

D-025

確率位置情報ストリームに対する履歴空間問合せ

早矢仕 新[†] 董 ていてい[†] 加藤 翔[†] 石川 佳治^{††}

^{††} 名古屋大学大学院情報科学研究科

^{††} 名古屋大学情報基盤センター ^{††} 国立情報学研究所

1 はじめに

近年, GPS や RFID などセンサで位置情報を取得し, アプリケーションに活用するケースが増加している. センサで位置情報を扱う際, センサは様々な影響による誤差を含むため, 確率的な表現を用いることが多い. そのため, 確率位置情報に対応した時空間データベースに関する研究が近年, 進められてきている [2]. また, 時空間データベース技術の応用として, オブジェクトの行動を検索するイベント問合せと呼ばれる問合せに関する研究も進められている [1]. 確率的時空間データベースを対象にオブジェクトの行動を検索する場合, 時系列を追って, 確率計算を行うため計算コストがかかるという問題がある.

本稿では, オブジェクトの確率位置情報の時系列データを対象に, オブジェクトの行動の問合せを行う手法と問合せの効率化のための索引データ構造について述べる.

2 問合せの定義

本論文では, オブジェクトの確率位置情報をパーティクルフィルタと呼ばれる位置推定手法によって得られたものと仮定する [3, 4]. パーティクルフィルタとは, ポピュラーな移動オブジェクトの位置推定手法であり, 確率分布を大量のサンプル集合によって表現するのが特徴である. そのため, オブジェクトの位置情報は図 1 のような形で表現される. 本論文で問合せの対象とする履歴データとは, サンプル集合と時間の情報を対応付けたデータである. その例が図 2 である.

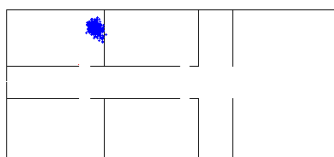


図 1: サンプル集合による確率分布の表現

この履歴データに対して, オブジェクトの挙動を問い合わせる方法について考える. 本論文ではオブジェクトの挙動として, 「ある確率以上で, 指定された時間

時刻	オブジェクト	パーティクル集合
11:11:11	o1	$S_1 = \{(613, -467), (583, -579), \dots, (649, -547)\}$
11:11:11	o2	$S_2 = \{(294, -304), (254, -318), \dots, (214, -369)\}$
11:11:13	o1	$S_3 = \{(93, -309), (121, -172), \dots, (151, 170)\}$
11:11:13	o2	$S_4 = \{(278, -283), (261, -219), \dots, (312, -161)\}$
11:11:15	o1	$S_5 = \{(604, 240), (454, 164), \dots, (335, 356)\}$

図 2: 履歴データの例

の間に, 指定された領域に到達した」というものを対象とする. 履歴データに対して, この条件を満たす行動をしたオブジェクトを検索する問合せを行動問合せと呼ぶ. この行動問合せを以下のように定義する.

定義 1 (行動問合せ) 問合せ矩形領域 r と時間区間 t と確率の閾値 θ が与えられた時, 行動問合せ $Q(r, t, \theta)$ は条件を満たすオブジェクトの ID を返す. □

以下では, この問合せの処理方法の効率化するためのアプローチについて述べる.

3 索引データ構造

行動問合せにおける確率計算は各時刻の確率位置情報とオブジェクトの移動確率から行われる. そのため, 行動問合せにおいて, 計算コストが最も大きいのがパーティクルフィルタによるサンプル集合の遷移のための計算となる. というのも, 次状態のサンプル集合を計算する際に, すべてのサンプルに対して状態遷移の計算と尤度の計算を行うからである.

本論文では上記の計算量を削減するために確率計算を近似的に行うための索引を作成するアプローチを取る. 使用する索引のデータ構造を示したのが図 3 である.

位置テーブル				遷移テーブル					領域テーブル		
t	object	C	P(C)	object	t	t'	C	C'	P(C' C)	Cell	Region
3	o1	C1	0.4	o1	3	4	C1	C1	0.25	C1	$\{(x1, y1), (x2, y2)\}$
3	o1	C2	0.3	o1	3	4	C1	C2	0.25	C2	$\{(x3, y3), (x4, y4)\}$
...	o1	3	4	C1	C3	0.5
4	o1	C1	0.5
...

図 3: 索引データ構造

索引は位置テーブル, 遷移テーブル, 領域テーブルの 3 つのテーブルから構成される. 位置テーブルは時刻とその時にオブジェクトがセルに存在した確率を関連付けたテーブルである. 位置テーブルの各データは, 履歴データから作成される. この時, 各セルでのオブ

Querying Spatio-Temporal Data on Probabilistic Spatial Data Streams

Arata Hayashi[†], Tingting Dong[†], Sho Kato[†], Yoshiharu Ishikawa^{††}

[†] Graduate School of Information Science, Nagoya University

^{††} Information Technology Center, Nagoya University and National Institute of Informatics

ジェットの存在確率は、セルに含まれるサンプルの数をサンプル集合に含まれるサンプルの総数で割ることによって求められる。遷移テーブルは時刻 t から時刻 $t+1$ へオブジェクトが移動した時に、どのセルからどのセルへどれだけ確率で移動したのかを保持したテーブルである。領域テーブルは空間を分割してできた各セルの番号と矩形領域を保持したテーブルである。

