

【ICDE2013勉強会】

## Session 5: Uncertainty in Spatial Data

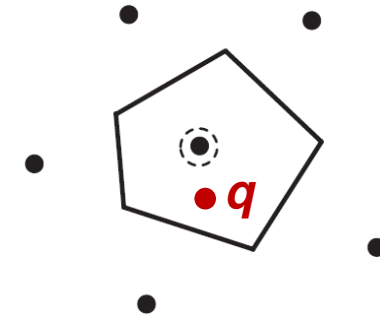
担当: 董・早矢仕(名古屋大学)

# Voronoi-based Nearest Neighbor Search for Multi-Dimensional Uncertain Databases

Zhang et al. (Hong Kong U., Ludwig-Maximilians-U. München)

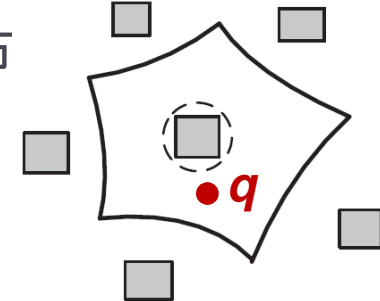
## ▶ ボロノイ図による最近傍問合せ

- ▶ 問合せ点  $q$  から一番近いオブジェクトを検索
- ▶ ボロノイ図を計算し, *Voronoi cell* が問合せ点  $q$  を含むオブジェクトが結果となる



## ▶ ボロノイ図による確率的最近傍問合せ (本論文)

- ▶ 想定: データが不確実で, 離散的に矩形内に分布
- ▶ *Possible Voronoi cell* (PV-cell) が問合せ点  $q$  を含むオブジェクトは結果となる可能性がある



#図は論文から引用

**問題点: PV-cellの計算コストが高い!**



PV-cell を *Uncertain Bounding Rectangle (UBR)* で近似し,  
UBRに基づく索引で問合せを処理する

## Uncertain Bounding Rectangle (UBR)の計算:

### ▶ **Shrink-and-Expand** algorithm

#### ▶ 矩形対による計算

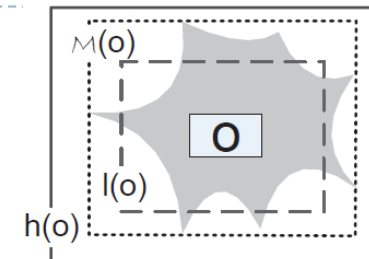
- ▶ Lower bound  $l(o)$ : **PV-cell**の**MBR**に含まれる
- ▶ Upper bound  $h(o)$ : **PV-cell**の**MBR**を含む

#### ▶ $||h(o) - l(o)|| < \Delta$ (指定閾値) まで反復

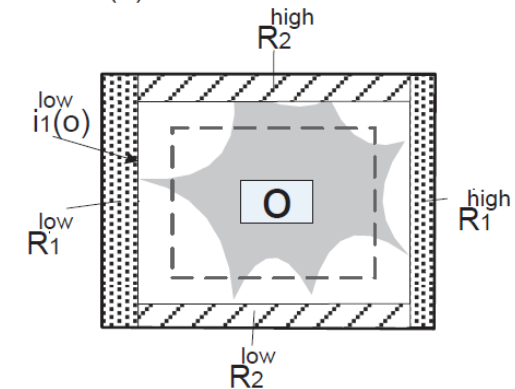
- ▶ 各次元各方向(左low, 右high) 毎に,  
 $h(o)$ と $l(o)$ の中間面 $i(o)$ と $h(o)$ の間の  
領域 $R(o)$ が**PV-cell**との交わりをチェック:

- 交わりがない:  $R(o)$ を削除し,  
 $h(o)$ を $i(o)$ まで**shrink**
- 交わりがある:  $l(o)$ を $i(o)$ まで**expand**

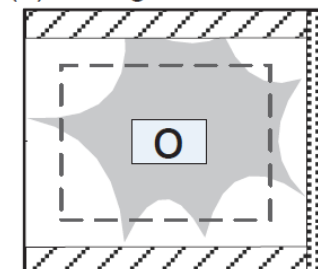
#### ▶ 計算した**UBR**は**PV-cell**の**MBR**より少し大きい



