

歩行者プローブデータ多面的活用のためのデータクリーニング手法に関する研究*

A study of the data cleaning methods for versatile practical use of Probe Person.*

薄井智貴**・三輪富生***・山本俊行***・森川高行****

By Tomotaka USUI**・Tomio MIWA***・Toshiyuki YAMAMOTO***・Takayuki MORIKAWA****

1. はじめに

人々の日常生活を豊かにする技術を提案する上で、その一人一人の詳細な行動を把握することは重要なファクターである。例えば、人が何を目的として行動しようとしているのか、実際にどういう行動をしたのか、行動した結果どうだったのか、という行動に関する一連の理解を深めるためには、詳細な実行動履歴データによる分析が必要不可欠である。

従来、これらのデータ収集手法としては個人へのアンケート形式やカメラ等による定点観測によるものが大半であった。しかし、近年の情報通信技術の発展とともに詳細な行動履歴をセンシングする機器も急速に普及してきており、その観測データを用いた様々な研究も行われてきている¹⁾。

この観測データは、特に交通計画分野においては、プローブ（探針）データと呼んでおり、歩行者や自転車、自動車など移動体の動きを観測する目的で収集されている。このプローブデータは、GPSや無線LAN、RSSI、RFIDなどの多様なセンシング技術を用いて観測された時空間情報であり、行動軌跡を時々刻々と記録したデータである²⁾。近年、様々な社会実験や研究活動においてこれらのプローブデータが大量に観測、収集され、それぞれの目的に即した分析が行われている。しかし、これらのプローブデータは、観測された状況下および目的でのみ意味のあるセンシングデータであり、行動軌跡自体の再現性はあるものの、他目的の分析に対する汎用性は低く、社会全体として有効活用できていない。そのため、プローブデータを社会の共有資産として、観測時目的以外において多面活用するための手法と用途を検討する必要がある。例えば、個々のプローブデータに対し行動の目的や状態などを各トリップ内に空間情報としてラベリングし、トリップ自体に付加価値

を与えることにより、汎用性のある行動データに拡張することができる。

一方、プローブデータには、GPSの精度による観測誤差や測位機器自体の誤差なども含まれているため、観測環境や機器の精度によっては、正確な位置情報が連続して収集できない場合がある。そのため、これらの誤差を含むデータに対し、外れ値や欠損値を正規化（クリーニング）し、汎用的に再利用可能なデータとして「デジタルコンテンツ化」することが、プローブデータの新しい価値を見出す上で重要となってくる。

以上の背景のもと、本研究ではプローブデータを「デジタルコンテンツ」データとして多面活用するためのデータクリーニング手法と、観測行動の意味付けのためのラベリング手法を提案し、プローブデータの新たな利活用方法を見出すことを最終目的としている。

2. 観測データの位置推定手法

交通計画分野においてプローブデータの分析を行う場合、まず、観測されたデータをクリーニングするため、道路ベクトルネットワークへのマップマッチング³⁾⁴⁾によって正規化を行い、データの位置や経路を推定することが大半である。このマップマッチングは、「移動体は道路上を移動する」という前提条件のもと、収集された移動データを道路ネットワーク上に補正するもので、観測位置推定のための誤差を排除することと、正確な行動経路を推定する目的で行われる。

この道路ネットワークへのマップマッチングは、実道路の中心線を結んだ線ベクトルデータへ観測したプローブデータを逐次吸着させ、経路を推定するものであるが、車両など主に道路を走行する移動体の位置推定には有効であるものの、歩行者や自転車のような路側移動体の位置推定には不向きである。というのも、歩行者の詳細な行動経路を検討する上では、道路上の左右どちらの路側を通行していたのか、路上の障害物を避けて移動したのか、信号で立ち止まっているのか、建物に立ち寄っているのか、

*キーワード：プローブデータ、マップマッチング、パーティクルフィルタ

**正会員，博(工)，東京大学空間情報科学研究センター
(千葉県柏市柏の葉5-1-5, TEL:04-7136-4308,
E-mail:usui@csis.u-tokyo.ac.jp)

***正会員，博(工)，名古屋大学工学研究科

****正会員，Ph.D.，名古屋大学大学院環境学研究科

これらの判断が分析の鍵となることが多く、道路ネットワークへのマッチングだけでは不十分といえる。

本研究では、課題であるこの道路ネットワークデータのような“線”へのマップマッチングに対し、車道や歩道、敷地、建物など移動可能空間全体に対して“面”へのマップマッチングを考える。しかし、現在普及しているGPS計測機器においては計測誤差に数十センチから数十メートルとばらつきが大きく、より正確な位置計測を必要とする“面”へのマップマッチングは非常に困難である。

