

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[†] 奥 健太[†]

[†] 龍谷大学理工学部 〒520-2194 滋賀県大津市瀬田大江町横谷 1-5

E-mail: [†]t18m060@mail.ryukoku.ac.jp, ^{††}okukenta@rins.ryukoku.ac.jp

あらまし 好きな地域と類似する地域を観光したい、地元と類似した雰囲気のある地域に住みたいなどの要求に対応するため、興味地域と特徴が類似する地域を推薦するシステムを提案する。このような地域推薦を実現するためには、地域の特徴化が課題となる。本研究では、Google Places から収集した Place データを基に地域の特徴化を行う手法を提案する。日本全国を第2次地域メッシュ（約10km×約10km）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周辺の類似地域をクラスタリングしておき、興味地域クラスタが入力されると、その地域と類似する地域を推薦する。

キーワード 地域推薦, 事例ベース推薦, 位置情報データ

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 [1], (b) 内容ベース推薦 [2], (c) 知識ベース推薦 [3] [4]。 (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^(注1)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、136 種類の種別が登録されている。

定義 3：地域クラスタ。 地域クラスタを $C_j \in \mathcal{C}$ 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 $|\mathcal{C}| \times |\mathcal{C}|$ 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。

1次メッシュ 1度毎の経線と2/3度(40°)毎の緯線によって、全国を分割して作られたのが、第1次メッシュ(約80km×約80km)である。標準メッシュコードは、区画南端の緯度を1.5倍した2桁の数字と、西端経度から100を引いた2桁の数字を順に並べ、4桁の数字で定義される。(例：5234)

第2次メッシュ 第1次メッシュを縦横に8等分して作られたのが第2次メッシュ(約10km×約10km)である。8等分された区画には、南から経線方向に、西から緯線方向に、それぞれ0から7の数字が振られており、この経線方向の数字「0」と緯線方向の数字「2」を順に並べ、第1次メッシュの数字の後ろにつけたものが標準メッシュコードとして定義される。(例：

(注1) : <https://developers.google.com/places/web-service/intro>

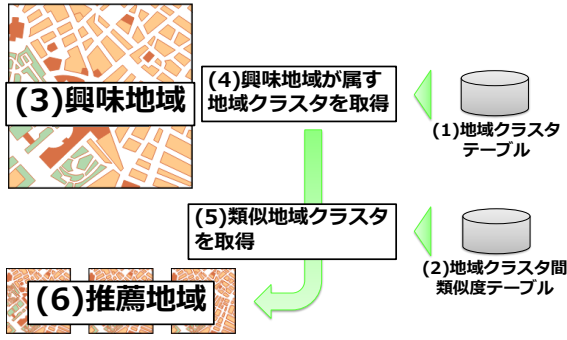


図1 類似地域推薦システムの構成図。

523402)

第3次メッシュ 第2次メッシュを縦横に10等分して作られたのが、第2次メッシュ(約1km×約1km)である。10等分された区画には、南から経線方向に、西から緯線方向に、それぞれ0から9の数値が振られており、この経線方向の数字「5」と緯線方向の数字「0」を順に並べ、第2次メッシュの数字の後ろにつけたものが標準メッシュコードとして定義される。(例：52340250)

3. システム

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。

3.1 システム概要

図1にシステムの構成図を示す。項目番号は図中の番号に対応する。

(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。

3.2 地域クラスタリング

3.2.1 メッシュの特徴ベクトル化

メッシュ r_i に含まれるPlaceデータの分布に基づき、メッシュ r_i を特徴ベクトル化する。特徴化したメッシュ r_i を特徴ベクトル $\mathbf{g}_i = (t_1, t_2, \dots, t_{135})$ で表現する。特徴ベクトル $\mathbf{g}_i$ は、135次元ベクトルで表現され、各要素 t_k はそのPlaceデータの種別が含まれる個数を表す。ただし、各要素は総和 $\sum_k t_k$ で除することで0から1の範囲になるように正規化しておく。な

