

# タクシープローブデータを利用した 旅行時間の予測

三浦 英俊

走行中に収集した情報を発信するシステムを備えたタクシーを「プローブタクシー」という。プローブタクシーがさまざまなところを走行すれば、タクシーから交通情報をリアルタイムで入手することができる。プローブタクシーから得られるデータの利用は将来さまざまな形で展開することが期待される。データの利用の1つとして、任意の2地点間の旅行時間予測について紹介する。旅行時間予測は道路リンクごとに付与された速度を使ってダイクストラ法など最短経路探索法を利用することが一般的であるが、ここではクリギングを使った手法について紹介する。

キーワード：プローブカー、旅行時間予測、クリギング

## 1. プローブカー

昔のスパイ映画などを見ると、ときおり自動車に発信器を仕掛けてその足取りを追う場面があった。

自動車の位置情報を収集・発信するシステムは、すでにスクリーンの世界を飛び出して、プローブカーシステム（プローブ（probe）、探査）あるいはフローティングカーシステムと呼ばれている。プローブカーはGPS（Global Positioning System、衛星を利用して位置を調べる装置）を利用して自分の位置を測定する。あわせて速度、加速度、ワイパーやABS（Antilock Brake System、タイヤ滑走防止装置）などのシステム稼働状態を計測してデータを送信する。このデータをプローブデータという。プローブデータを収集すれば、リアルタイムで道路混雑や天候について知ることが可能となる。収集されたデータから得られる有用な情報は、再びプローブカーへ、あるいは情報を必要とするところへ提供される。天気や渋滞情報はもちろんのこと、ABSの動作情報から冬季の道路凍結情報、エンジンやタイヤ情報からは自車のトラブルや適切なメンテナンス情報をフィードバックして得ることができる。

プローブカーは、移动通信を利用した自動車関連情報の送受信によって新しいサービスを提供するための基盤技術と位置づけられる。例えば、これまでは交通

情報は道路に設置された機器を利用して収集する方法が一般的であった。この方法だと機器が設置されていない道路については情報を得ることができなかった。プローブカーシステムの登場によって地域の道路情報を網羅的に取得できるようになった。

## 2. プローブデータを利用したサービス

プローブデータを利用した先駆的なサービスと実験についていくつか紹介する。

本田技研工業株式会社は、インターナビ・プレミアムクラブというサービスを提供している[4]。ホンダのカーナビゲーションユーザーを対象として、道路情報やメンテナンス情報などを提供するサービスである。クラブ会員の自動車はそれぞれプローブカー（インターナビ・プレミアムクラブでは「フローティングカー」と呼ばれる）となっており、走行データを共有することによって、詳細な交通情報を入手することができる。

株式会社ユビークリンクは、携帯電話を利用したナビゲーションシステムで、1万台以上のタクシーやケータイカーナビから収集したプローブデータを利用した交通情報を提供している[2]。

中国では、2008年の北京オリンピックにあたって、市内の渋滞や大気汚染の緩和、エネルギーの効率的利用を目的としてタクシープローブシステムを試験的に導入した。約6万6千台のタクシーへGPSの装着を行ってデータを収集し、ナビゲーションシステムや電光掲示板により情報の提供などの実験を行った実績がある[3]。

北京に先立つこと6年前の2002年に、わが国ではインターネット ITS 共同研究グループ（慶應義塾大学 SFC 研究所，トヨタ自動車株式会社，株式会社デンソー，日本電気株式会社）によって名古屋地区においてタクシーを利用したプローブカーシステムを含む将来の高度交通システムの実証実験が行われた。約1,500台（当時の名古屋市内のタクシーの2割）のタクシーを利用し，プローブデータの収集およびサービス提供の実験が実施された[7]。本稿で述べる旅行時間推定は，このときの名古屋のタクシーデータを使っている。

