

移動オブジェクトデータベースに 関する研究動向

石川 佳治

筑波大学大学院システム情報工学研究科

筑波大学計算科学研究センター

ishikawa@cs.tsukuba.ac.jp



概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 我々の研究



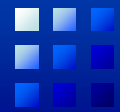
背景：位置ベースのサービス

- はじまり(US)
 - 911サービス
 - FCC(Federal Communications Commission)
 - 携帯キャリアに話者の位置特定機能を義務付け(1996年)
 - GPSを一般に開放：5～50メートルの精度
- 応用例
 - Mobile Resource Management (MRM)
 - 車両管理, 交通管理
 - Location-aware Content Delivery
 - 道路情報, 近隣情報
 - Digital Battlefield
 - 「10分以内にこの地域に到達できるヘリは？」



背景：移動体通信

- ユーザとセルの対応を管理：(key, cell-id)
- 2種類の基本処理
 - **point query**：電話の際に発生
 - 「キー123-456-789 をもつ移動オブジェクトの現在位置(セル)を求めよ」
 - **point update**：移動の際に発生
 - 「キー123-456-789の移動オブジェクトの位置(セル)を更新せよ」
 - 分散・複製・キャッシュ管理がキーポイント
- 一種の移動オブジェクトデータベース



移動オブジェクトデータベース(1)

- Moving Objects Database (MOD)
- 多様な問合せ
 - 空間データベースの拡張
 - 範囲問合せ (range query): 指定された範囲内のオブジェクトを検索
 - 「(10分以内に)つくば市を通過する移動オブジェクトを検索」
 - 最近傍問合せ (nearest neighbor query): 指定された点に最も近いオブジェクトを検索
 - k -最近傍問合せ (k-nearest neighbor query): 近いものから k 個
 - 結合問合せ (join query)
 - 「今後領域がオーバーラップする移動オブジェクトのペアを列挙せよ」
 - 「10分以内に10kmの距離に接近する飛行機のペアを列挙せよ」
 - closest-pair query
 - 「最接近時点での距離が近い順に k 個のオブジェクトのペアを示せ」
 - 類似検索 (similarity query)
 - 「オブジェクト o と類似した移動軌跡を持つオブジェクトは？」



移動オブジェクトデータベース(2)

- 多様な問合せ(続き)
 - 時間の扱いによる分類
 - past query
 - 「先月〇〇停留所で5分以上遅れたバスの件数は？」
 - set-oriented present query
 - 「ここから1km以内にタクシーをすべて求めよ」
 - future query
 - 「10分後のトラック#5の位置を予測せよ」
 - set-oriented future query
 - 「20分以内に目的地に到達するトラックをすべて求めよ」
 - アドホックな問合せ/連続的問合せ
 - trigger
 - 「目的領域にヘリが到達したらメッセージを私まで送信せよ」
 - incremental range/nearest neighbor query
 - 前回の検索結果を適宜更新



関連する研究領域

- 時空間データベース
 - 空間データベースの延長上
 - オブジェクトを点, 移動軌跡をベクトルなどでモデリング
 - 問合せ処理, 索引, 統計処理, データマイニング
- モバイルデータベース
 - モバイル環境＋データベース
- ユビキタスデータベース
- センサデータベース
- 時系列データベース



MODの応用(1)

- 時空間情報の検索
 - 「移動経路で2キロ以内に1泊1万円以内のホテルがあったら通知せよ」
 - location dependent query
- 交通管理
 - 「A地点に最も早く到達できるタクシーは？」
 - 「B地点への到達時間は？」
- モバイルEコマース
- Mobile workforce management
 - 「春日3-2-1の近くにいるサービスクルーは誰？」
- Digital battlefield
 - 「領域X内に味方の戦車は何台いるか？」



MODの応用(2)

- 位置とコンテキストに応じた情報提示
- 航空管理
 - 「将来衝突の可能性がある飛行機のペアがあるか？」
- 移動体通信のサポート
- モバイル環境の支援
 - mobile ad-hoc network (MANET): 位置情報を補うことによる効率化
- 商用ベースの応用
 - Mobitrac <http://www.mobitrac.com/>
 - Qualcomm <http://www.qualcomm.com/>
 - @Road <http://www.atroad.com/main.html>



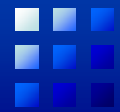
研究課題

- データモデル
- 索引手法
- 問合せ処理
- 位置ベースの情報サービス
- 時空間データマイニング



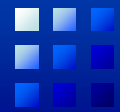
概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 我々の研究



研究事例: DOMINOプロジェクト

- Databases fOr MovINg Objects
- Wolfsonら (U. Illinois)
- 特徴
 - オブジェクトの現在, 過去, 未来の位置を追跡
 - 出発地・目的地情報を利用
 - 動的属性
 - 複雑な問合せ
 - DBMS上に実装
 - あいまい性を考慮



研究事例: CHOROCHRONOSプロジェクト

- 欧州の研究者を中心としたプロジェクト
 - 1996.8～2000.7
 - 移動オブジェクトデータベースを含む時空間データベースに関する研究
- 研究課題
 - オントロジー
 - モデルと言語: データ型, 問合せ言語
 - インタフェース
 - 問合せ処理
 - 索引手法, 格納手法
 - アーキテクチャ



CHOROCHRONOSの概要

- 時空間データのオントロジー
 - 時空間オントロジーと時空間DBの関連
- 時空間DBMSのためのモデルと言語
 - 時空間情報のためのデータ型
 - モデルと問合せ言語
 - 時空間データベースの設計手法
- 時空間データベースのインタフェース
 - GUI技術
 - 可視化
- 問合せ処理
 - 効率的な問合せ評価
 - 最適化処理
- 格納構造, 索引手法



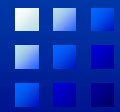
概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 我々の研究



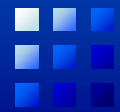
MODのデータモデル:分類

- 表現方式
 - 点データ, および点データの集まり
 - 移動軌跡(polyline)
 - 抽象データ型
 - 制約データモデル
- 問合せ対象
 - 過去の移動情報
 - 遡及的な問合せ・分析
 - 現在および今後の移動情報
 - リアルタイムの応答
 - 予測に基づく処理
 - 連続的問合せ
- あいまい性・不正確さの表現



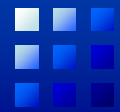
移動情報の表現方式: 点データ

- 位置を (l, t) のペアで管理
 - l : 位置 (座標値またはセルIDなど)
 - t : 時刻
 - DBMSによる管理が容易
- 問題点
 - 内挿, 外挿 (予測) が難しい
 - 移動軌跡の情報が欠落



移動情報の表現方式：移動軌跡データ

- 移動軌跡 (trajectory) を直接的に表現
 - 線分 (line segment) をつないだ 多角線 (折れ線, polyline) の表現が一般的
 - 例: $(x_1, y_1, t_1), (x_2, y_2, t_2), \dots$
 - 移動速度も (近似的に) 表現可能
 - 点データに比べ情報量大
- 問題点
 - 特別のアルゴリズム・データ構造が求められる
 - データ量



抽象データ型のアプローチ(1)

- Gütingらによる
- CHOROCHRONOSプロジェクトの一環
- 時空間データのためのデータ型を開発
 - 移動する点, 領域のためのデータ型
 - それらデータ型に基づく問合せ言語
- アプローチ
 - 時間の関数として時空間データを表現
 - リレーショナルモデルおよびSQLに埋め込み
 - 抽象化のレベル



抽象データ型のアプローチ(2)

- Moving Entities
 - 時刻をパラメータとする関数でモデル化

Moving point

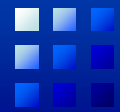
$\underline{mpoint} = \underline{time} \rightarrow \underline{point}$

Moving region

$\underline{mregion} = \underline{time} \rightarrow \underline{region}$

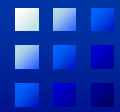
オペレーションの例:

inside: $\underline{mpoint} \times \underline{mregion} \rightarrow \underline{mbool}$ (a “moving boolean”)



抽象データ型のアプローチ(3)

- 抽象のレベル: moving pointを例に
 - time → point の関数
 - 3次元空間の多角線
 - 抽象的
 - 具象的
 - 連続的
 - 離散的
 - 無限の表現
 - 有限の表現
 - 概念レベルのモデル
 - 実装レベルのモデル



抽象データ型のアプローチ(4)

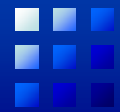
routes (airline: string, id: string, flight: mpoint)

–「500km未満の距離のJALのフライトは？」

- SELECT airline, id
FROM routes
WHERE airline = "JAL" AND length(trajjectory(flight)) > 5000

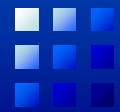
–「500メートル以内に接近するフライトのペアを求めよ」

- SELECT p.airline, p.id, q.airline, q.id
FROM routes p, routes q
WHERE val(initial(atmin(distance(p.flight, q.flight)))) < 0.5