お、本研究では、特徴ベクトル化の対象とするメッシュの大きさは2次メッシュとする。

3.2.2 メッシュの地域クラスタリング

まず、一つ目のクラスタ C_1 について考える。最北西のメッシュ r_1 を基準メッシュとし、次の手順で地域クラスタリングを実行する。

(1) r_1 をクラスタ C_1 に追加する。

(2) r_1 と隣接する四方のメッシュのうち、一つのメッシュに着目する。ここでは、東方向に隣接するメッシュ r_2 に着目する。

(3) r_1 と r_2 の特徴ベクトルの類似度 $\text{sim}(r_1, r_2)$ を算出する。

(4) $\text{sim}(r_1, r_2) \geq \alpha$ のとき、 r_2 をクラスタ C_1 に追加する。このとき、さらに r_2 に隣接するメッシュに着目し、(2)–(3)の処理を繰り返す。ただし、基準メッシュは r_1 で固定とする。一方で、この条件を満たさないときは、この方向の探索を打ち切る。その後、(2)に戻り、他の未探索のメッシュに着目して同様の処理を行う。

探索可能なメッシュが空になった時点でクラスタ C_1 に関するクラスタリングは終了する。つづいて、 C_2 について考える。クラスタ未割当のメッシュのうち、最北西のメッシュを基準メッシュとし、上記の処理を繰り返す。すべてのメッシュが、いずれかのクラスタに属したら、すべてのクラスタリング処理は終了となる。地域クラスタ C_j の特徴ベクトル $\mathbf{s}(C_j)$ は、そのクラスタに属するメッシュの特徴ベクトルの平均ベクトルで表す。すなわち次式で定義する：

$$\mathbf{s}(C_j) = \frac{1}{|C_j|} \sum_{i \in C_j} \mathbf{s}(e_i). \quad (1)$$

ここで、 $|C_j|$ は地域クラスタ C_j に属するメッシュの数を表す。

3.2.3 地域クラスタ間類似度行列の作成

地域クラスタ間類似度行列 $S = [s_{ij}]_{|C| \times |C|}$ を作成する。ここで、 s_{ij} は次式により算出する：

$$\cos(\mathbf{s}(C_i), \mathbf{s}(C_j)) = \frac{\mathbf{s}(C_i) \cdot \mathbf{s}(C_j)}{|\mathbf{s}(C_i)| |\mathbf{s}(C_j)|} \quad (2)$$

3.3 類似地域の推薦

興味地域と特徴が類似する地域クラスタを推薦する。地域クラスタ間類似度行列 S に基づき、興味地域と最も類似度が高い上位 n 件の地域クラスタを取得する。取得された地域クラスタを推薦地域として提示する。

4. 結果

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。

3.2節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

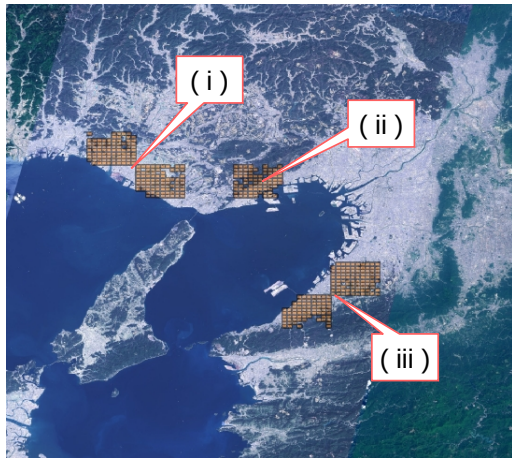


図 2 期待通りの結果が得られた事例。図中の地域 (i)–(iii) は、類似度が高かった地域クラスタを表す。写真は国土地理院撮影の電子国土基本図（オルソ画像）をキャプチャしたものを掲載。

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とを取りあげる。

図 2 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。各地域を拡大した図を図 3 に示す。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 4 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。各地域を拡大した図を図 5 に示す。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。

5. ま と め

本研究では、興味地域と特徴が類似する地域を推薦するシステムを提案した。Google Places の Place データを基にした結果より、期待通りの結果が得られた事例がある一方で、期待通りの結果が得られなかった事例もあった。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。

