

分散型移動オブジェクトデータベースにおける 予測型インデクスの性能評価

勝田 圭介[†] 柳沢 豊^{††} 佐藤 哲司^{†,††}

[†] 大阪大学 大学院情報科学研究科 〒 565-0871 吹田市山田丘 1-5

^{††} 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所 〒 619-0237 京都府相楽郡精華町光台 2-4

E-mail: †k-katuda@ist.osaka-u.ac.jp, ††yutaka@cslab.kecl.ntt.co.jp, †††satohtetsuji@lab.ntt.co.jp

あらまし 近年, GPS のコンパクト化, 精度向上などにより, 移動物体の位置を用いた様々なアプリケーションが開発され, 時空間インデクスを用いて移動物体の位置を高速に検索できる移動オブジェクトデータベースが注目を浴びている. しかし, センサネットワークの大規模化により, サーバで扱うデータ量が増大しサーバの負荷が高まっている. このサーバの負荷を軽減するために, 筆者らは分散型移動オブジェクトデータベースを開発している. また, データの分散化に伴うクエリ遅延問題に対し, 予測型インデクスを導入することでこの問題を解決できることを提案してきた. 本稿では, この予測型インデクスを構築するための物体の移動予測技術を新たに提案し, それぞれの有効性について評価する. 評価実験により, 予測型インデクスを導入して物体の移動を予測することでデータの分散化に伴う問題点を解消し, かつサーバの負荷を十分に軽減できることを示した. また, 物体の移動速度の分散値が移動予測の精度に大きな影響を与えることを示した.

キーワード 分散型データベース, 予測, 時空間データベース, センサネットワーク, 移動オブジェクト

Evaluation of Predictive Indexing in Distributed Moving Objects Databases

Keisuke KATSUDA[†], Yutaka YANAGISAWA^{††}, and Tetsuji SATOH^{†,††}

[†] Graduate School of Information Science and Technology, Osaka University
Yamadaoka 1-5, Suita-shi, Osaka, 565-0871 Japan

^{††} NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
Hikaridai 2-4, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan

E-mail: †k-katuda@ist.osaka-u.ac.jp, ††yutaka@cslab.kecl.ntt.co.jp, †††satohtetsuji@lab.ntt.co.jp

Abstract Recently, there are many location information services using the moving objects such as walking people, moving cars, and so on. To manage such moving objects efficiently, the moving objects database have been proposed. However, there are too much positioning data of moving objects for one database server to manage all of the moving objects. To manage such many objects efficiently, we are developing the distributed moving objects database which consists of several database systems. Moreover, we proposed a predictive indexing method to retrieve moving objects from distributed database systems. In this paper, we present several methods to predict moving objects for predictive index, and we describe the evaluation of these moving predictive techniques comparing with traditional predictive techniques using several types of moving objects data.

Key words distributed database, prediction, spatio-temporal database, sensor network, moving object

1. はじめに

近年, 携帯電話や車に GPS を取り付けてナビゲーションを自動化するなど, 物体の位置情報を利用したサービスが急速に普及している. また, 単に位置情報を利用だけでなく, 位

置情報と時刻の組をまとめて列にした移動軌跡データを大量に収集・分析する研究が盛んである. 例えば, 人の流れをパターン化してオフィス構造を最適化したり, 車の移動軌跡を分析して道路整備計画を立てる [1] といった過去の物体の位置情報を分析し, それを設計に利用するフィードバック研究がある. 他に

も、交通渋滞の予測やタクシーの配車管理 [2]、店頭広告自動切替など、過去、現在の位置情報から未来の位置情報を予測し、それをリアルタイムに利用する研究などもある。

こうした研究では、センサから取得したデータを一ヶ所に集めて集中管理することを前提としていた [3]~[5]。しかし、近年のセンサの廉価化、小型化にともない、多数のセンサを広範囲に配置することが可能になってきた。これにより、全てのデータを一元管理することが困難になり、データを分散管理する必要性が高まっている [6], [7]。そこで筆者らはこれまでに、分散型移動オブジェクトデータベースの提案を行っている [8]~[10]。

分散型移動オブジェクトデータベースは、位置情報をセンシングする多数のセンサと、それらセンサを管理するサーバにより構成される。各センサは、自らがセンシングしたデータを管理し、データを一定の間隔で間引きながらサーバへ送信する。サーバは各センサから受信したデータを用いて、どのセンサがどのようなデータを保持しているかを管理するための *Spatio-temporal index* (時空間インデクス^(注1)) を構築する。サーバはこのインデクスを用いることで、各センサが保持するデータを効率的に検索できる。このようにデータを分散化することにより、サーバの負荷軽減、センサの省電力などを実現することができる [11], [12]。

しかし、データベースを分散化することにより、サーバは各センサが保持するデータをリアルタイムに把握することができなくなり、クエリの即応性が失われるという問題が生じる。これは、各センサがデータを取得してから、サーバ側でそのデータに対応するインデクスが構築されるまでに遅延が生じるからである。これを解決するために、サーバが各センサからデータを受信する際に、それ以前のデータに対するインデクスを構築するだけでなく、それ以降のデータに対しても予測してインデクス(予測型インデクス)を構築する手法を提案してきた [8]~[10]。この予測型インデクスを構築するためには、未来の物体の移動を予測する必要があるが、筆者らはこれまでに過去の物体の移動速度 [8]、移動履歴 [9], [10] などを用いた物体の移動予測技術を提案してきた。これらの予測技術は共に人力車の移動軌跡データに対しては、有効な予測技術であった。しかし、それ以外の物体のデータに対するシミュレーション実験を行っていなかったため、予測型インデクスの一般的な有効性について示すことができなかった。