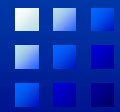
制約データモデルの利用(1)

- 制約論理式で表現
 - 制約データベース (constraint database) の枠組みを利用
 - 一般に線形制約 (linear constraint) を利用
 - 計算量の面で有利
- 理論的背景が存在, 多次元でも成立
- 計算量は一般に大
- 例: 移動軌跡の表現
 - $x = [2, -1, 0]^T t + [-40, 23, 30]^T \wedge 0 \leq t \leq 21$
 $\vee x = [0, -1, -5]^T t + [2, 23, 135]^T \wedge 21 \leq t < 22$
 $\vee x = [0.5, 0, -1]^T t + [-9, 1, 47]^T \wedge 22 \leq t$



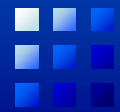
制約データモデルの利用(2)

- 問合せの例: 時刻 τ_1 から τ_2 の間につくば市の領域内に入った移動オブジェクトは?
 - $\{y \mid O(y) \wedge \exists (\tau_1 \leq t \leq \tau_2 \wedge \exists x (T(y, t, x) \wedge \varphi(x)) \wedge \exists t' (t' < t \wedge \forall t'' (t' < t'' < t \rightarrow \neg \exists x (T(y, t'', x) \wedge \varphi(x))))))\}$
 - $O(y)$: y が移動オブジェクトであることを表す述語
 - $T(y, t, x)$: 時刻 t におけるオブジェクト y の位置が x であることを表す述語
 - $\varphi(x)$: x がつくば市の領域内にあるとき真になる述語
 - 線形制約では書けないので, 外部述語として表現
 - 問合せ処理時には呼び出すだけ



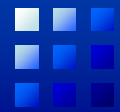
制約データモデルの利用(3)

- 問合せ処理
 - 問合せの制約式とある移動オブジェクト o に対する制約式を $\wedge$ でつないだ論理式を構築
 - その論理式に**充足解**があれば o は問合せを満たす
 - 線形計画法を利用
 - 1オブジェクトに対し多項式時間の処理
 - **限量子の除去** (quantifier elimination) : $\forall, \exists$ を取り除く
 - MOD中のすべての移動オブジェクトへの適用はコスト大
 - 線形制約でなければさらに計算量大



制約データモデルの利用(4)

- DEDALE
 - 制約データモデルに基づく時空間データベース
 - INRIAで開発
 - 線形制約を採用
 - トポロジカルな問合せを処理可能
 - 最適化に関する研究



時制論理の利用(1)

- Future Temporal Logic (FTL)
 - DOMINOで利用: 現在の情報をもとに未来を推論
- 述語の例
 - $f \text{ until } g$
 - g が真であるか, g が真となる直前まで f が真であるときに真となる
 - $\text{nexttime } f$
 - 次の時点で f が真となる
 - $\text{eventually } f (= \text{true until } f)$
 - f は将来的に真になる
 - $\text{always } f (= \neg \text{eventually } \neg f)$
 - 現在かつ将来において f は真



時制論理の利用(2)

- 述語の例(続き)
 - **eventually_within_c(g)**
 - 以降の c 単位時間の間に g は真となる
 - **eventually_after_c(g)**
 - 少なくとも c 単位時間の後には g は真となる
 - **always_for_c(g)**
 - 以降の c 単位時間の間 g はつねに真
 - **g until_within_c h**
 - 以降の c 単位時間の間に h が真でその直前まで g が真であるような状況がある
- 他に空間述語 (DISTANCE(o , P))なども提供



時制論理の利用(3)

- 以降の5単位時間の間に, ポリゴンP内に入るタクシーをすべて検索
 - RETRIEVE o
WHERE o.type = "Taxi" AND
eventually_within_5 INSIDE(o, P)
- 以降の5単位時間の間にポリゴンP内に入り, しかも3単位時間以上はP内に存在しているオブジェクトをすべて検索
 - RETRIEVE o
WHERE eventually_within5 (INSIDE(o, P)
AND always_for_3 INSIDE(o, P))



時制論理の利用(4)

- 以降の3単位時間の間に, ポリゴンPの中に入り2単位時間以上存在し, 少なくとも6単位時間後にはポリゴンO内にあるオブジェクトをすべて検索せよ
 - RETRIEVE o
WHERE eventually_within_3
[INSIDE(o, P) AND
always_for_2 (INSIDE(o, P) AND
eventually_after_6 INSIDE(o, O))]



あいまい性・不確かさの扱い

- 位置情報にはあいまい性が存在
 - あいまい性を減らすにはコストが発生（例：データ取得頻度の増加）
 - ある程度の不正確さを許容してもよい
 - may / must のセマンティクス，確率（存在確率，信頼確率）の提示
- コストの考慮
 - 更新コストと不正確さのコストのトレードオフ
 - 要求と対象に合った更新ポリシー



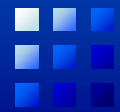
概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 我々の研究



MODの索引:分類

- 支援するデータモデル
 - 点データ
 - 移動軌跡
- 問合せ対象
 - 過去の移動情報
 - 現在および今後の移動情報
 - 連続的問合せ
- 更新の有無
- データ構造
 - 木構造: R-木, 四分木 (quad-tree) の拡張
 - その他 (空間分割, ハッシュなど)



空間索引の利用(1)

- 例: R-木
 - 多次元空間の点 (x, y, t) で点を表現
 - 移動軌跡 (polyline) の索引にも拡張可能
 - 頻繁な更新には向かない



空間索引の利用(2)

- R-木に基づくアプローチ
 - 初期のアプローチ
 - 時間をもう一つの次元と見る
 - 3D R-tree, RT-tree, 2+3 R-tree
- バイテンポラルなR-木
 - valid time と transaction time を扱うR-木
- Time-parameterized R-tree (TPR-tree)
 - 後述
 - TPR*-tree という拡張あり
- Trajectory bundle
 - 過去の移動軌跡(polyline)の検索に利用



空間索引の利用(3)

- 四分木に基づくアプローチ
 - MOF-tree
 - 時間的に連続した画像を扱うquad-tree
 - MOF+-tree
 - MOF-treeの拡張
 - Multiversion linear quadtree
 - 画像の時間的なシーケンスを格納
 - multiversion B-tree と linear region quadtree の組合せ
 - Overlapping linear quadtree
 - linear region quadtree の拡張



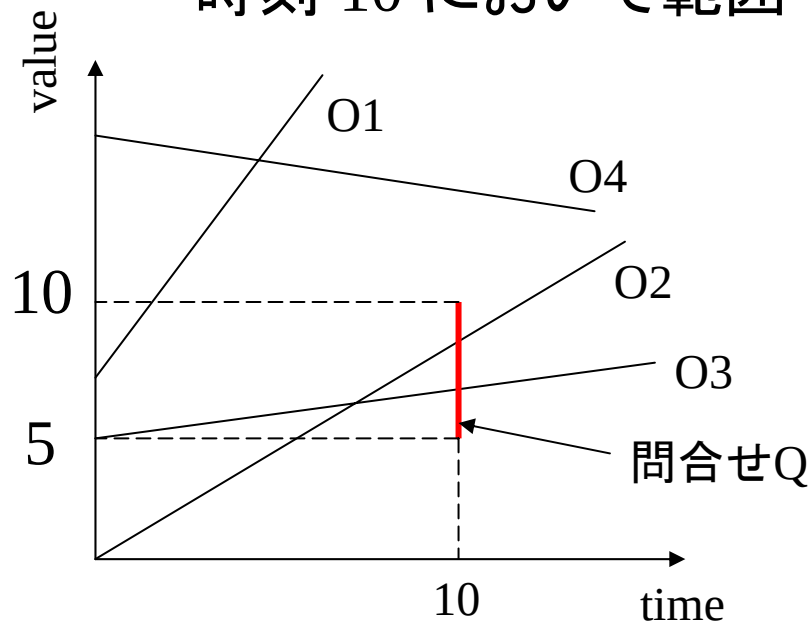
DOMINOの索引(1)

- アプローチ
 - 動的属性を幾何的に表現
 - 空間索引の利用
- 複数の表現形式
 - 値と時刻による表現 (value-time space)
 - 問合せ処理は効率的
 - 更新はあまり効率的でない
 - 切片と傾きによる表現 (intercept-slope space)
 - いわゆるハフ(Hough)変換
 - 問合せ処理はあまり効率的でない
 - 更新処理は効率的

