

ソーシャルメディアデータに基づく主要撮影地域の人気グルメ抽出

福田 哲也

熊野 雅仁

小野 景子

木村 昌弘

龍谷大学理工学部電子情報学科

1 はじめに

近年、撮影位置や撮影者などの情報をもつ膨大な写真データに基づいて、観光支援システムを構築する研究が注目されている [1]。一般に、人々が観光計画を立てる際、撮影スポットのような魅力的な地域の情報だけでなく、食文化や名物を含むグルメ情報も重要であることが多い。また、対象とする地域の大きさによって人気グルメ情報は変化することが考えられる。

本研究では、撮影位置と撮影者の情報をもつ膨大な写真データと、位置と料理ジャンルの情報をもつ膨大なレストランデータから、階層的な主要撮影地域とそこにおける人気グルメのジャンルを抽出し、可視化する手法を提案する。

2 提案法

2.1 階層的な主要撮影地域の可視化

Crandall ら [1] の手法にしたがって、位置情報を用いて写真データをクラスタリングすることにより、主要撮影地域を抽出する。写真 n の地理的位置を x_n ($n = 1, \dots, N$) とする。 $\{x_n\}$ に対して可変バンド幅 mean-shift 法 [2] を適用し、そのクラスタリング結果を $\{C_k^1\}$ とする。各 C_k^1 を第 1 階層の主要撮影スポットと呼ぶ。次に、各 C_k^1 に対して可変バンド幅 mean-shift 法を適用し、第 2 階層の主要撮影スポットを抽出する。この操作を繰り返して、階層的に主要撮影スポットを抽出する。

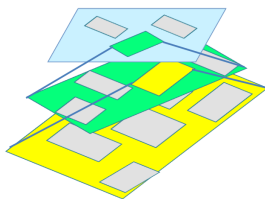


図 1: 主要撮影地域の階層的可視化

第 d 階層の主要撮影地域群を $\{C_k^d\}$ とする。各 C_k^d に対して、そこに含まれる写真の位置データ $\{x_n\}$ を主成分分析して、第 1 主成分方向と第 2 主成分方向を計算し、それらに平行な辺をもつ長方形のうち C_k^d を含む最小のもの R_k^d を求める。 R_k^d を、第 d 階層の主要撮影地域

と呼ぶ。階層ごとに各 R_k^d を地図上に配置することにより、主要撮影地域を階層的に可視化する (図 1 参照)。

2.2 レストラン数分布の可視化

各主要撮影地域 R_k^d に対して、その中に含まれるレストランの店舗数を高さとする直方体を地図上に配置することにより、主要撮影地域ごとのレストラン数の分布を可視化する。このような直方体をベースと呼ぶ。図 2 は、地図上に複数のベースを配置した例である。この分布から、レストランが密集する主要撮影地域の場所と、地域間のレストラン数に関する相対的な違いを把握することができる。

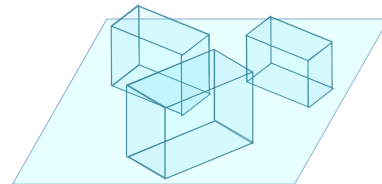


図 2: レストラン数分布の可視化

2.3 グルメジャンルの人気度の可視化

主要撮影地域における食文化や名物の検出やそれらの比較を容易にすることを旨として、各主要撮影地域 R_k^d の人気グルメのジャンルを可視化することを考える。料理ジャンル全体の集合を G とし、各料理ジャンル $g \in G$ に対して、 R_k^d における g をもつレストランの総数を $y^g(R_k^d)$ とし、全国の料理ジャンル g をもつレストランの総数を Y^g とする。また、 R_k^d におけるレストランの総数を $y(R_k^d)$ とし、全国のレストラン総数を Y とする。このとき、主要撮影地域 R_k^d の料理ジャンル g についての人気度 $P^g(R_k^d)$ を、

$$P^g(R_k^d) = \frac{y^g(R_k^d)}{y(R_k^d)} \bigg/ \frac{Y^g}{Y}$$

で定義する。 $P^g(R_k^d)$ が高いほど、 R_k^d において、料理ジャンル g の人気が高いことを示す。ここで、各 R_k^d ごとに、どの料理ジャンル g がどの程度の高い人気を得るかについて可視化を行うため、料理ジャンル g ごとの $P^g(R_k^d)$ を高さとする円柱を R_k^d に対応するベース上に配置することにより、主要撮影地域ごとの料理ジャンル人気度の分布を図 3 のように可視化する。本稿では、この円柱をファイバーと呼ぶ。ただし、ファイバーは、 R_k^d ごとに $P^g(R_k^d)$ の Top5 のみを可視化する。

Extracting popular local foods in major sightseeing areas based on social media data

