

空間連続データに適した補間リンク構造を 有する構造化 P2P ネットワーク

牛久保 辰典[†] 斉藤 裕樹^{††} 戸辺 義人^{††}

センサネットワーク技術の普及などにより位置情報サービス (Location-based Service, LBS) の利用分野が拡大している。本論文では、広域に配置された数多くのセンサによって構成された環境における、センシングデータを効率的に管理する構造化P2Pネットワークの設計について述べる。提案手法では、想定環境で取得されるデータには空間的な連続性を有することに着目し、データを保持するP2Pネットワークに空間補間を用いたリンク構造を与えることによって、冗長な情報を排しつつ空間的な計測値の保存を可能とする。

Interpolation Link Extended P2P Network for Spatially Continuous Data

TATSUNORI USHIKUBO[†] HIROKI SAITO^{††}
YOSHITO TOBE^{††}

We design the structured P2P Networks that efficiently manage sensing data in environment consisting of many sensors arranged in a large area. In the proposed method, we pay attention to the possession of spatial continuousness by the data acquired in assumption environment. Furthermore, we introduce the link structure to use space interpolation for the P2P networks where data was maintained. Therefore, it enables the preservation of a spatial measurement value rejecting redundancy information.

1. はじめに

近年、センサネットワーク技術の発展や GNSS (Global Navigation Satellite System) 機能を備えた小型デバイスの普及により、位置情報サービス (Location-Based Service, LBS) の利用分野が拡大している。このようなサービスでは、環境上に配置された大量のセンサから情報を取得した情報を扱い、大量の端末がサービスを利用するといったサービス形態が考えられる。また位置情報サービスの利用分野には、気象データや実世界の事象の空間分布といったような空間連続データを扱うサービスが考えられる。空間連続データは、実世界上の膨大な情報となるため、膨大なデータをネットワーク上で管理できるスケーラビリティのある手法が求められる。空間連続データを分散管理する手法として、実世界の構造を反映した P2P ネットワークが注目されている。

膨大な実世界の情報を扱う技術として、VoltDB [1]などの分散アーキテクチャに適したインメモリのリレーショナルデータベースシステムがある。VoltDB は範囲検索に対しても高い性能を有している。また、大規模なデータの格納に関する研究として M.Ahuja らの研究[2]がある。M.Ahuja らは 6PB もの大規模なデータをリレーショナルデータベースシステムで管理する手法を提案している。しかしながら、これらの研究では、スケーラビリティと柔軟性に関してあまり検討されていない。FunctionDB [3]と MauveDB [4]は線形補間や曲線回帰を用いたグラフ表示を補助する機能を有する。しかし、線形補間や曲線回帰を行うために近似値の計算を行う必要があるため、P2P アーキテクチャには適していない。TSAR [5]はセンサネットワークに適した 2 層の分散ストレージのアーキテクチャであり、Interval Skip Graph を利用している。しかしながら、TSAR は空間連続データには最適化されていない。

そこで、本論文では位置情報サービスに用いられる空間連続データを、データ補間技術を用いて管理する補間リンク構造を用いた構造化 P2P ネットワークの設計について述べる。本論文の構成は以下のとおりである。第 2 章では空間連続データを扱うために必要な要件を述べる。第 3 章では本研究の基盤となる Skip Graph [6]と空間充填曲線について解説した後、補間リンクの設計について述べる。第 4 章では補間リンクを用いた結果の補間精度について評価を行う。第 5 章で本論文のまとめと今後の課題について示す。

[†] 東京電機大学未来科学研究科情報メディア学専攻
Graduate School of Science and Technology of Future Life, Tokyo Denki University
^{††} 東京電機大学未来科学部情報メディア学科
Department of Information Systems and Multimedia Design, Tokyo Denki University

