12(2021).
- [3] 大塚公貴, 角谷和俊, "地物間の距離と関係性による地域特性分析", DEIM Forum 2019, P2-120(2019).
- [4] 安部克, 酒井聖, 松波友稀, 中島伸介, "レビュー評価項目別スコアを用いたホテル推薦手法の提案", DEIM Forum 2019, P1-134(2019).