そこで本稿では、人力車以外の人、モーションキャプチャなどのデータについて移動予測を行い、予測型インデクスが一般的に有効である事を示すための実験を行う。また、物体の移動に規則性がある場合に有効となる移動パターンに着目した予測技術を新たに提案する。物体の移動の様子は類似したものが多く、いくつかのパターンをもつものが多い。従って、過去の移動パターンから未来の移動パターンを予測できると考えられる。本稿では、確率モデルであるマルコフモデル [13], [14]、最長部分列を求めるアルゴリズムの LCSS [15], [16] を応用した移動予測技術を提案する。実験結果より、移動軌跡データの性質ごと

に最適な手法が異なることを示し、以上の移動予測技術を用いた予測型インデクスにより、十分にサーバの負荷を軽減させられることを示す。

2. 分散型移動オブジェクトデータベース

本章では、筆者らが提案してきた分散型移動オブジェクトデータベースの詳細を述べる。2.1 でその概要を、2.2 でデータの分散化にともなう問題点とそれに対処するアプローチについて、2.3 でアプローチである予測型インデクスの概要について述べる。

2.1 概要

移動オブジェクトデータベースとは、各センサが取得した物体の移動軌跡データを全てサーバに送信し、サーバが受信した全ての物体の移動軌跡データを管理するデータベースのことである。このデータベースにユーザが、時間、位置、物体の ID などの条件を指定した *window query*^(注2) を送ると、サーバが保持する全てのデータを用いて、それらの条件を満たすようなデータを検索する。このようなデータベースでは、サーバが全ての物体のデータを管理する必要がある。しかし、近年のセンサネットワークの大規模化によりセンサデータが増大し、サーバで管理するデータ量が莫大になりサーバの負荷が高まってきた。今後、センサネットワークが拡大するにつれて、サーバにより一元的にセンサデータを管理する方法はより困難になる。

このサーバの負荷集中を回避するために、筆者らは移動軌跡データを分散して管理する分散型移動オブジェクトデータベースを提案してきた。分散型移動オブジェクトデータベースでは、データは分散配置された各センサのローカルなデータベースが保持し、ローカルなデータベースはデータの一部のみをサーバに送信する。サーバは受信したデータを用い、各センサのもつローカルなデータに対するインデクスを構築する。サーバはユーザからのクエリを受け取ると、まずインデクスを用いて検索すべきデータを保持するセンサを特定し、次に特定したセンサに対しクエリを転送する。クエリが転送されたセンサでは、各センサで蓄積しているローカルなデータに対してクエリの処理を行い、答となるデータを検索する。そして最終的にユーザは、サーバを通して検索結果を得ることができる。

以下、図 1 を用いて、センサノード A, B, C とサーバからなる分散型移動オブジェクトデータベースの動作について説明する。

まず、サーバがインデクスを構築する様子を図 1(a) に例示する。それぞれのセンサは時刻 $t = 0, 1, 2, \dots, \tau$ における物体の位置を取得し、それぞれのローカルなデータベースに格納する。さらに、各センサは $0 \leq t \leq \tau$ の間にそれぞれの物体が移動した範囲を求め、サーバに送信する。サーバは、それぞれのセンサから受信した移動範囲 (R_A, R_B, R_C) を用いて、物体、時間、移動範囲を組にしたインデクス ($I_A = \{A, 0 \leq t \leq \tau, R_A\}, I_B = \{B, 0 \leq t \leq \tau, R_B\}, I_C = \{C, 0 \leq t \leq \tau, R_C\}$) を構築する。

(注1): 以下、インデクスと記載

(注2): 以下、クエリと記載

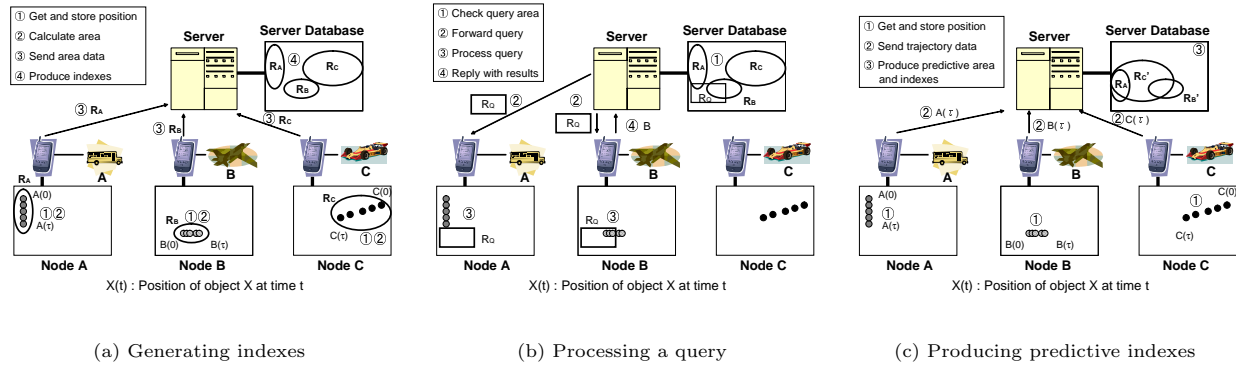


図 1 分散型移動オブジェクトデータベース
Fig. 1 Distributed moving objects database

次に、インデックスを用いることで、いかに効率的にクエリに対処できるかを図 1(b) に例示する。時刻 $0 \leq t \leq \tau$ において範囲 R_Q に位置する物体を問うようなクエリに対しては、まずインデックスを用いて検索すべきデータを保持するセンサを特定し（この例では A, B ）、次に特定したセンサに対しクエリを転送する。クエリが転送されたセンサでは、各センサで蓄積しているローカルなデータに対してクエリの処理を行い、答となるデータを検索する。そして最終的にユーザは、サーバを通して検索結果（この例では B ）を得ることができる。

