

索引付けされた移動軌跡データ からの効率的な移動統計量抽出法の評価

Evaluation of a Mobility Statistics Extraction Scheme for Indexed Spatio-Temporal Datasets

塚本 祐一¹ 石川 佳治² 北川 博之³

Yuichi TSUKAMOTO Yoshiharu ISHIKAWA
Hiroyuki KITAGAWA

空間情報利用の急速な進展および携帯機器などの普及から、時空間データベースの研究分野では、移動するオブジェクトやユーザの移動状況を、データベースに蓄積し効率よく管理するための研究が盛んに進められている。セルの集合に分割された空間上を時間の経過につれてオブジェクトが移動する状況を捉えるための統計量として、マルコフ連鎖モデルが存在する。我々は、蓄積された移動情報データからマルコフ連鎖モデルに基づき移動統計量を高速に抽出し、対話的な移動状況の分析を支援するための手法について研究を進めており、移動オブジェクトの移動軌跡データが空間索引R-木に蓄積されている際に、マルコフ連鎖の遷移確率を効率的に求めるための手法を開発している。本稿では実験に基づく提案手法の有効性の評価を行う。

With the recent progress of spatial information technologies and mobile computing technologies, spatio-temporal databases which store information on moving objects including vehicles and mobile users have gained a lot of research interests. Here we focus on a mobility statistics value called the Markov transition probability, which is based on a cell-based organization of a target space and the Markov chain model. We have proposed an algorithm to extract mobility statistics from indexed spatio-temporal datasets for the interactive analysis of huge collections of moving object trajectories. The proposed algorithm efficiently computes the specified Markov transition probabilities with the help of a spatial index R-tree. In this paper, we evaluate the effectiveness of proposed method based on an experiment.

¹ 学生会員 筑波大学大学院システム情報工学研究科
yuichi@kde.is.tsukuba.ac.jp
² 正会員 筑波大学電子・情報工学系
ishikawa@is.tsukuba.ac.jp
³ 正会員 筑波大学電子・情報工学系
kitagawa@is.tsukuba.ac.jp

1. はじめに

今日では空間情報利用の拡大や、モバイルユーザを対象としたデータベース技術が重要となっており、時空間データベースに関する研究が盛んに進められている[2]。時空間データベースからの統計情報の抽出に関する研究はその一つである。時空間データに関する統計量は、問合せ処理の効率化のみならず、蓄積された移動状況データをもとに移動分析を行う際においても有用である[4]。時空間データの利用拡大により、移動分析のための統計量を効率よく抽出することが今後より重要となると考えられる。

以上の背景をもとに、我々は時空間データベースから移動オブジェクトの移動状況に関する統計情報を抽出する手法の研究を進めている[5-7]。移動に関する統計量として、本研究ではマルコフ連鎖(Markov chain)モデルに基づく移動統計量を対象としている。

時空間データ分析におけるマルコフ連鎖モデルは、ある地域から別の地域へある期間内にどの程度の人口が移動したといったような、人や物の時空間的な移動傾向を把握するために用いられる[4]。この統計情報により、ある時点である位置にいるオブジェクトが次の時点でどこに移る可能性が高いといった予測が可能となる。

本研究では特に、移動オブジェクトの移動軌跡が空間索引R-木に蓄積されている状況を想定し、R-木から効率的にマルコフ連鎖の遷移確率を推定する手法を開発している。本稿では提案手法の実装と実験を行い、その有効性を示す。

2. マルコフ過程モデルに基づく移動統計量

図1に示すように、空間がセルに分割されているとする。各セルは矩形でなくてはならないが、サイズは必ずしも均一でなくてよいとする。各セルにはセル番号が付けられていて、番号によりセルを特定できるものとする。図1は、時刻 $t =$ でセル c_0 にいたオブジェクト A が、次の時刻 $t = t+1$ でセル c_1 に、そして $t = t+2$ の時点でセル c_2 に移動した状況を示している。

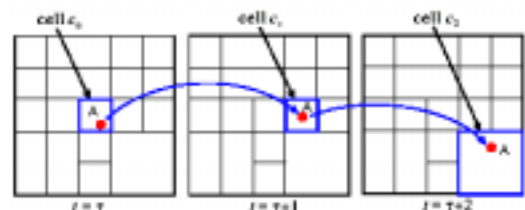


図1 マルコフ過程モデルの概念

Aのように、セル c_0 に現在いるオブジェクトが次の時点でセル c_1 に移動する確率 $\Pr(c_1|c_0)$ を時空間データベース中のデータを用いて予測したいとすると、以下のように計算できる。

$$\Pr(c_1|c_0) = \frac{\sum_{t=0}^{T-1} |\text{objs}(c_0, t) \cap \text{objs}(c_1, t+1)|}{\sum_{t=0}^{T-1} |\text{objs}(c_0, t)|} \quad (1)$$

