

情報の信頼性と収集頻度を考慮したプローブカーの最適配置計画*

Optimum location of probe-vehicles considering the data frequency and information reliability *

石黒洋介**・三輪富生***・山本俊行***・森川高行****

By Yousuke ISHIGURO**・Tomio MIWA***・Toshiyuki YAMAMOTO***・Takayuki MORIKAWA****

1. はじめに

昨今の IT の進展により、道路ネットワーク上を実際に走行する車両をセンサーとして利用するプローブカーシステムが実用化しつつあり、これにより高水準の旅行時間情報を得ることが可能となってきた。収集されたプローブカーデータから平均的な道路交通状況に関する情報（リンク旅行時間や走行速度等）を生成する場合、個々の車両の走行状態やデータ処理時のエラーなどの理由で生じるばらつきを考慮する必要がある。このため、プローブカーデータから信頼性の高い情報を生成するためには、ある程度のサンプルデータ数が必要となる。また、プローブカーシステムを新たな地域へ導入する場合には、どのエリアにどの程度のプローブ車両数が必要であるかを検討しなければならない。これまでの名古屋におけるタクシープローブシステムにおいては、各営業所への配車台数が詳細な検討をされないまま決められており、データ収集エリアにむらが生じている。

このような背景の下、本研究では、これまでに名古屋市内周辺で収集されたプローブカーデータを利用して、信頼性の高い情報を生成するために必要なプローブカーサンプル数（必要サンプル数）を分析するとともに、リンク属性を考慮したプローブカーのリンク通過頻度の分析を行う。さらに、得られた分析結果から、効率的なデータ収集のためのプローブカー最適配置計画を行う。

2. 使用データの概要

本研究で用いたプローブカーデータは、名古屋市内において、2002 年 10 月から 2003 年 3 月までの平日にデータ収集に用いられた 655 台のプローブカー（タクシー）データである。これを、著者ら¹⁾が開発したシステムにより、DRM 基本道路網上にマップマッチングしたデータを用いる。分析対象エリアは、名古屋市中心とした 2 次メッシュ 9 枚とし、交差点内リンクを除く全ての幹線道路（一般県道・指定市市道以上の道路）リ

ンク（リンク数：13116）である。なお、対象エリアには、分析データの収集に用いられたプローブカーが配車された、全てのタクシー営業所が含まれている。

また、以降の分析において使用する各 DRM リンクの沿道状況に関する情報は、GIS を用いて国勢調査データや商業統計データから集計している。

3. プローブカー通過頻度モデル

(1) 負の二項分布回帰モデル

ある一定期間中のある時刻帯に、プローブカーが各リンクを通過する回数は、リンクの属性、すなわち車線数やリンク交通量、沿道の店舗数や人口、配車されたタクシー営業所からの距離等によって変化すると考えられる。ここでは、これらリンク属性を考慮したプローブ車両通過頻度モデルを構築する。

車両のリンク通過発生がポアソン過程に従うと仮定する場合、あるリンク k を車両が y_k 台通過する確率はポアソン分布で表され、発生強度 λ_k にリンク間の非観測異質性を考慮した場合、負の二項分布が導かれる。非観測異質性にガンマ分布を仮定すると、負の二項分布は以下の式(1)ように表される。

$$p(y_k) = \frac{\Gamma(y_k + \theta) \left(\frac{\lambda_k}{\lambda_k + \theta} \right)^{y_k} \left(\frac{\theta}{\lambda_k + \theta} \right)^{\theta}}{y_k! \Gamma(\theta)} \quad (1)$$

ここに、 $1/\theta$ は非観測異質性の分散パラメータを表す。

本研究では、発生強度 λ_k を、リンクとタクシー営業所との関係を考慮して以下の式で表す。

$$\lambda_k = \exp \left(\sum_i \beta_{i,k} x_{i,k} \right) \times \sum_l \{ n_l \times \exp(-1 \times \beta_j \times L_{kl}) \} \quad (2)$$