このように、各センサは τ ごとに各物体が移動した範囲を送信するだけで、サーバがインデックスを用いて効率的にデータを検索することができる。

2.2 問題点とアプローチ

受信した物体の移動範囲を用いたインデックスだけでは、サーバは各センサのもつデータをリアルタイムに把握することができなくなり、クエリの即応性が失われる問題が生じる。これは、各センサが物体の位置を取得してから、サーバでその位置に対応するインデックスが構築されるまでに遅延が生じるからである。先ほどの例において、 $t = \tau + \omega$ ($0 < \omega < \tau$) のとき、現在時刻 $\tau < t < \tau + \omega$ に対するクエリに対しては、サーバは即答できない。なぜなら、このケースでは各センサは τ 毎にしか移動範囲を送信していないので、時刻 $t = \tau + \omega$ において、 $\tau < t < \tau + \omega$ に対応するインデックスを構築していないからである。つまり、クエリを受け取った時刻から $\tau - \omega$ 経過し $t = 2\tau$ まで待たなければ、各センサから $\tau < t \leq 2\tau$ における移動範囲を受信できず、インデックスも構築できないのである。

このクエリ即応性低下の問題を防ぐためには、データ受信時刻において、以降の物体の移動範囲を予測して、先読みのインデックスを構築する必要がある。先の例においてこの手法を導入すると、図 1(c) のように各センサから $0 \leq t \leq \tau$ の移動範囲を受信すると同時に、 $\tau < t \leq 2\tau$ における物体の移動範囲を予測して、予測移動範囲 (R'_A, R'_B, R'_C) を作成する。この予測移動範囲を用いて、サーバは先読みのインデックス $I'_A = \{A, \tau < t \leq 2\tau, R'_A\}$, $I'_B = \{B, \tau < t \leq 2\tau, R'_B\}$, $I'_C = \{C, \tau < t \leq 2\tau, R'_C\}$ を構築する。筆者らは、この先読みのインデックスのことを予測型インデックスとよんでいる。

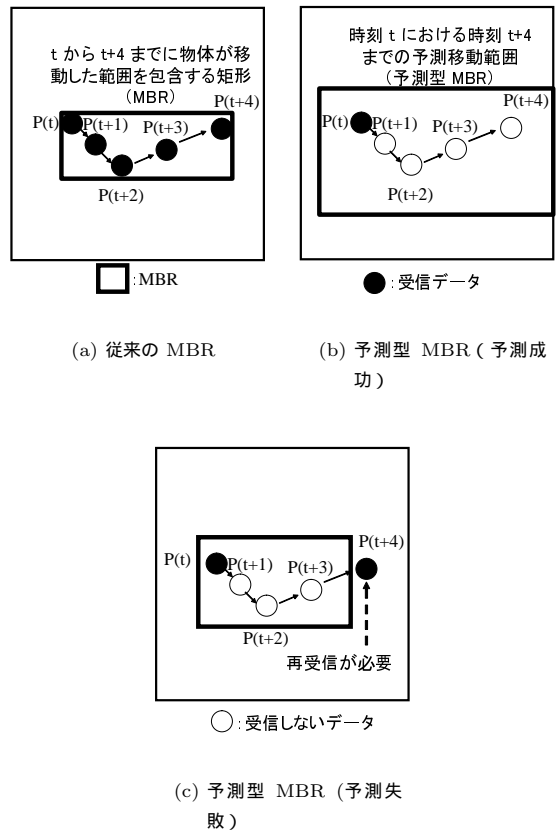


図 2 予測型 MBR の例
Fig. 2 An example of predictive MBR

このように予測型インデックスを用いることで、クエリの即応性を失わずに、データの分散化することの有効性を保つことができる。次節において、この予測型インデックスの概要について述べる。

2.3 予測型インデックスの概要

予測型インデックスとは、従来の空間インデックス [17] ~ [19] に時間経過による物体の移動先を予測する手法を導入したものである。本稿では、予測移動範囲を一般の空間インデックスにおける MBR (Minimum Bounding Rectangle) [20], [21] が一定時間経過後にどのような形に変形するかを、物体の過去の移動軌跡

データを用いて予測する．MBR とは，1つのインデクスに属する物体の位置を全て包含するような最小の矩形のことである．当然ながら物体の移動先は可能な限り移動予測先の MBR（予測型 MBR）に含める必要がある．仮に予測が当たらずに予測型 MBR の領域外に移動する物体があれば，クエリに正しく答えるためにその物体を管理しているセンサから位置データを再受信し，インデクスを再構築しなければならないからである．

図 2 に予測型 MBR の具体的な例を示す．図 2(a) は，時刻 t から $t+4$ までの物体の位置に従来の MBR を作成した様子である．図 2(b) は時刻 t において $t+1$ から $t+4$ までの予測が全て成功した場合の予測型 MBR の例で，図 2(c) は時刻 t において $t+1$ から $t+3$ までの予測は成功したが $t+4$ の予測が外れ $t+4$ における位置データの再受信が必要になる場合の予測型 MBR の例を示す．

このように，物体の移動予測が外れると，その都度センサが物体の位置を送信する必要があり，データを分散させる有効性が失われる．つまり，分散化の効率を上げるためには，物体の移動予測の精度を可能な限り向上させることが重要となる．一方，センサから受信するデータを減らしサーバの負荷を軽減させるためには，長時間先までの物体の移動を予測する必要がある．しかし，効率よく目的のセンサを検索するためには，予測型 MBR の面積をできるだけ小さくする必要がある．つまり，できるだけ多くの物体を予測型 MBR 内に包含し，かつ可能な限り物体の移動に沿った予測型 MBR を作成することが求められる．

次章で，この予測型 MBR を作成するための物体の移動予測技術について述べる．

3. 物体の移動予測技術

本章では，筆者らが提案してきた従来の移動予測技術 [8], [9] と本稿で提案する移動予測技術について述べる．3.1 において，従来の移動予測技術の概略を示し，3.2 で本稿で新たに提案する物体の移動パターンに着目した移動予測技術について説明する．本稿では，いくつかの移動予測技術の性能を公平に比較するため，予測型 MBR は全領域を面積の等しいセルに区切ったセル単位とする．