北澤ら⁵⁾は、これらの問題に対し、詳細な地図とジャイロコンパスを利用した歩行者の観測データを用い計測誤差を修正するマップマッチング手法を提案している。この手法は、独自の通路ネットワークデータと周辺の地物を境界線と領域で表した地図データを用い、壁面などとの交差判定を行いつつマップマッチングし、位置を推定する手法である。プローブデータの観測機器には、ジャイロコンパスと加速度計をGPSと組み合わせた特殊な端末を用いている。マップマッチングは、測位座標を軌跡ベクトルとして繋いでいき、ジャイロセンサーを用いて急激な進行方向の変化の修正や、壁面である外郭線との交点を逐次調べることにより、プロットデータを補正していく。結果として、建物内のマップマッチングに対し良好な結果を残しているが、進行方向の角度を計測するジャイロセンサーデータが必要なことや、周辺地物の詳細な外郭線データとネットワークデータのデータベース整備に多大な労力とコストがかかるため、GPS携帯等により取得された時空間情報のみの簡易なプローブデータや、詳細な地図データを持たない地域や建物内のデータに対してマッチングを行うには、敷居が高いと考えられる。

長坂ら⁶⁾は、アクティブRFIDタグを利用した移動体のリアルタイム位置推定にパーティクルフィルタを用いたアルゴリズムを考案している。この研究では、移動体に付与したタグから受信した電波強度を2台のRFIDタグリーダを用いて観測し、移動体の2次元平面上の位置をリアルタイムに推定しており、電波強度から拡散したパーティクルの尤度を計算し、尤度の高いパーティクルが集中する場所を推定値として検出している。

以上の既往研究を踏まえ、本稿では、まずプローブデータ正規化に必要な不可欠なマップマッチング手法を確立することを目的とし、長坂らのパーティクルフィルタアルゴリズムを援用した、面へのマップ

マッチングアルゴリズムに関して検討を行い、その手法を述べる。

3. プローブデータ位置推定アルゴリズム

(1) パーティクルフィルタ

パーティクルフィルタ^{7,8)}は、観測モデルとして非線形の任意のモデルを適用できるため、画像などの観測過程が複雑なコンピュータビジョン等で多く用いられている。近年では、主に画面上での対象物の追跡や輪郭抽出などにおいても適用され、注目を集めている手法である。本稿においては、2次元空間上で過去の観測データを事前分布として、多数のパーティクル（粒子）の集合を用いて現時刻での観測データを推定し、尤度の高さにより位置を補正することでベイズ的に経路を推定していく。以下、アルゴリズムについて述べる。

パーティクルフィルタは、時刻 t における観測値 θ_t から状態ベクトル x_t の事後確率である確率密度関数 $p(x_t | \theta_t)$ を推定する。この事後確率 $p(x_t | \theta_t)$ は、ベイズの定理により以下の式(1)に置き換えられる。

$$p(x_t | \theta_t) = \frac{p(y_t | x_t) \cdot p(x_t | \theta_{t-1})}{p(y_t | \theta_{t-1})} \quad (1)$$

$$p(x_t | \theta_{t-1}) = \int p(x_t | x_{t-1}) \cdot p(x_{t-1} | \theta_{t-1}) dx_{t-1} \quad (2)$$

ここで、 $p(y_t | x_t)$ は、ある状態 x_t のときに、観測値 y_t を得る尤度であり、 $p(x_t | \theta_{t-1})$ は、時刻 t における x_t の事前確率となっている。これらの式を用い、逐次的にパーティクル群集合を生成し、確率密度を求める。パーティクルフィルタのアルゴリズムを図1に示す。

まず、一時刻前 $t-1$ の状態をもとに抽出されたパーティクル群 $p(x_{t-1} | \theta_{t-1})$ を選択し、対象領域に拡散する。これを事前分布とする(1)。次に、運動モデル $p(x_t | x_{t-1})$ を用い、パーティクル群を一定の規則に基づいて移動させ、ランダム雑音を加える(2)。この運動モデルに関しては、後に説明する。さらに、各パーティクルに対し、その尤度を計算する(3)。本稿においては、各パーティクルから現時刻 t における観測値 x_t までの距離の逆数を尤度とした。最後に、パーティクルをフィルタリングする(4)。このフィルタリングでは、尤度の高いパーティクルのみ再びサンプリングし、低いものに関しては消滅する。このとき、尤度の最も高いパーティクルが集中する位置

