

ユーザー行動データによるサンプリング計算解の評価 Evaluation of Sampling solution based on user behavior data

岩崎 一輝[†]
Kazuki Iwazaki

鈴木優伽[‡]
Yuka Suzuki

斉藤和巳[‡]
Kazumi Saito

1. はじめに

本研究では、実道路網から構築したネットワーク上での施設配置問題を扱う。人々が目的施設に向かって移動する際に、目的施設以外に別の施設を寄り道することが多くあり、適切な寄り道施設の提示は人々の利便性等を高めると考えられる。そのためここでは、寄り道施設配置場所の最適化問題を扱う。我々はこれまでに、寄り道する際の容易さを示す指標である回遊中心性を提案し、適切な寄り道施設配置候補地集合の提示 [1] や、回遊中心性にユーザの回遊行動を組み込んだ行動回遊中心性 [3]、サンプリングを用いた大規模データでの多大な計算量を高速化する方法 [4] を提示してきた。しかし、従来の研究では、行動回遊中心性とサンプリング法をそれぞれ独立したものと扱ってきたが、実行に近い結果を得るためには双方を組み合わせて使う必要があると考える。

そのため本研究では、サンプリングによって求められた配置候補地の解精度をユーザの回遊行動データの観点から評価する。

2. 回遊中心性

回遊中心性 [1] は、ある地点から別の地点へ移動する際に、途中で立ち寄り容易さを示す指標として定義される。以下に、与えられた道路網での任意の地点間の移動において各地点の回遊中心性、及びに集合回遊中心性について述べる。与えられた地点の集合を $V = \{s, t, v, \dots\}$ 、地点 s から地点 t までの距離を $d(s, t)$ とし、地点 s から v を経由して t へ移動する際の最短寄り道距離を $D(s, t)$ とする。

$$D(s, t; v) = d(s, v) + d(v, t)$$

このとき、地点ペア集合 $C = \{(s, t); s, t \in V, s \neq t\}$ に対し、地点 v の回遊中心性 mBWC は次式で定義される。

$$\text{mBWC}(v; C) = \sum_{(s, t) \in C} \frac{d(s, t)}{D(s, t; v)}$$

ここで、地点集合 V の部分集合を $\mathcal{R} \subset V$ とし、地点 s から t への移動の際、 \mathcal{R} のうち最寄りの地点 r を経由した最短寄り道距離を $D(s, t; \mathcal{R})$ とする。

$$D(s, t; \mathcal{R}) = \min_{r \in \mathcal{R}} \{d(s, r) + d(r, t)\}$$

このとき、地点部分集合 \mathcal{R} の集合回遊中心性は以下で定義される。

$$\text{smBWC}(\mathcal{R}; C) = \sum_{(s, t) \in C} \frac{d(s, t)}{D(s, t; \mathcal{R})}$$

$\text{smBWC}(\mathcal{R}; C)$ において、 $\mathcal{R} = \{v\}$ とすれば、 $\text{mBWC}(v; C)$ に帰着されるので、集合回遊中心性は回遊中心性の自然な拡張である。

3. 行動データつき回遊中心性

回遊中心性に回遊者の行動データを組み込むことで、より回遊者の実行動にそった指標へと拡張する。回遊者の集合を M 、回遊者 $m \in M$ がステップ h で訪れたスポットを $r(m, h)$ とし、回遊者 m が訪れた総スポット数を H_m とすると、回遊者 m が滞在したスポット系列は $r(m) = \{r(m, 1), \dots, r(m, h), \dots, r(m, H_m)\}$ と書き表せる。このとき、行動データを組み込み拡張した回遊中心性を以下で定義する。

$$\text{mbBWC}(v) = \sum_{m=1}^M \max_{1 \leq h < H_m} \left\{ \frac{d(r(m, h), r(m, h+1))}{D(r(m, h), r(m, h+1); v)} \right\}$$

すなわち、回遊中心性では任意の 2 スポットを移動する際の寄り道の容易さを算出していたが、行動データつき回遊中心性では、実際に回遊者が移動した 2 スポット間での寄り道の容易さを算出している。そのため、スポット同士が隣接して存在し寄り道が容易であるが、実際には誰も訪れないようなスポット間に存在する寄り道スポットの抽出が抑えられる。また、行動データつき回遊中心性は回遊中心性と同様、集合回遊中心性へと拡張が可能であるため、行動データつき集合回遊中心性を以下の式で定義する。

$$\text{mbBWC}(\mathcal{R}) = \sum_{m=1}^M \max_{1 \leq h < H_m} \left\{ \frac{d(r(m, h), r(m, h+1))}{D(r(m, h), r(m, h+1); \mathcal{R})} \right\}$$