2. 連続データ処理におけるデータ補間

分散環境で空間連続データを管理するには2点の課題がある。広範囲検索におけるデータの冗長性と空間連続データが膨大な点である。本章では、連続データ処理の要件である広範囲検索におけるデータの冗長性について説明し、冗長かつ膨大なデータに適したP2Pアーキテクチャについて述べる。

2.1 連続データ処理の要件

本研究の目的は空間連続データからデータの特徴量をもとにデータ補間に必要なデータを抽出することである。本研究で扱う空間連続データとは、位置情報に対応する値を有し、その値が空間的に連続しているデータである。空間連続データは実世界上の非常に広範囲なデータであり、データに特徴のある範囲も存在するが、平坦でデータに変化の見られない範囲も多い。このような空間連続データにおいて広範囲に検索を行った場合、平坦で変化の見られない領域に存在する似た値のデータが多く取得され非効率的である。そのため、データの特徴から必要なデータのみを抽出することが必要となる。

2.2 P2P ネットワークによる連続データ処理手法

空間連続データは実世界上の膨大な情報であるため、スケーラビリティに優れたデータ管理手法として P2P アーキテクチャを適用する。本研究では、特に範囲検索に優れる Skip Graph を用いる。また、空間連続データを Skip Graph へ格納する際、二次元情報である位置情報を一次元情報へ変換する必要がある。そこで、二次元情報を一次元情報に変換する手法として、空間充填曲線を用いる。

3. データ補間を用いたスキップ構造を備える P2P ネットワーク

本研究では、2章で述べた問題点を解決するために、Skip Graph を拡張した補間リンクを提案する。補間リンクでは、広範囲検索におけるデータの冗長性を避けるために、大規模な空間連続データの中から補間に適したデータを抽出しリンクを形成する。本章では Skip Graph および Skip Graph を拡張したリンク構造である補間リンクについて述べる。

3.1 Skip Graph

Skip Graph とは、スキップリストを応用した構造化オーバーレイネットワークである。Skip Graph の構造の例を図1に示す。Skip Graph には key と value をそれぞれひとつずつ保持するノードが存在する。図1では数値の書かれた正方形がノードを表し、ノード内の数値が key を示す。なお本稿では、説明を容易にするために key には数値を用いる。key は順序付け可能とし、各ノードは与えられた key の順に整列する。各ノ

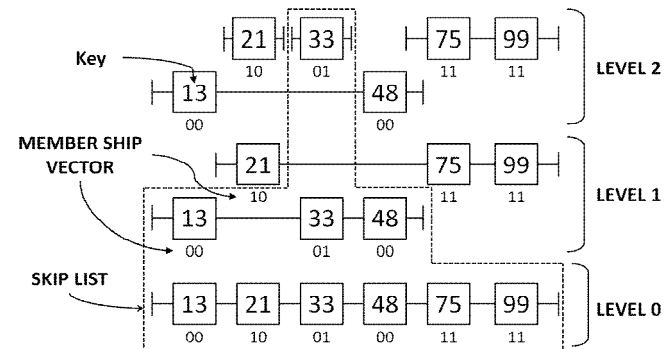


図1 SkipGraphs の構造

ドには membership vector と呼ばれる値が割り当てられている。各ノードは membership vector が先頭から i 文字 ($i = 0, 1, 2, \dots$) 等しいものと部分ネットワークを形成する。その各部分ネットワークにおいて、key の値が自身と前後に最も近いものと双方向の P2P 接続を確立する。特に、membership vector が頭から i 桁等しいものを Level i の接続と呼ぶ。Skip Graph で空間連続データを扱うためには、位置情報を key に格納する。位置情報である二次元情報を key へ格納するために、空間連続データの位置情報に対して空間充填曲線を用い、位置情報を一次元の情報へと変換することにより、位置情報を key へ格納することが可能となる[7]。

