

空間索引木における埋め込み型 MBR の数値評価

蒲原 智也[†] 井筒 大資^{††} 上島 紳一[†]

[†] 関西大学大学院総合情報学研究科

^{††} 関西大学総合情報学部

E-mail: [†]{fa8d001,fa60034,ueshima}@edu.kansai-u.ac.jp

あらまし 本稿では、我々が提案している空間索引木に対する新たな指標について提案する。提案している空間索引木は、距離をベースにボトムアップに構成している。この索引木では、最短経路問題のような距離ベースの 1 対 1 の問題には効率的に対応できる。しかし範囲問合せや、k 最近傍問題のような範囲が必要となる 1 対多の問題には、解を絞り込む段階に問題がある。そこで、我々は、MBR の概念を利用して、範囲問合せなどに対応するための領域の広さについての指標を提案する。国土地理院数値地図を用いて、実際にどのようなグラフが生成できるか確認し、利用時にどのような手順を踏んで探索するかを簡単に説明する。また、この MBR が有効に利用できるかの数値評価を行う。

キーワード 空間索引木, ネットワークボロノイ図,

Embedding MBR's in Spatial Index Tree for Road Network Applications

Tomoya KAMBARA[†], Daisuke IZUTSU^{††}, and Shinichi UESHIMA[†]

[†] Graduate School of Informatics, Kansai University

^{††} Faculty of Informatics, Kansai University

Takatsuki-city OSAKA JAPAN

E-mail: [†]{fa8d001,fa60034,ueshima}@edu.kansai-u.ac.jp

Abstract This paper proposes an extension of our spatial index tree and its generation method in order to enhance functions on range queries for road network applications. Our spatial index tree are to be generated in a bottom-up fashion, using Network Voronoi Diagram, with respect to distance information along road segments. Especially, it works efficiently for in point-to-point queries, such as shortest path problems. It, however, has a disadvantage in range queries, since road path length does not necessarily provide the extent of geographical regions. In fact, our previous spatial index, which is a distance-based index in each region possesses inaccuracies in refinement steps in range queries, such as Distant Range Queries and k-Nearest Neighbor queries. In this paper, we introduce MBR's (Minimum Bounding Rectangle) to our spatial index tree as a range index, and provide their bottom-up generation algorithm from base road network data. We also show the usage of MBR's to Geographical Survey Institute digital map, and discuss these range queries algorithms using them.

Key words Spatial Index Tree, Network Voronoi Diagram,

1. はじめに

近年、地理情報システムや道路網応用の研究分野で、最短経路探索、範囲問合せ (DRQ)、k-最近傍探索問題 (kNN) などに有効なスケーラブルな空間索引の研究が注目を集めている。

我々の先行研究 [1] では、空間索引木をボトムアップに生成した。また階層化した空間索引木は、距離の情報のみを元に空間索引木を構成した。空間索引木の生成では、基準となる点を、ランダムに選択しその点を中心にグラフを分割している。各部分領域は、領域内・隣接領域間の経路、領域の大きさの指標な

ど、距離を中心とした索引を持つ。これにより、距離ベースの最短経路探索には対応していたが、距離情報だけをもちいた範囲問合せや kNN などの空間に対する問合せに対しては、非効率的な面が予測できる。つまり、DRQ において距離次第では、大きさの指標の最大と最小の差が大きくなり、効率的でない。また、kNN においては、K 番目と推定するオブジェクトの位置がとても広くなり、問題がある。これは、距離情報のみを、領域の大きさの指標として得ることに由来する。そこで領域の広がりを示す指標として Minimum Bounding Rectangle(MBR) を木構造の中に組み込んで範囲探索に対応する。

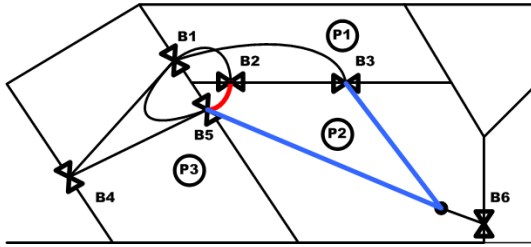


図 1 Δ と δ の定義

Fig. 1 The example which cannot use Δ and δ effectively

2 章では、我々の先行研究に存在する問題点について簡単に説明する。3 章では、先行研究の問題点に対する対策としての、空間索引木に対する MBR の詳細な生成法と利用方法を簡単に説明する。4 章では、生成した MBR についての評価を行なう。5 章で関連研究について簡単に説明する。

