

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 20 日現在

機関番号：32682

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24700075

研究課題名(和文)位置情報に最適化された空間構造化オーバーレイネットワーク

研究課題名(英文)Space Structured Overlay Network for Spatial Data

研究代表者

斉藤 裕樹 (SAITO, HIROKI)

明治大学・総合数理学部・准教授

研究者番号：70333127

交付決定額(研究期間全体)：(直接経費) 1,600,000円

研究成果の概要(和文)：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の利用分野が拡大している。しかし、分散計測環境では取得されるデータは不完全で不均一な情報であるため、有効に利用することが難しい。本研究では、物理的なセンシングレイヤを抽象化しデータの間接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管理機構および、空間構造化P2Pネットワーク上でのモデル表現レイヤ分散管理手法を検討した。国内の気温データの管理・検索に本手法を適用し、検索コストの減少に対して実用的な誤差範囲に収まり、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を確認した。

研究成果の概要(英文)：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 tends to be incomplete and imprecise, it is not suitable to present it to users or applications. In this work, 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.

研究分野：計算機システム・ネットワーク

キーワード：位置情報サービス オーバレイネットワーク 実世界情報処理 P2Pネットワーク

1. 研究開始当初の背景

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。

実世界データの情報処理には以上のような課題の解決が望まれている。

2. 研究の目的

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、センサの故障や停止、転送経路上でのエラーなどにより、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、利用することが難しい。一方、計測されるデータの質の面では、広範囲検索におけるデータ冗長性と空間連続データの膨大さが問題となる。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。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を実装することが求められてきた。

本研究の目的は、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、統計モデルを元にした空間連続データを管理できる機構を検討することである。このレイヤは、膨大なデータを扱うために構造化 P2P ネットワークモデルを用いた分散アーキテクチャとして検討を行った。

3. 研究の方法

(1) 実世界データの構造化の基本となる統計モデルに基づくモデル表現レイヤによるデータの抽象化について検討し、モデル表現レイヤの構造を定式化した。

(2) モデル表現レイヤのデータ構造に基づく分散コンピューティングモデルを検討した。本研究では、P2P ネットワークモデルによる階層型短縮経路を持つネットワークを構造化した。

(3) P2P 型分散システムとしてネットワークを構築するために必要な諸アルゴリズム、利用時の範囲検索・更新アルゴリズムを検討した。

(4) 実データを用いネットワークの特性評価を行い有効性を確認した。

4. 研究成果

(1) モデル表現レイヤによるデータの抽象化

空間連続データを統計モデルで扱うモデル表現レイヤを考案した。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。

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

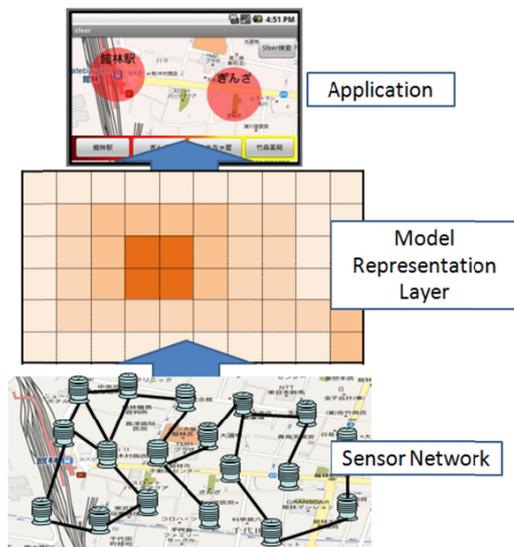


図1 モデル表現レイヤの構造

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。各格子上の点における値は、生の計測値から回帰分析または補間により推定が可能である。

(2) P2P ネットワークモデルによるモデル表現レイヤの実現

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 ネットワークである分散ハッシュ表 (DHT) 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が、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 はスキップリスト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 ID 空間に変換する手法とした。SkipGraph のように ID ご

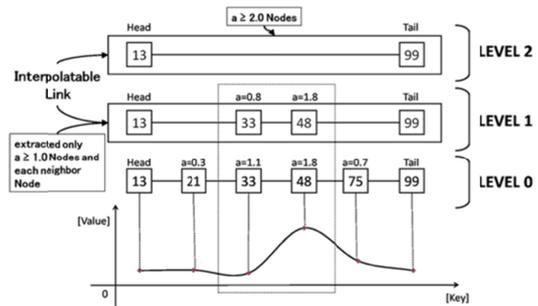


図2 補間リンクの構造

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リスト構造を構築していくものである。

(3) P2P リンク構築アルゴリズム

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。

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ことで、冗長な値を取り除くことができる。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

