

推薦論文

# 実世界データの空間連続性に基づく モデル駆動型P2Pネットワーク

斉藤 裕樹<sup>1,a)</sup> 牛久保 辰典<sup>2,†1</sup> 戸辺 義人<sup>3</sup> 鉄谷 信二<sup>4</sup>

受付日 2013年4月16日, 採録日 2013年11月1日

**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, 統計モデルをもと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, 十分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, 国内の気温データの管理・検索に本手法を適用した結果, 検索コストの減少に対して実用的な誤差範囲に収まり, 利用者は階層を指定することにより, トレードオフの関係にある検索コストと誤差から必要な情報を検索することが可能であることが確認された.

**キーワード:** 実世界情報処理, P2P ネットワーク, オーバレイネットワーク, センサネットワーク

## Model-driven Peer-to-peer Network Based on Spatially Continuous Real-world Data

HIROKI SAITO<sup>1,a)</sup> TATSUNORI USHIKUBO<sup>2,†1</sup> YOSHITO TOBE<sup>3</sup> NOBUJI TETSUTANI<sup>4</sup>

Received: April 16, 2013, Accepted: November 1, 2013

**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. 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, the proposed link structure allows us to obtain appropriate and accurate data by specifying hierarchy level.

**Keywords:** real-world data processing, P2P networks, overlay networks, sensor networks.

<sup>1</sup> 明治大学総合数理学部先端メディアサイエンス学科  
Department of Frontier Media Science, School of Interdisciplinary Mathematical Sciences, Meiji University, Nakano, Tokyo 164-8525, Japan

<sup>2</sup> 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, Adachi, Tokyo 120-8551, Japan

<sup>3</sup> 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, School of Science and Technology, Aoyama Gakuin University, Sagami-hara, Kanagawa 252-5258, Japan

<sup>4</sup> 東京電機大学未来科学部情報メディア学科

## 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、データの間接表現を行う新たなレイヤを導入し、統計モデルをもと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。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、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。4 章では、提案手法の評価実験と実験結果の考察を行う。5 章で関連研究について述べ、最後に、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類似し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。たとえば、天気の詳細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、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

Department of Information Systems and Multimedia Design,  
Science and Technology for Future Life, Tokyo Denki Uni-  
versity, Adachi, Tokyo 120-8551, Japan

<sup>†1</sup> 現在、日本ビジネスシステムズ株式会社  
Presently with Japan Business Systems, Inc.

<sup>a)</sup> hrskaito@meiji.ac.jp

本論文の内容は 2012 年 10 月の マルチメディア通信と分散処理  
ワークショップにて報告され、マルチメディア通信と分散処理研  
究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦さ  
れた論文である。

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、データの間中表現を行う新たなレイヤを導入し、統計モデルをもと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。

## 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抽象化し隠蔽する。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うえ、計測値を扱う際に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。

## 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。各格子上的点における値は、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

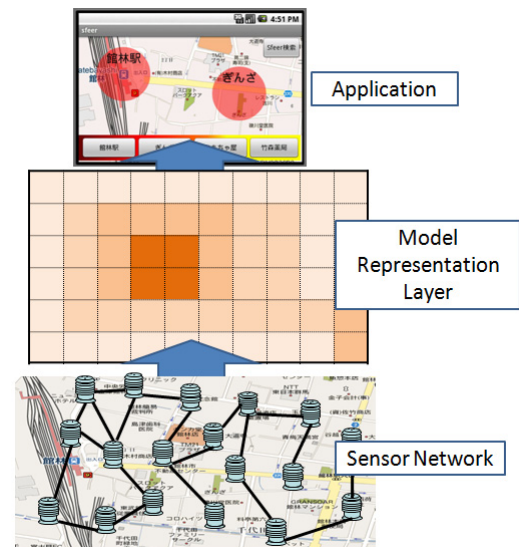


図 1 モデル表現レイヤの構成

Fig. 1 Structure of model representation layer.

