

道路地図データベースにおける空間索引機構の評価

2F-3

古本幸彦 大保信夫 木内範義 鈴木尚志

筑波大学

1 はじめに

近年、計算機の機能の向上と共に応用分野も拡大し、従来のデータベースシステムでは扱うことが困難であった、複雑、且つ大量の情報を効率的に扱うことが求められてきている。そのような応用として、地理情報システム (GIS) が挙げられる。当研究グループでは、GIS の一例として道路地図データベースシステムの研究開発を行っている。

地理情報システムにおいては“point型”、“line型”、“region型”などの、従来型のデータベースシステムが扱ってきたデータに比べて複雑な構造を持つ空間オブジェクトが存在する。そして、それらの空間オブジェクトを効率的に扱うために、空間索引機構を用いることが一般的である。

空間索引機構には様々なものが提案されているが、point型、line型、region型の全てを扱えるものは、その対象の最小包囲矩形をキーとするものがほとんどである。しかし、道路地図データベースに多く存在するline型のデータに対して、最小包囲矩形をキーとすることが効率的な検索に有効であるかは明らかでない。

本稿では、line型のデータに対して最小包囲矩形で索引を構築することが有効であるかを、検索コストの観点から検討する。

2 道路地図データベース

本研究では、以下に示す基本データ等からなる「全国デジタル道路地図データベース」[drm92]の一部を対象とし、使用している。

基本道路データ 道路網を表現する上での結節点 (交差点、起点、終点等) を示すノード、及びノード間を連結する線分 (道路等) を示すリンク等が含まれる

背景データ 施設の位置、地名等の表示位置を示す点及び、行政界、鉄道等を示す線、施設等の形状、水系等を示す面等が含まれる

これら表現する基本の空間オブジェクトとして、以下の3種類を考える。

- point型：ノード及び施設等の位置を示す
- line型：リンク及び境界線等を示す
- region型：交差点、施設等の形状を示す

Evaluation of a Spatial Index in a Road Map Database
Yukihiko FURUMOTO, Nobuo OHBO
Noriyoshi KIUCHI, Takashi SUZUKI
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305, Japan

3 空間索引機構

現在、空間索引機構としては様々なものが提案されている。しかし、PMR quadtree[Hos91]のように扱うデータの型を制限したものも多い。道路地図データベースには、point、line、regionの全ての型の空間オブジェクトが存在するため、それらを用いることはできない。

全ての型の空間オブジェクトを扱う空間索引機構としては、最小包囲矩形をキーとするものが挙げられる。ここでは、最も広く知られているR-tree[Gut84]を取り上げる。

R-treeでは、問い合わせ領域とキーとなる矩形の重なりによって検索を行なう。以下、line型のオブジェクト自身をlineと呼ぶこととすれば、図1において問い合わせ領域 Q に重なるlineを検索させた場合、1のように実際に Q に重なっているlineだけでなく、1'のように最小包囲矩形が重なっているだけでline自身は重なっていないものも検索されてしまう。1'のように、line自身

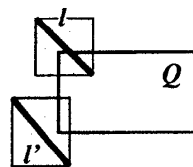


図1: 問い合わせ例

は重なっていないにも関わらず最小包囲矩形が重なっているために検索されてしまうことをFalse Dropと呼び、False Dropを起こさせる領域をエラー領域と呼ぶこととする。

検索コストを下げるためには、False Dropを少なくすることが有効であり、そのためにはエラー領域を小さくすることが有効である。そこで、lineを複数の短いlineに分割することが考えられる。図2のように、lineを分割することによってエラー領域を減少させることができる。