謝 辞

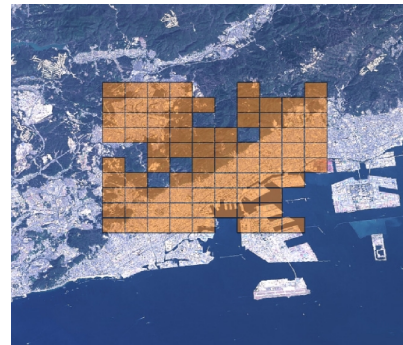
本研究は JSPS 科研費 JP15K12151, JP16H05932 の助成を受けたものです。ここに記して謝意を表します。

文 献

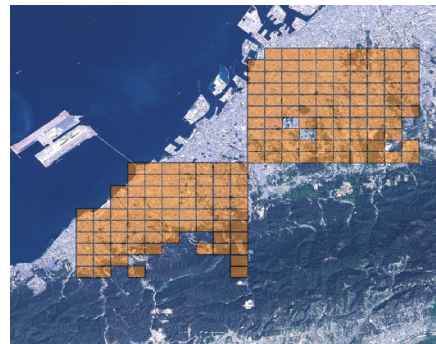
- [1] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5–53, 2004.
- [2] Iván Cantador, Alejandro Bellogín, and David Vallet. Content-based recommendation in social tagging systems. In *[RecSys2010]Proceedings of the 4th ACM conference on Recommender systems*, p. 237, 2010.
- [3] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor. *Recom-*



(a) 地域 (i).



(b) 地域 (ii).



(c) 地域 (iii).

図 3 図 2 の地域 (i)–(iii) の拡大図。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。

mender Systems Handbook. Springer, 2011.

- [4] Marius Kaminskas, Ignacio Fernández-Tobías, Francesco Ricci, and Iván Cantador. Knowledge-based music retrieval for places of interest. In *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies - MIRUM '12*, pp. 19–24, New York, New York, USA, 2012. ACM Press.

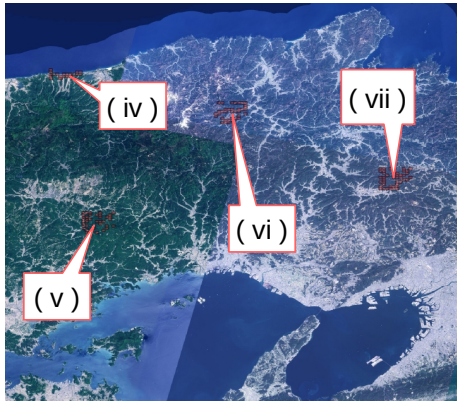
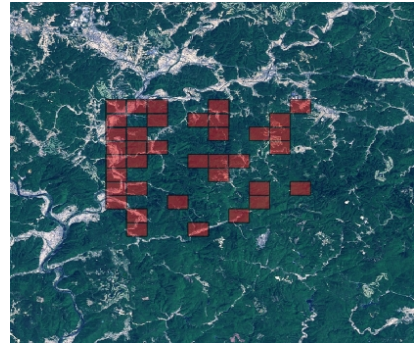


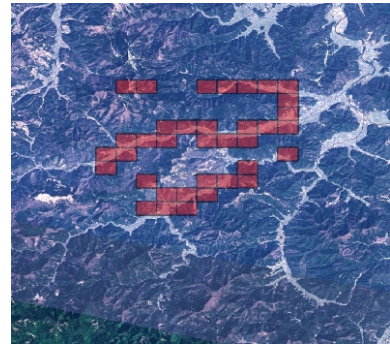
図 4 期待通りの結果が得られなかった事例。図中の地域 (iv)–(vii) は、類似度が高かった地域クラスターを表す。写真は国土地理院撮影の電子国土基本図（オルソ画像）をキャプチャしたものを掲載。



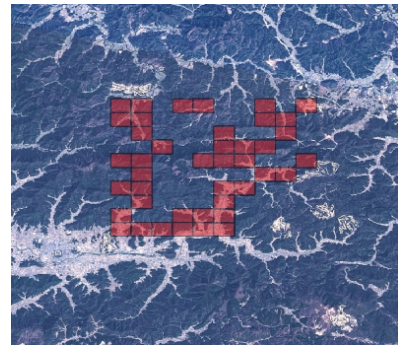
(a) 地域 (iv).



(b) 地域 (v).



(c) 地域 (vi).



(d) 地域 (vii).

図 5 図 4 の地域 (iv)–(vii) の拡大図。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。