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。

$$\begin{aligned} \text{regtemp}(x, y) = & w_1 + w_2x + w_3y + w_4x^2 + w_5y^2 \dots \\ & + w_{n-1}x^m + w_ny^m \end{aligned} \quad (1)$$

各項の重み  $w_1, \dots, w_n$  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  $\text{temp}(x, y)$  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。

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

この最適化問題を解く方法としては、ガウスの消去法 [5] がよく知られている。この結果得られた回帰関数  $\text{regtemp}(x, y)$  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。たとえば、 $x_1 \leq x_2 \leq x_3$  かつ  $y_1 \leq y_2 \leq y_3$  のとき、 $(x_2, y_2)$  の気温を補間するためには、すでに得られている  $(x_1, y_1)$ ,  $(x_1, y_3)$ ,  $(x_3, y_1)$ ,  $(x_3, y_3)$  におけるそれぞれの気温  $\text{temp}(x_1, y_1)$ ,  $\text{temp}(x_1, y_3)$ ,  $\text{temp}(x_3, y_1)$ ,  $\text{temp}(x_3, y_3)$  をもとに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。

$$\text{temp}(x_2, y_2) = (x_3 - x_2)(y_3 - y_2)\text{temp}(x_1, y_1)$$

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

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], Kademlia [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き分散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順序性の保存が必要である。SkipGraph [9] はスキップリスト [10] 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どうして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。

なお、モデル表現レイヤでは、格子上的すべての点における計測値を決定する必要がある。しかし、センサの分布に偏りがある環境では、1つの格子内に複数のセンサがある場合や、センサが存在せず値が欠落する場合が考えられる。これらの場合、以下の方法で各格子の値を決定する。

**1つの格子に対してセンサが複数存在する場合** 格子領域内に複数のセンサが存在する場合、P2P ネットワークに参加する前に格子領域を代表する値を1つ算出する。代表値を得る具体的な方法としては、複数のセンサから得られたデータをもとにグリッドの中心座標

の値を得ることや、平均値、中央値、または最頻値をとることなどが考えられる。これらは、格子に対応するノードが、格子領域内内のセンシングデータを取得することで計算を行う。たとえば、中心座標の値は、2.3 節の補間または回帰分析の手法を用いることで計算できる。

**1つの格子に対してセンサが存在しない場合** 格子領域内にセンサが存在しない場合、欠落した計測値を補う必要がある。そこで、この領域に対応するノードが、P2P ネットワーク構築中に計測値を有する周囲のノードに対して問合せを行うことで値の推定を行う。この場合、回帰分析は適用せず、局所的な情報から補間による計算を行う。具体的な処理手順は、3.3 節で P2P ネットワーク構築アルゴリズムとともに述べる。

#### 3.2 補間リンクの構築

モデル表現レイヤは、格子上的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、領域内が似た値を取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。

補間リンクによる冗長なデータの削減により、アプリケーションが P2P ネットワークからダウンロードするトラフィックを削減することができる。また、アプリケーション内にダウンロードされたデータは、その利用時には2次元のデータ構造に展開される場合もあるが、保存する際にはサイズの小さい状態で扱うことが可能となる。このため、特にデータを蓄積するアプリケーションにおいて使用する記憶域が少なく済む。

ここでは、補間リンクを形成するノードの選出方法を例を用いて説明する。図 2 は、気象庁のメッシュ気候値 2000 [11] の 1 km メッシュの気温データから、関東地方北

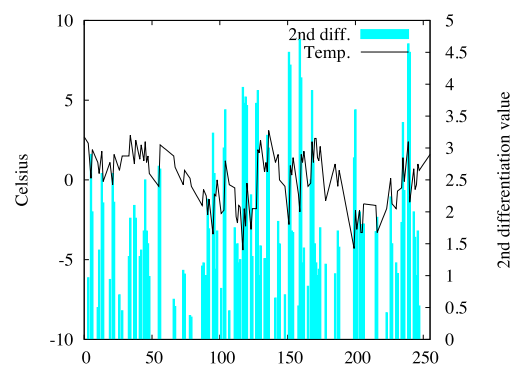


図 2 16 km 四方の気温の一次元化データとその二階微分値  
Fig. 2 One-dimensional data from temperature in 16 km sq. and its second degree differentiation.

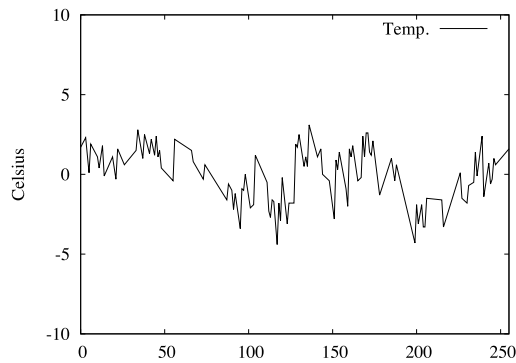


図 3 二階微分値 0.5 以上の点を結んだ近似曲線

Fig. 3 Approximated curve connected with second degree differentiation value of 0.5 or above.

