

センサ環境におけるオブジェクトの移動モデルの提案

堀田 孝司† 石川 佳治††

† 名古屋大学 情報科学研究科 †† 名古屋大学 情報基盤センター

1 まえがき

今日では、RFID タグなどの IC タグの活用がすでに広く行われている。ここでは、移動オブジェクト（人もしくは移動可能な物体）にタグを付与し、屋内の各所に設置されたタグリーダによりセンシングすることにより、その移動状況を収集し分析するためのフレームワークについて検討する。タグリーダによる読み取りは必ずしも正確であるとは限らず、ノイズやデータの欠損がしばしば発生するという問題があることから、曖昧な位置情報についての対応が重要となる。

このため、確率モデルを用いて、センシングされたデータを補完し、より正確な位置の推定をしようとするアプローチが存在する。たとえば、[1] では、粒子フィルタ (particle filter) [3] を用いて、各移動オブジェクトが各時点においてどの位置に存在していたかという存在確率、そして、次の時点でどの位置（タグリーダの場所）に移動するかという遷移確率の2つの確率を各時点について推定し、その結果を問合せ可能とするアプローチを提案している。

[1] の手法は、単一の移動オブジェクトのある移動に関する問合せに限られており、たとえば「Bob の移動履歴をもとに、オフィス1からオフィス2への移動がいつ発生したかを確率を付与して提示せよ」といった問合せが想定されており、大量の履歴データを集約して、主要な移動のパターンを捉えることは目的とされていなかった。これに対し本研究では、多数の移動オブジェクトの移動履歴を集約するためのモデルを提案し、集約処理のためのアプローチについて提案を行う。

2 確率モデルに基づく移動状況の表現

本研究では、[1] に従い、各時点におけるタグのセンシングにより得られた移動履歴情報を、確率モデルにより処理し得られる移動履歴データを対象とする。図1は、各オブジェクトがどの時点でどこに存在していたかを表現している。A, B はタグが付与されたオブジェクト（具体的には人物など）を表している。この表のエントリは、たとえば、時刻5においてAは位置H1（廊下1）、H2およびO1（オフィス1）に、それぞれ0.6, 0.3, 0.1でいたという確率を表している。

図2は隣接する時刻間における、位置間の遷移確率を表す。たとえば、時刻5にAが位置H1にいて、時

obj	t	l	$p_{x,t}^e(l)$	obj	t	l	l'	$p_{x,t}^e(l' l)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
A	5	H1	0.60	A	5	H1	H1	0.60
A	5	H2	0.30	A	5	H1	H2	0.20
A	5	O2	0.10	A	5	H1	O1	0.05
				A	5	H1	O2	0.15
B	5	O1	0.80	A	5	H2	H1	0.15
B	5	H1	0.15	A	5	H2	H2	0.70
B	5	H2	0.05	A	5	H2	O1	0.05
A	6	H1	0.45	A	5	H2	O2	0.10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

図1: 存在確率

図2: 遷移確率

刻6にH1にそのまま存在していた確率 $p_{A,5}^e(H1|H1)$ は0.6であり、H2に移動した確率 $p_{A,5}^e(H2|H1)$ は0.2である。これらのデータは、ノイズや欠損値を含むタグの読み取りデータをもとに、確率的モデルを用いて推定された結果である。本研究では、特定の推定手法には踏み込まず、このようなデータを前提として、どのような情報集約が可能かについて議論する。

3 基本移動パターンとそれに基づく集約

まず例を用いて説明する。ここでは、移動パターンを $l \rightarrow l'$ のように表記する。この移動パターンは、「位置 l から l' に1単位時間の中に移動した」という状況を表す。特に、後で出てくる移動パターンと区別して基本移動パターンとも呼ぶ。具体例としては、H1 → H2 があげられる。図1, 2のデータについてこのパターンとの関連を考える。時刻 $t=5$ においてオブジェクトAがH1にいた確率は $p_{A,5}^e(H1) = 0.6$ であり、これはこの移動パターンをルールと見たとき、前提部が成り立つ確率にあたる。また、時刻 $t=5$ でH1にいたオブジェクトAが次の時点 $t=6$ でH2に移動した確率は $p_{A,5}^e(H2|H1) = 0.2$ であり、これは前提部が成り立った場合に帰結部が成り立つ確率にあたる。

上記は特定の移動オブジェクトに対する移動に着目したものであるが、移動パターンの概念を用いて移動履歴の集約を行うことができる。まず、 \mathcal{O} をセンシング対象のオブジェクト全体の集合であると、 $\mathcal{O}_{[t_1, t_2]}$ を、時間区間 $[t_1, t_2]$ においてセンシング対象であったオブジェクトと呼ぶ。 $\mathcal{O}_{[t_1, t_2]} \subseteq \mathcal{O}$ が成り立つ。

ここで、時刻 t における移動パターン $q \equiv l \rightarrow l'$ の支持度 (support) を、

$$\text{sup}_t(q) = \frac{\sum_{o \in \mathcal{O}_{[t, t+1]}} p_{o,t}^e(l)}{|\mathcal{O}_{[t, t+1]}|}$$

と定義する。これは、時刻 t において位置 l にオブジェクトが存在している確率を表しており、相関ルー

A Proposal of Object Movement Model in Sensor Environments

Koji Hotta†, Yoshiharu Ishikawa††

† Graduate School of Information Science, Nagoya University

†† Information Technology Center, Nagoya University

ルマイニングやシーケンスマイニング [2] における支持度に対応する。一方、時刻 t における移動パターン $q \equiv l \rightarrow l'$ の確信度 (confidence) を、

$$\text{conf}_t(q) = \frac{\sum_{o \in \mathcal{O}_{[t, t+1]}} p_{o,t}^e(l) \cdot p_{o,t}^t(l'|l)}{\sum_{o \in \mathcal{O}_{[t, t+1]}} p_{o,t}^e(l)}$$