### 3. 名古屋のタクシープローブデータ

名古屋のタクシープローブデータについて紹介する。

図1は，2002年11月1日から30日までの1カ月間に，名古屋市役所前交差点（北緯35.18197度，東経136.90544度）を中心とする半径100mの円内から客を乗せて出発し，名古屋駅桜通口（北緯35.17078度，東経136.88401度）を中心とする半径100mの円内を目的地として走行した17本のタクシーの経路を図示したものである。プローブタクシーは，距離300mごとに，発進または停止時に，などいく

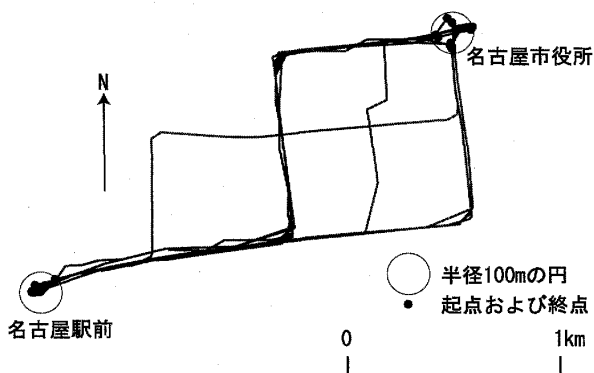


図1 名古屋市役所前から名古屋駅前までの17本の経路

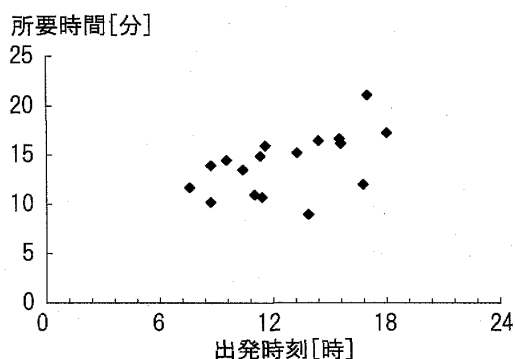


図2 出発時刻と所要時間

つかの状態に応じてデータを送信する仕掛けになっている[7]。ほとんどの場合数秒，長い場合には3分程度の間隔で送信が行われている。送信データには先に述べた位置や速度に加えて，空車/客扱いの状態も含まれているから，客を乗せて出発した地点と降ろした地点を知ることができる。

図1の2つの円の中心間直線距離（大圏距離）は2.31 km，17の経路距離は2.85 kmから3.06 km，起終点の直線距離に対する経路距離の比はおおよそ1.3倍である。図2はタクシーが客を乗せて出発した時刻と目的地までの所要時間の関係を表す。所要時間は最短9.0分，最長21.1分，平均14.1分でかなりのばらつきがあること，また午後の方が所要時間が長くなる傾向にあることが読み取れる。

### 4. クリギングを利用した旅行時間予測

クリギングという手法をタクシープローブデータに適用し，旅行時間を予測する方法について述べる。クリギングとは，空間データを対象として任意の地点の値を予測する地理統計手法である。対象とするデータに空間的な相関関係があるとき，つまり「データ間の距離が近いほど似た値となる」ようであれば，クリギングを適用して予測値を計算することができる。気象，資源，地質，環境分野など幅広く利用されている。

クリギングは未知データの予測値を既知の観測データ値の加重平均によって計算する。観測データに与える重みは，予測値の誤差の二乗の期待値を最も小さくするものが与えられる。未知データに距離が近い観測データほど大きな重みとなる。

クリギングを使ってどうやって旅行時間を予測するか？ 自動車の旅行時間に関する2つの性質：

1. 同じ時間帯に出発し，起点同士が近く，終点同士も近いならば，旅行時間は同じくらいになる。
2. 旅行時間は起点終点間の直線距離におおよそ比例する。

を利用して，クリギングの手法の1つである「普遍クリギング」を使う。まず，これら2つの性質が本当なのかどうか確かめる。

自動車の起点および終点を構成する4つの平面座標要素と出発時刻を軸とする5次元空間を考えれば，その中の点はある時刻にある場所を出発してから別の場所へのトリップ（移動）を表現していることになる。5次元空間内に相互の時空間距離が近い点群があれば，それらの起点同士，終点同士の距離は近く，出発時刻