部 16 km 四方の平面の 256 地点の気温を、ルベグ曲線による空間充填法 (Z-Ordering) で一次元に変換したデータおよび、これを連続した曲線と見なした際の二階微分値を图示したものである。

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ほど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、小さいほど傾きが変化しない、すなわち直線的な領域である。これにより、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 0.5 以上の点のみを結んだグラフである。元データの数 256 と比較して、抽出されたデータの数 110 であった。これより、特徴のある一部の点の情報から、元の情報に近似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、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。

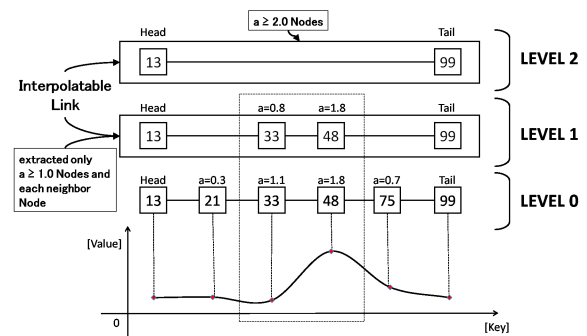


図 4 補間リンクの階層的構造

Fig. 4 Hierarchical structure of interpolatable link.

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させる働きをする。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させ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、ノードどうし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 節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 3 ノードから算出される二次関数  $y_i = a_i x^2 + b_i x + c_i$  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。

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \begin{pmatrix} K_{i-1}^2 & K_{i-1} & 1 \\ K_i^2 & K_i & 1 \\ K_{i+1}^2 & K_{i+1} & 1 \end{pmatrix}^{-1} \begin{pmatrix} V_{i-1} \\ V_i \\ V_{i+1} \end{pmatrix} \quad (4)$$

式 (5) 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、隣接するノードどうし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

**Algorithm: Link Construction****Inputs:**

$N_i$ :  $ID=i$  の新規参加ノード  
 $k$ : 参加レベル  
 $RT_{N_i,k}$ :  $ID=i$  のノードにおけるレベル  $k$  の経路表  
 $\Delta d$ : しきい値の増分

**Functions:**

$sendQuery(Node)$ : 引数に指定されたノードから値を取得する.  
 $makeTable(Routing\ Table, Level)$ : 経路表に指定レベルのエントリを作成する.  
 $ifExist(Node, Level)$ : 指定されたノードに問い合わせ、このノードが指定レベルに含まれているか判定する.  
 $addRoute(Routing\ Table, Direction, Node)$ : 経路表の指定された方向にノードのアドレスを登録する.  
 $sendAddRouteReq(To, Level, Direction, Node)$ : 指定されたノードに対し経路表への追加要求を送信する.

```

1   $k \leftarrow 0, d \leftarrow \Delta d$ 
2  loop
3     $V_{i-1} \leftarrow sendQuery(lookup\ left\ on\ RT_{N_i,k})$ 
4     $V_{i+1} \leftarrow sendQuery(lookup\ right\ on\ RT_{N_i,k})$ 
5     $a_i \leftarrow slope\ with\ N_i, V_{i-1}\ and\ V_{i+1}$ 
6    if  $a_i > d$  then
7       $makeTable(RT_{N_i,k+1})$ 
8       $j \leftarrow i + 1$ 
9      while  $j \leq N_{tail}$  do
10       if  $ifExist(N_j, k + 1)$  then
11          $addRoute(RT_{N_i,k+1}, right, N_j)$ 
12          $sendAddRouteReq(N_j, k + 1, left, N_i)$ 
13         break while
14       end if
15        $j++$ 
16     end while
17      $j \leftarrow i - 1$ 
18     while  $j \geq N_{head}$  do
19       if  $ifExist(N_j, k + 1)$  then
20          $addRoute(RT_{N_i,k+1}, left, N_j)$ 
21          $sendAddRouteReq(N_j, k+1, right, N_i)$ 
22         break while
23       end if
24        $j--$ 
25     end while
26   else
27     break loop
28   end if
29    $d \leftarrow d + \Delta d$ 
30    $k++$ 
31 end loop

```

図 5 補間リンク構築アルゴリズム

**Fig. 5** Interpolatable link construction algorithm.

計測値を要求するクエリを送信する.

