

時空間クリギングを用いた旅行時間予測

三浦英俊

Travel Time Prediction Using Space-Time Kriging

Hidetoshi MIURA

Abstract: This paper proposes a space-time kriging method to predict travel time with probe-vehicle data collected in Nagoya area. The prediction is developed in the five-dimensional space made with coordinates of start point and goal points, and start times. The result with 1000 observed probe-car trip data shows that proposed method compresses prediction error sum of squares less than 9 %.

keywords: クリギング (kriging) , プローブカーデータ (probe-vehicle data) , 旅行時間予測 (travel time prediction)

1 はじめに

本研究では、2002年に名古屋地区で収集されたタクシープローブデータに時空間クリギングを適用し、自動車による2地点間の旅行時間予測を試みる。クリギングとは、距離が近いほど似た値を示す性質をもつ空間データに対して、既知観測データの空間分布に基づいて未知データを推定する統計手法である。クリギングは鉱床の空間分布を推定する鉱床学の手法として発展し、空間分布を取り扱う分野で広く応用されている。経験的に出発地、目的地、時間帯が似た旅行時間が近い値になることが知られている。これを手がかりとして旅行時間予測にクリギングを適用することが可能である。

クリギングを用いた旅行時間予測手法はすでに三浦(2008)で提案されている。ただしこの研究はウェブサイトの提供する旅行時間予測データを使用して予測と結果の検証を行っているものであり、さらに移動が行われる時間は考慮されていない。本研究では実際の走行データとしてタクシープローブデータを使用し、旅行の出発時間を考慮した時空間クリギ

ングを実行する。

ダイクストラ法に代表される最短路探索手法を用いて旅行時間予測を行う場合、時間帯によって旅行時間が異なることを考慮するには、時間帯によって異なるリンク走行速度を与えればよい。リンク走行速度が時刻によって変化することを仮定した最短路問題は、“time-dependent shortest path problem”(時刻依存の最短路問題)として多くの既往研究がある(Chon et al., 2003)。

時空間クリギングによって同様に時刻に依存した予測が可能となる。出発時間が明示されたデータとしてタクシープローブデータを使う。タクシープローブデータとは、取り付けられた通信機を通じてタクシーから位置、速度、乗客の有無などの情報をリアルタイムで収集したデータである。収集したプローブデータを分析・活用する技術はすでに多くの提案がある。例えば時間帯ごとのリンク走行速度データを蓄積する手法はすでに実用化されている(三輪ほか, 2003)。詳細なリンク走行データは最短路検索手法を用いた旅行時間予測に大いに役立つ。一方クリギングはプローブデータによって得られる走行速度をリンク単位に分割せず、走行中の任意の2地点間の旅行時間データをそのまま予測に使用する。

*三浦英俊:明海大学不動産学部

〒279-8550 千葉県浦安市明海1-2, Tel. 047-355-5120
e-mail hmiura@meikai.ac.jp

2 使用データの概要

2002年11月1日に名古屋地区を走行した約1340台のタクシーから集めたタクシープローブデータを推定に使用する。このデータはトヨタ自動車、日本電気、デンソーによる「インターネットITSプロジェクト」で得た実験データの一部である。

乗客が乗車および出発してから目的地に到着して降車するまでを1トリップとする。乗車位置および降車位置が不明なデータ、乗車から降車までの走行距離が直線距離よりも短いなどエラーを含むデータは除外する。エラーのない23261トリップからさらに直線距離に対する走行距離の比が3倍以上の555データは目的地までの旅行時間推定にとって適当でないと判断して除外し、残る22706トリップを正常観測データとする。正常観測データのトリップ平均距離は直線距離で3.0km、走行距離で3.9kmである。平均速度は直線距離あたり16.6km/h、走行距離あたり20.9km/hである。時間帯別のデータ数は朝8~9時台が最も多く、次に夜間23時~0時台が多い(図1)。時間帯別の直線距離あたり平均速度は17時台が最も遅く13.2km/h、0時台が最速22.2km/hである。図2は使用した22897トリップの出発地の分布図である。大半のトリップは名古屋市内が乗車地点となっていることがわかる。