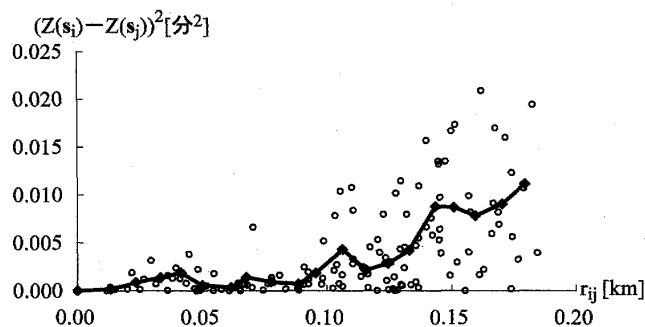


図3 17トリップの時空間距離と旅行時間の差の関係

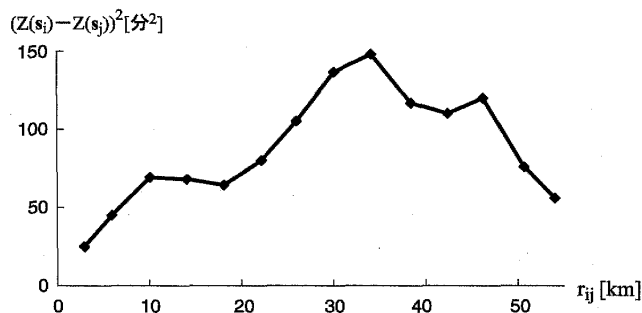


図4 1,000トリップについての時空間距離と旅行時間の差の関係

も似かよっている。そのような点群が5次元時空間距離が近いほど旅行時間の差が小さいならば、第1の性質が成り立っていることになる。第2の性質は、トリップごとに直線距離と旅行時間の散布図を描き、相関係数を調べることによって検討することができる。

3節で述べた事例で第1の性質を検証する。17の起終点ペアのうち第*i*番目のトリップの起点座標、終点座標をそれぞれ $(x_{i1}, y_{i1})$ ,  $(x_{i2}, y_{i2})$ , 出発時刻を $t_i$ として、これらをまとめたトリップベクトルを $\mathbf{s}_i = (x_{i1}, y_{i1}, x_{i2}, y_{i2}, t_i)$ とする。また起終点ペアの旅行時間を $Z(\mathbf{s}_i)$ とする。

2つの起終点ペア*i*, *j*の時空間距離 $r_{ij}$ を次のようにユークリッドノルムで測ることとしよう。

$$r_{ij} = \{(x_{i1} - x_{j1})^2 + (y_{i1} - y_{j1})^2 + (x_{i2} - x_{j2})^2 + (y_{i2} - y_{j2})^2 + \alpha^2(t_i - t_j)^2\}^{1/2} \quad (1)$$

ただし、 $\alpha$ は時間を距離尺度に変換するときの定数である。

図3は起終点ペア*i*, *j*の旅行時間の差の2乗 $(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ と5次元ユークリッドノルム $r_{ij}$ の関係を示す。 $\alpha$ は0.01[km/時]とした。1時間の出発時刻の違いが距離0.01 kmつまり10 mとして換算される。折線は $r_{ij}$ を20の区間に分割して、それぞれの区間において $(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ の区間別平均を求めてつないだものである。 $r_{ij}$ が大きくなるにつれて $(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ は大きくなる。すなわち、起終点の位置および出発時刻の違いが大きい起終点ペアほど旅行時間の差が大きい。

## 5. クリギングを用いた旅行時間予測

事例として扱ってきた17トリップは、起終点の位置がそれぞれ100 m以内ときわめて近い位置にあるトリップ群であった。起終点間の直線距離がほぼ等しいので第2の性質を調べるには適当ではない。