- (2) クエリを受け取った  $N_{i-1}$  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  $d$  を超えた場合は  $N_i$  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.

なお, 格子領域内にセンサが存在せず値が欠落しているノードについては, 補間リンク構築時に計測値の推定を行う必要がある. 値の欠落しているノードは, レベル 0 のリンクの構築を終えた段階で, 周囲のノードに問合せを行い,

得られた情報から補間により値を計算する. 周囲のノードへの問合せは, 二次元平面上の最近傍のノードから順に補間に必要な値が得られるまで行うが, 必要な計測値の数は, 適用する補間手法によって異なる. たとえば, 最近傍補間法では, 最も近いノードから得られた値を採用する. また, バイリニア法では, 周囲の 4 点の値を得た後に計算を行う. 以上の手順によりノードの値が決定した後, レベル 1 以上のリンクを構築する.

**3.4 補間リンクの特性**

補間リンクは, 二階微分値の大きい値を有するノードを上位レベルに選出するため, 一定の割合で確率的に上位レベルのノードを選択する Skiplist や Skip Graph とは異なり, データの傾き成分により各レベルに選出されるノード数が異なる. この特性を示すため,  $x$  の値域が  $[0, 2\pi]$  の次の単純なデータセットを考える.

$$f(x) = \sin x \quad (5)$$

このデータセットの二階微分値は以下の式で求められる.

$$\frac{d^2}{dx^2} f(x) = -\sin x \quad (6)$$

上位レベルに選出されるノード数は, 二階微分値の絶対値がしきい値  $d$  以上となる  $x$  の区間長に比例する. これは, 逆三角関数を用いて以下のように求めることができる.

$$length(d) = 2(\pi - 2\sin^{-1} d) \quad (0 \leq d \leq 1) \quad (7)$$

$$length(d) \propto -\sin^{-1} d \quad (0 \leq d \leq 1) \quad (8)$$

よって, 選出ノード数は  $-\sin^{-1} d$  に比例する.  $d = 0, 0.2, 0.4, 0.6, 0.8, 1.0$  と一定割合で増加させたときの  $(\pi - 2\sin^{-1} d)/\pi$  の値はそれぞれ約 1.00, 0.87, 0.74, 0.59, 0.40, 0.0 となり, レベルが上がるごとに選出されるノード数が減少することが分かる. なお, 任意のデータについても, フーリエ級数によって三角関数の項の和で表現されるデータは, 同様にして選出ノード数が逆三角関数の項の和に比例することがいえる.

以上から, 補間リンクは Skiplist や Skip Graph と同様に上位レベルのリンクが短縮経路となり, 検索の際のホップ数を低く抑えられることが分かる. なお, 補間リンクの範囲検索時の精度とホップ数については, 4.4 節で詳細な評価を行う.

**4. 評価実験**

本章では, 補間リンクのグラフ特性, 範囲検索コストおよび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 [11] 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

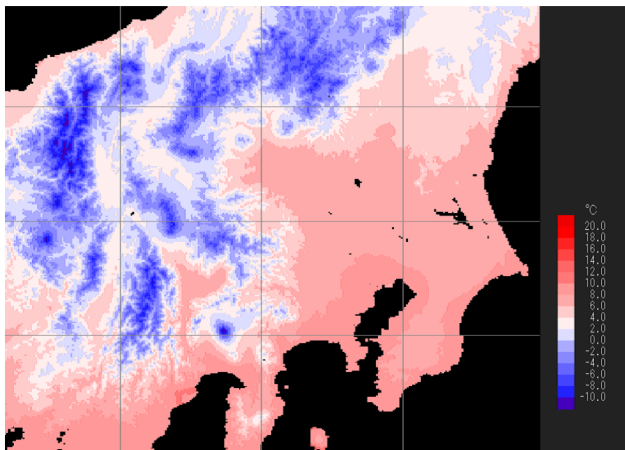


図 6 評価に用いたデータセット

Fig. 6 Dataset for evaluation.