3.1 従来の予測技術

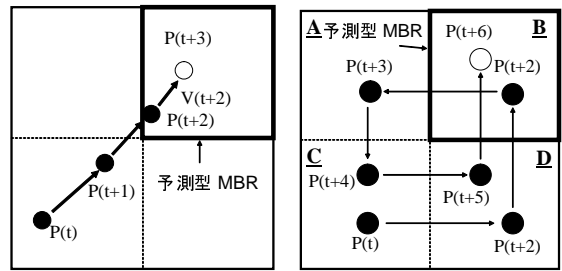
本節では，筆者がこれまでに提案した従来の物体の移動予測技術である物体の移動速度，移動履歴を用いた予測技術の概要について述べる．

(1) 移動速度を用いた予測

物体の移動が連続であることに着目して，直前の物体の移動速度で等速に移動すると予測して移動先を求める．図 3(a) のように，時刻 $t+3$ における物体の位置を予測する場合，時刻 $t+2$ における物体の速度で等速に移動すると予測し，予測先の位置を含むセル領域を予測型 MBR とする．

(2) 移動履歴を用いた予測

同じような移動を行う物体が多いことに着目して，現在の移動遷移列と一致する遷移列を過去の移動履歴の中から探し，得られた遷移列の中からその次の遷移として最も出現回数の多い



(a) 移動速度を用いた予測

(b) 移動履歴を用いた予測

図 3 従来の移動予測の例

Fig. 3 Examples of conventional moving predictions

遷移先を次の移動先として予測する．図 3(b) に移動履歴を用いた予測の例を示す．まず，予め全領域を A, B, C, D のように均等な矩形領域に分割しておく．時刻 $t+6$ における移動予測を行う場合， $t+4, t+5$ における移動履歴 ($C \rightarrow D$) とそれ以前の $t, t+1, t+2, t+3$ における移動履歴 ($C \rightarrow D \rightarrow B \rightarrow A$) を比較し，一致する履歴を探す．この例では $t, t+1$ においてのみ， $t+5, t+6$ と同様の移動を行っていたので， $t+2$ における遷移 B に移動すると予測し，予測先のセル領域を予測型 MBR とする．

移動速度を用いた予測は，直線運動のように速度がほとんど変化しない物体の予測に対しては有効であった．しかし，移動方向が頻繁に変わるような速度が著しく変化する物体の予測に対しては効果が期待できなかった．また，移動履歴を用いた予測は，移動に規則性がある物体に対しては有効であった．しかし，過去の物体の移動履歴の中に，最近の移動と全く同じ履歴がない場合，予測先を限定することができなかった．

そこで本稿では，移動履歴を用いた予測を拡張し，過去の移動履歴の中から類似する移動履歴を探し，それを移動予測に用いる移動パターンに着目した予測技術を新たに提案する．過去の移動履歴から最近の移動遷移に類似した遷移列を抽出する方法として，確率モデルであるマルコフモデル [13], [14] と最長共通部分列を求める LCSS [15], [16] を用いる．

3.2 移動パターンを用いた予測技術

予測手法の説明の前にまず，移動パターンの作成方法について説明する．まず，物体が移動する全範囲を均等に区切った全てのセルに Cell Identification (CID) を付与しておく．そして，各センサから物体の位置を受信するたびに物体が位置する CID を記録する．すると，物体が移動するにつれて物体が位置するセルも変化し，CID も変化することになる．この物体の CID の遷移列を物体の移動遷移と定義する．また，過去の全移動遷移を全て $n+1$ (定数) 個の CID 遷移となるように揃える．このとき，各移動遷移ごとに出現回数をカウントしておく．この移動遷移と出現回数を組にしたものを移動パターンと定義する．

以下，確率モデルであるマルコフモデルを移動予測に応用した方法，最長共通部分列を求めるアルゴリズムである LCSS を移

```

INPUT: int  $\gamma$ ; /*セルの数*/
CID_sequence_history S[ $\gamma$ ]; /*過去マルコフの全移動パターン*/
/* S[n]: 最終状態のCIDがnである移動パターン*/
int h[n]; /*予測対象物体の最新n状態の移動遷移*/
double Th; /*類似度の閾値*/
OUTPUT: int c; /*予測CID*/
CID_sequence_history{ /*移動パターンの構造体*/
int[n] history; /*n状態の移動遷移*/
int number; /*移動遷移の存在回数*/
}
int Prediction_using_Markov( $\gamma, S, h, Th$ ){
int Ppre[ $\gamma$ ]; /*予測CIDが $\gamma$ となる予測度*/
int tempN=0; /*一致回数*/
double Psim=0; /*類似度*/
for (int i=1; i $\leq$   $\gamma$ ; i++){ /*全セルに対して*/
foreach (CID sequence s in S[i]){
/* S[i]に存在する全移動パターンに対して*/
for (int j=1; j $\leq$  n; j++){ /*sの全状態に対して*/
if (h[j] == s.history[j]) then
/*予測対象パターンとの一致回数をカウント*/
tempN++;
}
Psim = tempN/p; /*類似度計算*/
if (Psim > Th) then { /*類似度が閾値を超えた*/
Ppre[i] += s.number * Psim;
/*予測度をインクリメント*/
tempN=0; /*初期化*/
}
}
return c where maximum Ppre[c] (for int c=1; c $\leq$   $\gamma$ ; c++);
/*予測度Ppre[c]が最大となるようなcを予測CIDとして返す*/
}

```

図4 マルコフモデルを用いた移動予測のアルゴリズム

Fig. 4 Algorithm of moving prediction using Markov model

動予測に応用した方法を説明する。また具体的な物体の予測例についても示し、マルコフモデルと LCSS の違いを明示する。

3.2.1 マルコフモデルを用いた予測

確率モデルであるマルコフモデルをそのまま移動予測に適用すると、予測対象の移動遷移と全く同じ移動遷移が過去の移動パターンに存在しなければ、移動先を予測することができない。そこで本節では、マルコフモデルを応用して、予測対象の物体の移動遷移に類似する移動遷移を予測に用いる方法について具体的に説明する。