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
 Δd : 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   $k \leftarrow 0, d = \Delta d$ 
2  loop
3     $V_{i-1} \leftarrow \text{sendQuery}(\text{lookup left on } RT_{N_i,k})$ 
4     $V_{i+1} \leftarrow \text{sendQuery}(\text{lookup right on } RT_{N_i,k})$ 
5     $a_i \leftarrow \text{slope with } N_i, V_{i-1} \text{ and } V_{i+1}$ 
6    if  $a_i > d$  then
7      makeTable( $RT_{N_i,k+1}$ )
8       $j \leftarrow i + 1$ 
9      while  $j \leq N_{\text{tail}}$  do
10       if ifExist( $N_j, k + 1$ ) then
11         addRoute( $RT_{N_i,k+1}, \text{right}, N_j$ )
12         sendAddRouteReq( $N_j, k + 1, \text{left}, N_i$ )
13       break while
14     end if
15      $j++$ 
16   end while
17    $j \leftarrow i - 1$ 
18   while  $j \geq N_{\text{head}}$  do
19     if ifExist( $N_j, k + 1$ ) then
20       addRoute( $RT_{N_i,k+1}, \text{left}, N_j$ )
21       sendAddRouteReq( $N_j, k + 1, \text{right}, N_i$ )
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3 P2P リンク構築アルゴリズム

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2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．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．

次に、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、前述の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

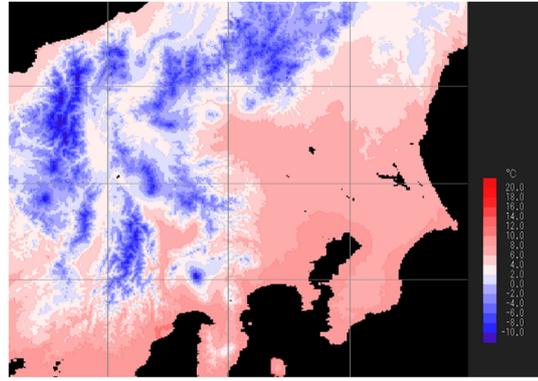


図4 実験データセット

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．

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

上式により算出された 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3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．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

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を行った．評価には、一般財団法人 気象業務支援センターより提供されている気象庁のメッシュ気候値 2000 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．本実験に用いたデータを可視化した図を図4に示す．ここに示すデータの中から、データ数を64 (8×8), 256 (16×16), 1024 (32×32), 4096 (64×64), 16384 (128×128), 65536 (256×256) の6とお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．

まず、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った (以下、ノード総数は N と表記する)．リンクを階層化する

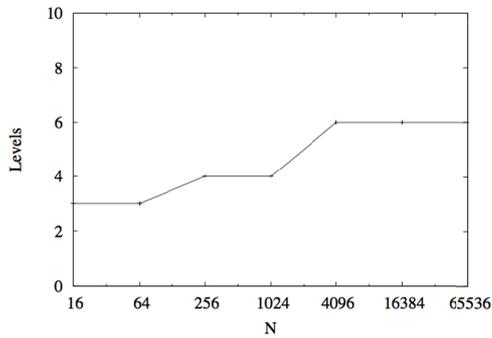


図5 Nと階層数

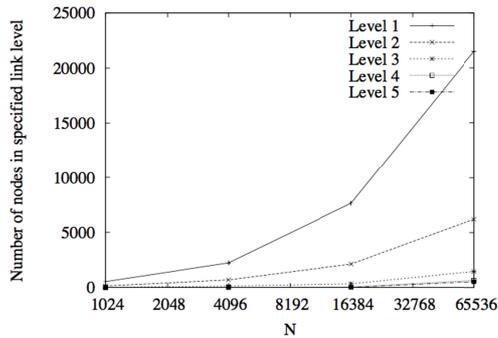


図6 Nと各レベルの抽出ノード数

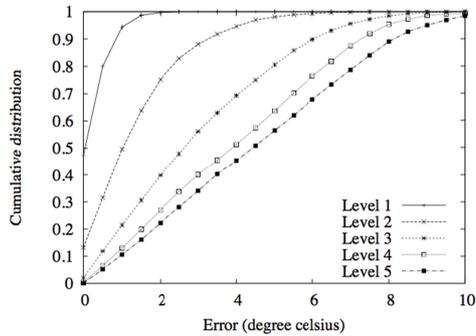


図7 各レベルにおける補間データの誤差

際のしきい値の増分を d (以下、しきい値を d , しきい値の増分を 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5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 d 増加することによって存在するノードが減少する。したがって、階層数は、 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。また、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、 $d=0.5$ 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 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。 $d=0.5$ における N に対する各レベルの抽出ノード数を図6に示す。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。