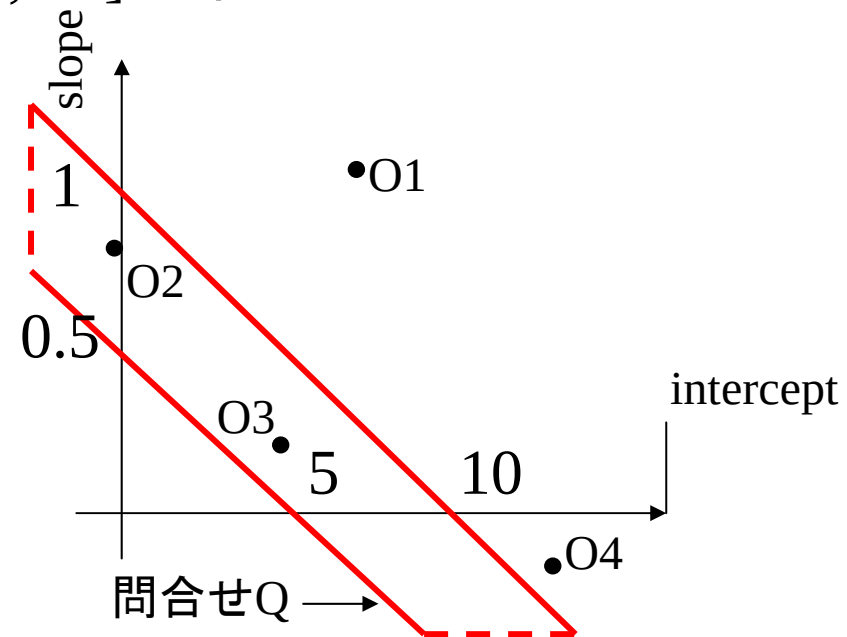


DOMINOの索引(2)

- 範囲問合せの例
 - 1次元空間の移動オブジェクト
 - 時刻 10 において範囲 [5, 10] に含まれるオブジェクトは?

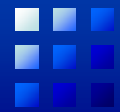


value-time space
における表現



slope-intercept space
における表現

d 次元空間の直線 $\Rightarrow$ $2d$ 次元空間の点



問合せの種類

- **タイプ1**: 時刻 t において与えられた矩形と交わるオブジェクトを検索
- **タイプ2**: 時区間 $[t_1, t_2]$ において与えられた矩形と交わるオブジェクトを検索
- **タイプ3**: 時区間 $[t_1, t_2]$ において与えられた移動矩形と交わるオブジェクトを検索



TPR-木(1)

- Time-Parameterized R-tree
 - R-木の拡張
- 移動する点データを **時刻をパラメータとする矩形で包囲**
 - 現在(**now**)を基点とする
 - 包囲矩形は時間とともに拡大(or 縮小)・移動
- 包囲矩形が小さくなるよう木を構築
 - 位置・速度が「近い」オブジェクトをグルーピング
 - 木の更新の際に再調整
- アルゴリズム
 - 検索: R-木がベース
 - 構築・更新: 大幅な拡張



TPR-木(2)

- velocity bounding rectangle (VBR)
 - R-木の範囲矩形(bounding rectangle)に対応
 - 位置と速度により移動オブジェクトをグルーピング
- できるだけ包囲矩形の「サイズ」を小さくしたい
- アプローチ: $[now, H]$ の時区間における包囲矩形の面積 $\int_{now}^{now+H} A(t)dt$ を最小化
 - H : 将来の修正時刻



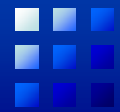
概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 我々の研究



MODにおける最近傍探索(1)

- 1-NN (nearest neighbor) に対するアプローチ
 - ボロノイ図を利用
 - いつまで検索結果が保証できるかも答える
 - 制約
 - 最大速度を仮定
 - 1-NNに限定
 - ボロノイ図の構築が必要
- 連続的な k -NN 問合せ
 - オブジェクトが q から q' に移動
 - 前回の問合せ結果をもとに, できるだけコストの低い問合せをしたい
 - 例: q に対する 2-NN が a , 4-NN が b のとき, $2 \text{ dist}(q, q') \leq \text{dist}(q, b)$
 - $\text{dist}(q, a)$ ならば, q' の 2-NN は q の 4-NN より近くにある



MODにおける最近傍探索(2)

- Time-parameterized NN query
 - 移動オブジェクトは一定速度で一方向に進む
 - 1-NNとその有効期限, 次の1-NNを連続的に提示
 - TPR-木を用いる拡張あり
 - 問合せ結果
 - 1-NN: i
 - 有効時間: 2
 - 次の1-NN: j
 - influence timeを用いた処理
 - 現在のNN(i)に対し, 各オブジェクトがいつNNに置き換わるか
 - influence time最小のオブジェクトが次のNN



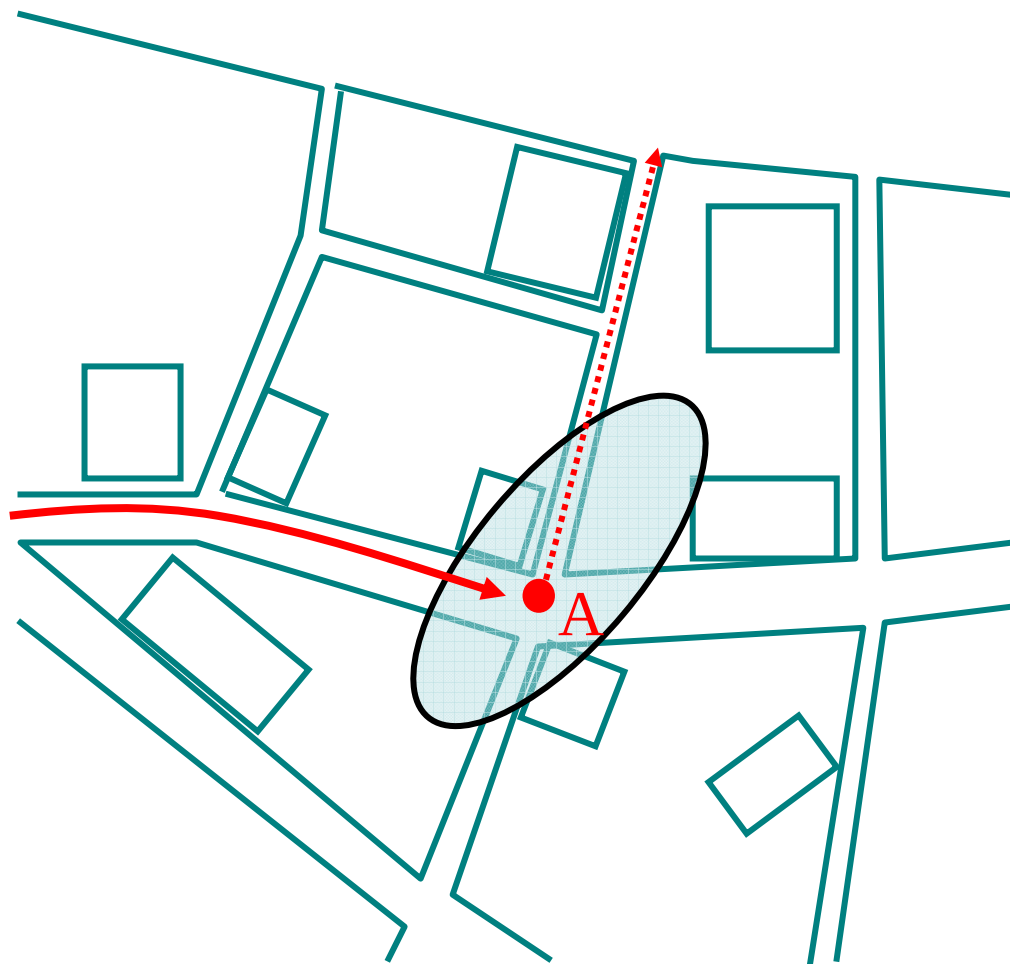
概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 時空間データマイニング
- 我々の研究
 - 移動オブジェクトに対する適応的情報提供
 - 大量の移動軌跡からの移動統計量抽出法
 - 移動ヒストグラムの構築・管理手法