そこで、より広い範囲から長短さまざまなトリップ

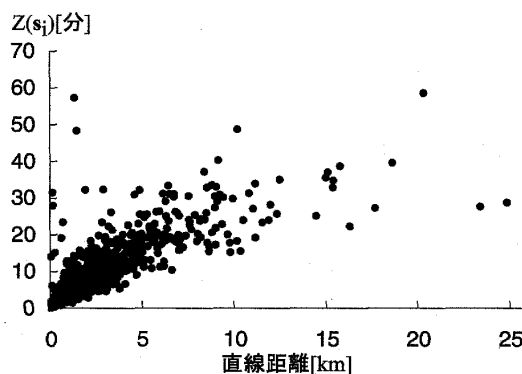


図5 直線距離と旅行時間

を取り出して2つの性質を検証する。図4は2002年11月1日に走行した名古屋のタクシープローブデータ27,654からランダムに選んだ1,000トリップに関する $(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ と $r_{ij}$ の関係を示す。折れ線は図3における区間別平均に相当する。起終点の位置が市内各地に広く分散しているため、図4は図3と比べて横軸縦軸とも大きな値を取っているが、全体の傾向はよく似ている。ただし、図4では $r_{ij}$ が30 kmを超えると $(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ が低下している。これらデータ間の5次元距離が遠いトリップは、起点から終点までの直線距離(大圏距離)が長距離となるものを多く含んでいる。長距離トリップは起終点間直線距離によって旅行時間はおおよそ説明できて、出発地目的地の違いはそれほど問題にならない。短距離トリップに比べて長距離トリップの旅行時間については起終点および終点の位置の影響が小さいことが、 $(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$ が低下する理由である。なお $r_{ij}$ が30 kmを超えるトリップペアは全トリップペア数 ${}_{1000}C_2 = 499,500$ のうち3%程度にすぎない。

図5は1,000トリップの直線距離と旅行時間の散布図である。いくつか短距離にもかかわらず時間を要しているトリップもあるが、おおよそ両者は比例関係に

あり、第2の性質を確かめるには十分であろう。相関係数は0.785である。

旅行時間を「普遍クリギング」によって予測する。普遍クリギングとはクリギングの手法の1つであり、予測値の期待値が対象空間内の地点別に異なっており、期待値が地点を変数とする関数によって記述できるときに適切な手法である。ここでは予測値すなわち旅行時間の期待値が直線距離の定数倍に比例することを利用する。

任意のトリップベクトル  $\mathbf{s}$  の旅行時間  $Z(\mathbf{s})$  が

$$Z(\mathbf{s}) = \beta f(\mathbf{s}) + \delta(\mathbf{s}) \quad (2)$$

という構造を持つと仮定する。  $f(\mathbf{s})$  を起点終点間の直線距離とする。  $\beta$  は比例定数である。  $\delta(\mathbf{s})$  は  $\mathbf{E}(\cdot)$  を平均、  $\text{Cov}(\cdot, \cdot)$  を共分散、  $\mathbf{h}$  を任意の5次元空間ベクトルとして、次の性質を仮定する。

$$\mathbf{E}(\delta(\mathbf{s})) = 0, \quad (3)$$

$$\text{Cov}(\delta(\mathbf{s} + \mathbf{h}), \delta(\mathbf{s})) = C(\mathbf{h}). \quad (4)$$

ここで、  $C(\mathbf{h})$  はコバリオグラムと呼ばれる関数である。(4)は、5次元時空間ベクトルで  $\mathbf{h}$  だけ異なる2つのトリップの  $\delta$  の共分散が  $\mathbf{h}$  を変数とする関数によって記述できることを表している。  $\mathbf{E}(\delta(\mathbf{s})) = \text{定数}$ 、および(4)が成り立つとき、  $\delta(\mathbf{s})$  は二次定常性を満たすという。コバリオグラム  $C(\mathbf{s})$  には使用する観測データに当てはまる適切な関数形を与えなければならない[1][5]。