$\text{obj}(c, t)$ は、時刻 t にセル c の領域に含まれているオブジェクトの集合を返す関数である。各時点 $t = 0, \dots, T-1$ で c_0 にいたオブジェクトの総数を分母とし、そのようなオブジェクトのうち、次の時点で c_1 に移ったオブジェクトの総数を分子として計算する。

確率 $\text{Pr}(c_1|c_0)$ は、現在の状態 (セル c_0) に依存して次の状態 (セル c_1) を予測するという意味で、1次のマルコフ過程の遷移確率に相当する。これを一般化すると、 n 次のマルコフ過程による遷移確率、すなわち、単位時間ごとに c_0, c_1, \dots, c_{n-1} とセルを移動したオブジェクトが次の時点でセル c_n を訪れる確率 $\text{Pr}(c_n|c_0, \dots, c_{n-1})$ と表され、その推測値は以下の一般形になる。

$$\text{Pr}(c_n | c_0, \dots, c_{n-1}) = \frac{\sum_{i=0}^{T-1} |\bigcap_{j=0}^n \text{objs}(c_j, t+i)|}{\sum_{i=0}^{T-1} |\bigcap_{j=0}^{n-1} \text{objs}(c_j, t+i)|} \quad (2)$$

なお、ここでは移動オブジェクトの移動が定常的 (stationary) な過程に従い、時刻によって遷移確率が変化しないと仮定している。

マルコフ遷移確率を効率的に求めることができれば、移動オブジェクトが次の時点でどのセルに移動する可能性が高いかという、一種の経路予測を行うことが可能となる。ある時刻における移動オブジェクト集合の移動状況データがまとめて与えられたとき、以降における移動状況をシミュレーションすることも可能となる。

3. 空間索引による索引付け

以下では、移動軌跡を点の集合として離散的に表現する場合の、本研究のアルゴリズムが想定する索引付け方式について述べる。線分を用いて移動軌跡を表現する場合の索引付け方式については、[5]でその方法を述べている。

例として、1次元空間でのオブジェクトの移動の例を考える。図2は、オブジェクト A, B が時刻 $t = 0$ から $t = T (= 8)$ まで x 軸上を移動する様子を示している。移動経路は曲線で表現されている。実際の移動軌跡はこのように複雑であるが、計算機上での表現には何らかの近似が必要である。図2に示した経路上の点は、各時刻において移動軌跡をサンプリングしたものである。この近似により、各オブジェクトの経路は時刻と位置のペアの列で表現できる。このような表現手法を点による表現と呼ぶ。GPS により一定期間ごとに移動オブジェクトの位置を検出する場合は、このような表現をすることが自然である。

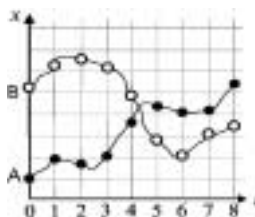


図2 点による移動軌跡の表現

点による表現を用いる場合、移動オブジェクト o のある時刻 $t = (= 0, 1, \dots, T)$ における位置が $[x_1, \dots, x_d]$ という d 次元空

間上の点で表される場合、 $d+1$ 次元空間のR-木を構築し、各時刻 t に対する o の情報を $[x_1, \dots, x_d, t]$ という $d+1$ 次元ベクトルと表現し、 o のオブジェクトIDとペアでR-木に挿入すると索引が構築できる。こうすることで、点による表現に対して時刻 t においてセル c に含まれている移動オブジェクトのIDを検索することが可能となる。

4. 素朴な遷移確率推定アルゴリズム

4.1 問題の定式化

時空間索引を用いて、式(2)に示した n 次のマルコフ遷移確率の推定を行うことを考える。ただし、特定のセルの組合せ c_0, \dots, c_n に対する遷移確率の推定ではなく、以下のように問題を一般化する。

遷移確率の推定問題

$n + 1$ 個のセルの集合 $C_0 = \{c_{0,1}, \dots, c_{0,|C_0|}\}, \dots, C_n = \{c_{n,1}, \dots, c_{n,|C_n|}\}$ が与えられたとき、任意のセルの組合せ (c_0, c_1, \dots, c_n) $C_0 \times \dots \times C_n$ に対し、 $\text{Pr}(c_n|c_0, \dots, c_{n-1})$ の値が未定義でなければその値を出力する。

