空間連続データに適したモデル駆動型 P2P ネットワークの設計

牛久保 辰典[1]　斉藤 裕樹[2]　戸辺 義人[3]　鉄谷 信二[2]

概要：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．センサネットワークのような分散計測環境では，取得されるセンサのありのままのデータは不完全で不均一な情報であるため，利用者が有効に利用することが難しい．情報の有用性を高めるためには，計測値に対して統計モデルや推計モデルなどを適用し有効な情報を抽出することが求められる．本論文では，物理的なセンシングレイヤを抽象化し，データの中間表現を行うモデル表現レイヤを提案し，統計モデルを元にした空間連続データを管理できる機構を検討する．次に，膨大な実世界情報を扱うため，P2P ネットワーク上でモデル表現レイヤを分散管理する手法を検討する．提案手法は，統計モデルに基づき有効な情報を抽出し，ノード間に階層的なリンク構造を構築することで，有効な精度での検索を可能にするものである．提案手法の評価のため，実世界の気温データの管理・検索に本手法を適用した結果，有効性が確認された．

Design of Model-driven Peer-to-peer Network for Spatially Continuous Data

Tatsunori Ushikubo[1]　Hiroki Saito[2]　Yoshito Tobe[3]　Nobuji Tetsutani[2]

Abstract: The advance technology for sensor networks has enabled Location-based Services (LBS), and many applications have been developed. Since real-world data especially generated by distributed measurement infrastructures such as sensor networks tends to be incomplete and imprecise, it is not suitable to present it to users or applications. To improve availability of real-world data, it is needed to apply statistical or probabilistic models that can provide robust interpretation of the data. In this paper, we define a new abstraction layer, called Model Representation Layer, that allows us to manage real-world data by using statistical representation. And then, we design the structured P2P network that efficiently manages sensing data in environment consisting of many sensors arranged in a large area. Furthermore, we introduce the link structure to use statistical method for the P2P network where data was maintained. Through our prototype implementation that manages temperature data in Japan, we confirmed its effectiveness.

[1] 東京電機大学大学院未来科学研究科情報メディア学専攻
Graduate School of Science and Technology for Future Life, Tokyo Denki University
5 Senju-asahi-cho, Adachi-ku, Tokyo 120–8551, Japan
[2] 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
5 Senju-asahi-cho, Adachi-ku, Tokyo 120–8551, Japan
[3] 青山学院大学理工学部情報テクノロジー学科
Department of Integrated Information Technology, Aoyama Gakuin University
5-10-1 Fuchinobe, Chuo-ku, Sagamihara-shi, Kanagawa 252–5258, Japan

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．センサネットワークのように環境上に広範囲に分散した計測基盤では，センシングデータの部分的な欠落やセンサの誤差，センサ密度の不均一さなどから，収集されたありのままの計測値は利用者が有効に利用することが難しい．一方，気温などの空間連続データでは，データの変化に特徴のある領域も見られるが，領域内での値が平坦でどのデータにもあまり違いが見られない場合も多いため，冗長な情報も多く含まれる．従来，このような実世界のセンシングデータはデータベースに格納され，利用者やアプリケーションは，不完全なセンシングデータから必要な情報を抽出する処理を行う必要があった．Matlab や，S や R などの統計分析ツールは，解析モデルを用いて，情報の抽出を行うことが可能であるが，データベースのように大量の情報を格納・管理・検索を効率的に行う機能は備わっていない．また，静的な情報を扱うため，リアルタイムに計測されるセンシングデータの処理には向いていない．

また，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，インメモリのリレーショナルデータベースシステム [1], [2] や，統計モデルを扱うことができるモデル駆動型データベース [3], [4] などが存在する．しかしながら，広範囲の情報を分散アーキテクチャで扱えるようなスケーラビリティと柔軟性についてはあまり検討されていない．

本研究では，これらの問題を解決するために，生のセンシングデータとその計測値を利用するアプリケーションの間に，データの中間表現を行う新たなレイヤを導入し，統計モデルを元にした空間連続データを管理できる機構を検討する．次に，P2P ネットワーク上でモデル表現レイヤの情報を分散管理するためのアルゴリズムについて検討を行う．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の要件と提案するモデル表現レイヤの原理について述べ，第 3 章では，P2P アーキテクチャを用いたモデル表現レイヤの実現手法とアルゴリズムについて述べる．第 4 章では，提案手法の評価実験と実験結果の考察を行う．第 5 章で関連研究について述べ，最後に，第 6 章で本論文のまとめと今後の課題を示す．