まず、過去の移動パターンを状態数が $n + 1$ のマルコフモデル（過去マルコフ）におきかえる。このとき、遷移回数から最終状態への推移確率を求めることができる。次に、予測対象となる物体の過去 n 状態の移動遷移（予測対象遷移列）と過去マルコフの最終状態以外の n 状態を比較し、それぞれの過去マルコフについて一致した状態の割合（類似度）を求める。類似度がある閾値をこえた場合のみ、その移動パターンが予測対象の移動パターンに類似したとみなし、最終状態への推移確率と類似度の積（予測度）を求める。最終的に予測度が最も高い最終状態を求め、その最終状態に対応する CID を予測 CID とし、予測 CID のセル領域を予測型 MBR とする。図4に、以上で説明したマルコフモデルを用いた予測のアルゴリズムを示す。

3.2.2 LCSS を用いた予測

LCSS (Longest Common Subsequence [15][16]) は DNA などのパターンマッチングによく用いられ、2つのシーケンスの

```

INPUT: int[] S, Sq;
OUTPUT: int sim; /*SとSqの類似度*/
int Calculate_LCSS(S, Sq){
int[m+1, n+1] C; /*LCSS テーブル*/
/* m=S, n=|Sq| */
/* LCSS tableの初期化 */
for (int i=0; i $\leq$  m; i++){
C[i, 0]=0;
for (int i=0; i $\leq$  n; i++){
C[0, i]=0;
/* SとSqの類似度を求める */
for (int i=1; i $\leq$  m; i++){
for (int j=1; j $\leq$  n; j++){
if (S[i] == Sq[j])
C[i, j] = C[i-1, j-1] + 1;
else
max(C[i-1, j], C[i, j-1]);
}
}
return C[m, n];
/* SとSqの最長部分列の長さ */
}

```

図5 LCSS を用いて類似度を求めるアルゴリズム

Fig. 5 Algorithm of obtaining similarity using LCSS

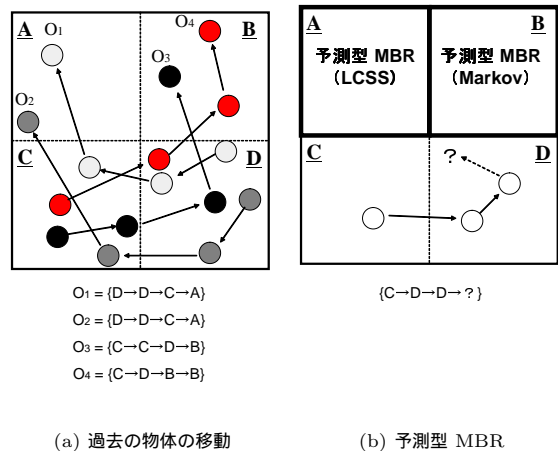


図6 移動パターンに着目した移動予測の例

Fig. 6 An example of moving prediction using motion patterns

最も長い共通部分を求めるアルゴリズムである。これを移動予測に適用して以下のように予測先を求める。

まず、過去の全物体の移動パターンと予測対象遷移列を比較し、LCSS を用いて最も類似度の大きい移動パターンを抽出する。以降、予測対象物体は抽出されたパターンどおりに移動すると予測する。最大類似度となるパターンが複数ある場合、その中で最も遷移回数が大きい遷移列を予測に用いる。図5に LCSS を用いて類似度を求めるアルゴリズムを示す。

3.2.3 予測型 MBR の作成例

図6は移動パターンに着目した移動予測の動作例である。図

表 1 評価用パラメータの定義

Table 1 Definition of parameters for evaluation

位置取得回数	単位時間当たりに各センサが物体の位置を取得する回数	N_s
送信間隔	各センサが 1 回データを送信するまでに物体の位置を取得する回数	τ
予測精度	予測 MBR に含まれる物体の位置割合	p

6(a) は、 2×2 のセル (CID=A, B, C, D) に分割されたグリッド上を物体 O_1, O_2, O_3, O_4 が移動する様子を示す。以下、この 4 つの物体の移動パターンのみを用いて、図 6(b) のように移動する物体の予測型 MBR を求める。

過去の移動パターン (遷移列, 出現回数) をまとめると以下のようになる。

$$\langle \langle D, D, C, A \rangle, 2 \rangle$$

$$\langle \langle C, C, D, B \rangle, 1 \rangle$$

$$\langle \langle C, D, B, B \rangle, 1 \rangle$$

以下、マルコフモデル, LCSS を用いた予測によって、それぞれの予測型 MBR を求める。

a) マルコフモデル

まず、予測対象遷移列である $h = \{C, D, D\}$ と移動パターンを用いて、それぞれの類似度 P_{sim} を求める。

$$\{D, D, C\} : P_{sim} = 1/3$$

$$\{C, C, D\} : P_{sim} = 2/3$$

$$\{C, D, B\} : P_{sim} = 2/3$$

次に、最終状態への推移確率との積をとった予測度 P_{pre} の和をそれぞれの CID について求める。

$$A : P_{pre} = 1 \times 1/3 = 1/3$$

$$B : P_{pre} = 1 \times 2/3 + 1 \times 2/3 = 4/3$$

$$C : P_{pre} = 0$$

$$D : P_{pre} = 0$$

以上より、予測度が最大となるのは CID=B のときであるので、予測 CID として B が求められる。つまり、マルコフモデルを用いた予測により得られる予測型 MBR は図 6(b) のように CID=B のセル領域全体となる。

b) LCSS

図 5 のアルゴリズムを用いて、予測対象遷移列である $h = \{C, D, D\}$ と過去のそれぞれの移動パターンとの最長部分列長を、以下のように類似度として求める。

$$\{D, D, C\} : P_{sim} = 2$$

$$\{C, C, D\} : P_{sim} = 2$$

$$\{C, D, B\} : P_{sim} = 2$$

類似度の最大値は 2 であるが複数存在するので、類似度だけでは決めることはできない。そこで、類似度が最大である移動遷移の中で、出現回数が最も多い $\langle D, D, C, A \rangle$ の最終状態 A を予測 CID とする。つまり、LCSS を用いた予測により得られる予測型 MBR は図 6(b) のように CID=A のセル領域全体となる。