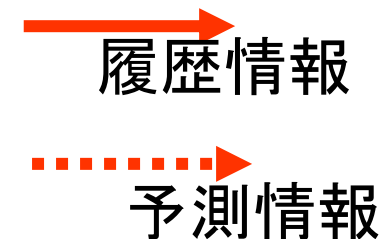


移動オブジェクトに対する適応的情報提供

利用者の移動経路(履歴情報・予測情報)を利用

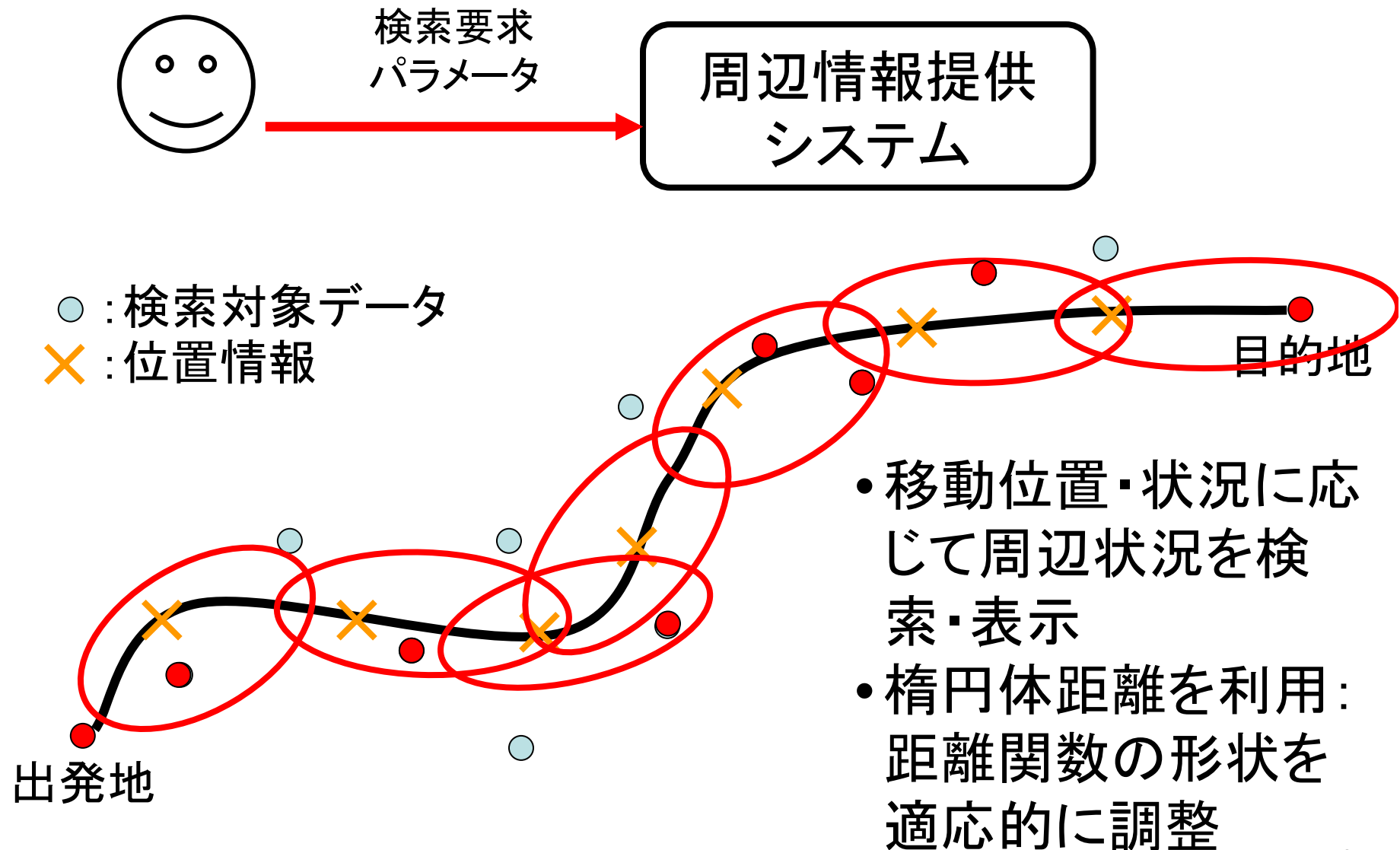


- ・例: 移動する各地点において周辺にあるコンビニを探したい
- ・移動状況(経路, 速度)を考慮した適応型の情報提供をしたい
- ・アプローチ: 楕円状の距離関数を利用





手法の概要





プロトタイプシステム

検索パラメータ設定

現在位置

問合せ中心

検索されたデータ

地図および検索結果を表示

検索結果表示

検索パラメータ設定

問い合わせ
☒ 範囲問い合わせ
☐ K近傍問い合わせ

問い合わせ中心
☐ 現在位置
☐ 現時点の着目地点
☒ 重み付け平均位置

重み付け
過去 0.2
未来 0.4
時刻 1

検索範囲 80
HBパラメータ 0.9

OK Cancel

Selected Features

theme: Pointdata1.shp

Shape	ID	type	X-coord	Y-coord
Point	1	ガソリンスタンド	535.08836	216.60
Point	2	その他	557.04981	230.5
Point	0	コンビニ	642.89913	188.6
Point	1	ガソリンスタンド	644.89562	216.6
Point	2	その他	600.97272	282.4
Point	0	コンビニ	596.97973	186.6

・プロトタイプシステムはArcView上に実現



概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- 時空間データマイニング
- **我々の研究**
 - 移動オブジェクトに対する適応的情報提供
 - **大量の移動軌跡からの移動統計量抽出法**
 - 移動ヒストグラムの構築・管理手法

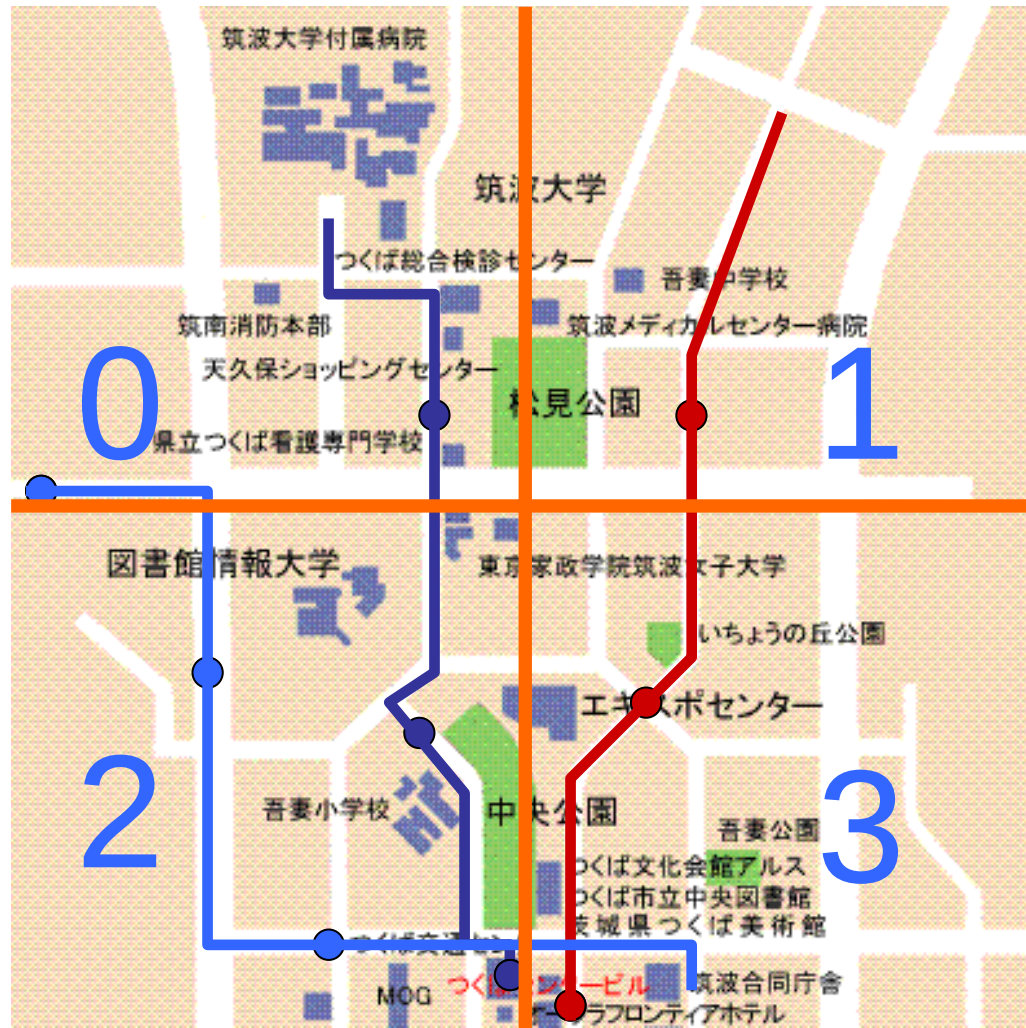


大量の移動軌跡からの移動統計量抽出法

- 目的
 - 大量の移動オブジェクトの移動情報を要約
 - 移動分析(mobility analysis), 移動予測シミュレーション
- アプローチ:
 - 空間的なマルコフ遷移確率の推定
 - 対象空間をセルの集合に分割
 - 移動軌跡データをもとにセル間の遷移確率を導出
 - 制約充足問題(CSP)として定式化
 - 空間索引の利用
 - R-木に移動軌跡データが蓄積
 - 空間索引の効率的な探索



移動状況データ(1)