2. モデル表現レイヤの提案

実世界の空間連続情報を効果的に管理し，利用者に対して良質なデータを提供するためには，計測インフラによるありのままの計測値の不完全な情報を補う必要や，広範囲領域処理における冗長なデータを排する必要がある．本章では，連続データ処理の要件とデータの抽象化手法について述べる．

2.1 空間連続データ処理の要件

センサネットワークなどで取得できる実世界のセンシングデータは，広範囲に分散した各々のセンサのありのままの計測値を集約したものである．このようなセンサによって計測されるデータは，センサ自身の故障や停止，データ転送経路上での通信エラーなどにより，データの部分的な欠落が発生しやすい．また，各々のセンサによる計測値の誤差や，計測を行う地点が不均一で標本の偏りがあることが多く，そのままでは有用なデータとして利用することが難しい．一方，計測されるデータの性質に着目すると，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が問題となる．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．具体的には，気温などの空間連続データは実世界の広範囲なデータであり，データの変化に特徴のある範囲も存在するが，全体的に計測値が平坦で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大量に取得され非効率的である．

以上の問題は，計測値をありのままに格納・検索を行うような従来のデータベースシステムをセンシングデータにそのまま適用していることに起因する．例えば，天気の予測や気温や降水量の分布，交通量や事故の分析を行うためには，「生のデータ」から統計的に有効な特徴を抽出する必要がある．しかしながら，従来のセンサネットワークシステムでは空間連続データを扱う効果的なデータ管理の仕組みがあまり検討されていないため，有効な分析には数学モデルによる解析ソフトウェアを用いることや，アプリケーションごとに統計処理フィルタを実装することが求められてきた．

本研究では，これらの問題を解決するために，生のセンシングデータとその計測値を利用するアプリケーションの間に，データの中間表現を行う新たなレイヤを導入し，統計モデルを元にした空間連続データを管理できる機構を検討する．提案手法は，膨大なデータを扱うために構造化 P2P モデルを用いた分散アーキテクチャで構成されるものである．

2.2 モデル表現レイヤによるデータの抽象化

本節では，空間連続データを統計モデルで扱うモデル表現レイヤについて説明する．モデル表現レイヤは，実世界の情報を領域ごとに分散管理する P2P アーキテクチャに適した構造であり，統計モデルを用いることでセンシングレイヤにおける物理的なセンサネットワークの構造やセンシングにおける不完全な計測値を隠蔽する．本手法の特徴

は以下のとおりである．

（１）独立性

モデル表現レイヤは，センシングレイヤにおける生の計測値およびアプリケーションから独立したものとして振る舞う．センサネットワークでは，ノードの増減やネットワーク構造の変化が想定される上，計測値を扱う際には欠落や標本の偏りに対応する必要がある．モデル表現レイヤにより，アプリケーションは，下位のセンシングレイヤの構造に依存せずに実世界情報を取得・利用することができる．

（２）統計モデルによるデータ解析

実世界の空間的連続データは，全体的に計測値が平坦で変化の見られない範囲が多く見られるため，これらの冗長なデータを省略するとともに，欠落した計測値や個別のセンサの誤差を補正するために，統計モデル表現によるデータの管理を行う．これにより，利用者がデータを利用した分析を行うのに有用な情報を提供することが可能である．

（３）動的計測への対応

実世界の情報は各々のセンサによりリアルタイムに取得されるため，モデル表現レイヤは，計測値の変化によって動的に保持するデータと表現モデルを更新する機構を有する．

2.3 モデル表現レイヤの構成

本節では，モデル表現レイヤのデータ構造について述べる．なお，ここでは，二次元平面上の気温を計測するセンサネットワークシステムを例に取り上げ，データ構造を説明する．計測データのスキーマは，(時刻, x 座標, y 座標, 気温, センサ ID) である．このデータが各センサノードからシステムにリアルタイムに送られるものとする．

二次元平面上に気温データを割り付け，統計モデルに基づき必要に応じて誤差の修正や計測値の予測をするために，均一な格子状のデータ構造を用いる．**図 1** は，格子状のデータ構造を備えたモデル表現レイヤの例である．格子状の表現を用いることにより，実世界の情報を行列構造で近似することが可能である．なお，格子の分割密度は想定されるセンシング範囲やデータの質により決定される．各格子上の点における値は，生の計測値から推定が可能である．推定の手法としては，回帰分析によって曲線近似を行う手法と，補間により生の計測値から格子上の点の値を得る手法がある．

