

複数の探索木を用いた地図データの空間探索*

2R-9

横川完治 鈴鹿豊明

日立ソフトウェアエンジニアリング（株）

1 はじめに

空間探索とは、空間上の探索範囲と呼ばれるある範囲の領域を指定し、その領域と交わりのある全ての図形データを取り出す処理のことである。ここで空間とは一般に n 次元空間を指すが、以下では 2 次元空間のこととする。空間探索に用いられる図形データの探索木とは、図形データの外接長方形の中心点の集合に対する探索木で、その各ノードに対して外接長方形を対応させたものである。葉ノードの外接長方形はそのノードに対応した図形データを含む最小の長方形である。また、非葉ノードの外接長方形はそのノードの下位にある葉ノードの全ての図形データを含む最小の長方形である。通常、外接長方形の各辺は x 座標軸または y 座標軸のいずれか一方と平行であるとする。図形データの探索木の外接長方形と探索範囲の交わりの判定はその下位の葉ノードの図形データと探索範囲の交わりの可能性の判定として役立つ。すなわち、あるノード p の外接長方形と探索範囲が交わらなければ、この外接長方形はノード p の下位の葉ノード q の外接長方形を含むので、葉ノード q の図形データと探索範囲の交わりはない。このことが木探索のアルゴリズムと結び付いて空間探索を高速にする。以上の空間探索は確立された技術である [1]。

2 大きな線状の図形データの問題

探索対象に比較的小さい図形データのみが含まれる場合、従来の空間探索方式で十分である。しかし、大きな線状の図形データが存在する場合、この方式は効果的ではない。なぜならば、大きな線状の図形データと探索範囲が交わらないにもかかわらず、その外接長方形と探索範囲は交わることが多く、したがって、探索木を辿って、その図形データに対応する葉ノードまで無駄にアクセスしてしまうからである。このように、大きな線状の図形データの存在は空間探索の妨げになるという問題がある [2]。実際の例として、住宅地図の空間探索が挙げられる。住宅は比較的小さい図形データであり、住宅地図の図形データのほとんどを占めている。しかし、中には通常少数であるが道路や境界線のように大きな線状の図形データがあり、従来の空間探索方式ではうまくいかない。

3 複数の探索木

前節の問題を解決するために、図形データの大きさにより図形データを 2 個の探索木に振り分け、それぞれの探索木について従来の空間探索を行うという方法を提案する。図形データの大きさについて具体的に言うと、図形データを囲む外接長方形の幅と高さが共にあるシキイ値より小さければ、その図形データは小さいと判定する。このシキイ値は前もって統計的処理によって決められたものである。実際の応用の場合、大きな図形データの数は小さな図形データの数よりもかなり少ないことが多い。その時、

*Spatial Search of Map Data Using Two Search Trees

Kanji Yokokawa, Toyooki Suzuka

Hitachi Software Engineering Co., Ltd.

大きな図形データを対象にする探索木は比較的小さな木構造であるため、その探索木を頻繁にアクセスしても計算コストが小さくて済む。また、小さな図形データを対象とする探索木において葉ノードを無駄なくアクセスすることができる。したがって、2個の探索木を用いた空間探索は1個の探索木の場合よりも全体としてより効率的に行われる。

4 空間探索の例

単純化した空間探索の例を挙げる。ここでは探索木は領域2等分割木、k-d木、BD木等の2分木とする。この木構造は N 個の葉ノードを持つとすると、木構造全体で $2N-1$ 個のノードがあり、根ノードから葉ノードまでのパスの長さは $\log_2 N$ 程度である。図形データに関して次の仮定を置く。 N 個の葉ノードの内、 N_1 個は小さい図形データに対応し、残り N_2 個（但し、 $N_1 + N_2 = N, N_1 \gg N_2$ ）は大きい図形データに対応するものとする。探索範囲と全ての大きい図形データの外接長方形は交わるものとする。また、探索範囲は十分小さく、それとただ1個の図形データが交わるものとする。

まず1個の探索木を使用する場合を考える。探索範囲と交わる小さい図形データに対応する葉ノードをアクセスする時、全ての大きい図形データに対応する葉ノードも辿られてしまうため、 $2(N_2 + 1)\log_2 N$ 個のノードをアクセスする。また、探索範囲と交わる大きい図形データに対応する葉ノードをアクセスする時、同様に、 $2N_2 \log_2 N$ 個のノードをアクセスする。したがって、平均のコストは

$$I = \frac{2N_1(N_2 + 1)\log_2 N + 2N_2^2 \log_2 N}{N}$$

となる。

次に、2個の探索木を用いる場合を考える。図形データの大きさで分類してできる2個の探索木の葉ノードの数はそれぞれ N_1 個と N_2 個になる。探索範囲と交わる小さな図形データに対応する葉ノードをアクセスする時、 $2\log_2 N_1$ 個のノードをアクセスする。また、探索範囲と交わる大きな図形データに対応する葉ノードをアクセスする時、木構造全体を巡ってしまうため、 $2N_2 - 1$ 個のノードをアクセスする。したがって、平均のコストは

$$J = \frac{2N_1 \log_2 N_1 + N_2(2N_2 - 1)}{N}$$

となる。

$N \approx N_1 \gg N_2$ として近似すると、

$$I - J \approx 2N_2 \log_2 N$$

となる。つまり、2個の探索木を用いることによって、平均 $2N_2 \log_2 N$ 個のノードのアクセスを省略することができる。

5 おわりに

様々な大きさの大量の図形データを含む対象に、図形データの大きさに基づいて構築された複数の探索木を使用することによって、空間探索の高速化を実現した。

参考文献

- [1]横川 鈴鹿：地図情報システムのためのBD木を用いた空間探索，情処第50回全大2N-8，1995
- [2]古本 他 3：道路地図データベースにおける空間索引機構の評価，情処第50回全大2F-3，1995