2. 先行研究での問題点

我々の先行研究 [1] では、境界点間経路を中心とする距離を重要視した空間索引木を構成しているため、領域の広がりが必要とする問題には、大量の処理が必要となる。本章ではそのことについて簡単に説明する。

図 1 は、ある母点領域の境界点間経路を表している。図中の $\times$ は、境界点を示す。空間索引は距離をベースとしているので、母点領域の形はどのような形をしているかはわからない。また、 $B^2 - B^5$ の境界点間経路のように極端に短い経路を δ^{P2} として持ち、 $B^3 - B^5$ のように長い経路を Δ^{P2} として P2 の母点領域の索引として持つことがある。P1 の領域にあるノードに対して、B2 からの経路が $\Delta^{P2} - \delta^{P2}$ より小さかった場合、 Δ^{P2} の経路は、P2 を横断する経路の索引として問題を持つ。

この探索手法を DRQ や kNN などの空間に対する問合せに直接的に適用すると、手数が増え効率的でない。ある母点領域内のオブジェクトに対する複数の領域を経由した Δ と δ を見積もり情報として利用した場合、とても大きく見積もってしまう、実際に探索する範囲がとても広がってしまうことが予想される。

たとえば、図 2 は、DRQ において、探索が必要となる範囲を例示したものである。図中でオブジェクトの位置は下限の値は経由してきた領域の δ の値の合計で求められ、上限の値は経由してきた領域の Δ の値で両矢印を書いている。経由領域が増えたとこの矢印の幅は広がり、最短経路探索時にとても有効であった長距離では、非常に使いにくくなる。この現象に対する対策として MBR を利用することで、探索範囲を減少させることを考える。この方法は次章で詳細に説明する。

3. MBR の生成と利用

3.1 生成法

道路網において、交差点をノード、道路片をエッジとみなした無向グラフ $G = (N, E)$, (N : ノード集合, $E \subseteq N \times N$: エッジ集合) を考える。更に、各エッジは重みを持ち、各ノード n は 2 次元での座標 $(n.x, n.y)$ を持つものとする。

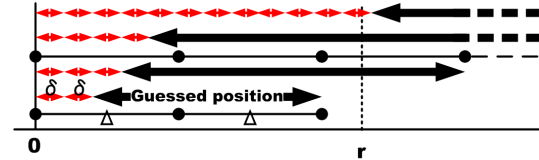


図 2 Δ と δ のもつ可能性

Fig. 2 The guessed position of objects

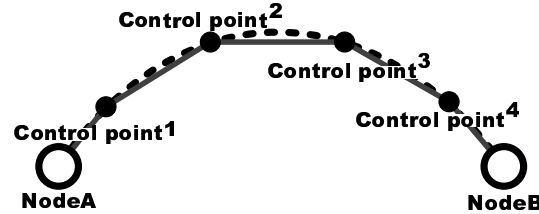


図 3 曲線の直線近似例

Fig. 3 The example of straight-line approximation

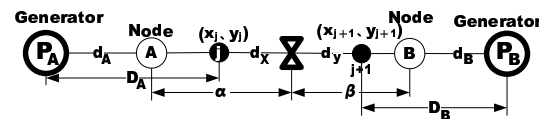


図 4 数式で利用する各用語

Fig. 4 The parameters used with expression

[厳密な境界点位置の特定]:

前章での境界点の位置はレベル 1 での 2 母点間をつなぐ経路上の中間の位置にある点である。このとき、境界点は、2 次元上での位置という物は必要なかった。なぜなら境界点は、ただ単に領域の端を示すと同時に、母点領域間をつなぐエッジであり、2 母点の位置がその役割を果たしていた。しかし、母点領域の空間の広がりを与える MBR では、2 次元上での母点領域内すべてのノードの位置・エッジの曲線の頂点・境界点の位置が必要となってくる。そこで本項では、それぞれの算出法について述べる。

[エッジの構成]