回帰による分析

回帰分析では，(x, y) の点の気温を近似するために以下のような関数を用いる．

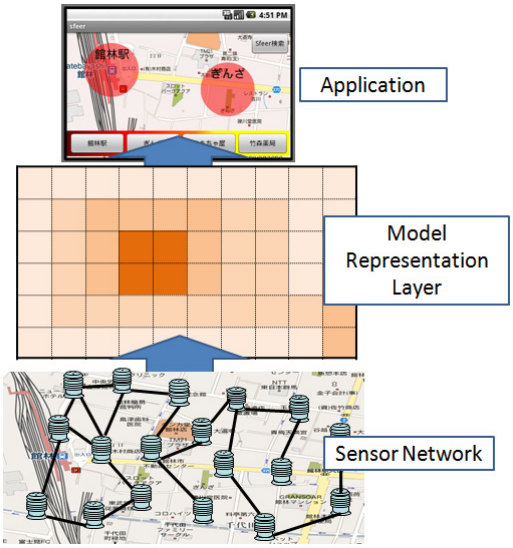

図 1 モデル表現レイヤの構造
Fig. 1 Structure of Model Representation Layer.

$$regtemp(x, y) = w_1 + w_2 x + w_3 y + w_4 x^2 + w_5 y^2 ... (1)$$

各項の重み $w_1, \ldots, w_n$ を，生の計測値との差が小さくなるようにする．一般的には，実測値 $temp(x, y)$ を用いて，以下の式で平均二乗誤差を求めたとき誤差が最小となるように決定する．

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}(regtemp(x_i, y_i) - temp(x_i, y_i))^2} \qquad (2)$$

この最適化問題を解く方法としては，ガウスの消去法 [5] が良く知られている．この結果得られた回帰関数 $regtemp(x, y)$ によって，格子上の任意の点の気温を計算することが可能となる．

補間による分析

補間は，実測値から欠落した値を推定する手法であり，最近傍補間法，線形補間法，スプライン補間法，クリギング法などが知られている．例えば，(x_1, y_1)，(x_1, y_3)，(x_3, y_1)，(x_3, y_3) におけるそれぞれの気温 $temp(x_1, y_1)$，$temp(x_1, y_3)$，$temp(x_3, y_1)$，$temp(x_3, y_3)$ が得られており，$x_1 \le x_2 \le x_3$ かつ $y_1 \le y_2 \le y_3$ のとき，(x_2, y_2) の気温を補間するためには，バイリニア補間を用いると以下の式で推定値を得ることができる．

$$\begin{aligned} temp(x_2, y_2) = {} & (x_3 - x_2)(y_3 - y_2)temp(x_1, y_1) + \\ & (x_3 - x_2)(y_1 - y_2)temp(x_1, y_3) + \\ & (x_1 - x_2)(y_3 - y_2)temp(x_3, y_1) + \\ & (x_1 - x_2)(y_1 - y_2)temp(x_3, y_3) \end{aligned} \qquad (3)$$

なお，これら以外にも統計分析の分野では，カーネル補間法，ロジスティック回帰，ノンパラメトリック法といっ

た多くの回帰分析および補間手法が知られている．提案するモデル表現レイヤでは，このような様々な回帰分析および補間手法が適用可能である．

3. P2P アーキテクチャによるモデル表現レイヤの実現手法

3.1 空間連続データの P2P での実現

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構造化 P2P アーキテクチャを適用する．代表的な構造化 P2P ネットワークには，Chord[6], CAN[7], Kademilia[8] などの分散ハッシュ表（DHT）が知られている．DHT では，ハッシュ変換に基づいて分散されたノードで情報を管理し，ルーティングテーブルを構築することにより効率的なデータの分散管理と検索を実現している．しかし，構造化 P2P ネットワークを用いて空間連続データを扱うためには，連続量を扱った範囲検索を可能とする必要があることからハッシュ変換を行わずに情報の持つ順序性を保存することが必要である．SkipGraph[9] はスキップリスト [10] の構造を分散環境で実現している手法であり，ハッシュ変換を用いた DHT の代わりに，連続した値を ID に用いることにより範囲検索を可能にしている．一方，位置情報のような多次元空間の情報を P2P ネットワーク上の ID のような一次元空間にマッピングするための手法としてルベーグ曲線（Z-Ordering），ヒルベルト曲線，シェルピンスキー曲線などの空間充填曲線が知られている．

本研究では，P2P ネットワーク上でモデル表現レイヤの情報を分散管理するため，まずモデルで表現された実世界の多次元の位置情報を空間充填曲線を用いて一次元の ID 空間上に変換する．次に SkipGraph のように一次元化された ID ごとにノードを割り当て，ID の順番ごとに前後に隣接するノード同士で双方リンクを形成し，ID で整列された一次元の線形リスト構造を構築する．