3 時空間クリギング

本節では普遍クリギングを時空間に適用した旅行時間予測手法について述べる。

予測対象地域を平面領域と仮定する。対象領域内の出発地座標を (x_1, y_1) 、目的地座標を (x_2, y_2) 、出発時刻を t としたとき、これらの変数によって構成される5次元ユークリッド空間の部分集合を D とする。任意の出発地・目的地・出発時刻の組 $s = (x_1, y_1, x_2, y_2, t) \in D$ について、出発地目的地間の旅行時間 $z(s)$ が確定項と確率項の和として

$$z(s) = \beta f(s) + \delta(s) \quad (1)$$

という構造を持つと仮定する。 β はパラメータであ

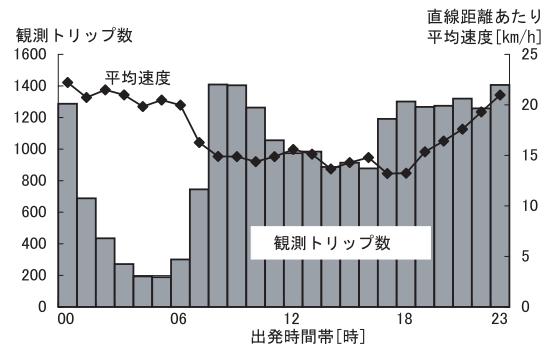


図1：正常観測データの時間帯別トリップ数と平均速度

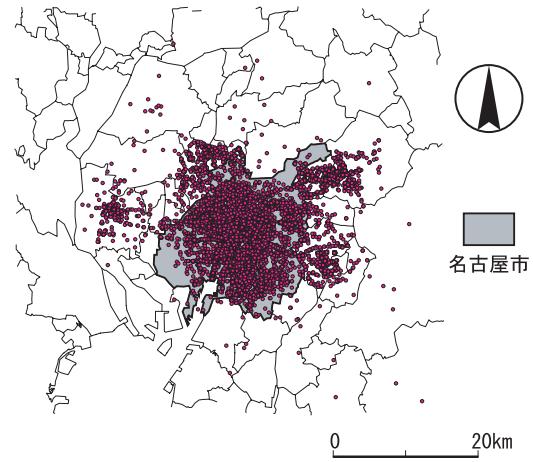


図2：正常観測データの出発地分布

る。 $f(s)$ は s に対応する出発地から目的地へその時刻に出発したときの平均旅行時間である。図1の折れ線グラフで示した出発時間帯別の直線距離あたり平均速度を $\bar{v}_t(t=0, \dots, 23)$ として、

$$f(s_i) = \frac{\text{出発地目的地間の大圏距離}}{\bar{v}_t}$$

とする。 $E(\cdot)$ 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、コバリオグラムを $C(\mathbf{h})$ として、 $\delta(s)$ は次の性質を満たすと仮定する。

$$E(\delta(s)) = 0, \quad (2)$$

$$\text{Cov}(\delta(s + \mathbf{h}), \delta(s)) = C(\mathbf{h}). \quad (3)$$

$E(\delta(s)) = \text{定数}$ 、および(3)が成り立つとき、 $\delta(s)$ は二次定常性を満たすという。これは任意の $s \in D$ について $\delta(s)$ の平均が一定であり、相対的位置が \mathbf{h} 離れたときの $\delta(s)$ と $\delta(s + \mathbf{h})$ の共分散がコバリオ

グラムと呼ばれる関数 $C(\mathbf{h})$ で記述できることを意味する。

コバリオグラムの関数型にはいくつかの定型が提案されている。本研究では、コバリオグラムが等方的(距離 $\|\mathbf{h}\|$ のみに依存する)であることを仮定し、指数型と呼ばれる次の関数形を使用する。

$$C(\mathbf{h}) = \theta_1 e^{-\|\mathbf{h}\|/\theta_2}. \quad (4)$$

θ_1, θ_2 はパラメータである。 $\theta_1 \geq 0, \theta_2 > 0$ とする。 x_1, y_1, x_2, y_2 の値は経度緯度を平面直角座標系(第VII系)に変換したものを使う。D内の任意の2点 $\mathbf{s}_k = (x_{k1}, y_{k1}, x_{k2}, y_{k2}, t_k), \mathbf{s}_l = (x_{l1}, y_{l1}, x_{l2}, y_{l2}, t_l)$ に対して5次元空間内の距離 $\|\mathbf{h}\| = \|\mathbf{s}_k - \mathbf{s}_l\|$ は空間軸に対する時間軸の比を α [時間/km](>0)として次のように定義する。ただし位置座標の単位はキロメートル、出発時刻の単位は時間とする。

$$\begin{aligned} \|\mathbf{h}\| = & \{(x_{k1} - x_{l1})^2 + (x_{k2} - x_{l2})^2 \\ & + (y_{k1} - y_{l1})^2 + (y_{k2} - y_{l2})^2 + (\alpha t_k - \alpha t_l)^2\}^{\frac{1}{2}}. \end{aligned}$$