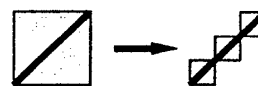


図2: lineの分割例

4 実験と評価

lineの分割によるアクセスコストの変化を調べるために、実際の道路地図データを用いてシミュレーションを行なった。

本実験においては、道路地図データからリンクを表すlineのみを取り出して用いた。その総数は3328で、平均の長さは約109であった。そして、それらをそのまま

R-tree に格納した場合と、2 等分、4 等分、8 等分して R-tree に格納した場合で、アクセスコストを比較した。

4.1 オブジェクトへのアクセスコスト

まず、オブジェクト(リンク)を分割することによって、アクセスされるオブジェクト数がどのように変化するかを調べた。その結果を図3のグラフに示す。

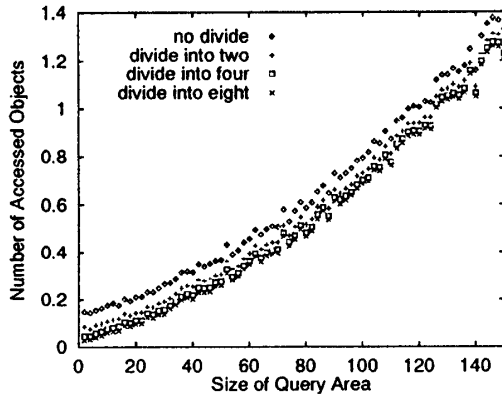


図3: アクセスされるオブジェクト数

分割を行なうことによって、アクセスされる実データ数が減少していることが分かる。また、検索領域が小さい場合にその差が顕著となっている。

4.2 空間索引機構の走査コスト

次に、R-tree の操作の段階におけるアクセスコストを調べた。バッファが十分に存在する場合は、少なくとも非葉ノードは常にバッファ内に保持しておき、ディスクアクセスを行わないことが一般的である。そこで、まず非葉ノードはバッファ中に存在しているものとみなし、葉ノードのアクセスのみを調べた。その結果を図4のグラフに示す。更に、非葉ノードへのアクセス数も考慮す

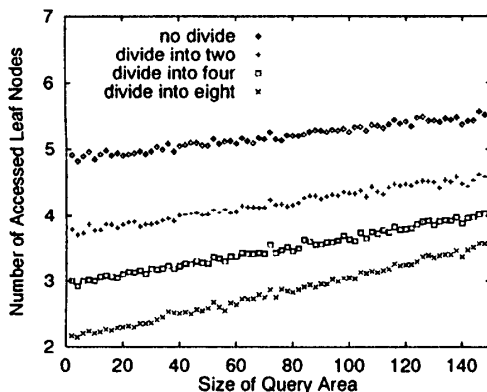


図4: アクセスされる葉ノード数

ると図5のグラフのようになった。

図4では、分割を行なうほど葉ノードへのアクセスは減少しているが、図5では必ずしもそうはなっていない。

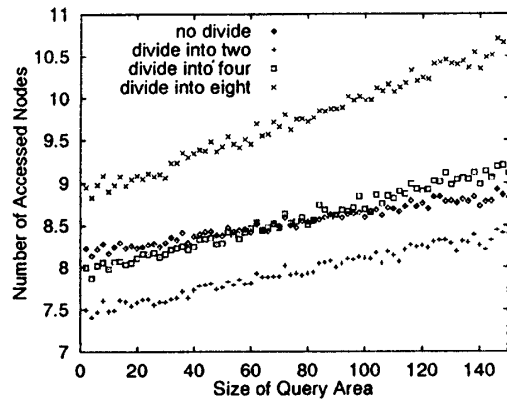


図5: アクセスされるノード数

これは、分割による無駄なノードへのアクセスの減少よりも、R-tree に格納されるエントリ数の増大によるノード数の増加が勝っているためである。

4.3 考察

以上の結果より、line を分割して格納することによって実データ、及び葉ノードへのアクセス数は減少することが分かった。しかし、バッファが十分に存在しない状況を想定すると、非葉ノードの走査にもコストがかかると考えられるため、必ずしもアクセスコストが下がるとはいえないことが分かる。

5 おわりに

本稿では、道路地図データベースに多く存在する line に対しては、line を細かく分割し、それらの最小包圍矩形の集合をキーとする索引を構築することが、検索コストの観点から有効であることを確認した。ただし、十分な大きさのバッファが確保できない場合には、以上の議論が成り立たない場合もあることが分かった。

今後は、line 型のデータだけでなく、point 型、region 型のデータも存在する状況でのアクセスコストの評価を行なってゆく。そして、R-tree 以外の空間索引機構でも同様の実験を行ない、道路地図データベースに最も適した空間索引機構を特定したい。

参考文献

- [Gut84] Guttman, A., R-Trees: A Dynamic Index Structure for Spatial Searching, Proc. ACM SIGMOD Int. Conf. on Management of Data, 1984 pp.47-57
- [Hos91] Hoel, E. G. and Samet, H., Efficient Processing of Spatial Queries in Line Segment Databases, Proc. 2nd Symp. on Large Spatial Databases, in LNCS Vol.525, 1991, pp.237-256
- [drm92] (財)日本デジタル道路地図協会, 全国デジタル道路地図データベース標準 第2.2版, Nov. 1992.