と定義する。この式は、時刻 t から $t+1$ にかけて l から l' への移動が、どの程度の確率で生じるかを表している。こちらはデータマイニングにおける確信度に対応する。これら支持度、確信度を用いれば、ある時刻 t における移動状況を要約でき、また、その移動パターンがどの程度有用であるかを評価できる。支持度が低い移動パターンはそもそも適用できる状況がまれであり、確信度が低い移動パターンはその信頼性が低いことになる。

これまではある時点 t を対象としていたが、上記の考え方を時区間に拡張することができる。時区間 $[t_s, t_e]$ における移動パターン $q \equiv l \rightarrow l'$ の支持度、確信度を

$$\text{sup}_{[t_s, t_e]}(q) = \frac{\sum_{t=t_s}^{t_e-1} \sum_{o \in \mathcal{O}_{[t, t+1]}} p_{o,t}^e(l)}{\sum_{t=t_s}^{t_e-1} |\mathcal{O}_{[t, t+1]}|}$$

$$\text{conf}_{[t_s, t_e]}(q) = \frac{\sum_{t=t_s}^{t_e-1} \sum_{o \in \mathcal{O}_{[t, t+1]}} p_{o,t}^e(l) \cdot p_{o,t}^t(l'|l)}{\sum_{t=t_s}^{t_e-1} \sum_{o \in \mathcal{O}_{[t, t+1]}} p_{o,t}^e(l)}$$

と定義する。これらの指標を用いることで、たとえば「昨日の午前中の主要な基本移動パターンは？」といった問いに答えることができる。逆に、このような情報を集約しておけば、頻度がまれな移動パターンが発生したときに、それを検出して通知することも可能となる。

4 移動パターンの一般化

これまでは1ステップの移動パターンのみを考えていたが、これを拡張して $q \equiv l \rightarrow l' \rightarrow l''$ という2ステップの移動パターンを考える。このパターンを、 l, l' と移動したオブジェクトが次に l'' に移動すると解釈する。まず、支持度については、前提となる $l \rightarrow l'$ という移動がどのくらいの確率で生じたかを考えればよいので、上記の拡張で

$$\text{sup}_{[t_s, t_e]}(q) = \frac{\sum_{t=t_s}^{t_e-2} \sum_{o \in \mathcal{O}_{[t, t+2]}} p_{o,t}^e(l) \cdot p_{o,t}^t(l'|l)}{\sum_{t=t_s}^{t_e-2} |\mathcal{O}_{[t, t+2]}|}$$

と定義できる。一方、確信度については、

$$\text{conf}_{[t_s, t_e]}(q) = \frac{\sum_{t=t_s}^{t_e-2} \sum_{o \in \mathcal{O}_{[t, t+2]}} p_{o,t}^e(l) \cdot p_{o,t}^t(l'|l) \cdot p_{o,t+1}^t(l''|l')}{\sum_{t=t_s}^{t_e-2} \sum_{o \in \mathcal{O}_{[t, t+2]}} p_{o,t}^e(l) \cdot p_{o,t}^t(l'|l)}$$

となる。このような考え方を発展させることで、さらに長い移動パターンも考えることができる。なお、

$(H1)^2 \rightarrow H2 \equiv H1 \rightarrow H1 \rightarrow H2$ のような略記も許すことにする。

さらに、位置 l から次の時点に任意の位置 x を経由して l' に至ったような移動パターンを $l \xrightarrow{2} l'$ のように表す。“2” という数値は、2回の遷移を行っていることを表している。これはより上位の移動パターンであり、その支持度・確信度は、 $l \xrightarrow{1} x \xrightarrow{1} l'$ という移動パターンに関する集約を、すべての x について実行することで得られる。

5 移動パターンの集約問合せとマイニング

本研究では、ユーザから指定された時区間において、与えられた移動パターンがどの程度の支持度、確信度を持つかを問い合わせることを可能にしたいと考えている。このような問合せは、一種の集約問合せと考えられる。一方、ユーザから指定された支持度、確信度の閾値を満たすような移動パターン検出するタスクも考えられる。これを本研究では移動パターンマイニングと呼ぶ。

ただし、容易に分かるとおり、ありふれた短いシーケンスの移動パターン (例: $H1 \rightarrow H1$) は支持度、確信度ともに高いことから、ユーザにとって有用性の高い移動パターンを検出するには、支持度、確信度以外の異なる指標による評価や、事後処理を行うことによる、列挙された移動パターンのさらなる集約処理などが必要となると考えられる。

また、センサ情報から推定される図 1, 2 のデータは、蓄積された形ではなく、ストリームの形式で与えられることも考えられる。ストリームデータを逐次的に処理するようなタスクについても検討を行いたい。そこでは、データが得られるごとの、低コストでインクリメンタルなパターンの集約するが求められる。

6 まとめと今後の課題

本稿では、RFID タグなどをセンシングして、移動オブジェクトの位置を取得し蓄積するような応用における、移動パターンの表現モデルについて検討した。確率的に推定された位置情報、遷移情報をもとに、集約処理をどのように表現すればよいかについて議論した。今後は移動パターンのさらなる拡張や、実装アルゴリズムについて検討したい。

謝辞

本研究の一部は、科学研究費 (19300027, 21013023) の助成による。

参考文献

- [1] J. Letchner, C. Ré, M. Balazinska, and M. Philipose. Access methods for Markovian streams. In *ICDE 2009*, pp. 246–257, 2009.
- [2] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, 2005.
- [3] S. Thrun, W. Burgard, and D. Fox. 確率ロボティクス. 毎日コミュニケーションズ, 2007.