トリップ  $\mathbf{s}_0$  の旅行時間  $Z(\mathbf{s}_0)$  が未知であるとき、クリギングを利用した予測値を  $\hat{Z}(\mathbf{s}_0)$  とする。ここで  $n$  個の観測データが利用可能であるとき、それらのトリップベクトルを  $\mathbf{s}_i (i=1, \dots, n)$ 、旅行時間データを  $Z(\mathbf{s}_i)$  とすれば、  $\hat{Z}(\mathbf{s}_0)$  は観測データ値の重み付き平均

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i z(\mathbf{s}_i) \quad (5)$$

によって推定される。

重みパラメータ  $\lambda_i$  は、推定誤差二乗の期待値を最小にする次の最適化問題を解くことによって得られる。

$$\min_{\lambda_i} \mathbf{E}((Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0))^2) \quad (6)$$

$$= C(\mathbf{0}) - 2 \sum_{i=1}^n \lambda_i C(\mathbf{s}_0 - \mathbf{s}_i) + \sum_{i,j} \lambda_i \lambda_j C(\mathbf{s}_i - \mathbf{s}_j),$$

$$\text{s. t. } \sum_{i=1}^n \lambda_i f(\mathbf{s}_i) = f(\mathbf{s}_0). \quad (7)$$

ここで(7)は推定値が不偏すなわち  $\mathbf{E}(\hat{Z}(\mathbf{s}_0)) = \beta f(\mathbf{s}_0)$  であるための条件である。

## 6. 旅行時間予測結果の比較

実際に旅行時間の予測を行い、実データと比較する[6]。2002年12月6日(金)に計測されたタクシープローブデータのうちトリップ大圏距離5km未満、5km以上10km未満、10km以上の3カテゴリから100ずつランダムに取り出した300トリップ  $\mathbf{u}_i$  の旅行時間の予測値  $\hat{Z}(\mathbf{u}_i)$  を計算する。300トリップそれぞれについて、トリップ  $\mathbf{u}_i$  からの5次元時空間距離が近い500個の観測トリップデータを  $Z(\mathbf{s}_i)$  として使う。  $\hat{Z}(\mathbf{u}_i)$  と「正解」である旅行時間実データ  $Z(\mathbf{u}_i)$  を比較する。さらに同じ300トリップデータをウェブサービスとして広く利用されている2つの経路探索サービス、Google マップルート・乗換案内、ナビタイム車ルート検索と比較する。図6は、それぞれクリギ

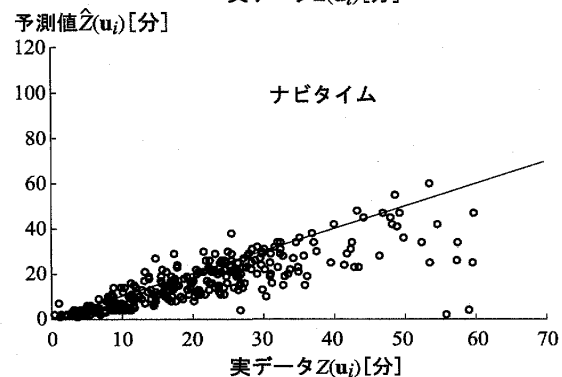
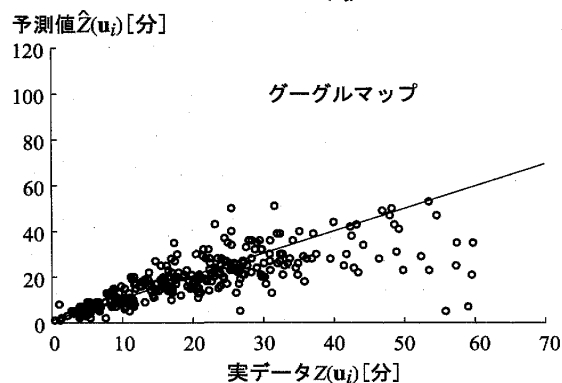
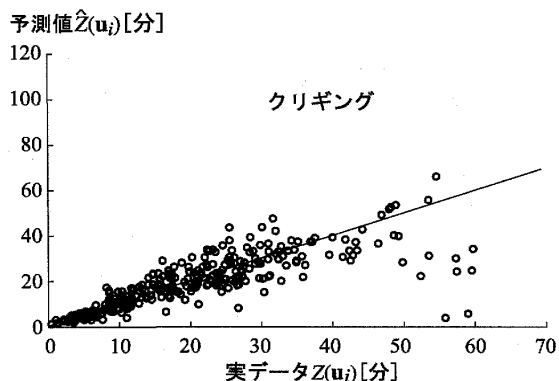


