

地物間の距離と関係性に基づく地域特性分析

大塚 公貴[†] 角谷 和俊[†]

[†] 関西学院大学総合政策学部 〒 669-1337 兵庫県三田市学園 2 丁目 1

E-mail: [†]{ezn23068,sumiya}@kwansei.ac.jp

あらまし 近年、旅行者が旅先を決定する際じゃらんなどの観光情報サイトを利用する例が増えている。しかし、既存の観光サイトでは観光地の情報は入手できても、その地域の情報を入手することはできない。そこで本研究では、地物 A に対する周辺の状況から地域的な特性を抽出する。地域特性を抽出することにより、ユーザが訪れたことのない土地に訪れる際、その土地の特性を容易に把握可能となる。

キーワード 地理情報, 情報推薦, 機械学習

1. はじめに

近年、旅行者が旅先を決定する際じゃらんなどの観光情報サイトを利用する例が増えている。しかし、既存の観光サイトでは地物の人気ランキングやレビューなどの情報が全面に押し出され、地域的な特性を記載している例は少ない。本研究では地物に対する周辺の地物から地域的な特性を抽出する。これによりユーザは訪れたことのない場所のイメージが容易に行えるようになる。例えば、繁華街のラーメン屋と田舎にあるラーメン屋では同じラーメン屋でも周囲にある地物は異なっている。つまりある地物に対する周囲にある地物を分析することにより、その周辺が繁華街であるのか田舎であるのかなどの特性を分析できると我々は考えた。本研究はこのような地物間の関係性を複数集めることにより、地域的な特性を抽出することを目的とする。2 章では関連研究から本研究の位置づけを説明し、3 章では提案手法について述べる。4 章ではカテゴリ間の関係性について説明し、5 章ではその実験と考察を行う。6 章でまとめと今後の課題について述べる。

2. 関連研究

地理情報に関する研究は数多く行われてきた。成川らは略地図に関する研究で地物の重要度を判定した。[1] 地物間の関係性を抽出する際に参考になっている。また、吉田らはタグなどのメタデータでは推測困難な「夜景が美しい」、「待ち合わせ場所に向いている」などの主観的特徴を考慮した観光スポット検索システムを提案した。[2] ユーザレビューを用いる上で参考になっている。地域分析において野守らは、日本全国の観光地の口コミデータを用いて、観光客が話題にする観光テーマを確率的に抽出し、そのテーマを軸として各観光地の特徴を定量的に評価した。口コミデータを用いる点は類似しているが、カテゴリ構造を用いるという部分で本研究とは異なっている。[3]

3. 提案手法

3.1 カテゴリ分類

本研究では地域特性の抽出にあたって地物のカテゴリ構造とじゃらんのユーザレビュー、距離の 3 つを用いる。まずは地物のカテゴリ構造について説明する。地物を、タウンページの電話帳カテゴリを用いて、カテゴリへ分類した。以下はその例である。

- ・旅行・宿泊>ホテル・ペンション
- ・ショッピング>コンビニ・スーパー・デパート>コンビニエンスストア
- ・旅行・宿泊>観光・温泉・銭湯>温泉浴場
- ・レジャー・スポーツ>博物館・美術館>博物館
- ・レジャー・スポーツ>動物園・植物園・水族館
- ・グルメ・飲食>飲み屋
- ・旅行・宿泊>おみやげ
- ・自動車・バイク>駐車場・パーキング

図 1 カテゴリ分類

このようなカテゴリを用いることにより、地物どうしの関係からカテゴリ間の関係を抽出可能となる。例えば、カテゴリ「レジャー・スポーツ」とカテゴリ「ショッピング」の関係性を求めたい際には、レジャー・スポーツに属する実際の地物「万博記念公園」とショッピングに属する実際の地物「エキスポシティ」の関係性をレビューから抽出する。このような地物間関係を膨大な数集めることによりカテゴリ「レジャー施設」とカテゴリ「ショッピング」の関係性抽出が可能となる。次に影響度について述べる。

3.2 影響度

本研究における影響度とは、地物のレビュー内にある他カテゴリの地物を指す。例えば、ホテルに関するレビュー中に「コンビニ」に関する記述があった場合、ホテルがコンビニに影響を与えているとし、影響度があると考ええる。今回は一つの地物に対して 100 件のレビューを目視で確認し、その中出现する他カテゴリの地物の数を数えた。影響度判定を行う際には、じゃらんのレビューを用いた。具体的には以下ようになる。

ホテル	コンビニエンスストア	影響度
ホテルアルファワン横浜関内	コンビニエンスストア	33
京王プレッソイン新宿	コンビニエンスストア	29
京王プレッソイン日本橋茅場町	コンビニエンスストア	24
チェックイン四条烏丸	コンビニエンスストア	23
ヴィアイン名古屋新幹線口	コンビニエンスストア	21

図 2 ホテルがコンビニに与える影響度

このデータの影響度とは、各ホテルのレビュー中に記載されているコンビニエンスストアに関する記述のことを指す。細かな表記ゆれ等は主観で修正を行った。上記の表では5つのデータのみが記載されているが、本来のデータでは50個のデータセットで一つのカテゴリ間関係を表すシートとなっている。次カテゴリ間の相関関係の分析について述べる。

