

## プローブデータのクレンジング処理と車種別の運行特性分析\*

Cleansing Procedure for Probe Vehicle Data and Analysis on Travel Behavior by Vehicle Types \*

堀口良太\*\*, 清水行晴\*\*\*, 金崎貴文\*\*\*, 岡村明彦\*\*\*\*, 上田憲道\*\*\*\*, 杉崎康弘\*\*\*\*

By Ryota HORIGUCHI\*\*, Yukiharu SHIMIZU\*\*\*, Takafumi KANEZAKI\*\*\*,

Akihiko OKAMURA\*\*\*\*, Norimichi UEDA\*\*\*\* and Yasuhiro SUGISAKI\*\*\*\*

### 1. はじめに

本稿では、一連の連続した移動軌跡として記録されるプローブデータに含まれるデータ欠損区間を検出し、除去あるいは補完するとともに、その連続軌跡をトリップに分割するクレンジング処理手順について述べる。また、クレンジング処理の過程で抽出された情報をもとに、車種別の運行特性を考察する。

旅行時間推計を目的とする場合、理想的なプローブ走行軌跡として、次のようなものが望まれる。

データが通信エラーや GPS 衛星補足状況の影響を受けることなく、正常に記録されていること。  
トリップエンドで駐停車しているデータ区間が適切に除去され、トリップ中の走行状態のデータだけで構成されていること。

それらのデータが記録された地点および走行している方向が、正しくデジタル道路地図のリンク上に照合されていること。

しかしながら、実際のデータには様々な不正・誤差要因が含まれており、それらを適切な手順で除去、あるいは補完しなければ、いかなる推計アルゴリズムの適用も意味をなさない。ここでは、平成13年2月の約1ヶ月間にわたって横浜市で実施された、プローブ車両走行実験<sup>1)</sup>のデータ(以下 IPCar12)を用い、現実のプローブデータに含まれる不正・誤差要因に対するクレンジング手順を以下のように構築する。IPCar12ではタクシー、バス、塵芥車、その他商用車それぞれ約50台、合計約200台のデータをオンライン収集している。

### 2. クレンジング処理手順とその効果

#### (1) データ記録ギャップの検出

旅行時間推計を目的とする場合、プローブデータは時空間上での走行軌跡を表せる連続した「線」の情報として記録されなければならない<sup>2)</sup>。IPCar12では30秒ごとに周期的にデータを記録する仕様であるが、実際には、(A)トリップエンドでのエンジン停止に伴う車載機電源 OFF や、(B)通信エラーによって、データの記録間隔がこの周期通りにならず、連続性が保証されない区間がある。このような不連続データ区間を「記録ギャップ」と呼んでいる。

Duration vs Distance of Recording Gaps  
Dur < 300 sec., IPCar2000

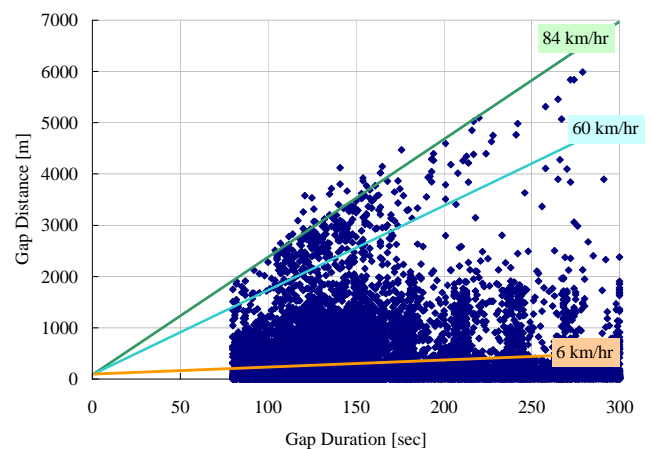


図1: 記録ギャップの継続時間と距離の関係(5分以下)

図1は80秒以上300秒以下の記録ギャップについて、その継続時間とギャップ前後の直線距離との関係を示すものである。明らかに(A)のトリップエンドでの記録ギャップは、その距離が0m±GPS計測精度の範囲にななければならないが、(B)の通信エラーによるギャップも多数検出されていることが示されている。従って、記録ギャップの検出には継続時間だけでなく速度(ここでは直線距離÷継続時間)を使って、次のようなクレンジングの戦略が適用される。