以上の事柄がクリギング適用に必要な仮定である。なお、本研究では時空間クリギングのために提案されたコバリオグラム(例えば Gneiting(2002))を使用しなかった。これらの時空間分析用コバリオグラムは、一定時間間隔で集計された空間データつまりパネルデータを対象としたものであり、本研究で使用するデータには適当でないと考えたからである。

さて、 n 組の出発時間・出発地・目的地の組についての旅行時間観測データが得られているとしよう。第 i 番目のデータの出発地・目的地ベクトルを $\mathbf{s}_i \in D$ 、旅行時間を $z(\mathbf{s}_i)$ とする($i = 1, \dots, n$)。クリギングによる任意の $\mathbf{s}_0 \in D$ に関する旅行時間予測値 $\hat{Z}(\mathbf{s}_0)$ は、観測データ $z(\mathbf{s}_i)$ の重み付き平均

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

によって得られる。ここで重みパラメータ λ_i は $\sum_{i=1}^n \lambda_i = 1$ を満たすとする(ただし λ_i は正値とは限らない)。パラメータベクトル $\lambda = (\lambda_1, \dots, \lambda_n)'$ は

平均二乗予測誤差

$$\begin{aligned} \sigma^2(\mathbf{s}_0) &= E((z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0))^2) \\ &= C(\mathbf{0}) + \mathbf{\lambda}' \Sigma \mathbf{\lambda} - 2\mathbf{\lambda}' \mathbf{c} \end{aligned} \quad (6)$$

を最小にする値として求められる。ここで $\Sigma = \{C(\mathbf{s}_i - \mathbf{s}_j)\}_{ij}$, $\mathbf{c} = (C(\mathbf{s}_0 - \mathbf{s}_1), \dots, C(\mathbf{s}_0 - \mathbf{s}_n))'$ と置いた。(6)にラグランジュ乗数 m を導入して、最小化すべき目的関数を

$$\phi(\mathbf{\lambda}, m) = C(\mathbf{0}) + \mathbf{\lambda}' \Sigma \mathbf{\lambda} - 2\mathbf{\lambda}' \mathbf{c} - 2m \left\{ \sum_{i=1}^n \lambda_i f(\mathbf{s}_i) - f(\mathbf{s}_0) \right\}$$

とする。 $\phi(\mathbf{\lambda}, m)$ を $\mathbf{\lambda}, m$ に関して偏微分して得られる正規方程式を解いたならば、

$$\begin{aligned} \mathbf{\lambda} &= \{\mathbf{c} + \mathbf{X}' \Sigma^{-1} \mathbf{X}\}^{-1} (f(\mathbf{s}) - \mathbf{X}' \Sigma^{-1} \mathbf{c})' \Sigma^{-1}, \quad (7) \\ m &= -(f(\mathbf{s}) - \mathbf{X}' \Sigma^{-1} \mathbf{c})' (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \end{aligned} \quad (8)$$

を得る。ただし $\mathbf{X} = \{f(\mathbf{s}_1), \dots, f(\mathbf{s}_n)\}'$ とする。 $\mathbf{z} = \{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}'$ として(7)を(5)に代入すると予測値 $\hat{Z}(\mathbf{s})$ の計算式を得る。

$$\begin{aligned} \hat{Z}(\mathbf{s}) &= \sum_{i=1}^n \lambda_i z(\mathbf{s}_i) = \mathbf{\lambda} \mathbf{z} \\ &= \hat{\beta}_{gls} f(\mathbf{s}) + \mathbf{c}' \Sigma^{-1} (\mathbf{z} - \hat{\beta}_{gls} \mathbf{X}) . \end{aligned} \quad (9)$$

ただし

$$\hat{\beta}_{gls} = (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{z} \quad (10)$$