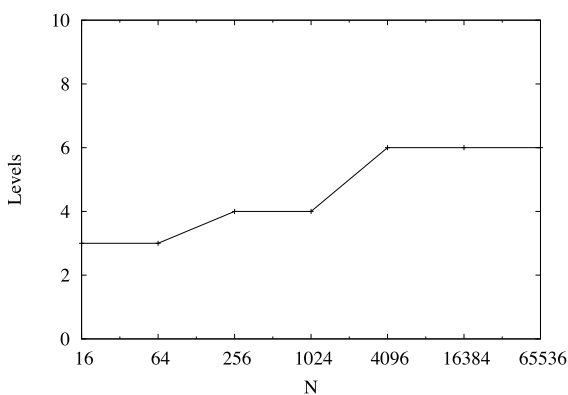


図 7 ノード総数  $N$  と階層数

Fig. 7  $N$  and levels.

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や湖上となる地点には、近傍地点の値を適用させた。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 64 ( $8 \times 8$ ), 256 ( $16 \times 16$ ), 1024 ( $32 \times 32$ ), 4096 ( $64 \times 64$ ), 16384 ( $128 \times 128$ ), 65535 ( $256 \times 256$ ) の 6 通りとなるように領域を指定し、1 データに対して 1 ノードを割り当て実験を行った。また、二次元の位置情報は空間充填曲線 Z-ordering を用いて一次元情報へと変換した。

#### 4.2 グラフ特性の評価

本節では、ノード総数  $N$  に対する補間リンクの諸特性について評価を行う（以下、ノード総数は  $N$  と表記する）。リンクを階層化する際のしきい値の増分  $\Delta d$  を 0.5（以下、しきい値を  $d$ 、しきい値の増分を  $\Delta d$  と表記する）に設定した際のノード数ごとに階層化した後の階層数を図 7 に示す。各レベルにおいて  $d$  以上の値を有するノード間でリンクが形成されるため、レベルが 1 上がり  $d$  が  $\Delta d$  増加するごとに存在するノードが減少する。したがって、階層数は、

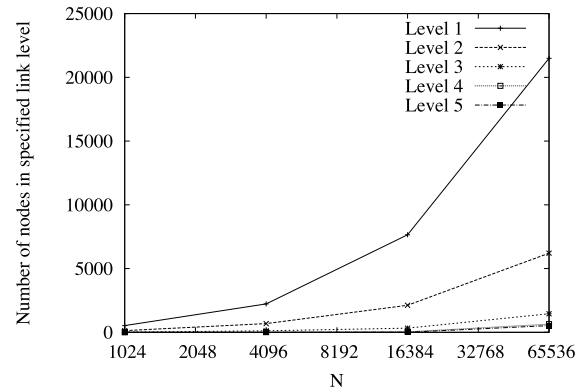


図 8  $N$  と各レベルの抽出ノード数

Fig. 8  $N$  and extracted nodes in each level.

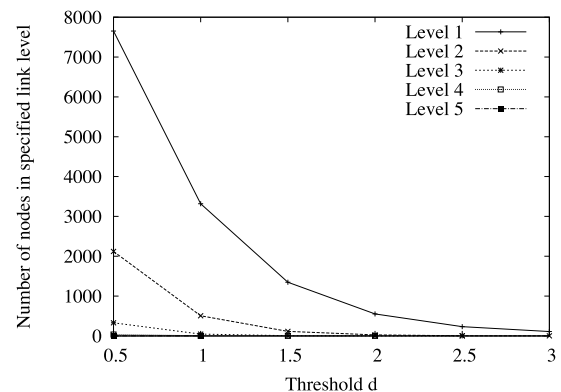


図 9  $\Delta d$  と各レベルの抽出ノード数

Fig. 9  $\Delta d$  and extracted nodes in each level.

$N$  の指数的な増加に対して一定の増加に抑えられている。

次に、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関して評価を行う。データセットに対して、 $\Delta d = 0.5$  に設定した際の  $N$  に対する抽出ノード数、および  $N = 16384$  のデータに対して  $\Delta d$  を 0.5 から 3.0 まで 0.5 間隔で増加させた際のレベルに対する抽出ノード数の比較を行う。 $\Delta d = 0.5$  における  $N$  に対する各レベルの抽出ノード数を図 8 に、 $N = 16384$  のとき  $\Delta d$  を変化した際の各レベルでの抽出ノード数を図 9 にそれぞれ示す。結果から、抽出ノード数は  $N$  に比例して増加し、 $d$  に反比例して減少することが分かる。