3.2 補間リンク

大規模な空間連続データを有するノードの中から冗長な値を排し、有効な値のみ抽出しリンクを形成する補間リンクを提案する。補間リンクとは、変化が少なく近隣のデータと同様の値を持つデータを省くことで補間に適したデータのみを抽出し双方向リンクを形成したものである。補間に適したデータとは、周囲のデータに比べて隣接したノードが有するデータから大きく変化したデータであると考えられる。また、近隣のデータと比べて変化が小さく、平坦なデータは補間を行う際は必要ではない。

ここで、 N_i : i 番目のノード、 K_i : N_i の key、 V_i : N_i に格納された value、 a_i, b_i, c_i : N_{i-1}, N_i, N_{i+1} の 3 ノードから算出される二次関数の各係数とする。 a_i は二次関数の二次項の係数であり、 a_i は二次関数の傾きを表す。傾き a_i が大きい場合データの変化は急になり、 a_i が小さい場合データの変化は緩やかとなる。そのため、 a_i の大きさに基づくリンクを生成することによって、補間に適したデータを容易に抽出することができる。ここで、 a_i, b_i, c_i は以下の行列演算を解くことで算出される。

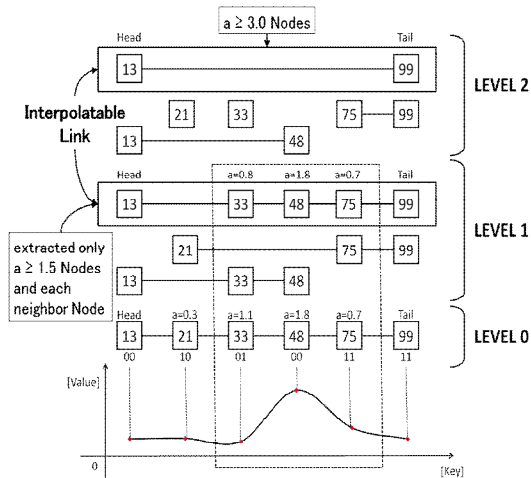


図 2 補間リンクの構造

```

Algorithm 1: newNodeInsert()
1:  $N_i \leftarrow$  New Node;  $d \leftarrow$  threshold;
2: SkipGraphRouting;
3: sendQuery to  $N_{i-1}, N_{i+1}$ ;
4: receiveKeyValue from  $N_{i-1}, N_{i+1}$ ;
5:  $a_i \leftarrow$  calcSlope;
6: IF  $a_i > d$  THEN
7:    $N_i$  join to Next Level;
8:    $N_{i+1}$  join to Next Level;
9:    $N_{i-1}$  join to Next Level;
10: END IF
11: WHILE  $j \leq$  tailNodeIndex DO
12:   IF  $N_j$  joined to Next Level THEN
13:     setNextLevelRight( $N_j$ );
14:   END IF
15:    $j = j + 1$ ;
16: END WHILE
17: WHILE  $k \geq$  headNodeIndex DO
18:   IF  $N_k$  joined to Next Level THEN
19:     setNextLevelLeft( $N_k$ );
20:   END IF
21:    $k = k - 1$ ;
22: END WHILE
    
```