3.2 補間リンクの構築

モデル表現レイヤは，格子上のすべての点における計測値を有するものである．空間内の計測値は，領域内で変化に特徴が見られる領域も存在するが，領域内が似た値をとり周囲と比較してほとんど変化のない領域も多い．広範囲検索を行う際に，このような冗長なデータが取得されることは非効率的である．したがって，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，有効な値のみを抽出しリンクを形成する．この構造を補間リンクと呼ぶ．利用者は，補間リンク上のノードの情報を用いて検索を行い，得られた検索結果にもとづき補間を行うことで，実際の計測値に近い結果をより少ないデータ量で得ることができる．

ここでは，補間リンクを形成するノードの選出方法を例

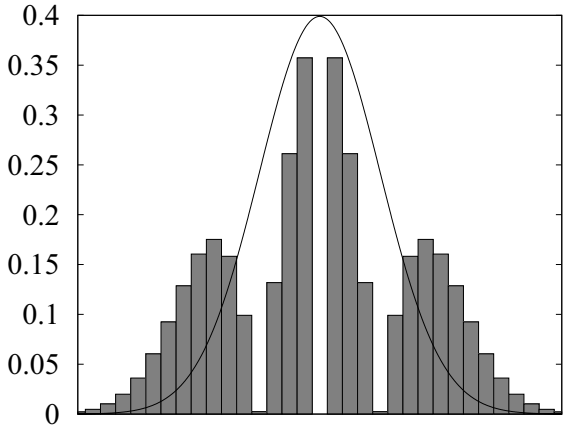

図 2 ガウス分布およびその二階微分値
Fig. 2 Gaussian Distribution and Its Decond Degree Differentiation.

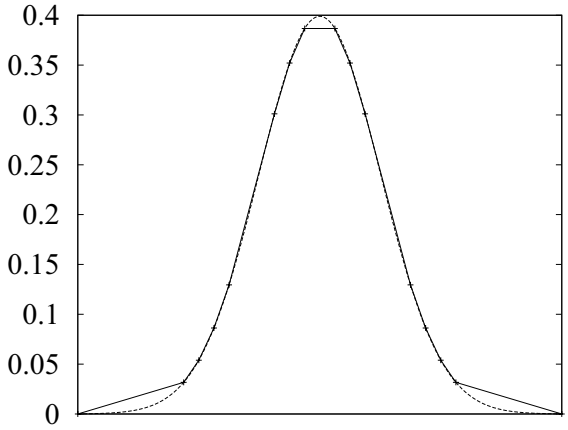

図 3 二階微分値 0.1 以上の点を結んだ近似曲線
Fig. 3 Approximated Curve Connected with Second Degree Differentiation Value of 0.1 or above.

を用いて説明する．正規分布に代表されるガウス分布関数は，実世界の分布を統計モデルで解析する際によく用いられる．**図 2** は以下のガウス分布関数 $N(\mu,\sigma)$ および，一定区間ごとの二階微分値 $\frac{d^2}{dx^2}N(\mu,\sigma)$ を図示したものである．

$$N(\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{4}$$

曲線が連続な関数で表現されているとき，二階微分値は曲率を示し，一定区間において値がどの程度変化しているかを知ることができる．二階微分値が大きい程傾きの変化の度合いが大きい，すなわちカーブがきつい領域であり，小さい程傾きが変化しない，すなわち直線的な領域である．このことから，二階微分値の大きな領域のデータのみを残し，小さな領域のデータを排することで，冗長な値を取り除くことができる．**図 3** は，二階微分値が 0.1 以上の点のみを結んだものと，元の関数を比較したグラフである．これより，特徴のある一部の点の情報から，元の関数に近似

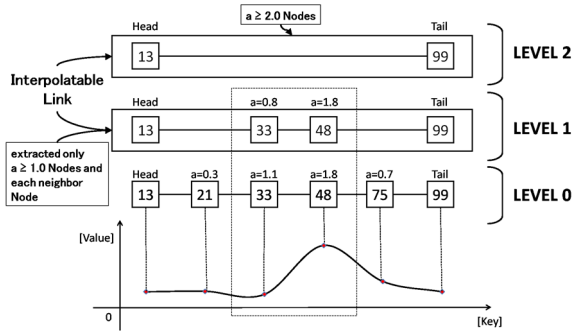

図 4 補間リンクの階層的構造
Fig. 4 Hierarchical Structure of Interpolatable Link.