オブジェクトA
の移動状況

2→2→0→0

オブジェクトB
の移動状況

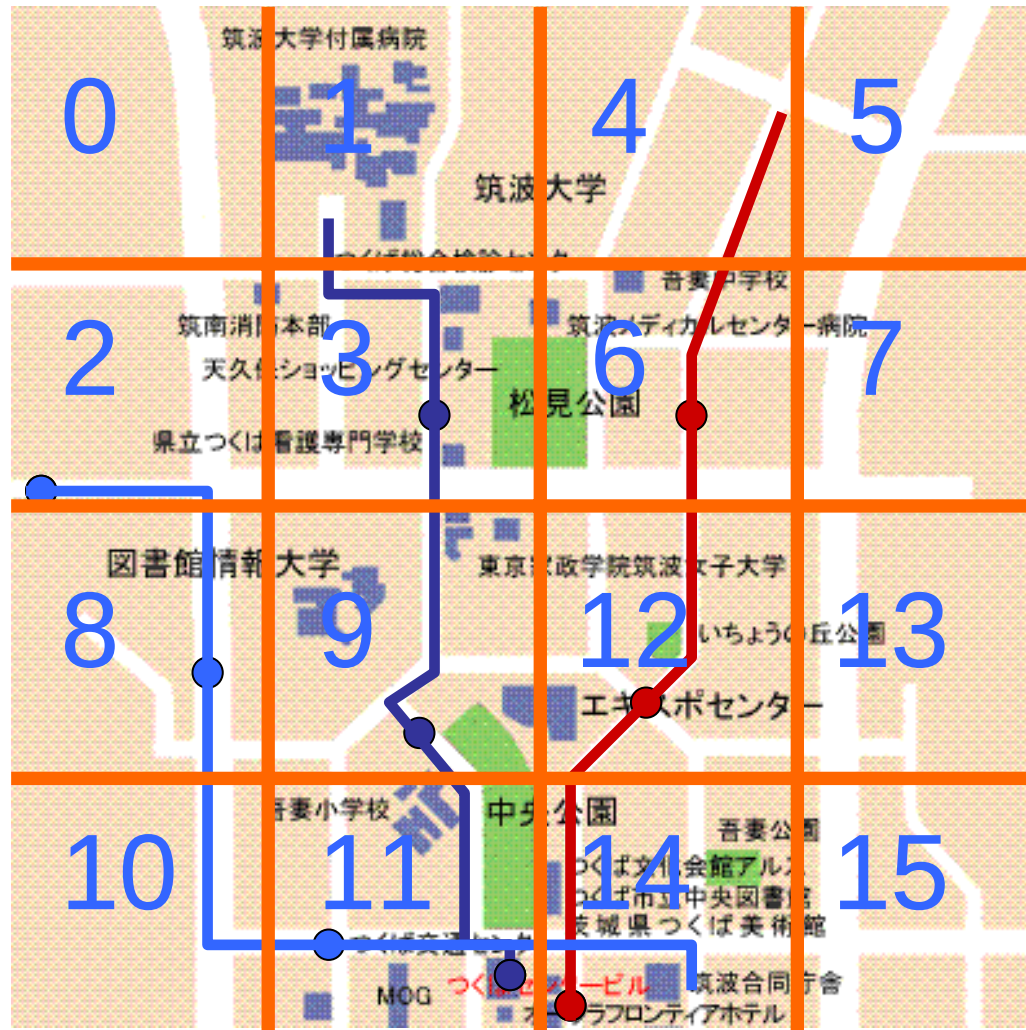
3→3→1→1

オブジェクトC
の移動状況

0→2→2→3



移動状況データ(2)



オブジェクトA
の移動状況

11→9→3→1

オブジェクトB
の移動状況

14→12→6→4

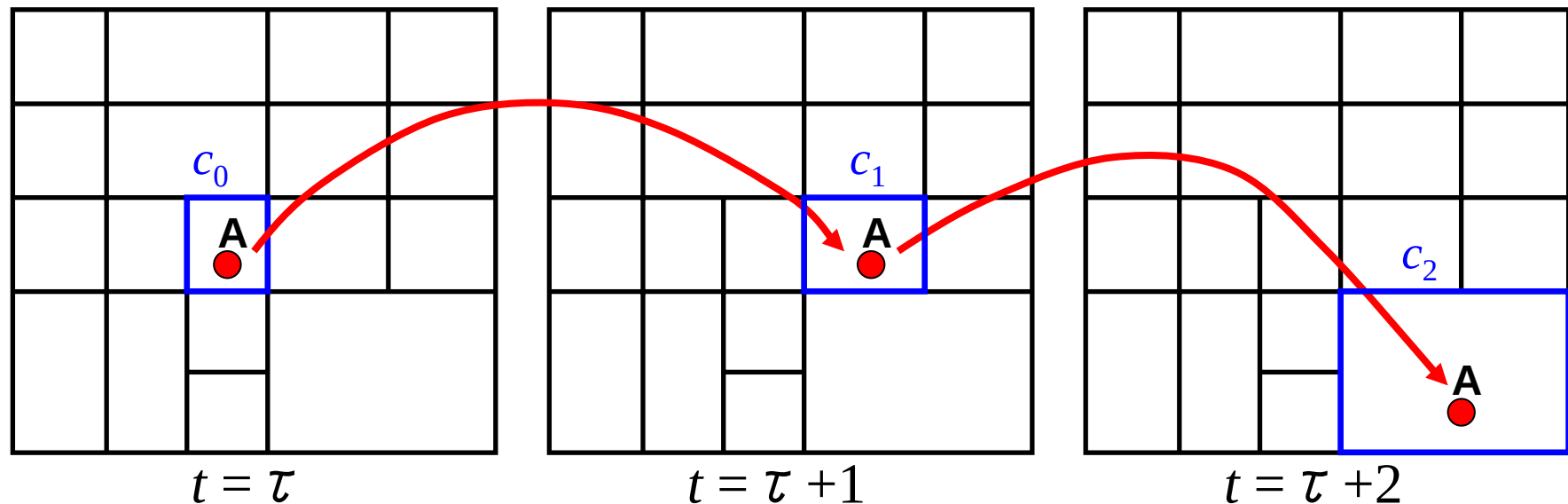
オブジェクトC
の移動状況

2→8→11→14



マルコフ遷移確率(1)

- 例: セル c_0 から c_1 へ単位時間で移動したオブジェクトが, 単位時間後に c_2 に存在する確率は?



- 2次の遷移確率 $\Pr(c_2|c_0, c_1)$
- n 次のマルコフ遷移確率 $\Pr(c_n|c_0, \dots, c_{n-1})$ への拡張は容易



マルコフ遷移確率(2)

- 移動分析(交通工学)における古典的手法
- 特殊な相関ルール
 - 遷移確率が信頼度(confidence)に対応
 - 順序が存在
- 適用分野
 - 移動経路の推定
 - 移動オブジェクトの将来の移動先を推定
 - 移動状況のシミュレーション
 - 時刻 $t = \tau$ における移動状況が与えられたとき, $t = \tau + 1, \tau + 2, \dots$ における変化を推定する

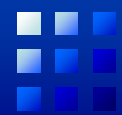


確率の定式化

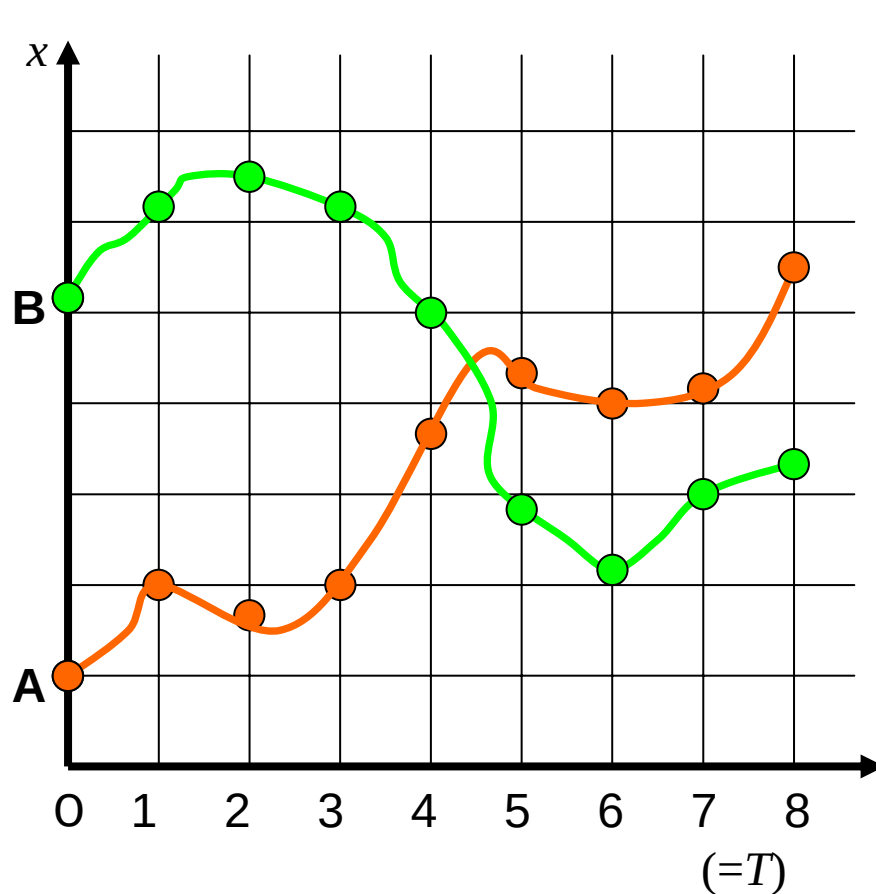
- 対象データ: $t = 0$ から $t = T$ の移動軌跡
- n 次のマルコフ遷移確率の定義

$$\Pr(c_n | c_0, K, c_{n-1}) = \frac{\sum_{t=0}^{T-1} |I_{i=0}^n \text{objs}(c_i, t+i)|}{\sum_{t=0}^{T-1} |I_{i=0}^{n-1} \text{objs}(c_i, t+i)|}$$

