

道路ネットワークにおける複数評価視点を考慮した 空間情報検索アルゴリズムの検討

中山 泰宏[†] 佐藤 秀樹[†]

[†]大同大学大学院 情報学研究科

1. はじめに

位置属性を持つ空間データを複数視点から評価可能とする情報検索問題を取り上げる。例えば、不動産物件の検索では、「長男が通う中学校と次男が通う小学校から近い物件」、「主婦が日常利用するスーパーマーケット、銀行、クリーニング店から近い物件」等を同時に考慮することが求められる。問題の具現化のため、各視点の評価尺度として複数問合せ点に対する集約距離を、複数視点の評価をまとめるためにスカイライン演算[1]を考える。本稿では、問題の求解法として、CE(Collaborative Expansion)アルゴリズム[2]を示し、次に処理時間短縮化のための改良策を検討し、実験により改良 CE アルゴリズムを評価する。

2. 空間情報検索問題

空間データの集合を D 、評価関数の集合を F とすると、式①が定義する ϕ は F に関する D のスカイライン集合を求める。関係 $>_F$ は、式②が定義する F に関する優越関係(dominance relation)である。一般性を失うことなく、式②では小さい値が大きい値より優越していることとする。

$$\phi(D, F) = \{p \in D \mid \nexists p' \in D, p' >_F p\} \dots \textcircled{1}$$

$$p >_F p' \Leftrightarrow (\exists f_i \in F, f_i(p, Q_i) < f_i(p', Q_i)) \wedge (\forall f_j \in F, f_j(p, Q_j) \leq f_j(p', Q_j)) \dots \textcircled{2}$$

関数 f_i は、データ p の任意の属性に基づき評価値を返す。属性には空間属性ならびに非空間属性が含まれるが、本稿では空間属性に限定し、 f_i を複数問合せ点の集合 Q_i に関する集約距離(SUM, MAX 等) [3]とする。

3. 空間情報検索アルゴリズム

本章では、CE アルゴリズムとその改良策を述べる。

3.1 CE アルゴリズム

図1はCEアルゴリズムの擬似コードであり、filtering フェーズ(1-10行目)と refinement フェーズ(11-22行目)から成る。前者は最初のスカイライン・データ s を、後者は残りのスカイライ

ン・データを探索する。

```

CE(D, R, F)
  D:data set; R:R-tree on D; F:function set;
  S:skyline list; C:candidate list; fi ∈ F;
1.  i:=0;
2.  while(true){
3.    <p, daggN(p, Qi)>:=ANN(D, R, fi(p, Qi));
4.    if (p is visited by all fi){
5.      enqueue(S,p); dequeue(C,p); break;}
6.    else if (p is returned for the first time)
7.      enqueue(C,p);
8.    i++;
9.    if (i==|F|) i:=0;
10. }
11. while(notempty(C)){
12.   <p, daggN(p, Qi)>:=ANN(D, R, fi(p, Qi));
13.   if (p is visited by all fi){
14.     if (p ∈ C and (p isn't dominated by s(∈S))){
15.       enqueue(S,p); dequeue(C,p);
16.       for (p' ∈ C)
17.         if (p' is dominated by p) dequeue(C,p);
18.     }
19.   }
20.   i++;
21.   if (i==|F|) i:=0;
22. }
23. return S;
    
```

図1. CEアルゴリズム

図1のANN (Aggregate Nearest Neighbor) (3, 12行目)は、漸進的に集約最近傍を探索する。本稿では、ANNとして、R-木に基づくIER (Incremental Euclidean Restriction)アルゴリズム[3]を用いる。図2は、IERアルゴリズムの擬似コードである。Hは、R-木のノードまたはデータ・オブジェクトと問合せ点集合Qとの間の集約距離 $d_{agg}(e, Q)$ に従う優先度キューである。また、 $d_{agg}^E(e, Q)$ は集約ユークリッド距離、 $d_{agg}^N(e, Q)$ は集約ネットワーク距離であり、この間では[補題1]が成立する。このため、図2の11行目では集約ネットワーク最近傍となるデータ・オブジェクト e が返される。