ここでの、エッジとは 2 つのノードをつなぐものであり、地図では道路片にあたる。国土地理院数値地図 2500(空間データ基盤)では、データは XML に記述されており、一本の曲線は、直線近似にて複数の直線によって構成されている。また、その直線は、本稿では節点と呼ぶノードから成り立っている。また、節点も 2 次元での座標 (x, y) をもつ。つまり各曲線は図 3 のように表現される。

[境界点位置の算出]

ここで段階で既知であるのは、ノード A からノード A の母点までの距離・ノード B からノード B の母点までの距離である。境界点の位置は下記の式から算出することができる。また図 4 に数式中で利用する各用語を簡単に示す。

ノード A からノード B へのエッジが n 本の直線近似によって構成されている場合

$$\begin{cases} d_A + \alpha = d_B + \beta \\ \alpha + \beta = \overline{AB} = \sum_{i=0}^n \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \end{cases} \quad (1)$$

ここでの x_0 はノード A の x 座標であり x_n はノード B の x 座標である。

$$(*) \begin{cases} \alpha = \frac{d_B - d_A + \overline{AB}}{2} \\ \beta = \frac{d_A - d_B + \overline{AB}}{2} \end{cases} \quad (2)$$

ノード A から j を探す。このとき式 (3)(4) を同時に満たす必要がある。

$$\sum_{i=0}^j \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \leq \alpha \quad (3)$$

かつ

$$\sum_{i=0}^{j+1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} > \alpha \quad (4)$$

(x_{j+1}, y_{j+1}) から母点 P_B までの距離を

$$\sum_{i=j+1}^n \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} + d_B \equiv D_B.$$

j 番目のエッジの端点 $(x_j, y_j)(x_{j+1}, y_{j+1})$ からの距離を (d_X, d_Y) とおけば, (*) と同様に

$$\begin{cases} d_X = \frac{D_B - D_A + \sum_{i=0}^j \sqrt{(x_{i+1} - x_i)^2 + (y_{j+1} - y_j)^2}}{2} \\ d_Y = \frac{D_A - D_B + \sum_{i=j+1}^n \sqrt{(x_{i+1} - x_i)^2 + (y_{j+1} - y_j)^2}}{2} \end{cases} \quad (5)$$

また境界点 $\bar{X}$ の位置座標 $(x_{\bar{X}}, y_{\bar{X}})$ は

$$\begin{cases} x_{\bar{X}} = \frac{d_Y * x_j + d_X * x_{j+1}}{d_X + d_Y} \\ y_{\bar{X}} = \frac{d_Y * y_j + d_X * y_{j+1}}{d_X + d_Y} \end{cases} \quad (6)$$

となる。

これらの式により, 境界点の 2 次元座標上での位置を算出することができる。

[MBR の生成]:

原地図より生成する第一層を G_{l1} と呼び, 以降の第 n 層を G_{ln} と呼ぶ。

[G_{l1} の生成時] G_{l1} 生成時には, 各母点領域で, すべてのノード座標, すべてのエッジ上の複数ある節点, すべての境界点の座標より, X について最大・最少の値, Y について最大・最少の値をもとめ, 母点ノードに関連付けを行い保存する。

[G_{l2} 以上での生成時] 上位層 G_{li} を生成するときに, G_{li-1} での母点ノードの中で XY の値の最大・最少の値を自身に所属するノードより集め, その中での最大・最少の値をとることで, G_{li} での自身の MBR を生成することができる。

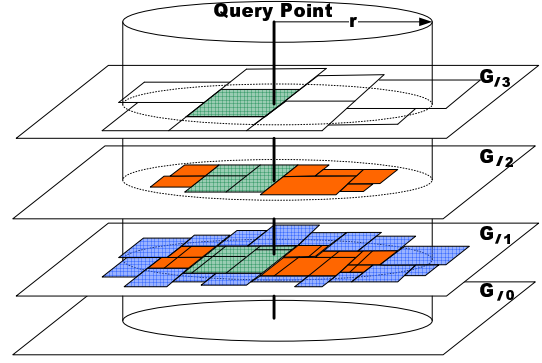


図 5 DRQ での MBR 利用の概念
Fig.5 Concept of the MBR use in DRQ

3.2 MBR の利用方法

本節では, ここまでで説明した空間索引木とそれに付加する MBR を用いてどのように探索に役立てるかを簡単に説明する。

[前提条件]:

- 探索時にオブジェクト集合 O ・探索開始地点 q が与えられる。
- 各母点領域にオブジェクトがいくつ含まれるかを, 空間索引木に対して付加する。
- 検索範囲として DRQ では, 距離 R が, kNN では個数 k が与えられる。

[DRQ での利用]

図 5 に, DRQ での MBR の利用方法の図を示す。

第一段階として, q 点より半径 R で円を描く。その R の円に一部でも含まれる MBR を持つ母点のみ精査する。図中では, 青い領域に相当する。このとき MBR が R の円に完全に含まれる母点領域にあるオブジェクトに対して精査しない。図中では, オレンジの領域に該当する。また一部分も R の円内に MBR が含まれない領域は, 精査する必要がない。

第二段階では第一段階で精査する必要があると認められた母点領域に対して 空間索引の δ, Δ を利用し, オブジェクトの位置を推定することができる。このとき精査される領域は R の円周上にあり G_{l1} の母点領域になる。

第三段階では, 詳細な位置が必要とみられたオブジェクトに対して探索を行い, R の円に含まれるオブジェクトをすべて知ることができる。

[kNN での利用]

図 6 に kNN での MBR の利用方法の図を示す。

第一段階として, q 点の含まれる MBR をすべての階層で取り出す。下位の階層より各 MBR のもっとも遠い頂点までの q からの平面距離で円を描き, 一部でも含まれる MBR に含まれるオブジェクトの個数が K 以上になるレベル・領域を抽出する。円を書く MBR の頂点と同じレベルに属す MBR だけで判定する。たとえば, 図 6 は, G_{li} での MBR を緑の領域で示す。 q がある MBR の頂点までの距離で描いた円を点線で示している。この円に一部でも含まれる MBR 領域を破線で示した。

この破線に含まれるオブジェクトの個数が K に達したとき次の段階に進む。 K に達しなかったとき, 上位の階層である青い

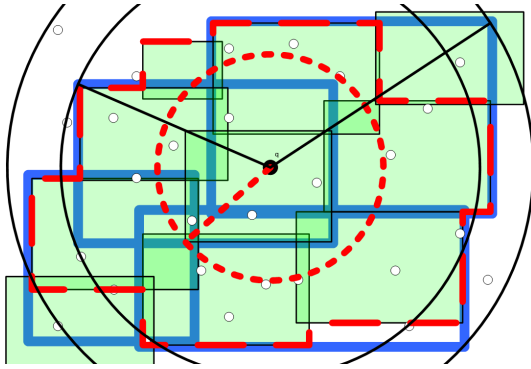


図 6 kNN での MBR 利用の概念
Fig. 6 concept of the MBR use in kNN

領域の q が存在する MBR の頂点まで近い頂点の円 (黒線内側) で同様に数える。K に足りなかったときは、外側でも行い、K に達するまで繰り返す。

第二段階として、母点間の隣接関係のみで抽象化したグラフを用いて経路する領域の δ, Δ を用い計算して、各母点に所属するオブジェクト群の推定位置を推測する。

第三段階として、k 番目のオブジェクトのあるグループの推定位置と重複する部分のあるオブジェクト集合の精査処理を行い k 個のオブジェクトを抽出する。

4. 評価

提案方式のシミュレータを実装し、ノード数の異なるグラフを用いて、空間索引木を生成し、その空間索引木に付加した MBR の大きさなどを数値評価する。

4.1 実データへの適用

ここでは、評価対象の地図として、国土地理院数値地図 2500(空間データ基盤, 1/2500 縮尺) 近畿 - 1 に含まれる京都府と和歌山県のデータを用いる。このデータは、府県の全域にわたってデータとして収録されていないため、ダイクストラ法を用いて、ノードがエッジで連結されている部分のみを抽出した。この抽出により、複数の連結グラフに分けることができる。