4. サンプリング回遊中心性

与えられた道路網で寄り道が容易な $|\mathcal{R}| = K$ 箇所の地点選定は、集合回遊中心性の目的関数 $\text{smBWC}(\mathcal{R}; C)$ の最大化問題として定式化できる。本研究では、最大化法として貪欲法に焦点を絞る。貪欲法では、 $\mathcal{R}_0 \leftarrow \emptyset$ 、 $k \leftarrow 1$ で初期化し、 $k = K$ となるまで、既に選定した \mathcal{R}_{k-1} を固定し、

$$\hat{r}_k \leftarrow \arg \max_{r \in S \setminus \mathcal{R}_{k-1}} \text{smBWC}(\mathcal{R}_{k-1} \cup \{r\}; C) \quad (1)$$

により地点を求め、 $\mathcal{R}_k \leftarrow \mathcal{R}_{k-1} \cup \{\hat{r}_k\}$ でその地点を解集合に追加し、 $k \leftarrow k+1$ とすることを繰り返す。ただし、貪欲法の計算量は $O(|V|^3)$ となり、大規模データへの適用は一般に困難となる。そこで本研究では、サンプリングによる近似解法を用いる。地点ペア集合 C の部分集合 $B \subset C$ をランダムに選択し、上記貪欲法に

[†]静岡県立大学 経営情報学部

[‡]静岡県立大学 経営情報イノベーション研究科

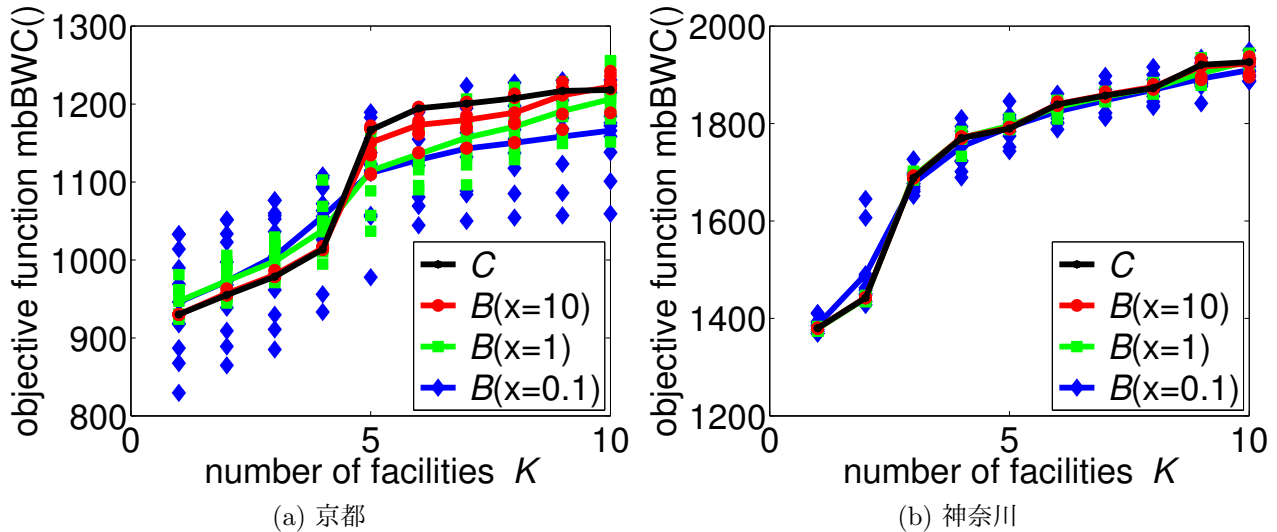


図1: サンプル抽出での評価結果

において

$$\hat{r}_k \leftarrow \arg \max_{r \in \mathcal{V} \setminus \mathcal{R}_{k-1}} \text{smBWC}(\mathcal{R}_{k-1} \cup \{r\}; \mathcal{B}) \quad (2)$$