記録ギャップのうち、時間がT1秒(IPCar12では

\* キーワード: プローブ, データクレンジング, 走行特性, データ処理

\*\* 正員, 工博, (株)アイ・トランスポート・ラボ, 〒162-0824 東京都新宿区揚場町2-12-404, Tel: 03-5261-3077, E-mail: horiguchi@i-transportlab.jp

\*\*\* 日本電気(株), ITSソリューション推進本部

\*\*\*\* NECソフト(株), 第一SI事業部 ITS部

【第 26 回土木計画学研究発表会講演集 岩手大学 2002 年 11 月】  
 300 秒)以上で、かつ距離が 0m±GPS 精度未満のものは、トリップエンドでのエンジン OFF 駐車と見なし、プローブデータを区切る  
 時間が T1 以上で上記以外のものは、通信エラーによる記録ギャップと見なし、走行軌跡を区切る  
 時間が T1 未満の記録ギャップは、連続した軌跡の情報として、以降の処理を適用する。

( 2 ) GPS 検出異常区間の検出

図 2 はマルチパスが発生する関内周辺の、IPCar12 での GPS 検出異常イベント位置と、200m メッシュでの発生割合を濃淡で示したものである。多くのメッシュで GPS 検出異常発生割合が 20% 以内になっている。

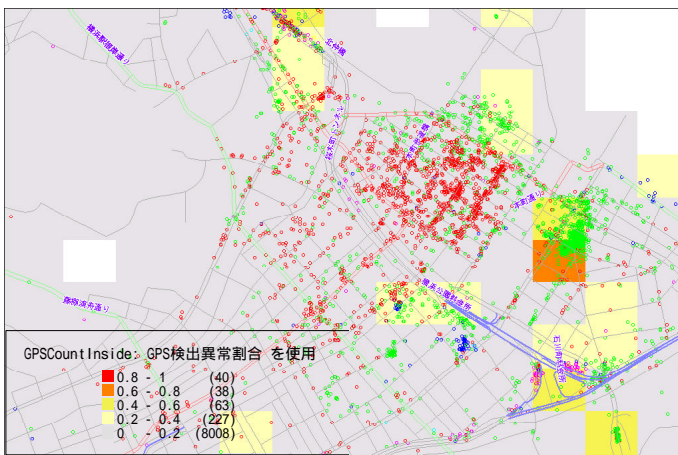


図 2：関内エリアでの GPS エラー発生状況

Duration vs Distance of GPS Error Gaps  
 Dur < 36,000 sec., IPCar2000

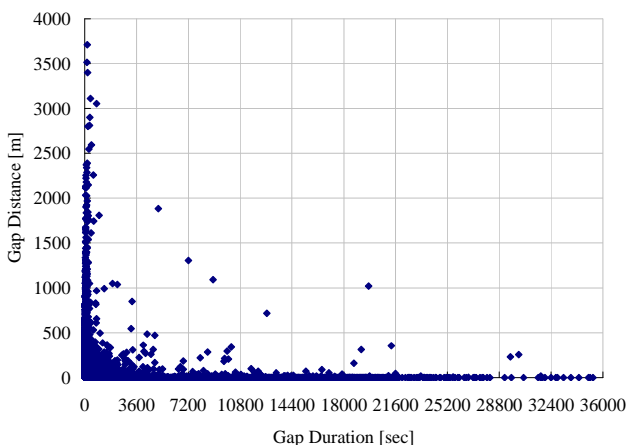


図 3：GPS 検出異常区間の継続時間と距離の関係

一定時間以上 GPS 検出異常の状態が続く場合、その区間で位置を同定して経路に照合することは困難となるため、こういった区間を検出し補完するか、あるいはそこで軌跡を区切るかを判断する必要がある。