図 7.8 の部分に該当する市町村は、京都府京丹後市網野町である。このグラフは 960 ノード, 1325 エッジより成り立っている。また図 9.10 は京都府船井郡京丹波町である。このグラフは 1572 ノード, 2079 エッジより成り立っている。図 7.9 は数値地図 2500 でのすべての情報を描いたものである。図 8.10 は、空間索引木のグラフ生成の母点選択確率は、1/2 において生成し、ある階層での MBR を描いたものである。このとき、図 8 は G_{110} であり母点数 3 だった。また、図 10 は G_{15} であり母点数 45 だった。

図 8.10 での表示は、赤がノード、エッジは青、境界点は黒丸、母点が緑、桃色での四角が MBR を示す。

4.2 評価

生成した空間索引木にたいして、比較および考察する。

評価 1: MBR がどのように各階層において地図そのものの一つの MBR として考えたときの広さに対して、各母点領域の MBR の合計がどう推移していくかをノードの分布が全体的に

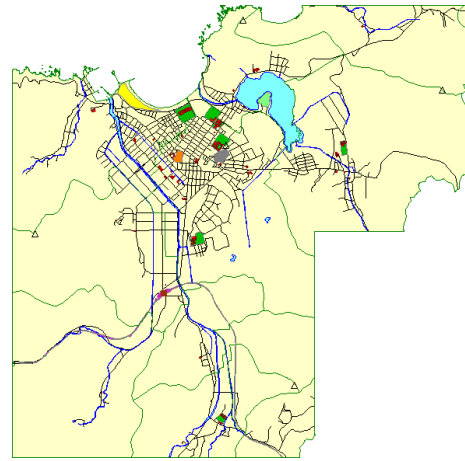


図 7 数値地図 (京丹後市網野町)
Fig. 7 Digital Map of Kyotango City Amino-cho

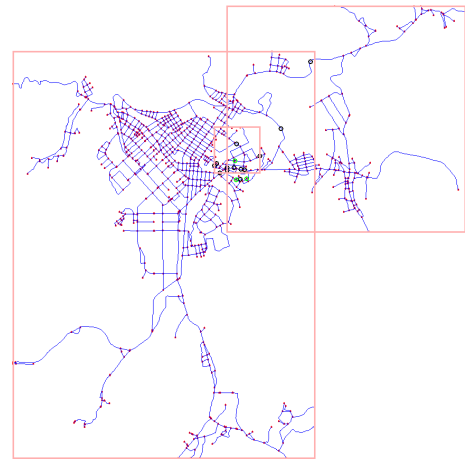


図 8 MBR 生成後の地図 (京丹後市網野町)
Fig. 8 MBR of Kyotango City Amino-cho

ある、図 8 の空間索引木の情報から比較する。

評価 1 の結果 (図 11): 図 11 は、地図全体を 100%としたとき、MBR の面積の単純な合計・MBR が他の MBR と重複する部分の合計・MBR が一度でも被覆した部分の割合を示す。

母点となる確立の高い階層が低い場合、グラフから遠くない部分のみを MBR は覆い、母点のとなる確率の低い階層では、グラフと関係ない部分まで覆うため、全体に対する被覆率が右肩上がりとなる。また MBR の重複はレベル 6 のあたりを頂点に減少を始めている。これはグラフ全体をある一定のサイズ以上の MBR で構成することで、他の領域と重複することが少なくなっていくことによると思われる。

評価 2: MBR の持つ領域の広がり、各階層においてどう広がっていくかをレベル 1 の時の広さを 1 として、6 種類の地図で各階層での大きさを比較評価する。

評価 2 の結果 (図 12): どの地図でも、指数的に一定の増加で領域が広がっていくことがわかる。各種の地図で最上位の階層とその一つ前の階層で領域の広がりが、ほとんど変化しなかったのは、母点ノードの減少が 2 分の 1 でなかったことによると思われる。

