

# 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\*

Development of on-line map-matching algorithm for low-cost probe-vehicle data \*

三輪富生\*\*・木内大介\*\*\*・山本俊行\*\*\*\*・森川高行\*\*\*\*\*

By Tomio MIWA\*\*・Daisuke KIUCHI\*\*\*・Toshiyuki YAMAMOTO\*\*\*\*・Taka MORIKAWA\*\*\*\*\*

## 1. はじめに

プローブカーシステムは路側観測機器などのハード的インフラを必要としないため、詳細な交通情報データを比較的安価に収集可能である。しかし、より大規模に、またはより長期間データを収集する場合には、データ収集費用が実験の継続への支障となり得る。特に、大規模なデータ収集において、データをリアルタイム利用する場合には、データ送信コストを抑えるために、送信頻度を低く設定せざるを得ない。また、タクシー会社や運送業者等が運用する車両管理システムは、情報量が少なく（車両位置座標と時刻のみなど）、また低頻度なデータ収集（1分以上や数百m間隔）ではあるが、FM無線やデジタルMCA等によるリアルタイムな収集であり、この車両管理データをプローブカーデータとして利用できれば、情報収集エリアを大きく拡大できる。

このような低頻度かつ情報量の少ないプローブカーデータ（低コストプローブカーデータ）を、オンラインで（またはリアルタイムに）マップマッチングを行う場合には、生成される情報の精度をなるべく低下させないようなマッチング手法が必要となる。特に、大都市の都心部のように細街路が多い地域においては、プローブカー位置座標間を最短距離経路で結ぶ方法では、細街路を多く通過したり、右左折が多いなど、非現実的な経路を走行経路として特定する可能性が高い。

そこで本研究では、上記のような低コストに収集されるプローブカーデータを、オンラインでマップマッチングする手法を開発することを目的とする。

## 2. 既往研究

プローブデータをオンラインでマップマッチングする手法はこれまでも提案されている<sup>1)2)</sup>。特に、近年では、カーナビなどの車載システムに依存しない交通情報収集の必要性から、誤差が大きく、かつ低頻度な場合にも適用可能なマッチングアルゴリズムの必要性を示す研究もある<sup>3)</sup>。ここでは、低コストプローブカーデータを処理することが可能な幾つかの既往研究について概説し、それらの解決すべき問題点を明らかにする。

小島ら<sup>4)</sup>は、オンラインでのマップマッチングにおいて、GPS誤差によるマッチングミスの発生を軽減するため、既に走行位置が確定したプローブデータ位置座標（ここでは第1プロットと呼ぶ）から、距離閾値を超える2つの新たな位置座標（第2、第3プロットと呼ぶ）を取得後、第3プロットを近くのリンクに仮にマッチングし、第1プロットと第3プロットの間を結ぶ最短距離経路上に第2プロットをマッチングする方法を示している。これにより、第2プロットをGPS誤差の影響で座標近くのリンクに誤ってマッチングすることを防ぐことができる。しかし、探索は最短距離経路としており、細街路を多く含む都心部のネットワークに対しては、実際より細街路を多く通過したり、右左折が多いなど、非現実的な経路にマッチングされる可能性がある。また、高速道路が一般道路と並行して立体配置されている区間では、第3プロットを最も近いリンクに仮にマッチングすることで、第2プロットのマッチング精度を低下させる恐れがある。この点については4. で詳しく述べる。

Yang et al.<sup>5)</sup> は、リンクが密集した交差点付近でのマッチングエラーを防ぐため、各交差点ノードに対して一定の半径を持つNodeBufferを設定し、各プロットがこれに含まれる場合は当該ノードにマッチングする方法を示している。それ以外の場合は、近接のリンクにマッチングを行うが、マッチング候補とすべきリンクが複数存在する場合は、マッチングを保留し、以降のプロットがマッチングされたリンクと先の候補リンク間で最短距離経路を探索し、長さの短い経路上にマッチングするとしている。この方法は、収集したデータを即座にマッチングすることを想定しており、比較的リアルタイム性が高く、また、交差点付近でのマッチングミスは少ない。しかし、都心部のネットワークにおいては、各プロットに対する候補リンクが複数存在する 경우가多く、最短経路探索をいくつかの区間に分けて行うため、おのずと処理が煩雑になることが予想される。さらには、リンク旅行時間を正確に算出するためには、NodeBufferにマッチングされたプロットを再度経路上にマッチングし直す必要がある。