IPCar12 では、5 分間以上 GPS 検出異常が継続する場合は、そこで軌跡を区切ることにした。

( 3 ) アイドリング駐車検出

プローブ車両の停止状態を検出することで、エンジン ON 状態での駐車をトリップエンドとして検出できる。IPCar12 では GPS 位置のみを使い、5 分間以上半径が GPS 位置精度の円内にとどまっている場合、この区間をトリップエンドとしてデータを区切っている。図 4 にアイドリング駐車区間の分布を示す。タクシーの空車待ちが多いと思われる商業地区に集中していることがわかる。

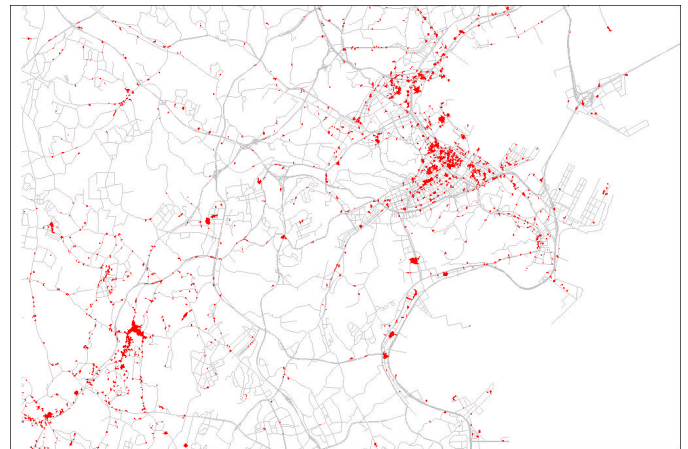


図 4：アイドリング駐車区間の分布

平成 13 年度の IPCar 実験(IPCar13)では、停止状態をショートストップとして記録するとともに、パーキングブレーキなどの車両運行状態を表す情報を並行イベントとして収集している。これらを手がかりとして、アイドリング駐車状態を検出することも有効である<sup>2)</sup>。

( 4 ) U ターンやループ軌跡のトリップエンド検出

IPCar12 では以上の処理を経ても、タクシーが利用客乗降のために短時間停車するデータ区間は検出されない。そこで、タクシーは空車待ち拠点をもち、その地点に回帰する性質を持つと仮定して、トリップエンドを検出する。すなわち、GPS 軌跡が U ターンやループする場合は、その区間のどこかで利用客が乗降し、トリップ目的地が切り替わったと想定する。図 5 はそのような軌跡が検出されたとき、当該区間の中間地点で軌跡を区切った場所を示したものである。画面左下に戸塚駅があり、IPCar12 でのプローブはこの駅を拠点とするものが多いが、検出されたトリップエンドは戸塚駅を中心とした一定距離内に分布している。

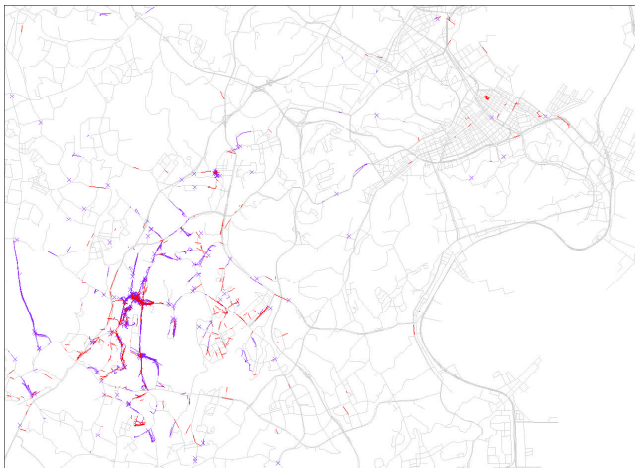


図 5：U ターンとループする軌跡を切断した箇所の分布

ここでの処理は、トリップエンドの検出という意味の他、後述するマップ照合不正区間の補正処理に適したデータを得るという意味もある。

#### (5) 細街路走行区間のクレンジング

IPCar12 では、プローブ車両に自律走行機能やマップ照合機能を持たせず、リアルタイムでデータセンターのサーバが GPS 位置をマップ照合している。このとき、処理負荷を軽減するため、照合対象リンクをデジタル道路地図(DRM)の基本道路リンクに限定している。このため、基本道路リンクに含まれない細街路を走行するプローブ軌跡は、マップ照合不可の区間となるため、クレンジング処理で除外する。

