

Place 種別の分布に基づく 地域特徴を用いた 類似地域推薦システムの提案

鈴木 芳光

Yoshimitsu SUZUKI

情報メディア学専攻修士課程 1年

1. はじめに

居住したい地域を選ぶ際には、その地域の交通利便性や生活利便性、雰囲気等、いくつかの要素を基準にしながら検討する。しかしながら、候補となる地域が膨大に存在する場合、その中から最適な地域を見つけ出すことは困難である。そこで、ユーザが求める環境に合った地域を推薦する地域推薦システムへの需要が高まる。推薦システムの基本方式として、大きく次の三つの方式がある：(a) 協調ベース推薦、(b) 内容ベース推薦、(c) 知識ベース推薦。(a) および (b) の方式には、大量の履歴データが必要となる。また、(a) では新規ユーザおよび新規アイテムのコールドスタート問題がある。(b) では、新規アイテムのコールドスタート問題については解消されるものの、新規ユーザのコールドスタート問題は依然として残る。これらに対し、(c) は履歴データを必要とせず、新規ユーザに対しても推薦可能である。

本研究では、履歴データをもたない新規ユーザにも推薦可能なシステムを目指し、知識ベースの地域推薦システムを提案する。知識ベース推薦には、制約ベース推薦方式と事例ベース推薦方式がある。制約ベース推薦は、ユーザが提示した制約を満たすアイテムを推薦する。事例ベース推薦は、ユーザが例示したアイテムや属性に類似するアイテムを推薦する。事例ベース推薦方式により推薦システムを実装する際には、アイテム間の類似性尺度の定義が必要となる。本研究では、地域内に含まれる施設や店舗等の位置情報に関連するデータ (Place データとよぶ) の分布を基に地域を特徴化し、その特徴を基に

地域間の類似性尺度を定義する。ユーザが興味のある地域を例示することで、その興味地域と類似する地域を推薦する。

2. 定義

本章では関連する定義やデータセット等について説明する。

定義1: 地域メッシュ. 地域メッシュとは、地表面を一定のルールに従い、多数の正方形などに分割したものをメッシュという。このメッシュを標準化したものを標準地域メッシュといい、代表的なものとして第1次メッシュ、第2次メッシュ、第3次メッシュがある。経度・緯度を用いて特定の位置を表現できるように標準メッシュコードを用いれば特定のエリアを表現することができる。

定義2: Place データ. Place $p_i \in P$ は、施設や店舗など位置情報に関連するデータである。Place p_i は、その Place の種別 $p_i.type$ 、地理座標系の位置座標として経度 $p_i.x$ および緯度 $p_i.y$ をもつ。本研究では、Google Places API により、Place データを収集する。Google Places に登録されている Place 種別として、cafe, food, museum など、135種類の種別が登録されている。

定義3: 地域クラスタ. 地域クラスタ $c_j \in C$ と表す。 c_j は特徴ベクトルが類似するメッシュの集合で表される。

定義4: 地域クラスタ間類似度行列. 地域クラスタ間類似度行列を $S = [s_{ij}]$ と表す。類似度行列 S は $|C| \times |C|$ の対称行列である。 s_{ij} は地域クラスタ C_i と C_j の間の類似度を表す。

3. システム概要

本章では、提案システムである類似地域推薦システム構成について説明する。

(1) 日本全国を第2次地域メッシュに分割し、各メッシュに含まれる Place 種別の分布に基づき地域特徴ベクトルを算出する。算出された地域特徴ベクトルを基に地域のクラスタリングを行う。

- (2) 地域クラスタ間類似度を算出する。算出された地域クラスタ間類似度を基にクラスタリングを行う。
- (3) 興味地域の緯度・経度を抽出する。
- (4) 地域クラスタテーブルから興味地域が属す地域クラスタを取得する。
- (5) 地域クラスタ間類似度テーブルから類似地域クラスタを取得する。
- (6) 取得したクラスタを基に推薦地域を提示する。

4. 評価

提案手法による類似地域推薦結果を確認する。

本稿では、対象を関西圏に限定し、関西の領域を第2次地域メッシュに分割した。また、Google Places APIにより、Google Placesから関西圏のPlaceデータを収集した。収集したPlaceデータ数は588,926件であった。

3章で述べた手法により、各地域メッシュを特徴ベクトル化し、地域クラスタリングを行った。また、得られた地域クラスタ間で類似度行列を作成した。ここでは、地域クラスタ間で最も類似度が高かったクラスタ群のうち、期待通りの結果が得られた事例と、そうでなかった事例をとりあげる。

図1は、期待通りの結果が得られた事例である。図中の地域(i)~(iii)は、類似度が高かった地域クラスタを表す。地域(i)~(iii)はいずれも海岸付近の地域であり、住宅が多く分布している。

一方で、図2は、期待通りの結果が得られなかった事例である。図中の地域(iv)~(vii)は、類似度が高かった地域クラスタを表す。地域(v)~(vii)が山林地域であるのに対し、地域(iv)は海岸付近というように、地域の特徴が大きく異なっているといえる。本手法では、Google Placesに登録されているPlace種別を基に地域の特徴化を行ったが、今回用いたPlace種別だけでは十分に地域の特徴を抽

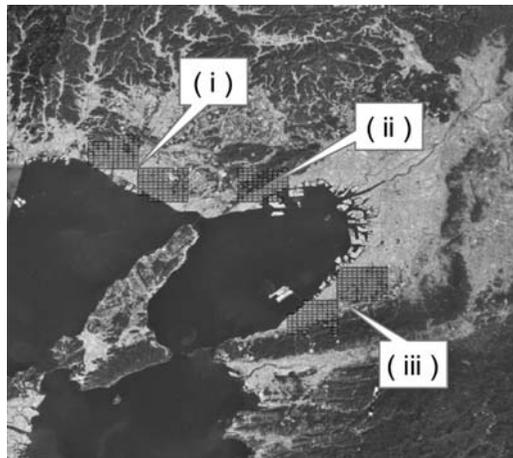


図1 期待通りの結果が得られた事例。図中の地域(i)~(iii)は、類似度が高かった地域クラスタを表す。写真は国土地理院撮影の電子国土基本図(オルソ画像)をキャプチャしたものを掲載。

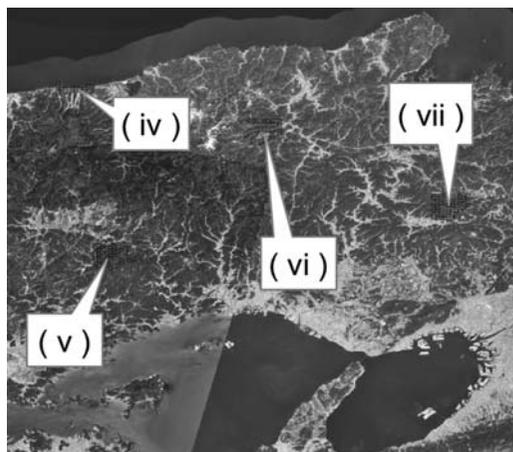


図2 期待通りの結果が得られなかった事例。図中の地域(iv)~(vii)は、類似度が高かった地域クラスタを表す。写真は国土地理院撮影の電子国土基本図(オルソ画像)をキャプチャしたものを掲載。

出できていないと考えられる。今後は、特徴化するための情報源として、Place種別以外の要素も検討する。