本研究では、これらの研究の問題点を踏まえ、細街路や立体的な高速道路が多く配置された都心部においても、精度の高いマッチング方法を開発する。ただし、本研究では低コストプローブデータを対象としているため、多くの既存研究で用いられているような、車両の瞬間速度や進行方向等の情報は用いず、位置座標および時刻のみからの走行経路特定が可能であることを前提とする。

\*キーワード：交通管理，交通情報，ITS

\*\*正員，博士（工学），名古屋大学大学院環境学研究科  
（名古屋市千種区不老町，TEL: 052-789-3565，  
E-mail: miwa@trans.civil.nagoya-u.ac.jp）

\*\*\*正員，工修，三菱商事株式会社

\*\*\*\*正員，博士（工学），名古屋大学大学院工学研究科

\*\*\*\*\*正員，Ph.D.，名古屋大学大学院環境学研究科

### 3. データの概要

#### (1) プローブカーデータ

本研究で使用するデータは、2002年10月～2003年3月に名古屋都市圏で収集されたプローブカーデータのうち、試験的に50mおよび5秒間隔といった高頻度な収集を行ったデータである。この高頻度プローブカーデータから、分析に適切と思われるデータ区間を抽出し、これをデジタル道路地図（DRM）基本道路網上に手作業でマッチングを行うことで、予め正しい走行経路を特定した。さらに、その後、データを間引くことで低頻度データを作成した。使用した高頻度データを表-1に示す。

分析に適切とする基準は、高速道路を少なくとも1度は使用していることや、適度なうろつき走行を行っていることなどであり、実空車に関しては抽出の基準とはしていない。また、低頻度データの作成は、50m間隔のデータからは、50mごと間隔を広げ最大で500m間隔のデータを、5秒間隔のデータからは、5秒～15秒ごとに最大で90秒間隔のデータを作成した。

表-1 使用した高頻度データ

車両ID	収集頻度	収集年月日	取得時刻
1006	50 m	2002/11/15	8:00～17:00
3014	50 m	2002/10/31	13:00～23:00
1173	5 sec	2002/12/25	11:00～16:00
1173	5 sec	2002/10/29	19:00～25:00

#### (2) GPS誤差と移動閾値

プローブカーが同一地点に滞在しているのか、もしくは移動しているのかを判断するためには、GPS座標がそもそも有する誤差を把握する必要がある。そこで、前述の高頻度データと正解経路を用いて、GPS誤差を確認する。また、これにより、プローブカーが移動しているとみなせる距離閾値（以降、移動閾値と呼ぶ）を確認する。

図-1は、使用データの各位置座標から正解経路までの距離より計算された、GPS誤差の発生確率を示している。図より、データの95%が正解経路から30m以内に位置することが分かる。これは、連続する2つのプロットが30mより離れた場合は、95%の確率で移動したとみなせることを意味している。したがって、本研究では、移動閾値を30mとする。ただし、DRMから作成された正解経路は道路幅員等を考慮していないため、図-1が厳密な意味でのGPS誤差を表すわけではない。

#### (3) マッチング精度指標

既存研究でも頻繁に使用されるマッチング精度指標、プロットの中率は、正しいリンクにマッチングされたプロット数と総プロット数の割合である。しかし、低頻度プローブデータでは、プロットのみを正しくマッチングできても、その間の走行経路を間違えば、生成されるリンク旅行時間等の交通情報は誤った情報となる。そこで、本研究では以下のような2つの精度指標を定義する。

経路一致率(%) =

正しく特定された区間長 / 正解経路長 × 100

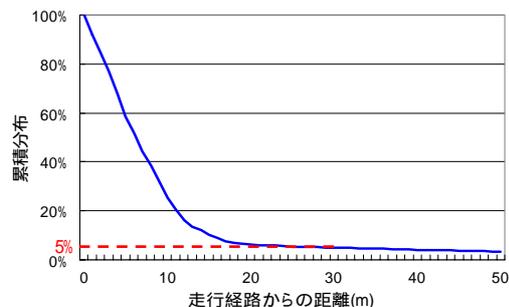


図-1 GPS座標が走行経路から離れる確率

経路迂回率(%) =

誤って特定された区間長 / 特定された経路長 × 100

経路一致率<sup>1)</sup>は、間違った区間が含まれる場合でも特定された経路に正解経路がすべて含まれれば100%となる。一方、経路迂回率は、マッチングによって特定された全経路長のうち、誤って特定された経路の割合を表す。したがって、完全に正しく経路が特定されれば、経路一致率 = 100%、経路迂回率 = 0%となる。