#### (6) リアルタイムマップ照合不正区間の補正

前述の通り、リアルタイムでのマップ照合は、データ量や計算量に制約があり、あまり複雑なアルゴリズムを適用できない。このため、実際には走行しないリンクに照合される状況が頻繁に発生する。これは上下線が分離されて DRM に記述されていたり、あるいは細街路が基本道路リンクに沿っていたりする区間で、とくに顕著となる。

分析や情報提供はリンク単位でなされるので、不正照合は入念に補正されなければならない。IPCar12 の解析では、オフラインで走行軌跡全体の位置データを用いて、もっともらしい経路に照合した。オフラインマップ照合のアイデアは次の通りである(図 6)。

トリップ軌跡の近傍にあるリンクほど小さいコストを与える。ある距離以上離れたリンクには無限大のコストを与える。

そのコストを元に、トリップ起点から終点まで最

短経路探索を実施する。

求められた経路上のリンクに、トリップ軌跡を照合する。

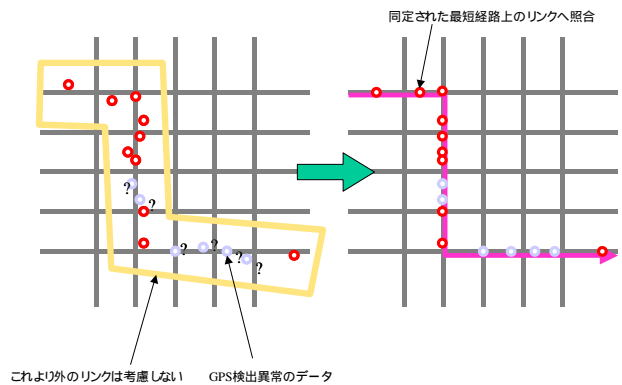


図 6：オフラインマップ照合の考え方

すなわち、連続した GPS 移動軌跡に沿って低いペナルティが与えられた場の中で、起点から終点へ経路探索を行うものであるが、明らかに軌跡が U ターンやループする区間は、照合された経路から抜け落ちてしまう。従って、ここでのオフライン照合には、前出(4)のクレンジング処理が必須となる。

図 7にオフライン照合で補正された例を示す。図は保土ヶ谷バイパスと中原街道が立体交差する地点である。DRM リンクは上下線分離で表現されており、リアルタイム照合では最も近接していた反対方向のリンクに照合されているが、オフライン照合により補正されていることがわかる。

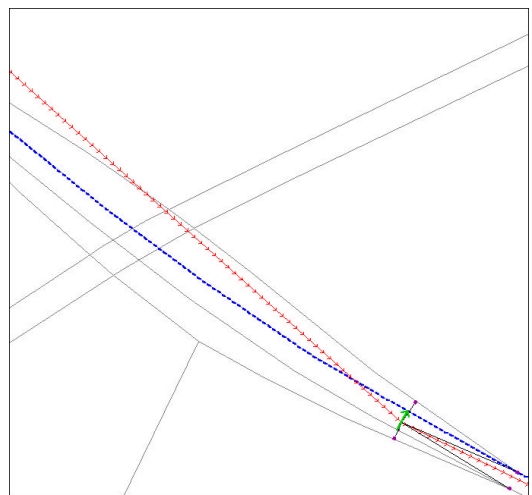


図 7：リアルタイム照合ミスの補正

### 3. 車種別プローブの運行特性分析

ここでのクレンジング処理は、旅行時間の解析に都合がよい連続した走行軌跡を得るためのものであるが、各段階で除去されたデータ区間を集計、解釈することで、プローブ運行の特性がわかる。

図8は30日間のタクシープローブの稼働時間と休止時間を車両ごとに集計したもので、平均70%の稼働率でデータ収集ができることを示している。また、図9は収集されたデータに、前出のクレンジング処理を施した過程で除去されたデータ量を示している。GPSエラーの時間とエンジンONでの駐車時間、および細街路を走行している時間の合計は、車両に因らずほぼ同じくらいの割合となっている。