した結果が得られることが分かる．このような手法により変化量に注目しノードを抽出することで，補間リンク形成に必要なノードを選出する．

また，補間リンクでは，ノードを選出する際の二階微分値のしきい値 d により，元のデータとの近似の度合いと選出されるノードの割合が決定される．d が小さければ，近似の精度が高い反面，選出ノード数が多くなり，d が大きければ，近似の精度が低い反面，選出ノードが少なくて済む．補間リンクでは，d を複数設定したリンクを階層的に構築する．これにより，利用者は階層を指定することで，必要に応じた精度のデータを得ることを可能とする．範囲検索の際には，指定されたレベルのノード間のリンクをたどることで，目的の領域の近似値を得ることができる．

図 4 は，モデル表現レイヤの計測値と補間リンクの構造を例示したものである．まず，モデル表現レイヤのすべての計測値から，すべての隣接ノードを結ぶレベル 0 の双方向リンクを構築する．次に，二階微分値を求め，しきい値 1.0 以上のノードである 13, 33, 48, 99 を選出し隣接するノードとの間でレベル 1 の双方向リンクを構築する．さらに 2.0 以上の二階微分値を持つノードによりレベル 2 のリンクを構築し，以下同様の手順で高レベルのリンクを構築する．

このような上位の補間リンクは，レベル 0 のリンクに対して，短縮経路として検索に必要なホップ数を低減する働きをする．すなわち，任意のノードの情報を検索する際には，上位レベルのリンクによりルーティングを行うことで，検索に要するホップ数を低減することが可能である．

3.3 ノード選出アルゴリズム

本節では，ノード同士の通信によって上位レベルのノードを決定するアルゴリズムについて述べる．P2P アーキテクチャを用いたアルゴリズムでは，3.2 で述べた二階微分近似値を局所的なノード間の通信で算出する必要がある．そのために，隣接ノードの値を用いて 2 次関数による曲線近似を行う．具体的には，N_i:i 番目のノード，K_i: N_i の ID，V_i: N_i に格納された計測値，a_i, b_i, c_i: N_{i-1}, N_i, N_{i+1} の

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

図 5 補間リンク構築アルゴリズム
Fig. 5 Interpolatable Link Construction Algorithm.

3 ノードから算出される二次関数 $y_i = a_i x^2 + b_i x + c$ の各係数とすると，a_i は導関数の定義よりこの二次関数の二階微分値となる．ノード選出にはこの値を用いる．

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．

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \begin{pmatrix} K_{i-1}^2 & K_{i-1} & 1 \\ K_i^2 & K_i & 1 \\ Ki - 1^2 & K_{i+1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} V_{i-1} \\ V_i \\ V_{i+1} \end{pmatrix} (5)$$

式 (5) により選出された 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．またレベル 0 リンクの始点のノードと終点のノードも含める．これにより抽出されたノードについて，隣接するノード同士を双方向リンクで結ぶ．以上のリンク構築をしきい値 d を正の整数倍した値について行い，しきい値 d に対して n 倍した際のリンクを「レベル n」のリンクとして階層化する．

図 5 は，新規ノード追加時の補間リンク構築アルゴリズムを示したものである．図 5 において，レベル $k+1$ 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は以下のとおりである．

(1) N_i は，レベル k において N_{i-1} と N_{i+1} に対して自身の ID と計測値を送信すると同時に，それぞれの ID と計測値を要求するクエリを送信する．
(2) クエリを受けとった N_{i-1} と N_{i+1} は，N_i へ計測値を送信する．
(3) ID と計測値を受け取った N_i は，式 (5) から二階微分

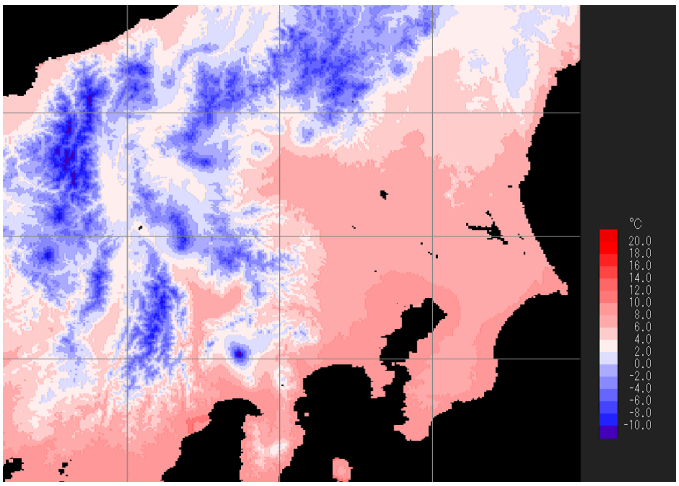

図 6 評価に用いたデータセット
Fig. 6 Dataset for Evaluation.