### 4. オンラインマップマッチング手法の開発

ここでは、低コストプローブカーデータを対象としたマップマッチング手法の開発を行う。ここでは、まず開発のベースとなるアルゴリズムを適用し、問題点を明確にした後、より精度の高いマッチング手法を開発する。

#### (1) ベースアルゴリズム

ここでは、開発の基本となるアルゴリズムを示す。これは小島ら<sup>4)</sup>とほぼ同様であり、処理手順は以下の通り。

**Step1:** 最初のプロットを最も近いリンクにマッチングする。以後、最新のマッチング済みプロットを第1プロット、マッチングされたリンクを第1リンクと呼ぶ。

**Step2:** 第1プロット以降に取得され、第1プロットから移動閾値を超えるプロットを第2プロットとする。また、移動閾値を超えないために、第2プロットとはならない（複数の）プロットを第2待機プロットと呼ぶ。

**Step3:** 同様に、第2プロットから移動閾値を超えるプロットを第3プロットとし、第2プロットから移動閾値を超えない（複数の）プロットを第3待機プロットとする。第2プロットからの進行方向を考慮して、第3プロットを近接リンクに仮にマッチング（仮マッチング）し、このリンクを第3リンクと呼ぶ。

**Step4:** 第1リンクと第3リンクの間で最短距離経路を探索し、この経路上に、第2待機プロット、第2プロットをマッチングする。

**Step5:** 第2プロットがマッチングされたリンクを第1リンク、第2プロットを第1プロット、第3（待機）プロットを第2（待機）プロットとし、Step3へ戻る。

ベースアルゴリズムによるマッチング精度を図-2に示す。図より、データ収集頻度が低くなるにしたがって経路一致率が大幅に低下しており、特に、距離周期データと時間周期データでは、経路一致率低下の傾向に違い

がみられる。つまり、距離周期データの方が、頻度低下に伴う経路一致率の低下が顕著となっている。これは、時間周期データでは、低速走行や交差点での停止等により実質的にはプロット間の移動量が少ないために、経路一致率がそれほど低下しないことを意味している。

一方、経路迂回率は、距離間隔データが常に大きな値を示しているのに対して、時間間隔データでは低頻度化によってむしろ精度が向上している。これは、高頻度の時間周期データは、交差点などのリンクが密集した地点でGPS誤差を含む多くのデータが存在すると、かえって精度が低下してしまうことを示唆している。

精度低下の原因をより詳細に調べると、図 - 3 a, b に示すようなマッチングエラーが多くみられた。図 - 3 a は、第1プロットと第3プロット間で最短距離経路を用いたために、第2プロットから離れた経路を特定している。図 - 3 b は、第3プロットを一般道路と立体的に配置する都市高速道路に仮マッチングしたために、探索経路がオンランプを通過するような、非現実的な走行経路が特定された結果である。

## (2) リンクコストの考慮と複数経路からの特定

### a) リンクコストの設定

第1, 第3リンク間の経路探索において、第2プロット的位置を考慮しないリンクコストを用いることによるマッチングエラー(図 - 3 a)は、リンクコストを適切に設定することで解決できる。そこで、本研究では以下のような2つのリンクコストを設定した。

$$cost_i = D_{i2} \times l_i \quad (1.a)$$

$$cost_i = \ln(Prob_{i2}) \times l_i \quad (1.b)$$

ここに、それぞれリンク  $i$  について、 $cost_i$  はコスト、 $l_i$  はリンク長、 $D_{i2}$  は最も近い第2(待機)プロットまでの距離、 $Prob_{i2}$  は最も近い第2(待機)プロットがリンク  $i$  に存在する確率(図 - 1 より)。

### b) 複数の候補経路からの選別

第3プロットの仮マッチングリンク(第3リンク)を、第3プロットから最も近いリンクとすることによるマッチングエラー(図 - 3 b)は、第3リンクを複数設定することで解決できる。ここでは、第3リンクを最大3リンクまで設定し、それぞれについて第1リンクから経路探索を行い候補経路集合とする。しかし、候補経路集合には、図 - 3 bに示すような非現実的な経路も含まれる。そこで、第1プロットと第3プロット間の時刻差  $T$  と下式で表される  $T_{limit}$  を比較し、 $T_{limit} > T$  の場合はこれを候補経路としない。

$$T_{limit} = L_e / v_e + L_g / v_g \quad (2)$$

ここに、ある候補経路に対して、 $L_e$  ( $L_g$ ) は高速道路(一般道路)区間長、 $v_e$  ( $v_g$ ) は高速道路(一般道路)の速度閾値であり、本研究では、GPS誤差による見かけ上の走行速度誤差を考慮して120(80) km/h とした。