ただし、 β は未知パラメータ、 x_{ik} は説明変数を表し、 n_l は営業所 l の配車台数、 L_{kl} はリンク k と営業所 l との距離を表す。

説明変数 x_{ik} として、①高速道路ダミー、②多車線ダミー、③12 時間交通量、④都心（栄交差点）に対する上り方向ダミー、⑤沿道店舗数、⑥沿道店舗の総売上額、⑦人口集中地区ダミー、⑧沿道就業者人口、⑨沿道夜間人口、⑩リンクと各タクシー営業所との距離、を用いた。ただし、⑤沿道店舗数、⑥沿道店舗総売上額に関しては

*キーワード：プローブカー、最適配置、GA

**学生員，工修，名古屋大学大学院工学研究科
(名古屋市千種区不老町，TEL: 052-789-3729，

E-mail: yosuke@trans.civil.nagoya-u.ac.jp)

***正員，博士（工），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 次メッシュ内集計値, ⑧沿道就業者人口, ⑨沿道夜間人口に関しては, 4 次メッシュ内集計値として得られているため, メッシュ内での一様性を仮定し, 各メッシュ内集計値に対して, “リンク k の長さ/各メッシュ内に含まれるリンクの総延長”を乗じて算出した.

(2) パラメータ推定結果

2002 年 11 月の平日に収集されたプローブデータから, 0 時から 23 時までの各 1 時間帯の最初の 10 分間 (すなわち, 0:00~0:10, 1:00~1:10, ..., 23:00~23:10) に観測された各リンクのプローブカー通過台数 (台/10 分・月) を集計した. その上で, 各リンクに対して式(1), 式(2)から算出される 1 ヶ月間の各時間帯最初の 10 分間の通過回数が, 集計された通過回数と一致する確率の全リンクの積が最大となるよう, 最尤推定法によりパラメータを推定した.

推定結果は, 24 の時間帯ごとに得られるが, 例として, 表-1 に 8:00~8:10 および 0:00~0:10 の推定結果を示す. 都心上り方向ダミー及び就業人口のパラメータが朝ピーク時間帯では正に, 深夜では負に推定されており, 朝ピークでは都心に向かうリンクや商業エリアでプローブ通過頻度が高いことが分かる. また, 夜間人口についてはその逆となっている. さらに, タクシー営業所からの距離のパラメータが深夜ほど小さくなっていることから, 深夜ほど営業所から遠くのリンクで通過頻度が相対的に高くなることが分かる.

表-1 プローブカー通過頻度モデルの推定結果

説明変数	8:00~8:10		0:00~0:10	
	推定値	(t 値)	推定値	(t 値)
定数項	-5.39	(-61.7)	-5.43	(-61.1)
高速ダミー	-1.08	(-11.3)	-1.81	(-19.2)
多車線ダミー	0.99	(24.8)	0.45	(12.5)
12 時間交通量 (100 台)	2.34	(16.5)	4.78	(32.3)
都心上り方向ダミー	0.08	(2.9)	-1.23	(-47.3)
店舗数	4.37	(5.4)	3.95	(5.6)
沿道総売上額 (万円/年)	2.01	(6.3)	1.65	(5.9)
DID ダミー	0.85	(13.1)	1.10	(20.1)
就業者人口 (千人)	1.67	(1.1)	-3.89	(-3.1)
夜間人口 (千人)	-1.94	(-2.5)	1.30	(2.1)
タクシー営業所距離 (km)	0.32	(43.4)	0.18	(28.8)
サンプル数	40002		73001	
AIC	22739.4		28435.3	