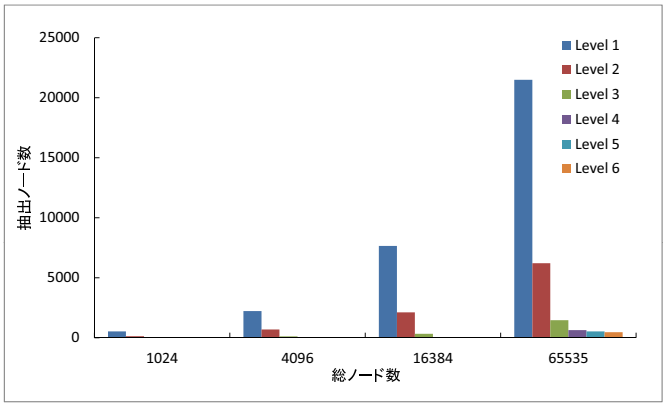

図 7 総ノード数 N と各レベルの抽出ノード数
Fig. 7 Number of Nodes and Extracted Nodes in Each Level.

近似値 a_i を算出する．

(4) 算出した a_i をしきい値 d と比較し，a_i が d を越えた場合は N_i はレベル $k+1$ の補間リンクに選出される．

(5) N_i はフラッディングを行い，レベル $k+1$ の中で両隣となるノードを探索し，双方向のリンクを確立する．

上記の 5 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，低レベルから高レベルに向けて反復し，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を構築した際の範囲検索コストの評価および，補間リンクを用いた補間データの誤差の評価について述べる．

4.1 評価用データセット

評価には，気象庁のメッシュ気候値 2000[12] を用いた．メッシュ気候値とは，年間の平均気温や最高気温などの気候情報を 1km メッシュで推定した気象データである．本実験では，1971 年〜2000 年の間の関東地方全域の 3 月の気温の平均値データを用いた．データ領域内で海上となる地点には，現実に存在しない値をダミーデータとして適用させた．本実験に用いたデータを可視化した図を**図 6** に示す．図 6 に示すデータの中から，データ数を 1024（32 × 32），4096（64 × 64），16384（128 × 128），65535（256 × 256）の 4 とおりとなるように領域を指定して実験に用いた．二次元の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．

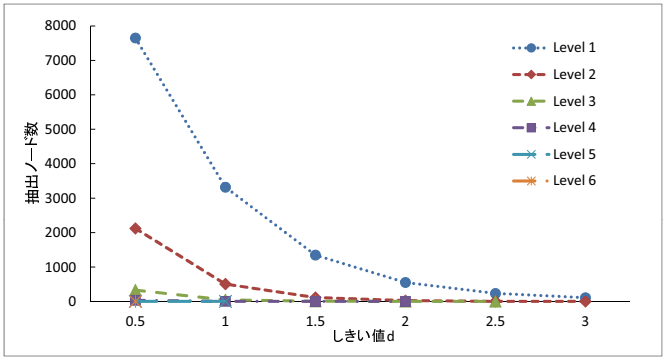

図 8 しきい値 d と各レベルの抽出ノード数
Fig. 8 Threshold d and Extracted Nodes in Each Level.

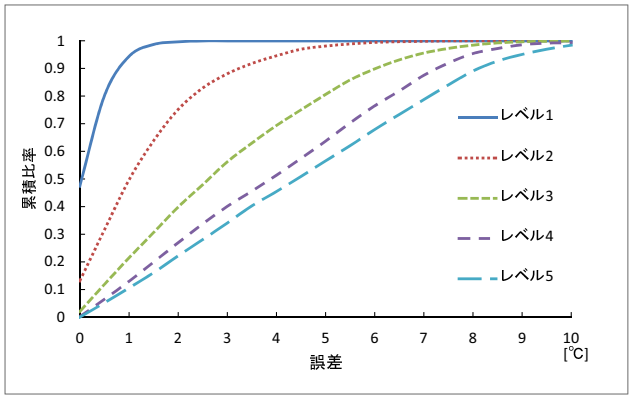

図 9 各レベルにおける補間データの誤差評価
Fig. 9 Error of Interpolated Data in Each Level.

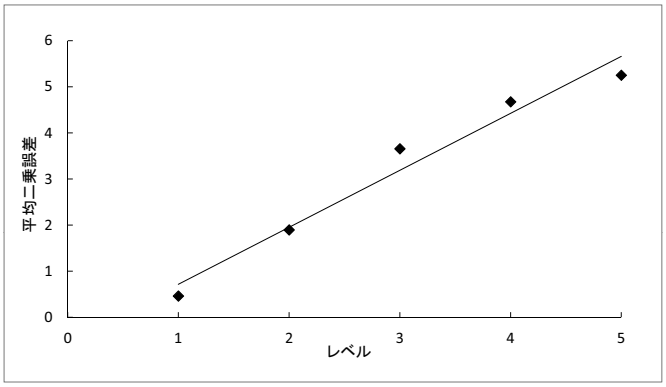

図 10 レベルと平均二乗誤差
Fig. 10 Level and Error of mean square.