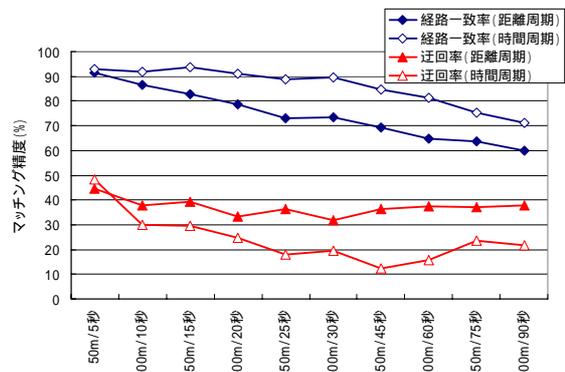


図 - 2 ベースアルゴリズムによるマッチング精度



図 - 3a 最短距離経路によるミスマッチング



図 - 3b 高速道路へのミスマッチング

### c) マッチング経路の特定

複数の候補経路からマッチング経路を特定するための評価値を以下のように設定する。

$$\min E(k) = j_2 d_{j_2,k} + d_{2,k} + j_3 d_{j_3,k} + d_{3,k} \quad (3.a)$$

$$\max E(k) = j_2 \ln(Prob_{j_2,k}) + \ln(Prob_{2,k}) + j_3 \ln(Prob_{j_3,k}) + \ln(Prob_{3,k}) \quad (3.b)$$

ここに、候補経路  $k$  に対して、 $E(k)$  は経路  $k$  の評価値、 $d_{2,k}$  ( $d_{j_2,k}$ ) は第2プロット(各第2待機プロット)からの距離、 $Prob_{2,k}$  ( $Prob_{j_2,k}$ ) は第2プロット(各第2待機プロット)が経路  $k$  上に位置する確率(図 - 1 より)。

### d) マッチング精度

ここまで示したリンクコストおよびマッチング経路評価式はそれぞれ2種類であるが、ここではマッチング精度が最も高かった、式(1.a), (2), (3.a)を適用した場合の、距離周期データに対する結果のみを示す。

ベースアルゴリズムでは、500m間隔データに対して経路一致率が約60%であったが、図 - 4より、経路一致率の精度向上には、第2プロット的位置を考慮したリンクコストの適用が効果的であることが分かる。経路迂回

率においては、第2プロットの位置を考慮したリンクコストの適用は低頻度データに、複数経路の考慮は高頻度データにそれぞれ効果が大きくなっている。これは、高頻度データにおいては、第2プロットの位置を考慮したリンクコストの適用のみではGPS誤差の影響によるマッチングミスを防ぐことができないことを示している。いずれにしても、ここで示した手法により、マッチング精度が大きく改善することが示された。

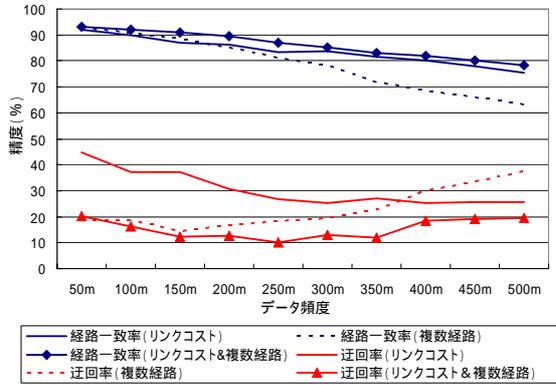


図 - 4 リンクコスト/複数経路適用後のマッチング精度

### (3) ドライバーの経路選択行動を考慮した方法

低頻度プローブカーデータでは、連続するプロット間に走行可能経路が複数存在する。特に、都心部のような細街路が密なネットワークにおいては、位置座標のみからは走行経路を特定することが難しい。ここでは、候補経路からマッチング経路を特定する際に、ドライバーの経路選択行動を考慮することで精度の向上を図る。

#### a) 経路選択モデル

ここでは、ドライバーの経路選択行動をロジットモデルで表現し、効用関数を下式のように表す。

$$V_k = \sum_i \left\{ \left( \beta_{i,0} + \sum_j \beta_{i,j} \bar{x}_j \right) x_{k,i} \right\} \quad (4)$$