4.2 素朴なアルゴリズム

素朴な方式として、特定のセルの組合せ c_0, \dots, c_n について $\text{Pr}(c_n|c_0, \dots, c_{n-1})$ の推定を行うことを考える。この確率は以下の2つの集合 S, Q を求めることで計算できる。

- (1) ある時刻 $t = (= 0, 1, \dots, T-n)$ においてセル c_0 にいて、かつ、 $t = +1$ においてセル c_1 にいて、...、かつ $t = +n-1$ においてセル c_{n-1} にいたオブジェクトの集合 S
- (2) S に含まれるオブジェクトのうち、 $t = +n$ においてセル c_n にいたオブジェクトの集合 Q

先の推定問題に対する素朴なアルゴリズムはこのアイデアをもとに導ける。各時刻 t について、R-木を繰り返し検索することで、各時刻について上記の集合 S, Q を決定し、それを集計することで求める確率を決定する。この方式は、アルゴリズムは比較的単純であるが、R-木の検索数が T に比例するという性質があり[5-7]、移動軌跡を索引付けする時間が長くなれば長くなるほど、検索コストが高くなるという問題点がある。

5. 効率的な遷移確率推定アルゴリズム

次に効率的なアルゴリズムについて述べる。このアプローチでは、 $\text{Pr}(c_n|c_0, \dots, c_{n-1})$ の推定を以下の3ステップで行う。

- (1) 単位時間ごとに c_0, \dots, c_{n-1} に存在したオブジェクトの総数をR-木を用いてカウント
- (2) 単位時間ごとに c_0, \dots, c_n に存在したオブジェクトの総数をR-木を用いてカウント
- (3) ステップ2のカウント数 ÷ ステップ1のカウント数により $\text{Pr}(c_n|c_0, \dots, c_{n-1})$ を計算

R-木の探索はステップ(1), (2) でそれぞれ1回ずつ、合計2回しか行われなため、先の素朴な手法に比べ効率的な処理が達成できる。

各ステップでオブジェクトの集計処理を行う関数は R-木を用いた空間的制約の充足解の探索手法[3]を拡張したものである。このアプローチでは R-木上をルートからリーフ方向に時間的な制約を満たす解を探していく。その際、より効率的にR-木を探索するためにバックトラックや枝刈り条件によって解候補の絞込みを行い、充足解をみれなく探索する。セルに対する空間的制約も、解候補の絞込みに役立つ。

詳しいアルゴリズムについては、DBSJ Letters 論文[6]でも述べているため省略する。詳細については論文[5-7]を参照していただきたい。

6. 実験

本節では 移動オブジェクトのシミュレーションデータを用いて、素朴な遷移確率推定アルゴリズムと提案手法との処理時間の比較を行う。

6.1 実験用データ

実験ではBrinkoffにより作成された移動オブジェクトデータ生成ソフトウェア[1]を用いる。これは実際の市街地の道路ネットワーク上を自動車などが移動する際の移動の状況をシミュレーションするシステムであり、これによって生成された移動データを実験において使用する。今回扱うデータは、このシステムで提供されているドイツOldenburg市の中心部(約2.5×2.8km)の道路ネットワークにおいて、パラメータを適宜設定して生成したものである。図3にこのシステムのインタフェース画面を示す。地図上に見える丸い点が移動オブジェクトを示している。

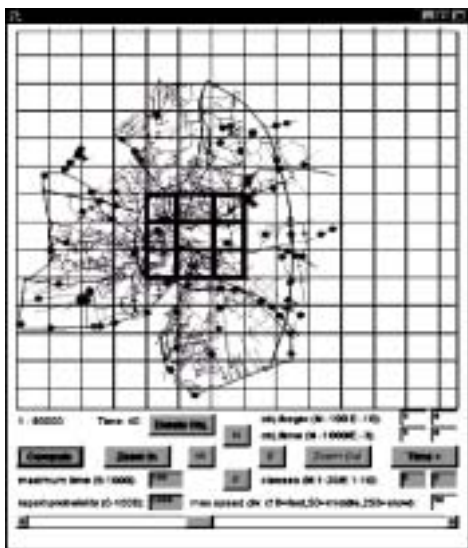


図3 データ作成ソフトウェア

移動データ生成では、初期移動オブジェクトの個数を5件とし、1分ごとに5件ずつ移動オブジェクトが地図上にランダムに生成・配置されるものとした。各移動オブジェクトには目的地が設定されており、目的地に達した場合は地図上から削除される。また、移動オブジェクトが地図領域の外に出た場合にも地図上から削除される。