を推定値としてプローブデータの位置を確定する。以後、この(1)~(4)をプローブデータサンプル数分繰り返す、プローブデータの経路を確定していく。

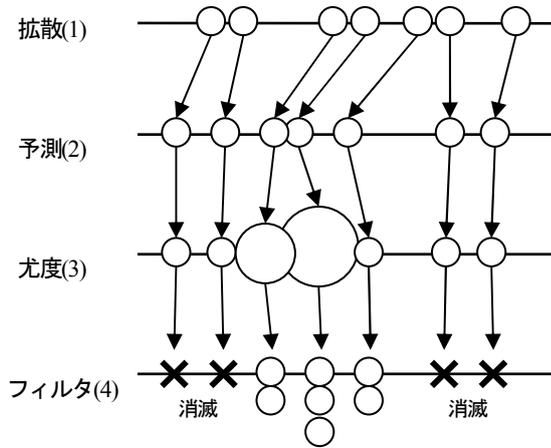


図-1 パーティクルフィルタ概念図

(2) 運動モデル

パーティクルフィルタでは、推定対象の動作特性をモデル化したもので、モデルの精度を左右するものである。本研究においては、歩行者プローブデータを想定しているため、移動方向や移動速度が不規則に変化することが考えられ、運動モデルの予想が困難である。そこで、本稿においては運動モデルにランダムウォークを仮定し、長坂ら⁶⁾の論文においても高い位置推定精度が報告されている式(3)を用いた。

$$x_t^i = x_{t-1}^i + \varepsilon \quad (3)$$

$$\varepsilon = \begin{bmatrix} d \cdot \cos(\theta) \\ d \cdot \sin(\theta) \end{bmatrix} \quad (4)$$

ただし、 $d \sim N(0, \sigma^2)$ とし、現時刻 t における i 番目のパーティクルの状態ベクトル x_t^i は、一時刻前の状態ベクトル x_{t-1}^i を、 θ 方向に、 d 距離だけ拡散させる。また、 d は、平均 0 、分散を σ^2 とする正規乱数である。

しかしながら、今回用いた運動モデルは拡散方向 θ に一様乱数のみ生成しているため、パーティクル移動空間内の障害物等は考慮されていない。そのため、実道路空間への適用の際は、建築物の壁面や路側帯などの障害物も考慮するモデルを構築する必要がある。こ

れらの実装に関しては今後の課題としたい。

4. まとめ

本稿においては、歩行者プローブパーソンデータのクリーニング手法の一つとして、プローブデータの道路面へのマップマッチングを目的とした、ベイズの定理をもとにしたパーティクルフィルタの援用について、アルゴリズムの検討を行った。本稿においては、実装の段階でパラメータ値の設定とプログラミングにおいて初歩的なミスが見つかり、実プローブデータの実装までには至らなかった。

今後の課題としては、GPS携帯端末より得られた実プローブデータを地図上の道路領域においてマッチングを行い、アルゴリズムの有用性を検証するとともに、データクリーニング手法の基盤アルゴリズムとなるよう改良を重ねていく予定である。

なお、今回のアルゴリズムの実装結果の一部については、発表時までには修正し当日紹介する。

参考文献

- 1) 原田昇, 吉井稔雄, 牧村和彦: プローブデータを取り巻く動向と課題, 土木計画学研究発表会講演集, Vol.31, 2005.
- 2) 国土交通政策研究所: 次世代マルチモーダルITS研究会報告書, 2004.
- 3) 小島英史, 羽藤英二: プローブパーソンデータによるオンラインマッチングアルゴリズム, 土木計画学研究発表会講演集, Vol. 29, 2004.
- 4) 三輪富生, 木内大介, 山本俊行, 薄井智貴, 森川高行: 低コストプローブカーデータのオンラインマップマッチング手法の開発, 交通工学, 2009年(印刷中).
- 5) 北澤桂, 小西勇介, 柴崎亮介: Personal Positioning System におけるマップマッチング法の提案, 地理空間情報フォーラム, 2001.
- 6) 長坂康史, 金子尚人: アクティブRFID技術を用いた移動体位置推定アルゴリズムに関する研究, 広島工業大学紀要研究編, 第43巻, pp299-304, 2009.
- 7) Michael Isard, Andrew Blake: Condensation-conditional density propagation for visual tracking, International Journal of Computer Vision, Vol.29, No.1, pp.5-28, 1998.
- 8) 加藤丈和: パーティクルフィルタとその実装法, 情報処理学会研究報告(CVIM), 2007.