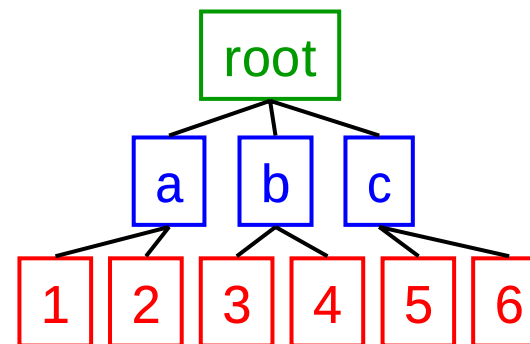
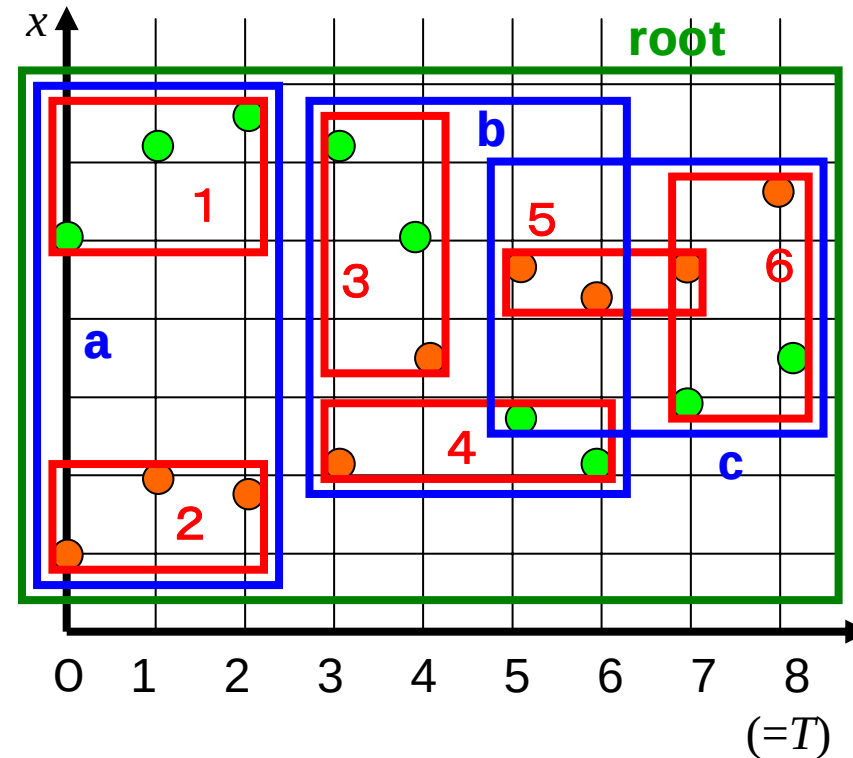
- $\text{objs}(c_i, t)$: 時刻 t で c_i にいたオブジェクトの集合
- 分母: 時刻 t ($0 \leq t \leq T-1$) においてセル c_0 にいて, $t+1$ では c_1 にいて, ..., 時刻 $t+n-1$ では c_{n-1} にいたオブジェクトの数
- 分子: 分母に含まれ, かつ, 時刻 $t+n$ で c_n に移動したオブジェクトの数



移動軌跡の表現: $(d + 1)$ -次元のR-木

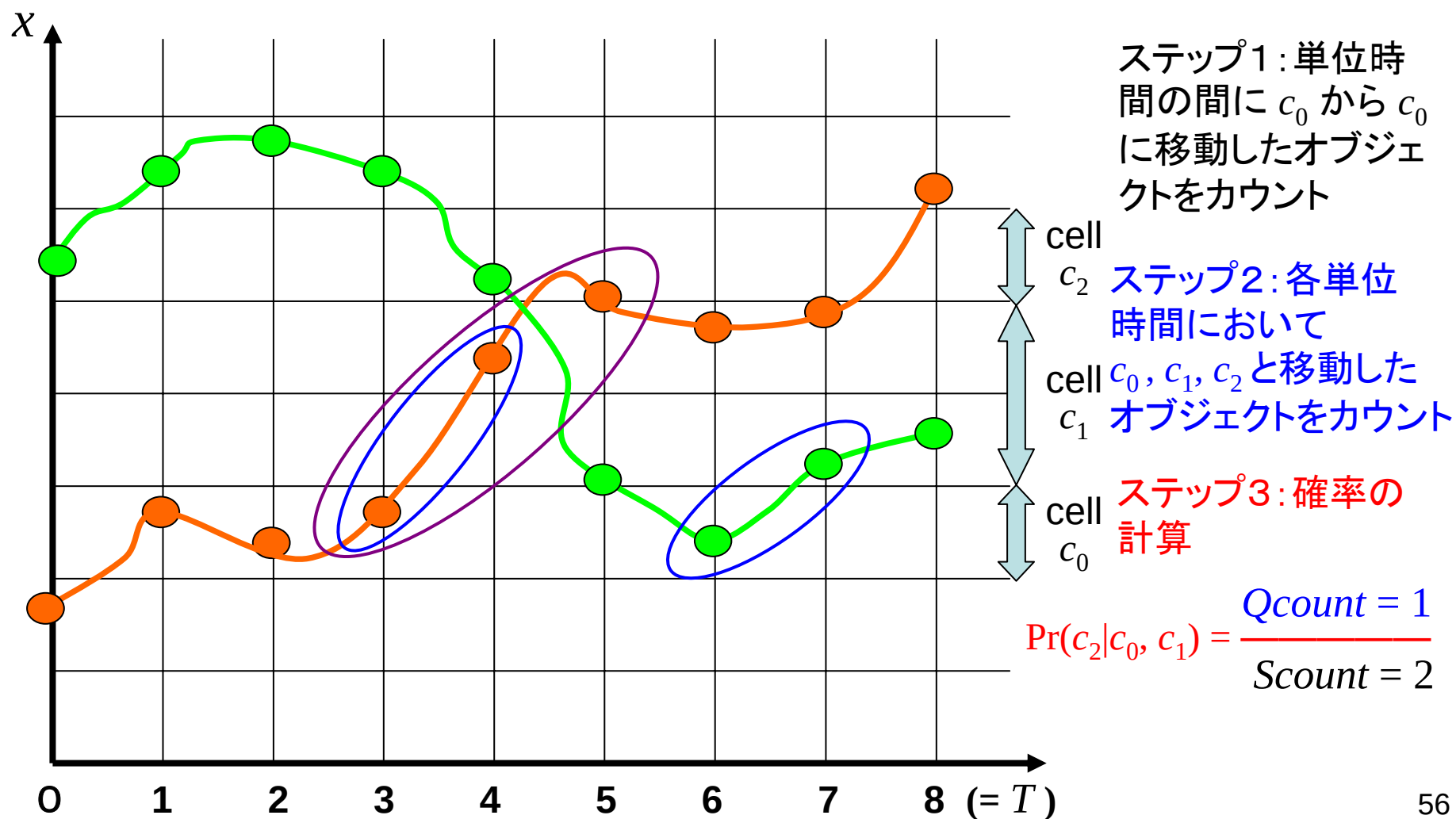


サンプリングに基づく表現



制約充足問題(CSP)に基づく処理

例: $\Pr(c_2|c_0, c_1)$ の推定





R-木を用いたカウント処理

- アイデア: 問題を制約充足問題 (constraint satisfaction problem, CSP) として定式化
- 制約を満たすオブジェクトは, ある τ について以下の制約をすべて満たす
 - $t = \tau$ においてセル c_0 にいた
 - $t = \tau + 1$ においてセル c_1 にいた
 - ...
 - $t = \tau + n$ においてセル c_n にいた
- $n + 1$ 個のすべての制約条件を満たすオブジェクトを探索
- R-木による空間制約充足解の探索手法 (Papadias, VLDB98) を拡張