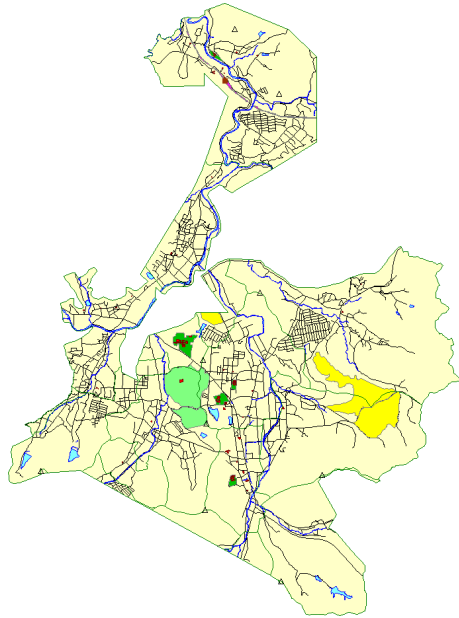


図 9 数値地図 (船井郡京丹波町)
Fig. 9 Digital Map of Funai-gun Kyotanba-cho

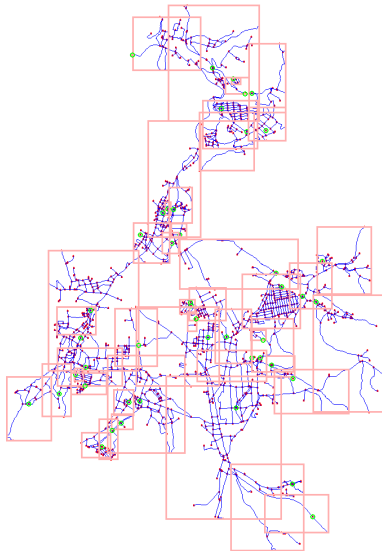


図 10 MBR 生成後の地図 (船井郡京丹波町)
Fig. 10 MBR of Funai-gun Kyotanba-cho

評価 3: ある MBR が他の MBR とどの様に重複する部分をもつか、ノード数が 40323 の地図で、空間索引木を構成したときの情報から確認する。このとき図 13 に示すような形で、確認する MBR が黒実線である場合、他 MBR の重複部分を表記する。また MBR(A) と MBR(B) の関係において、 $A \rightarrow B$ という確認を行なった場合、 $B \rightarrow A$ の確認は行なわないものとする。

各 case について簡単に説明する。case1 から 4 までは、四隅のいずれかのみ重複している。case5 から 8 までは、1 辺の上に他の領域が重複している。case9 は確認する MBR の中に他の MBR が完全に包含している。case10・11 は、確認する MBR を他の MBR が横断または縦断している。case12・13・14・16 は確認する MBR が他の MBR より小さく一辺以上を他の領域に覆われている (case5 から 8 の逆パターンにあたる)。case15