4. 評価実験

本章では、筆者らが提案する移動予測技術の有効性を示すための評価実験の概要、及び結果について述べる。以下、4.1 で本実験における分散型移動オブジェクトデータベースのモデルを定義し、4.2 で実験の概要、4.3 で実験結果の考察について述べる。

4.1 モデル定義

予測型インデクスは、過去の物体の移動の様子から未来の物体の移動を予測して作成される予測移動範囲を用いて構築される。しかし予測であるので、実際には予測移動範囲の領域外に移動する物体が存在する可能性がある。予測移動範囲の領域外に物体が移動した場合、つまり予測が 100 % 正確ではない場合、その領域外に移動した物体の位置をただちに送信する必要がある。そのためには、サーバは予測型 MBR が作成されるとすぐに各センサにそれぞれの物体の予測型 MBR を送信しなければならない。このように、データの分散化によるデメリットも発生する。

そこで、どのような条件を満たせば分散化が有効になるのかを求める。その前に評価に用いる変数を表 1 に定義する。従来の集中型移動オブジェクトデータベースは各センサが物体の位置を取得するたびにサーバへデータを送信していたので、表 1 の変数を用いて総送信回数は $\phi_s = nN_s$ ($n \in \mathbb{N}$) となる。一方、提案する分散型移動オブジェクトデータベースでは、ある一定間隔 i 毎に位置データを送信し、さらに予測が外れた場合に再送信する必要があるので、総送信回数は

$$\phi_c = n \left(\frac{N_s}{\tau} + N_s(1-p) \right)$$

となる。分散型移動オブジェクトデータベースを提案する目的は、各センサの送信回数を減らすことによるサーバ負荷の軽減であるので、送信回数削減率

$$P_R = \frac{\phi_s - \phi_c}{\phi_s} = \left(p - \frac{1}{\tau} \right) \cdot 100$$

を可能な限り 100% に近づけることが目標となる。つまり、送信間隔 τ と予測精度 p を大きくすることが送信回数削減率を大きくするための要件となる。一方、効率よく目的のセンサを検索するためには、予測型 MBR の面積を可能な限り小さくする必要があるが、本実験では、各予測技術を公平に比較するためにそれぞれの移動軌跡データにおいて予測型 MBR の面積を等しくした。

4.2 評価実験内容

本実験では、実際に GPS を用いて取得した人力車、センサを用いて取得した人、モーションキャプチャのデータについて、各予測技術を用いた場合の送信回数削減率を計測した。表 2 に

表 2 評価用移動軌跡データ

Table 2 Moving trajectory data for evaluation

移動物体	人力車	人	モーションキャプチャ
移動物体の数	のべ 39 台	のべ 49 人	のべ 20 個
センサの数	1 個/1 台	1 個/1 人	37 個/1 台
データ量	平均 800KB	平均 100KB	平均 750KB
データ取得時間	平均 8 時間	平均 5 分	平均 30 秒
データ取得間隔	1 秒毎	1/600 秒毎	1/3 秒毎
セルの個数	30 × 30	30 × 30	15 × 15
物体の移動範囲	奈良市内 (2km × 2km)	フォーラム会場 (30m × 30m)	半径 4m の円

評価に用いた移動軌跡データの詳細を示し、以下に比較を行った五種類の予測技術を示す。

- 新たに提案した予測技術
 - マルコフモデルを用いた予測 (Markov)
 - LCSS を用いた予測 (LCSS)
- 従来の予測技術
 - 移動速度を用いた予測 (Velocity)
 - 移動履歴を用いた予測 (History)
- 停止予測 (Stay)

停止予測は予測技術の比較を行うための参考手法である。停止予測は、物体が現在位置するセルそのものを予測型 MBR とする手法である。つまり、物体は移動せずに停止し続けると予測していることとなる。

以下、各移動予測技術を比較するための実験内容について示す。

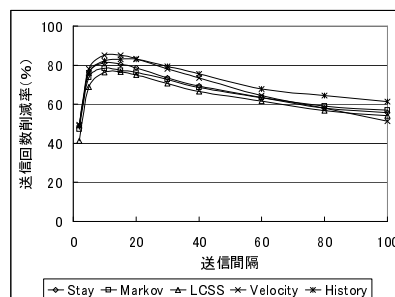
(1) 送信間隔に対する予測精度 (図 7)

送信間隔を変化させることで、最適な予測方法が変化するかどうかを調べる。それぞれの移動軌跡データに対し、送信間隔を変化させながらそれぞれの移動予測技術により得られる予測精度を計測する。

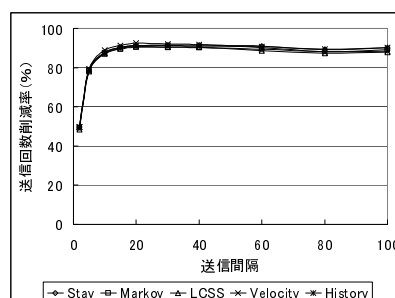
(2) 予測対象セル数に対する送信回数削減率 (図 8)

物体の移動の性質の違いによって、予測対象となるセルの数を増加させたときの送信回数削減率の減少率に差が生じるかを調べる。それぞれの移動軌跡データについて、予測対象セル数を増加させながらそれぞれの移動予測技術により得られる送信回数削減率を計測する。最も手法間の差が顕著に得られるように送信間隔は $\tau = 30$ とした。予測対象セル数の増加方法は、移動パターンに着目した予測では予測度が高い上位 n 個のセルを予測 CID として選択し、移動履歴に着目した予測では、一致した履歴の中で、遷移回数の多い上位 n 個を予測先として選択する。