#### 4.3 補間リンクの誤差

本節では、補間リンクのデータの誤差の評価について述べる。本実験では、 $N = 16384$  のデータセットを用い、 $\Delta d = 0.5$  に設定した。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。

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 10 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、レベルの上昇にともない、誤差の分布が広がり、

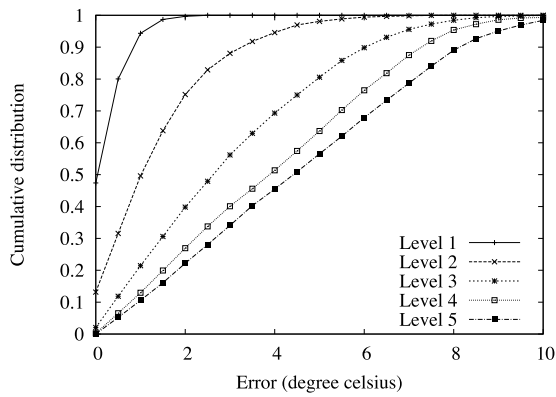


図 10 各レベルにおける補間データの誤差

Fig. 10 Error of interpolated data in each level.

高い誤差の計測値の割合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。

#### 4.4 利用時の効果

補間リンクでは、検索精度と検索コストはトレードオフの関係にある。そこで、利用時の性能評価のため、任意の 2 ノード間の範囲検索を行いホップ数と誤差の評価を行う。

範囲検索では、検索範囲の始点ノードに到達した後、階層の指定により検索精度を指定し、始点ノードから指定された精度に対応する階層レベルを上限に補間リンクをたどり検索範囲の終点のノードに到達するため、全体として検索コストは、始点ノードに到達するコストと検索範囲内の補間リンクをたどるコストの和となる。始点ノードに到達するまでのコストは、SkipGraph と同様の確率的選択に基づく階層化短縮経路を本 P2P ネットワークに導入することにより、 $O(\log N)$  で実現できることが知られている。そこで、本実験では、検索範囲をたどるホップ数を計測するため、補間リンクにおいて任意の 2 ノードを選択し、レベルを指定して一方のノードからもう一方のノードまでリンクをたどった際のホップ数の計測を行った。

$\Delta d = 0.5$  とし  $N$  を変化させ、検索精度（階層レベル）を指定しランダムな 2 ノードの組合せで計測を 1000 回行ったホップ数の平均を図 11 に示す。レベル 1～3 では、ホップ数が指数的に減少しているが、レベル 4 以上の低い精度の検索ではホップ数が収束することが分かる。これは、階層化時にレベルの増加とともにリンクを構成するノードが減少していくが、高いレベルにおいては  $d$  が増加するのみで選出ノードが変化しなかったためと考えられる。

また、 $N = 65536$  での各レベルのレベル 0（元データ）に対するホップ数の割合と、レベル 5 に対する各レベルの平均二乗誤差の割合を図 12 に示す。レベルの増加に反比例してホップ数が減少する様子と、レベルの増加に比例して

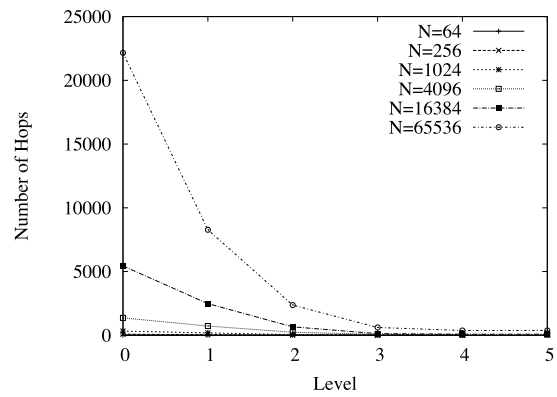


図 11 精度（レベル）ごとのホップ数

Fig. 11 Number of hops.

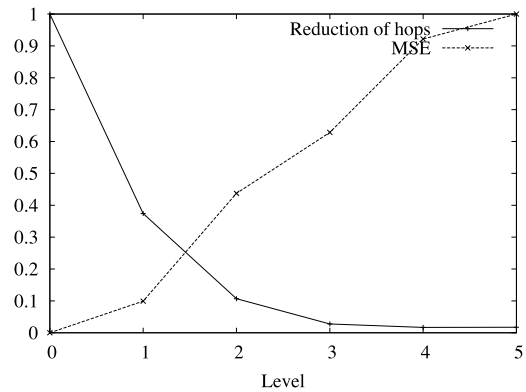


図 12 ホップ数と平均二乗誤差の比較

Fig. 12 Comparison number of hops with MSE.