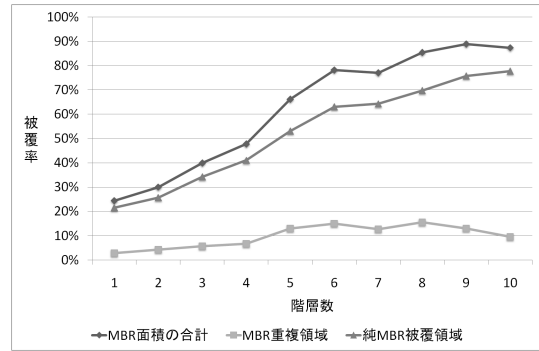


図 11 地図全体にたいしての MBR の広さ
Fig. 11 Area of MBR for the whole map

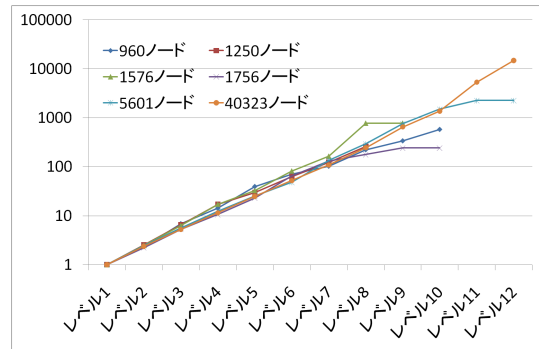


図 12 MBR の持つ領域の広がり
Fig. 12 The expanse of the domain at MBR

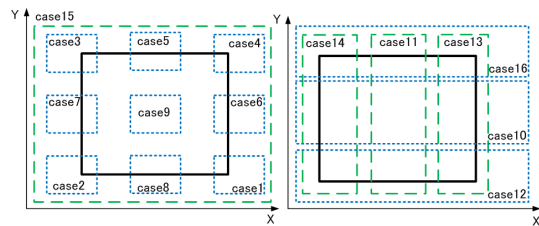


図 13 MBR 重複のパターン
Fig. 13 The patterns of the MBR overlap

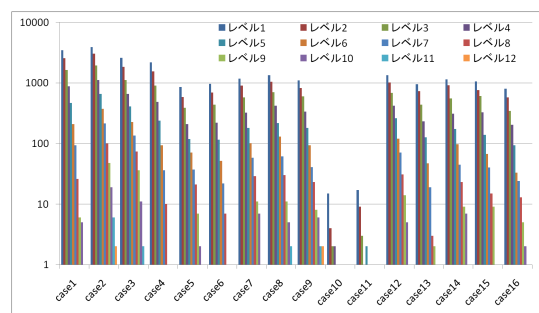


図 14 MBR 重複箇所の合計
Fig. 14 The number of the MBR overlap point

は、確認する MBR が完全に他の MBR に含まれている (case9 の逆パターン)。

評価 3 の結果 (図 14): 縦軸は対数軸で、個数を示しているが、類似した case ごとに同様の数値が出ていることがわかる。case11 と 12 は、存在することが、とても少なく他のノード数

の空間索引木では、確認することが出来なかった。四隅いずれかの部分のみの重複がもっとも多く、残りの重複方法は同等の値を示していることがわかる。

5. 関連研究

Kolahdouzan ら [2] は、ネットワークボロノイ図による領域分割を用いて、k-NN 問題の効率的な計算法を提案している。同手法では、索引化の対象の位置を基準にグラフを生成している。このため索引の対象の種類毎にグラフを生成する必要がある。また、索引化の対象毎のグラフは互いにまったく異なり、階層化などを適用する構造となっていないため、k-NN 以外の道路網応用へ容易に適用できない。

Kriegel ら [3] は、前処理として、参照点をグラフに埋め込み、すべての索引化の対象からの距離を予め計算して、ベクトル空間に写像することで、DRQ と k-NN に適用している。問合せに対しては、参照点までの距離を基準に距離を見積り、フィルタリングと精選の 2 段階探索により高速化している。同様に、Hjaltason ら [4] は、距離空間に一般化してリブシット埋め込みと多次元ベクトルの低次元化の観点から、類似探索についてまとめている。

Cheema ら [5] は、オブジェクト・質問点が矩形領域内にあると仮定して、矩形を利用して half-space と呼ばれる領域を生成し、遠方にある候補となるオブジェクトを持つ矩形領域を刈り込むことで効率的な逆 k 最近傍に適用している。

6. おわりに

本稿では、我々の提案している空間索引木について簡単に説明し、範囲問合せや K 最近傍問題に対応するための各母点領域に対する領域の広がりを示すための MBR の提案を行い、その簡単な数値評価を行なった。

今後の課題として、本手法での包括的評価や DRQ や kNN においての本手法の評価、空間索引木の並列生成法の提案などが考えられる。

文 献

- [1] 蒲原 智也, 上島 紳一 **道路網応用のための空間索引木の提案と最短経路探索への応用**, 情報処理学会論文誌: データベース Vol.2, No.2, pp10–28, 2009.
- [2] M. Kolahdouzan and C. Shahabi, *Voronoi-based K nearest neighbor search for spatial network databases*, Proceedings of the 30th International Conference on Very Large Data Bases . pp849–851, 2004.
- [3] H-P. Kriegel, P. Kroger, P. Kunath, M. Renz and T. Schmidt, *Proximity Queries in Large Traffic Networks*, Proceedings of the 15th ACM International Symposium on Advances in Geographic Information Systems . 2007.
- [4] G.R. Hjaltason and H. Samet *Properties of Embedding Methods for Similarity Searching in Metric Spaces*, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.25 No.5, pp.530–549, 2003.
- [5] M.A. Cheema, X. Lin, Y. YZhang, W. Wang and W. Zhang *Lazy Updates: An Efficient Technique to Continuously Monitoring Reverse kNN.*, Proceedings of the 35th International Conference on Very Large Data Bases(VLDB09), pp.1138–1149, 2009.