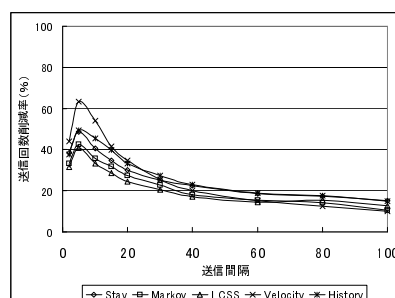
実験結果は、いずれも人力車、人のデータは 20 種類、モーションキャプチャデータは 500 種類をランダムに抽出し、それぞれ 5 回の試技の平均値をプロットした。送信回数削減率を 0% 以上にすることができれば、集中型データベースよりも分散型データベースにおける送信回数の方が少ないということであるので、サーバの負荷を軽減でき分散化の有効を示すことができたといえる。図 7 に示すように、いずれの予測技術でも、またいずれのデータであっても送信回数削減率を 0% 以上にす



(a) 人力車のデータ



(b) 人のデータ

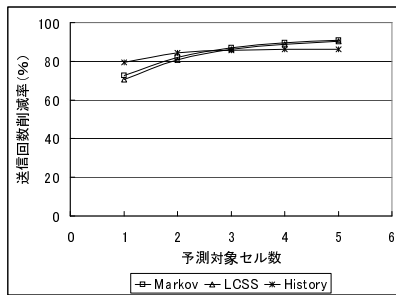


(c) モーションキャプチャのデータ

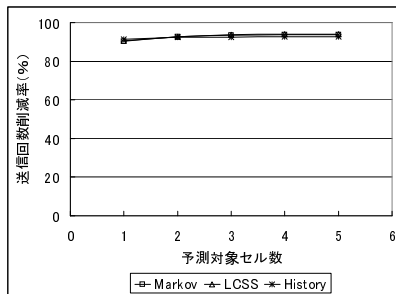
図 7 送信間隔に対する送信回数削減率

Fig. 7 Transmission reduction ratio for transmission interval

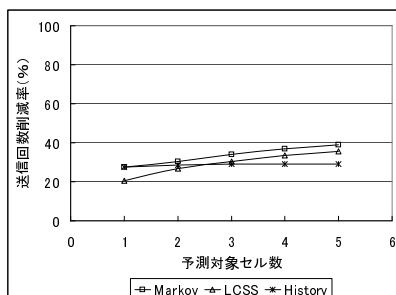
ることができた。つまり、予測型インデックスの導入により、遅延の問題をある程度解消し、データ分散化の利点を生かすことができるといえる。



(a) 人力車のデータ



(b) 人のデータ



(c) モーションキャプチャのデータ

図 8 予測対象セル数に対する送信回数削減率

Fig. 8 Transmission reduction ratio for the number of predictive cells

4.3 実験結果の考察

まず、予測対象セル数が1のとき、つまり予測型 MBR の面積が比較的小さい場合において、移動物体の種類ごとに有効な予測技術について図7より考察する。

人力車のデータでは、移動速度を用いた予測が最も有効な方法で、送信間隔が15回のときに送信回数削減率を85%にまで増加させることができた。これは、人力車のデータは一定速度で移動することが多く、直前の速度に依存して移動することが多いためと考えられる。しかし、逆に送信間隔を大きくすると移動速度を用いた予測は送信回数削減率が急減する。これは、人力車は等速で移動し続けるのは15秒までで、それ以降は移動速度が変化することが多いことを示している。一方、送信間隔を大きくすると移動履歴を用いた予測によって最も送信回数

削減率を大きくできる。つまり、人力車のデータはある周期で繰り返される移動パターンをもつことを示している

また、人のデータは、予測技術に関わらず総じて送信回数削減率を90%程度にまで増加させることができた。また、送信間隔を大きくしても送信回数削減率がそれ程大きくならず、停止予測も有効であることから、時間経過にかかわらずほとんど移動しないデータであったことが分かる。

モーションキャプチャのデータは、セルの区切り方が最も粗いにもかかわらず他のデータに比べ送信回数削減率が最も低くなった。これは、モーションキャプチャのデータは移動予測が難しく、予測型 MBR の領域外に移動する物体が増え、再送信の回数が増加したためと考えられる。しかし、送信間隔が小さいときは、移動速度に着目した予測は送信回数削減率を65%程度にまで増加させることができた。これは、移動予測が難しい移動軌跡データであっても移動の連続性から、近い未来であればある程度の予測が可能であることを示している。

このように、予測対象のセル数が1つであるときのように予測型 MBR の面積が小さい場合は、移動速度や移動履歴を用いた予測が有効であることが示された。次に、予測型 MBR の面積が比較的大きい場合の有効な予測技術を図8より考察する。予測対象セル数を増加することで、予測精度が向上し送信回数削減率を増加させることができると考えられるが、予測型 MBR の面積が増加するので目的のセンサを検索する効率が悪化する。

人力車やモーションキャプチャのデータは、予測対象セル数を増加させたときの送信回数削減率の増加が最も顕著である。特に、本稿で提案する移動パターンに着目した予測を用いた場合に、送信回数削減率の上昇率は顕著である。つまり、予測対象セル数を増加させるような場合には、従来の移動履歴を用いた予測技術に対する今回新たに提案する移動パターンを用いた予測技術の優位性を述べるができる。

一方、以上のような実験から移動軌跡データによって送信回数削減率に大きな違いが現れることが分かる。つまり、データによって予測の難易度が大きく異なることを示している。この予測難易度を左右する原因となる移動性質として物体の速度の分散値が考えられる。速度の分散値が大きいということは、移動のランダム性が高いということなので移動予測は難しいと考えられる。そこで、それぞれの移動軌跡データについて、送信間隔ごとに速度の分散値の計測しそれらの平均値を移動範囲の大きさ N_s によって正規化した値を速度分散として求めると、以下ようになる。