図6 旅行時間予測結果の比較

ング、グーグルマップ、ナビタイムによる旅行時間予測結果を比較するために、実データと予測値のプロットを散布図で示したものである。実データと予測値の誤差二乗平均はクリギング 63.2[分<sup>2</sup>]、グーグルマップ 75.8[分<sup>2</sup>]、ナビタイム 77.4[分<sup>2</sup>]である。わずかながらクリギングの予測精度が最も良い。ただし、グーグルマップとナビタイムが予測に使用したデータは名古屋のタクシープローブデータではない。名古屋のタクシープローブデータと合わせて予測すれば、両サービスともさらに良い予測を行うことができたと考えられる。

図6の3つの図には、実データ50分以上のところに予測のあてはまりが悪いトリップをいくつか見ることができる。これらは移動距離が短く、何らかの理由で旅行時間が長時間となったトリップであり、あてはまりが悪いのはやむを得ない。ランダムに選んだ300トリップに紛れ込んだものであり、除外すべきであったかもしれない。同じように、観測データにも除外した方が適切なデータが入っていた可能性がある。実際にクリギングを利用するには、適切な観測データを選んでおくことも必要となるだろう。

## 7. おわりに

本稿ではタクシーを利用したプローブデータについて紹介し、データの旅行時間予測への利用について述べた。タクシーはプローブカーとなって各地の情報を収集発信し、道路の混雑緩和や交通安全に寄与する新しい公共的な役割を担うことになる。

旅行時間の予測は、ダイクストラ法に代表される経路探索法によって行われることが多い。経路探索法は

よく知られているようにネットワークモデルを利用し、リンク別に与えられる速度が大きな役割を果たす。タクシープローブデータからリンクごとの速度データを得る試みはすでにさまざまな形で行われており、リンク別速度を利用した経路探索法による旅行時間予測でも良い結果が得られている[7]。クリギングはネットワークモデルの代わりに高次元空間におけるデータの空間的相関を手がかりとする。時空間を使って人や自動車の移動や流動を考察することは、これまでと異なった角度から問題をとらえるやり方の1つであろう。

## 参考文献

- [1] Cressie, N. A. C., *Statistics for Spatial Data Revised Edition*, Wiley-Interscience (1993).
- [2] 株式会社ユビークリンク, 全力案内, <http://www.z-an.com/>, 2009年12月1日にアクセス確認。
- [3] 日経BP社, 北京市, 五輪に向けて交通渋滞対策 新システム「フローティングシステム」に期待, 日経ビジネスオンライン, 2008年7月31日付配信。
- [4] ホンダ技研工業株式会社, インターナビ・プレミアムクラブ, <http://www.premium-club.jp/index.html>, 2009年12月1日にアクセス確認。
- [5] 間瀬茂, 武田純, 空間データモデリング—空間統計学の応用, 第6章, 共立出版 (2001)。
- [6] 三浦英俊, 時空間クリギングとプローブタクシーデータを利用した都市内の自動車旅行時間推定, 都市計画論文集, 44-3, 793-798 (2009)。
- [7] 三輪富生, 境隆晃, 森川高行, 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シンポジウムプロシーディングス, 277-282 (2003)。