Fukuda TETSUYA, Masahito KUMANO, Keiko ONO and Masahiro KIMURA
Ryukoku University

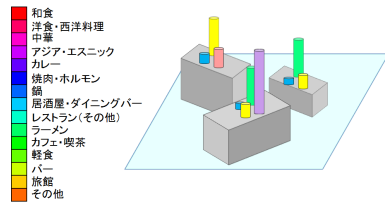
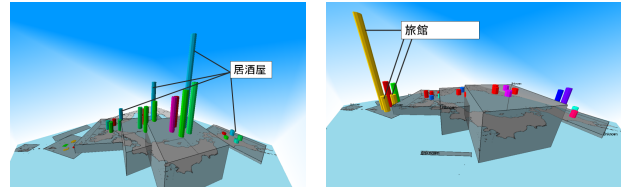


図3: グルメジャンルの人気度の可視化



(a) 比較法 (b) 提案法

図5: 人気グルメジャンル可視化

3 実験

3.1 実験データ

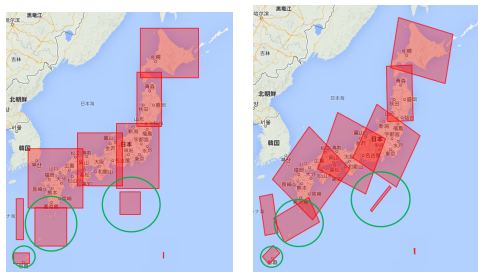
写真共有サイト Flickr から 2010 年度に撮影された日本の位置情報付き写真 162,933 枚を収集し、また、グルメクチコミサイト、食べログから、423,711 店舗の位置情報、ならびに料理ジャンル情報を収集した。食べログの料理ジャンルは、図3左側の15分類を用いた。

3.2 比較法

各地方では、一般的な人気レストランよりもご当地の名物、名産物や特産物を扱う店舗が多い傾向があると考えられるため、各ジャンル g のレストラン数 $y_g(R_k^d)$ を比較法として用いる。

3.3 実験結果

階層的に主要撮影地域を抽出したところ、最も深い最下層は $d = 16$ となった。また、最下層 $d = 16$ において、432 の主要撮影地域が得られた。



(a) 比較法 (b) 提案法

図4: 主要撮影地域範囲の抽出

日本地図上に主要撮影地域を可視化する場合、主要撮影地域に含まれる写真データの緯度・経度情報に基づいて、単純に西南の端、北東の端を対角線とする長方形として可視化した場合を図4(a)、主成分分析を用いて求めた R_k^d を可視化した場合を図4(b)に示す。図4(b)中、特に、丸で示す部分においては、図4(b)における R_k^d の形状が明確さを増していることが分かる。

次に、抽出された階層 $d = 1$ における各主要撮影地域を対象として、比較法によるベースとファイバーを可視化した結果を図5(a)、提案法によるベースとファイバーを可視化した結果を図5(b)に示す。比較法では、日本で最もレストランが多いと考えられる関東のベー

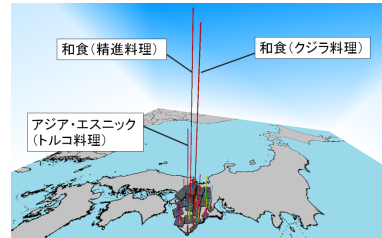


図6: 提案法による人気グルメ可視化

スが尤もらしく高い傾向を示しているが、人気のあるグルメジャンルはどの地域でも同じく一般的な人気を得る居酒屋が突出し、各地の独特のグルメジャンルが得られてないことがわかる。一方、提案法では、例えば九州地方で、旅館で食べる食事に人気があるなど、地域独特の特徴が視認できることがわかる。

また、図6は、関西地域を抜粋して階層 $d = 3$ を可視化した例である。極めて高いファイバーに着目し、食べログの料理ジャンルにおける詳細な206種類を用いて各特長を調べた。例えば、和歌山ではクジラ料理や魚介類を生かしたトルコ料理が上位を占めた。このように、地域独特の特徴を捉えていることから、地域全体を俯瞰して地域の食文化や名物が視認できる提案法の有効性が示唆される。

4 まとめ

本研究では、観光支援に向けて食事という観点から、写真データとレストランデータという2つのソーシャルメディアデータを統合することで、各地域独特の食文化や名物について、全体を俯瞰し、他の地域と比較分析できる可能性を示した。今後は、各主要撮影地域においてシーズンによって変化する食文化や名物等を抽出する問題に取り組む予定である。

参考文献

[1] David J. Crandall, Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. Mapping the world's photos. In *Proceedings of the 18th International Conference on World Wide Web*, pp. 761–770, 2009.

[2] Dorin Comaniciu and Peter Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603–619, 2002.