(a) Before the iteration



(b) During the iteration



(c) After the iteration

#図は論文から引用

# UBRに基づく索引: *PV-index*

## ▶ 索引構造: 空間分割

### ▶ 主索引

#### ▶ 多次元octree(2D: quadtree)

- **UBR**に基づき, IDを管理
- 非リーフノード: 主記憶
- リーフノード: ディスク

### ▶ 二次索引: ディスク

#### ▶ ハッシュ表

- 各オブジェクトのID, **UBR**, 不確実情報を管理

#図は論文から引用

## ▶ 問合せ処理

- ▶ 主索引で**UBR**が $q$ を含むオブジェクトIDを検索
- ▶ 二次索引のオブジェクト情報で正確な確率を計算

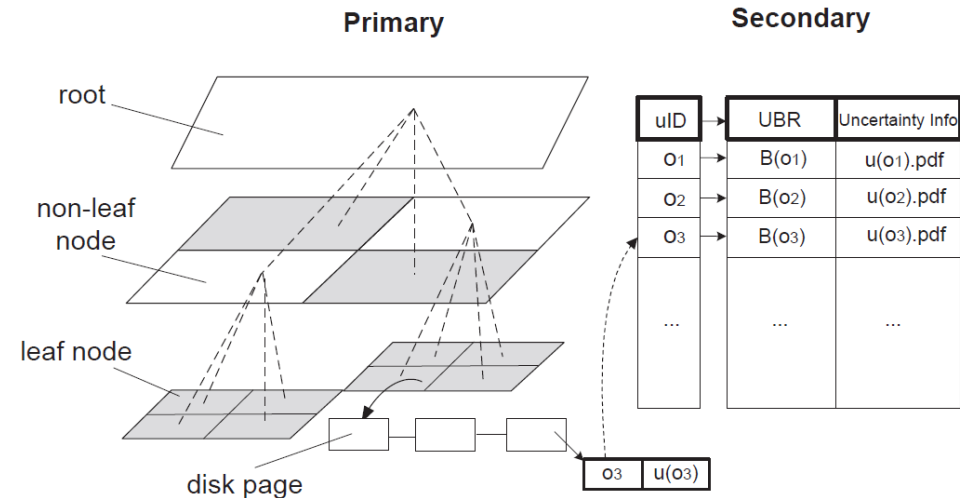


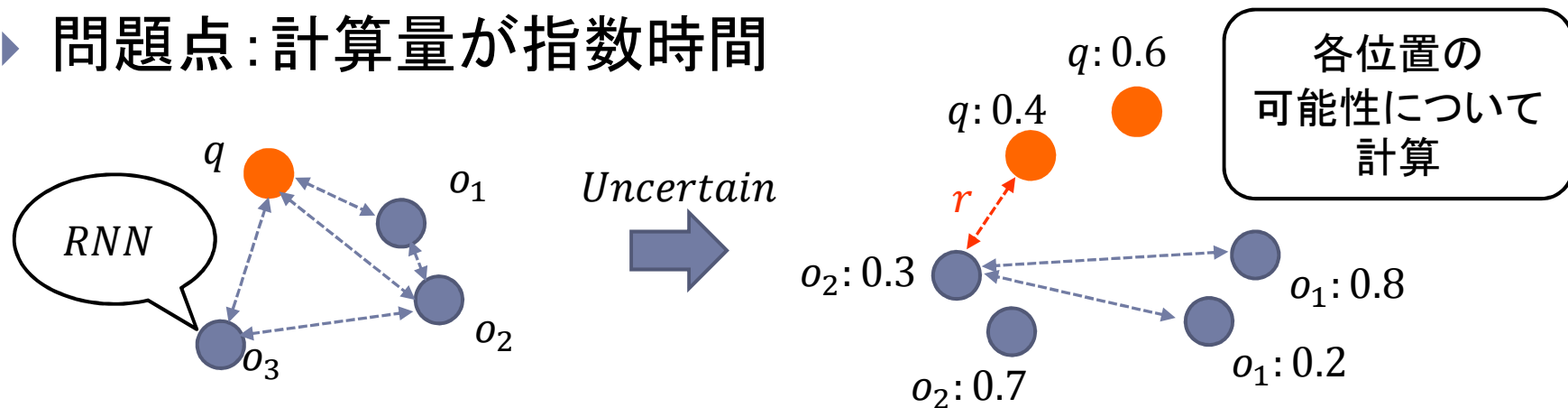
Fig. 7. Illustrating the PV-index (2D).