4. 必要サンプル数に関する分析

(1) 必要サンプル数の考え方

道路区間 (リンク i) の時刻別平均旅行時間などの道路交通情報生成において, ある信頼水準を満たすための必要サンプル数 (ここでは, Ne_i と表記する) は, Cheu ら²⁾ のサンプル数の違いによる幹線道路の車両走行速度の推定の研究に示されるとおり, 各リンクの真の母平均旅行時間の区間推定式より許容誤差を ε , 信頼水準を $(1-\alpha)\%$ とすると以下のように求められる.

$$Ne_i \geq \left\{ s_i \times t_{(N-1)} \left(\frac{\alpha}{2} \right) / \varepsilon \bar{x}_i \right\}^2 \quad (3)$$

ただし, s_i はリンク i で観測されたプローブリンク旅行時間の不偏標準偏差, $t_{(N-1)}(\alpha/2)$ は自由度 $(N-1)$ の t 分布の $\alpha/2\%$ 点, $\bar{x}_i$ はサンプル平均旅行時間である. ここで, t 分布は必要サンプル数 Ne によって変化するため, 繰り返し計算による必要サンプル数 Ne の算出が必要となる.

各リンクの必要サンプル数の計算は, 2002 年 10 月から 2003 年 3 月までの 6 ヶ月間に, 各 DRM リンクを 5 ~ 80km/h で走行したプローブカーデータを用い, 各時間帯最初の 5 分間 (すなわち, 0:00~0:05, 1:00~1:05, ..., 23:00~23:05) の観測プローブカー通過台数とその観測リンク旅行時間から, 信頼水準 90%, 許容誤差 10% として繰り返し計算により求めた. ここで, 観測リンク旅行時間を 5 ~ 80km/h の範囲にした理由は, GPS エラーやマップマッチングエラーなどにより発生する, 不適切なばらつきを排除するためである.

ここで, 必要サンプル数は式(3)により求められるが, この値は不偏標準偏差 s_i の影響を強く受けることに注意が必要である. さらに, この不偏標準偏差はサンプル数の影響を強く受ける. したがって, 収集されたサンプル数の違いによる不偏標準偏差の信頼性を考慮する必要がある. そこで, χ^2 分布による母分散の信頼性を評価し, 分析に用いるデータを選別する.

真の母標準偏差を σ とすると, $(N-1)s^2/\sigma^2$ は自由度 $N-1$ の χ^2 分布 ($\chi^2_{(N-1)}$) に従い, 信頼水準 $(1-\alpha)\%$ では上側確率を χ^2_U , 下側確率を χ^2_L とすると次式が成り立つ.

$$\chi^2_{L(n-1)} \leq (n-1)s^2/\sigma^2 \leq \chi^2_{U(n-1)} \quad (4)$$

これを, σ^2 について解くと, 真の母分散 σ^2 の信頼水準 $(1-\alpha)\%$ の信頼区間は, 次式のように表すことができる.

$$\frac{(n-1)}{\chi^2_{U(n-1)}} s^2 < \sigma^2 < \frac{(n-1)}{\chi^2_{L(n-1)}} s^2 \quad (5)$$

この式を用いて, 比較的母亲標準偏差と不偏標準偏差の誤差が小さく, かつ, 6 ヶ月間データからなるべく多くのリンクが分析対象リンクとなるよう, サンプル数 (プローブカー通過台数) が 100 以上のリンクにより必要サンプル数の分析を行うこととする. サンプル数 100 のとき, 信頼水準 90% の母標準偏差と不偏標準偏差の関係は $0.90s_i < \sigma_i < 1.13s_i$ となる. すなわち, リンク i において, サンプルから求めた不偏標準偏差は, 真の母標準偏差の前後約 10% の区間に存在することになる.