その結果、定常状態では平均して100件強の移動オブジェクトが地図上を移動することとなった。本実験では、このようにして生成される多数の移動オブジェクトの移動状況を、1分刻みで $T=1000$ 分の期間に対して生成した。この結果、(オブジェクトID, 時刻, x座標, y座標)という形式のデータが124,752件生成された。これらのデータを3次元のR-木に登録した。

ここで問合せとして用いた、マルコフ遷移確率の推定問合せについて述べる。図に示すように、一部の空間をランダムに選び、一定サイズのセルに分割し、遷移確率の推定に要した計算時間を計測し比較する。図4では、ランダムに3×3の空間を選んだ時のセル領域と、導かれる遷移確率の例を示したものである。この場合は1次のマルコフ遷移確率の値を示している。

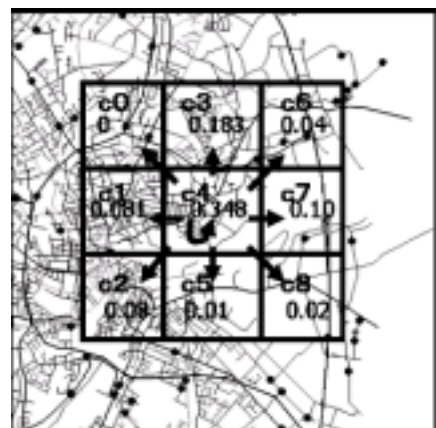


図4 マルコフ遷移確率の例

6.2 実験結果

図5は、空間を30×30のセルに分割し、ランダムに3×3のセルを選んだ時の1次の遷移確率(81通りの組み合わせ)の推定時間の比較を示す。縦軸には計算時間を示す。横軸は実験用データとして用意した $T=1000$ 分うち、何分までのデータに対して確率推定を行うかを示している。 $T=1000$ の時、用意したデータをすべて使用している。

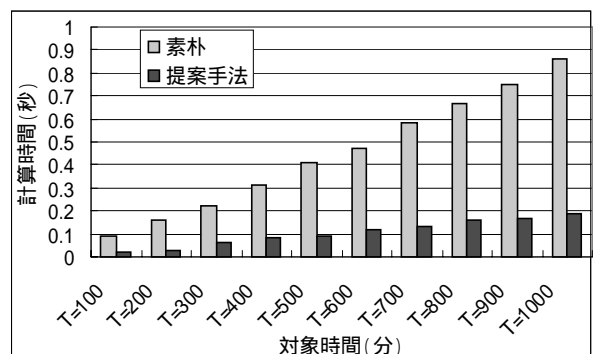


図5 3×3のセルによる1次のマルコフ遷移確率

図5から分かるように、提案手法の方が素朴な手法よりも、約4倍近い速度を示している。

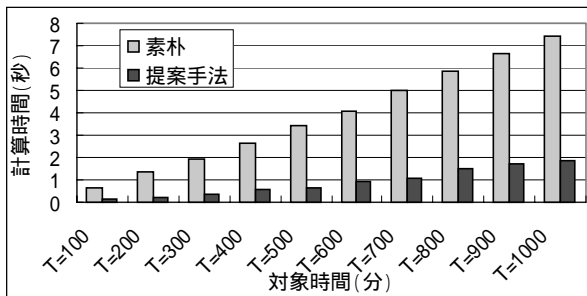


図6 3×3のセルによる2次のマルコフ遷移確率

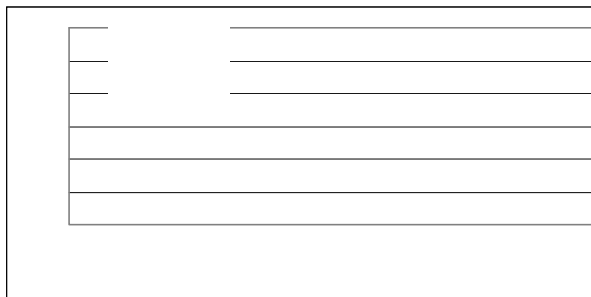


図7 3×3のセルによる3次のマルコフ遷移確率

図4, 5, 6が示すように、素朴なアルゴリズムに比べて効率的な遷移確率推定アルゴリズムの方が良い結果であることがわかる。

次に、効率的なアルゴリズムが不利な状況の例を示す。これは、空間を20×20のセルに分割し、ランダムに3×3のセルを選んだ時の2次の遷移確率（729通りの組み合わせ）の推定時間の比較である。