# Interval Reverse Nearest Neighbor Queries on Uncertain Data with Markov Correlations

Chuanfei Xu, Yu Gu, Lei Chen, Jianzhong Qiao, Ge Yu

(Dept. of Information Science and Engineering Northeastern University, China / Dept. of Computer Science and Engineering Hong Kong University of Science and Technology, Hong Kong, China)

- ▶ 目的: 曖昧なオブジェクトのIRNN問合せの高速化
- ▶ IRNN(Interval Reverse Nearest Neighbor)
  - ▶ **IRNN( $q, \Delta, D$ ): 時刻 $\Delta$ の間でRNNを満たす確率が最大となる回数が最大のオブジェクトを返す**
    - ▶ RNN: クエリオブジェクト $q$ を最近傍点とするオブジェクトを見つける
    - ▶ RNNの確率  
 $q$ との距離が $r$ かつ他のオブジェクトとの距離が $r$ 以上となる確率
- ▶ 問題点: 計算量が指数時間



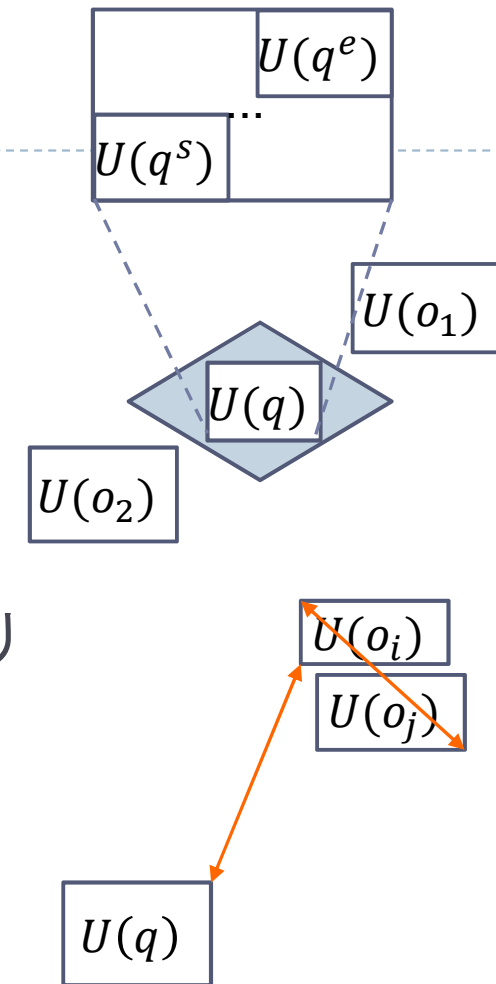
# 枝刈りによる高速化

## ▶ 空間条件による枝刈り

- ▶  $\Delta$ の間の移動軌跡からMBRを作成
- ▶  $q$ のMBRと近傍のMBRの二等分線から領域を作成
  - ▶ 領域外のオブジェクトは枝刈りされる
- ▶ MBR間の最小距離, 最大距離による枝刈り
  - ▶  $\min\_dist(U(q), U(o_i)) > \max\_dist(U(o_i), U(o_j))$ 
    - $o_i$ は $q$ の最近傍とはならないので枝刈り