図 3 補間リンク構築アルゴリズム

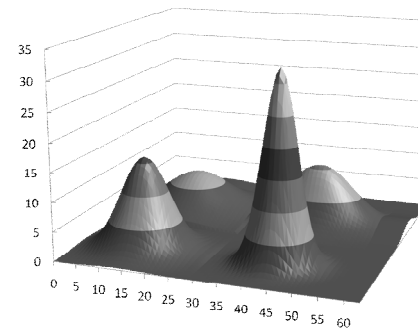


図 4 補間リンク構築アルゴリズム

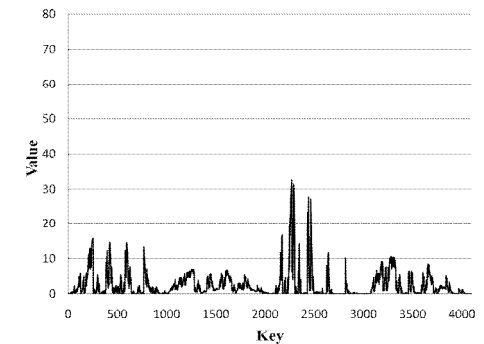


図 5 一次元情報に変換した元データ

と key=99 が無条件に抽出される．最後に、しきい値によって抽出された key=48 の両隣の key=33 と key=75 が抽出され、各ノードは双方向に P2P 接続を確立し、Level 1 の補間リンクが生成される．Level 1 内のノード間で傾きの計算を再度行い、傾き a を更新する．更新された a を参照して Level 2 のしきい値 3.0 でノードの抽出を行う．図 2 のように抽出されるノードが無い時点で補間リンクの生成を終了する．

3.3 補間リンク構築アルゴリズム

本節では、補間リンクが導入された Skip Graph における新規ノード追加時の補間リンク構築アルゴリズムについて述べる．アルゴリズムを図 3 に示す．図 3 において、level k に新規追加されるノードを N_i とする．新規ノード追加時の補間リンク構築アルゴリズムの手順について以下に記す．

- 1) N_i は既存の Skip Graph の新規ノード追加アルゴリズムを使用してノードを追加
- 2) N_i は level k において N_{i-1} と N_{i+1} に対して自身の key と value を送信すると同時にそれぞれの key と value を要求するクエリを送信
- 3) クエリを受け取った N_{i-1} と N_{i+1} は N_i へ key と value を送信
- 4) key と value を受け取った N_i は式(1)から傾きとなる a_i を算出
- 5) 算出した a_i をしきい値 d と比較、 a_i が d を超えた場合は N_i とその左右のノードである N_{i-1} と N_{i+1} を level $k+1$ の補間リンクへ抽出
- 6) N_i はフラッディングを行い level $k+1$ の中で両隣となるノードを探索、フラッディングで発見した両隣のノードと双方向の接続を確立

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \begin{pmatrix} K_{i-1}^2 & K_{i-1} & 1 \\ K_i^2 & K_i & 1 \\ K_{i+1}^2 & K_{i+1} & 1 \end{pmatrix} \begin{pmatrix} V_{i-1} \\ V_i \\ V_{i+1} \end{pmatrix} \quad (1)$$

式(1)により算出された a_i を用いて、しきい値 d を超える変化量の大きいノードを抽出する．また、抽出されたノードの左右のノードと Level 0 のリンクにおける先頭ノードと終端ノードは無条件に補間リンクとして抽出する．上記の手法で抽出されたノードで双方向の P2P 接続を確立し補間リンクを形成する．しきい値 d はアプリケーション依存とし、アプリケーション開発者が適宜設定する．また、Level i ($i = 0, 1, 2, \dots$) 時のしきい値を d_i としたとき、 $d_i > d_{i-1}$ となるように設定する．補間リンクは各レベルで形成され Level i の補間リンクは Level $i-1$ の補間リンクからノードを抽出し形成される．

補間リンク生成手順の例について、図 2 に示す．図 2 において、しきい値 d を 1.5 としたとき、Level 1 の補間リンクはしきい値 1.5 で間引かれる．その結果、 $a=1.8$ である key=48 のノードが抽出される．その後、Level 0 における始端と終端である key=13

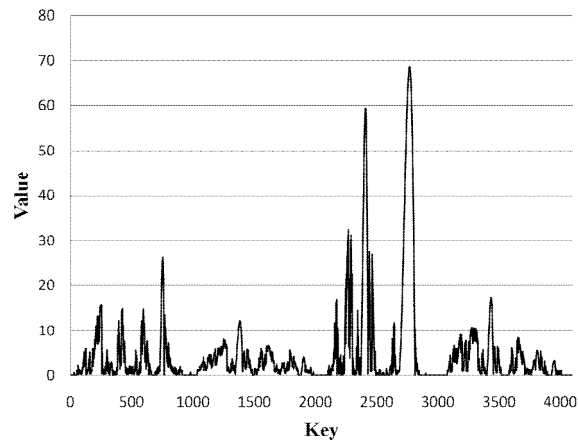


図 6 三次スプライン補間の結果

上記の6ステップを行うことで補間リンクの構築が完了する．本アルゴリズムはlevel $k+n$ ($n = 1, 2, 3, \dots$)において，補間リンクに抽出されるノードが存在しなくなり次第終了する．