により候補地点を求める方法である。明らかに、サンプリングによる集合回遊中心性の解精度や安定性は、道路網全体の地点数 $|\mathcal{V}|$ や、サンプリング集合要素数 $|\mathcal{B}|$ に依存する。以下では、 $x \in \{0.1, 1, 10\}$ をパラメータとし、サンプリング数を全地点数の線形オーダーとなるように $|\mathcal{B}| = x|\mathcal{V}|$ で制御する。全交差点で求めた解集合 $\mathcal{R}_k(C)$ に対し、サンプリングで求めた解集合 $\mathcal{R}_k(B)$ を目的関数

$$\text{mbBWC}(\mathcal{R}) = \sum_{m=1}^M \max_{1 \leq h < H_m} \left\{ \frac{d(r(m,h), r(m,h+1))}{D(r(m,h), r(m,h+1); \mathcal{R})} \right\}$$

に適用し、サンプリング計算解の精度を評価する。

5. 実験による評価

本実験では、写真共有サイト Flickr * から、ユーザの移動地点のデータをクロールして所得し、地点集合とした。詳細には、Flickr に投稿された写真に付随する緯度経度情報から、ユーザの移動地点の集合 \mathcal{M} を所得し、それらを Mean-Shift でクラスタリングした結果で地点集合 \mathcal{V} を構築した。また、 \mathcal{M} から、ユーザの任意の移動地点ペア (s, t) が分かるため、移動地点ペア全体の集合を \mathcal{C} とした。

選定する地点数を $K \in \{1, \dots, 10\}$ と変化させ、各パラメータ $x \in \{0.1, 1, 10\}$ での独立な 10 試行で $\mathcal{R}_k(B)$ を算出した際の、目的関数 $\text{smBWC}(\mathcal{R}; C)$ の値を図 1 にマーカーで示し、それらの平均をラインで繋いで示した。

京都府の評価結果 (図 1(a)) では、パラメータ $x = 0.1$, $x = 1$ では全体的にばらつきがみられ、解精度が不安定であったが、 $x = 10$ においては、4 つ目まではすべての組み合わせでの結果と同様の結果が得られ、それ以降で多少のばらつきは見られるものの、最終的にはすべての組み合わせの結果と同様の結果を得ることができた。

これに対し、神奈川県の評価結果 (図 1(b)) では、パラメータ $x = 0.1$ においては多少のばらつきが見られたものの、 $x = 1$, $x = 10$ においてはすべての組み合わせの結果と同様の結果を得ることができた。これらの結果から、回遊行動データをサンプリングして得ることのできる解の解精度は、地域や、回遊行動データの違いによって多少影響されるが、パラメータ設定を $x = 10$ 程度にすれば安定した結果が得られると考える。

6. おわりに

本研究では、京都府と神奈川県を対象に、サンプリングを用いて抽出された施設配置候補地を回遊行動中心性の観点から評価した。配置候補地としてそれぞれ 10 か所の候補地を抽出する評価実験結果では、パラメータの設定だけでなく、地域の違いや、そこを訪れる回遊者の行動の違いによって解精度や安定性に影響が確認された。今後の課題として、より多くの地域での実験や、多様な回遊者の行動データを収集したのち、再度パラメータを設定したうえで実験を行い、サンプリング法の有効性を評価していく。

謝辞 本研究は、科学研究費補助金基盤研究 (C)(No. 26330345) の補助を受けた

参考文献

- [1] 伏見 卓恭, 齊藤 和巳, 武藤 伸明, 池田 哲夫, 風間 一洋, “実距離を考慮した重要観光スポットの抽出”, 第 102 回 知識ベースシステム研究会 (SIG-KBS), 2014.
- [2] 鈴木優伽, 伏見卓恭, 齊藤和巳, 風間一洋, “最尤推定によるユーザー行動分析”, 第 12 回観光情報学会全国大会 (STI2015), 2015.
- [3] 岩崎一輝, 鈴木優伽, 齊藤和巳, 風間一洋, “ユーザー行動データによる回遊中心性の拡張”, 第 14 回情報科学技術フォーラム (FIT2015), 2015.
- [4] 岩崎一輝, 鈴木優伽, 齊藤和巳, “サンプリングによる回遊中心性の解精度評価”, 情報処理学会第 78 回全国大会 (IPSI2015), 2015.
- [5] D.Comanicu, P.Meer, “Mean Shift: A Robust Approach Toward Feature Space Analysis”, IEEE Transactions on PAMI 24:603-619, 2002.

*<https://www.flickr.com/>