次に、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

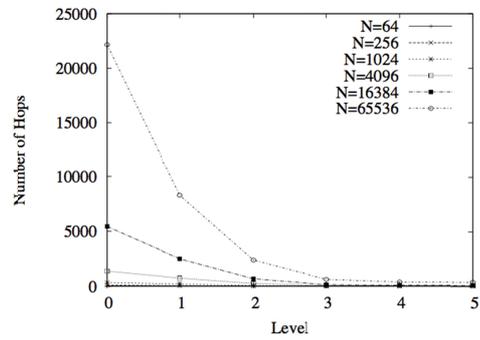


図8 精度(レベル)ごとのホップ数

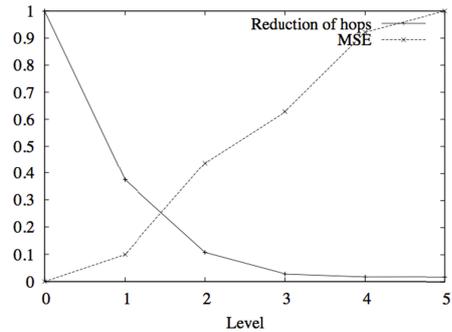


図9 ホップ数と平均二乗誤差の比較

タセットを用い $d=0.5$ 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7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。

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った。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辿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辿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、検索範囲を辿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辿った際のホップ数の計測を行った。ランダムな2ノードの組み合わせで計測を1000回行ったホップ数の平均を図8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。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9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確認された。

5. 主な発表論文等

〔雑誌論文〕(計5件)

(1) Yuki Yagi, Naofumi Kitsunezaki, Hiroki Saito, Yoshito Tobe, RWFS: Design and Implementation of File System Executing Access Control Based on User's Location, Proceedings of International Workshop of Data Delivery and Security Techniques in Wireless and Mobile Networks (ADST-WMN 2014), 査読有, 2014, pp. 805-810.

(2) Hiroki Saito, Tatsunori Ushikubo, Multi Precise Interpolation-available KVS for Geographical Data Management, Proceedings of ECDM 2014, 査読有, 2014, pp. 1-4.

(3) Hiroki Saito, Tsubasa Takayama, Learning from Location History and Context for Collaborative People Behavioral Prediction, Proceedings of ECDM 2013, 査読有, 2013, pp.1-4.

(4) 斉藤 裕樹, 牛久保 辰典, 戸辺 義人, 鉄谷 信二, 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, 情報処理学会論文誌, 査読有, Vol. 55, No. 2, 2013, pp. 1079-1088.

(5) Hiroki Saito, Kazuki Yamada, Collecting Sharing and Inferring User Behavior for Mobile Recommendation with Location History and Its Context, Proceedings of ECDM 2012, 査読有, 2012, pp.221-224.

〔学会発表〕(計4件)

(1) 八木 佑侑季, 狐崎 直文, 斉藤 裕樹, 戸辺 義人, RWFS: 位置情報による空間アクセス制御を行うファイルシステムの設計と実装, 情報処理学会 マルチメディア, 分散, 協調とモバイル (DICOMO) シンポジウム 2014, pp. 1158-1165, 2014年7月, 新潟県月岡市.

(2) 伊藤 翼, 八木 佑侑季, 斉藤 裕樹, 狐崎 直文, 戸辺 義人, 近距離無線電波強度とグループ行動知識モデルを用いた測位手法の提案, 情報処理学会マルチメディア通信, 分散, 協調とモバイルシンポジウム (DICOMO) 2013, pp. 204-207, 2013年7月, 北海道河東郡音更町.

(3) 高山 翼, 山上 慶, 斉藤 裕樹, 戸辺 義人, 鉄谷 信二, マイクロブログのコンテンツを用いた行動予測への確率過程モデルの適用と評価, 情報処理学会マルチメディア通信と分散処理研究会報告, 2013-DPS-154(24), 2013年3月, 東京都足立区.

(4) 牛久保 辰典, 斉藤 裕樹, 戸辺 義人, 鉄谷 信二, 空間連続データに適したモデル駆動型構造化P2Pネットワーク, 情報処理学会マルチメディア通信と分散処理ワークショップ (DPSWS) 2012, pp.65-72, 2012年12月, 愛媛県松山市.

6. 研究組織

(1) 研究代表者

斉藤 裕樹 (SAITO HIROKI)

明治大学・総合数理学部・准教授

研究者番号: 70333127