4. 評価実験

本章では，提案する補間リンクの補間精度について評価について述べる．実験では，補間リンクを用いた結果と元データとの比較を行った．

評価には人工的に生成した空間連続データを用いた．評価に用いたデータを図4に示す．データ数は4096 (64×64)とし，空間正規分布を含む．二次元情報である位置情報は空間充填曲線の一種であるZ-orderingを用いて一次元情報へと変換した．位置情報を一次元情報に変換した後のkeyとvalueの値の関係のグラフを図5に示す．補間リンクに抽出されたノードの数は元データ数の約50%とした．

生成された補間リンクのデータに対して三次スプライン補間を用いて復元を試みた補間データに関するグラフを図6に示す．元データ(図5)と補間データ(図6)を比較した結果，補間データが元データの特徴をおおまかに再現していることが分かった．また，補間データと元データとの誤差平均は3.96であった．このことから，補間リンクは元データの特徴を抽出する手法として適しているということがわかる．

5. おわりに

本稿では，位置情報サービスに用いられる空間連続データを管理する構造化P2Pネットワークとして，Skip Graphを拡張した補間リンクを提案し，評価を行った．提案手法を用いることで元データから補間に適切なデータを抽出可能であることが評価によって示された．今後は，時系列データを考慮した手法の検討を行っていく．

本研究は，東京電機大学総合研究所課題Q11-J06として実施したものである．

参考文献

- 1) R. Kallman, H. Kimura, J. Natkins, A. Pavlo, A. Rasin, S. Zdonik, E. P. C. Jones, S. Madden, M. Stonebraker, Y. Zhang, J. Hugg, and D. J. Abadi. "H-store: a high-performance, distributed main memory transaction processing system", *Proceedings of VLDB Endow.*, vol. 1, iss. 2, pp. 1496-1499, 2008.
- 2) M. Ahuja, C. C. Chen, R. Gottapu, J. Hallmann, W. Hasan, R. Johnson, M. Kozyrczak, R. Pabbati, N. Pandit, S. Pokuri, K. Uppala. "Peta-Scale Data Warehousing at Yahoo!", *Proceedings of the 35th SIGMOD international conference on Management of data*, pp. 855-861, 2009.
- 3) A. Thiagarajan, and S. Madden. "Querying continuous functions in a database system", *Proceedings of the ACM SIGMOD international conference on Management of data*, pp. 791-804, 2008.
- 4) A. Deshpande, and S. Madden. "MauveDB: Supporting Model-based User Views in Database Systems", *Proceedings of the ACM SIGMOD international conference on Management of data*, pp. 73-84, 2006.
- 5) P. Desnoyers, D. Ganesan, and P. Shenoy. "TSAR: A Two Tier Sensor Storage Architecture Using Interval skip Graphs", *Proceedings of ACM Sensys 2005*, pp. 39-50, 2005.
- 6) J. Aspnes and G. Shah, "Skip graphs", *Fourteenth ACM SIAM Symposium on Discrete Algorithms (SODA)*, pp. 384-393, 2003.
- 7) Y. Shu, B. C. Ooi, K.-L. Tan, and A. Zhou, "Supporting multi-dimensional range queries in peer-to-peer systems", *Proceedings of the Fifth IEEE International Conference on Peer-to-Peer Computing (P2P'05)*, pp. 173-180, 2005.