[補題1] $d_{agg}^E(e, Q) \leq d_{agg}^N(e, Q)$ □

図2のSPQ (Shortest Path Query) (8行)は、2点間の最短ネットワーク経路を探索する。本稿では、SPQとして、A*アルゴリズムを用いる。

3.2 CEアルゴリズムの改良策

CEアルゴリズムの refinement フェーズで候補リスト C にないデータ・オブジェクトは、

filtering フェーズで見つけられる唯一のネットワーク・スカイライン・データに優越され、スカイライン・データには成り得ない。また、refinement フェーズで見つけられるネットワーク・スカイライン・データ s に優越されるデータ・オブジェクト p' も候補リスト C から削除される(図1の17行目)。IERアルゴリズムは最短ネットワーク経路の計算(図2の8行目)後、集約ネットワーク最近傍 p を返す。 p が最終的にネットワーク・スカイライン・データに成り得るか否かは、図1の14行目で判定される。上記の最短ネットワーク経路は高い計算コストを要するため、CEアルゴリズム(図1の14行目)が行う条件判定 $p \in C$ を IERアルゴリズム(図2の8行目の直前)で行うことにより、不必要な最短ネットワーク経路計算の削減を図ることができる。

```

IER(D, R, dagg(e, Q))
  D:data set; R:R-tree on D; dagg(e, Q):agg-function;
  H:priority queue in the order of agg-distance
  1. for(e ∈ root(R)) enqueue(H, <e, daggE(e, Q))
  2. while(notempty(H)){
  3.   <e, dagg(e, Q)>:=dequeue(H);
  4.   if (e is a R-tree node)
  5.     for(e' ∈ e) enqueue(H, <e', daggE(e', Q))
  6.   else if (e is a data object in D){
  7.     if (dagg(e, Q) is daggE(e, Q)){
  8.       daggN(e, Q):=agg(SPQ(e, qi(∈ Q)));
  9.       enqueue(H, <e, daggN(e, Q)>);}
  10.    else{//dagg(e, Q) is daggN(e, Q)
  11.      return <e, daggN(e, Q)>;}
  12.  }
  13. }
    
```

図2. IERアルゴリズム

4. 評価実験

本章では、改良 CE アルゴリズムの処理時間の評価実験に関して述べる。プログラムは Java 言語により実装し、実験は Intel Core i5-6600 PC(クロック周波数 3.3GHz,メモリ 8GB)を使って行う。道路ネットワークデータは愛知県データ[4]を、空間データは愛知県の賃貸物件情報(約32万件)[5]を用いる。ある地点を中心とする500m四方の範囲内で問合せ点を無作為に生成した同一問題に関して、CEアルゴリズムと改良CEアルゴリズムの処理時間を比較する。

図3は、集約距離SUMとMAXが各2個、計4個の評価関数、 $|Q_i|=3$ の下での、犬山駅、大同町駅、名古屋駅の各地点における実験結果である。改良CEアルゴリズムの処理時間は、CEアルゴリズムの30%~42%に削減された。各地点を中心とする1km²あたりの空間データ数は各々34, 150, 1230であり、これが処理時間に影響している。

図4は、評価関数の数が2, 4, 6、集約距離SUMとMAXが各々1, 2, 3、 $|Q_i|=3$ とする大同町駅の実験結果である。改良CEアルゴリズムの処理時間は、CEアルゴリズムの33%~44%に削減された。

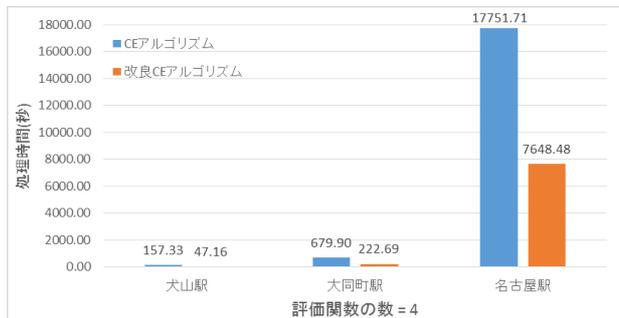


図3. 異なる地点における処理時間

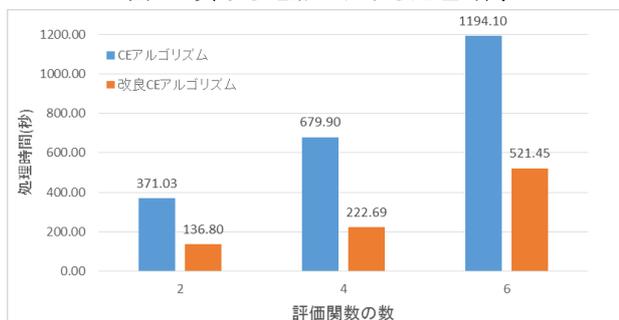


図4. 評価関数の数に対する処理時間(大同町駅)

5. おわりに

本稿では、空間データを複数の視点から評価可能とする情報検索問題の解を求めるため、改良CEアルゴリズムを提案し、処理時間がCEアルゴリズムの40%程度に削減されることを実験により確認した。今後、さらなるアルゴリズムの処理時間の改善を図っていく。

謝辞 本研究では、株式会社ネクストが国立情報学研究所の協力により研究目的で提供している「HOME'S データセット」を利用させていただきました。

参考文献

[1] S. Borzsonyi, et al: "The Skyline Operator", ICDE, pp.421-430, 2001
 [2] K. Deng, et al: "Multi-source Skyline Query Processing in Road Networks", ICDE, pp.796-805, 2007
 [3] Yiu, et al: "Aggregate Nearest Neighbor Queries in Road Networks", IEEE TKDE, Vol.17, No.6, pp.820-833, 2005
 [4] 国土交通省国土地理院: 数値地図2500(空間データ基盤)中部-2
<http://www.gsi.go.jp/geoinfo/dmap/dm2500sdf/>
 [5] 国立情報学研究所 情報学研究データリポジトリ: HOME'S データセット
<http://www.nii.ac.jp/dsc/idr/next/homes.html>