- 人力車：0.05
- 人：0.031
- モーションキャプチャ：0.327

送信回数削減率が高く、移動予測が比較的容易であることが分かる人のデータは速度分散が最も低く、逆に、移動予測が比較的難易であることが分かるモーションキャプチャのデータは速度分散が最も高くなる。つまり、移動軌跡データに対する速度分散を予め計測しておくことで、ある程度送信回数削減率を見積もることができる。

5. おわりに

本稿では、分散型移動オブジェクトデータベースにおける予測型インデクス構築のための物体の移動予測技術の提案、及び性能評価を行った。物体の移動パターンに着目して2つの物体の移動予測技術を提案し、これらの予測技術と従来の予測技術により、どんな移動物体に対しても、センサの送信回数を減らすことができ、サーバの負荷を減少させることができることを示した。また、速度の分散値が移動予測の難易度に大きな影響を与えることを得た。移動軌跡データの性質から最適な予測技術を求める定量的評価、また、速度の分散値から予測精度を求める式の実験的評価は今後の課題とする。

謝 辞

本研究を進めるにあたって多大なご協力を頂いた大阪大学情報科学研究科の村田正幸教授をはじめ、大阪大学の諸先生方には、この場を借りて厚く御礼申し上げます。

文 献

- [1] 杉野 勝敏, 朝倉 康夫, 羽藤 英二, “移動体位置情報を用いた交通行動シミュレーションモデルの開発,” 土木計画学研究・講演集 (CD-ROM), vol. 26, Nov. 2002.
- [2] M. Vazirgiannis and O. Wolfson, “A spatio temporal model and language for moving objects on road networks,” in *Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases (SSTD)*, pp. 20–35, July 2001.
- [3] M. Hadjieleftheriou, G. Kollios, V. Tsotras, and D. Gunopulos, “On-line discovery of dense areas in spatio-temporal databases,” in *Proceedings of the 8th International Symposium on Spatial and Temporal Databases (SSTD) Conference*, pp. 306–324, July 2003.
- [4] G. Kollios, V. J. Tsotras, D. Gunopulos, A. Delis, and M. Hadjieleftheriou, “Indexing animated objects using spatiotemporal access methods,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 5, pp. 758–777, 2001.
- [5] M. Song, K. Park, J. Ryu, and C.-S. Hwang, “Modeling and tracking complexly moving objects in location-based services,” *Journal of Information Science and Engineering*, vol. 20, pp. 517–534, 2004.
- [6] 白石 陽, 安西 祐一郎, “分散センサデータベースの閲覧のためのインクリメンタルなデータ提供方式,” トランザクション「データベース」, vol. 44, 2003.
- [7] Y. Cai, K. A. Hua, and G. Cao, “Processing range-monitoring queries on heterogeneous mobile objects,” in *Proceedings of 2004 IEEE International Conference on Mobile Data Management (MDM'04) Conference*, pp. 27–38, Jan. 2004.
- [8] 勝田 圭介, 柳沢 豊, 佐藤 哲司, “分散型移動軌跡データベースのためのインデクス構築手法,” 第132回データベースシステム第7回放送コンピューティンググループ合同研究会報告書, pp. 17–24, Jan. 2004.
- [9] 勝田 圭介, 柳沢 豊, 佐藤 哲司, “分散型移動軌跡データベースのための移動の特徴に基づく予測型インデクス構築手法,” マルチメディア, 分散, 協調とモバイル (DICOMO2004) シンポジウム論文集, pp. 397–400, July 2004.
- [10] Keisuke Katsuda, Yutaka Yanagisawa, and Tetsuji Satoh, “A pattern-based predictive indexing method for distributed trajectory databases,” in *Proceedings of the International Conference on Information Networking (ICOIN) 2005*, pp. 745–754, Jan. 2005.
- [11] Y. Xu, J. Winter, and W.-C. Lee, “Prediction-based strategies for energy saving in object tracking sensor networks,” in *Proceedings of 2004 IEEE International Conference on Mobile Data Management (MDM'04) conference*, pp. 346–357, Jan. 2004.
- [12] D. Papadias, J. Zhang, N. Mamoulis, and Y. Tao, “Query processing in spatial network databases,” in *Proceedings of the 29th Very Large Data Bases Conference (VLDB2003)*, pp. 12–23, Sept. 2003.
- [13] F. J. Damerau, *Markov Models and Linguistic Theory: An Experimental Study of a Model for English*. Janua Linguarum Series Minor No 95, Mouton De Gruyter, 1971.
- [14] L. C. G. Rogers, D. Williams, and D. Williams, *Diffusions, Markov Processes, and Martingales*. Cambridge Univ Pr (Txp), 2000.
- [15] Y.-T. Tsai, “The constrained longest common subsequence problem,” *Information Processing Letters*, vol. 88, pp. 173–176, Nov. 2003.
- [16] L. Bergroth, H. Hakonen, and T. Raita, “A survey of longest common subsequence algorithms,” in *Proceedings of the Seventh International Symposium on String Processing Information Retrieval*, pp. 39–48, Sept. 2000.
- [17] P. K. Agarwal, L. Arge, and J. Erickson, “Indexing moving points,” in *Proceedings of the ACM SIGMOD/PODS 2000 conference*, pp. 175–186, May 2000.
- [18] S. Saltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez, “Indexing the positions of continuously moving objects,” in *Proceedings of the ACM SIGMOD/PODS conference 2002*, pp. 331–342, May 2000.
- [19] S. Saltenis and C. S. Jensen, “Indexing of moving objects for location-based services,” in *Proceedings of the 18th International Conference on Data Engineering (ICDE2002)*, pp. 463–486, Feb. 2002.
- [20] Y.-J. Choi and C.-W. Chung, “Selectivity estimation for spatio-temporal queries to moving objects,” in *Proceedings of the ACM SIGMOD/PODS 2002 conference*, pp. 440–451, June 2002.
- [21] Y. Tao, J. Sun, and D. Papadias, “Selectivity estimation for predictive spatio-temporal queries,” in *Proceedings of International Conference on Data Engineering (ICDE2003)*, pp. 417–428, March 2003.