ここに、 $x_{k,i}$  は経路  $k$  の  $i$  番目属性値、 $\beta_{ij}$  は未知パラメータ、 $\bar{x}_j$  は走行済み経路の  $j$  番目属性値である。

ここで、式(4)は、直前まで走行していた経路が以降の経路選択行動に影響を与えるとしている。また、モデル構築用データは、使用データの各プロットに対して式(1.a)、(2)を適用した候補経路と正解経路である。

推定結果を表 - 2 に示す。すべてのパラメータが有意に推定され、モデル適合度も高いことが分かる。また、直前経路の属性に関するパラメータが有意に推定されたことから、直前に走行していた経路の属性が以降の経路選択行動に影響を与えることが示された。

#### b) マッチング精度

次いで、距離間隔データに対して、上記の経路選択モデルを適用した結果を示す(図 - 5)。この結果より、(2)で開発した方法と比較して、若干ではあるがマッチング精度が向上していることが分かる。特に、経路迂回率の減少は、非現実的な細街路通過や右左折の減少に

表 - 2 経路選択モデル推定結果

説明変数		パラメータ (t値)	
高速道路区間長 (m)		-0.500	(-6.93)
一般幹線道路区間長 (m)	定数項	-0.157	(-2.04)
	直前経路 $\beta_{i-1}$	4.814	(4.36)
細街路 細街路の右左折数		-59.242	(-5.16)
細街路 細街路以外の右左折数		-32.206	(-2.93)
幹線 細街路 切替り回数		-64.891	(-4.65)
単位距離あたりの旅行時間 (sec/m)	定数項	-4.412	(-6.86)
	夜 $\beta_{i-1}$	1.727	(2.86)
	直前細街路 $\beta_{i-1}$	1.902	(2.84)
	直前高速 $\beta_{i-1}$	-10.078	(-1.97)
サンプル数		427	
修正 R 値		0.608	
的中率		0.867	

\*1 直前経路の最後のリンクが一般幹線の場合1, それ以外の場合は0  
\*2 PM7:00 - AM7:00において1, それ以外の場合は0

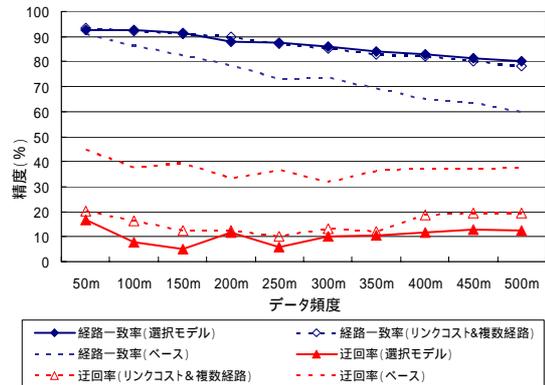


図 - 5 経路選択モデルを用いたマッチング結果

よる結果である。ただし、ここではモデル推定と適用に同じデータを用いているため一般性は乏しいが、経路選択モデルを適用することの可能性は示せたといえよう。

### 5. おわりに

本研究では、低頻度かつ情報量の少ないプローブデータにたいして、オンラインに実行可能なマップマッチング手法を開発した。既存の手法をベースとして、GPS誤差を考慮したリンクコストの適用や複数の候補経路からの選別、特定により、既存手法を大幅に改良可能であることを示した。また、ドライバーの経路選択行動を考慮することでも、精度向上の可能性を示した。今後は、実際の低コストプローブデータへの適用を通じて、マッチングエラー等の詳細な検討を行った上で、より精度の高いマッチング手法を確立する必要がある。

#### 参考文献

- White, C.E., et al.: Some Map Matching Algorithms for Personal Navigation Assistants, Transportation Research Part C 8, pp.91-108, 2000.
- Greenfeld, J.S.: Matching GPS Observations to Locations on a Digital Map, 81st Transportation Research Board, CD-ROM, 2002.
- Hellinga, B., et al.: Obtaining Traveller Information via Mobile Phone Location Referencing - Challenges and Opportunities, 2003 TAC (Transportation Association of Canada) annual conference proceedings HP.
- 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プマッチングアルゴリズム, 土木計画学研究・講演集, Vol.29, CD-ROM, 2004.
- Yang, J., et al.: The Map Matching Algorithm of GPS Data with Relatively Long Polling Time Intervals, Journal of Eastern Asia Society for Transportation Studies, Vol.6, pp.2561 - 2573, 2005.