図8：タクシーの稼働時間と休止時間

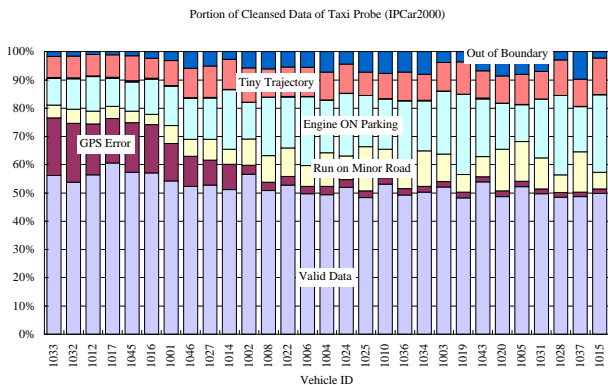


図9：クレンジングで除去されたデータ量内訳

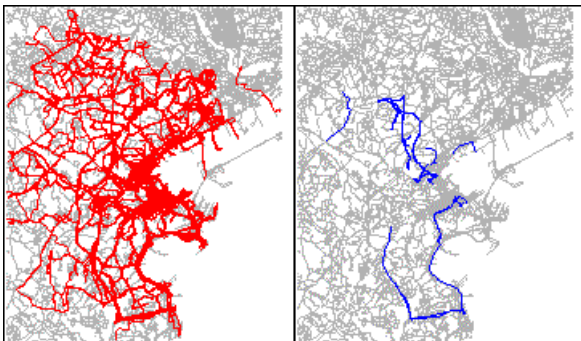


図10：塵芥車の走行範囲(左：午前中、右：夕方)

図10は塵芥車の時間帯別走行範囲を地図上に示したものである。左の午前中では横浜市全域に広がっているが、右の夕方では、ほとんどの車両が運行していないことが示されている。

車種ごとの一般的な運行特性を表1にまとめる。タクシーの稼働時間が群を抜いており、クレンジング処理を経ても、全体の35~40%(70%×50%強)で有効な走行軌跡が得られる。しかしながら、空間的な広がりが小さく、昼間よりも夜・深夜のトリップが多くなる。

バスも比較的有効な軌跡を効率よく得られる車種といえるが、走行範囲は当然ながら営業路線、営業時間帯に限定される。よって、タクシーとバスを相互に補完する形でプローブを構成すれば、時間的に偏りが少ないデータ収集システムを構築できる。

塵芥車の走行軌跡は、横浜市の場合午前中の収集時間に集中することと、細街路を走行する割合が多いことがデメリットであるが、市街全域に広がっている。また、その他業務車両は相対的に稼働時間が少なく、データ収集効率という面ではデメリットがある。

	タクシー	バス	塵芥車	その他	
稼働時間比率	70%	40%	30%弱	10%弱	
内訳	アイドリング駐車	15-20%	20%弱	20-30%	15%
	細街路走行	10%	5%弱	20-25%	5%
	範囲外走行	4-8%	2-3%	-	-
	短距離軌跡	10%	5-8%	5-10%	10%
	GPSエラー	10%弱	10%弱	4-5%	10%弱
有効軌跡	50%強	70%	50%弱	60%強	
空間的な広がり	拠点を中心に5km位	営業路線に限定	横浜市域全体	横浜市外へも広がる	
時間的な分布	夜、深夜に多くなる	営業時間に限定	午前中が中心	業務時間に限定	

表1：車種別の平均的な運行特性

#### 4. まとめ

以上において、プローブデータのクレンジング処理とその過程で得られる情報から車種別の運行特性を分析した。

謝辞：本研究の実施にあたり、データを提供いただいた(財)自動車走行電子技術協会、並びにIPCarプロジェクト参加者各位に謝意を表します。

#### 参考文献

- 1) 自動車走行電子技術協会：「ITSの社会的有効性に係るシステムの実証」,平成14年3月
- 2) 堀口良太：走行イベント単位でのプローブデータ記録方式の効用,第26回土木計画学研究発表会講演集(今回発表予定),2002