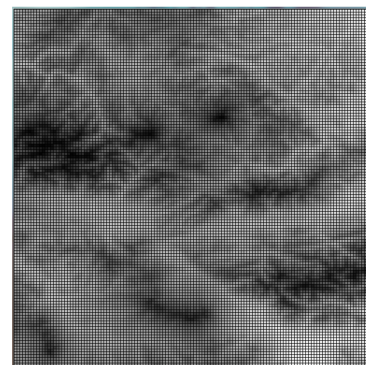


図 13 元データ（Level 0）

Fig. 13 Original data in level 0.

平均二乗誤差が増加している様子が分かる。レベル 1 ではホップ数の割合が約 0.374 であり平均二乗誤差が約 0.533 度、レベル 2 ではホップ数の割合が約 0.107 であり平均二乗誤差が約 2.35 度であった。このように、検索コストの減少に対して実用的な誤差範囲に収まっていることが分かる。以上から、利用者は階層を指定することにより、トレードオフの関係にある検索コストと誤差から、必要な情報を取得することが可能であると考えられる。

補間リンクのデータと元データを視覚的に比較するため、 $N = 65536$  の元データを図 13 に、 $\Delta d = 0.5$  とした

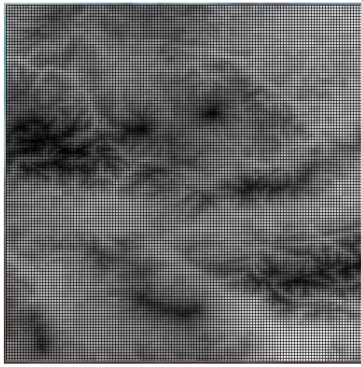


図 14 補間データ Level 1

Fig. 14 Interpolated data in level 1.

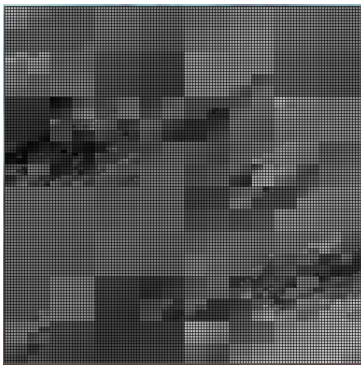


図 15 補間データ Level 3

Fig. 15 Interpolated data in level 3.

補間リンクのレベル 1 と 3 での補間結果をそれぞれ図 14 と図 15 に示す。図 13 の元データと図 14 のレベル 1 を比較すると、元データをおおまかに再現できることが分かる。一方、図 15 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、元データの特徴を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。よって、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と考えられる。

## 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 Ahuja らの研究 [2] がある。Ahuja らは 6 PB 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 [12] 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従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。

今後の課題として、動的計測への対応があげられる。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、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検討する必要がある。補間リンクのトポロジの更新には、センサの計測値に変更があった時点で補間リンクを更新する方法と、利用者からの検索があった時点で検索領域の補間リンクを更新する方法があり、計測値の更新頻度と検索頻度によって適切な方法を選択する必要がある。今後は、運用時におけるモデル表現レイヤの変更コストに関する評価を行う。また、本研究では空間充填曲線により 2 次元空間を 1 次元のデータ構造に変換し P2P ネットワークの構造化を行っているが、ドロネー構造をもとに 2 次元空間をトポロジとした P2P ネットワーク [13] も存在する。しかし、ドロネー構造は動的計測に対応する場合にトポロジの変更が高コストであることが知られており [14]、今後の評価によりトポロジ表現と変更コストの特性を明らかにする必要があると考えられる。また、さらに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 科研費若手研究 (B) 24700075 の助成を受けたものである。

## 参考文献

- [1] Kallman, R., Kimura, H., Natkins, J., Pavlo, A., Rasin, A., Zdonik, S., Jones, E.P.C., Madden, S., Stonebraker, M., Zhang, Y., Hugg, J. and Abadi, D.J.: H-store: A High-performance, Distributed Main Memory Transaction Processing System, *Proc. VLDB Endow.*, Vol.1, No.2, pp.1496–1499 (2008).
- [2] Ahuja, M., Chen, C., Gottapu, R., Hallmann J., Hasan, W., Johnson, R., Kozyczak, M., Pabbati, R., Pandit, N., Pokuri, S. and Uppala, K.: Peta-Scale Data

