

地理カテゴリとユーザの行動目的を考慮した地理検索システムの提案

大塚 公貴(関西学院大学), 北山 大輔(工学院大学), 角谷 和俊(関西学院大学)

研究背景

従来の地理情報検索では

既存の検索システムではオブジェクトに対する**利用目的を考慮した検索**ができない

検索の例



ユーザ

- 大阪で**食べ歩き**に利用できる観光オブジェクトはどこ?
- 東京で**宿泊に便利な**観光オブジェクトはどこ?

研究目的

地理オブジェクトの**利用目的を考慮した検索システムの構築**

同じ地理カテゴリに属している地物どうしであってもその**利用目的**は異なっている



このような利用目的を用いて地理情報検索の精度向上を図る

清水寺



利用目的

- 参拝
- 観光**
- etc

湊川神社



利用目的

- 参拝
- 食べ歩き**
- etc

提案手法

1. ユーザの移動の抽出

4travelの旅行ブログからユーザの移動を取得



旅行ブログはオブジェクト間の移動を取得できる

ユーザの移動に重要度を付与

$$hop(x, c) = \sum \left(\frac{1}{hop_{num}(y, c)} \right)$$

2. iUF値の算出

頻出するカテゴリ間の移動の重みを下げる



ユーザの移動を取得した際に駅などの公共交通機関にあたる地理オブジェクトは出現しやすい

ユーザの出現頻度を表す

$$UF(\text{User Frequency}) = \frac{\text{該当ユーザの出現回数}}{\text{全ユーザ数}}$$

UF値の**逆数**を表す

$$iUF(\text{inverse User Frequency}) = \frac{1}{UF}$$

3. 共起度の算出

共起度：対象のオブジェクトがどのくらい他のカテゴリと共に利用されるかを表す

$$hop(x, c) = \sum \left(\frac{1}{hop_{num}(y, c)} \right) * iUF$$

実験により算出された共起度

地理カテゴリ/地理オブジェクト	日泰寺	伊勢神宮	清水寺	金剛寺	浅草寺
駅	11.30	30.63	18.23	23.01	7.34
温泉	0.00	0.00	0.00	0.00	0.00
空港	7.89	0.00	0.00	8.75	10.63
グルメ・レストラン	18.21	38.05	24.36	18.24	43.13
公園・博物館	23.11	0.00	2.26	6.79	18.11
市場・商店街	10.48	0.00	0.00	0.00	51.67
自然・景観地	5.01	37.65	14.12	14.90	12.16
ショッピングモール	0.00	0.00	0.00	0.00	0.00
スーパー・コンビニ・量販店	0.00	0.00	0.00	0.00	24.00
専門店	15.32	16.00	40.00	26.67	34.67
テーマパーク	0.00	7.06	0.00	0.00	54.71
寺・神社	39.81	39.05	119.98	163.46	36.63
動物園	0.00	12.63	0.00	0.00	8.42
美術館・博物館	0.00	14.40	9.60	9.60	19.20
名所・史跡	11.15	39.10	36.63	33.70	60.58
宿・ホテル	3.49	27.41	19.75	21.23	26.42

日泰寺や清水寺は「寺・神社」との共起度が高い

4. カテゴリの組み合わせによる利用目的の推定

対象のオブジェクトが属しているカテゴリ 共起度が高くなるカテゴリ



利用目的: **食べ歩きに適したお寺**
Ex. 伊勢神宮

対象のオブジェクトが属しているカテゴリ 共起度が高くなるカテゴリ



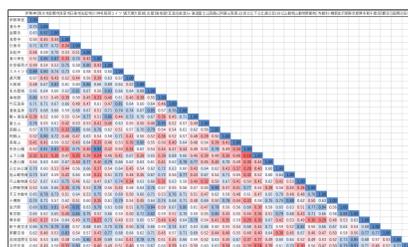
利用目的: **観光に適したお寺**
Ex. 清水寺

5. オブジェクト間の類似度の算出

コサイン類似度を用いてオブジェクトどうしの類似度の算出

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|} = \frac{\sum_{i=1}^{|\vec{q}|} q_i d_i}{\sqrt{\sum_{i=1}^{|\vec{q}|} q_i^2} \cdot \sqrt{\sum_{i=1}^{|\vec{d}|} d_i^2}}$$