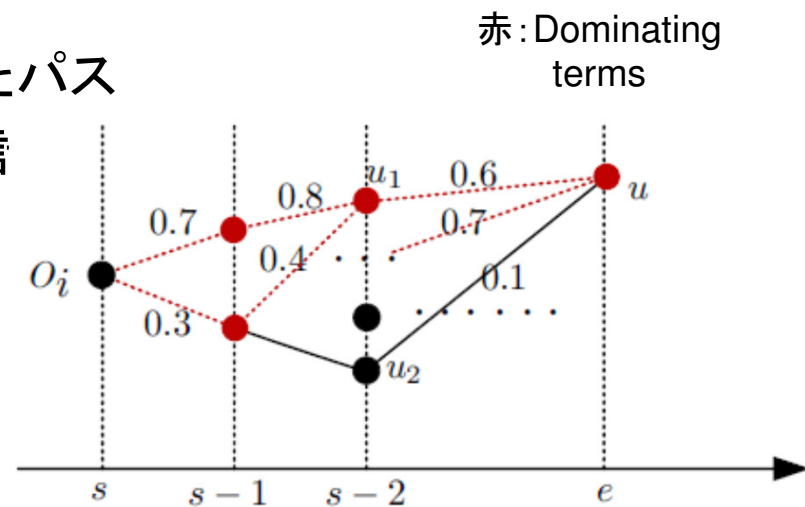
## ▶ 確率条件による枝刈り

- ▶  $q$ が最近傍になる確率の上限と下限を計算  
上限が他のオブジェクトの最近傍確率の下限より小さい時,  
候補から除外



# 各状態の確率の計算の高速化

- ▶ 各状態の確率は条件付き確率表を用いて計算
- ▶ Probability Decomposition Verification(PDV)
  - ▶ 確率の低い状態を近似的に計算
    - ▶ IRNNでは確率の高い状態を求めることが目的
  - ▶ 2つの状態に分けて計算
  - ▶ Dominating terms
    - ▶ 確率の高いtop-kの状態を辿ったパス
    - ▶ 条件付き確率表に従い, 確率を言
    - ▶ 最近傍確率の近似計算に利用
  - ▶ Remainder terms:
    - ▶ Dominating以外の状態
    - ▶ 正確な確率の計算時に利用



図は論文より引用

# Efficient Tracking and Querying for Coordinated Uncertain Mobile Objects

Nicholas D. Larusso, Ambuj Singh

(Department of Computer Science, University of California)