は β の一般化最小二乗推定量である。

予測値の計算に必要な(1)のパラメータ β と(4)のパラメータ θ_1, θ_2 は制限付き最尤推定法によって推定する。

4 パラメータの推定

正常観測データ 22706 トリップからランダムに選んだ 1000 トリップを観測データとして、(1)のパラメータ β , (4)のコバリオグラム $C(\mathbf{h})$ の θ_1 と

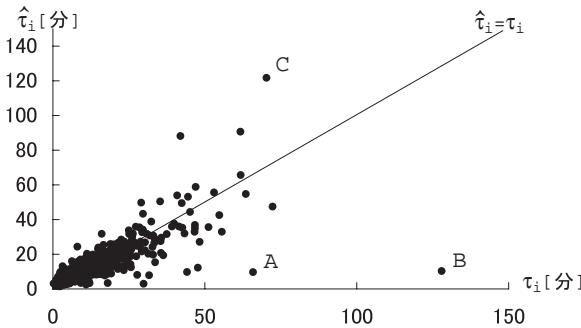


図 3: 実旅行時間と推定値の散布図

θ_2 , および α を推定する . $\delta(s)$ が正規分布に従うとする仮定する . α を 0 から 0.1 刻みで変化させつつ , 制限付き最尤推定法によって θ_1, θ_2 を推定する (Cressie,1993) . 次に (10) によって得られる β_{gls} を β の推定値とする .

得られたパラメータ推定値は $\alpha = 2.5$ [時間/km] , $\theta_1 = 49.9, \theta_2 = 4.10$ $\beta_{gls} = 0.6996$ である .

5 旅行時間の推定

観測データ収集の 1 週間後 11 月 8 日の正常トリップの 21256 データからランダムに選んだ 1000 トリップを検証データとする . 前節で推定したパラメータを用いて検証データの旅行時間推定値 $\hat{\tau}_i$ を計算し , 検証データの実際の旅行時間 τ_i と比較する ($i = 1, \dots, 1000$) . 図 3 は横軸に実際の旅行時間 , 縦軸に推定値を取ったものである . 実データ旅行時間 30 分以内の推定精度はそれほど悪くないが , 50 分以上のものはばらつきが非常に大きい . これには 3 つの理由が考えられる . (1) 観測データに 50 分以上の旅行時間であるものが少ない . (2) 実旅行時間の長いトリップは途中で (あるいは乗車地点降車地点いずれかで) 長時間の停止時間を含んでいるものがある (図 3 の A,B など) . (3) 高速道路で長距離移動したトリップは移動速度が速いので一般道利用トリップを多く含む観測データを用いた推定は精度が悪い (図 3 の C) .

本研究で使用したデータは前述したように観測用・検証用とともに道路距離 / 直線距離の比が 3 以上

の迂回が大きかったものを除外したのみである . より精密な旅行時間推定および予測のためには , 長時間停止を含むトリップは取り除く , 高速道を利用したトリップデータは高速道を利用する場合の旅行時間予測に限定して使用するなどのデータ吟味が必要である .

6 おわりに

検証データ推定値の誤差二乗和 $\sum_i (\tau_i - \hat{\tau}_i)^2$ は 42179.87 である . 出発時刻データを使用しない空間クリギングの誤差二乗和 48652.33 と比べて 9 % 程度減少した .

今後は天候別 , 一般道利用・高速道路利用などデータを分類して時空間クリギングの有効性を確かめるとともに , より適切なコバリオグラム関数について考察したい .

参考文献

- Chon, H. D., Agrawal,D. and Abbadi, A. E.(2003) *FATES: Finding A Time dEpendent Shortest path*, Berlin: Springer .
- Cressie, N.(1993) *Statistics for Spatial Data. Revised Edition*, John Wiley & Sons, 29–170.
- Gneiting, T.(2002) Nonseparable, Stationary Covariance Functions for Space-Time Data, *Journal of the American Statistical Association*, 97, 458, 590-600.
- 三浦英俊 (2008) 普遍クリギングを用いた最短旅行時間予測 . オペレーションズ・リサーチ , Vol.53, No.3 , pp. 173-178.
- 三輪 富生 , 境 隆晃 , 森川 高行 (2003) 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 .
- 三輪 富生 , 森川 高行 (2004) プローブカーデータを用いたドライバーの選択経路データの収集と分析 . 交通工学 , Vol.39, No.2 , pp. 49-56 .