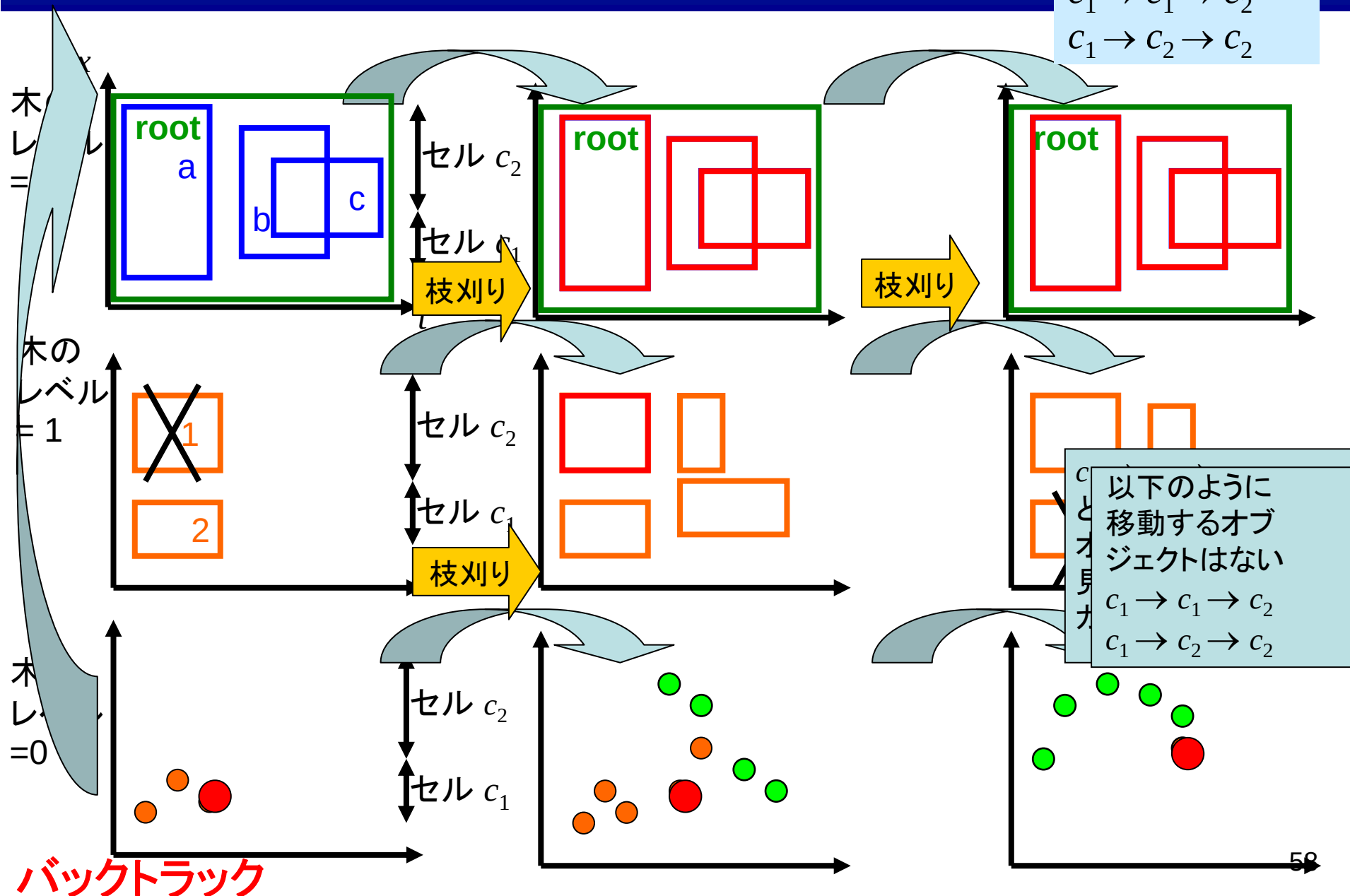


問合せ処理の例

ターゲット

$C_1 \rightarrow C_1 \rightarrow C_2$

$C_1 \rightarrow C_2 \rightarrow C_2$





概要

- 移動オブジェクトデータベース(MOD)とは
- 研究事例: DOMINOプロジェクト
- MODのデータモデル
- MODの索引
- 問合せ処理
- **我々の研究**
 - 移動オブジェクトに対する適応的情報提供
 - 大量の移動軌跡からの移動統計量抽出法
 - **移動ヒストグラムの構築・管理手法**

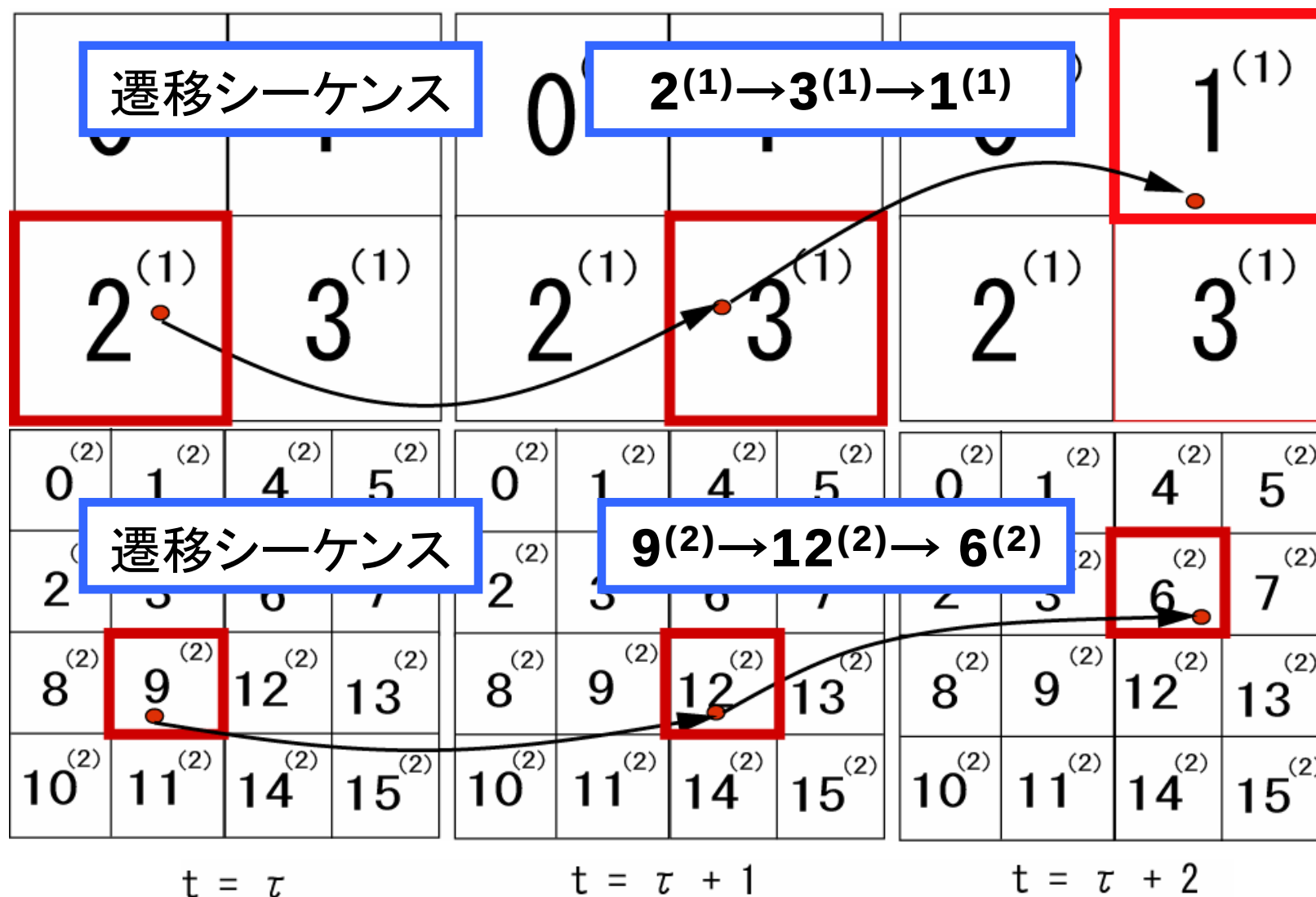


移動ヒストグラム (mobility histogram)

- 移動パターンをコンパクトに要約する一種のヒストグラム
- 与えられた移動パターンに対し、予測される出現数(選択率)を返す
- マルコフ連鎖モデルを用いた移動パターンの表現
- 表現
 - 論理レベル: データキューブ
 - 物理レベル: 木構造