4. カテゴリ間の相関

実験にはレビューから抽出した影響度と、GoogleMapで測定した地物どうしの距離の二つを用いた。地物Aのレビュー中に見られる、他カテゴリの地物に関する記述の頻出値を影響度とした。地物Aに対する地物Bの影響度が1以上となった場合、GoogleMapを用いて地物Aと地物Bの距離を測定する。この際、対象とする地物Bは地物Aに最も近傍にあるものとし、このようなデータを50個ほど作成した。次に影響度と距離を用いて、カテゴリどうしの相関を求めた。その際、距離には対数を用い数値の正規化を行った。尚、影響度が0となった場合、距離の測定は行わずデータの作成は行っていない。



図 3 イメージ

5. 実験と考察

地物どうしの距離とレビューでの影響度からカテゴリ間の相関関係について分析を行った。分析の結果、カテゴリ間の関係は「距離が近づけば影響度が増加する関係」と「距離に影響を受けない関係」、「距離が近づけば影響度が減少する関係」の3つに定義することができた。本節では三つの関係について例を用いながら説明し、その考察について述べる。

5.1 距離が近づけば影響度が増加する関係

下表を参照すると、カテゴリ「温泉浴場」に近づけば近づくほど、カテゴリ「おみやげ」の影響度が増加していることが分かる。このことからカテゴリ「温泉浴場」このような結果となった要因としては、温泉浴場にはお土産屋が隣接していることが多いからだと考えられる。他にも「ホテル・ペンション」と「コンビニエンスストア」、「レジャー・スポーツ」と「ショッピング」で同様の関係が確認できた。

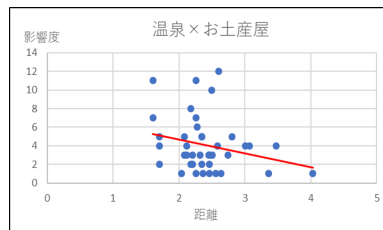


図 4 増加関係

5.2 距離によって影響を受けない関係

下表からカテゴリ「博物館」に近づいたとしても、カテゴリ「駐車場」の影響度は増加も減少もしないということが分かる。しかし、このデータではデータ数が少なかった為に相関が出なかったとも考えられるため、検証が必要である。この他には「博物館」と「コンビニエンスストア」で同様の関係がみられた。

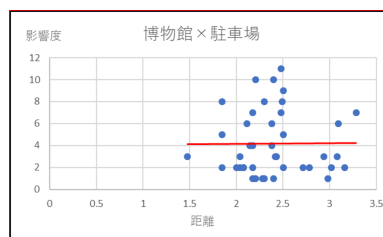


図 5 関係なし

5.3 距離が近づけば影響度が減少する関係

図5からカテゴリ「ホテル・ペンション」に距離が近くなればなるほど、カテゴリ「駐車場」の影響度は減少することが分かる。このような結果となった理由として、使用したホテルのデータがビジネスホテルに偏っていたということが考えられる。この他にはカテゴリ「動物園・植物園・水族館」と、カテゴリ「コンビニエンスストア」で同様の関係がみられた。

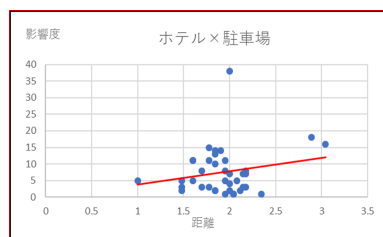


図 6 減少関係

6. まとめと今後の課題

地物Aに対する周囲の地物から地物Aの特徴を分析するというのが本研究のアプローチである。本稿ではユーザーレビューと地物どうしの距離からカテゴリ間関係の抽出を行った。その結果カテゴリ間の関係を3種類に定義することが可能となった。今後の課題として、まず地域特性の分析が挙げられる。本稿ではカテゴリ間の関係性の定義にとどまっているため、今後はデータを集め地域特性の分析を行う必要がある。二つ目は影響度として用いているレビューサイトがじゃらんであるため、観

光地などの偏ったカテゴリ間関係しか抽出することができない点である．今後は他のレビューサイトを用いて、様々なカテゴリでの組み合わせを検討する必要がある．三つめは使用しているカテゴリについてである．現在カテゴリの階層については考慮せず、カテゴリ間の関係を抽出している．今後は階層を考慮したカテゴリ間の関係の抽出が課題となる．

文 献

- [1] 成川健太郎, 北山 大輔, 角谷 和俊. 略地図分析に基づくアナロジーを用いた POI 提示システム第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2018), No. H5-2, 2018.
- [2] 吉田朋史, 北山大輔, 中島伸介, 角谷和俊. ユーザレビューの分散表現を用いた主観的特徴の意味演算による観光スポット検索システム. 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2017), No. P6-5, 2017.
- [3] 野守耕爾, 神津友武: 口コミデータに PLSA を適用した観光客目線による観光地分析, 2015 年度人工知能学会全国大会論文集, 2015.
- [4] D. Kitayama and K. Sumiya. A deformation analysis method for artificial maps based on geographical accuracy and its applications. In Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality, pp. 1926. ACM, 2012.
- [5] 石垣司, 竹中毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係の自動抽出法, オペレーションズ・リサーチ, Vol. 56, No. 2, pp. 77-83, 2011.