- Warehousing at Yahoo!, *Proc. 35th SIGMOD International Conference on Management of Data*, pp.855–861 (2009).
- [3] Thiagarajan, A. and Madden, S.: Querying Continuous Functions in a Database System, *Proc. ACM SIGMOD International Conference on Management of Data*, pp.791–804 (2008).
- [4] Deshpande, A. and Madden, S.: MauveDB: Supporting Model-based User Views in Database Systems, *Proc. ACM SIGMOD International Conference on Management of Data*, pp.73–84 (2006).
- [5] Lay, D.C.: *Linear Algebra and Its Applications 4th Edition*, Pearson Education (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 Mazières, D.: Kademlia: A Peer-to-peer Information System Based on the XOR Metric, *Proc. International workshop on Peer-To-Peer Systems (IPTPS 2002)* (2002).
- [9] Aspnes, J. and Shah, G.: Skip Graphs, *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2003)*, pp.384–393 (2003).
- [10] Pugh, W.: Skip Lists: A Probabilistic Alternative to Balanced Trees, *Proc. Workshop on Algorithms and Data Structures*, pp.437–449 (1989).
- [11] 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.jmbc.or.jp/index.html>) (参照 2013-11-5).
- [12] Desnoyers, P., Ganesan, D. and Shenoy, P.: TSAR: A Two Tier Sensor Storage Architecture Using Interval skip Graphs, *Proc. 3rd ACM Conference on Embedded Networked Sensor Systems (Sensys 2005)*, pp.39–50 (2005).
- [13] 小西佑治, 寺西裕一, 竹内 亨, 春本 要, 下條真司, 西尾章治郎: センサ観測値分布の概要把握を可能とする階層化ドロネーオーバーレイネットワーク構築手法, *情報処理学会論文誌*, Vol.51, No.2, pp.624–634 (2010).
- [14] Tagami, A., Ano, S. and Tomiura, Y.: Simulation Analysis of Moving Peer Influence on Location-Aware P2P Network, *Proc. 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA 2010)*, pp.1121–1127 (2010).

## 推薦文

本論文では、センサネットワークによる空間連続データの分散測定を行う際に、データの間中表現を行うモデル表現レイヤを構築し、統計モデルをもとにした空間連続データの管理を実現することで、指定された精度を保ちつつ、伝送データを削減するモデル駆動型 P2P ネットワークを提案している。提案手法は、統計モデルを導入したセンサデータの処理を行うところに特徴があり、センサネットワークにおけるデータ処理技術の向上に貢献する提案である。以上より本論文は推薦に値する。

(マルチメディア通信と分散処理研究会主査 勝本道哲)



齊藤 裕樹 (正会員)

2001 年明治大学大学院理工学研究科基礎理工学専攻博士課程修了。博士(工学)。明治大学、東京電機大学を経て、2013 年より明治大学総合数理学部准教授。オーバーレイネットワーク、センシングとネットワークのソフトウェアの研究に従事。日本ソフトウェア科学会、IEEE、ACM 各会員。



牛久保 辰典

1988 年生。2011 年東京電機大学未来科学部情報メディア学科卒業。2013 年同大学大学院未来科学研究科情報メディア学専攻修士課程修了。同年より日本ビジネスシステムズ株式会社勤務、現在に至る。実世界の空間情報の利活用とデータベースに興味を持つ。



戸辺 義人 (正会員)

東芝、慶應義塾大学、東京電機大学を経て、2012 年青山学院大学理工学部情報テクノロジー学科教授。センサネットワーク、参加型センシングの研究を進めている。IEEE、電子情報通信学会、計測自動制御学会、人間情報学会各会員。



鉄谷 信二

1980 年北海道大学大学院修士課程修了。同年電電公社(現、NTT)入社以来、ファクシミリにおける画像信号処理、電子写真記録および立体表示技術等の研究・実用化に従事。2000 年 ATR 国際電気通信基礎技術研究所に出向、環境生成技術に関する研究、五感メディアの研究に従事。2004 年 4 月東京電機大学理工学部教授、2008 年 4 月同大学未来科学部教授、マルチメディアに関する研究に従事、現在に至る。工学博士。