(2) 必要サンプル数の推定

分析に際しては, サンプル数が十分でないリンクについても, 必要サンプル数が必要となる. そこで, 前述の議論を踏まえ, 各時間帯でサンプル数 100 以上のリン

クを用いて、必要サンプル数モデルを構築する。この際、時刻の変化に伴う走行状態の変化により、必要サンプル数が異なることを踏まえ、時間帯を考慮したモデルとする。本研究では、0~6時、7~9時、10~16時、17~19時、20~23時の5つに時間帯を分け、次式のようなモデルによりリンク*i*の必要サンプル数 Ne_i を推定可能とする。

$$\ln(Ne_i) = \sum_j \left(\bar{\beta}_j + \sum_k \beta_{j,k} \delta_k \right) x_j + \varepsilon_i \quad (6)$$

ここに、 x_j は j 番目説明変数、 δ_k は時間帯 k のダミー変数、 ε_i は誤差項である。

推定結果を表-2に示す。昼間の時間帯や朝ピーク時の都心上り方向リンクなど、混雑した時間帯ほど、リンク通過旅行時間のばらつきが小さくなるため、必要サンプル数も少なくなることが分かる。また、就業者人口や交通量が多い都心部ほど必要サンプル数が多く、夜間人口が多い郊外部ほど少なくなることが分かる。

表-2 必要サンプル数モデルの推定結果

説明変数	パラメータ	(t 値)
定数項 (共通)	3.186	(12.4)
10~16 時	-1.099	(-2.3)
17~19 時	-1.072	(-2.7)
20~23 時	-0.532	(-2.3)
都心上り方向ダミー	-0.087	(-2.8)
7~9 時	-0.261	(-2.0)
10~16 時	0.440	(3.4)
17~19 時	0.443	(4.1)
20~23 時	0.271	(4.6)
ln(店舗数)	0.188	(6.4)
7~9 時	0.179	(2.7)
10~16 時	-0.288	(-2.7)
17~19 時	-0.332	(-3.7)
20~23 時	-0.152	(-2.8)
ln(就業者人口)	0.342	(2.5)
ln(夜間人口)	-0.436	(-3.3)
ln(12h 交通量)	0.198	(7.7)
流出リンク数	0.253	(1.9)
10~16 時	-0.150	(-1.5)
17~19 時	-0.105	(-2.7)
20~23 時	-0.121	(-4.9)
サンプル数	3444	
修正済 R^2 値	0.338	

5. GAによるタクシープローブカーの最適配置計画

前章までに示した、プローブカー通過頻度モデルおよびプローブカーデータ必要サンプル数モデルを用いて、プローブカーの最適配置計画手法を考える。ここでは、655台のプローブカー（タクシー）を名古屋市周辺のタクシー営業所へと配車するものであり、予め設定したデータ収集期間中に、なるべく多くのリンクで必要とする信頼性水準を満たすよう配車計画を行う。

実際のデータ収集において、プローブカーが配車された営業所は41営業所であり、事業者の別や各営業所の配車容量を考慮しない場合は、655台を配車する全配車パ

ターンは 1.87×10^{65} パターン (= ${}_{695}C_{40}$) であり、全てを比較することは不可能である。また、各リンクでのプローブサンプル数は、式(6)により算出される必要サンプル数以上は必要としないこととする。つまり、各リンクで得られた情報の信頼性を表す指標 r_i (リンクカバー率と呼ぶ) を図-1のように考えると、この関数は、サンプル数=必要サンプル数において微分可能ではない。そこで、以降では、遺伝的アルゴリズム (GA) を用いてプローブカーの最適配車パターンを求める。

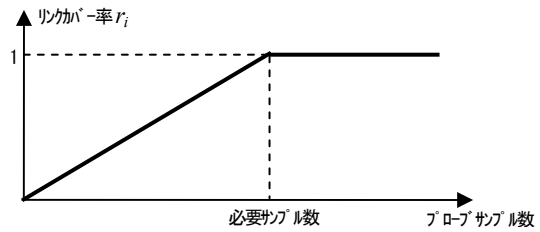


図-1 リンクカバー率

(1) 遺伝的アルゴリズムの概要

適用した遺伝的アルゴリズムの流れは以下の通り。

<Step 1> 41個のセルを持つ遺伝子配列に、合計655台となるようにランダムにプローブカーを配車する。このような遺伝子を100個作り、初期遺伝子集団とする。