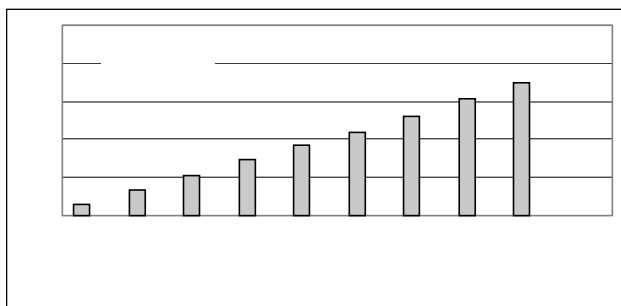


図8 3×3のセルによる2次のマルコフ遷移確率

この原因としては、セルの分割が粗いため、枝刈りをしても解候補をあまり絞り込むことができなかったというのが原因の1つとして考えられる。

今回の実験の詳細な分析で分かったことは、構築されたR-木の形が性能に大きな影響を及ぼすことである。今回の実装で用いたR-木の挿入アルゴリズムよりもさらに良いもの（例：R*-木）を使用すれば、さらに計算時間は短縮できると考えられる。また、入力データに応じた適当なセルの分割が重要であることも分かった。

7. まとめ

本稿では、時空間データベースの情報からマルコフ過程モデルに

基づく移動統計量を推定するためのアルゴリズムを実装し、有効性を示すための実験を行った。空間索引の内部情報を利用し、計数処理の効率化を図った点が特徴となっている。今後の課題として、更に詳しい実験や、1) データベース中のデータの分布に合わせたセルの適合的な分割、2) 非定常なマルコフ過程への拡張などが挙げられる。

【謝辞】

本研究の一部は、日本学術振興会科学研究費基盤研究(B)(12480067),若手研究(B)(14780316),および文部科学省科学研究費特定領域研究(14019009) 財団法人セコム科学振興財団研究奨励金による。

【文献】

- [1] Brinkhoff, "A Framework for Generating Network Based Moving Objects", *GeoInfomatica*, 6(2), pp. 153-180, 2002.
- [2] C.S. Jensen (ed.), "Special Issue: Indexing of Moving Objects", *IEEE Data Engineering Bulletin*, 25(2), Jun. 2002.
- [3] D. Papadias, N. Mamoulis, and Vasilis Delis, "Algorithms for Querying by Spatial Structure", *Proc. of VLDB*, 1998.
- [4] G.J.G. Upton and B. Fingleton, "*Spatial Data Analysis by Example, Volume II: Categorical and Directional Data*", John Wiley & Sons, 1989.
- [5] 石川, 塚本, 北川, 索引付けされた移動軌跡データからの移動統計量の抽出法について, 第14回データ工学ワークショップ (DEWS 2003), 2003年3月.
- [6] 塚本, 石川, 北川, 索引付けされた移動軌跡データからの効率的な移動統計量抽出法, 日本データベース学会Letters Vol.2, No.1, 2003年5月.
- [7] 塚本, 石川, 北川, 索引付けされた移動軌跡データからの移動統計量の抽出法の評価, 情報処理学会研究報告, 2003-DBS-131-30, 2003年7月

塚本 祐一 Yuichi TSUKAMOTO

筑波大学大学院システム情報工学研究科在学中 2002年筑波大学第三学群情報学類卒業。時空間データベースの研究に従事。日本データベース学会学生会員。

石川 佳治 Yoshiharu ISHIKAWA

筑波大学電子・情報工学系講師。1994年筑波大学大学院博士課程工学研究科単位取得退学。博士(工学)。空間データベース、文書データベース、情報検索などに興味を持つ。ACM, IEEE-CS, 情報処理学会, 電子情報通信学会, 日本データベース学会, 各会員。

北川 博之 Hiroyuki KITAGAWA

筑波大学電子・情報工学系教授。1980年東京大学大学院理学系研究科修了。理学博士(東京大学)。異種情報源統合、文書データベース、WWWの高度利用等の研究に従事。著者「データベースシステム」(昭晃堂), 「Unnormalized Relational Data Model」(共著, Springer-Verlag)等。ACM, IEEE-CS, 日本データベース学会, 情報処理学会, 電子情報通信学会, 日本ソフトウェア科学会, 各会員。