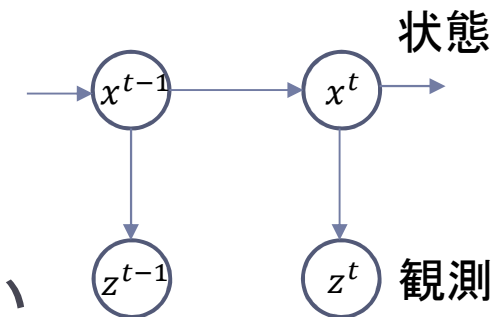
## ▶ 目的

### ▶ 1: 曖昧な移動オブジェクトの位置推定精度の向上

#### ▶ 従来法: Kalman Filter

- ▶ 隠れマルコフモデルの特殊なケース
- ▶ 状態(位置), 観測をガウス分布で表現

#### ▶ 問題点: グループ内の関連を利用していない

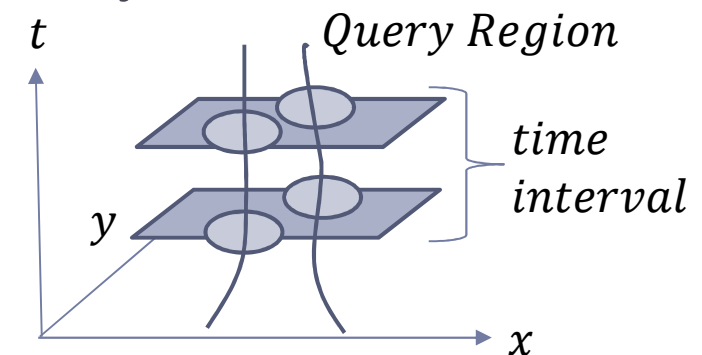


### ▶ 2: Probabilistic Threshold Range Query問合せの高速化

#### ▶ PTRQ(Probabilistic Threshold Range Query)

- ▶ 時間区間, 領域, 確率閾値を指定
- ▶ 条件を満たすオブジェクトを返す

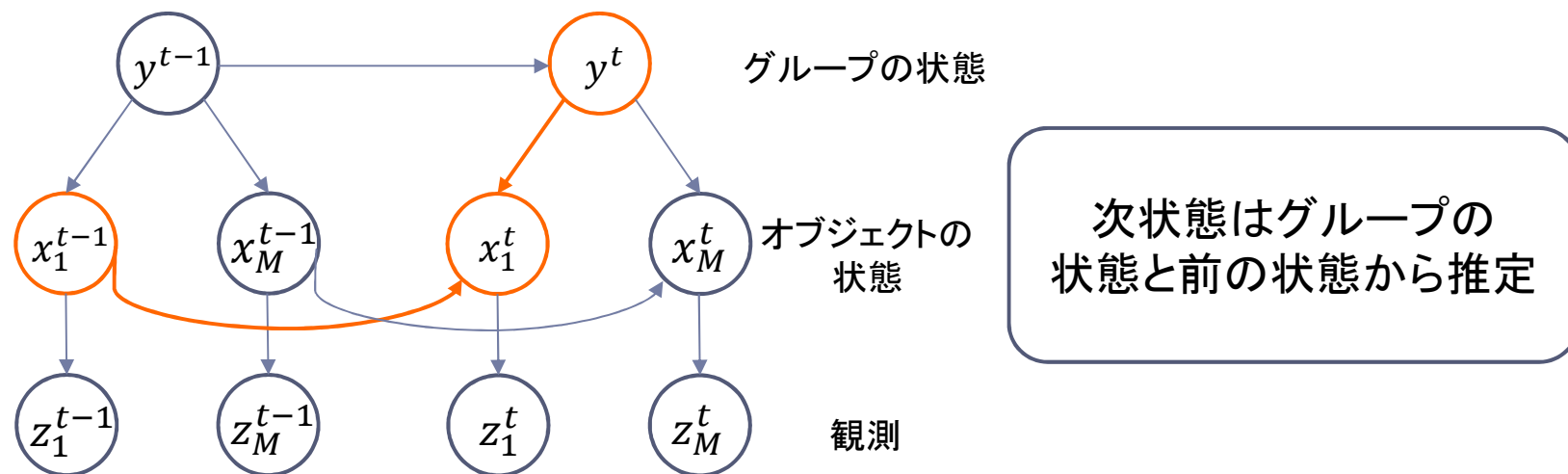
#### ▶ 問題点: オブジェクトを索引構造で扱うとメンテナンスコストがかかる





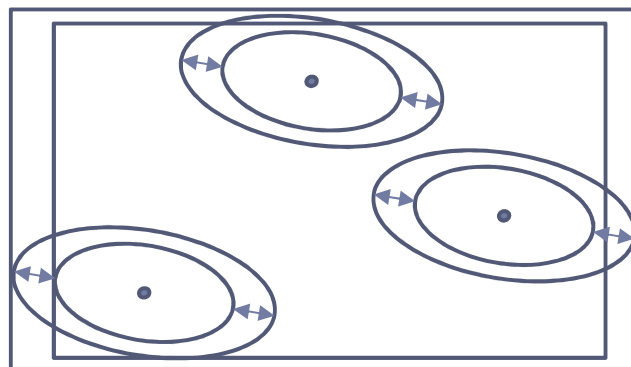
# Clustred Kalman Filter(ClustKFM)

- ▶ ClustKFM: グループの重心の状態を追加したカルマンフィルタモデル
  - ▶ グループの状態を元に各オブジェクトの状態を推定
  - ▶ グループはk-meansによるクラスタリング
- ▶ ノイズによるデータ欠損や異常な観測値をグループの状態から補正→精度の向上



# 問合せの高速化

- ▶ クラスタのBounding Region(BR)を作成
  - ▶ 確率の閾値により大きさが変動
  - ▶ オブジェクトの移動速度の誤差と位置の曖昧性によるBRの変動を考慮
- ▶ BRに重ならないオブジェクトグループを枝刈り
- ▶ 既存のクラスタをそのまま利用するためデータ構造の更新のコストが低い



BR