<Step 2> 各遺伝子の配車パターンに対して適応度を計算する。ここで、各リンクのプローブカー通過頻度は式(2)より求められる。各遺伝子の適応度は、道路総延長に対するプローブカバー率を表す式(7)、もしくは道路網上の交通量に対するカバー率を表す式(8)と定義する。

$$CO = \sum_i (r_i \times l_i) / \sum_i l_i \quad (7)$$

$$CO_q = \sum_i (r_i \times l_i \times q_i) / \sum_i (l_i \times q_i) \quad (8)$$

$$\begin{cases} r_i = S_{ki} / Ne_i & \text{if } S_{ki} < Ne_i \\ r_i = 1 & \text{otherwise} \end{cases} \quad (9)$$

ここに、 S_{ki} は配車パターン k の時のリンク i のプローブ通過台数期待値、 Ne_i はリンク i の必要サンプル数、 l_i はリンク i のリンク長、 q_i はリンク i の12時間交通量である。また、 r_i は必要サンプル数に対する通過台数の割合（リンクカバー率）である。ただし、信頼水準を満たす平均旅行時間を得るには必要サンプル数を超えていけばよいので、 $r_i > 1$ ならば $r_i = 1$ とする。

<Step 3> 適応度の高い順から10%はそのまま次世代に残し(エリート保存方式)、残り90%は100個の遺伝子集団より適応度の高い遺伝子が高確率で選択されるように、適応度に比例して遺伝子を選択する。

<Step 4> 2つの遺伝子のセル同士を交換する(交叉)。交叉は一様交叉とし、1/2の確率でそれぞれのセルを交叉する。ここで、交叉後の遺伝子の総車両数は655台には

ならないので交叉後の遺伝子の総車両数を N とすると、それぞれの遺伝子のセル内の車両数に“ $655/N$ ”を乗じて総車両数を655台に合わせることにする。

<Step 5> 遺伝子集団が多様性を持つように、新たな遺伝子を発生させる(突然変異)。ここでは、遺伝子集団からランダムに選んだ5%を新たな遺伝子で置き換える。

以上の <Step 2>～<Step 5>を繰り返し、100世代連続で適応度の増加(成長)がみられなかった場合に終了とし、最適配車パターンとする。また、3章において、通過台数の期待値は1ヶ月間の各時間帯10分間の期待値としたので、これを2で除すことにより5分間の期待値とする。最適配置はデータ収集期間を1ヶ月、3ヶ月、6ヶ月、12ヶ月の場合を設定し、各リンクの通過台数は、期待値を期間に応じて倍することによって求める。

(2) 結果と考察

表-3に最適配置後の2つのカバー率を示す。リンク交通量を含む適応度を用いた方がカバー率が高くなっているが、交通量は都心部や大きな幹線道路などに集中しているためである。また、最適配置によってカバー率が向上していることも分かる。つまり、実際の収集に際しての配車パターンが最適ではないことが分かる。

表-3 各期間での最適配車パターンのカバー率

設定収集期間	CO		CO _q	
	最適化前	最適化後	最適化前	最適化後
1ヶ月	0.040	0.058	0.058	0.086
3ヶ月	0.109	0.137	0.153	0.193
6ヶ月	0.192	0.215	0.255	0.287
12ヶ月	0.298	0.320	0.372	0.397