マルコフ連鎖モデル

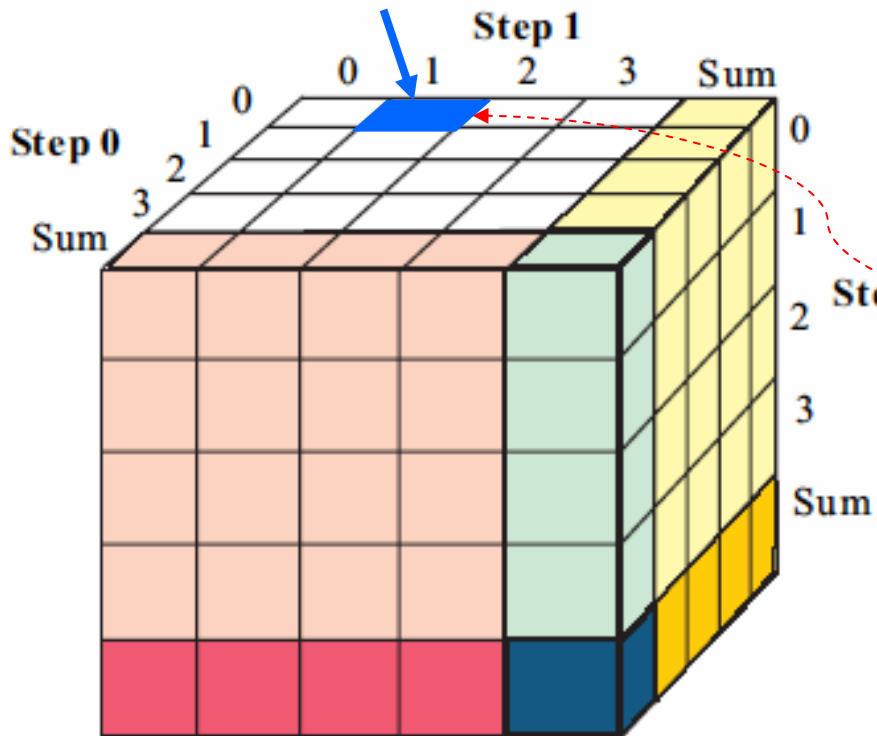


マルコフ連鎖モデルの概念(2次のマルコフ連鎖)

ステップ0のセル → ステップ1のセル → ステップ2のセル

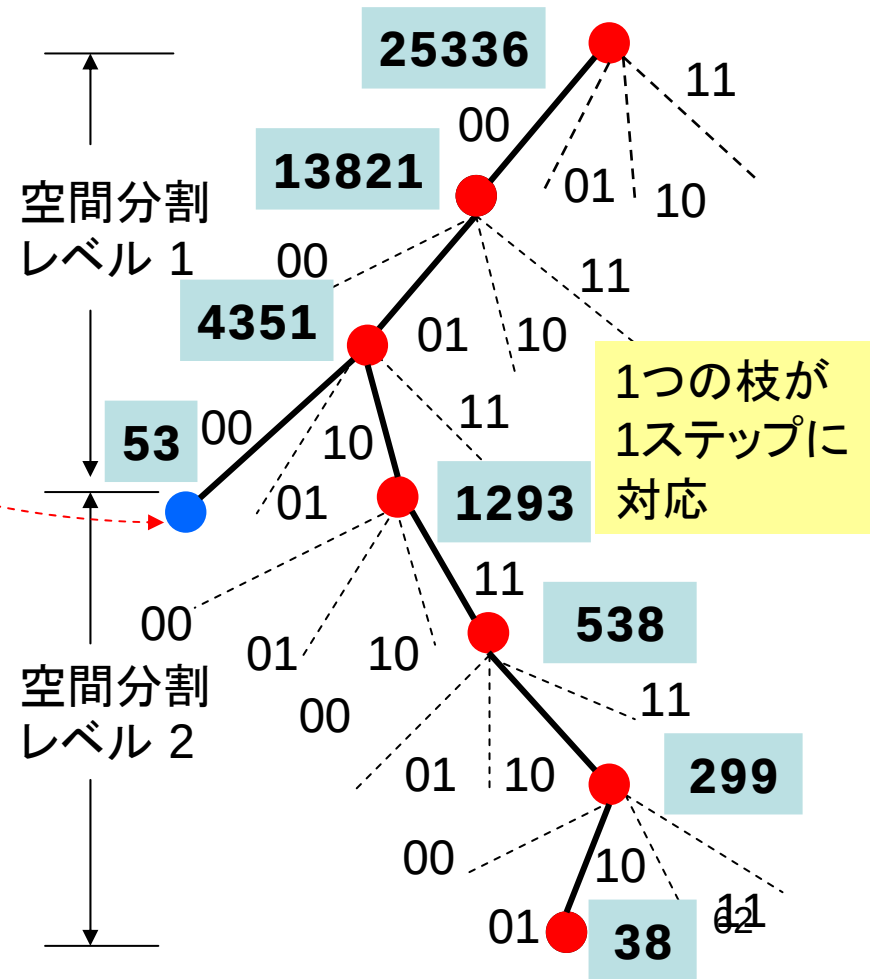
論理的なヒストグラム

- n次元のマルコフ連鎖をn+1次元のデータキューブで表現
- 0~3はグリッドセルの番号
 - セルの粒度は細分化可能
- 例: $0 \rightarrow 1 \rightarrow 0$ という遷移の総数を矢印のキューブセルに格納



物理的なヒストグラム

- 四分木(quadtrees)とk-d-木を統合した構造: 各ノードにカウント保持
- 適応的に木を拡張・縮退

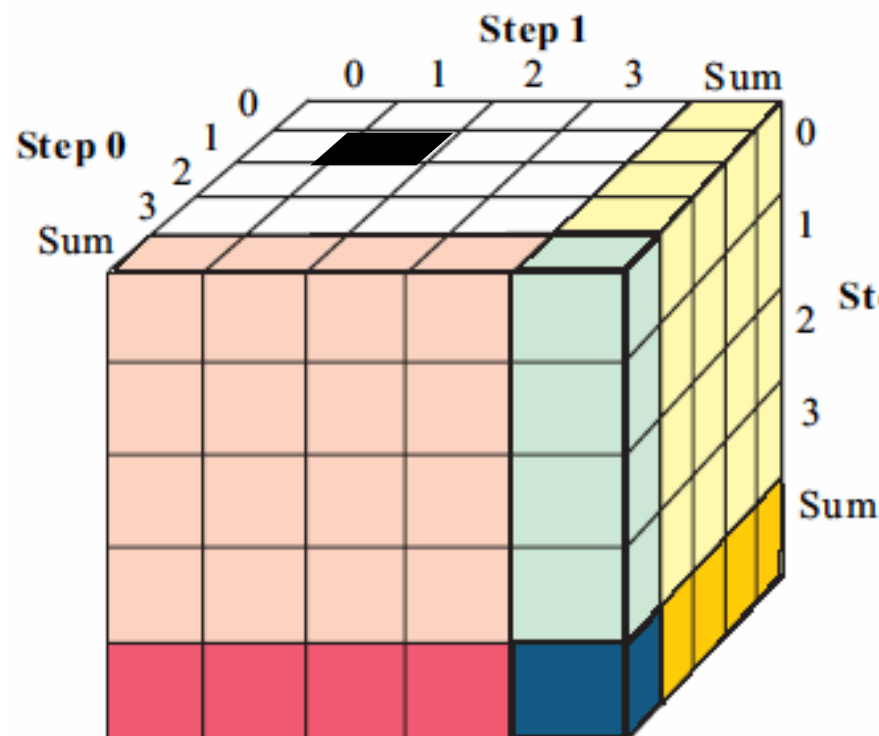




ヒストグラムの論理構造のイメージ

n 次のマルコフ連鎖 $\rightarrow$ $n+1$ 次元のデータキューブ

Sumは各次元の値ごとの和を表す



1 $\rightarrow$ 1 $\rightarrow$ 0に対応する

step0		step1		step2	
0	1	0	1	0	1
2	3	2	3	2	3

空間を4分割した時の遷移



移動ヒストグラムの物理構造

- **メモリを効率的に利用**

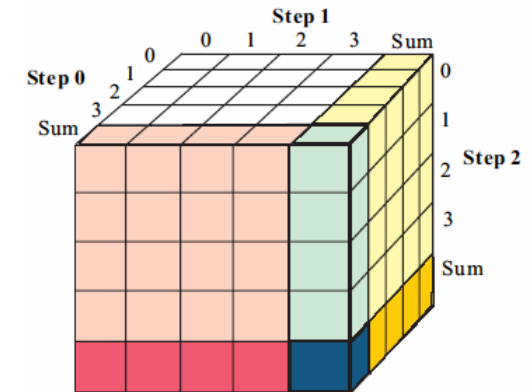
- データキューブの管理法の場合

- 複数個のデータキューブが必要
- セル数が多くなる
(32×32 のセルで2次のマルコフ遷移
すると, 10億個のキューブセルが必要)
- 出現しないシーケンスのセルの値は0

- **出現したところだけ**にメモリを割り当てる

- **木構造**で表現

- 木の各ノードは0個以上4個以下の子ノード
- 各ノードにカウンタ





移動ヒストグラムの利用