算出したコサイン類似度



計: 31のオブジェクト

青: 類似度が高い

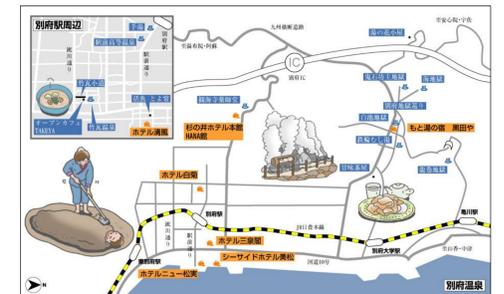
赤: 類似度が低い

6. 略地図による評価

略地図のラベルを用いた利用目的の評価

略地図に出現するオブジェクトはラベルに基づいて出現している

ラベル: 「別府温泉散策マップ」



評価実験

実験の流れ

- 略地図中のオブジェクトのiUF値を算出
- 各オブジェクト間のコサイン類似度を算出
- ラベルに基づき判定

竹瓦温泉のスコアシート

Hop(竹瓦温泉)	U(各カテゴリへの全てのユーザの移動数)	iUF(各カテゴリの全てのユーザの移動数/総ユーザ数)	iUF	Hop*iUF	総ユーザ数
駅	277.00	0.35	2.82	28.19	781
温泉	89.00	0.09	11.92	114.13	781
空港	153.00	0.20	5.00	5.36	781
グルメ・レストラン	481.00	0.62	1.62	44.25	781
公園・博物館	105.00	0.13	7.44	6.20	781
市場・商店街	37.00	0.05	21.11	38.94	781
自然・景観地	137.00	0.18	5.70	2.85	781
ショッピングモール	7.00	0.01	111.57	0.00	781
スーパー・コンビニ・量販店	6.00	0.01	130.17	0.00	781
専門店	27.00	0.03	28.93	0.00	781
テーマパーク	28.00	0.04	27.89	0.00	781
寺・神社	206.00	0.26	3.79	4.11	781
動物園	21.00	0.03	37.19	21.69	781
美術館・博物館	84.00	0.11	9.30	7.75	781
名所・史跡	447.00	0.57	1.75	36.11	781
宿・ホテル	277.00	0.35	2.82	54.51	781

両オブジェクトのHop*iUF値

両オブジェクトのHop*iUF値	竹瓦温泉(Hop*iUF(全てのオブジェクト))	竹瓦温泉(Hop*iUF(全てのオブジェクト))
駅	19.74	28.19
温泉	150.92	114.13
空港	9.36	11.46
グルメ・レストラン	28.82	44.25
公園・博物館	0.00	6.20
市場・商店街	0.00	38.94
自然・景観地	8.55	2.85
ショッピングモール	0.00	0.00
スーパー・コンビニ・量販店	0.00	0.00
専門店	0.00	0.00
テーマパーク	6.97	0.00
寺・神社	11.06	4.11
動物園	49.59	21.69
美術館・博物館	0.00	7.75
名所・史跡	82.56	36.11
宿・ホテル	25.14	54.51

コサイン類似度



両オブジェクトは散策という点で類似

考察

オブジェクトの類似度を略地図のラベルを用いて評価を行った

→ 略地図に出現するオブジェクトは観光地などが多く、偏りがある

→ データ数の不足により類似した可能性がある

今後の課題

- データ数を増やす
- 別の評価手法の検討
- 他の移動データを用いた特徴ベクトルの算出
- UIの検討