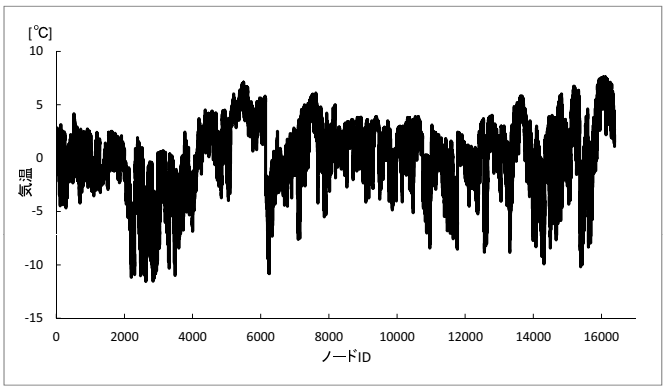

図 11 元データの一次元表現
Fig. 11 One Dimensional Data in Level 0.

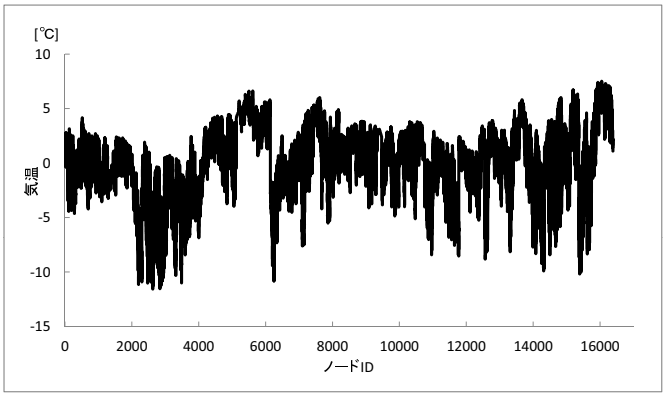

図 12 補間データ Level 1
Fig. 12 Interpolated Data in Level 1.

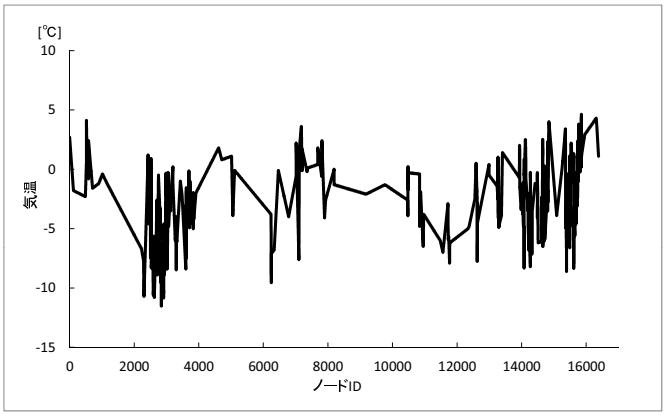

図 13 補間データ Level 3
Fig. 13 Interpolated Data in Level 3.

4.2 範囲検索コストに関する評価

本節では，補間リンクにおける範囲検索コストの評価について述べる．補間リンクにおいて精度と範囲を指定して検索を行うためには，指定されたレベルのノードのリンクを辿って範囲検索を行う．よって，補間リンクにおける範囲検索のコストは，各レベルの抽出ノード数，すなわちネットワークサイズに比例すると考えられる．そこで，補間リンクのネットワークサイズに関して評価実験を行った．

4.2.1 評価方法

1024（32 × 32），4096（64 × 64），16384（128 × 128），65535（256 × 256）の 4 とおりのデータセットに対して，しきい値 d を 0.5 から 3.0 まで 0.5 間隔で増加させ，各レベルでノードを抽出し補間リンクを構築する．構築した補間リンクから，総ノード数 N と各レベルでの抽出されたノード数，およびしきい値 d と各レベルで抽出されたノード数の比較を行う．

4.2.2 評価結果

総ノード数 N に対する各レベルの抽出ノード数を**図 7** に示す．結果から，N に比例して抽出ノード数が増加していることが分かる．

また，しきい値 d による各レベルでの抽出ノード数の変化を**図 8** に示す．図 8 は，$N = 16384$ のデータに対して，d を 0.5 から 3.0 まで 0.5 間隔で増加させた際に各レベルに抽出されるノード数をグラフに示したものである．グラフより，抽出ノード数は d に反比例することが分かる．