図-2に、データ収集期間を3ヶ月及び12ヶ月とした場合の最適配車図を示す。当然のことながら、収集期間の拡大とともに、信頼性の高い情報が取得できるリン

ク数も増加している。また、収集期間が短い場合はリンクの多い都心部に、収集期間が長いほど郊外に分散して配車することが効率的なデータ収集を可能とすることが分かる。

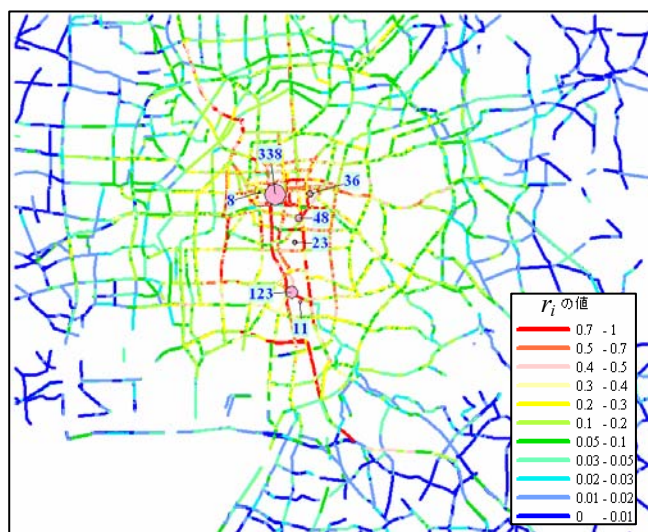
6. おわりに

本研究では、プローブカーの通過頻度モデルおよび必要サンプル数モデルを構築し、これらを用いて遺伝的アルゴリズムによりプローブカーの最適配置を行う手法を示した。これにより、データ収集期間や、必要とするデータの信頼性、総プローブ車両数等を入力値として、適切なプローブカー配置を計画することが可能となった。

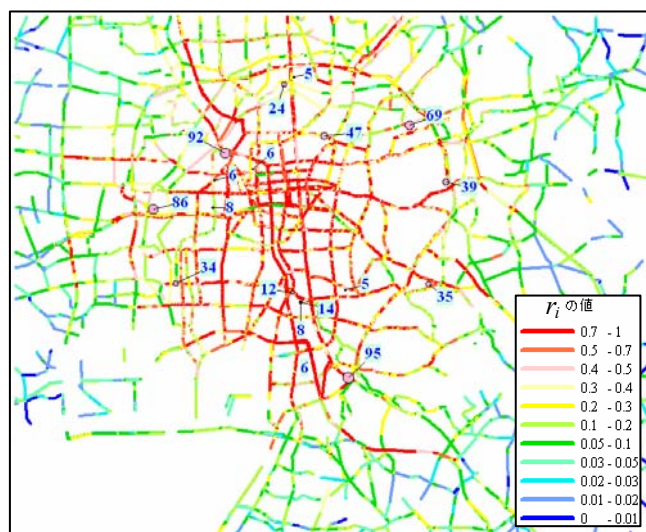
ただし、本文中にも述べたように、必要サンプル数の計算は観測データばらつきに強い影響を受ける。本研究においても、より詳細にプローブカーデータのエラーを排除する作業を行えば、最終的に算出されるカバー率や配車位置等が大きく異なることも予想される。このため、プローブカーデータのデータ処理などの理由で起こるエラーデータの除去には十分な配慮が必要となる。また、収集期間を予め定めた場合に、必要なカバー率を得るためのプローブカー台数を求める手法も開発する必要がある。つまり、短期間での情報収集が必要な場合には、このようなプローブカー必要台数の分析も実用的な着眼点であり、今後はさまざまな視点からプローブカーの配置台数の検討を行う予定である。

参考文献

- 1) 三輪富生ほか：プローブカーデータを用いた経路特定手法と旅行時間推定に関する研究，第2回ITSシンポジウム2003 Proceedings, pp.277-282, 2003.
- 2) Cheu, R.L. *et al.*: Probe vehicle population and sample size for arterial speed estimation, Computer-Aided Civil and Infrastructure Engineering 17, pp.53-60, 2002.



<データ収集期間：3ヶ月>



<データ収集期間：12ヶ月>

図-2 プローブカーの最適配置図(適応度CO)