以上が問合せ処理に使用する索引データの構造である。

4 問合せ処理アルゴリズム

このアルゴリズムは主に 3 つのステップで行われる。

第 1 ステップでは位置テーブルの情報を使い、問合せ領域に包含されるセルの存在確率を調べ、確率の閾値による判定を行う。判定の状況を示したのが図 4 である。問合せ領域に含まれたセルに対応するオブジェクトの存在確率を位置テーブルから取得し、各オブジェクトについて存在確率の総和を求めて閾値判定を行う。この時、指定された時刻の間で一度でも閾値の条件を満たしたならば、そのオブジェクトは問合せ結果に含まれる。問合せ結果に含まれなかったオブジェクトを対象に次のステップの処理を行う。

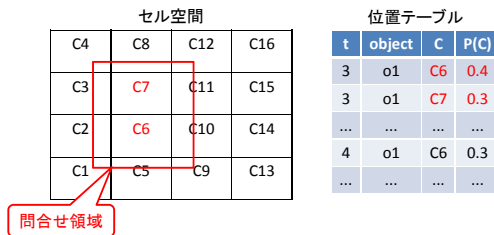


図 4: 問合せ領域とセル情報からの確率計算

第 2 ステップでは問合せ領域に包含されるセルへの到達確率を遷移確率から計算する。到達確率の計算はまず、指定された時刻の始点に当たる時刻のオブジェクトの存在確率を位置テーブルから取得する。そして、遷移テーブルから遷移確率を取得し、存在確率と掛けあわせて次の時刻の存在確率を計算する。この確率計算を時間区間の最後まで繰り返す。セル間のオブジェクトの移動の例を表したのが図 5 である。図の丸は各セルを、矩形は問合せ領域を、矢印はオブジェクトの移動を表す。この時重要なのが、一度問合せ領域に該当するセルに到達した場合、そのセルからの遷移は行わないという点である。つまり、図 5 の点線の矢印で表された遷移による確率計算は行わないということである。これは一度到達した軌跡を重複して調べ、確率計算を行うことを防ぐためである。つまり、セルを $C1, C1, C2$ と移動して $t = k + 2$ で問合せ領域に到達する軌跡の確率と、そこから更に $t = k + 3$ で $C2$ または $C3$ に移動した軌跡の到達確率を足しあわせるのを防ぐためである。軌跡の重複を除去することで、最終

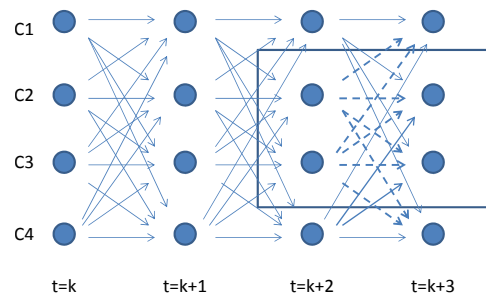


図 5: オブジェクトの遷移とセルの到達確率

的な到達確率はオブジェクトが初めて問合せ領域に到達した時の到達確率の総和として求められる。到達確率を各時刻の到達した時の確率の和として計算できるため、途中で和が閾値以上になった時、オブジェクトは条件を満たしたと判定し、処理を打ち切ることが可能となる。これによって、計算の高速化が可能である。

第 3 ステップでは、パーティクルフィルタを実行し、指定された区間の各時刻でのサンプル集合を計算し、各時刻でのオブジェクトの到達確率を求める。パーティクルフィルタで逐次、サンプル集合の状態遷移を行う際、一度問合せ領域に到達したパーティクルは到達確率の計算には用いず、到達確率は未到達のパーティクル集合を母集団とし、各時刻で到達したパーティクルの比から計算する。最終的な到達確率は指定区間中の到達確率の総和として求められる。そして、到達確率を閾値で判定し、条件を満たしたオブジェクトを問合せ結果に含める。

以上の 3 つの各ステップで得られた問合せ結果の和集合が最終的な問合せ結果となる。

以上が問合せ処理アルゴリズムである。

5 まとめと今後の課題

本稿では、確率位置情報の時系列データを対象に、オブジェクトの到達を調べる行動問合せのための索引データ構造と問合せ処理アルゴリズムについて述べた。今度の課題として、問合せ処理の実証実験と性能評価が挙げられる。

謝辞

本研究の経費の一部は科学研究費 (23650047) および内閣府最先端研究開発プロジェクト (FIRST) による。

参考文献

- [1] Magdalena Balazinska, Christopher Ré, Julie Letchner, and Dan Suciu. Event queries on correlated probabilistic streams. *in Proc. of ACM SIGMOD*, 2008.
- [2] Tobias Emrich, Hans-Peter Kriegel, Nikos Mamoulis, Matthias Renz, and Andreas Züfle. Querying uncertain spatio-temporal data. *in Proc. of ICDE*, 2012.
- [3] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. 確率ロボティクス. 毎日コミュニケーションズ, 2007.
- [4] 樋口知之. 予測にいかす統計モデリングの基本 - ベイズ統計入門から応用まで. 講談社, 2011.