これらの結果から，補間リンクの範囲検索コストは N に比例し，d に反比例することが示された．

4.3 補間精度

本節では，補間リンクの各レベルにおける補間精度の評価について述べる．

4.3.1 評価方法

本実験では，総ノード数 $N = 16384$ のデータセットを用いた．しきい値 d を 0.5 から 0.5 間隔で増加させ補間リンクを構築した．補間リンクのレベル 1 からレベル 5 までの各レベルのノードに対して線形補間を行った結果に対し，元データと比較し誤差の評価を行った．

4.3.2 評価結果

レベル 1 からレベル 5 までの各レベルにおける誤差の累積分布を**図 9** に示す．レベル 1 では，ほぼすべてのノードが誤差 1.0 度以内である．レベル 2 からレベル 5 を比較すると，レベルの上昇に伴い，誤差の分布が広がり、高い誤差の計測値の度合いが増加していることが分かる．各レベルでの平均誤差は，レベル 1 で 0.26 度，レベル 2 で 1.36 度，レベル 3 で 2.97 度，レベル 4 で 3.98 度，レベル 5 で 4.51 度となった．平均誤差からも下位レベルの誤差が小さく，上位レベルが大きくなっていることが分かる．

また，各レベルにおける平均二乗誤差の値を**図 10** に示す．各レベルでの平均二乗誤差は，レベル 1 で 0.46 度，レベル 2 で 1.90 度，レベル 3 で 3.65 度，レベル 4 で 4.67 度，レベル 5 で 5.25 度となった．図 10 に示すとおり，レベルの上昇に対して平均二乗誤差は比例して増加している．そのため，平均二乗誤差からも下位レベルの誤差が小さく，上位レベルが大きくなっていることが分かる．

また，補間リンクのデータと元データを比較するため，元データを**図 11** に，補間リンクのレベル 1 と 3 での補間

結果をそれぞれ**図 12** と**図 13** に示す．図 11 の元データと図 12 のレベル 1 を比較すると，元データと違いが無く精密に再現できることが分かる．一方，図 13 のレベル 3 では，精度が低下するものの特徴となる一部の点は抽出でき，元データをおおまかに再現していることが確認できる．

これらの結果から，構築された補間リンクにおいて，精度が必要な場面では下位レベルを用いることで精密な補間が可能である．また，上位レベルを用いた場合，より少ないデータ量で元の計測結果の大まかな再現が可能となる．データの利用者の目的や状況に合わせた適切な検索が可能であることが示された．

5. 関連研究

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 [11] 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．

6. 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モデル駆動型構造化 P2P ネットワークである補間リンクを提案し，補間リンクのネットワークサイズと補間精度に関する評価を行った．提案手法を用いることで，補間に適切なデータを抽出可能であり，範囲検索コストが総ノード数 N に比例し，しきい値 d に反比例することが評価実験によって示された．また，上位レベルから下位レベルに下がるにしたがって補間精度が上昇することが評価実験によって示された．以上から，利用者の目的や状況に合わせた適切な検索が可能であることが示された．今後は，データ更新時における補間リンクモデル更新手法や時系列データを考慮した構造の設計を行っていく．

謝辞 本研究は JSPS 科研費 24700075 の助成を受けたものである．

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] Lay, C., D., Linear Algebra and Its Applications (4th Edition), Addison Wesley (2011).

[6] Stoica, I., Morris, R., Karger, D., Kaashoek, M., Dabek, F. and Balakrishnan, H.: Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications, *Proc. ACM SIGCOMM' 01*, pp.149–160 (2001).

[7] Ratnasamy, S., Francis, P., Handley, M. and Karp, R.: A Scalable Content Addressable Network, *Proc. ACM SIGCOMM' 01*, pp.161–172 (2001).

[8] Maymounkov, P. and Mazieres, D.: Kademlia: A Peer-to-peer Information System Based on the XOR Metric, *Proc. IPTPS' 02* (2002).

[9] Aspnes, J. and Shah, G.: Skip Graphs, *14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp.384–393 (2003).

[10] Pugh, W.: Skip lists: A probabilistic alternative to balanced trees, *Workshop on Algorithms and Data Structures*, pp.437–449 (1989).

[11] 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.

[12] メッシュ気候値 2000, 一般財団法人 気象業務支援センター（オンライン）, 入手先 < http://www.jmbsc.or.jp/index.html >（参照 2012-7-6）.